

Audio Visual Synchronization of Rhythm

Andrew Godbout and Jeffrey E. Boyd
University of Calgary
Calgary, Canada

agodbout@ucalgary.ca, jboyd@ucalgary.ca

Abstract

The use of sound feedback to assist in learning and correcting motor skills has some innate advantages. For example, sound leaves a person's eyes free for important tasks like avoiding obstacles, and ample psychological evidence shows a connection between the synthesis and perception of motion and sound. When motion is periodic, as it is with most forms of human locomotion, the sound feedback should also be periodic, i.e., rhythmic and synchronized to a person's motion. We present a method to track an athlete performing a known movement in real-time using an RGB-D camera. We use a Kalman filter in conjunction with iterative closest point to synchronize incoming data against a known movement pattern and produce a musical rhythm. The athlete and the rhythm move in phase with one another. We call this process rhythmic sonification. We demonstrate our method by tracking and rhythmically sonifying an advanced in-line speed skating movement called double-push. Rhythmically sonified feedback provides a new opportunity to teach and reinforce this and other athletic skills.

1. Introduction

It is now common to mount cameras on people's bodies, on flying drones, on vehicle dashboards and on specialized robots. Capturing footage of athletes in their environment is among the most prevalent applications provided by the increased portability and inexpensive nature of these cameras. An interesting feedback loop can be created whereby an athlete is able to receive computer analysis and feedback of their movements in real-time in their environment. We present a method to track an athlete performing a known repetitive movement and produce a rhythmic sound feedback synchronized with the rhythm of the movement. Our method is robust and provides ease of use and setup. We demonstrate the method with speed skating athletes performing an advanced technique on a speed skating treadmill. The rhythmic sound is designed to reinforce the movement pattern by creating an interactive feedback loop.

Examples of the human ability to synchronize their motion with sound are found throughout many cultures. We see examples when an audience claps their hands with a piece of music, when dancers move together during a waltz, and when soldiers march in step to a band. The association is profound and can be exploited to help people acquire and perfect new motor skills, or to recover normal motor skills in clinical treatment.

In this paper, we build on prior work in the field of sonification - the non-verbal representation of data in sound. Most human locomotion, and many human activities, involve periodic motions. We exploit this by creating sonified feedback that is both rhythmic and synchronized to a person's motion. Figure 1 illustrates the concept. The process begins with measurements of a person performing a periodic motion acquired in real-time. These measurements can come from any number of sensors including sensors worn on the body like accelerometers or goniometers, skeletal pose angles obtained from a motion capture system, or (as proposed in this paper) from an RGB-D camera. A computer model synchronizes its simulated motion with the measurements from the person, which in turn provides audio feedback. As the feedback helps the person to learn or perfect a motor skill, the feedback also allows the person to synchronize their motion with the computer model. In effect, a process of mutual synchronization emerges that allows the person and model to walk in step just as two or more people walking side-by-side can adjust their gaits to do the same.

The computer model can come from any number of sources, but it must allow synchronization based on the real-time measurements. Typically, we acquire a model from another person who can already perform the motor skill - we record the skilled performers movements to get typical examples of the skill being performed correctly.

The emphasis on synchronization and rhythmic feedback represents a change from conventional approaches to human motion analysis and pose estimation. The usual goal in human motion analysis is pose estimation. The reasonable assumption is that from a correctly computed pose, one can

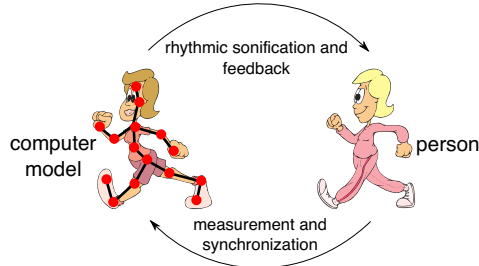


Figure 1. An illustration of the concept of synchronous sonified feedback. A computer model synchronizes to real-time measurements of a person engaged in a periodic motion. The system produces sound feedback to help the person learn or restore the motion. Because the person hears the sound feedback, they can synchronize themselves to the model and sound. The process is analogous to soldiers adjusting their strides to march in step.

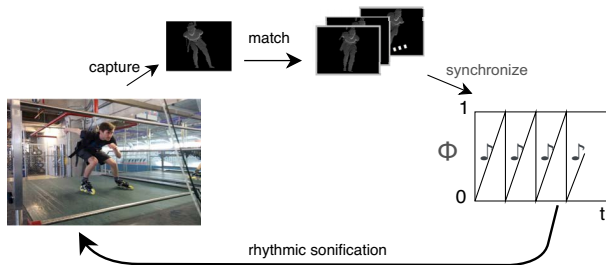


Figure 2. A skater on the treadmill is recorded with a kinect camera, each new incoming point cloud is compared against point clouds in the model movement library using the Iterative Closest Point algorithm. The comparisons are combined into a single phase estimate, ϕ , which is tracked over time and drives rhythmic auditory feedback completing the interactive feedback loop.

derive whatever they need. In fact, we can synchronize from skeletal pose angles from a Vicon [30] system as shown by Boyd and Godbout [5], but the measurements need not be so complex. Since our singular requirement is to produce a reliable phase measurement, we can do this from other measurements, obviating the need for a solution to pose estimation. However, the requirement for sonified feedback introduces a new constraint - the phase estimate must be accurate enough to keep timing errors in the rhythm within human tolerances for the perception of rhythm. Given that humans can hear small timing differences, this is a challenging requirement.

Figure 3 describes the process that produces rhythmic audio feedback. The synchronization process produces a phase signal, ϕ that cycles from zero to one, repeating with the period of the motion. We set one or more phase thresholds, ϕ_T . As ϕ crosses the phase threshold, the sonification system produces a sound. We avoid continuous sounds - they do not work well for audio feedback - so the sound will be transient like a single note played on a piano, or a single drum beat. The number and placement of the phase

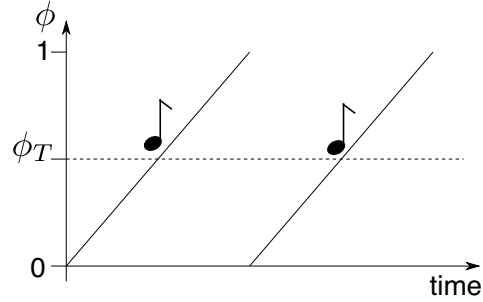


Figure 3. Rhythmic sonification: as a person performs a periodic motion, the phase ϕ ramps from zero to one repeatedly. Each time the phase crosses the phase threshold ϕ_T , the systems triggers a sound. Because of the periodic nature of the phase, the resulting sound pattern is rhythmic, and synchronized to the person's motion.

thresholds determines the rhythmic pattern that the sonification system plays. Variations in the pitches associated with the thresholds controls the repeating melody. This allows us to design a sound feedback pattern that reinforces correct performance of a motion, emphasizing important elements of timing.

2. Background

2.1. Perception and Synthesis of Motion

The study of human motion perception through vision provides insight that is useful in the design of motion feedback systems. Johansson's seminal work with moving light displays demonstrated that the visual perception of human motion is reliable and fast: "after the first one or two steps: this is a walking human being! This perceptual effect has been observed without exception" [18]. Blake and Shiffrar [4] provide a survey of results covering several decades of research in this area. They point to the body of evidence regarding the robustness of human motion perception in the presence of numerous confounding factors. Among the few factors that can interfere with this robust phenomenon is a distortion of the timing or *phase* relationships in the motion [1] – when the temporal phase of the motion of moving lights is scrambled, the perception of human motion is lost.

Blake and Shiffrar [4] also point to "converging lines of evidence" that the human ability to perceive human motion results at least in part from experience in producing one's own motion. Neurological evidence points to the existence of mirror neurons that activate both when an action is performed, and when the same action is perceived in others [22, 21]. Furthermore, these mirror neurons also respond to the *sound* associated with an action, even in the absence of the corresponding visual stimulus [21]. Bidet-Caulet et al. [3] identified a similar response to auditory footsteps. They draw the conclusion that there is a

“supramodal area” of the brain that is “involved in human motion recognition, irrespective of the sensory modality input.”

In the context of the work presented here, we highlight the following salient points.

1. Humans have a specialized ability to perceive human motion.
2. This specialized ability is connected to our experience synthesizing our own motion.
3. The perception of motion extends beyond vision to include sound.

2.2. Sonification

Sonification refers to the non-verbal representation of data with sound [15]. It is a diverse field involving psychology, music, kinesiology, computer science, and engineering.

Within the portion of sonification that involves sport, the concept is particularly attractive because communicating through sound does not interfere with an athlete’s vision, maintaining required visual stimuli for performance and safety. Examples of sonification in sport are numerous. Godbout and Boyd [11] introduced sonification to the sport of speed skating, describing an application that provided real-time feedback to correct lost cross-over technique of a specific skater. Stienstra et al. [27] demonstrated a sonification technique that mapped data from force sensors embedded in speed skates to audio feedback for the skater to improve technique. Godbout et al. [12] demonstrated sonification of accelerometer data from a consumer-grade wearable sensor, the *Armour 39* [29].

In the sport of running, Eriksson et al. [9] describe a sonification system that provides real-time feedback to help improve *running economy*. Hockman et al. [17] use a runner’s pace to control the time scaling of a phase vocoder to synchronize music playback. This enables synchronized music playback to become part of a runner’s workout. Godbout et al. [12] use a phase synchronization system to sonify *virtual footfalls*. Their goal is to encourage runners to increase their cadence and shorten their stride, thereby reducing heel striking. As the runner approaches the desired cadence, the virtual footfalls coincide with the actual footfalls, reinforcing the correct motion.

Schaffert et al. [25, 24] apply sonification to rowing by sonifying force feedback to the athletes. Their research shows significant improvements in mean boat velocity, and anecdotally this research is associated with a successful Olympic rowing team. Wolf et al. [32] sonify a motion error signal to provide feedback. However, they observe that the sonified error signal does not help one to learn motor skills.

We note that this seems to be true in general of error sonification, possibly because sonified errors do not normally tell a person how to correct the error.

In addition to these sports, karate [28], swimming [16] and golf [26] have all been successfully sonified as part of an interactive experience.

2.3. Rhythmic Sonification

Parameter mapping in sonification refers to the mapping of values from input sensors to one or more audio parameters for data display [14]. For example, one could map the force measured by a sensor to the pitch of a sound, or map the distance to an object to loudness. The technique is common in sonification for sports. For example, consider accelerometer data – in golf [26], karate [28] and rowing [25, 24] the mapping of acceleration to musical pitch allows the users to perceive the level of force being applied. Sensor data has been parameter mapped to sound characteristics to allow users to perceive changes in their posture [9] or orientation of their equipment [32].

However, parameter mapping does not innately capture the important phase and timing data that is critical in human motion. For example, consider the rowing sonification system described by Schaffert and Mattes [24]. The system maps a number of force and acceleration measurements to sound parameters. Although the mapping does not explicitly consider rhythm, the sounds produced are rhythmic because the human motion of rowing is periodic, which leads to periodic measurements, which finally leads to periodic sounds (i.e., rhythms). Schaffert and Mattes report that the success of their system is at least in part due to rhythmic sounds. The rowers can synchronize to the sound, and thereby row in unison to increase efficiency and speed.

Godbout and Boyd [11] proposed *rhythmic sonification*, which differs from simple parameter mapping by explicitly modelling the periodicity of human motion and sensor data derived from human motion [11]. Figures 1 and 3 in Section 1 describe the concept. To illustrate the process, we recap some results from Godbout and Boyd [11]. The context was a speed skater that had lost his ability to do crossovers while in corners on an oval track – for unknown reasons he was dropping his right toe as he set the skate down, which caused instability and loss of speed. Referring to Figure 1, the model was a set of ankle angle measurements representing a single cycle of a corner speed skating stride as demonstrated by an accomplished skater. The measurements came from a simple goniometer that could measure ankle movement in real-time, worn by the skater. The subject skater then wore the goniometer as a small computer carried by the skater synchronized the live measurements with the recorded model. The synchronization algorithm was a brute-force search over phase and cadence space to maximize the cross correlation between model and live data

at any moment in time. The algorithm could therefore adjust to a skater's change in cadence while producing a phase signal like that illustrated in Figure 3.

The sonification has two components. First, a set of four phase thresholds (ϕ_T in Figure 3) trigger notes on a C-major arpeggio. So when the subject skater matched the model well, he would hear the C-major arpeggio played in four-quarter time. This rhythmic feedback had the most significance, and quickly helped the skater to recover a correct rhythm in his stride. Second, the sonification system produced a buzzing sound when the ankle angle was too large just before the skate contacts the ice. This is trivial with the synchronization system – phase is known, so it is easy to identify the moment before the blade sets down, and apply parameter mapping to the ankle angle at that specific moment.

Variations on the theme of rhythmic sonification including using multiple sensor data streams, such as skeletal pose angles from a motion capture system [5], or using data from consumer grade wearable sensors [12]. The key features of rhythmic sonification are:

1. synchronization of sensor data to model data to produce a phase signal, and
2. mapping of phase threshold to trigger sounds.

In short, rhythmic sonification does not simply map a data value, but also maps temporal information in the data to sound. When viewed in the context of the psychological and neurological evidence cited in Section 2.1, rhythmic sonification is an ideal tool for assisting people to learn or recover motor skills.

While it is still a subject of research, rhythmic sonification has potential beyond sports in the area of clinical treatments, as seen in research on treatments for Parkinson's patients. For example, del Olmo and Cudeiro [8] showed the coefficient of variability in the gait of Parkinson's patients improved with rhythmic sound stimuli. Willems et al. [31] investigate the use of rhythmic auditory cues for Parkinson's patients. Nombela et al. [20] present a survey of results in the application of rhythm to the treatment of Parkinson's.

3. Visual Synchronization with RGB-Depth Camera

In recent years RGB-Depth (RGB-D) cameras have become common. The addition of depth/disparity data to conventional video offers many advantages. RGB-D data improves segmentation, view invariance and robustness in phase estimation. The double-push stride in in-line skating serves as a context for demonstration. Capturing and analyzing speed skating athletes presents an interesting application domain for visual tracking. Speed skating uses a

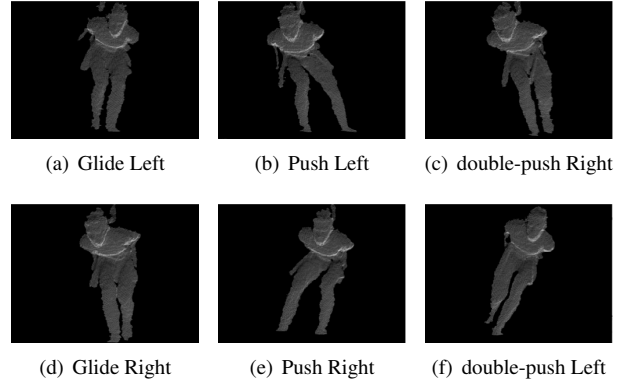


Figure 4. Phases of Speed Skating with double-push. Traditional skating has glide and push phases (a, b, d, e). double-push skating adds in an extra push using the exterior edge of the blade (c, f).

push with the blades to the side of the body to produce forward propulsion and a subsequent side to side swaying of the athletes.

We use speed skating to refer to both in-line speed skating (roller skates with wheels in alignment) and ice speed skating (traditional ice skating). The two types of speed skating share many of the same characteristics, however *double-push* is typically only executed within in-line skating. The components of the double push are outlined in Figure 4.

We hypothesize that rhythmic sonification based on synchronization between the skater and the particular model skater they wish to emulate will aid the skater. While teaching and training the rhythm of double-push is a practical and desirable goal, it is easy to see that the same techniques can apply to other periodic movements in which a subject is attempting to mimic an established or known movement pattern.

3.1. Methods

The work presented here uses a specialized speed skating treadmill. Ultimately, we seek to implement our method using a mobile apparatus such as that demonstrated by Boyd, et al [6]. This apparatus featured an RGB-D camera mounted on a mobile autonomous vehicle to both follow and capture depth images of speed skating athletes. The depth camera allows for easy segmentation using a connected components approach, and a best hypothesis multiple-target tracker [7] is able to track objects given the segmented object's centroid. Because the vehicle is designed to keep the target in the same relative position with respect to the camera, the best hypothesis multiple target tracker can use a Kalman filter with a constant-position model to resolve the trajectories. The work here uses the same system to track objects, but focuses on tracking a skater on a treadmill with a stationary camera. While there

may be small movements within the view frame due to the nature of the speed skating movement the same constant-position model assumption of the object centroid is a reasonable approximation.

3.1.1 Phase from Depth

In prior work producing phase from periodic motion [11, 12, 5], the incoming data is compared against a set of models built using the same sensor setup. The best matching model influences the current phase estimate in the movement. Iterative Closest Point (ICP) is a method for registering point clouds [2, 33, 13]. The ICP algorithm finds corresponding points between two point clouds and creates a rigid body transformation to minimize the distance and then iteratively continues until convergence at a local minimum.

The termination criteria of ICP, minimized euclidean squared distance between sets of matching points in two point clouds, is a measure of the similarity between the two point clouds. The Point Cloud Library (PCL) has implemented a version of ICP that takes advantage of a number of optimizations and estimations designed to reduce computational complexity [23], however, the algorithm remains computationally intensive. ICP returns the similarity, f , between two point clouds (p_1 and p_2): $f = ICP(p_1, p_2)$.

We use a Microsoft Kinect to record a model of the double-push movement. The model is a set of point clouds, G , recorded from an experienced in-line skater during one period of the repetitive movement. We search within G using ICP to find model point clouds that match the current data, and from the matches infer the current phase. Due to the computational complexity of the ICP algorithm we rely on a modified Kalman filter [10] to track and predict the phase of the movement.

Following the algorithm presented in Alg. 1, let $\mathbf{x}_k = [\phi_k, \dot{\phi}_k]^T$ be the state and $\mathbf{z}_k = [\bar{\phi}_k, \dot{\phi}_0]^T$ be the measurement of a skater at time k . ϕ is the phase of the skater within the double-push movement, $\bar{\phi}_k$ is the measured phase, $\dot{\phi}_k$ is the rate of change of ϕ and $\dot{\phi}_0$ is a typical cadence, a tuneable parameter. We have the following process and measurement models:

$$\mathbf{x}_k = F\mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \text{ and} \quad (1)$$

$$\mathbf{z}_k = H\mathbf{x}_k + \mathbf{v}_k, \quad (2)$$

where

$$F = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \quad (3)$$

$$H = I, \quad (4)$$

and $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ are the process and measurement noise respectively. The state covariance is:

$$P_k = \begin{bmatrix} \sigma_{\phi_k}^2 & 0 \\ 0 & \sigma_{\dot{\phi}_k}^2 \end{bmatrix}. \quad (5)$$

Algorithm 1: Tracking Point Clouds with Kalman filtering and ICP

input : p_k : current point cloud sample,

$G = \{(\phi_1, g_1), \dots, (\phi_m, g_m)\}$: Model set of (phase, point cloud) pairs.

output: Updated State Estimate $\hat{\mathbf{x}}_k^+$, Updated Error Covariance P_k^+

1. Extrapolate:

$$\hat{\mathbf{x}}_k^- = F\hat{\mathbf{x}}_{k-1}^+ \\ P_k^- = FP_{k-1}^+F^T + Q$$

2. Monte Carlo Measurement:

2.1 Select Sample Domain:

if $\sigma_{\phi_k}^- < \sigma_{\phi_{min}}^-$ **then**
 $\phi_{min} = \phi_k^- - \frac{5}{2}\sigma_{\phi_{min}}^-, \phi_{max} = \phi_k^- + \frac{5}{2}\sigma_{\phi_{min}}^-$
else
 $\phi_{min} = \phi_k^- - \frac{5}{2}\sigma_{\phi_k}^-, \phi_{max} = \phi_k^- + \frac{5}{2}\sigma_{\phi_k}^-$

2.2 Monte Carlo Sampling:

for $i = 1 \rightarrow n$ **do**
 select random j s.t. $(\phi_j, g_j) \in G$ and
 $\phi_{min} \leq \phi_j \leq \phi_{max}$
 $f = ICP(g_j, p)$
 $\beta'_i = \exp(-f^2/2\sigma_f^2)$
 $\phi_i = \phi_j$
end

$$\beta_i = \frac{\beta'_i}{\sum_{i=1}^n \beta'_i}, i = 1, \dots, n$$

2.2 Weighted Average:

$$\mathbf{z}_k = \begin{bmatrix} \sum_{i=1}^n \beta_i \phi_i \\ \phi_0 \end{bmatrix}$$

3. Update:

$$K = P_k^- H^T [H P_k^- H^T R]^{-1} \\ \hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + K[\mathbf{z}_k - H\hat{\mathbf{x}}_k^-] \\ P_k^+ = [I - KH]P_k^-$$

Consider p , a point cloud obtained from an RGB-D camera monitoring a subject performing a periodic movement. To obtain a measurement vector for p , the algorithm compares p to a set of model point clouds, G , using ICP. Each $\gamma \in G$ is a pair: (ϕ, g) containing a model point cloud, g , and the corresponding phase, ϕ .

We can perform n calls to the ICP algorithm within a fixed period of time, to determine a best match, $n \leq |G|$. We estimate a search domain within the phase domain of G as $\phi_k^- \pm \frac{5}{2}\sigma_{\phi_k}^-$ and sample that portion of the ϕ domain in a Monte Carlo fashion to obtain a search set, J , with size

n . Note, it is essential that the range of ϕ values sampled should not be too small, since this is equivalent to testing only a single model point cloud. Therefore, we force $\sigma_k^- = \sigma_{\phi_{min}}$ if $\sigma_k^- < \sigma_{\phi_{min}}$ (Algorithm 1: 2.2). We set n as large as possible while maintaining real-time performance.

For each phase ϕ_j , we select g_j , the model point cloud nearest to ϕ_j , and compute $f = ICP(g_j, p)$. f is weighted according to a normal distribution: $\beta' = \exp(-f^2/2\sigma_f^2)$. σ_f is estimated based on the range of ICP . We normalize β' to get β such that $\sum_{i=1}^n \beta_i = 1$. Finally, our measurement, $\mathbf{z}_k$, is the sum of the weighted similarities:

$$\mathbf{z}_k = \left[\begin{array}{c} \sum_{i=1}^n (\beta_i \phi_i) \\ \phi_0 \end{array} \right] \quad (6)$$

$\dot{\phi}_0$ is a constant parameter representing a typical cadence. The state extrapolation ($\hat{\mathbf{x}}_k^-$), error covariance extrapolation (P_k^-), gain (K) and updates ($\hat{\mathbf{x}}_k^+$, P_k^+), follow the standard kalman filter [10]:

$$\hat{\mathbf{x}}_k^- = F\hat{\mathbf{x}}_{k-1}^+, \quad (7)$$

$$P_k^- = FP_{k-1}^+F^T + Q, \quad (8)$$

$$K = P_k^- H^T [HP_k^- H^T R]^{-1}, \quad (9)$$

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + K[z_k - H\hat{\mathbf{x}}_k^-], \text{ and} \quad (10)$$

$$P_k^+ = [I - KH]P_k^-, \quad (11)$$

where Q is the process noise covariance, and R is the measurement noise covariance:

$$R = \left[\begin{array}{cc} \sum_{j=1}^n (\phi_j - \bar{\phi}_k)^2 & 0 \\ 0 & \sigma_{\phi_0}^2 \end{array} \right], \quad (12)$$

where $\sigma_{\phi_0}^2$ is the uncertainty in $\dot{\phi}_0$, set as $\sigma_{\phi_0}^2 = 0.1$. The system is initialized with:

$$\hat{\mathbf{x}}_0 = \left[\begin{array}{c} 0.5 \\ \dot{\phi}_0 \end{array} \right], P_0 = \left[\begin{array}{cc} 0.1 & 0 \\ 0 & 0.01 \end{array} \right]. \quad (13)$$

A successful tracking of a periodic movement produces a consistent ramping of phase, ϕ , from 0 to 1 and a tracking of $\dot{\phi}$ that is biased towards $\dot{\phi}_0$ but responds well to changes in pacing.

3.1.2 Sound from Phase

Rhythmic feedback is created by triggering sounds at evenly distributed intervals as the phase ramp crosses predefined thresholds in the phase space. Based on the duration of the double-push movement we divide the movement into four sections placing sonic triggers at each phase threshold between the sections, as in Figure 5. The four musical tones are based on a D major arpeggio, with notes in order: D_4 , $F_4^\#$, A_4 and D_5 .

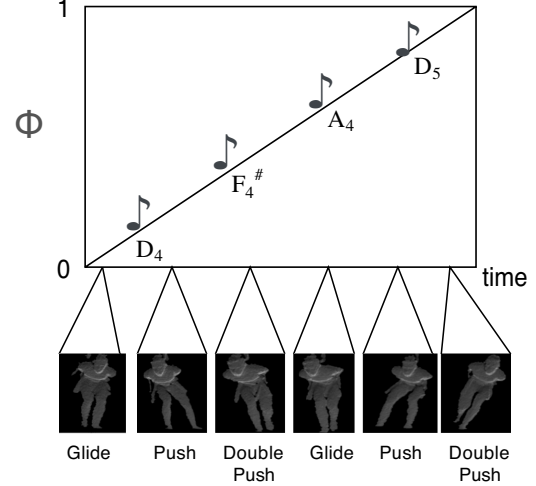


Figure 5. Rhythmic feedback is synchronized to the skating movement. As the subject progresses through the movement, ϕ advances in sync and as it passes each of the four thresholds the corresponding musical note is broadcast.

The notes are evenly spaced in the model movement and so a perfect matching of subject to model results in the musical notes being triggered at even intervals. The arpeggio sonically implies an order to the feedback.

3.2. Testing

We tested our system using a skating treadmill with two accomplished speed skating athletes, one acting as the model and the other as the subject. A kinect camera is set up at the front of the treadmill and the rhythmic auditory feedback is provided via wireless headphones. We instruct our skaters that a perfect matching of their movement to model movement produces an even distribution in the 4-note rhythmic auditory feedback. We are looking for a consistent phase ramp in sync with the subject's movement.

As detailed in Section 2.1, humans are innately able to perceive synchronization between movement and sound and especially one's own movement. We rely on this ability and consider it as the gold standard for determining synchronization between rhythm and movement. We determine that our phase estimate is synchronized with a subject by asking the subject to assess the registration between the sounds they hear and the movements they perform, i.e., we ask them if the musical rhythm occurs at the same time in their movement each repetition. As observers we do the same, listening to the rhythm while monitoring the movements of the subject. We used the following test procedure:

1. **Record the model:** The skating treadmill is set to 22 kilometers per hour, a relaxed, realistic pace for in-line skating. The first skater performs numerous repetitions of the double-push skating movement, which

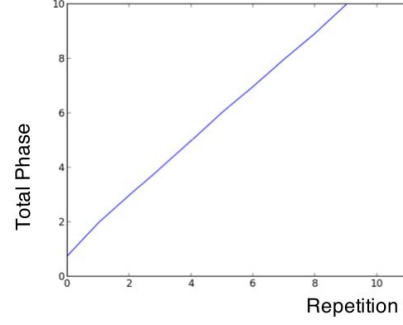
is recorded with the depth camera. A speed skating coach picks out one cycle of the movement to be the model, G .

2. **Synchronize model skater vs recorded model:** The first skater again skates on the treadmill. Now the incoming depth images are compared against the previously recorded model and the phase of the movement is computed. When the phase crosses the evenly spaced phase thresholds the corresponding musical note is broadcast. The skater changes pace of movement and the treadmill belt pace is adjusted between 18 and 24 kilometers per hour. The skater's perception of synchronization between movement and rhythm is held as the standard for evaluating the system. The signals are considered to be synchronized if the rhythmic auditory feedback occurs at perceptually consistent phases of the movement for both the skater and observers.
3. **Synchronize subject skater vs recorded model:** A second skater, the subject, skates on the treadmill performing in-line double-push. The treadmill speed varies between 18 and 24 kilometers per hour and the skater varies their cadence. Again, the skater's perception of synchronization between movement and rhythm is held as the standard for evaluating the system. The signals are considered to be synchronized if the rhythmic auditory feedback occurs at perceptually consistent phases of the movement for both the skater and observers.

3.3. Observations

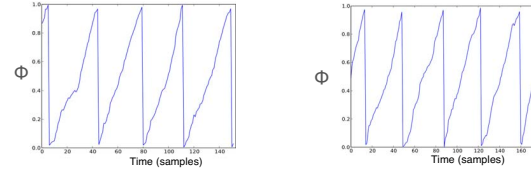
In the case of each subject, the system was able to synchronize against their movements and thus produce rhythmic auditory feedback (Figure 7). After the tests we manually observe the video frame by frame and select an identifiable point in each skating stride repetition the corresponding total phase (unnormalized) is then observed and presented in Figure 6. The subjects each commented that they could easily determine that the rhythmic auditory feedback was synchronized with, and under the control of their movements. The subjects each indicated that the system was easy to use and easy to understand.

One subject described the system as “simple yet surprisingly immersive” noting that he quickly became entrained with and focused on the rhythm. The lowest note (D_4) occurred at the start of the left foot push and one subject mentioned the feedback occurred when his wheels went over a ridge in the treadmill belt and it gave the sensation that his push was casting aside some object. The other skater mentioned he could hear differences in the rhythm between left and right double-push allowing him to establish an even rhythm and symmetry in his stride.



Rep	Phase
1	0.762
2	1.974
3	2.989
4	3.974
5	4.989
6	6.030
7	6.981
8	7.980
9	8.930
10	9.980

Figure 6. Manual visual selection of left foot set-down and the computer generated total phase value (unnormalized). The left foot set-down was selected manually in a video sequence of 10 skating strides and the subsequent computer generated phase is examined. A strong correlation is observed.



(a) Model Skater: Approx. 22km/hr (b) Subject Skater: Approx. 22km/hr

Figure 7. Example phase plots of the two subjects. Phase ramps from 0 to 1 as the subjects proceed through the double push. (a) is our model skater synchronized against his own model recording. (b) is our subject skater synchronized against the model recording. The sample rate is approximately 18Hz.

The phase estimate was responsive to the changes in pacing, including the typical rests, starts and stops the athletes undertook. There were no identifiable synchronization issues associated with the changes in treadmill belt speed, nor changes in individual subject cadence. However, maintaining a minimum σ_{ϕ_k} (Eq. 5) is essential. Without this requirement the tracker can become too confident in its predications when synchronization has occurred

for some duration, i.e., σ_{ϕ_k} becomes too small. Enforcing $\sigma_{\phi_k} \geq \sigma_{\phi_{min}}$, allows the tracker to continue sampling from a broad enough phase range to ensure phase lock. This is especially important when a subject abruptly changes pace, such as during resting, starting and stopping.

3.4. Discussion

As was the case with previous work in rhythmic sonification, the model requirements are not strict - almost any reasonable example of a cycle of the motion will do. Although there is variation in the ways people perform periodic motions, these variations are not large enough to prevent these (and other) synchronization methods from estimating a phase trajectory that is accurate to within the requirements of human perception.

We have not provided any empirical evaluation of the phase signal derived by our algorithms. This is in part due to a lack of ground truth data - we do not know the true phase and therefore cannot compare. However, since our ultimate goal is to provide sonic feedback that a person will perceive as synchronized, we have relied on the subjective reports of participants in our experiments. We maintain that this is sufficient, particularly because the human ability to perceive synchronization is well documented in science and culture. The ultimate test will be in patient and athlete outcomes that are beyond the scope of this paper.

One of our subjects has requested additional training time with the system. This person, a high-performance athlete, has expressed difficulty in maintaining and self-evaluating his technical efficiency during extreme training efforts. Outside of the scope of our original vision for best uses of rhythmic auditory feedback, the subject notes an interest in “listening” to how his skating technique degrades as fatigue sets in, something he mentions is difficult to currently assess and correct in real-time.

While we have shown how to produce rhythmic feedback that reinforces correct movement, we have yet to establish protocols for clinical treatment, nor for teaching a motor skill to an athlete. With severe gait disorders, a patient might not be able to walk well enough to synchronize with a model gait. In this case, we speculate that a progression of models designed to challenge a patient incrementally might work better than pushing a patient to synchronize with a normal gait. The same may also be true for teaching an athletic motor skill. For example, the double-push might be taught by synchronizing to an intermediate model that exaggerates a glide phase before turning that glide into an extra push.

Currently, clinical apparatus to provide sonic feedback to patients is large and expensive. For example, Motek Medical’s [19] CAREN system requires a major infrastructure installation limiting it to inpatient use. The work presented here offers the potential for sonified feedback for outpa-

tients. A computer, camera and treadmill are well within what is affordable and physically practical for use at home.

4. Conclusion

We have demonstrated a system to synchronize and create rhythmic sonic feedback of human locomotion based on depth/disparity cameras. The system exhibits an ability to synchronize sound to within tolerances for humans to perceive that the sound is synchronized to their motion. The rhythmic feedback reinforces the correct performance of the motion, and opens the door to applications for clinical treatments, and for new training methods for athletes.

References

- [1] B. I. Bertenthal and J. Pinto. Complementary processes in the perception and production of human movements. In L. B. Smith and E. Thelen, editors, *A Dynamic Systems Approach to Development: Applications*, pages 209–239. MIT Press, Cambridge, MA, 1993. 2
- [2] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992. 4
- [3] A. Bidet-Caulet, J. Voisin, O. Bertrand, and P. Fonlupt. Listening to a walking human activates the temporal biological motion area. *NeuroImage*, 28(1):132–139, October 2005. 2
- [4] R. Blake and M. Shiffrar. Perception of human motion. *The Annual Review of Psychology*, 58:47–73, 2007. 2
- [5] J. E. Boyd and A. Godbout. Multidimensional synchronization for rhythmic sonification. In *International Conference on Auditory Display*, pages 68–74, Atlanta, GA, June 2012. 1, 3, 4
- [6] J. E. Boyd, A. Godbout, and C. Thornton. In situ motion capture of speed skating: Escaping the treadmill. In *Computer and Robot Vision (CRV), 2012 Ninth Conference on*, pages 460–467. IEEE, 2012. 4
- [7] I. J. Cox and S. L. Hingorani. An efficient implementation of reid’s multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(2):138–150, 1996. 4
- [8] M. F. del Olmo and J. Cudeiro. Temporal variability of gait in parkinson disease: effects of a rehabilitation programme based on rhythmic sound cues. *Parkinsonism and Related Disorders*, 11(1):25–33, January 2005. 4
- [9] M. Eriksson, K. A. Halvorsen, and L. Gullstrand. Immediate effect of visual and auditory feedback to control the running mechanics of well-trained athletes. *Journal of Sports Sciences*, 29(3):253–262, 2011. PMID: 21170792. 3
- [10] A. Gelb. *Applied optimal estimation*. MIT press, 1974. 5
- [11] A. Godbout and J. E. Boyd. Corrective sonic feedback for speed skating: a case study. In *International Conference on Auditory Display*, pages 23–30, Washington, DC, June 2010. 3, 4
- [12] A. G. Godbout, C. Thornton, and J. E. Boyd. Mobile synchronization for athletes: a case study in commercialization

- of sonification. In *International Conference on Auditory Display*, New York, NY, June 2014. 3, 4
- [13] D. Grest, J. Woetzel, and R. Koch. Nonlinear body pose estimation from depth images. In *Pattern Recognition*, pages 285–292. Springer, 2005. 4
- [14] F. Grond and J. Berger. *The Sonification Handbook*, chapter Parameter mapping sonification, pages 363–397. Logos, 2011. 3
- [15] T. Hermann, A. Hunt, and J. G. Neuhoff, editors. *The Sonification Handbook*. Logos Verlag, Berlin, Germany, 2011. 2
- [16] T. Hermann, B. Ungerechts, H. Toussaint, and M. Grote. Sonification of pressure changes in swimming for analysis and optimization. In *International Conference on Auditory Display*, Atlanta, Georgia, 2012. Georgia Institute of Technology. 3
- [17] J. A. Hockman, W. M. M., and I. Fujinaga. Real-time phase vocoder manipulation by runner’s pace. In *NIME ’09: Proceedings of the 2009 conference on New interfaces for musical expression*, 2009. 3
- [18] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14(2):201–211, 1973. 2
- [19] Motek Medical Inc. CAREN (computer assisted rehabilitation environment). Retrieved April 13, 2015, from <http://www.motekmedical.com/products/caren/>. 8
- [20] C. Nombelaa, L. E. Hughes, A. M. Owenc, and J. A. Grahnc. Into the groove: Can rhythm influence parkinsons disease? *Neuroscience and Biobehavioral Reviews*, 37(10 Pt 2):2564–2570, December 2013. 4
- [21] G. Rizzolatti and L. Craighero. The mirror-neuron system. *Annual Review of Neural Science*, 27:169–192, July 2004. 2
- [22] G. Rizzolatti, L. Fogassi, and V. Gallese. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2:661–670, September 2001. 2
- [23] R. B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1–4. IEEE, 2011. 4
- [24] N. Schaffert and K. Mattes. Interactive sonification in rowing: An application of acoustic feedback for on-water training. *IEEE Multimedia*, 22(1):58–67, January 2015. 3
- [25] N. Schaffert, K. Mattes, and A. O. Effenberg. Examining effects of acoustic feedback on perception and modification of movement patterns in on-water rowing training. In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, AM ’11, pages 122–129, New York, NY, USA, 2011. ACM. 3
- [26] Sonic Golf Inc. Sonic golf. <http://www.sonicgolf.com/>. 3
- [27] J. Stienstra, K. Overbeeke, and S. Wensveen. Embodying complexity through movement sonification: case study on empowering the speed-skater. In *Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity*, CHIItaly, pages 39–44, New York, NY, USA, 2011. ACM. 3
- [28] M. Takahata, K. Shiraki, Y. Sakane, and Y. Takebayashi. Sound feedback for powerful karate training. In *NIME ’04: Proceedings of the 2004 conference on New interfaces for musical expression*, pages 13–18, Singapore, Singapore, 2004. National University of Singapore. 3
- [29] Under Armour, Inc. Retrieved January 20, 2014, from <http://www.underarmour.com/>. 3
- [30] Motion capture systems from vicon. Retrieved January 23, 2012, from <http://www.vicon.com>. 1
- [31] A. M. Willems, A. Nieuwboer, F. Chavret, K. Desloovere, R. Dom, L. Rochester, D. Jones, G. Kwakkel, and E. van Wegen. The use of rhythmic auditory cues to influence gait in patients with parkinson’s disease, the differential effect for freezers and non-freezers, an explorative study. *Disability and Rehabilitation*, 28(11):721–728, June 2006. 4
- [32] Wolf, Peter, Sigrist, Roland, Rauter, Georg, and Riener, Robert. Error sonification of a complex motor task. *BIO Web of Conferences*, 1:00098, 2011. 3
- [33] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International journal of computer vision*, 13(2):119–152, 1994. 4