A Phase-Entrained Particle Filter for Human Motion Synchronization

Manikashani, Peyman

http://hdl.handle.net/11023/2463

Downloaded from PRISM Repository, University of Calgary
Abstract

Humans have a specialized ability to synchronize their movements with a perceived auditory feedback. The association between human locomotion and sound feedback is especially profound when the movement has a periodic nature and the feedback is rhythmic. Dancing or marching with rhythmic music are examples of such association. This thesis presents a human motion feedback system capable of generating rhythmic auditory feedback synchronized to a periodic movement. We demonstrate the synchronization in the context of an everyday, yet complex example of human locomotion: normal walking gait. Our system relies on visual measurements provided by a conventional RGB camera and utilizes a phase-entrained particle filter to synchronize an oscillator with a subject’s movement in real-time. A long-term prospect of this research is that rhythmic sonic feedback in sync with walking gaits has potential to be used in therapeutic applications for clinical treatment of gait disorders.
Acknowledgements

My most sincere thanks to my family,
whose support and encouragement have always motivated me to follow my goals,
and to my supervisor, Dr. Jeffrey E. Boyd,
whose exceptional guidance and support encouraged development of this thesis.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>i</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>ii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation and Problem Statement</td>
<td>1</td>
</tr>
<tr>
<td>1.2 System Overview</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Contributions</td>
<td>3</td>
</tr>
<tr>
<td>1.4 Collaborations</td>
<td>5</td>
</tr>
<tr>
<td>1.5 Road Map</td>
<td>5</td>
</tr>
<tr>
<td>2 Background</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Human Motion Perception</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Rhythmic Auditory Feedback and Synchronization</td>
<td>10</td>
</tr>
<tr>
<td>2.3 Auditory displays and Sonification</td>
<td>13</td>
</tr>
<tr>
<td>2.4 Human Motion Acquisition</td>
<td>18</td>
</tr>
<tr>
<td>3 Temporal Estimation Fundamentals</td>
<td>21</td>
</tr>
<tr>
<td>3.1 Stochastic State Estimation Framework</td>
<td>21</td>
</tr>
<tr>
<td>3.2 Bayesian Estimation Framework</td>
<td>23</td>
</tr>
<tr>
<td>3.2.1 Statistical Preliminaries</td>
<td>24</td>
</tr>
<tr>
<td>3.2.2 Recursive Bayesian Filtering</td>
<td>25</td>
</tr>
<tr>
<td>3.3 Kalman Filter</td>
<td>28</td>
</tr>
<tr>
<td>3.3.1 Problem Definition</td>
<td>28</td>
</tr>
<tr>
<td>3.3.2 Propagation Mechanism</td>
<td>29</td>
</tr>
<tr>
<td>3.4 Summary</td>
<td>33</td>
</tr>
<tr>
<td>4 Synchronization and Rhythmic Sonification for Periodic Motor Movements</td>
<td>34</td>
</tr>
<tr>
<td>4.1 Walking Gait Example</td>
<td>35</td>
</tr>
<tr>
<td>4.2 Visual Localization and Phase Estimation</td>
<td>36</td>
</tr>
<tr>
<td>4.2.1 Condensation Framework</td>
<td>36</td>
</tr>
<tr>
<td>4.2.2 Phase-Entrained Particle Filter</td>
<td>41</td>
</tr>
<tr>
<td>4.3 Rhythmic Sonification</td>
<td>46</td>
</tr>
<tr>
<td>5 Implementation</td>
<td>49</td>
</tr>
<tr>
<td>5.1 Segmentation</td>
<td>49</td>
</tr>
<tr>
<td>5.2 Hausdorff Distance Computation</td>
<td>50</td>
</tr>
<tr>
<td>5.3 Sonification Schemes</td>
<td>53</td>
</tr>
<tr>
<td>5.4 System Dynamics Control Parameters</td>
<td>54</td>
</tr>
<tr>
<td>5.5 Initialization</td>
<td>55</td>
</tr>
<tr>
<td>5.6 General System Development</td>
<td>56</td>
</tr>
<tr>
<td>6 Testing</td>
<td>60</td>
</tr>
<tr>
<td>6.1 Walking Gait Experiment</td>
<td>60</td>
</tr>
<tr>
<td>6.2 Observations</td>
<td>61</td>
</tr>
<tr>
<td>6.2.1 Model vs Model</td>
<td>61</td>
</tr>
</tbody>
</table>
List of Tables

6.1 Walking Gait Experiment Statistical Results ......................... 66
List of Figures and Illustrations

1.1 Motion Feedback Process Overview ............................................. 2
2.1 The Necker Cube Illusion .................................................. 7
2.2 Point Light Display Transformation ........................................ 8
2.3 Auditory Display as a Musical Instrument .................................. 14
3.1 Stochastic Filtering Diagram ............................................... 22
3.2 Prior Versus Posterior Probability Distribution ......................... 24
3.3 The Kalman Filter Propagation Process ................................... 31
4.1 Synchronization Method and Walking Gait Experiment Setup ............. 36
4.2 Factored Sampling Concept ............................................... 38
4.3 The Propagation Process in the CONDENSATION ...................... 40
4.4 Phase-Entrained Particle Filter’s State Vector .............................. 41
4.5 Hausdorff Distance Concept ............................................ 45
4.6 The Scale Parameter’s Role in Particle’s Weighting Function .......... 45
4.7 Rhythmic Sonification Process .......................................... 47
5.1 Segmentation .............................................................. 51
5.2 Hausdorff Distance Implementation ...................................... 52
5.3 Sonification Schemes for Walking Gait Experiment .................... 54
5.4 Phase Drift and Diffusion ............................................... 55
5.5 Initialization .............................................................. 56
5.6 Walking Gait Model Silhouettes ......................................... 58
5.7 Pure Data Sonification Patch ............................................ 59
6.1 Walking Experiment Estimated Phase Signals ........................... 64
6.2 Walking Experiment Scattered Plots of Estimated Phase Signals .......... 65
6.3 Auditory Feedback’s Timing Error ...................................... 66
Chapter 1

Introduction

1.1 Motivation and Problem Statement

The human brain has developed highly specialized skills for analyzing sounds and extracting valuable information from auditory cues. Music perception is among the most compelling examples that is evident in humans from early infancy when mothers sing lullabies to their infants. As a key component of musicality, rhythm represents the temporal structure of a musical piece and it is often expressed through periodic beats. Given the critical role of temporal relations in coordinating human motor movements, such periodic pulses not only enhance the musical experiences, but also provide the brain with an elegant tool well suited for guiding motor skills. The association between auditory rhythms and human locomotion is profound and prevalent through numerous examples such as dancing, marching or exercising in sync with rhythmic music. Utilizing this association, auditory rhythms have the capacity to provide an assistive mechanism that facilitates learning or improving certain repetitive motor movements. This capacity motivates development of this thesis.

Periodicity is an inherent characteristic of most human locomotion. We exploit this feature to build a human motion feedback system capable of generating rhythmic auditory cues synchronized to a periodic movement. We demonstrate the effectiveness of our system in the context of an everyday, yet complex example of a periodic motor skill: normal walking gait. An inspirational benefit of choosing this example is that rhythmic auditory feedback synchronized to gaits, has potential to enhance clinical treatments of gait disorders. Although evaluating the effectiveness of our system in clinical settings is beyond the scope of this thesis, the therapeutic applications are a long-term prospect of this research.
Figure 1.1: The high-level overview of the motion feedback process. The computer model exploits real-time visual measurements, provided by a conventional camera, to synchronize its motion with subject’s movement. Consequently, our system utilizes the synchronization data to generate auditory feedback in form of a synchronous sonic rhythm.

1.2 System Overview

Figure 1.1 illustrates the high-level concept of the motion feedback process. Our system uses a conventional camera for obtaining real-time visual measurements of a subject performing a periodic movement. A synthetic computer model utilizes the acquired measurements to synchronize an oscillator with a subject’s movement. Consequently, the system generates feedback in the form of an auditory rhythm by producing a non-verbal sonic representation of the synchronization data – a process we refer to as rhythmic sonification. Since the whole process takes place in real-time, as the computer model uses measurements to synchronize its motion with the subject, the subject has the opportunity to exploit the provided feedback to synchronize his motion with the synthetic model. This mutual synchronization process helps the subject to adjust and improve his movement based on the provided feedback.
Our process for generating synchronized motion feedback has the following two main steps that are described in depth in Chapter 4:

1. **Synchronization.** Since we are acquiring measurements from a visual sensor, i.e. a conventional camera, localization of the movement at each frame of the input video sequence is a necessity in our synchronization process. In addition, for the purpose of generating synchronized feedback, we exploit the temporal relations inherent in human movements. In the case of periodic movements, the key to accessing such temporal information is embedded in the phase of the movement—defined as the elapsed fraction of a period of the motion. As a result, we are interested in a mechanism that is capable of estimating both location and phase during the course of the movement. Section 4.2.2 describes our proposed mechanism: the phase-entrained particle filter.

2. **Sonification.** As the synchronization step produces the phase estimates, the sonification process exploits the estimated results to generate auditory feedback in sync with the phase data. Given the common periodic nature of both rhythm and human movement, we employ a rhythmic sonification strategy to generate the feedback as an auditory rhythm synchronized to the movement. Section 4.3 describes the details of the sonification process.

### 1.3 Contributions

In this thesis, we show by virtue of a demonstration that visual measurements of a human locomotion suffice for synchronizing a periodic motor movement against a computer model of that movement. Our synchronization method utilizes a particle filter to track a person while synchronizing an oscillator to the cycle of the gait. This synchronization process is referred to as “phase entrainment”. Thus we refer to this method as a “phase-entrained particle filter”. Additionally, based on this synchronization, we use an existing sonification system to generate rhythmic auditory feedback in sync with the movement. Such feedback has
potential to be utilized for training or clinical purposes and facilitate learning or improving a periodic motor movement.

Our motion feedback system has the following significant features:

1. **Real-time.** The system is capable of generating auditory feedback synchronized to the movement in real-time.

2. **Reliability.** The produced feedback provides a reliable auditory representation of progression through the movement.

3. **Affordability.** The system avoids the costly motion capture technologies by using a conventional camera for obtaining visual measurements.

4. **Availability.** The prevalence of conventional cameras in commonplace technologies, such as smartphones, enhances the availability of our system.

5. **Non-intrusiveness.** Unlike most wearable motion capture technologies, the visual measurements do not interfere with user’s performance during the movement.

In conventional visual approaches towards human motion analysis, the human pose is often defined as a relative configuration of body parts that is represented through joint angles and positions. Many pose estimation strategies require using multiple cameras for obtaining complex measurements from the human subject during action. For periodic movements, our synchronization process introduces an alternative view of human pose in the context of phase and uses only a single camera to obtain measurements. This view arises from the fact that each period of the movement can be discretized by a finite sequence of poses and each pose can be represented by a single phase value. As a result, estimating the phase of the movement at a given time is in essence equivalent to finding the associated pose at that time.
1.4 Collaborations

The original idea of this thesis, using sound feedback for training purposes in periodic motor movements, was suggested to me by Dr. Boyd, my supervisor. Prior research in the field of sonification, “Corrective Sonic Feedback For Speed Skating: A Case Study” [10], conducted by Godbout and Boyd, provided key inspirations for development of this thesis. During my studies as a Master’s student at University of Calgary, I had the opportunity to consult with Mr. Godbout, my laboratory colleague, and utilize his feedback together with Dr. Boyd’s feedback to improve the quality of my research. While I developed the whole synchronization system independently in C++, for development of sonification system in Pure Data, I collaborated with Dr. Boyd who provided me with helpful Pure Data patches for rhythmic sonification.

1.5 Road Map

Chapter 2 reviews prior research in the field of human motion perception and discusses the implications of rhythmic feedback and synchronization in such perception. We continue with a review of inspiring sonification strategies employed in selected studies and compare different motion acquisition technologies to justify our choice of conventional cameras for collecting visual measurements in our motion feedback system. In Chapter 3 we present the general temporal estimation framework in the context of recursive Bayesian filtering problem and review a well-known optimized solution of this problem in certain circumstances: the Kalman filter. Chapter 4 presents the details of our synchronization method, the phase-entrained particle filter, and describes a rhythmic sonification process for transforming the synchronization results (phase estimates) into rhythmic auditory cues. In Chapter 5 we elaborate the implementation details of our system. Chapter 6 demonstrates the results of evaluating our motion feedback system in the context of the walking gait experiment. Finally, Chapter 7 provides a conclusion of our study together with future directions of this research.
Chapter 2

Background

2.1 Human Motion Perception

In the design of human motion feedback systems, it is crucial to understand how humans perceive motion. Although this thesis does not focus on cognitive neuroscience or psychology, the results of extensive investigations in human motion perception from such disciplines, provide significant insights for development of our motion feedback system. This section reviews some fundamental studies on human motion perception.

Human sensory modalities provide the brain with the primary means of perceiving events and objects it encounters in everyday life. However, these modalities are not totally independent and indeed, their integration often plays a substantial role in forming a coherent and comprehensive perception of the world [57]. For example, speech perception is a common daily routine for humans that involves a multisensory integration process. In this routine, the visual sense associated to reading lips interacts with the auditory sense of hearing [75]. This interaction is generally beneficial and often results in a more coherent perception of speech. Nonetheless, there are instances where such integration results in an illusory perception such as a phenomenon known as the McGurk effect. McGurk and MacDonland [60] observed that when normal adults hear the auditory syllable “ba” dubbed on a video of a mouth movement associated to “ga”, they report hearing the syllable “da”. However, when they only hear the sound and do not see the video, they correctly report hearing “ba”.

Although sensory modalities have an important contribution in building perceptual experiences, human perception is not limited to the sensory inputs. As Jean-Jacques Rousseau [76] elegantly puts it, “Our sensations are purely passive, while all our perceptions or ideas are born out of an active principle which judges”. Berthoz [12] argues that perception is an
“internal simulation of action” and it goes beyond our sensory inputs. In a sense, he views perception as the brain’s hypothesis about an external stimuli rather than the exact signals received through our senses. Perceptual illusions are especially interesting from this point of view, because they take place when our internal perception fails to capture the reality of the external object or event as observed by our senses. Figure 2.1 depicts a visual illusion example, known as the Necker cube [65], where the designated point on the corner can either be perceived on the inner or outer face of the cube. An intriguing category of perceptual illusions in the study of human motion perception is the illusory self-motion, known as vection. In this illusion, our brain postulates self-movement when we are in fact stationary. A compelling example of vection is the “train illusion”; when we are in a stationary train, watching a nearby moving train in an adjacent track can induce a perception that our own train is moving. Self-motion illusion is not limited to the sense of vision. Several studies have reported similar effects in terms of auditory vection [89]. For instance, a spatialized rotating auditory cue around subjects may result in the self-rotation or circular vection perception [56].

It has long been known that rapid succession of still images that contain same objects with slight displacement, can produce the perception of a smooth motion. This visual illusion is referred to as Apparent Motion and is commonly used in creation of animations in
videos. More than 40 years ago, Gunnar Johansson’s fundamental study of Point-Light (PL) displays [52], opened new horizons in the field of human motion perception through vision. In his study, he attached small light bulbs to the main joints of a walking actor dressed in black. Filming the movement with an applied video technique, created an animation of bright set of point lights moving against a black background. The results of the study showed that the subjects who observed the PL animation were able to quickly perceive a walking human figure after the initial steps of the actor. However, they did not perceive the static PL display as a human figure. Figure 2.2 illustrates the concept of PL display transformation. Blake and Shiffrar [15] present evidence in support of the reliability of human motion perception in spite of perturbations in the pattern of PL displays. Several years of research in this area suggests that humans have a distinctive specialization in recognizing motion and particularly human movements. Various studies suggest that this expertise is indeed innate, such that one-month-old infants have a preference for a moving visual cue over a stationary one [19] and infants of four months of age or older pay longer attention to PL displays associated to human movements in comparison to random lights [10, 34]. Despite the general stability of the perception of a walking human figure in PL displays, an
important observation reveals that spatiotemporal perturbation in the phasic structure of
the motion, is among the few confounding factors that can disturb such perception through
a PL display animation [11, 43, 46].

In an apparent motion animation of human movements, if the timing between the onsets
of the frames is long enough, subjects have a tendency to perceive the direction of motion in
forms of an “anatomically plausible” path [79, 80]. As Blake and Shiffrar [15] discuss, human
experience in performing one’s own movements, plays a part in shaping one’s expertise for
perceiving others’ movements. Hommel et al. [48] describe this idea in their “Theory of Event
Coding (TEC)”, which implies that perception of an action and planning an action are in-
tertwined and they share a common code or sensory representation. Support for this theory
comes from neurophysiological evidence. Electrophysiological recordings in monkeys show
that a special group of sensorimotor neurons known as the “Mirror Neurons” fire both when
monkeys perform an action and when they observe that action in others [71, 72]. Increas-
ingly, functional Magnetic Resonance Imagining (fMRI) also reveals that neural resources
involved during action planning and action perception have partial intersections [28].

The neurophysiological support for TEC is not limited to the sense of vision. Bidet-Caulet
et al. [14] investigated the neural circuits involved in hearing the sound of walking. They
observed that auditory cues associated to footfalls provoked an activation in “the temporal
biological motion area of the brain” which was previously known to be associated to visual
stimuli. Not surprisingly, the cross-modality interactions in the brain is not even limited to
the audio-visual interactions. For instance, Graziano and Gross [42] discuss that the neurons
which fire when a specific part of a monkey’s body is touched, also fire when a visual stimuli
is present in a nearby area of that specific limb.
2.2 Rhythmic Auditory Feedback and Synchronization

As mentioned in Section 2.1, perturbations in the timing of the PL apparent motion have potential to cause confusion in perception of the human motion [15]. This observation suggests that the temporal information embedded in a movement, plays a critical role in understanding the dynamics of that movement. Humans are sensitive to such temporal relations. Rhythm perception is a notable example. Thaut [87] describes rhythm as the “explicit divisions of time or space into intervallic time systems, recurrent and often (but not always) characterized by periodicity”. From this perspective, rhythm can be considered as the prime medium for encoding and transferring the temporal information in a periodic system. Our brain has indeed developed highly specialized skills for perceiving rhythms. For instance, the auditory system in the brain allows for rapid detection of the temporal relations embedded in an audio signal [64]. Rhythm, from another point of view, is a crucial component of musicality. Human expertise in perceiving music and rhythmic melodies is emergent from early infancy [70]. Vocal rhythmic melodies and lullabies that mothers sing for their infants are profound examples [51]. Scientific findings support that infants have a preference for rhythmic melodies over non-rhythmic ones [88]. To realize the importance of music perception in brain, several studies point out to linkages between music perception and the fundamental task of language processing, and discuss the connections in terms of similar regions in the brain that are involved during both tasks [58, 54].

The association between rhythms and synchronization is profound. Our daily activities are replete with examples of such association. Dancing to a beat, marching with a rhythm or exercising with music are just a few examples. Given the periodic nature of human locomotion, as shown by McGeer [59], and periodicity of rhythms, this association may not be surprising for humans. However, the synchronous behavior is also apparent in animals through many examples, such as synchronous chorusing of some insects or frogs [39]. Even some fireflies have exhibited synchronous behavior by flashing at a rhythmic constant pace [86].
of these examples, there is one crucial intersection: rhythm. In fact, creation and perception of rhythms play a critical role in both encoding and conveying the temporal information required for a synchronous behavior in biological systems. However, synchronization is not even limited to biological systems. For instance, the phase lock loop [13] (PLL) is a well-known mechanism for synchronizing two oscillators by adjusting the frequency of an internal oscillator to an external one.

The evidence for correlation between rhythmic auditory cues and sensorimotor responses in humans is strong and abundant through different disciplines. From neuroscience, there is evidence in support of linkages between rhythmic audio and motor areas in the brain. The human brain exhibits periodic neural oscillations at different frequencies. Fujioka et al. [36] observed that when healthy adults with no intention to move were presented with a rhythmic auditory cue, modulations in beta-band frequencies, $\beta$ ($\approx 20$ Hz), which are known to have correspondence with motor activities (see Pfurtscheller and Da Silva [67]), also showed correspondence to the auditory rhythm. Using electroencephalography (EEG) recordings, Snyder and Large [84] showed that when subjects were listening to a rhythmic auditory pattern with an occasional missed tone, the evoked phase-locked gamma-band activity, $\gamma$ (20-60 Hz), corresponded to the played pattern and was weakened at the omissions. Another study [35] suggests that both $\beta$ and $\gamma$ oscillations play distinct roles during perception of rhythms and interactions of audio and motor movements. In research conducted by Chen et al. [25], subjects were asked to listen to a rhythmic music and were later asked to tap in synchrony to that rhythm in the context of two experiments. In one experiment they knew that they would later be asked to tap along with music. However, in the other experiment, they were not anticipating such request. The results of fMRI analysis in both experiments revealed neural activity in motor regions of the brain when listening to the rhythmic audio before tapping, regardless of subjects’ anticipation of tapping.

Synchronization is of principle importance in musical performances as well, and rhythmic
feedback plays a central role in achieving it. Goebl and Plamer [41] studied the effects of auditory feedback among pianists in an ensemble performance. Two pianists were asked to perform a duet such that the leader played the higher notes as a soloist and, the follower played the lower notes as an accompanist. Three type of feedbacks were presented to the pianists: full-feedback such that both pianists could hear each other playing, half-feedback such that leader could only hear himself and follower could hear both leader and himself playing, and self-feedback such that the pianists could only hear themselves. The results indicated that reduction in feedback led to asynchronous performance and also as the auditory feedback decreased, pianists showed a tendency towards synchronizing visual cues such as head movements instead of auditory cues. Dancing is another example that requires synchronization between perceived auditory rhythms and motor movements. Although young children may have difficulties in adjusting their movements to a rhythmic beat [69], there is evidence in support of their inclination towards rhythmic movements. In a study by Eerola et al. [30], two to four years old children were presented with a rhythmic music that contained sudden tempo changes. In the meantime, children’s movements were monitored by a marker-based optical motion capture system. The statistical analysis of motion data revealed that the movement pattern of children had a periodic characteristic which was at times in sync with the played music.

Music therapy and clinical treatments in motor rehabilitation is an important application that arises from the correlation between motor movements and rhythmic auditory cues. For instance, Bradt et al. [20] suggests that rhythmic auditory cues may help in recovery of stroke patients by improving their gait velocity, strength, cadence and symmetry. Given the periodic nature of gaits, numerous studies investigate the effectiveness of rhythmic auditory feedback in treatment of gait disorders especially in Parkinson’s patients [29, 66]. Sacks [77] mentions a number of stories in regards to patients who have recovered their motor abilities through music therapy, such as a violin concerto record that helped a man regain his walking
ability after an injury. The literature in music therapy applications is vast, Davis, Gfeller and Thaught [27] provide an overview and evolution of music therapy over several years of research in this area.

2.3 Auditory displays and Sonification

Auditory display research investigates how to exploit the capabilities of the human auditory system so that audio can serve as the principal means of data representation while reflecting the associated changes in the data infrastructure. As Figure 2.3 illustrates, information source, information communicator and information receiver are the key components of an auditory display system [55]. The sonification process plays a central role in auditory displays by determining the underlying principles of transforming data into sound [45]. Sonification research is indeed highly interdisciplinary and covers the interactions between numerous fields such as music, computer science, cognitive sciences, psychology, linguistics and audio engineering.

As Figure 2.3 shows, a classic example of sonification is the translation of music scores into actual musical pieces, when the visual symbols of musical notes with certain frequencies will be sonified as musical tones or pitches. In this process, the musical instrument is in essence a medium for transforming player’s gestures to sonic pitches. From this perspective, the combination of haptic interactions with auditory displays can be viewed as a musical instrument [90].

Kramer [55] mentions a number of benefits for auditory displays. The following advantages are of particular interest to us for the purpose of developing our motion feedback system for certain human movements:

1. **No interference with visual perception:** This feature is especially helpful for performance in active tasks such as sports, where visual analysis of events is of prime importance in action.
Figure 2.3: Main components of an auditory display system [55]: information source (e.g. a musical score), information communicator (e.g. a musical instrument) and information receiver (e.g. a person listening to music).

2. **Accurate representation of temporal relations in data:** Given the human auditory system’s high sensitivity to rhythms, as described in Section 2.2, this feature provides the opportunity to exploit such innate skills through rhythmic sonification paradigms.

3. **Concurrent broadcast of auditory streams through different channels:** Since humans can listen to a mixture of input auditory signals in parallel, this feature allows for a more holistic and comprehensive perception of information. Furthermore, given humans’ synchronization abilities, concurrent broadcast has potentials in reinforcing synchronous behavior in certain tasks that require real-time synchronization.

There are different approaches and techniques for sonification. Blattner et al. [16] explain the concept of *earcons* in the context of interactions between users and computers, especially in graphical user interface (GUI) domains, with the goal of conveying feedback to the user. To create earcons, they propose using *motives* in the form of short sequential pitches with a rhythmic pattern. Some examples include the sounds associated to arrival of new messages and e-mails, operating system start up and shut down, and error messages due to failure in certain operations. Simple construction and abstract representation, such that association between sounds and object features is not essential, are among advantages of earcons. However, learning auditory patterns as they get longer and more complex may be problematic [9].
Gaver [37] studied the use of sounds in computer interfaces through auditory icons. In contrast to earcons, as Gaver explains, auditory icons take advantage of intuitive sounds in correspondence to the nature of events or objects. The file deletion sound, resulted by putting an item in the recycle bin in operating systems, is a compelling example of auditory icons. Although auditory icons have the advantage of being familiar to the user, they still may face problems similar to earcons when it comes to ease of learning auditory cues that are associated to virtual tasks with no inherent auditory feature [9].

Parameter Mapping is a dominant sonification technique which refers to mapping information to an auditory parameter such as intensity, pitch or duration [45]. Unpleasantness of generated sounds is a potential drawback for this technique [9]. The examples of parameter mapping are numerous. For instance, in a study conducted by Effenberg [31], subjects were presented with visual, auditory and audio-visual representations of counter movement jumps (CMJs) respectively in form of video, sonified audio or both cues. They were then asked to estimate the height of perceived CMJs and reproduce them. Sonification process involved measuring the exerted force during jumps using a force plate and converting the measurements to amplitude and frequency modulations. Although some subjects did not like the sound, the statistical analysis of the results showed that the combined audiovisual feedback resulted in least absolute error for estimations.

In general, sounds often provide valuable information in sports. For example, the intensity of a ball kick sound may be an indication of the level of strength exerted in shooting. Similarly, the sound that a golf shot produces has potential to provide valuable information to the golfers in regards to the shot or cue characteristics and feeling [73]. As previously mentioned, sonification has an advantage of not interfering with visual stimuli. This benefit is especially tempting for applying sonification in sports, where visual processing of events is critical. Additionally, given the crucial importance of timing in sports, the usefulness of athletic sonic feedback can be further enhanced if it reflects the temporal information of the
activity as well. As previously discussed, rhythm provides an elegant way of incorporating such temporal relations into the sonic feedback.

In the context of running, Hockman et al. [47] designed a real-time auditory feedback system in sync with a runner’s pace. To estimate the frequency of strides, they interpreted the data collected by an accelerometer that was attached to the runner’s shoes. The estimated pace then served as a time scale controller of a phase vocoder, by means of which, they could synchronize the pace of running to the tempo of a generated music. In another study, Wijnalda et al. [92] also employed sensors to measure and monitor dynamic features such as heart rate or pace, and consequently utilized the measurements for adjusting or selecting music.

In the sport of rowing, Schaffert et al. [78] applied a parameter mapping scheme for sonifying the acceleration in the direction of boat movement. They mapped the changes in acceleration to pitch modulations, such that increased acceleration resulted in higher frequencies, and communicated online sonic feedback to the rowers. The periodic nature of rowing and boat acceleration trends lend itself nicely to the concept of rhythms and as the authors discuss, the constructed sonic feedback helps towards an enhanced realization of the rhythm of the actual movement.

Stienstra et al. [85] investigated the effectiveness of sonification by parameter mapping in the context of speed skating. They acquired force measurements from sensors attached to the skates and interpreted the acquired data to extract the desired movement characteristics such as speed, acceleration and changes in acceleration. They conveyed sonic feedbacks to the speed skater in a continuous manner by means of modulations in a pink noise based on interpreted data. The results indicated improvements in the studied speed skating experience in presence of sonic feedback.

In another speed skating example, which was especially inspiring for development of this thesis, Godbout and Boyd [40] demonstrated a rhythmic sonification method for improving
the cross-over technique that is performed in corners of a speed skating oval. Their subject was an accomplished speed skater who had lost his ability to perform a proper cross-over, due to a condition known as “Lost Move Syndrome”. The proposed method by the authors involved a model based real-time comparison between a perfect cross-over as the model, performed by another athlete, and the subject’s cross-over. The model movement consisted of a set of measurements representing the ankle extension during a period of the cross-over. The model was then compared to the subject’s ankle angle measurements in real-time, derived from a wearable sensor carried by the subject. The synchronization process involved synchronizing subject’s stride to a model stride by means of a cross correlation technique. As for the sonification, the perfect matching between subject and model cross-over resulted in a rhythmic sonic feedback. Nonetheless, in the case of large differences between model and subject ankle extensions, a sawtooth tone with an intensity proportional to the amount of error was communicated to the subject before footfalls. The authors found this method helpful for the subject in making improvements towards recovering from the lost ability.

Boyd and Sadikali [18] investigated a synchronization and sonification paradigm for human gaits without using wearable sensors. They used a conventional camera to capture the video of a person walking on a treadmill. For the synchronization, they exploited Boyd’s [17] video phase locked loop (VPLL) method, which involved an array of independent phase locked loops (PLLs) corresponding to pixel intensities in the captured video image. The periodic nature of gaits led to the phase locking of all PLLs in the VPLL to the fundamental gait frequency. Their approach resulted in phase images constructed from the phase measurement array. To sonify the phase data, they generated a rhythmic auditory pattern constructed from percussion sounds that were triggered in coincidence with salient poses through the movement.

The sonification literature is replete with applications in sports. In the aforementioned examples, we reviewed a few studies employing synchronized sonic feedback in different sports
such as running [47], rowing [78], speed skating [40, 85], and walking [18]. An important outcome of these studies for our motion feedback system is that a sonification that conveys temporal relations of a movement, especially in terms of rhythms, can produce effective feedback for learning certain periodic motor skills.

2.4 Human Motion Acquisition

Motion capture systems are diverse both from underlying technology and application point of view. The diversity in applications includes, but is not limited to, control examples such as interactive gaming interfaces and virtual environments, or analysis examples such as clinical treatment and athletic performance improvements [62, 63]. In spite of the variety of motion capture technologies, it is worthwhile to mention that there is no best choice of technology for all scenarios, and each technique has its own cons and pros in any situation. This section presents a brief review of the most notable technologies in the field of human motion acquisition and justifies our choice for our motion feedback system.

Optical motion capture systems, as described in [82, 44], typically utilize two or more cameras, and a set of small reflective markers mounted on the relevant parts of the body. The setup in such systems allows for an accurate estimate of 3D position by triangulation. In general, calibration is required in the initialization stage. In these systems, usually a virtual skeleton model represents the kinematic model of human body. Vicon [8] and PhaseSpace [7] motion capture products are two examples of such systems. Despite their accurate 3D position estimates, such systems may not be the best choice for our purpose for the following reasons. First, due to the number of cameras and requirements for calibration, portability is a major problem with optical motion capture systems. Second, the necessity to wear markers can limit performance in active tasks. Third, optical motion capture systems are usually expensive. Although these systems work well in laboratory environments, the aforementioned issues restrict the consumers who may seek to use our system for general rehabilitation
purposes outside a laboratory.

Inertial measure units (IMUs) are another group of systems used for capturing motion. They operate by means of sensors attached to some parts of the body and take advantage of accelerometers and gyroscopes for measuring movements. Xsens MVN [71] is an example of such systems. Drift is a major concern with IMUs due to accumulation of gyroscope errors.

Mechanical motion capture suits are another group of sensor-based systems that are basically exoskeletons worn by the subjects. The Gypsy 7 motion capture system by Meta Motion [6] is an example. A major problem with using mechanical suits for our purpose is the restriction of a subject’s movement.

The advent of motion capture technologies in gaming industry, such as Microsoft Kinect [61], has opened new doors to computer vision applications due to the industrial push for technological advancements, while reducing the prices to an affordable range. Shotton et. al [81] presented a method for real-time prediction of 3-D human body poses using the depth data provided by Microsoft Kinect. Their method uses a large training dataset that reduces the challenging pose estimation problem to a simpler classification problem. Although Microsoft Kinect is a tempting choice for our motion feedback system, it may not be our best choice due its limited viewpoint.

Marker-less motion capture strategies using conventional cameras are especially enticing for the purpose of our motion feedback system. They do not require using any wearable sensors or markers, and thus do not interfere with the subject’s performance during the movement. These strategies can be classified as two major categories of model-free or model-based approaches. While the former does not involve using any a-priori model, the latter relies on some sort of a human model [62]. Brand [21] presented a model-free approach for finding the 3-D body pose from 2-D silhouettes of a movement without using any articulated model of the body. He employed a machine learning technique for training a hidden Markov model based on 2-D silhouettes and then determined the 3-D pose as the path with maximum
Gavrila and Davis [38] presented a model-based approach for tracking the 3-D pose of a human movement using multiple conventional cameras for capturing the motion. They used edge features of the multi-view images and modeled the body as segments of superquadrics. They formulated the 3-D pose recovery problem as a search problem that involved finding the pose parameters of the model that best matched the actual subject. Similarly, Bregler and Malik [22] presented a model-based movement tracking method for recovering the 3-D pose of an articulated human configuration by introducing and utilizing twist motion models and exponential maps. They used conventional cameras for capturing the motion and ellipsoids for modeling the human body. Their method required initialization of the 3-D model at the first frame and worked for either single or multiple cameras.

For the purpose of our motion feedback system, we are seeking an efficient, affordable and non-intrusive technology that is available to general consumers. Given the brief review of motion capture technologies presented in this section, we believe conventional cameras are the best choice for the specific needs of our application. Chapter 4 describes how we employ a single conventional camera for collecting visual measurements for our motion feedback system.
Chapter 3

Temporal Estimation Fundamentals

In development of our human motion feedback system using conventional cameras, we need to localize the subject at each frame of the input video sequence to extract the required motion information for the synchronization process. A rudimentary way of achieving such localization is to apply successive detections by searching each frame of the captured video sequence to identify the regions or objects of interest. However, this approach is prone to causing instabilities in our synchronization process due to inaccurate detections or missing the target in absence of desirable conditions. As a result, this strategy is impractical for our purpose. In general, a major drawback of consecutive detections that can in turn exacerbate inaccuracies is neglecting the temporal information available through previous frames. To overcome this issue, we seek a mechanism capable of representing and propagating the knowledge about a particular visual entity through time. That is why we employ a temporal estimation strategy instead of serial detections. This section provides an introduction to the principal concepts of temporal estimation in the context of recursive Bayesian filtering and reviews a seminal solution to this problem in certain circumstances, known as the Kalman filter.

3.1 Stochastic State Estimation Framework

The state estimation problem, sometimes referred to as filtering, involves propagation of certain knowledge about a particular target object through time. The information to be propagated about the target is often modeled as an $N_x$ dimensional vector $\mathbf{x} = \{x_1, x_2, ..., x_{N_x}\}$ referred to as the state vector, where $x_{1:N_x}$ represent the desired aspects of the information such as position, velocity, acceleration, size, color or any other quantity of interest.
In a computer, in the propagation process time must be modeled as a discrete entity \( t \) \[50\]. In the filtering problem, as illustrated in Figure 3.1, the main objective is to approximate the state \( x_{t+1} \) of the modeled object at time \( t + 1 \), given the state vector \( x_t \) at time \( t \), together with a set of \( N_z \) dimensional measurements \( z_{1:t+1} \) from time 1 to \( t + 1 \). A dynamic model and a measurement model are two key components of the temporal estimation system. A dynamic model, also known as temporal model, is a potentially time-varying function that encapsulates the knowledge about the system dynamics such as motion kinematics, e.g. constant velocity, constant acceleration, periodic motion, etc. A measurement model, on the other hand, provides a mapping from the state vector space into the measurement vector space and evaluates the likelihood of an observation (or measurement) given a state \[68\].

Before stepping into mathematical details, it is important to distinguish the difference between three categories of state estimation known as filtering, prediction and smoothing \[26\]. The filtering problem involves computing a representation of the conditional probability distribution function \( P(x_t|z_{1:t}) \) over the state \( x_t \), given all the measurements \( z_i \) such that \( 1 \leq i \leq t \). This process is referred to as prediction, when computing the \textit{a priori} probability form of a future state using measurements up to time \( t \). Alternatively, the process is called smoothing, when calculating the \textit{a posteriori} probability form of a past state given present and past observations. In all of these problems, the following assumptions are considered in practice to simplify the computations \[33\]:

\[
\begin{align*}
\mathbf{x}_1 & \quad \cdots \quad \mathbf{x}_{t-1} & \quad \mathbf{x}_t & \quad \mathbf{x}_{t+1} & \quad \cdots \quad \mathbf{x}_n \\
\mathbf{z}_1 & \quad \mathbf{z}_{t-1} & \quad \mathbf{z}_t & \quad \mathbf{z}_{t+1} & \quad \mathbf{z}_n
\end{align*}
\]
1. An observation at time \( t \) depends only on the state at time \( t \) and is conditionally independent of other states. In other words:

\[
P(z_t|x_1, ..., x_n, z_1, ..., z_n) = P(z_t|x_t).
\] (3.1)

2. Each state is conditionally independent of all previous states, except its immediate predecessor. This Markov chain assumption is mathematically expressed as:

\[
P(x_t|x_1, ..., x_{t-1}) = P(x_t|x_{t-1}).
\] (3.2)

In a stochastic filtering system, the dynamic model and the measurement model respectively represent the state transition probability \( P(x_t|x_{t-1}) \) and the measurement likelihood probability \( P(z_t|x_t) \). These components can be mathematically defined by the following vector valued functions [26]:

\[
x_{t+1} = f(x_t, v_t), \quad \text{and}
\]

\[
z_t = g(x_t, u_t),
\] (3.3)

where \( f: \mathbb{R}^{N_x} \to \mathbb{R}^{N_x} \) and \( g: \mathbb{R}^{N_z} \to \mathbb{R}^{N_x} \) are respectively the state and measurement update equations, \( v_t \) is the process noise and \( u_t \) is the measurement noise. Figure 3.1 illustrates the generic graphical model of state transition given the aforementioned assumptions and equations.

### 3.2 Bayesian Estimation Framework

Bayesian conditioning presents one of the most fundamental mechanisms for making inferences and drawing conclusions about world entities and events under uncertainty by fusing new observations and prior knowledge about a particular quantity of interest. This section provides a brief overview of Bayesian filtering based on Chen [26].
3.2.1 Statistical Preliminaries

Before proceeding through the Bayesian filtering problem, it is worthwhile to review some important statistical definitions in Bayesian conditioning:

1. Bayes’ Rule. Given two random variables $X$ and $Z$, let $P(X|Z = z)$ represent the conditional probability distribution function (PDF) of $X$ given the measurement $z$. Since this estimate is achieved after acquiring the observation, it is called the posterior distribution and is calculated based on the Bayes’ rule [24] as follows:

$$P(X|Z) = \frac{P(Z|X) \times P(X)}{P(Z)} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}},$$

where $P(X)$ is called the prior distribution, $P(Z|X)$ is the likelihood and $P(Z)$ is referred to as the evidence. As illustrated in Figure 3.2 in Bayesian temporal estimation approaches, the role of new observations is to reduce the amount of uncertainty in the posterior PDF in comparison to the prior distribution. This effect can be reflected in the shape of posterior distribution in terms of sharper peaks around the true target.

2. Normalization. As a probability theory axiom, any probability density function must
have a unity integral. This condition for the posterior PDF in Bayesian conditioning is
guaranteed by the evidence, which is indeed a normalizing factor that scales the integral
of the PDF to unity. The normalized posterior PDF is computed as:

\[
P(X|Z) = \frac{P(Z|X)P(X)}{\int_X P(Z|X)P(X)dx},
\]

where \( \int_X P(Z|X)P(X)dx \) is the normalization factor or evidence.

3. **Marginalization.** Having the posterior joint likelihood of two parameters \((X, \theta)\) where
the probability distribution of \(X\) conditioned on \(\theta\) is identifiable, the marginal posterior
provides the opportunity to estimate the conditional PDF of \(X\) given \(Z\) as follows:

\[
P(X|Z) = \int_{\theta} P(X, \theta|Z)d\theta = \int_{\theta} P(X|\theta, Z)P(\theta|Z)d\theta.
\]

4. **Expectation.** The conditional posterior PDF provides all the required information for
inferring any desired average statistics suitable for the problem of choice. Such average
can be computed as:

\[
\mathbb{E}_{P(X|Z)}[f(X)] = \int_X f(X)P(X|Z)dx.
\]

3.2.2 Recursive Bayesian Filtering

Let \(x_t\) represent the state at time \(t\) and \(Z_t = \{z_1, z_2, ..., z_t\}\) denote a set of observations
from time 1 to \(t\). As Figure 3.2 inherently implies, the main goal in Bayesian filtering
involves utilizing the prior predictions and finite observations to yield more accurate posterior
estimations. From the mathematical point of view, at a given time instance \(t\), this objective
can be accomplished by exploiting the fusion of posterior estimate \(P(x_{t-1}|Z_{t-1})\) at time
\(t - 1\) and an observation \(z_t\) at time \(t\) to recursively derive the posterior density \(P(x_t|Z_t)\).
The details of this recursive process, as this section describes, play a fundamental role in
understanding the sequential filtering methods discussed later in this thesis.

In general, the recursive Bayesian filtering process involves two major steps, prediction
and correction. In the former, the target is to predict the probability distribution of the
current state $x_t$ given all the previous measurements $Z_{t-1} = \{z_1, ..., z_{t-1}\}$, yielding the prior distribution $P(x_t|Z_{t-1})$ at time $t$. As for the correction, the new observation $z_t$ is utilized to improve the predicted estimate for the current state, resulting in the posterior distribution $P(x_t|Z_t)$ at time $t$. The following calculations elaborate these steps.

To predict the current state using previous measurements, marginalization (Equation 3.7) over all past states yields:

$$P(x_t|Z_{t-1}) = \int_{x_{t-1}} P(x_t|x_{t-1}, Z_{t-1})P(x_{t-1}|Z_{t-1})dx_{t-1}. \quad (3.9)$$

Consequently, to take into account the role of new observation in correction stage, a straightforward manipulation of the Bayes’ rule (Equation 3.5) implies:

$$P(x_t|Z_t) = \frac{P(Z_t|x_t)P(x_t)}{P(Z_t)} = \frac{P(z_t, Z_{t-1}|x_t)P(x_t)}{P(z_t, Z_{t-1})},$$

$$= \frac{P(z_t|Z_{t-1}, x_t)P(Z_{t-1}|x_t)P(x_t)}{P(z_t|Z_{t-1})P(Z_{t-1})},$$

$$= \frac{P(z_t|Z_{t-1}, x_t)P(x_t|Z_{t-1})P(Z_{t-1})P(x_t)}{P(z_t|Z_{t-1})P(Z_{t-1})},$$

$$= \frac{P(z_t|Z_{t-1}, x_t)P(x_t|Z_{t-1})}{P(z_t|Z_{t-1})}. \quad (3.10)$$

Given the conditional independence assumptions for the states and measurements presented by Equations 3.1 and 3.2, the above PDFs can be simplified as:

$$P(x_t|Z_{t-1}) = \int_{x_{t-1}} P(x_t|x_{t-1})P(x_{t-1}|Z_{t-1})dx_{t-1}, \quad \text{and} \quad (3.11)$$

$$P(x_t|Z_t) = \frac{P(z_t|x_t)P(x_t|Z_{t-1})}{P(z_t|Z_{t-1})}. \quad (3.12)$$

where the posterior $P(x_t|Z_t)$ is comprised of three main components: prior, likelihood and evidence. The prior knowledge is expressed by the term $P(x_t|Z_{t-1})$ and can be calculated as per Equation 3.11, utilizing the dynamic model probability distribution $P(x_t|x_{t-1})$ governed by Equation 3.3. The probability distribution $P(z_t|x_t)$ represents the likelihood that is, in essence, the measurement model characterized by Equation 3.4. Finally, the probability
distribution \( P(z_t|Z_{t-1}) \) describes the evidence that is indeed a normalization factor computed based on Equation 3.6 as:

\[
P(z_t|Z_{t-1}) = \int_{x_t} P(z_t|x_t)P(x_t|Z_{t-1})dx_t.
\] (3.13)

The posterior probability distribution \( P(x_t|Z_t) \) provides all the required knowledge for inferring the current state \( x_t \) at time \( t \). This inference can be achieved using some statistical measure of optimality. For instance, based on the Maximum a Posterior (MAP) estimation approach, the objective is to find the state that maximizes the posterior probability distribution, i.e. the statistical mode of the PDF, that is:

\[
\hat{x}_t^{MAP} = \arg\max_{\inf} P(x_t|Z_t).
\] (3.14)

Alternatively, Minimum Mean Square Error (MMSE) method utilizes the measurement history and prediction error to infer the state as the conditional mathematical expectation of \( x_t \) given \( Z_t \). Utilizing Equation 3.8 the MMSE estimate can be calculated as:

\[
\hat{x}_t^{MMSE} = E(x_t|Z_t) = \int_{x_t} x_t P(x_t|Z_t)dx_t.
\] (3.15)

When comparing MAP and MMSE methods, it is worthwhile to note that each approach has its own cons and pros in different conditions. For instance, since MAP does not require calculating the integral in the denominator of the posterior PDF, it is generally considered to be a less computationally expensive approach when compared to MMSE which requires computation of evidence as well as prior and likelihood. Nonetheless, MAP has a major disadvantage that may lead to inaccurate estimates in certain circumstances such as high dimensional spaces. For instance, given a narrow tall spike with small width in the PDF of a multidimensional space, although the spike may have a high density, the probability of the state estimate belonging to that spike is not high [26].

This section described the conceptual approach for estimating the posterior and prior probability densities in recursive Bayesian filtering. Nonetheless, in general, calculating
the integrals in Equations 3.11 and 3.12 is an intractable problem. Section 3.3 describes an important special case where this problem has an analytical solution in the context of the Kalman filter. However, in more complex circumstances where the conditions for an analytical solution are not met, numerical methods such as those discussed in Chapter 4 provide an alternative solution to the filtering problem.

3.3 Kalman Filter

An important analytical solution to recursive Bayesian filtering, known as the Kalman filter [53, 91], arises from a special criteria where the state prediction and measurement update Equations 3.3 and 3.4 are governed by linear stochastic dynamics with Gaussian noises. Given that normal distributions are preserved through linear transformations, this criteria implies that the system dynamics and measurement model probability densities, $P(x_t|x_{t-1})$ and $P(z_t|x_t)$, have normal distributions. A significant advantage of having normal densities in filtering processes is that maintaining the first two moments, the mean and the covariance, can completely express the probability distribution and suffice for its propagation over time. As a result, the discussed criteria simplify the filtering problem to propagating a representation of these moments through the temporal estimation process. This section reviews Kalman filtering based on Welch and Bishop [91].

3.3.1 Problem Definition

The general objective of a Kalman filter is to estimate an $N_x$ dimensional state $x_t$ at time $t$ given an $N_z$ dimensional measurement $z_t$, in a stochastic process defined by the linear dynamic model:

$$x_t = Fx_{t-1} + v_t,$$  \hfill (3.16)

and measurement model:

$$z_t = Gx_t + u_t,$$  \hfill (3.17)
where $v_t \sim \mathcal{N}(0, Q)$ and $u_t \sim \mathcal{N}(0, R)$ represent zero mean Gaussian noises respectively with process and measurement covariance matrices $Q$ and $R$. The matrix $F_{N_x \times N_x}$, defines the state transition parameters from $x_{t-1}$ to $x_t$ in the dynamic model and the matrix $G_{N_z \times N_x}$, provides a mapping from the state space into the observation space in the measurement model. From the recursive Bayesian filtering point of view, Equations 3.16 and 3.17 respectively demonstrate the state transition $P(x_t|x_{t-1})$ and likelihood $P(z_t|x_t)$ probability densities.

### 3.3.2 Propagation Mechanism

The Kalman filter assumptions for the dynamic and measurement model defined in Section 3.3.1 necessitate a normal distribution for both prior and posterior distributions in the context of recursive Bayesian filtering (Equations 3.11 and 3.12). As previously mentioned, the mean and the covariance completely express the Gaussian distributions. Given that linear transformations can preserve normal distributions, the Kalman filter propagation process reduces to determining these parameters through the temporal estimation process.

Let $x_t^-$ denote the mean of the prior distribution $P(x_t|Z_{t-1})$ that is the a priori state estimate at time $t$ computed as:

$$x_t^- = \mathbb{E}[x_t|Z_{t-1}] = Fx_{t-1}^+,$$

where $x_{t-1}^+$ is the a posteriori state estimate at time $t - 1$ that represents the mean of the posterior $P(x_{t-1}|Z_{t-1})$. Consequently the prior covariance $P_t^-$ at time $t$ is:

$$P_t^- = \mathbb{E}[(x_t - x_t^-)(x_t - x_t^-)^T|Z_{t-1}] = FP_{t-1}^+F^T + Q,$$

where $P_{t-1}^+$ is the posterior covariance at time $t - 1$. In accordance with Bayesian filtering framework, the Kalman filter improves the a priori estimate $x_t^-$ by exploiting the difference between an actual observation $z_t$ at time $t$ and the prediction of that observation $Gx_t^-$. This difference is often referred to as measurement innovation. To calculate the a posteriori state
estimate \( x_t^+ \), the Kalman filter employs a linear weighted blend of a priori state estimate and measurement innovation as:

\[
x_t^+ = x_t^- + \mathcal{K}_t (z_t - Gx_t^-),
\]

(3.20)

where \((z_t - Gx_t^-)\) represents the measurement innovation and \(\mathcal{K}_t\) is called the Kalman gain, which indeed targets to minimize the posterior covariance \(P_t^+\) defined as:

\[
P_t^+ = \mathbb{E}[(x_t - x_t^+)(x_t - x_t^+)^T | Z_{t-1}].
\]

(3.21)

By substituting equation 3.20 into 3.21 for \(x_t^+\), it can be shown (as in [23]) that the Kalman gain \(\mathcal{K}_t\) which minimizes the resulting equation can be calculated as:

\[
\mathcal{K}_t = \frac{P_t^- G^T}{GP_t^- G^T + R}.
\]

(3.22)

The posterior covariance \(P_t^+\) can then be computed with [23]:

\[
P_t^+ = (I - \mathcal{K}_t G_t)P_t^-,
\]

(3.23)

where \(I\) is an \(N_x \times N_x\) identity matrix.

An important corollary implied by Equation 3.22 is that Kalman gain plays a decisive role in the estimation process by distributing the amount of trust between dynamic and measurement model based on the following mechanism: as the prior covariance \(P_t^-\) approaches zero, so does the Kalman gain, \(\mathcal{K}_t\), implying that the prior estimate, \(x_t^-\), gains more weight. In contrast, as the measurement noise covariance, \(R\), approaches zero, the Kalman gain puts more weight towards the measurement innovation.

To summarize the propagation process, Figure 3.3 illustrates the high level concepts of the steps discussed in this section. At a given time step \(t\), the process begins with drifting the previous posterior at time \(t - 1\), utilizing the system dynamics matrix \(F\) introduced in Equation 3.18. Consequently, the process noise expressed by \(Q\) in Equation 3.19 will be applied on the predicted PDF. Finally, to estimate the posterior distribution at time
Figure 3.3: The Kalman filter propagation process: given the previous posterior PDF at time $t - 1$, to estimate the next posterior at time $t$, the Kalman filter utilizes the dynamic model to predict the effective prior and consequently uses the measurement model to correct and improve the predicted prior. (This figure is inspired by Figure 1 in [50])
In short, the Kalman filter temporal estimation process can be described as an ongoing repetition of predictions followed by corrections; the two main steps, upon which, the equations presented in this section can be classified. At a given time step \( t \), after mean and covariance of the prior distribution are estimated in the prediction step, the correction phase provides corrective feedback for improving the predicted moments and determines the effective posterior mean and covariance that will in turn be used in calculation of the next prior at time \( t + 1 \). Algorithm 1 formalizes this procedure.

**Algorithm 1: The Kalman Filter Propagation Algorithm**

| input  | The posterior mean \( x_{t-1}^+ \) and covariance \( P_{t-1}^+ \) at time \( t - 1 \), The observation \( z_t \) at time \( t \) |
| output | The posterior mean \( x_t^+ \) and covariance \( P_t^+ \) at time \( t \) |

**A. Prediction:**

1. a priori mean estimate: \( x_t^- = Fx_{t-1}^+ \)
2. a priori covariance estimate: \( P_t^- = FP_{t-1}^+ F^T + Q \)

**B. Corrective Feedback:**

1. Kalman gain: \( K_t = \frac{P_t^- G^T}{GP_t^- G^T + R} \)
2. a posteriori mean estimate: \( x_t^+ = x_t^- + K_t(z_t - Gx_t^-) \)
3. a posteriori covariance estimate: \( P_t^+ = (I - K_tG_t)P_t^- \)
3.4 Summary

The Bayesian conditioning framework provides the formal definition of general state estimation problem in the context of recursive Bayesian filtering and Kalman filtering is a well known optimal solution under certain circumstances. Despite Kalman filter’s ease of implementation and efficiency, unfortunately, the conditions required by this filtering approach, Gaussian and linear dynamic and measurement models as well as Gaussian noises, may not always be met in practice. As a result, in the absence of such conditions (as is the case for our problem of interest in this thesis), to solve the filtering problem, we need to exploit numerical sampling approaches such as particle filters described in the Chapter 4.
This chapter presents our method for developing a human motion feedback system capable of generating auditory rhythms synchronized to periodic human movements. To develop our method, we rely on two salient points explained in depth in Chapter 2:

1. As described in Section 2.1, temporal relations embedded in a human movement play a critical role in perception of that movement. In the case of periodic movements, such temporal relations can be expressed through the *phase* of the movement, defined as the elapsed fraction of the motion in a single period.

2. As described in Section 2.2, rhythm provides an elegant way of representing the temporal relations in a motion. In addition, humans have an specialized ability to perceive rhythms and exploit its temporal information for exhibiting synchronized motor responses.

In accordance with the above points, our method has two main steps:

1. **Phase estimation**: Given a video sequence of a periodic human motion, the first step is to visually localize the movement and estimate the associated phase signal of the movement.

2. **Rhythmic sonification of the estimated phase**: Given the phase estimate of the movement, the second step is to sonify the phase signal such that consecutive cycles of the movement lead to a periodic rhythmic auditory pattern.

In pursuit of an efficient visual localization method, Chapter 3 introduced the temporal estimation framework in the context of recursive Bayesian filtering problem and reviewed one
of its analytical solutions under certain circumstances, known as the Kalman filter. However, in the case of our motion feedback system, the conditions required by the Kalman filter may not always be met. In other words, there is no guarantee that a periodic human movement will follow linear dynamics with Gaussian noises in practice. Unlike Kalman filters, the particle filter provides a non-analytical approximate solution to recursive Bayesian filtering problem without necessitating any condition for the system dynamics, measurement and noise.

In this chapter, we present the details of our phase estimation strategy in the context of phase-entrained particle filter and then describe a sonification scheme for translating the estimated phase signal to rhythmic auditory feedback in sync with the movement.

4.1 Walking Gait Example

We demonstrate our synchronization method by virtue of an everyday, yet complex example of a periodic movement: normal walking gait. As described in Section 2.2, rhythmic auditory cues have potential to be beneficial in clinical treatment of gait disorders. Inspired by this hypothesis, our motion feedback system presents a method for creating a rhythmic sonification synchronized to walking gait.

Figure 4.1 illustrates a conceptual overview of the process as well as our walking experiment setup. While the subject is walking on a treadmill, a phase-entrained particle filter localizes the subject’s silhouette in a side-view video sequence provided by a conventional RGB camera. At the same time, the particle filter also estimates the phase signal \( \phi \) associated with the movement. Subsequently, a sonification scheme exploits the phase estimate, \( \phi \), to generate rhythmic auditory feedback synchronized to gaits.
Figure 4.1: A conceptual overview of synchronization method and gait experiment setup: An RGB camera provides the side-view video sequence of a subject walking on the treadmill. A phase-entrained particle filter estimates the phase signal, $\phi$, associated to gaits, which will in turn be used by the sonification scheme to generate synchronized rhythmic auditory feedback.

4.2 Visual Localization and Phase Estimation

To visually localize the movement and estimate the associated phase signal, $\phi$, we employ a particle filter as per the CONDENSATION algorithm introduced by Isard and Blake [50]. Section 4.2.1 reviews the general framework of particle filtering in the context of the CONDENSATION as described by Isard and Blake [50]. Consequently in Section 4.2.2 we present our customization of this general framework, introducing the phase-entrained particle filter.

4.2.1 CONDENSATION Framework

Let $x_t$ define the state vector at time $t$ with trajectory $X_t = \{x_1, ..., x_t\}$. Similarly, let $z_t$ denote a measurement at time $t$ with the history $Z_t = \{z_1, ..., z_t\}$. In compliance with the
state and measurement independence assumptions presented by Equations 3.2 and 3.1, as for the Condensation, the conditional independence assumption for the states implies that system dynamics forms a temporal Markov chain such that:

\[ P(x_t | x_{t-1}) = P(x_t | x_{t-1}). \]  

Likewise, for the measurements, an observation at a given time depends only on the state at that time and is conditionally independent of other states and measurements. Probabilistically, this condition implies that:

\[ P(Z_{t-1}, x_t | X_{t-1}) = P(x_t | X_{t-1}) \prod_{i=1}^{t-1} P(z_i | x_i), \quad \text{and} \quad \]  

\[ P(Z_t | X_t) = \prod_{i=1}^{t} P(z_i | x_i). \]  

Similar to recursive Bayesian filtering, the propagation problem for the Condensation can be defined as an inference problem with prior distribution \( P(x_t | Z_{t-1}) \) such that:

\[ P(x_t | Z_t) \propto P(z_t | x_t) P(x_t | Z_{t-1}), \quad \text{and} \quad \]  

\[ P(x_t | Z_{t-1}) = \int_{x_{t-1}} P(x_t | x_{t-1}) P(x_{t-1} | Z_{t-1}) dx_{t-1}, \]  

where the models for dynamic process and measurement update are probabilistically represented as \( P(x_t | x_{t-1}) \) and \( P(z_t | x_t) \), respectively.

To calculate the posterior \( P(x_t | Z_t) \), the Condensation utilizes a Monte Carlo approach that approximates the probability distribution using a factored sampling technique, illustrated in Figure 4.2. Consider the standard definition of Bayes’ rule:

\[ P(x | z) \propto P(z | x) P(x). \]  

In the event that it is hard to calculate the closed form of the posterior, \( P(x | z) \), factored sampling provides an alternative approximate representation of the posterior based on a set of weighted sample points \( s_i \), randomly sampled from the prior \( P(x) \). These points are depicted as circles with variable areas in Figure 4.2. The area of a circle (or sample point
Figure 4.2: Factored sampling concept: a set of weighted samples, depicted as circles, are sampled from the prior distribution and evaluated based on the likelihood model to receive their associated weights. Such weights that are reflected as the area of circles, i.e. the bigger blob has a higher weight and vice versa. The resulting blobs provide an approximate representation of the posterior density (inspired by Figure 3 in [50]).

$s_i$ represents its weight, denoted as $\pi_i$, and it is proportional to the measurement model likelihood $P(z|x = s^k)$, calculated as:

$$\pi_i = \frac{P(z|x = s_i)}{\sum_{j=1}^{N_s} P(z|x = s_j)}, \tag{4.7}$$

where $N_s$ is the total number of sample points. The weights generated in this fashion approximate the posterior $P(x|z)$.

Figure 4.3 summarizes the propagation process in the CONDENSATION algorithm. From a probabilistic perspective, Figure 4.3(a) describes a process similar to the Kalman filter propagation scheme depicted in Figure 3.3. At a given time $t$, the CONDENSATION utilizes a known representation of the posterior at $t - 1$ to infer the prior distribution at time $t$ by applying the dynamic drift and noise on the previous posterior. Once the prior has been estimated, the measurement model exploits new observation to correct the prior estimate. However, unlike the Kalman filter, in the CONDENSATION, during the propagation process the probability densities may have non-Gaussian distribution and undergo nonlinear trans-
formations. Figure 4.3(b) elaborates this process from the factored sampling point of view. Given the sample point representation of the posterior at time \( t - 1 \), to generate a new sample set, the **Condensation** performs a re-sampling such that samples with greater weights have a higher chance of being chosen. Consequently the new samples go through the drift and dispersion steps to estimate the effective prior at time \( t \). Finally the posterior at time \( t \) will be approximated based on the likelihood of new observation given the prior estimate.

Let \( S_{t-1} = \{(s^i_{t-1}, \pi^i_{t-1}, c^i_{t-1}), i = 1, ..., N\} \) be a set of weighted samples at time \( t-1 \), where \( s^i_{t-1} \) is a sample, also known as particle, drawn from the posterior at time \( t-1 \) and \( \pi^i_{t-1} \) is its associated weight. Utilizing the cumulative weights \( c^i_{t-1} \) for performing an efficient re-sampling by binary subdivision, Algorithm 2 formalizes the details of the propagation process. In the following (Section 4.2.2), we present our customization of this algorithm.

**Algorithm 2:** The Condensation algorithm [50].

**input**: Sample set \( S_{t-1} = \{(s^i_{t-1}, \pi^i_{t-1}, c^i_{t-1}), i = 1, ..., N\} \) at time \( t-1 \) & Observation \( z_t \) at time \( t \).

**output**: Sample set \( S_t = \{(s^i_t, \pi^i_t, c^i_t), i = 1, ..., N\} \) at time \( t \).

for \( i = 1 \rightarrow N \) do

1. **Re-sample**: select \( s'^i_t \) as follows:
   
   1. \( r \leftarrow \) a random number \( \in \) uniform distribution \( U([0,1]) \)
   2. \( j \leftarrow \) find smallest \( j \) by binary search such that \( c^j_{t-1} \geq r \)
   3. \( s'^i_t \leftarrow s^j_{t-1} \)

2. **Predict**: sample from \( P(x_t| x_{t-1} = s'^i_t) \) using a linear stochastic differential equation \( s^i_t = A s^i_t + Q w_t \) where \( A \) is the system dynamics, \( w_t \) is a vector of standard normal random numbers and \( QQ^T \) is the process noise covariance.

3. **Measure**: assign a weight \( \pi^i_t = P(z_t| x_t = s^i_t) \) to each new sample based on measured feature \( z_t \).

end

4. **Normalize**: the weights so that \( \sum_{i=1}^{N} \pi^i_t = 1 \) and update the cumulative weights such that \( c^0_t = 0 \) and \( c^i_t = c^{i-1}_t + \pi^i_t \), \( i = 1, ..., N \).

5. **Estimate**: Infer the mean of the estimated sample set \( S_t \) as \( E[S_t] = \sum_{i=1}^{N} \pi^i_t s^i_t \).
Figure 4.3: The propagation process in the CONDENSATION: (a) probabilistic point of view. (b) factored sampling point of view.
Figure 4.4: An illustration of the state space parameters in a sample video frame: \((x,y)\) represent the coordinates of top left corner on the rectangular region of interest, while \(\dot{x}\) and \(\dot{y}\) are their associated velocities. \(\alpha\) is a parameter showing the scale of the rectangle with a fixed aspect ratio. \(\phi\) represents the phase of the movement that can be used for synchronizing the subject’s motion with the computer model.

4.2.2 Phase-Entrained Particle Filter

To employ a particle filter as per the CONDESATION framework, we need to customize Algorithm 2 to make it suitable for the purpose of our motion feedback system. This section presents such customization specifically by defining the state vector parameters, dynamic process and the measurement model.

4.2.2.1 State Vector Definition

As previously mentioned, to generate rhythmic auditory feedback, our filtering system must localize the movement and estimate its associated phase. Therefore, in the walking gait example, the phase of the gait, denoted as \(\phi\), is part of the state space. Let \(s_t = [x_t, y_t, \dot{x}_t, \dot{y}_t, \alpha_t, \phi_t]^T\) be the 6-dimensional vector representing the state of a particle at time \(t\). As Figure 4.4 illustrates, while \(x_t\) and \(y_t\) define the coordinates of the top left
corner of the bounding box, \( \dot{x}_t \) and \( \dot{y}_t \) represent their associated velocities in \( x \) and \( y \) directions, respectively. The scale parameter, \( \alpha_t \), allows the bounding box to change size with a fixed aspect ratio. The phase dimension, \( \phi_t \in [0, 1] \), represents the phase of the movement, although this is difficult to represent in a static, two-dimensional figure.

4.2.2.2 Dynamic Model

To predict the state of the particles at time \( t \) (Algorithm 2, step 2), we model the dynamics of the gait motion as a second order process. In particular, we build the \( i^{th} \) of \( N_s \) new samples using the stochastic differential equation:

\[
s_i^t = As_i^{t'} + Bu + Qw_i^t
\]

where \( A, B \) represent the dynamic model, \( u \) is the control vector, \( w_i^t \) is a vector of standard normal random numbers, and \( QQ^T \) is the process noise covariance. Specifically, for our phase-entrained filtering we use:

\[
A = \begin{pmatrix}
1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
B = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
\]

\[
u = \begin{pmatrix}
\Delta \phi
\end{pmatrix},
Q = \begin{pmatrix}
\sigma_x & 0 & 1 & 0 & 0 & 0 \\
0 & \sigma_y & 0 & 1 & 0 & 0 \\
0 & 0 & \sigma_{\dot{x}} & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma_{\dot{y}} & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma_\alpha & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma_\phi
\end{pmatrix}.
\]
Therefore, the deterministic drift assumes constant velocity in subject position, constant scale, and phase drift equal to $\Delta \phi$. Elements on the diagonal of $Q$ determine the amount of dispersion for each state dimension.

4.2.2.3 Measurement Model

To compute the weight of particles (Algorithm 2, step 3), the observation model estimates the conditional likelihood of new observation at time $t$ given the predicted particles. In other words, it assigns the particle $s^i_t$, the weight $\pi^i_t$ proportional to $P(z_t|x_t = s^i_t)$, where $i = 1, ..., N_s$. To evaluate each particle, in short, our measurement model exploits a set of integrated model gait silhouettes to perform a comparison between particles and integrated models. The partial weight is therefore a measure of particle’s conformity to a model silhouette. In our phase-entrained particle filter, the key to evaluate each particle $s^i_t$ is embedded in its phase dimension $\phi^i_t$, as it determines the associated model against which the particle should be compared.

Algorithm 3 shows more details of our measurement process. First, the algorithm segments the new observed image at time $t$ using chromakeying and background subtraction to extract the silhouette of the subject. Section 5.1 describes the details of our segmentation process. Second, the phase $\phi^i_t$ serves as an index to retrieve the corresponding silhouette from the set of model silhouettes. Consequently, the algorithm computes a similarity measure between the model silhouette and the measured silhouette that is cropped from the observed frame based on the rectangle defined by $x$, $y$, and $\alpha$. Finally, our algorithm assigns a weight to the particle based on this similarity. The Hausdorff distance [49] provides a convenient similarity measure at this step. Figure 4.5 illustrates how this distance can be calculated by superimposing the particle and model silhouettes on top of each other. Section 5.2 elaborates the implementation details of this process. The Hausdorff metric, in essence, measures the distance between foreground points of two binary images as follows. Let $P = \{p_1, ..., p_n\}$ and $Q = \{q_1, ..., q_n\}$ denote two finite sets of points. The Hausdorff distance between these
points sets can be defined as:

\[
d(P, Q) = \max[h(P, Q), h(Q, P)], \quad \text{where} \\
h(P, Q) = \max_{p \in P} \min_{q \in Q} \|p - q\|, \quad \text{and} \\
\tag{4.9}
\]

when \(\|\cdot\|\) is an underlying norm on points of \(P\) and \(Q\).

Algorithm 3: Observation model

**input:**
1. \(z_t\): video frame at time \(t\).
2. \(s_i^{t}\): predicted particle at time \(t\).
3. \(M = \{I_1, \ldots, I_m\}\): Image sequence of model silhouettes

**output:** particle weight \(\pi_i^{t} = P(z_t|x_t = s_i^{t})\) at time \(t\)

1. **Segment** the input video frame \(z_t\) using chromakeying and background subtraction and store the segmented image as \(z_t'\).
2. **Retrieve** the image of the particle’s rectangular region of interest and its associated model as follows:
   (a) extract the image of rectangular region of interest defined by \(s_i^{t}\) from \(z_t'\) and store it as \(P\).
   (b) compute the index of particle’s model as \(k = \lfloor \phi_i^{t} \times (m + 1) \rfloor\).
   (c) retrieve the associated model’s image and store it as \(Q = I_k\).
3. **Measure** the similarity between \(P\) and \(Q\) using the Hausdorff distance \([49]\) metric and store the results as \(d_i^{t} = d(P, Q)\).
4. **Assign** particle’s weight in proportion to \(d_i^{t}\) as \(\pi_i^{t} \propto e^{-(d_i^{t}/\sqrt{2}\sigma_d)^2}\) where \(\sigma_d\) is a constant scaling parameter.

It is worthwhile to mention that the last step of our measurement process (Algorithm 3 step 4) involves assigning a particle \(s_i^{t}\) the weight \(\pi_i^{t}\) by utilizing the calculated Hausdorff distance \(d_i^{t}\) as:

\[
\pi_i^{t} \propto e^{-(d_i^{t}/\sqrt{2}\sigma_d)^2}, \\
\tag{4.11}
\]

44
Figure 4.5: An illustration of the Hausdorff distance by overlaying model (yellow) and particle (blue) silhouettes. The green area represents the intersection and the red arrows represent sample distance between two pairs of points chosen from particle point set $P = \{p_1, ..., p_n\}$ and model point set $Q = \{q_1, ..., q_n\}$.

Figure 4.6: The role of scaling parameter in particle’s weighting function
where $\sigma_d$ is a constant scaling parameter. The choice of $\sigma_d$ controls the sensitivity level of our measurement process in measuring the particle’s conformity to a model silhouette. As Figure 4.6 illustrates this concept, $\sigma_d$ indeed represents the slope of the weighting function given by Equation 4.11.

4.3 Rhythmic Sonification

In our path towards generating synchronized auditory feedback, previous section described the first step of our method: phase estimation. In this section, we describe a sonification process for transforming the estimated phase signal to rhythmic auditory feedback synchronized to gaits. This step is developed on the basis of a prior work in this area presented by Godbout and Boyd [40].

Let $\phi \in [0, 1]$ denote the phase signal estimated by our phase-entrained particle filter. In the case of normal walking, the periodic nature of subject’s gaits, necessitates $\phi$ to follow a periodic pattern too. In other words, repeating with each period of the subject’s gait, the synchronized phase estimate $\phi$ cycles from zero to one as well. As previously described in Section 2.2, periodicity lends itself nicely to definition of rhythms. In the following, we discuss a mechanism to take advantage of the periodic nature of $\phi$ in creation of rhythms.

Figure 4.7 describes our rhythmic sonification mechanism. We divide a full period of the estimated phase signal into $N_{\phi_T}$ sections using $N_{\phi_T}$ phase thresholds denoted as $\phi^i_T$ where $i = 1, ..., N_{\phi_T}$. In the case of walking gaits, we choose to divide each period into four equal sections with thresholds $\phi^i_T$ equally placed through the period as:

$\phi^1_T = 0.25, \quad \phi^2_T = 0.5, \quad \phi^3_T = 0.75, \quad \phi^4_T = 1,$

where $\phi^4_T = 1$ indicates the end of a period. Each phase threshold $\phi^i_T$, in essence, represents a particular pose from a full period of the movement such as those depicted in Figure 4.7.

For instance, in our model sequence for walking gait, where a full period is defined by a half stride, the phase threshold $\phi^2_T = 0.5$ represents a pose where the legs are about to align. As
the subject progresses towards the movement, the estimated phase signal passes the phase thresholds. The passage of phase signal $\phi$ over phase thresholds $\phi_i^T$ triggers auditory cues, such as certain pitches, in the sonification system. In our sonification system, we assign each phase threshold an auditory pitch as the following:

$$\phi_1^T \rightarrow C_4, \quad \phi_2^T \rightarrow E_4, \quad \phi_3^T \rightarrow G_4, \quad \phi_4^T \rightarrow C_5.$$ 

Considering this sonification strategy, repetitive cycles of the movement, such as consequent strides during walking, result in a periodic sequence of auditory pitches— in our case a repetitive C-Major arpeggio. Such auditory repetition can quickly be perceived as an steady rhythm, especially if the subject follows the periodic pattern of the movement steadily. Given that $\phi$ itself is synchronized to subject’s movement, the created auditory rhythm will be synchronized to the movement as well. For instance, in the case of walking gaits, slower strides would elicit a slower rhythm and vice versa.

Figure 4.7: Rhythmic Sonification Process
It is worthwhile to note that, in general, the number of phase thresholds $N_{\phi_T}$ and their placement throughout a period of the movement, depends on the nature of the periodic movement and the goal of the synchronized rhythmic feedback. For instance, in walking gait example, if our intention is to provide auditory feedback to help subjects maintain uniform strides, it may be beneficial to put a phase threshold at the pose associated with footfalls. Similarly, if we are aiming to teach a certain motor skill for a specific movement such as a periodic dance move, it may be helpful to put the phase threshold at critical poses in such movement so that subjects can identify whether they copy the model move correctly.
Chapter 5

Implementation

In the previous chapter, we introduced our phase-entrained particle filtering method for estimating a phase signal synchronized to the course of a periodic human locomotion such as walking gait. Subsequently, we described a rhythmic sonification strategy that uses the phase to generate auditory feedback in sync with the movement. A key component at the heart of this synchronization process is our measurement model – as it is the only internal component of the motion feedback system that analyzes the external data acquired from the actual movement. Section 4.2.2.3 defined the measurement model in general terms. This section explains in more detail how the acquired information from the observations are processed within the main components of the measurement model. In addition, by focusing specifically on the gait synchronization example, we present the sonification schemes of our system. Furthermore, we discuss how to fine tune the control parameters of the system dynamics and also describe the initialization process of our phase-entrained particle filter. Finally, we provide the implementation details of the whole system.

5.1 Segmentation

As Figure 4.1 illustrated, our apparatus for the gait experiment utilizes a walking treadmill as well as a chromakey green screen. Figure 5.1(a) shows a sample observation $z_t$ in the system provided by a conventional RGB camera. As Algorithm 3 described, the first step of the measurement process is to segment this video frame to extract the silhouette of the walking subject. To do this, we first employ chromakeying to remove the green screen and then use background subtraction together with a set of smoothing operations to remove the treadmill. The process is described in detail as follows.
1. **Chromakeying**: Chromakeying is a standard technique for identifying a foreground of interest when a chromakey screen is used as a background. Algorithm 4 describes our chromakeying procedure. Given the RGB input image $z_t$ at time $t$, the process first transforms the image coloring scheme to a Hue-Saturation-Value (HSV) color space to extract the hue information embedded in the image. Utilizing the hue range of chroma screen, Algorithm 4 then thresholds the hue to generate a binary image. Figure 5.1(b) shows the results of chromakeying on the sample RGB input illustrated in Figure 5.1(a).

2. **Background Subtraction**: Since we are only interested in the binary silhouettes of the subject, we need to discard the treadmill from the foreground of the binary images. To do this, we pre-compute the location of treadmill by applying chromakeying on a frame with no subjects. Once this location (as shown in Figure 5.1(c)) is determined, to extract the silhouette of the walking subject, we subtract treadmill’s precomputed location from the results of chromakeying.

3. **Smoothing**: A set of morphological operations eliminate sporadic foreground pixels. Specifically, the operations include repetitive erosions followed by dilations. Figure 5.1(d) shows the final binary gait silhouette resulted by our segmentation process.

5.2 Hausdorff Distance Computation

The measurement model in Algorithm 3 measures the similarity between model and subject gait silhouettes using the Hausdorff distance. Figure 5.2 illustrates the details of how we calculate the Hausdorff distance. Let $P$ and $Q$ represent the images of the character “C” respectively in non-italic and italic formats as depicted in Figure 5.2. In practice, to calculate the Hausdorff distance efficiently, we take advantage of distance transform matrices of $P$ and $Q$. In computation of the distance transform matrices, we assume 8-neighborhood connectivity and an $L_\infty$ norm to find the shortest distance of each pixel to the closest
Figure 5.1: (a) A sample video frame observed by the system. (b) Partially segmented sample frame resulted by chromakeying only. (c) The pre-computed location of treadmill. (d) Final segmented frame after chromakeying, background subtraction and smoothing.

Algorithm 4: Chromakeying Algorithm

\begin{verbatim}
input : Color Image $I_{RGB}$, Background Color $c$, Threshold $\tau$
output: Segemented Binary Image $I_S$

$I_{HSV} \leftarrow \text{ColorTransform}(I_{RGB})$
$I_H \leftarrow \text{ExtractHue}(I_{HSV})$

\begin{algorithmic}
  \FOR {row $i \in I_H$}
    \FOR {column $j \in I_H$}
      \IF {$|I_H(i, j) - c| > \tau$}
        $I_S(i, j) \leftarrow 1$
      \ELSE
        $I_S(i, j) \leftarrow 0$
      \ENDIF
    \ENDFOR
  \ENDFOR

return $I_S$
\end{algorithmic}
\end{verbatim}
Figure 5.2: Hausdorff distance implementation using $L_{\infty}$ norm and distance transform matrices
foreground pixel. Once the distance transforms have been calculated, to realize Equation 4.9, we first define a region of interest in distance transform of $Q$ by superimposing the foreground region of $P$ on top of the distance transform of $Q$. The partial Hausdorff distance, $h(P,Q)$, is then computed as the summation of all the values in the defined region of interest in the distance transform of $Q$. Likewise, to calculate $h(Q,P)$, we first construct a region of interest in distance transform of $P$ in a similar fashion and then compute $h(Q,P)$ as the summation of all values in such region. Once both $h(P,Q)$ and $h(Q,P)$ are determined, in compliance with equation 4.9, the Hausdorff distance between $P$ and $Q$ will be defined as:

$$d(P,Q) = \max[h(Q,P), h(P,Q)].$$

5.3 Sonification Schemes

Section 4.3 described our rhythmic sonification process in general terms. For the gait synchronization example, we generate the rhythmic auditory feedback based on the following sonification schemes:

1. **Single Note**: A single $C_5$ note plays at the last beat of a four-quarter rhythm depicted in Figure 5.3. This note signals the end of each gait cycle in coincidence with footfalls.

2. **Ascending Arpeggio**: As Figure 5.3 illustrates, an ascending sequence of pitches $C_4, E_4, G_4, C_5$ (a C-Major chord) broadcast during the course of a period that is evenly partitioned with a four-quarter rhythm, such that each pitch corresponds to a beat. This scheme provides subjects with continuous feedback throughout the gait cycle.

In accordance with the process explained in Section 4.3, as the phase signal $\phi$ passes each phase threshold $\phi_T$, the subjects receive auditory feedback by playback of associated pitches depicted in Figure 5.3.
5.4 System Dynamics Control Parameters

Section 4.2.2.2 introduced the dynamic model of our phase-entrained particle filter. For the deterministic drift and diffusion of phase dimension, we set the phase drift parameter $\Delta \phi = 0.4 \text{ radians}$ and the dispersion as $\sigma_\phi = \frac{\Delta \phi}{4} = 0.1 \text{ radians}$ (with the processing speed of 15 frames per second). These values define a dynamic range broad enough to account for individual differences between the majority of subjects in terms of height and gait pattern. However, if the subject’s height or gait pattern is an extreme outlier, we need to adjust $\Delta \phi$ accordingly to account for the unusual stride length not covered within the expected range of our system. As Figure 5.4 illustrates, at a given system time $t + 1$, $\Delta \phi$ introduces a deterministic drift from the previous phase estimate $\phi_t$ to the current prior phase estimate $\phi_{t+1}^-$. The stochastic diffusion is illustrated by a normal distribution around the mean $\phi_{t+1}^-$ with the standard deviation $\sigma_\phi$. 

![Figure 5.3: Sonification schemes for walking gait experiment](image)
Figure 5.4: An illustration of the phase drift $\Delta \phi$ and diffusion $\sigma_\phi$ at a given system time $t + 1$

5.5 Initialization

Figure 5.5 illustrates the initialization procedure of our phase-entrained particle filter, which has two major steps:

1. The segmentation process (described in Section 5.1) extracts the silhouette of a walking subject. This segmentation establishes an initial bounding box around the silhouette. The red rectangle in Figure 5.5 depicts an example initial bounding box.

2. We instantiate $N = 250$ particles. The $x$ and $y$ values of the particles are sampled from two univariate Gaussian distributions centered at the position of the bounding box with the standard deviation values $\sigma_x = 20$ and $\sigma_y = 10$ pixels in $x$ and $y$ dimensions, respectively. Since the subject position is not changing on the treadmill, we sample values of $\dot{x}$ and $\dot{y}$ from two Gaussian distributions with zero mean and standard deviation approximately equivalent to 2 pixels. We set the scale of the particles as a constant value equal to the scale of the initial bounding box. As for the phase values of the particles, we
exploit a uniform distribution to spread the particles across the normalized phase range, \( \phi \in [0, 1] \). The yellow rectangles in Figure 5.5 illustrate example particles projected onto the image plane.

Once all \( N \) particles are instantiated, we assign a uniform weight to each particle equal to \( \frac{1}{N} \). These \( N \) initial particles are sufficient to initialize the Condensation algorithm at step 1.

5.6 General System Development

We implemented our synchronization system in C++ utilizing OpenCV library [1]. OpenCV has a strong focus on optimized performance that is especially helpful for real-time applications and it also provides support for hardware acceleration capacities.

For sonification, we used Pure Data [2], a visual programming environment that facilitates sound synthesis. Although pure data supports custom extensions to some extent, it is not
a strict programming language. The core logic and flow of a Pure Data program is formed
and controlled through patches – an organization of visual objects and routines that can be
connected to each other by visual links. Figure 5.7 shows a screen shot of our Pure Data
patch created for displaying phase signal and generating rhythms at real-time.

The communication between our synchronization system implemented in C++ and soni-
ification patch developed in Pure Data is established through Open Sound Control (OSC)

The gait model incorporated in our measurement model consists of a set of silhouettes
produced with the gait synthesis tool in Poser [83]. As Figure 5.6 shows the model silhou-
ettes, our model represents a single gait period. Because it is difficult to disambiguate left
and right strides from a side-view silhouette, we define the period in our system based on a
single step, left or right.
Figure 5.6: Walking gait model: the silhouettes represent a cycle of gait model. The model walker progresses in the direction of arrows. $\phi$ indicates the associated phase of each model silhouette through the cycle.
Figure 5.7: Pure Data sonification patch
Chapter 6

Testing

6.1 Walking Gait Experiment

We test the ability of our system to synchronize against the gait model by conducting the following experiments:

1. **Model vs Model.** Given the single gait period represented by system’s incorporated model silhouettes, we created a periodic gait animation by concatenating multiple repetitions of the model sequence. We then used this animation as an input to the system to see if our method is capable of generating synchronized rhythmic feedback against its own model.

2. **Model vs Subjects.** We invited 8 subjects (5 male, 3 female), all with normal gaits, to use our system while requiring them to walk normally and not wear colors that might confound chromakeying. Each subject walked on the treadmill for approximately three to four minutes. We informed the subjects that they would hear auditory feedback while walking and instructed them to change the treadmill speed when requested. All the changes were applied in steps of 0.1 miles per hour (mph) in the range from 1.2 mph to 2.2 mph.

In both experiments we varied the pattern of broadcasting auditory rhythm based on the *single note* or the *ascending arpeggio* sonification schemes described in Section 5.3.

Our evaluation scheme is twofold, qualitative assessment and quantitative analysis. For the qualitative assessment we rely on subjective verification of synchronization between gaits and perceived auditory feedback. Sections 2.1 and 2.2 reviewed evidence in support of humans’ expertise to accurately perceive motion and auditory rhythms and exhibit motor
movements synchronized to rhythms. This ability allows humans to quickly verify the synchronization between their motor movements and a perceived auditory rhythm. We utilized this innate ability to evaluate phase signal $\phi$, and consequently the rhythmic auditory feedback. In addition, as the quantitative analysis, we employ a numerical method for verifying the correlation between phase estimates and ground truth data.

6.2 Observations

6.2.1 Model vs Model

In the first experiment, synchronization of model gait versus itself, our subjective observations confirmed synchronization between model walker’s footfalls and the played $C_5$ note (depicted in Figure 5.3) in both sonification scenarios. In the case of ascending arpeggio feedback, we perceived a consistent and stable four-quarter rhythm as a C-Major arpeggio in perfect sync with our gait animation. We categorized a rhythmic auditory cue as a stable feedback if the perturbations in the tempo of the rhythm was negligible during the course of simulated walking. The reason for this categorization is that the synthetic walker produces identical strides during the animation. Such strides should in turn be sonified as a constant paced rhythm.

Given that this experiment synchronizes the gait model to itself, we may expect the estimated phase signal $\phi$ to exhibit exact phase ramps. However, as Figure 6.1(a) shows the phase estimation results, this expectation is not completely fulfilled. The main reasons for this follow.

1. The phase-entrained particle filter employs a stochastic framework to provide an approximate solution to the temporal estimation problem. Such stochastic characteristic introduces a certain amount of randomness into the estimation process to account for non-deterministic nature of observations. This randomness, in part, results in non-exact estimates.
2. The Hausdorff distance does not vary smoothly in response to smooth changes in the inputs. This is partially due to the pixel-wise accumulation of distances from the foreground regions when computing the Hausdorff distance. Such abrupt changes result in some perturbations in the estimation process that in turn lead to non-exact phase ramps.

Despite these imperfections, the estimated phase signal, $\phi$, still serves as a reliable means of sonification. In general, we determine the reliability of an estimated phase signal $\phi$ as its level of conformity to a hypothetical perfect linear phase ramp that represents a period of the movement – such that higher conformity results is more reliability and vice versa. As a result, strictly ascending ramps are essential components of a reliable phase signal $\phi$. Given the rhythmic sonification process described in Section 4.3, a non-reliable phase estimate often results in erroneous triggering of auditory cues – such that subject’s pose is not in sync with the intended pose as implied by phase thresholds $\phi_T$.

6.2.2 Model vs Subjects

For the second experiment, synchronization of model gait versus real subjects, single note sonification strategy resulted in subjects’ confirmation that the $C_5$ note played in sync with their footfalls for all variations in speed of the treadmill. In the ascending arpeggio case, after a few seconds, the subjects perceived that the high $C_5$ note synchronized with their footfalls and they reported an association between the tempo of the arpeggio and their pace, i.e. they felt faster cadence resulted in higher tempo of the auditory rhythm. When we asked the subjects to compare different feedback strategies, they showed a general preference towards the ascending arpeggio scheme due to its more elaborate feedback and aesthetically pleasing nature.

As observers, we noticed that in all cases, regardless of the pattern of auditory feedback or treadmill speed, the reference $C_5$ note (depicted in Figure 5.3) was always in sync with
the footfalls. Figures 6.1(b) and 6.1(c) illustrate the estimated phase signal $\phi$ for a typical participant walking at two different speeds. The phase ramps cycle from zero to one repeating with each stride. Figure 6.1(d) depicts the phase signal associated to the case where the subject’s movement is not categorized as walking, e.g. standing.

To evaluate the reliability of generated phase ramps, we need to compare the estimated phase signal with some sort of true phases as the ground truth assumption. Although we cannot measure the true phase directly, we can estimate it by utilizing the periodic nature of gaits. We build our ground truth data by assuming perfectly linear phase ramps between consecutive footfalls of the subjects. Figures 6.2(a) through 6.2(h) illustrate sample scattered plots of phase estimates versus the ground truth for four subjects.

6.3 Discussion

The correlation between phase estimates and true phase values defines a quantitative measure of reliability of our phase estimates. Given the discrete phase estimates $\hat{\phi}_t$ and the true phase values $\phi_t$ (ground truth) for a certain system time interval, we calculate the correlation between estimates and true values using the Pearson $r$ correlation coefficient [32]. As described in Section 6.2.1, we defined the reliability of an estimated phase signal $\phi$ as its level of conformity to perfectly linear ramps, i.e. true phase values. The Pearson $r$ value in essence measures this conformity such that higher values of $r$, as it approaches to one, correspond to a higher conformity. Likewise, as $r$ approaches to zero, the correlation, and thus the reliability of generated phase signal, decreases.

Table 6.1 presents the Pearson $r$ values for all the subjects involved in our walking gait experiment for a slow and fast speed respectively denoted as $r_{slow}$ and $r_{fast}$. As the average results show, the correlation between our estimates and ground truth data is close to one. This quantitative analysis, together with our subjective perception of synchronizations described in Section 6.2 confirm the reliability of the generated phase signal $\phi$. 
Figure 6.1: Estimated phase signal for (a) the model versus model synchronization, (b,c) subject #1 walking respectively at a slow and fast speed on the treadmill and (d) subject #1 during a non-gait activity. In all figures φ is the phase normalized to $0 \leq \phi \leq 1$. 
Figure 6.2: Scattered plots of phase estimates, $\phi$, versus ground truth, $\phi_{true}$. The depicted line in each figure is the line that fits to the scattered points best.
Table 6.1: Walking gait experiment statistical results: A summary of quantitative analysis of results for the walking gait experiment. The statistical measures include Pearson $r$ values for walking at slow and fast speeds, as well as the mean and the covariance of timing errors, denoted by $\mu_{\Delta t}$ and $\sigma_{\Delta t}$ respectively.

<table>
<thead>
<tr>
<th>Subject #</th>
<th>$r_{\text{slow}}$</th>
<th>$r_{\text{fast}}$</th>
<th>$\mu_{\Delta t}^{\text{slow}}$ (s)</th>
<th>$\sigma_{\Delta t}^{\text{slow}}$ (s)</th>
<th>$\mu_{\Delta t}^{\text{fast}}$ (s)</th>
<th>$\sigma_{\Delta t}^{\text{fast}}$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>0.9854</td>
<td>0.9854</td>
<td>0.0530</td>
<td>0.0432</td>
<td>0.0530</td>
<td>0.0432</td>
</tr>
<tr>
<td>1</td>
<td>0.9932</td>
<td>0.9917</td>
<td>0.0308</td>
<td>0.0228</td>
<td>0.0246</td>
<td>0.0211</td>
</tr>
<tr>
<td>2</td>
<td>0.9797</td>
<td>0.9834</td>
<td>0.0604</td>
<td>0.0494</td>
<td>0.0445</td>
<td>0.0324</td>
</tr>
<tr>
<td>3</td>
<td>0.9791</td>
<td>0.9779</td>
<td>0.0692</td>
<td>0.0529</td>
<td>0.0351</td>
<td>0.0258</td>
</tr>
<tr>
<td>4</td>
<td>0.9774</td>
<td>0.9727</td>
<td>0.0746</td>
<td>0.0575</td>
<td>0.0449</td>
<td>0.0397</td>
</tr>
<tr>
<td>5</td>
<td>0.9501</td>
<td>0.9510</td>
<td>0.0820</td>
<td>0.0608</td>
<td>0.0836</td>
<td>0.0623</td>
</tr>
<tr>
<td>6</td>
<td>0.9725</td>
<td>0.9852</td>
<td>0.0721</td>
<td>0.0561</td>
<td>0.0424</td>
<td>0.0305</td>
</tr>
<tr>
<td>7</td>
<td>0.9774</td>
<td>0.9680</td>
<td>0.0595</td>
<td>0.0489</td>
<td>0.0454</td>
<td>0.0358</td>
</tr>
<tr>
<td>8</td>
<td>0.9877</td>
<td>0.9910</td>
<td>0.0320</td>
<td>0.0198</td>
<td>0.0314</td>
<td>0.0211</td>
</tr>
<tr>
<td>Average</td>
<td>0.9775</td>
<td>0.9784</td>
<td>0.0592</td>
<td>0.0457</td>
<td>0.0449</td>
<td>0.0346</td>
</tr>
</tbody>
</table>

Figure 6.3: Auditory feedback’s timing error
Figure 6.3 introduces a scheme for evaluating the accuracy of phase estimates. Let us assume that we have a hypothetical phase threshold $\phi_T$ as depicted in Figure 6.3. Based on the ground truth phase signal (red line), as the true phase passes $\phi_T$, an associated auditory response will be generated at time $t_j$. However, the estimation error in our phase estimate (blue line) results in triggering the same auditory cue as the estimated phase signal passes $\phi_T$ at time $t_i$. The error measure $\Delta t_i = |t_i - t_j|$ therefore indicates the timing error associated with auditory cue. We use this method to measure the accuracy of generated feedback by calculating the average error measure, $\mu_{\Delta t}$, and the standard deviation, $\sigma_{\Delta t}$, of all such timing errors through the consecutive periods of the movement. Table 6.1 presents the calculated timing error measures for all of our subjects walking at a slow and fast speed. As the results show, the average timing error is less than 60 milliseconds. To have an idea of the magnitude of this error, let us consider a standard march tempo that is approximately equivalent to 120 beats per minute. This tempo suggests that a standard march step takes about half second. Thus, our timing error of 60 milliseconds is equal to $\frac{1}{8}$ of a step. In musical sense, given that each step is quarter note, our average timing error will be less than a $32^{nd}$ musical note that is considered as a short time interval within the sensitivity of human ears.
Chapter 7

Conclusion

The work presented in this thesis demonstrates a method for generating rhythmic auditory feedback synchronized to a periodic human movement in real-time. A computer, a treadmill and a conventional RGB camera makes our system fairly affordable and available to general consumers. A key advantage of our system is minimizing interference with user’s performance during the movement by relying on marker-less visual measurements for synchronization. Our qualitative and quantitative evaluation scheme confirms that visual measurements obtained from a conventional RGB camera suffice for synchronizing a subject’s walking gait against a computer model. Although we demonstrate the reliability, stability and accuracy of our motion feedback system in the context of walking gait example by using an incorporated gait model, there is no strict requirement for the type of periodic movement or model that can be used within our system. Indeed, as long as our measurement model is able to reasonably measure the distance between model and subject silhouettes using the Hausdorff distance, we expect our synchronization scheme to provide reliable phase estimates. As a result, we believe that our system is not limited to walking gait and can account for a broader category of periodic human movements.

One area that has room for improvement in our phase-entrained particle filter is our measurement model. Although the Hausdorff distance provides a reasonable measure of similarity between particle and model silhouettes for the walking gait example, in the case of more complex movements, such as specific dance moves, our measurement model may fail to measure the distances with the precision required by our synchronization scheme. To overcome this limitation, we speculate that replacing the Hausdorff distance with a more accurate measure of similarity may be helpful.
While our synchronization system introduces a necessary precursor to using rhythmic auditory feedback for clinical treatment of gait disorders, evaluating the effectiveness of our method in therapeutic applications is beyond the scope of this thesis. A future path of this research may include customizing the presented motion feedback system for clinical use so that it accounts for the specific therapeutic needs. For instance, our system may fail to synchronize to gait patterns resulting from specific gait disorders that are too far from the model. A potential solution to account for such differences is to use a hybrid gait model capable of choosing the best set of models that fit the gaits. To utilize our system for refining a gait pattern, one may need to devise a specific sonification scheme tailored based on the nature of a gait disorder so that the feedback encourages progressive corrections. In conclusion, although the existing system needs certain refinements for therapeutic usages, we believe the work presented here takes rhythmic sonification one step closer to clinical applications.
Bibliography


[65] Louis Albert Necker. Observations on some remarkable optical phenomena seen in switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid. 1832.


