

Design of a Percussion Robot with Dynamic Timbre-Production

Author Names Omitted for Anonymous Review. Paper-ID [add your ID here]

Abstract—This paper introduces the design strategy behind a percussion robot named Kiki that can dynamically produce a range of timbres. It focuses on the physical characteristics of striking mechanisms that influence timbre, and the dynamics and kinematics of the way humans and robots strike drums. This is presented as a case-study in the approach to timbre in musical robots.

I. INTRODUCTION AND PREVIOUS WORK

Attempts to automate percussion playing date back at least to the Islamic Golden Age [3][2][1]. In these automata, a striking mechanism is driven into a drum via a complex system of levers, water-wheels, and gravity. However, due to limitations in the mechanics, the striker always falls upon the drum in the same way, producing the same timbre. By contrast, the dynamic manipulation of timbre is central to how humans play drums; for example, the famous djembefola Famoudou Konaté reports being able to produce approximately twenty-five distinct timbres [15]. In fact, the word ‘timbre’ comes from the Greek word from ‘drum’, and for many drums, especially hand drums, timbre is the primary parameter that the player manipulates while playing (as opposed to pitch for the majority of orchestral instruments). In recent decades, many more percussion robots and automata have been built [24][9][23][26][11][20][27][21][18][5][10][4]. Although levers, cables, and gravity have been replaced by solenoids, servos, and electricity, the majority of these still drive fixed-position drumsticks or mallets into a drum with little or no control over timbre. One notable exception is Haile [26], one of whose strikers can move along the drum’s radial axis, striking it in different locations. Prior to the addition of this capability, the authors reported that “the main mechanical caveats mentioned were Haile’s limited timbre and volume control” [25]; based on the analysis below it is not likely that the addition of this one extra degree of freedom greatly increased the timbral range. Another notable exception is MIT’s Cog, which, when outfitted with special arms with compliant actuators [27], was able to exploit the arms’ natural dynamics in striking a snare drum, perhaps modeling more closely the way humans strike snare. However, the timbre of sound produced by this method was not within the purview of the study and was not assessed. To the best of the author’s knowledge, no percussion robot has been designed specifically with the aim of dynamically producing a variety of timbres that are similar to those produced by human players. Therefore, the author has built a djembe-playing robot named Kiki (shown in Figure 1), specifically with this goal in mind. Here, the thought-process behind Kiki’s design is presented as a case-



Fig. 1. Kiki, a robotic percussionist with dynamic timbre-production capabilities.

study in how timbre might be approached more generally in musical robots. The first half of this paper will focus on the material properties of the striking mechanism which influence timbre, including the solution eventually used in Kiki. The second half shall focus on the dynamics and kinematics of actuating the striking mechanism, given the particular challenges associated with dynamic timbre production.

II. STRIKING MECHANISMS

A. Djembe Strokes

Insofar as the goal is to produce human-like timbres, it is fruitful to take a biomimetic approach, and examine how humans achieve different timbres. As previously mentioned, a



Fig. 2. Mamady Keita demonstrating (a) bass and (b) tone strokes. Source: [13]

skilled djembe soloist may produce a wide variety of distinct timbres; certainly this repertoire could be expanded even further if the possibility of striking the drum with arbitrary objects were included. However, djembe accompaniment technique comprises three core strokes with aurally distinct timbres: bass, tone, and slap. Players typically evaluate the sound of a djembe by evaluating these three strokes [22]. It is therefore justifiable to focus on building a machine that can reproduce the timbre of these strokes, although the goal is to do so *dynamically* (i.e. rather than using three fixed beaters) so that the machine will be capable of playing other intermediate timbres and searching its timbre space for particular sounds. Below are descriptions of the striking technique for each stroke, which are informed both by the descriptions given in [15], an analysis of a video of the djembefola Mamady Keita demonstrating the strokes [13], and on the author's own experience playing the instrument.

1) *Bass*: The center of the drum is struck with an open hand, as is illustrated in Figure 2(a). The fingers are slightly hyperextended so that the pressure is concentrated on the palm. The precise location of impact depends on the precise shape of the drum, the speed of the passage being played, and the player's preference, but the overall goal appears to be to excite the head in its first radial normal mode. The result is a deep, resonant, sustained bass sound with few audible higher partials.

2) *Tone*: Many verbal descriptions of djembe technique often indicate that tone is played by striking the drum such that palmar-digital crease falls upon the rim of the drum. In the video used for this analysis, however, the player's entire hand is shifted more towards the center of the drum, such that the medial extremity of his proximal palmar crease is clearly seen in contact with the rim. This is shown in Figure 2(b). The four fingers are held straight and somewhat rigidly. In the video, the palmar crease is seen contacting the rim slightly before the the fingers contact the head. The four fingers are parallel to the head at the moment of impact and strike it with uniform pressure across their length.

3) *Slap*: The hand strikes the drum at a slight angle, with the fingers held loosely in a slightly curved position. In slow motion, the stroke has two discrete components, illustrated in Figure 3. First, the distal palmar crease and the medial extremity of the proximal palmar crease come into contact with the rim. Second, inertia causes the fingers to bend about the metacarpophalangeal joint so that, a very short time later,

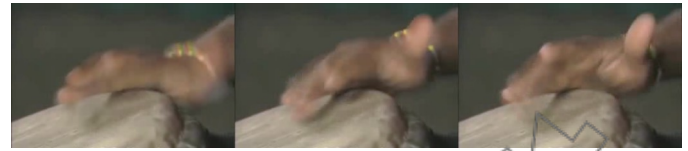


Fig. 3. Three consecutive frames of Mamady Keita demonstrating slap stroke. They show the frame just prior to impact, at the moment impact, and immediately after. Source: [13]

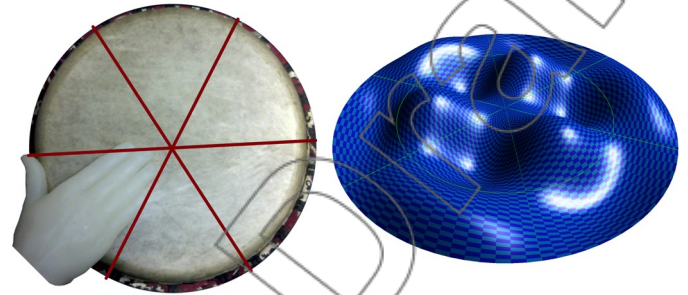


Fig. 4. Normal modes of a vibrating circular membrane, showing how the shape of the beater might influence the resulting timbre

the fingertips and only the fingertips contact the drum head. Video analysis reveals the palm touching the rim in one frame, and the fingertips have just rebounded by the next frame, so the interval is on the order of about 40 milliseconds.

B. Factors Influencing Timbre

Percussion robots are often built to be capable of striking the drum at several radial distances from the center of the head [23][12][25]. However, the foregoing analysis makes it clear that the strokes in question differ by more than just the impact location. Below is a discussion of some other factors that may contribute to the distinct timbre of each stroke in.

1) *Hand Rigidity*: The hyperextension of the fingers during bass indicates that the fingers are rigid, while during slap the fingers must be loose so that they may be under the control of inertia just before and after impact. During tone, the fingers appear to be have an intermediate rigidity. This suggests that the a robotic beater could benefit from variable rigidity.

2) *Hand Morphology*: Fourier analysis of the time evolution of vibrating bodies in general makes it clear that the shape of the perturbation that set the body in motion plays a large role in the frequency content of the resulting sound. The same principals hold for vibrating membranes as for strings. A circular membrane with a zero-displacement boundary condition about its circumference has two types of nodal lines, which define the normal modes [7]. The first type forms concentric circles about the center of the drum, and the height of any concentric circle varies sinusoidally. The other type of nodal line runs radially outwards from the center of the head. The height of any other radial line forms a Bessel function of the first kind with a zero-crossing falling on the drum's boundary. This is illustrated in Figure 4.

The part of the human hand used in the tone stroke is roughly wedge shaped, and roughly four inches in breadth

at its base. A typical djembe (including the one used in this study) has a circumference of about forty inches. The hand therefore fits comfortably into a radial sector of the head that is about a sixth of its total surface area. One may therefore hypothesize that the sixth radial normal mode plays a prominent role in the sound of the tone. If the hand were rectangular, so that the index, ring and pinkie finger were equal in length to the middle finger, and the fingers were not tapered, and so forth, then the extreme end of the hand would cross the nodal lines of the sixth normal mode, thereby suppressing it. The hand, in turn must suppress even higher radial normal modes, which an overall narrower object would not. Likewise, the palm of the hand is roughly round, and roughly a quarter of the diameter of the head. During the bass stroke it tends to push the head downward roughly into the shape of a parabolic dish, exciting the first and possibly second concentric normal modes while suppressing the radial normal modes and higher order concentric modes. A much smaller circular object would allow higher concentric modes to sound. Any fat, rigid object with corners would exert more force at the corners than its center as the head deforms downward, which would tend to excite the radial normal modes. It thus follows that an object whose surface of contact is similar in shape to the hand will be better suited to recreating the timbre of the hand than an arbitrarily-shaped object. Note that the foregoing analysis focuses on the normal vibrational modes, which are steady-state solutions to the wave equation. However, because of the steep amplitude envelope of percussive sounds which focuses much of the sound's power in the first few milliseconds of vibration, transient solutions to the wave equation may play a large role in the perceptual qualities of the stroke. Although the transient motion of a circular elastic plate with a zero-displacement boundary condition along its circumference in response to loading on a radial sector has been studied [6], it is not clear what frequency content emerges from this motion. However, the nature of the transient motion will still be determined by the shape of the initial perturbation, and consequently the shape of the hand is likely to be important in determining the resulting sound.

C. Additional Considerations for Slap

Slap is often considered the most difficult of the three strokes for a beginning human player to produce, and has proven difficult to mechanize. We thus present some additional information that may be useful in this regard.

1) *Hand size:* Drum-to-hand size ratio is important for the production of slap. Sunkett makes the following observation.

This [slap] is not an easy sound to achieve on every drum, and the ability to do so is often related to the diameter of the drumhead and the size of the player's hands. If you have small hands, the drumhead diameter does not have to be very large to achieve the sound without too much effort. Larger hands require larger head diameters... There are perceivable frequency differences in the resultant sounds. The highest overtones used to produce a

dynamic slap are most easily activated near the edge of the drum [22].

Presumably the drum to hand ratio must be large so that the hand can excite the higher radial normal modes while suppressing the lower ones, and the edge of the drum is used to excite the higher concentric normal modes.

2) *Open Fingers:* Beginners are sometimes taught to play tone with the fingers together, and slap with the fingers apart. This artifice is perhaps designed to regulate the flexibility of the fingers, taken as a single unit. Keita reports that although he teaches the strokes this way, he plays both strokes with his fingers slightly apart [14].

3) *Sticks in Sabar:* In Sabar ensembles of Northern Senegal, a variety of open-bottom drums are played which are roughly similar, in limited respects, to djembe. A consideration of their technique lends insight into how a the slap sound on a djembe may be mechanized. (The following discussion is the result of personal correspondence with the late Dr. Mark Sunkett). In contrast to the Djembe, which is played with two bare hands, the drums of the Sabar ensemble are played with one bare hand (traditionally the left) and one stick, known in the Wolof language as 'galan', held in the other (right) hand. The bare hand generally plays the three strokes associated with djembe, using similar technique, while the stick typically plays only one stroke. Anecdotally, native players report that the sound of the slap played by the bare hand should sound identical to the sound produced by the stick. The stick is made either of Tamarind, which is a hardwood of the Leguminosae family, or an indigenous wood, called 'sump' in Wolof, which is somewhat softer and more flexible. The stick is typically about sixteen inches long and very roughly 3/8 inch in diameter, although the ideal diameter varies somewhat proportionally to the size of the drum being played. The stick is prepared for use by removing the bark and rounding the ends with a knife. Ideally, the stick is slightly bowed on one end and, while playing, the stick contacts the drum head along the convex edge of the bowed segment. The stick is held loosely in the hand, oriented perpendicular to the fingers. It is actuated by rotating the forearm about the roll axis, so that the stick moves similar to a windshield-wiper. This arrangement is certainly mechanizable, although subsequent analysis reveals that it may not be dynamic.

D. Timbral Evaluation

The goal of the foregoing discussion was to consider what factors might contribute to an object's timbre when used as a striking mechanism, and in particular, what objects might sound most like the hand or produce the greatest range of timbres when striking the drum. In order to provide greater insight into how these factors might influence the design of such a beater, timbral evaluation of several objects was carried out. The purpose of this study was exploratory, and no hypothesis is proposed.

1) *Methodology:* In this study, various objects, including human hands, were used to strike the drum in various ways. The resultant sounds were recorded and compared against each



Fig. 5. Some of the materials used in timbral assessment of striking mechanisms

other. This was carried out as follows: A particular object and method of striking the drum with it were casually identified as being worthy of analysis by the author on account of the foregoing discussion; Several recordings were made of the object striking the drum in a particular way; All recordings were made during the same recording session, with the same placement of microphones and drum, so as to control for the placement of microphone, and acoustics of the room; The recordings were edited such that the first sample in the file corresponds to the zero-crossing marking the onset of the sound; The recordings were then analyzed and compared.

2) *Materials Used*: A variety of objects of different materials were tested during this study. They included various drumsticks, mallets, pieces of foam, rubber, cork, and linoleum. They were at times used alone, and at times mounted to a flat or convex wood or rubber block. Only a subset of these objects are reported here. The materials reported are a hickory drumstick, a sheet of 1/4-inch black rubber cut roughly to the outline of a human hand, and a large piece of foam rubber in the shape of a fist. These objects are depicted in Figure 5. They were used to strike three locations on the drumhead, corresponding roughly the three strokes under consideration. The locations were the center of the drum, the ‘edge’ of the drum (approximately three inches from the boundary of the drumhead) and the ‘rim’ (approximately one inch from the boundary).

3) *Centroid*: Machine representations of timbre which correspond to perception are an area of ongoing research. One seminal study [16] found that humans rate timbral similarity according to three dimensions, corresponding roughly to attack quality (explosiveness), the temporal evolution of spectral components, and brightness (strength of higher partials). Djembe players almost universally describe bass, tone, and slap as being low, mid, and high, respectively, suggesting a difference in perceptual brightness. Furthermore, a drumhead’s normal modes are of course determined by its geometry; perturbing it in a particular way merely distributes the energy amongst those modes in a particular way. Spectral analysis as can be seen in Figure 6 confirms this, indicating that for bass, relatively little of the energy is in the higher partials. Tone is intermediate, and slap has relatively little energy in the lower partials. This suggests the use of spectral centroid (weighted average) [19] as a preliminary measure of timbral similarity, which has also been shown to correlate with perceptual bright-

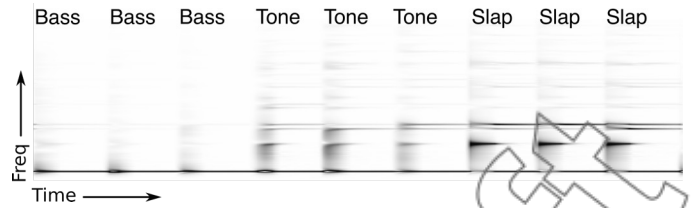


Fig. 6. Sonogram of djembe strokes played by human, showing different energy distributions for different strokes.

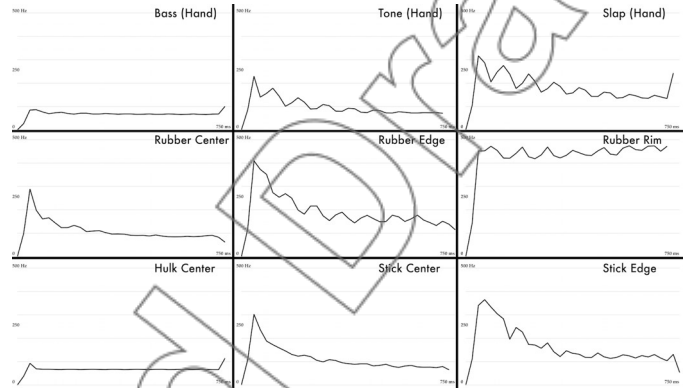


Fig. 7. Centroid plotted as a function of time for the first 750 ms of the strokes and objects considered in this study.

ness [8]. It is important to point out that this by itself would not be an appropriate way of comparing timbre across instruments because two sounds with dramatically different frequency distributions could produce the same centroid. In the case of the membranophones, however, it does not appear to be possible to control the frequency distribution independently of the centroid. Therefore centroid is used as follows.

Given a discrete signal x of a drum sound, x is separated into M windows $W = \{w_1, \dots, w_M\}$, each containing N consecutive samples and each successively translated in x by a hop-size of h samples. First, the Fourier Transform X of each window is computed. Then, the spectral centroid C for a given window $w \in W$ is the amplitude-weighted average of X across all frequencies ω .

$$X(w, \omega) = \sum_{n=0}^{N-1} w[n] e^{-j\omega \frac{2\pi n}{N}}; C(w) = \frac{\sum_{\omega=1}^N X(w, \omega) * \omega}{\sum_{\omega=1}^N X(w, \omega)}; \quad (1)$$

In particular this study uses a window size $N = 1024$ and a hop-size of $h = 512$ samples. For each stroke, only the first 750 milliseconds of audio after the onset were used because, although the drum still audibly resounds for some time beyond that, the signal-to-noise ratio becomes too low and the variance in the centroid becomes high. The sound of a particular stroke may have a certain amount of variability, as it cannot be performed identically each time. In order to address this, for each object and strike location, $C(w)$ is computed for three separate instances of the stroke and averaged point-wise over W . The results are plotted in Figure 7.

4) *Comparisons*: Given two drum sounds each separated into their respective windows W_1 and W_2 , the sounds are compared using the standard deviation σ of one with respect to the other over the windows w . Additionally, the base-2 logarithm and 12^{th} power of the centroid, \mathbb{C} , are used so that the result is expressed in semitones.

$$\sigma = \pm \sqrt{\frac{1}{M} \sum_{i=1}^M (\mathbb{C}(w_i \in W_1) - \mathbb{C}(w_i \in W_2))^2} \quad (2)$$

$$\mathbb{C}(w) = \log_2(C(w)^{12}) \quad (3)$$

Additionally, the positive or negative solution to the square root is chosen according to

$$\sigma = \begin{cases} + & \text{if } \sum_{i=1}^M \mathbb{C}(w_i \in W_1) > \sum_{i=1}^M \mathbb{C}(w_i \in W_2) \\ - & \text{otherwise} \end{cases}$$

This allows the metric to retain some information about which sound is perceptually higher. In this manner each sound is compared to each other sound. Comparing the average of three tone strokes to the average of three separate tone strokes yielded $\sigma = 0.9$. This was taken to be the resolution of measurement and all values were rounded to the nearest integer. The results are shown in Table I.

TABLE I
TIMBRAL COMPARISON OF STRIKING MATERIALS. TO DETERMINE THE SIGNS, ROWS WERE USED AS W_1 AND COLUMNS AS W_2

	Bass (Hand)	Tone (Hand)	Slap (Hand)	Rubber Center	Rubber Edge	Rubber Rim	Hulk Center	Stick Center	Stick Edge
Bass (Hand)	0	-8	-17	-9	-18	-32	2	-10	-16
Tone (Hand)	8	0	-11	-3	-13	-27	8	-3	-10
Slap (Hand)	17	11	0	9	-3	-16	17	10	4
Rubber Center	9	3	-9	0	-11	-25	10	-2	-17
Rubber Edge	18	13	3	11	0	-15	19	10	4
Rubber Rim	32	27	16	25	15	0	33	26	19
Hulk Center	-2	-8	-17	-10	-10	-33	0	-10	-16
Stick Center	10	3	-10	2	-10	-26	10	0	-8
Stick Edge	16	10	-4	17	-4	-19	16	+8	0

5) *Discussion*: There were far too many confounding variables in this study to make predictions about what timbre will be produced by a given object. In addition to the considerations in Section II-B, other properties such as mass, softness, coefficient of friction, impact velocity etc. . . appear to be important. Nonetheless, the aim of this study was only to provide the tools and a starting point for exploring those properties in greater detail. In any case, a few observations may be made regarding the above data. The stick is capable of producing sounds that are relatively similar to tone and slap. This is consistent with the discussion of sabar technique above. However, it cannot produce a variety of timbres. In this study, the hand had a range of about 17 semitones (bass compared to slap), whereas the stick's range was less than half or that (edge compared



Fig. 8. The several layers of the hand used in Kiki showing, from left to right, the aluminum and spring-steel 'forearm'; the vinyl core; silicone with an embedded anchor near the fingertips; and latex 'skin'.

to center). In particular, the stick could create a sound with a low centroid similar to bass stroke, which is consistent with the observations about hand morphology. The rubber sheet could also produce sounds similar to tone and slap, and additionally had a range of 32 semitones (center to rim), which is nearly twice as large as the hand. However this range extended the range of the hand in the direction of increasing centroid, and so the rubber also could not excite the lower normal modes. Although it was approximately the correct size and shape to excite the fundamental, it was perhaps too flimsy and lightweight to do so effectively. The foam rubber fist was taken as an extreme example of an object that is large enough, sturdy enough, and the correct shape to excite the fundamental. It produced a timbre quite similar to, and even slightly lower than the bass stroke. Due to its large size it was incapable of producing any sound aside from this.

E. Kiki's Hand

Using insights gleaned from the foregoing study, a number of prototype striking mechanisms were built. Notable amongst them was a fully lifelike silicone rubber hand, made by alginate casting a human hand. The result was a copy accurate down to the level of detail of the fingerprints. This prototype produced a satisfying range of sounds, however it also had a few problems. It was too heavy to be actuated by practical means; the slap was not quite crisp enough at low amplitudes; and it was somewhat too floppy and in certain scenarios the fingertips would jiggle and bounce on and off the head making several onsets when only one was intended. So Kiki's final hand, whose construction is depicted in Figure 8 was made to address these issues.

The entire hand is made upon an aluminum rod which serves as its 'forearm', and which extends several inches into the hand. At the very interior of the hand, two gracile but very rigid lengths of spring steel transect the aluminum rod, which serve to make the 'palm' very rigid and inflexible, thereby improving the bass stroke. The steel cross-pieces were then embedded into a piece of black vinyl sheet-rubber that had been cut roughly into the shape of a human hand, but somewhat smaller than the desired final hand size. This vinyl is less floppy than silicone, and prevents the 'fingers' from bouncing on the drumhead. A small mold was then built

that was somewhat larger in all dimensions than the vinyl cutout, and this was used to encase the vinyl in silicone. This gives the hand enough weight and softness to play the bass and tone strokes. Additionally, a small metal anchor was embedded in the silicone such that a wire loop protruded from the top of the hand near the ‘fingertips’. Attaching a cable to this loop allows the fingers to be hyperextended, effectively controlling the rigidity of the hand. The silicone was dipped into liquid latex rubber, which cured and formed a thin skin around the entire hand. Latex has a somewhat harder surface texture than silicone, which improves the slap sound. Finally, the latex was coated with chalk dust to remove the tack from its surface texture. This hand was not evaluated using the methods outlined in the foregoing study. A more robust method of comparing the timbres of hand-drum strokes has been presented in [17]; relating that more strongly to human perception will provide a method of comparing the similarity of human- and robot-produced strokes, and this has been left as future work.

III. ARM

A synthetic hand, even a very good one, does not, by itself, guarantee a satisfying range of timbres; it must also be driven into the drum in an appropriate variety of ways. Following is an analysis of the dynamics and kinematics of a robotic arm designed to drive a synthetic hand into the head of a hand-drum for musical purposes. Implicit in the analysis is the goal of producing subtle timbral variations as previously discussed. The analysis reveals the necessity that the arm have three degrees of freedom. Specific actuators are analyzed for suitability in this application, and the optimum length for each segment is calculated. The analysis also solves the inverse kinematics problem for the proposed arm.

A. Dynamics

1) *Human Dynamics*: Many of the most satisfying synthetic hands studied were relatively massive, including the one ultimately used. Simple, inexpensive actuators such as hobby servos and small solenoids proved incapable of driving these hands, so other solutions are necessary. A simple study was therefore conducted to assess the physical requirements of the sought actuators. The primary constraint is the impact velocity of the hand against the drum, which will be higher for louder sounds, according to the type of stroke. An actuator, in order to be suitable for the current application, needs to be capable of driving the hand at both the highest and lowest velocities (or approximately so). In order to estimate these velocities, the following measurements were made. Sixteenth notes were played on the drum by a human at a tempo of 120 beats per minute, measured with a metronome, using the tone stroke and alternating hands. This tempo was chosen as the basis for analysis because it is fast enough that there is no time for extraneous or wasted movement in the player’s body, but it is not so fast as to limit the player’s dynamic range. The interval Δt between note onsets was thus

$$\Delta t = 0.125 \text{ s} \quad (4)$$

Because the hands alternate, this is the duration of half of the period of the stroke, namely the part of the stroke during which where a hand moves from its highest point above the drum head to the point of contact with the head. The return to the highest point requires another Δt seconds, but that part of the stroke is not of interest here. In one condition of the experiment the drum was played pianissimo, and in the other condition it was played fortissimo. In each condition, the height, h_p and h_f , of the hand above the drum head was measured during that phase of the stroke at which the distance was maximum. These were approximately

$$h_p \approx 0.03 \text{ m (1.25 in)}; h_f \approx 0.13 \text{ m (5 in)} \quad (5)$$

Given h , it is possible to derive the average acceleration a during the interval Δt .

$$a = \frac{2h}{r^2}; a_p = 3.84 \text{ m/s}^2; a_f = 16.64 \text{ m/s}^2 \quad (6)$$

If it assumed that the hand undergoes constant acceleration over the interval Δt , the impact velocity, v , may be derived.

$$v = at; v_p = 0.115 \text{ m/s}; v_f = 2.08 \text{ m/s} \quad (7)$$

If the acceleration is not, in fact, constant, then v could be any value given Δt and h , but the foregoing values will be taken as a reasonable approximation.

2) *Servo Dynamics*: Previous projects have successfully employed inexpensive hobby servomotors [23] or solenoids [12] to drive lightweight drumsticks. Although such were not able to drive the present (heavy) mechanism to the desired velocity, higher quality robotics servos can meet these requirements. Because it is trivial to actuate a servo slowly, the following analysis shall focus on the maximum velocity constraint, v_f . A servo with no load will rotate at its published maximum angular velocity ω_{noload} . A massless arm (with no end effector) of length r , attached at one end to the horn of the servo, will have an instantaneous velocity v at its extremity $v = \omega r$. However, if a load of constant mass m (i.e. a synthetic hand) is attached to the extremity of the arm, the servo’s actual angular velocity ω will decrease as the motor struggles to move it. In this case, as r increases, the applied torque $mgr\cos\theta$ increases, which causes ω to decrease until $mgr\cos\theta$ is equal to the published stall torque T_{stall} , at which point the velocity will be naught. The precise curve of ω as a function of applied torque is not typically published, but for the sake of analysis it will be assumed that it is linear, as defined by the known points, at which either ω or $mgr\cos\theta$ is zero.

Substituting these points into the definition of a line gives the following.

$$\omega = \omega_{noload} \left(-\frac{mgr^2\cos\theta}{T_{stall}} + r \right) \quad (8)$$

In other words, increasing r causes v to increase according to $v = \omega r$, but it also causes the applied torque to increase, resulting in a decrease of ω and consequently v . So what value of r , r_{opt} , will result in maximum v , v_{max} ? First we substitute



Fig. 9. Impact angle of Mamady Keita's hand while playing strokes

v/r for ω and solve for v . Then, we search for the extrema by finding the derivative of v with respect to r , searching for the value of r (r_{opt}) at which the derivative is naught, and solving for r_{opt} .

$$\frac{d}{dr}f(v) = \omega_{noload} \left(-\frac{2mgr\cos\theta}{T_{stall}} + 1 \right) \quad (9)$$

$$0 = \omega_{noload} \left(-\frac{2mgr_{opt}\cos\theta}{T_{stall}} + 1 \right) \quad (10)$$

$$r_{opt} = \frac{T_{stall}}{2mg\cos\theta} \quad (11)$$

In other words, the maximum velocity of a particular end-effector (synthetic hand) given the constraints of a servo is achieved when the arm is just such a length that the applied torque is half of the published stall-torque. Because the applied torque depends upon the angle with respect to gravity, the highest velocity could be achieved with a telescoping arm that shortens in length when it becomes more horizontal. This is not expected to be necessary in the current application and will not be further considered.

B. Three Segment Arm

Initially, it may seem that the arm under discussion would need two degrees of freedom: one to control the radial distance of the hand from the center of the drum, and another roughly analogous to the flexion and extension of the human elbow, in order to strike the drum. However, informal experimentation with natural and synthetic hands revealed that different timbres can be produced by striking the drum at different angles, which necessitates a third degree of freedom. In particular, during the tone stroke, the wrist lies approximately on (or slightly below) the plane of the drum head at the moment of impact, but for slap, the wrist is considerably below it. For bass the wrist obviously must be above the plane of the drum. These informal findings are confirmed by scrutinizing a video of djembefola Mamady Keita, as seen in Figure 9. His hand, in addition to being less rigid during slap, strikes the drum from a lower angle.

a) Impact Angle: In practice, these degrees of freedom will not be entirely orthogonal, as a stroke will involve raising the synthetic hand above the drum by flexing the 'elbow', which changes the angle and radial distance of the end effector from the center of the drum as well. For the sake of the present analysis, however, an orthogonal system will be imposed. The

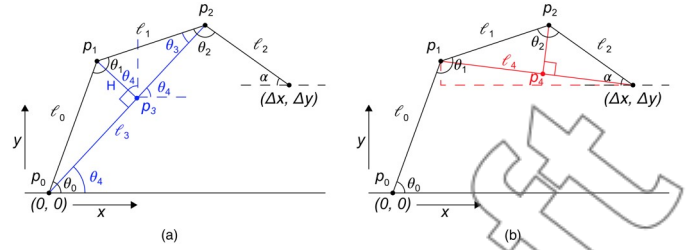


Fig. 10. Inverse Kinematics for a three-segment robotic arm, showing the variable names used in the analysis. The black segments depict the arm, and the colored parts represent quantities used in intermediate calculations.

degrees of freedom under consideration will thus be the radial distance Δx , the height Δy of the end effector above the drum head, and the angle α of the hand with respect to the plane of the drum head. In the case of a robot, the useful range of locations and angles of attack will depend on the morphology of the particular synthetic hand and the particular drum. In this case, informal experimentation with the preferred synthetic hand has shown the following ranges to be useful at the moment of impact.

$$\Delta x \approx 0 \sim 30cm \quad (0 \sim 12in) \quad (12)$$

$$\alpha \approx -0.4 \sim 0.4rad \quad (-23^\circ \sim 23^\circ) \quad (13)$$

Here, Δx represents the distance between the fingers striking the edge of the drum and the palm striking the center of the drum, for a 16 inch drum and 4 inch fingertip-to-palm distance. Estimating the necessary height Δy is more difficult because at slow tempi human players consume time by bringing their hand higher (and accelerating more slowly) than is physically necessary to achieve the desired impact velocity. This additional height may provide meaningful visual cues to human musicians. Based on the analysis in III-A1 above, it is expected that the following should be ample.

$$\Delta y \approx 0 \sim 30cm \quad (0 \sim 12in) \quad (14)$$

C. Inverse Kinematics

Given the desired coordinates $(\Delta x, \Delta y)$ of the arm's end-point and the angle α that the last segment makes with respect to the plane of drum head, we wish to know the appropriate angle of each servo, $\theta_0, \theta_1, \theta_2$ (refer to Figure 10 for the variable names used in the following analysis). To find these, it is first necessary to calculate the two dimensional position, p_0, p_1, p_2 , of each servo. The coordinate system shall be defined such that p_0 lies at $(0, 0)$. Since the length of each arm segment, ℓ_0, ℓ_1, ℓ_2 , is constant, the position of p_2 is easy to calculate.

$$p_0 = (\Delta x - \ell_2 \cos\alpha, \Delta y + \ell_2 \sin\alpha) \quad (15)$$

Calculating the position of p_1 is somewhat more involved. First, the distance ℓ_3 between p_0 and p_2 must be calculated (As depicted in blue in Figure 10(a)).

$$\ell_3 = \sqrt{p_{2x}^2 + p_{2y}^2} \quad (16)$$

Here, the subscripts x and y indicate the x and y coordinates of the point. Note that a solution to the inverse kinematics will exist if, and only if $\ell_3 \leq \ell_0 + \ell_1$. Given that, the position of p_1 is found as follows. There exists a point p_3 that lies upon ℓ_3 and is the shortest distance, H , from p_1 . The precise location of p_3 along ℓ_3 depends upon the relative lengths, λ , of ℓ_0 and ℓ_1 (given here without proof).

$$\lambda = 0.5 + \frac{\ell_0^2 - \ell_1^2}{2\ell_3^2}; p_3 = (\lambda p_{2x}, \lambda p_{2y}) \quad (17)$$

The length of H is found by first finding the angle θ_3 between ℓ_1 and ℓ_3 , using the Law of Cosines:

$$\theta_3 = \arccos \frac{\ell_1^2 + \ell_3^2 - \ell_0^2}{2\ell_1\ell_3} \quad (18)$$

This allows H to be found using the definition of sine.

$$H = \ell_1 \sin \theta_3 \quad (19)$$

However, on modern computers, trigonometric functions are typically implemented using successive approximation (i.e. Taylor series) and consequently have a high time complexity. Timing is highly important in musical applications, so it is therefore desirable to simplify trigonometric expressions where possible. Since $\sin(\arccos(\theta)) = \sqrt{1 - \theta^2}$, Equation 18 and Equation 19 may be simplified as follows.

$$\theta'_3 = \frac{\ell_1^2 + \ell_3^2 - \ell_0^2}{2\ell_1\ell_3}; H = \ell_1 * \sqrt{1 - \theta'^2_3} \quad (20)$$

Here, the prime indicates that θ is no longer a valid angle, just an intermediate result. H is at some angle θ_4 with respect to vertical, which must be found in order to separate H into its components. Because H is perpendicular to ℓ_3 , θ_4 is also the angle θ_4 which ℓ_3 makes with the horizon. This angle can be found using the definition of \cos , but since $\arccos \theta$ is only defined for $0 \leq \theta < \pi$, angles on the interval $-\pi \leq \theta < 0$ must be deduced manually, according to the position of p_{2y} .

$$\theta_4 = \begin{cases} \arccos \frac{p_{2x}}{\ell_3} & \text{if } p_{2y} > 0 \\ -\arccos \frac{p_{2x}}{\ell_3} & \text{otherwise} \end{cases} \quad (21)$$

This allows the calculation of p_1 .

$$p_1 = (p_{3x} - H \sin \theta_4, p_{3y} + H \cos \theta_4) \quad (22)$$

Again, these trigonometric functions can be simplified, allowing the calculation of p_1 as follows.

$$\theta'_4 = \frac{p_{2x}}{\ell_3}; p_{1y} = p_{3y} + H\theta'_4 \quad (23)$$

$$p_{1x} = \begin{cases} p_{3x} - H\sqrt{1 - \theta'^2_4} & \text{if } p_{2y} > 0 \\ p_{3x} + H\sqrt{1 - \theta'^2_4} & \text{otherwise} \end{cases} \quad (24)$$

There are, in fact, two solutions for the position of p_1 . The other is the reflection of p_1 about ℓ_3 , and can be solved by using $-H$ in place of H . The solution given here, chosen arbitrarily, puts p_1 farther from the body of the drum during normal operation. The angles θ_1 , and θ_2 shall be calculated using the Law of Cosines, which means that, in addition to

ℓ_3 , the length ℓ_4 of the third side of triangle ℓ_1, ℓ_2 will need to be known (as depicted in red in Figure 10(b)).

$$\ell_4 = \sqrt{(\Delta x - p_{1x})^2 + (\Delta y - p_{1y})^2} \quad (25)$$

Furthermore, in order to find whether $-\pi \leq \theta_2 < 0$, it will be necessary to determine whether p_2 lies above or below ℓ_4 . This may be accomplished by defining the point on ℓ_4 which lies nearest to p_2 . This may be found analogously to Equation 17.

$$\lambda' = 0.5 + \frac{\ell_1^2 - \ell_2^2}{2\ell_4^2}; p_{4y} = \lambda'(\Delta y - p_{1y}) \quad (26)$$

The x coordinate of p_4 is not needed. It is now possible to calculate the sought angles θ_0, θ_1 , and θ_2 , using the Law of Cosines and the definition of cosine, again manually correcting for negative angles.

$$\theta_0 = \begin{cases} \arcsin \frac{p_{1y}}{\ell_0} & \text{if } p_{1x} > 0 \\ \pi - \arcsin \frac{p_{1y}}{\ell_0} & \text{otherwise} \end{cases} \quad (27)$$

$$\theta_1 = \arccos \frac{\ell_0^2 + \ell_1^2 - \ell_3^2}{2\ell_0\ell_1} \quad (28)$$

$$\theta_2 = \begin{cases} -\arccos \frac{\ell_1^2 + \ell_2^2 - \ell_4^2}{2\ell_1\ell_2} & \text{if } (x > p_{1x}) \text{ XOR } (p_{2y} > p_{4y}) \\ \arccos \frac{\ell_1^2 + \ell_2^2 - \ell_4^2}{2\ell_1\ell_2} & \text{otherwise} \end{cases} \quad (29)$$

Q.E.D.

IV. FUTURE WORK

An optimal striking algorithm for the arm remains an open area for future study. The idea is to use a closed-loop controller to bring the hand in contact with the drum at the correct location and time and with the correct velocity, and to do so by moving along a path that can be easily interpreted visually by human interactors. The current solution implemented in Kiki involves some simplifying assumptions and works acceptably well only provided that there is not a great change in velocity between adjacent strokes. Further research is also needed to more rigorously assess the degree to which this robot achieves its stated purpose, i.e. how perceptually similar its sounds are to a human player.

ACKNOWLEDGMENTS

Thank you to Garth Paine for generous support of some of the research presented in this paper.

REFERENCES

- [1] Ibn al-Razzaz al Jazari. Oxford. MS Graves 27.
- [2] Ibn al-Razzaz al Jazari. *Miniature from a manuscript source of Kitb al-jmi' bayan al-'ilm wa al-'amal al-nfi' f sin'at al-hiyal*. Freer Gallery of Art, Smithsonian Institution, Inventory Number F 1942.10, 1315. URL http://www.qantara-med.org/qantara4/admin/pics_super_zoom/1118Clepsidre-aux-musiciens-Egy.jpg. [Online; accessed 13-July-2015].

- [3] Ibn al-Razzaz al Jazari. *The Book of Ingenious Mechanical Devices (Kitab fi ma'rifat al-hiyal al-handasiyya)*. D. Reidel, Dordrecht, Holland, 1974.
- [4] Nicholas A. Baginsky. The three sirens: A self learning robotic rock band. URL <http://www.the-three-sirens.info/>. [Online; accessed 01-September-2015].
- [5] Compressorhead. Compressorhead ace of spades. URL https://www.youtube.com/watch?v=3RBSkq-_St8. [Online; accessed 11-January-2016].
- [6] S. S. Dey and V. Tirumala Rao. Transient response of circular plates and membranes: a numerical approach. *International journal of mechanical sciences*, 39(12): 1405–1413, 1997.
- [7] William Cronk Elmore and Mark A. Heald. *Physics of waves*. Courier Corporation, 1969.
- [8] Antti Eronen. Comparison of features for musical instrument recognition. In *2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 19–22. IEEE, 2001.
- [9] Guy Hoffman and Gil Weinberg. Interactive improvisation with a robotic marimba player. *Autonomous Robots*, 31(2-3):133–153, 2011.
- [10] Ajay Kapur. A history of robotic musical instruments. In *Proceedings of the International Computer Music Conference*, pages 21–28. Citeseer, 2005.
- [11] Ajay Kapur. *Digitizing North Indian Music: Preservation and Extension using Multimodal Sensor Systems, Machine Learning and Robotics*. PhD thesis, University of Victoria, 2008.
- [12] Ajay Kapur, Eric Singer, Afzal Suleman, and George Tzanetakis. A comparison of solenoid-based strategies for robotic drumming. *ICMC Copenhagen*, 2007.
- [13] Mamady Keita. Djembe: Bass tone and slap by Mamady Keita, . URL <http://www.youtube.com/embed/NI05CDVV8XQ>. [Online; accessed 17-December-2012].
- [14] Mamady Keita. Mamady Keita on djembe technique, . URL <http://www.youtube.com/embed/qOs732jFwcE>. [Online; accessed 14-December-2014].
- [15] Famoudou Konaté, Thomas Ott, and Thomas Goldhahn. *Rhythms and Songs from Guinea*. Lugert, 2000.
- [16] Carol L Krumhansl. Why is musical timbre so hard to understand. *Structure and perception of electroacoustic sound and music*, 9:43–53, 1989.
- [17] Michael Krzyzaniak and Garth Paine. Realtime Classification of Hand-Drum Strokes. In *International Conference on New Interfaces for Musical Expression*, 2015.
- [18] Ye Pan, Min-Gyu Kim, and Kenji Suzuki. A Robot Musician Interacting with a Human Partner through Initiative Exchange. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 166–169, 2010.
- [19] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, Ircam, 2004.
- [20] Byeong Sam Jeon. Telematic Sound Improvisation Using Human-Controlled Robotic Percussion Instruments. Master's thesis, University of California, Irvine, 2008.
- [21] Eric Singer, Jeff Feddersen, Chad Redmon, and Bill Bowen. LEMUR's Musical Robots. In *International Conference on New Interfaces for Musical Expression*, 2004.
- [22] Mark Sunkett. *Mandiani drum and dance: Djimbe performance and black aesthetics from Africa to the new world*. White Cliffs Media Co, 1995.
- [23] Assaf Talmudi, Jonathan Rubin, and Liat Segal. Robotic drums. URL <http://www.youtube.com/embed/uqecHPnvb6c>. [Online; accessed 19-September-2015].
- [24] Carlos Andres Velez, Diego Valencia, Ibrahim Tansel, and Sabri Tosunoglu. Design of a MIDI compatible drumming robot. In *Proceedings of ECTC 2008 ASME Early Career Technical Conference*, pages 1.1–1.6. ACM, 2008.
- [25] Gil Weinberg and Scott Driscoll. Toward robotic musicianship. *Computer Music Journal*, 30(4):28–45, 2006.
- [26] Gil Weinberg, Scott Driscoll, and Mitchell Parry. HAILE—AN INTERACTIVE ROBOTIC PERCUSSIONIST. Ann Arbor: Scholarly Publishing Office, University of Michigan Library, 2005.
- [27] Matthew M. Williamson. *Robot Arm Control Exploiting Natural Dynamics*. PhD thesis, Massachusetts Institute of Technology, 1999.