ETH zürich

ETH Library

Predicting audio step feedback for real walking in virtual environments

Working Paper

Author(s):
Zank, Markus; Nescher, Thomas; Kunz, Andreas

Publication date:
2014

Permanent link:
https://doi.org/10.3929/ethz-a-010277703

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Predicting Audio Step Feedback for Real Walking in Virtual Environments

Markus Zank
Innovation Center Virtual Reality (ICVR)
Institute of Machine Tools and Manufacturing
ETH Zurich Tel. (+41)44 633 83 78
email: zank@iwf.mavt.ethz.ch

Thomas Nescher
Innovation Center Virtual Reality (ICVR)
Institute of Machine Tools and Manufacturing
ETH Zurich Tel. (+41)44 632 46 69
email: nescher@iwf.mavt.ethz.ch

Andreas Kunz
Innovation Center Virtual Reality (ICVR)
Institute of Machine Tools and Manufacturing
ETH Zurich Tel. (+41)44 632 57 71
email: kunz@iwf.mavt.ethz.ch

November 21, 2014

Abstract

When navigating in virtual environments by using real walking, the correct auditory step feedback is usually ignored, although this could give more information to the user about the ground he is walking on. One reason for this are time constraints that hinder a replay of a walking sound synchronous to the haptic step feedback when walking. In order to add a matching step feedback to virtual environments, this paper introduces a calibration-free system which can predict the occurrence time of a step-down event based on an analysis of the user's gait. For detecting reliable characteristics of the gait, accelerometers and gyroscopes are used that are mounted on the user's foot. Since the proposed system is capable of detecting the characteristic events in the foot's swing phase, it allows a prediction that gives enough time to replay sound synchronous to the haptic sensation of walking. In order to find the best prediction regarding prediction time and accuracy, data gathered in an experiment is analyzed regarding reliably occurring characteristics in the human gait. Based on this, a suitable prediction algorithm is proposed.

Introduction

Increasing immersion in virtual environments is an important goal in VR research. Usoh et al. [1] and Ruddle et al. [2] showed that for increasing the feeling of presence, real walking as a navigational method is superior to stepping in place, and joystick or keyboard interaction. In such systems, a head-mounted display is used to visualize the virtual environment while walking around. The user's position and orientation is tracked, which allows adapting the visual feedback accordingly. Therefore, the user experiences a self-motion that matches the motion seen by him in the virtual environment. However, such tracking systems do not give any information about the user's foot placement and thus cannot be used to trigger a correctly synchronized replay of walking sounds.

To further increase immersion for real walking in a virtual environment, an auditory component could be added which would give information about the ground the user is walking on. Depending on the current virtual environemnt, the sound could be different, such as walking on concrete, gravel, or snow. Moreover, the acoustic characteristics of the environment could also be included, e.g. reverb effects in a cathedral. Nordahl et al. showed in a study that a correct auditory feedback can significantly increase immersion [3]. Thus, an immersive VR system must able to block the real sound of the walking step, while replaying a synthetic sound instead that exactly fits to the experienced VR environment regarding sound character and timing. This imposes the following requirements on the system:

- Headphones are required to block the real step sound and to provide a synthetic one instead.
- The sound must be replayed at the correct time so that it is synchronous to the haptic step sensation.
- The sound signal must match the virtual environment regarding sound characteristics and echo, but also the physical properties of the ground the user is walking on.
- The system should work reliably for any user and ideally without preliminary calibration or training phase.

Compared to real world, there are certain latencies in such a system as shown in Figure 1.

While in the real world, the auditory step feedback would have a latency of $4 - 6$ ms, the virtual environment has a much higher latency, consisting of three main parts: sensor delay given by the sensor, sensor update rate and used connection (~ 6 ms), the used audio hardware (~ 35 ms) and the software used for replaying the sound. While not based on exactly the same setting, the measurements done by Wang et al. [4] illustrate the underlying problem regarding latencies in consumer grade audio hardware that is also the cause for the 35 ms latency in our case. We therefore need a system that is capable of determining the right time for an auditory step feedback, but can also predict

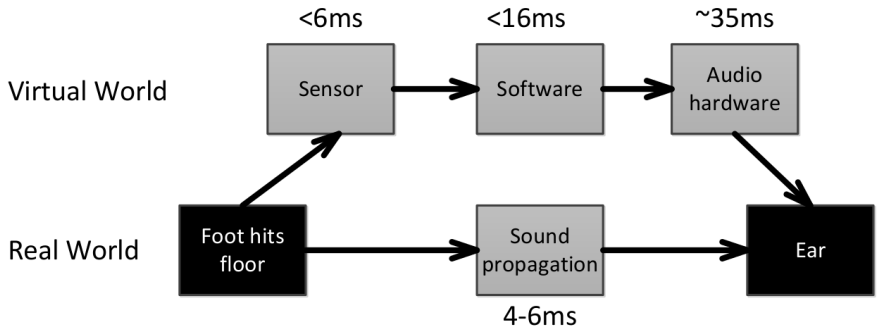


Figure 1: Comparison between real and virtual world regarding occurring latencies.

it early enough and with a sufficient precision to guarantee that the timing of the synthetic sound matches the real one.

Related Work

The sound of our steps gives us information about the material and structure of the ground we are walking on. Giordano et al. researched the ability of people to identify ground materials by non-visual means [5]. While the amount of information depends on the simulated material, the distinction between solid (wood, concrete,...) and aggregate surfaces (gravel) seems to be very easy even if only auditory cues are available. Serafin et al. even showed that users perform better at identifying ground materials if only auditory cues are provided instead of haptic ones [6].

Increasing the immersion of a virtual environment by generating such a synthetic auditory sound feedback poses the problem of step detection and sound synthesis. Within the research field of physically based sound synthesis, numerous sound synthesis models were already presented. There was a model presented by Avanzini et al. [7], which was used to generate synthetic step sounds by Turchet et al. [8]. Step detection on the other hand is mainly done in the medical research field, and in particular in gait analysis. Pappas et al. [9] also designed a step phase detection system for a functional electrical stimulation.

Turchet et al. used shoes equipped with force sensitive resistors to demonstrate the viability of an auditory step feedback [8]. Another approach was introduced by Nordahl et al. [10], who used an array of microphones that were integrated in the floor the user was walking on. Law et al. presented another floor based system that is used in a CAVE system and provides a visual, haptic and auditory virtual ground [11].

As shown above, a number of systems exist that provide auditory feedback for walking in virtual environments. However, none of these systems is capable

of predicting the occurrence time of the auditory step feedback, since they use force or acoustic measurements, such as microphone arrays, force sensor plates, or custom-built shoes that are equipped with sensors. All systems have in common that they measure the real step-down time and thus typically do not leave enough time to synchronously replay an artificial sound.

The so-called "feeling of agency" is a psychological measure for a person claiming resposibility for certain events; in this case having caused the step sound with their walking. This feeling of agency was investigated by Menzer et al., who measured the influence of an artificially introduced delay between the haptic feedback of the step-down event and the acoustic sensation [12]. They showed that the acceptance of a sound sensation decreases with an increasing delay between the step and the acoustic feedback. But even for a delay of 100 ms, 90% of the participants still accepted the sound as their own. In another user study, Nordahl found out that users started to notice the time difference between haptic and auditory feedback once the delay was above 60.9 ms [13]. However, these findings are in contrast to research by Occelli et al. who also performed studies on temporal order judgment [14]. They found the perception threshold for the delay in the audio-tactile perception to be between 20 - 75 ms. The difference might be explained by the fact that in contrast to Nordahl's [10, 13] and Menzer's [12] work, these values were not from experiments with walking, but with various other tactile stimuli.

To overcome the problem of delayed sound replay, this paper introduces a system that uses accelerometers and a gyroscope together with suitable prediction algorithms which allow for a synthetic auditory feedback being replayed at the exact moment when the real auditory feedback should occur during human gait. This is possible since the system can measure data during any phase in human gait and not only during the stance phase (see Figure 2). In addition, the proposed system does not need any user calibration and is low-cost.

Gait Event Predictor

Sensors and Hardware

Since we want a wearable system that is able to predict the time of the auditory step feedback, 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 3 and also [15, 16] for similar setups). The sensor is connected to a backpack worn laptop which runs the prediction software, the rendering engine, and provides the auditory feedback to the user wearing headphones.

The used sensor is an Xsens MTx inertial measurement unit running at 200 Hz connected via USB to a notebook with an i7-2760QM quad core cpu @ 2.4 GHz and 8 GB main memory.

Figure 4 shows the system with all components, including the head-mounted display, the headphones and the tracking system.

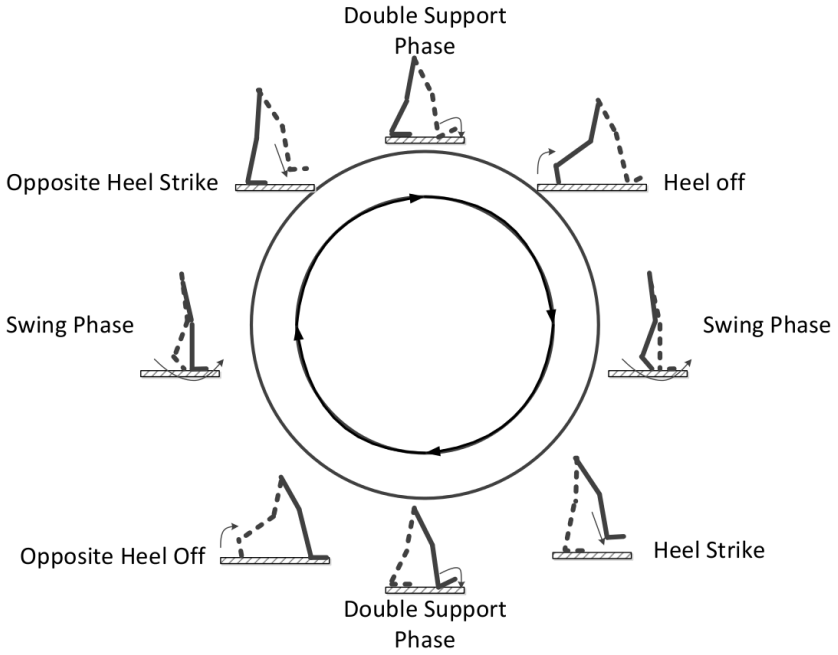


Figure 2: The human gait cycle as used by Wendt et al. [19] based on the definition by Inman et al. [21]

Gait Pattern

Human locomotion has been a research topic for a long time. It is essentially a cyclic process as depicted in Figure 2. This means that there is a basic pattern that is repeated in each step, alternating between left and right. Pappas et al. [9] and Willemsen et al.[17] both divided the step into four phases: Stance, heel-off, swing, and heel-strike. While this cycle is not completely identical for different people, it is very similar [18].

This repetitive gait pattern should also be visible in the signals measured with the sensors mentioned above. In Figure 6, the sensor signals for one single step in regular forward walking are depicted together with the corresponding phases in the foot movement. The solid line shows the signal from the gyroscope, measuring the foot roll rate. The dashed and the dotted lines show the signals from the accelerometers, measuring the foot's forward and upward accelerations.

Predictor Realization

Wendt et al. showed that the duration of the swing phase scales linearly with the step duration [19]. Based on this observation, we propose an approach for predicting the time of the step sound based on a set of person-independent,

Figure 3: Used sensor setup

reliably occurring and unambiguous "events".

Based on these events, we look for a relation between them that allows us to predict the time the auditory step feedback should begin at. The output of the predictor is the remaining time to the auditory step feedback (RTF) after the latest used event. Figure 5 shows the design principle.

Gait events

The gait events used for prediction have to fulfill the following criteria:

- They have to occur for every person
- They have to occur in every step
- Based on their time of occurrence it has to be possible to estimate the time until the step sound occurs
- They can be detected robustly

To find events that fulfill those requirements, we limit ourselves to forward walking at normal speed and users with healthy gait.

There are a number of points in the gait cycle one might consider for gait events. In the following section, we will present some of them and discuss their suitability for being used as gait events. The most obvious events are maxima or minima in the measured signal. However, there are a number of difficulties in using them. First, we need to be sure that a given point is not just a local maximum (Figure 7a), since this would result in a wrong prediction time. Therefore a certain waiting time is required to be sure that no other maximum

Figure 4: VR system

would occur. However, this would add an additional delay to the predictor. Moreover, it would be difficult to define an optimal wait time. When using maximum values as characteristic events, another problem is that the measured signals do not always possess a distinct maximum, i.e. a peak value that could be easily detected. Instead, signals have a flat maximum (see solid line for the angular velocity of the roll rate in Figure 6) which makes it difficult to define the exact occurrence time of such a maximum (see Figure 7b). Furthermore, if such a signal is noisy, determining the exact occurrence time becomes even more imprecise. A peak that is easy to detect would be a high narrow one as in Figure 7c. However, these peaks often occur in groups at the beginning and end of the step. The ones at the end are after the auditory step feedback and therefore useless for a prediction. For both cases, it is unclear which peaks belong to the characteristic gait cycle and which ones do not.

Another possibility could be to define a certain threshold and using the crossing of this threshold as event. However, this poses the question of a good choice of the threshold. Although the basic locomotion pattern is similar between people, the amplitude of the walking pattern differs. Therefore, it is difficult to define a threshold that is triggered by everyone even for normal walking.

Thus the most suitable approach is to use zero crossings. In general, if the zero crossing occurs with high gradient, there will be only one distinct zero crossing even with sensor noise or small jitters in the movement. Figure 7d shows a zero crossing from actual walking which exihibits this behavior due to

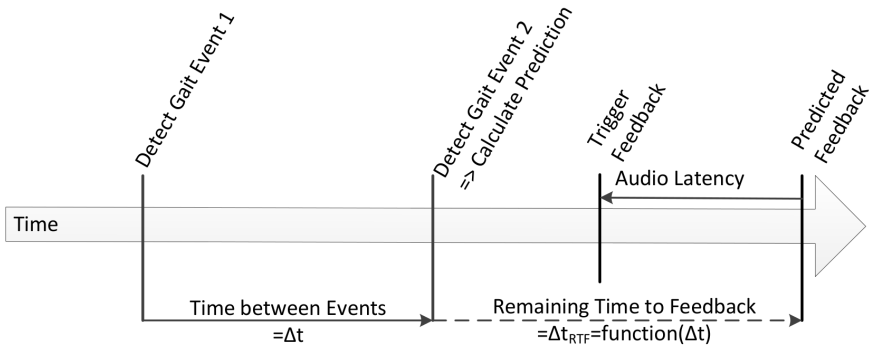


Figure 5: For predicting the feedback based on the time difference between events, the triggering time has to be earlier due to the audio system's latency.

their location in the gait cycle. This makes zero crossings a good choice for gait events.

We therefore define the following four events :
① Foot roll rate downwards zero crossing
② Forward acceleration zero crossing
③ Up acceleration zero crossing
④ Foot roll rate upwards zero crossing

Figure 6 shows a typical step, the corresponding foot movements, and the four events defined above. These four events will be used to define a suitable prediction algorithm that will be introduced in the next paragraph.

Prediction

The goal of the prediction is to estimate the time of the auditory step feedback t_{RTF}. Instead of calculating this as an absolute time, we calculate $\Delta t_{RTF} = t_{RTF} - t_k$ where t_k is the time of the last occuring event used in the prediction (cf. Figure 5). Since the prediction is calculated immediately after all necessary events occured, Δt_{RTF} is the time from the moment the prediction is done until the step feedback has to be audible. To calculate Δt_{RTF} a standard linear regression with basis functions is used (1) (defined for example in [20]). Here, a_i is a constant scalar weighting factor and $c_i = f_i(t_m, t_n)$ are the basis functions. t_m and t_n are the absolute times of any two events.

$$\Delta t_{RTF} = \mathbf{a}^T \cdot \mathbf{c} = a_1 \cdot c_1 + a_2 \cdot c_2 + ...a_N \cdot c_N \tag{1}$$

From all possible choices for basis functions f_i, we select the polynomial ones defined in Table 1 for an in-depth evaluation. The use of linear terms is motivated by Wendt's finding of a linear relation between step frequency and time spent in a certain step phase relative to the step duration [19]. Additionally

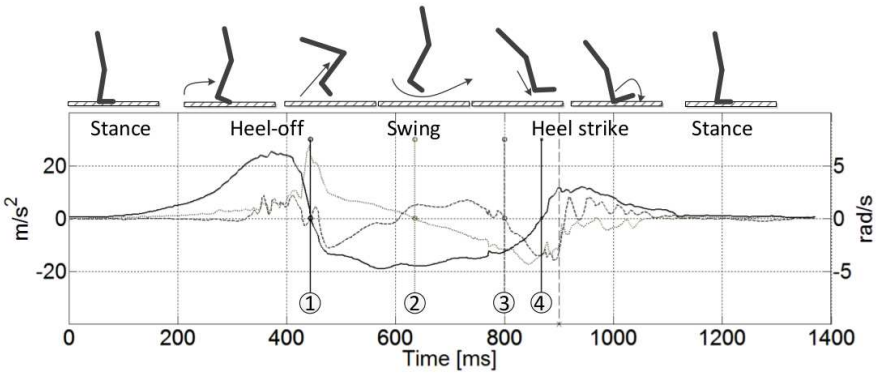


Figure 6: The plot shows the upward acceleration (dashed), forward acceleration (dotted) and roll rate (solid) of a single step together with the step phases. The upper part shows the corresponding foot movements. ①-④ mark the locations of the person invariant gait events and the beginning of the auditory step feedback (thin dashed).

quadratic terms are added in order to evaluate if adding basis functions of higher order can improve the predictor performance.

$\mathbf{a}$ is calculated using real world walking data where the absolute times of all events and the auditory audio feedback are known. Using this data, we can calculate $\mathbf{a}$ using linear least squares (2) with $\mathbf{D} = [\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_M]^T$ and $\boldsymbol{\Delta t}_{RTF} = [\Delta t_{RTF,1}, \Delta t_{RTF,2}, ..., \Delta t_{RTF,M}]^T$ where M is the number of used steps.

$$\mathbf{a} = (\mathbf{D}^T \cdot \mathbf{D})^{-1} \cdot \mathbf{D}^T \cdot \boldsymbol{\Delta t}_{RTF} \tag{2}$$

In order to reduce the number of possible predictors, every predictor has to fulfill the following conditions:

- Not all c_i have to use the same f_i
- $t_m > t_n$
- All c_i use the same t_n
- Multiple c_i can, but do not have to, use the same t_m

Experiment

In order to evaluate the predictors, an experiment was conducted to gather real world data to compare their performance. 10 people (2 female, 8 male) were recruited to perform a walking task. They wore the sensor as depicted in Figure 3 and the laptop from the VR setup (Figure 4) for data recording.

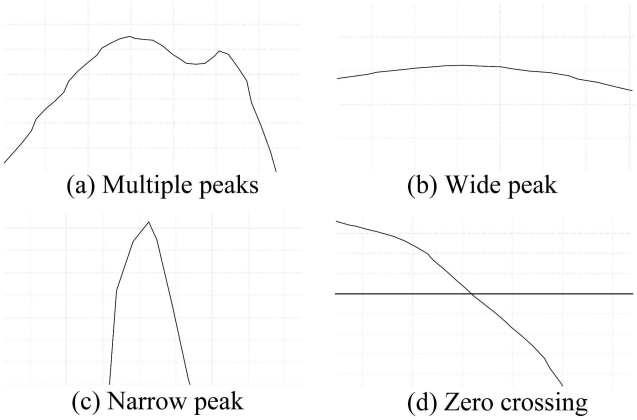


Figure 7: Different cases for defining events

The tracking system, head-mounted display and headphones were not used for the experiment. In order to measure at which point in the step the real sound occurs, we attached an additional microphone at the user's ankle to acoustically determine the true time of the step sound (see Figure 8).

For the experiment, the participants were asked to walk about 24 meters in four runs with sensor and audio recording running. They were instructed to walk in a natural fashion and speed, but were asked not to talk during the experiment because of the audio recording.

Figure 8: Sensor setup and microphone attached to the ankle

Results

The time of the audio feedback time was determined manually. First, the audio data was filtered with a band-pass filter with a pass band from 330 to 10000 Hz to suppress noise and low frequency distortions caused by the leg movement. In the resulting signal, the beginning instance of the step sound was tagged manually. In order to keep the classification robust, all ambiguous steps were discarded. Also the parts in between the four walking parts in the experiment were discarded. This provided a total of 154 steps for the analysis, in which every participant contributed at least 11 steps.

Predictor Performance

Using the approach presented before, any combination of the proposed gait events was evaluated. As an additional variation parameter, polynomials of degree one (e.g. $\Delta t_{RTF} = a_1(t_m - t_n) + a_0$) and two (e.g. $\Delta t_{RTF} = a_2(t_m - t_n)^2 + a_1(t_m - t_n) + a_0$) were used, including combinations of more than two events (e.g. $\Delta t_{RTF} = a_1(t_k - t_n) + a_2(t_m - t_n) + a_3$), resulting in a total of 87 evaluated predictors. For every predictor, $\mathbf{a}$ was calculated using linear least squares. Then, the deviation of the Δt_{RTF} from the actual remaining time was evaluated and the overall standard deviation σ of this prediction error was 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.

As a second condition, a cross validation (CV) was conducted, using the data of 9 users to determine $\mathbf{a}$ which was then applied to the 10^{th} user. This was done for every user and the results were combined.

Since there are a lot of possible event combinations, four predictors were chosen as a selection of representative predictors (Table 2). For a comparison, Table 3 states the error between the Δt_{RTF} and the real remaining time until feedback.

Discussion

Figure 9 gives an overview of all tested predictors in relation to the error threshold of 60 ms (based on [13]) and the required prediction time of 35 ms. It is important to keep in mind that the prediction error is based on human perception, whereas the required average prediction time is based on hardware and software latencies. In general, there is the tendency that the prediction errors become larger, the longer the prediction time is. Thus, a suitable tradeoff has to be found between the maximum acceptable error and shortest feasible prediction time. There are three distinct groups of classifiers visible in Figure 9, each centered at a certain prediction time. This means that for the group with a prediction time of about 225 ms (last event = ②), the prediction error can

be so large that it is noticeable by the user, making these predictors unsuited even though the prediction time is very good. However, the predictors with a prediction time of about 80 ms (last event = ③) fulfill both requirements. The ones with a prediction time around 25 ms (last event = ④) have an even lower error, but cannot meet the prediction requirements of our hardware.

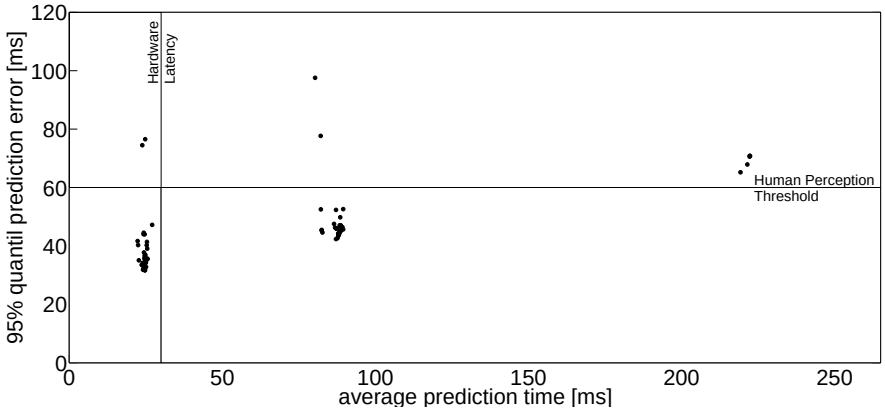


Figure 9: The plot shows the average prediction time and the 95% quantil of the prediction errors where every point represents a predictor. We assume 60 ms as the upper limit for the error and a minimum prediction time of 35 ms. This means that only predictors in the lower right part fulfill both requierments.

Figure 10 shows the four selected predictors, their standard deviation and maximum errors overlaid on an actual step from our experiment. The most precise predictors (I and IV) reach a standard deviation σ of around 16 ms. If we compare this result to the limits stated in the literature, all predictors fulfill the robustness requirements very well. Thus, the second criteria, the prediction time, will be discussed next.

In contrast to σ, the mean Δt_{RTF} depends only on the used events. Predictors using event ④ have an average Δt_{RTF} of 23.8 ms. Depending on the used hard- and software, this may or may not offer enough time to generate and trigger an audio playback in time. However, since σ is so small, even if the feedback is delayed, it should not be noticeable by the user, even though the average error for the replay time is not zero, under the condition that the overall system latency is small enough. In our case, with an audio latency (L_A) 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 (3).

$$L_A - \Delta t_{RTF} \pm 3 \cdot \sigma = 35 - 23.6 \pm 3 \cdot 15.7 \tag{3}$$

The predictor II uses event ③ as last event and therefore has a much higher expected Δt_{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

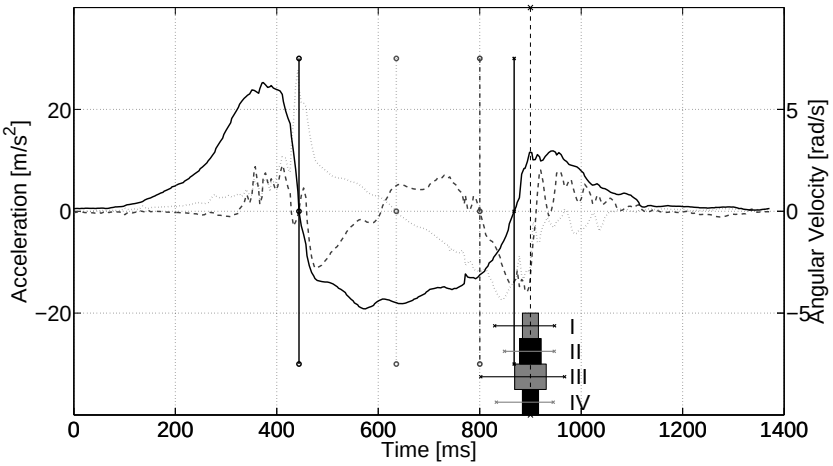


Figure 10: The plot shows a data from a single step from the experiment and the variance, minimum and maximum error of the predictors presented in Table 2 centered around the true feedback time.

order to get a higher Δt_{RTF}. When looking at the predictor using only events ① and ②, this behavior is confirmed, with an expected Δt_{RTF} of 220 ms, σ is 31 ms (predictor III). With this standard deviation, it is possible that the classification error is so large that it can be noticed by the user, if 60 ms is assumed to be the limit. However, the 100 ms boundary based on Menzer's work [12] is still achieved.

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

Table 4 shows the standard deviation per user for the predictors. The individual standard deviation per user is smaller than the standard deviation over all users, which means a user calibration could improve the result, even though it is not required to reach the necessary prediction performance. The high standard deviation of user 8 is caused by a single outlier, due to the small number of samples per user. If this one sample is omitted, σ is reduced to a value normal for the respective predictors.

Figure 11 shows the relation between the degree of the polynomial and the resulting standard deviation for both, the cross validation and non-cross validation condition. The non-cross validation case shows no change in standard deviation depending on the degree of the polynomial, whereas for the cross validation the standard deviation is in some cases much higher. The high difference between non-cross validation and cross validation condition implies some kind of overfitting at higher degrees, because there are certain users for who the

prediction fails completely.

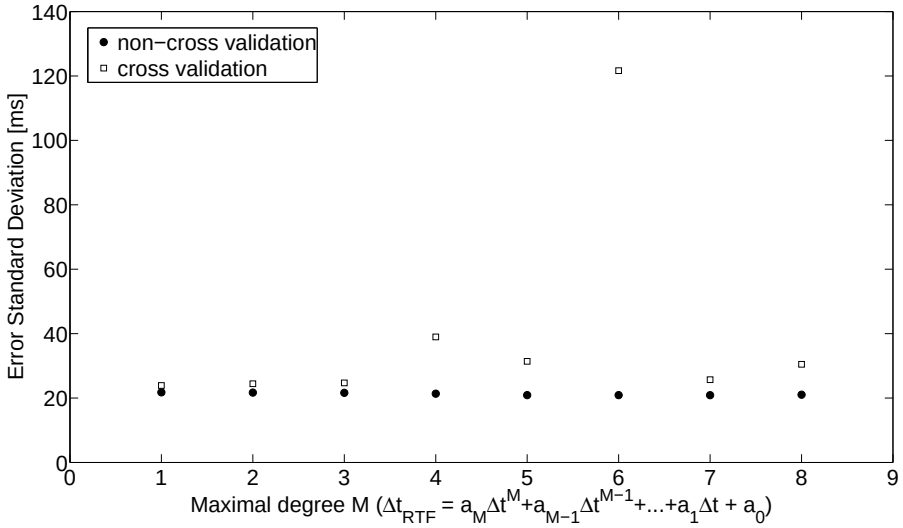


Figure 11: The plot shows the relation between maximal degree of the polynomial and the standard deviation of the prediction error for the cross validation and non-cross validation case.

Table 1: Used choices for c_i. One or more c_i together model the relation between the time of the gait events m and n and the RTF.

c_i	Description
1	constant offset
$t_m - t_n$	time difference of events m and n
$(t_m - t_n)^2$	squared time difference of events m and n

Table 2: Predictor comparison. The table shows the used events and the resulting equation for Δt_{RTF} with t_i = time of event i.

Predictor	events used	Δt_{RTF} [ms]
I	$\Delta t = t_4 - t_2$ [ms]	$\Delta t_{RTF} = -0.0025 \cdot \Delta t^2 + 2.0187 \cdot \Delta t - 78.1424$
II	$\Delta t = t_3 - t_1$ [ms]	$\Delta t_{RTF} = 1.1581 \cdot \Delta t + 66.3783$
III	$\Delta t = t_2 - t_1$ [ms]	$\Delta t_{RTF} = -0.0049 \cdot \Delta t^2 + 2.0656 \cdot \Delta t + 207.9707$
IV	$\Delta t = t_4 - t_1$ [ms]	$\Delta t_{RTF} = -0.0018 \cdot \Delta t^2 + 2.4747 \cdot \Delta t - 279.3835$

Also a comparison with a Gaussian process model showed very similar performance (see Table 5). The Gaussian process used the same time differences

Table 3: Predictor comparison. The table shows the mean Δt_{RTF} and the standard deviation σ of the Δt_{RTF} from the real remaining time until the auditory step feedback. The last column shows the error's mean and standard deviation from the cross validation. See Table 2 for the definition of the predictors.

Predictor	mean Δt_{RTF} [ms]	σ [ms]	mean(error) $\pm\sigma$(error)
I	23.6	15.7	0.4 ± 16.8
II	88.0	21.3	0.9 ± 23.8
III	218.8	31.1	2.5 ± 34.3
IV	23.6	16.0	0.0 ± 17.7

Table 4: Standard deviation per user for all 4 predictors. The higher standard deviation of user 8 is caused by a single outlier (error = -70ms for predictor I).

	User:	1	2	3	4	5	6	7	8	9	10
σ(error)[ms]	I	15.0	8.9	17.3	11.9	7.7	8.9	12.4	26.4	8.7	7.3
	II	11.4	8.9	17.1	9.2	11.8	11.0	20.0	23.0	19.1	7.9
	III	15.6	16.5	20.9	16.3	11.3	27.5	19.4	32.9	35.4	9.7
	IV	14.6	9.6	17.5	12.1	7.7	9.1	11.5	25.3	10.4	7.8

as input and output as the regression approach did. In this case c_i, as defined in (1), becomes $c_i = e^{-\beta||x-b_i||^2}$, where b_i are the center vectors of a Gaussian radial basis function and a_i defines the weight of the respective basis. However, the Gaussian process model has a higher complexity and does not achieve a better prediction performance.

Table 5: Comparison of the demonstrated approach and a Gaussian process model

Predictor	Regression: σ [ms]	Gaussian process: σ [ms]
I	16.8	16.7
II	23.8	25.2
III	34.3	34.6
IV	17.7	16.7

The requirements for user independence and calibration-free operation are also fulfilled, since the evaluation of the cross validation shows that the predictors can reach the required precision and prediction times even for unknown users.

Because the design of these predictors is tailored on the walking pattern for forward walking, we cannot expect them to work for completely different types of walking. For backwards walking for example, the gait cycle is completely different and therefore the events the prediction is based on will not occur in

the expected order if they occur at all.

However, the presented approach of finding key events in the gait cycle and using the time difference between these events in a simple regression model is very flexible and could therefore be adapted to cover other types of walking.

Conclusion and Future Work

This paper presented a system that uses accelerometers and gyroscopes for predicting the correct time for an auditory step feedback in human gait. The system does not require any calibration and is able to reduce the overall latency for an auditory step feedback.

Two characteristic gait events of healthy forward walking define a time difference, which is the basis for the prediction. The prediction algorithm is capable of achieving a prediction error that is below the human perception threshold. From the possible characteristic features of human gait, the zero crossings of the measured signals performed best for a reliable and robust approach. From the combination of different events - resulting from foot accelerations and angular velocity - one of the predictors with a good performance relies on the foot roll rate only and thus only requires one single-axis gyroscope per foot. However, for this predictor the prediction time is shorter than for the other ones, and thus it could be used only if shorter prediction times are feasible. For achieving longer prediction times, both the acceleration and the foot roll signal have to be used. Although these predictors are not so precise, their prediction is still within the tolerable limits. However, these predictors require two different input signals from an accelerometer and a gyroscope.

The design of the predictor was chosen in such a way that it matches the time for walking on a flat rigid surface. In this case, the real step sound can correctly be replaced by a virtual one. However, for other real surfaces like tall grass or snow, for which the real sound could occur earlier, the predictor needs to be adapted and retrained.

Future work should focus on detecting and predicting other steps than straight forward walking, such as walking backwards, stomping, sneaking, or turning on the spot. More parameters of the human gait could also be evaluated in order to use them for a physically-based synthetic sound generation. By adjusting the possible prediction time, the user acceptance regarding the auditory step feedback could be analyzed more in detail. Here, maximum acceptable time differences between real and synthetic sound should be investigated, including the effect of an early compared to a delayed auditory feedback.

Acknowledgements

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

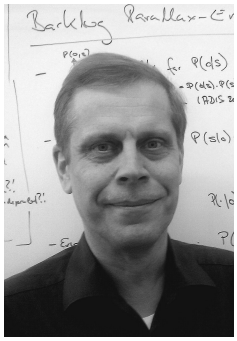
Author's Biographies

Markus Zank received his M.Sc. degree in Mechanical Engineering from ETH Zurich in Switzerland in 2013. He has been a student at the ICVR group (Innovation Center Virtual Reality) at ETH since 2012 where he is now a Ph.D. student. His research interests include real walking in virtual environments, human locomotion planning and human interaction with virtual environments.

Thomas Nescher is a researcher in the ICVR group (Innovation Center Virtual Reality) at ETH Zurich in Switzerland. He holds a M.Sc. degree in Computer Science with specialization in Visual Computing from ETH. Thomas is currently a Ph.D. candidate student at ETH, examining optimization strategies for navigation in immersive virtual environments. His research interests cover Human Computer Interaction fields, ranging from remote collaboration to virtual reality applications and navigation in virtual environments.

Andreas Kunz was born in 1961, and studied Electrical Engineering in Darmstadt/Germany. After his diploma in 1989, he worked in industry for 4 years. In 1995, he became a research engineer and Ph.D. student at ETH Zurich/Switzerland in the Department of Mechanical Engineering. In October 1998, he finished his Ph.D. and established the research field "Virtual Reality" at ETH and founded the research group ICVR (Innovation Center Virtual Reality). In 2004, and became a private docent at ETH. Since July 2006, he is Adjunct Professor at BTH. Since 1995, he has been involved in students' education and since 1998 he has been giving lectures in the field of Virtual Reality. He also gives lectures abroad, in Switzerland as well as in other countries such as Germany, Sweden, USA, Romania, etc. Dr. Kunz published in IEEE Virtual Reality and PRESENCE, and reviews papers for several IEEE conferences.

References



[1] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P. Brooks, Jr. Walking > walking-in-place > flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive*

techniques, SIGGRAPH '99, pages 359–364. ACM, 1999.

[2] Roy A. Ruddle and Simon Lessels. The benefits of using a walking interface to navigate virtual environments. *TOCHI '09: Transactions on Computer-Human Interaction*, 16(1):1–18, 2009.

[3] R. Nordahl, S. Serafin, N.C. Nilsson, and L. Turchet. Enhancing realism in virtual environments by simulating the audio-haptic sensation of walking on ground surfaces. In *Virtual Reality Workshops (VR), 2012 IEEE*, pages 73–74. IEEE, 2012.

[4] Yonghao Wang, Ryan Stables, and Joshua Reiss. Audio latency measurement for desktop operating systems with onboard soundcards. In *Audio Engineering Society Convention 128*. Audio Engineering Society, 2010.

[5] Bruno L Giordano, Yon Visell, Hsin-Yun Yao, Vincent Hayward, Jeremy R Cooperstock, and Stephen McAdams. Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditions. *The Journal of the Acoustical Society of America*, 131:4002, 2012.

[6] Stefania Serafin, Luca Turchet, Rolf Nordahl, Smilen Dimitrov, Amir Berrezag, and Vincent Hayward. Identification of virtual grounds using virtual reality haptic shoes and sound synthesis. In *Proceedings of Eurohaptics Symposium on Haptic and Audio-Visual Stimuli: Enhancing Experiences and Interaction*, pages 61–70, 2010.

[7] Federico Avanzini, Stefania Serafin, and Davide Rocchesso. Interactive simulation of rigid body interaction with friction-induced sound generation. *Speech and Audio Processing, IEEE Transactions on*, 13(5):1073–1081, 2005.

[8] Luca Turchet, Rolf Nordahl, Stefania Serafin, Amir Berrezag, Smilen Dimitrov, and Vincent Hayward. Audio-haptic physically-based simulation of walking on different grounds. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pages 269–273. IEEE, 2010.

[9] Ion PI Pappas, Milos R Popovic, Thierry Keller, Volker Dietz, and Manfred Morari. A reliable gait phase detection system. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 9(2):113–125, 2001.

[10] Rolf Nordahl, Luca Turchet, and Stefania Serafin. 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, 2011.

[11] Alvin W Law, Benjamin V Peck, Yon Visell, Paul G Kry, and Jeremy R Cooperstock. A multi-modal floorspace for experiencing material deformation underfoot in virtual reality. In *Haptic Audio visual Environments and Games, 2008. HAVE 2008. IEEE International Workshop on*, pages 126–131. IEEE, 2008.

[12] Fritz Menzer, Anna Brooks, Pär Halje, Christof Faller, Martin Vetterli, and Olaf Blanke. Feeling in control of your footsteps: Conscious gait monitoring and the auditory consequences of footsteps. *Cognitive Neuroscience*, 1(3):184–192, 2010.

[13] Rolf Nordahl. Self-induced footsteps sounds in virtual reality: Latency, recognition, quality and presence. *Presence*, pages 353–354, 2005.

[14] Valeria Occelli, Charles Spence, and Massimiliano Zampini. Audiotactile interactions in temporal perception. *Psychonomic bulletin & review*, 18(3):429–454, 2011.

[15] E. Foxlin. Pedestrian tracking with shoe-mounted inertial sensors. *Computer Graphics and Applications, IEEE*, 25(6):38–46, 2005.

[16] Ross Stirling, Jussi Collin, Ken Fyfe, and Gérard Lachapelle. An innovative shoe-mounted pedestrian navigation system. In *Proceedings of European Navigation Conference GNSS*, 2003.

[17] Antoon Th. M. Willemsen, Fedde Bloemhof, and Herman BK Boom. Automatic stance-swing phase detection from accelerometer data for peroneal nerve stimulation. *Biomedical Engineering, IEEE Transactions on*, 37(12):1201–1208, 1990.

[18] Christopher L Vaughan, Brian L Davis, and Jeremy C O'connor. *Dynamics of human gait*. Human Kinetics Publishers USA, 1992.

[19] J.D. Wendt, M.C. Whitton, and F.P. Brooks. Gud wip: Gait-understanding-driven walking-in-place. In *Virtual Reality Conference (VR), 2010 IEEE*, pages 51–58. IEEE, 2010.

[20] Christopher M Bishop et al. *Pattern recognition and machine learning*, volume 1. springer New York, 2006.

[21] V.T. Inman, H.J. Ralston, and F. Todd. *Human walking*. Williams & Wilkins, 1981.