Robust Prediction of Auditory Step Feedback for Forward Walking

Markus Zank,[*] Thomas Nescher,[†] Andreas Kunz[‡]
Innovation Center Virtual Reality (ICVR)
Institute of Machine Tools and Manufacturing
ETH Zurich

Abstract

Virtual reality systems supporting real walking as a navigation interface usually lack auditory step feedback, although this could give additional information to the user e.g. about the ground he is walking on. In order to add matching auditory step feedback to virtual environments, we propose a calibration-free and easy to use system that can predict the occurrence time of stepping sounds based on human gait data.

Our system is based on the timing of reliably occurring characteristic events in the gait cycle which are detected using foot mounted accelerometers and gyroscopes. This approach not only allows us to detect but to predict the time of an upcoming step sound in real-time. Based on data gathered in an experiment, we compare different suitable events that allow a tradeoff between the maximum precision of the prediction and the maximum time by which the sound can be predicted.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality H.1.2 [Models and Principles]: User/Machine Systems—Human factors;

Keywords: step prediction, gait, human walking, virtual reality, auditory feedback, step sound

1 Introduction

One of the goals in virtual reality (VR) is to present an immersive environment that resembles reality in every possible way. Previously, it has been shown that a user's feeling of presence in the virtual environment is much higher in VR systems that offer real walking based navigation instead of joystick or keyboard as a navigation method [Usoh et al. 1999; Ruddle and Lessels 2009]. In such systems, the user can see a virtual environment through a head mounted display and move around in the environment by walking in the real world. Since the user is tracked (head position and orientation), the experienced self motion matches the motion seen by the user in the virtual environment.

However, there is also an auditory component to walking that depends on the virtual ground and the surroundings. If, for example, the user is walking on virtual gravel, the system should play a matching sound for his steps, or if he is moving in a large open space like a cathedral, an appropriate reverb effect should be added. To realize this, the real auditory step feedback has to be replaced by a synthetic one, necessitating the following three requirements: First, the user has to wear headphones to block out the real step sound and to provide the synthetic one. Second, the correct time for the step sound has to be found, and third, a sound matching the step and the virtual surroundings has to be generated.

There are certain latencies in such a system, especially a delay between the triggering of the sound playback and the actual playback. We therefore need a system that is not only capable of detecting the right time for an auditory step feedback, but one that can predict it early enough and with sufficient precision that the timing of the synthetic step sound matches the real one. Also such a system should work for different users without calibration and therefore must be adaptive to differences in walking patterns between these people.

2 Related Work

Generating such a synthetic auditory step feedback for a user in a virtual environment basically consists of two separate problems: step detection and sound synthesis. Physically-based sound synthesis is an independent topic and a number of models were already presented, for example in [Avanzini et al. 2005] where a model has been presented which has previously been used to create synthetic step sounds [Turchet et al. 2010]. Step detection is closely related to medical research and gait analysis, where for example Pappas et al. designed a step phase detection system for functional electrical stimulation [Pappas et al. 2001]. In [Turchet et al. 2010], the viability of auditory step feedback systems using shoes equipped with force sensitive resistors has been demonstrated and in [Nordahl et al. 2011], a system based on an array of microphones located on the floor around the user was used.

Menzer et al. investigated the feeling of agency over step sounds in relation to an artificially introduced delay [Menzer et al. 2010]. The results show that the acceptance of the sound decreases with increasing delay between the step and its auditory feedback. For a delay of 100 ms the acceptance is still around 90%. Nordahl stated in a different study that users started to notice a time difference between haptic and auditory step feedback once the delay was bigger than 60.9 ms [Nordahl 2005]. Occelli et al. summarized different results of studies on temporal order judgement and found values ranging from 20 to 75 ms for the audio-tactile condition [Occelli et al. 2011]. However, in constrast to Nordahl's and Merzer's work, these values are not from experiments with walking but with various other tactile stimuli.

Altough there are systems for generating auditory feedback for walking in virtual environments, none of the previously available systems is capable of predicting the occurrence of the auditory step feedback because of the use of force or acoustic measurements. Furthermore, they employ ground based equipment, like a microphone array or force sensor plates, or custom-built shoes which limits the ease of use and portability. Therefore, we propose an accelerometer and gyroscope based system together with suitable prediction algorithms that allow for a synthetic auditory feedback to be played at the time at which the real auditiory feedback would

[*] e-mail:zank@iwf.mavt.ethz.ch
[†] e-mail:nescher@iwf.mavt.ethz.ch
[‡] e-mail:kunz@iwf.mavt.ethz.ch

VRST 2013, October 6–9, 2013, Singapore.
2013 Copyright held by the Owner/Author. Publication rights licensed to ACM.
ACM 978-1-4503-2379-6/13/10 $15.00

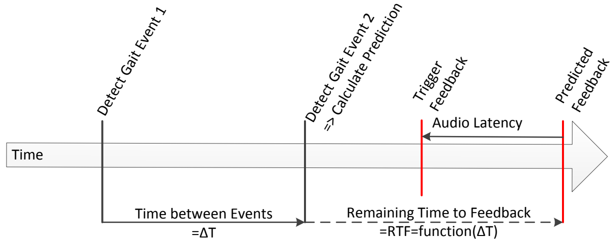


Figure 1: *For predicting the feedback based on the time difference between events, the triggering time has to be earlier due to the audio system's latency.*

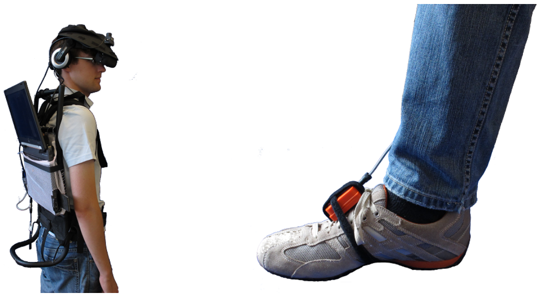
occur in human gait. The system does not need any user calibration and is low cost.

3 Gait Event Predictor

3.1 Sensors and Hardware

Since we want a wearable system that is able to predict the time of the auditory step feedback, it is not possible to use a setup based on force sensitive resistors like [Nordahl et al. 2011] did. Instead, an inertial measurement unit equipped with a 3D accelerometer, gyroscope and magnetometer is used. It is attached to the top of the user's shoe (see Figure 2b and see [Foxlin 2005; Stirling et al. 2003]). The sensor is connected to a backpack worn laptop which runs the necessary software and provides the auditory feedback to the user wearing headphones.

Figure 2 shows the current setup of our VR system in which this auditory step feedback will be used.



(a) *VR system* **(b)** *Used sensor setup*

Figure 2: *VR system, consisting of backpack carried laptop, head mounted display, headphones and head tracking system (a) and the used sensor setup (b).*

3.2 Gait Pattern

The topic of human locomotion has been researched for a long time. Essentially, it is a cyclic process with the same basic pattern repeated every step. In [Pappas et al. 2001] and [Willemsen et al. 1990], the step was divided into four phases: Stance, heel-off, swing, and heel-strike.

Figure 3 shows a typical step measured using the sensor setup described above as well as the corresponding step phases and foot movements. On the one hand, there are differences in the gait cycle that are characteristic for the person. On the other hand, there is a clear structure in the gait cycle that is independent of the person.

Table 1: *Possible choices for c_i. One or more c_i together model the relation between the time of the gait events a and b and the RTF*

c_i	Description
1	constant offset
$T_a - T_b$	time difference of events a and b
$(T_a - T_b)^2$	squared time difference of events a and b

3.3 Predictor

Wendt et al. showed that the duration of the swing phase scales linearly with the step duration [Wendt et al. 2010]. Based on this, we try to find points within the common structure of the step that show a similar behaviour. Using the occurrence time of those events, we have to find a relation to the time the auditory step feedback begins at (Figure 1). If we succeed in finding such events, we will be able to calculate the remaining time to the auditory step feedback (RTF) based on those events. By this means, we are able to predict the time of the auditory feedback independent of the user and without any calibration.

In order to achieve a reliable prediction for any user, the events used for the prediction have to be user invariant. This makes events based on thresholds unsuited, because, due to the differences in gait between people, thresholds must either be user dependent or very low and therefore prone to noise and misdetections. Instead, we use the zero crossings in following measurements: Forward acceleration, upwards acceleration, and the roll rate around the mediolateral axis. The zero crossings are easy to detect by the change in the sign and appear reliably in every step.

Since the used zero crossings are part of a transition between two peaks, the data does not oscillate around zero. Only for the foot roll rate, three individual zero crossings within a few milliseconds can occur (up to approximately 15 ms), in which case only the first one is used and the others are rejected. Figure 3 shows a typical step, the corresponding foot movements and the following four events based on zero crossings:

① Foot roll rate downwards zero crossing
② Forward acceleration zero crossing
③ Up acceleration zero crossing
④ Foot roll rate upwards zero crossing

Based on the time between two or more of those events, we predict the RTF after the latest event (1).

$$RTF = a_1 \cdot c_1 + a_2 \cdot c_2 + ...a_N \cdot c_N \qquad (1)$$

c_i can be the time difference between two of the above events, a function of a time difference, or a constant (see Table 1). The constant factors a_i are determined using a standard linear least squares approach based on training data with known feedback times (2), where the columns in matrix C correspond to different c_i of the same step and every row to a different step (3). $\overrightarrow{RTF}_{Training}$ is a vector containing the corresponding true time to feedback.

$$A = [a_1, a_2, ...a_n]^T = (C^T \cdot C)^{-1} \cdot C^T \cdot \overrightarrow{RTF}_{Training} \qquad (2)$$

$$C = [\overrightarrow{c_1}, \overrightarrow{c_2}, ..., \overrightarrow{c_N}] \qquad (3)$$

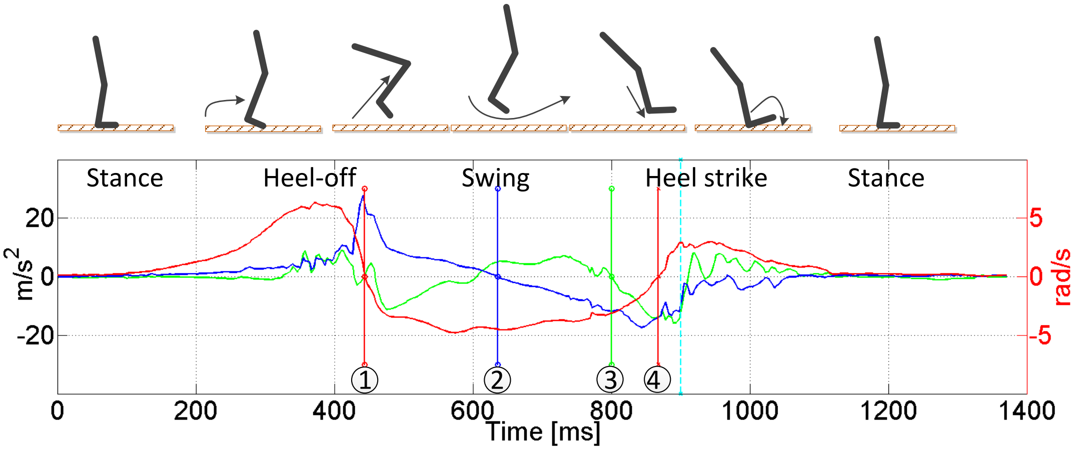


Figure 3: *The plot shows the upward acceleration (green), forward acceleration (blue) and roll rate (red) of a single step together with the step phases. The upper part shows the corresponding foot movements. ①-④ mark the locations of the person invariant gait events and the beginning of the auditory step feedback (cyan, dashed).*

4 Experiment

To gather data for training and evaluating the predictors, an experiment was conducted. 10 participants (2 female, 8 male) took part in this experiment. No participant wore shoes that have caused difficulties in attaching the sensor or an unusual auditory feedback. The VR setup described in section 3 was used. However, for this experiment the head mounted display, headphones and tracking system were omitted. Additionally, a microphone was attached to the right foot to record the real step sound for determining the true time of the feedback.

The participants were asked to walk to the other side of the room and back in normal speed on a straight line, while their movements and sounds were recorded. They were informed that there was an audio recording running and that they should not talk during the walking. For every participant the walking was conducted twice.

5 Results

The audio data from the experiment was filtered using a bandpass filter to remove noise and low frequency distortions caused by the movement. In the resulting signal, the auditory step feedback was tagged manually at the beginning of the sound and only steps that had an unambiguous sound feedback and a clear beginning were included. The turn steps at the far end of the room were excluded. This provided a total of 154 steps for the analysis, in which every participant has at least 11 steps.

5.1 Predictor Performance

Using the approach presented in section 3, different combinations of the proposed gait events are evaluated. Constant, linear and quadratic terms are included for c_i and the factors a_i are calculated. Then, the deviation of the RTF from the actual remaining time is evaluated and the overall standard deviation σ of this prediction error is calculated as well as the mean RTF. Since the mean error is zero due to the least squares approach, σ^2 is also the mean squared error of the predictor. This provides a measure for the robustness and the prediction capability of the predictor.

Since there are a lot of possible event combinations, Table 2 shows a selection of the best predictors. The table also states the error between the RTF and the actual remaining time until feedback. This is evaluated using a leave-one-out cross validation where the predictor is applied to one participant after the other, using the other 9 to train the factors A.

6 Discussion

The most precise predictors (I and IV) reach a σ of around 16 ms. If we compare this result to the limits stated in section 2, those predictors fulfil our robustness requirements very well. In constrast to σ, the mean RTF depends only on the used events. Predictors using event ④ have an average RTF of 23.8 ms. Depending on the used hard- and software, this may or may not offer enough time to generate and trigger a playback in time. However, since σ is so small, even if the feedback is delayed, it should not be noticeable by the user, if the overall system latency is small enough. In our case with a audio latency (AL) of 30 to 40 ms, this should still be acceptable. For more than 98% of the steps, the prediction error is within $\pm 3 \cdot \sigma$. The error can therefore be expected to be between -35.7 and 58.5 ms (4).

$$AL - RTF \pm 3 \cdot \sigma = 35 - 23.6 \pm 3 \cdot 15.7 \tag{4}$$

The predictor II uses event ③ as last event and therefore has a much higher expected RTF of around 87 ms, but it also has a higher σ. This means that, compared to the predictor including event ④, we have to accept a higher σ in order to get a higher RTF. When looking at the predictor only using events ① and ②, this behaviour is confirmed, at an expected RTF of 220 ms, σ is 31 ms (predictor III in Table 2). With this standard deviation, the users might notice a delay in the auditory feedback for their steps, but the upper limits of the acceptance range stated above can still be met.

Moreover, such a high RTF will usually not be necessary for an auditory step feedback and even if this is the case, it could be considered to use this only as a rough estimate for the initial feedback preparations and use a later event for the actual triggering of the feedback.

The user independance and calibrationlessness requirements are also fulfilled, since even for the cross validation condition, where the user is unknown to the predictor, the prediction error was below the maximal acceptable value for every participant.

Table 2: *Predictor comparison. The table shows the used events, the formula for the RTF, the mean RTF and the standard deviation σ of the RTF from the true remaining time until the auditory feedback with T_i = time of event i. The last column shows the error mean and standard deviation from the cross validation.*

Predictor	events used	$RTF = A^T \cdot C$ [ms]	mean RTF [ms]	σ [ms]	$mean(error)$ $\pm\sigma(error)$
I	$\Delta T = T_4 - T_2$	$RTF = -0.0025 \cdot \Delta T^2 - 1.0187 \cdot \Delta T - 78.1424$	23.6	15.7	0.4 ± 16.8
II	$\Delta T = T_3 - T_1$	$RTF = -0.1581 \cdot \Delta T + 66.3783$	88.0	21.3	0.9 ± 23.8
III	$\Delta T = T_2 - T_1$	$RTF = -0.0049 \cdot \Delta T^2 - 1.0656 \cdot \Delta T + 207.9707$	218.8	31.1	2.5 ± 34.3
IV	$\Delta T = T_4 - T_1$	$RTF = -0.0018 \cdot \Delta T^2 - 1.4747 \cdot \Delta T - 279.3835$	23.6	16.0	0.0 ± 17.7

7 Conclusion and Future Work

The presented approach for the prediction of auditory step feedback based on accelerometers and gyroscopes is calibration-free, needs no stationary equipment or custom made shoes. It is capable of predicting the time of the step feedback which allows reducing the overall system latency.

The prediction is based on the time difference between characteristic gait events and works well for healthy forward walking. It is possible to achieve a prediction error that is below the value that is noticeable by the user (see section 2). This shows that choosing zero crossings of measurement values as gait events is a reliable and robust approach. One of the best predictors has the additional advantage of only using measurements of the foot roll rate and therefore requires only one single-axis gyroscope per foot, although the time by which the step sound can be predicted is short. However, if the prediction is required earlier, it is possible to use events based on the upward and forward acceleration. Those predictors are less precise, but still within the required limits. But since they are based on the upward and forward acceleration, it is necessary to have additional sensors.

By design, the predicted time corresponds to the beginning of the acoustic step feedback on a flat, rigid surface. For those surfaces, only the chosen sound file has to be replaced. For surfaces that can generate sound before the foot hits the floor, like tall grass or snow, it would be necessary the adapt and retrain the predictor.

In future work, the number of different detectable step types can be improved, including e.g. backwards walking, stomping, sneaking, or turning on the spot. Also more parameters of the step could be estimated with the goal of using them as input for a physically-based synthetic sound generation. Furthermore, the user acceptance of the auditory step feedback should be analysed in detail. Especially the maximum acceptable time difference between real and synthetic sound as well as the effects of early feedback compared to late feedback.

Acknowledgements

The authors would like to thank the Swiss National Science Foundation (project number 127298) for funding this work.

References

AVANZINI, F., SERAFIN, S., AND ROCCHESSO, D. 2005. Interactive simulation of rigid body interaction with friction-induced sound generation. *Speech and Audio Processing, IEEE Transactions on 13*, 5, 1073–1081.

FOXLIN, E. 2005. Pedestrian tracking with shoe-mounted inertial sensors. *Computer Graphics and Applications, IEEE 25*, 6, 38–46.

MENZER, F., BROOKS, A., HALJE, P., FALLER, C., VETTERLI, M., AND BLANKE, O. 2010. Feeling in control of your footsteps: Conscious gait monitoring and the auditory consequences of footsteps. *Cognitive Neuroscience 1*, 3, 184–192.

NORDAHL, R., TURCHET, L., AND SERAFIN, S. 2011. Sound synthesis and evaluation of interactive footsteps and environmental sounds rendering for virtual reality applications. *Visualization and Computer Graphics, IEEE Transactions on 17*, 9, 1234–1244.

NORDAHL, R. 2005. Self-induced footsteps sounds in virtual reality: Latency, recognition, quality and presence. *Presence*, 353–354.

OCCELLI, V., SPENCE, C., AND ZAMPINI, M. 2011. Audiotactile interactions in temporal perception. *Psychonomic bulletin & review 18*, 3, 429–454.

PAPPAS, I. P., POPOVIC, M. R., KELLER, T., DIETZ, V., AND MORARI, M. 2001. A reliable gait phase detection system. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 9*, 2, 113–125.

RUDDLE, R. A., AND LESSELS, S. 2009. The benefits of using a walking interface to navigate virtual environments. *TOCHI '09: Transactions on Computer-Human Interaction 16*, 1, 1–18.

STIRLING, R., COLLIN, J., FYFE, K., AND LACHAPELLE, G. 2003. An innovative shoe-mounted pedestrian navigation system. In *Proceedings of European Navigation Conference GNSS*.

TURCHET, L., NORDAHL, R., SERAFIN, S., BERREZAG, A., DIMITROV, S., AND HAYWARD, V. 2010. Audio-haptic physically-based simulation of walking on different grounds. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, IEEE, 269–273.

USOH, M., ARTHUR, K., WHITTON, M. C., BASTOS, R., STEED, A., SLATER, M., AND BROOKS, JR., F. P. 1999. Walking > walking-in-place > flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, ACM, SIGGRAPH '99, 359–364.

WENDT, J., WHITTON, M., AND BROOKS, F. 2010. Gud wip: Gait-understanding-driven walking-in-place. In *Virtual Reality Conference (VR), 2010 IEEE*, IEEE, 51–58.

WILLEMSEN, A. T. M., BLOEMHOF, F., AND BOOM, H. B. 1990. Automatic stance-swing phase detection from accelerometer data for peroneal nerve stimulation. *Biomedical Engineering, IEEE Transactions on 37*, 12, 1201–1208.