Walker Speed Adaptation in Gait Synthesis

Joëlle Tilmanne and Thierry Dutoit

TCTS Lab / Numediart Institute, University of Mons (UMONS), Mons, Belgium
joelle.tilmanne@umons.ac.be

Abstract. This paper presents a Hidden Markov Model based walk synthesizer, where the velocities of the synthesized walks are interpolations of the walk paces present in the training database. In a first stage, Trajectory Hidden Semi Markov Models (THSMM) of gait are trained for different subjects and different velocities, using a database of motion capture walk sequences. In a second stage, the parameters of the individual speed models are interpolated in order to synthesize walks with speeds not present in the training database. A qualitative user evaluation of the synthesized sequences shows that the naturalness of motions is preserved after linear interpolation and that the evaluators are sensible to the interpolated speed of the motion.

Keywords: motion synthesis, walk, style, HMM.

1 Introduction

Human walk is a complex phenomenon even if it is something we do every day without even thinking of it. The parameters that modify the way we walk are numerous and their respective influence can hardly be differentiated. Furthermore, as natural experts of the human walk we are highly sensitive to any inconsistency in synthesized walks. Modeling and synthesizing gait that looks natural and humanlike is hence a challenging task for computer animation.

Motion capture (mocap) sequences are the most accurate numerical representations of human motion we can get. However, a mocap sequence is a frozen and non interactive representation of the motion: once it is recorded, it cannot be edited easily. If the user needs a longer walk sequence, another style or a different speed, the sequence cannot be used and a new one has to be recorded. However, recording a database with all the style options for all the motions is impossible, and recording new motions each time a new animation has to be produced is costly and often not materially possible. Our goal is to find ways of parameterizing the "stylistic" components of the motion independently from the functional part of the same motion, in order to give the user some kind of interactive control on the style of the output motion.

We adopt a model based approach for synthesizing natural looking walk with controllable pace preserving the walker's identity. We use statistical learning techniques, and more precisely Trajectory Hidden Semi Markov Models

M. Kallmann and K. Bekris (Eds.): MIG 2012, LNCS 7660, pp. 266–277, 2012.

(THSMMs) to automatically extract the underlying rules of human walk, directly from training on mocap data. In [1], we presented a continuous control of exaggerated walk styles from one single actor (sad, tiptoe, catwalk, manly, etc.). In the present work we consider speed and individuality as other aspects of "motion style". The control of these stylistic characteristics is built based on sequences from different subjects walking at different speeds. In addition to the synthesis of realistic walks with controlled speed, the difference in the speed expression across subjects can be analyzed based on the trained models.

2 Related work

In order to produce motion sequences presenting a new style, two mocap based approaches can be taken. One consists in editing existing motions and the second one uses mocap as training data and builds models that can be used to synthesize new stylistic walks by changing the parameters of the original model. This second approach is the one we have taken in this work. Statistical learning techniques can learn without any prior knowledge, to model the style component independently from the functional part of the motion. The stylistic variations are then embedded in the model parameters. By modifying these parameters, the user can control the style of the synthesized motion.

Several studies [2,4,3] have used PCA not only for reducing the dimensionality of the data, but to separate the influence of style from the motion. In a similar way, ICA has also been used [5] to differentiate style and functional motion. In [6], a multilinear motion analysis separates style and individuality variations and in [7], a linear time-invariant procedure learns the translation between two styles from motions with the same content.

A common approach consists in using Hidden Markov Models (HMMs) to synthesize motion, and to integrate a style variable into the model. One of the advantages of using HMMs is that they are well suited to time series modeling. There is no need to use time warping procedures as the time variability of the motion is integrated directly in the HMM structure. Wang et al. [8] present a parametric HMM incorporating a "style" parameter in the probability density functions. In their "Style Machine", Brand and Hertzmann [9] include a style variable which is automatically extracted during training and which can be chosen before the synthesis process. However their style variable is not explicit and changes some intrinsic style-related parameters, which can make it hard to use as a style controller. Yamazaki et al. [10] model walk using a Hidden Semi-Markov Model (HSMM) approach similar to the one we use. Their model takes into account speed and stride length as a "style" variation using multiple regression. However, this method can only be used to model quantitative variations, and is thus not suited to model emotions or expressivity that can hardly be described by numerical values. In their approach, the whole training has to be done again if one wants to add a new style in the model. This is not the case in our work, where a new style can be added by a single adaptive training using the new style data, without retraining the previously existing styles.

3 Training Data

The training data comes from the *eNTERFACE'08 3D walk* mocap database [11] that we recorded previously, and which contains walk sequences for 41 subjects at different speeds. Each performer was given four different instructions, and each one had to be repeated three times. The instructions were: "Walk (normally)", "You are in a park, the weather is nice, you take your time", "You are going to work, but you are not late", "You have to catch a train, you are very late but you are not allowed to run". Once the instruction was given, the subject was free to choose his speed (no treadmill was used). The absolute position of the root is not taken into account in the modeling, as it can be recalculated afterwards. The rotations of the body represented by an 18-joint skeleton were captured, giving 54 values per frame to describe the motion. The original Euler angle data was converted into exponential map which is locally linear and where singularities can be avoided [13].

Motion speed is one of the major parameters that influence motion. Walking slower or faster implies more than just modifying the frame rate of a "normal" walk. Fig. 1 illustrates this difference by comparing the steps of normal, slow and fast walks for one subject. All the steps displayed are linearly resampled to the same number of frames. It can be noticed that the evolution of the angles values is clearly different for the three walk speeds. Playing the same angle sequence faster or slower is therefore not sufficient for accurate walking speed synthesis. When they walk, humans change their posture according to their speed, and the duration of each gait phase is adapted in a non-linear manner.

4 Modeling and Synthesis Method

The stylistic modeling problems encountered when studying motions are similar to the ones encountered for years in speech modeling. In this work, we take advantage of procedures developed for speaker adaptation in speech synthesis [18] and adapt them to our stylistic motion problem. The training and synthesis algorithms are based on functions originally implemented for speech within the "HMM-based Speech Synthesis System" (HTS) framework, publicly available on the HTS website [14]. The adaptation of this HMM-based procedure to the motion problem is presented in more details in [15]. The dynamical aspect of the data is taken into account by integrating the first and second derivatives of our parameters (Trajectory HMM (THMM) [16]) into the model. By adding these derivatives to our 54 original parameters, we obtain a 162 dimensional vector of observations to model. The time spent in each state of the HMM is explicitly modeled in duration probability density functions thanks to Hidden Semi-Markov Models (HSMM) [17].

The first stage of the procedure consists in training an average "reference" model. Both steps (left or right) are modeled by separate left-right five-states THSMMs.

In the second stage of the training, the reference model is adapted to each specific style, using the walk sequences from one single style at a time.

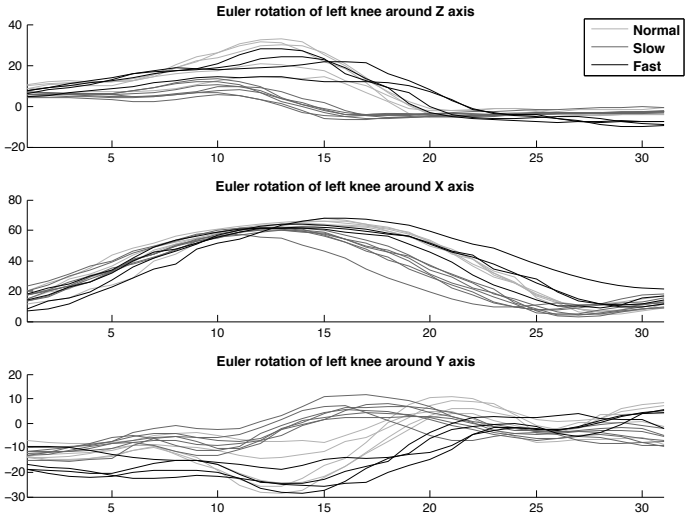

Fig. 1. Left knee angles, linearly resampled to 31 frames (original number of frames of first normal step), for the left leg steps of one complete walk sequence of normal (blue), slow (green) and fast (red) walks of subject number one from the database. The values of the Euler angles are displayed separately for the Z , X and Y axis (three subplots).

Using an adaptive training procedure [18,15] enables all of our models to be aligned without any time warping, and to correspond to the same internal structure, which will be needed for the interpolation process. Using these models, new sequences of walk can be synthesized for each of our individual walk styles. The synthesis consists in first concatenating the desired number of left and right step to form a complete walk sequence model. The parameter sequence is then calculated by finding the amximum likelihood parameter sequence corresponding to this complete model. The continuity of the synthesized sequence is ensured as the dynamics of the original data is respected thanks to the trajectory HMM structure. The synthesized sequence consists in the 54 angles values that are modeled. The absolute displacement of the skeleton can be calculated in a post processing step. The step boundaries are known, and calculating the height of each foot thanks to the fixed skeleton size, we determine which foot is in contact with the ground. The position of the whole body can then be calculated using forward kinematics until the other foot becomes the reference, and so on.

5 Continuous Control of Walk Speed

So far, each speed is modeled separately, and in the synthesis step, the user's control on the output sequence is only the choice of one among the individual speed models. We want to give the user the ability to have more control on the output sequence and to synthesize speeds that were not present in the original

training data. In the model training stage presented in Section 4, each walk speed is modeled by six five-states HSMMs, and each HSMM contains both state duration and 162-dimensional observation modeling. For each state of each HSMM, duration is modeled by one Gaussian (mean and variance) and observations are modeled by single Gaussians (multidimensional Gaussian with diagonal covariance matrix). One whole speed model is hence represented by 4890 Gaussian probability density functions (pdfs).

These model parameters are then modified in order to control the speed of the synthesized walks and to obtain speeds not present in the original training data. The new models are obtained by linear combinations of the probability density functions of each speed-specific model. The 4890-dimensional model parameter space is hence considered as a continuous space in which new speeds can be produced.The known speeds are used as landmarks in the continuous model parameter space, an no abstract control parameter has to be involved.

6 Results

As presented in Section 4, the first part of the procedure consists in training models for each particular speed. Models of slow, normal and fast walks were calculated for each subject of the database. First an average walk model was calculated using as training data the three normal speed walk sequences from all the 41 subjects of the database, and second an adaptive training was performed to obtain three different models for slow, normal and fast walk of each single subject. Applying the adaptation procedure for each of the 41 subjects and each of the three speeds of the database, we obtain 123 speed and individuality specific THSMMs. Fig. 2 compares three channels of the original data of normal, slow and fast walks of subject number one to examples of walk steps synthesized with the corresponding normal, slow and fast models trained on the same original data. It can be noticed that the three models are properly adapted to the different speeds they represent, both in duration and in amplitude.

Since we have slow, normal and fast data for 41 different subjects, we can also study how the personal variations can influence motion changes linked to speed. Thanks to the shared structure of our models and their common training basis, their parameters are aligned and we can easily compare the stylistic transforms. These transforms can hence be used not only for interpolation but also for comparison across styles or across subjects. We could observe that the subject individualities were expressed more strongly in the slow walk than in the other walks, and that even if there is a strong correlation among the subjects for each speed, some people display a stronger personal style than the others. These observations support the choice of modeling the speed transform individually for each subject.

Our approach to continuously control the synthesized walk speed consists in linearly interpolating the set of parameters between two speeds from the same subject:

$$pdf_{new} = pdf_{speed1} + interp * (pdf_{speed2} - pdf_{speed1}). \quad (1)$$

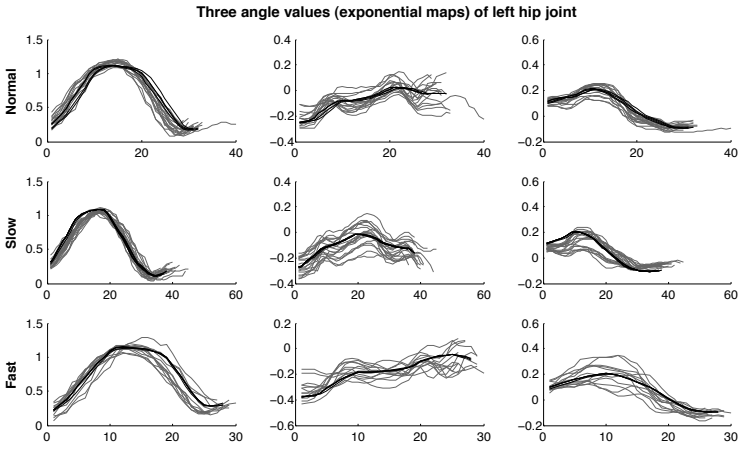

Fig. 2. Rotation (exponential maps) of the left hip joint during left leg steps for normal, slow and fast walks. Time is displayed in number of frames. The original steps are displayed in grey and examples of synthesized steps in black.

where *pdf* is the vector of the probability density function means (i.e. the parameters of the model), and *interp* is a scalar value that gives the interpolation ratio between *speed* 1 and *speed* 2. If this approximation was perfect, the normal walk could be obtained as a combination of slow and fast walk. However, as illustrated in Fig. 1, normal walk is not a linear interpolation between slow and fast walks. Our three walk speed models can be considered as three landmarks that approximate the evolution of walk according to speed for one particular subject. Linear interpolation between these three landmarks produced natural looking walk sequences, as was assessed in the user evaluation.

In addition to interpolation between two existing walk speeds, we can also create exaggerated variations of the speeds present in our original database. By giving the *interp* factor values higher than one, and taking the normal walk model as a reference, we can easily build models of exaggerated fast or slow walks. Depending on the individuality of the subjects, these exaggerations remained natural for different ranges of values of *interp*: higher values if the two interpolated speeds were close and lower values if they were very different from each other. The quality of the synthesized walk depended on the original styles as impossible values of angles appeared when the difference were exaggerated too much (knees bending backwards or awkward bending of the spine for instance).

A third interesting interpolation result is obtained by giving to *interp* values smaller than zero. The difference between the controlled style (speed2) and the chosen reference speed walk (speed 1) model is substracted to that reference walk instead of being added. The new style obtained that way presents characteristics at the opposite of the controlled style. For instance the fast walk which is faster than the neutral model and where the pose of the character tends to bend inwards with respect to the average posture will give an opposite model where

the character walks slower and looks much more "open" by its posture. We are hence able to synthesize styles that do not appear in our database but that show style characteristics that are opposite to the ones of recorded styles. Fig. 3 presents some examples of synthesized slow and fast original speeds along with interpolated variations. The figure shows that both poses (for instance the hand balancing) and durations (number of frames for the same step) are affected by the interpolation process.

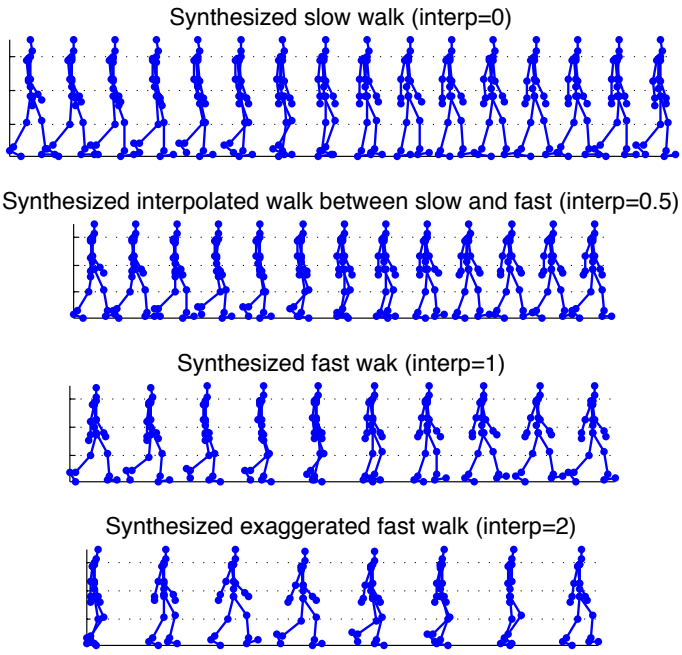

Fig. 3. Synthesized left step for original and interpolated variation of slow and fast speeds of subject 4. The figures display in order the original slow speed, a speed between slow and fast (interp=0.5), the original fast speed and an exaggerated fast speed (interp=2). The stick figures are displayed every three frames.

Following the same interpolation and extrapolation procedures, we can also control the style intensities between subjects, by exaggerating for instance the stylistic difference between two persons, or by synthesizing a new style which is a mix between several subjects.

7 User Evaluation

The quality of the synthesis was evaluated through a user study. The evaluation was done based on walks from 4 subjects among the 41 subjects of the eNTERFACE'08 database so as to reduce the number of videos to be evaluated. As it

has been said in the introduction, humans are natural experts of human motions and are highly sensitive to any discrepancy in an animation. They are therefore the best judges of the naturalness of the motion. Our evaluation consisted in three tests, evaluating respectively the quality of the synthesis compared to original training data (Section 7.1), the perception of the interpolation between two speeds (Section 7.2), and the motion naturalness perceived (Section 7.3).

Participants accessed to the evaluation tests through a web browser. Once their answer was selected and saved, they could not come back to previous videos. If they did not complete the test thoroughly, they could come back later. They had to start the videos themselves by clicking on it, and could watch them as many times as they wanted. In the video sequences, motion was performed by a blue stick-figure character as shown in Fig. 3.

Twenty-four naive evaluators took part in the evaluation, 10 females and 14 males, from 24 to 66 years old, with an average and standard deviation of 36 and 13 years, respectively. The videos were randomly picked by the evaluation program, leading to different evaluation sets for each evaluator. The results of each test were taken into account only if the user completed all the evaluations of the given test. The final number of evaluators taken into account is not the same for all test as some users dropped the evaluation procedure before completing the three tests.

7.1 Style Modeling Evaluation

In this first test, the evaluator was presented two videos at the same time, one displaying the original walk and the other the synthesized walk sequence corresponding to the same subject (1 to 4) and the same speed (normal, slow or fast). The original and synthesized sequences did not contain the same number of steps. The evaluation set consisted in 12 video pairs and 110 comparison tests were performed (5 by each evaluator). No interpolation was performed in this test that aimed at assessing the accuracy of the models compared to the original data they were trained on. The order of the two videos was randomly determined by the program. Evaluators were asked to choose between five possible qualifications ranging from “identical” to “nothing in common”. The test validated the synthesis based on models trained on mocap data. The evaluators saw that the walk sequences were not the same, but they did find that they were not very different.

7.2 Interpolation Factor Recognition

In this second test, the evaluator was presented two videos at the same time and the test was completed after ten evaluations. The first video presented two walks in a row, with the second walk faster than the first one. Those two walks were synthesized from the original speed models. The second video displayed one single walk sequence that corresponded to an interpolation of the walk speeds presented in the first videos. The interpolation factor present in the evaluation were -1, -0.5, 0, 0.5, 1, 1.5 and 2. The evaluator was asked to position the speed

of the third walk compared to the speeds of the first two walks, thanks to a continuous slider that ranged from -1.5 to 2.5 (these values were not apparent on the slider, where only the position of the first walk (interp =0) and of the second walk (interp=1) were displayed). Two hundred and forty evaluations were performed on the set of 84 evaluation video pairs. This test investigates if the evaluators are sensitive to the interpolation factor and how that interpolation is perceived. Fig. 4 presents the results of the interpolation factor recognition. We observe that the evaluators were good at recognizing the interpolation factors between the existing styles (interpolation values of 0, 0.5 and 1). For extrapolations, even if the trend is respected and the evaluators are able to differentiate the extrapolation values (-1 seems lower than -0.5 and 2 seems higher than 1.5), the extend of the extrapolation is underestimated. The perception of the interpolation by the evaluator was hence linear inside the actual speeds boundaries but became non-linear outside. We should take that factor into account for the synthesis step: when the user wants to synthesize a walk with an interpolation factor of 1.5, he should take an interpolation factor of 2 rather then 1.5 if he wants the viewer to perceive the interpolation the way he imagined it.

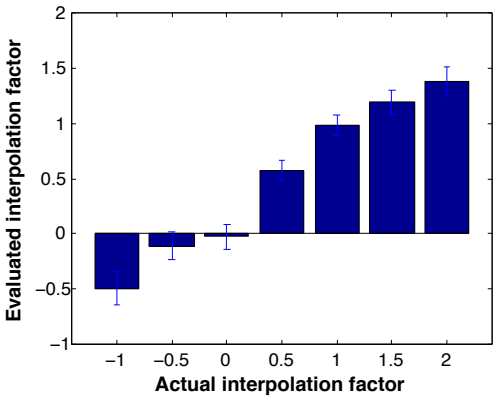

Fig. 4. Actual interpolation factor compared to the evaluated interpolation factor. The error bars describe the .95 confidence interval.

7.3 Naturalness Evaluation

In this last test, only one walk sequence was displayed. The test was completed after ten evaluations and 240 evaluations were performed on a set of 96 walk sequences. The set of the evaluated walks contained original sequences from the database, synthesized sequences corresponding to the original speeds of the database and synthesized interpolated sequences. For each video, the evaluator was asked to assess the naturalness of the walk sequence, on an open continuous scale ranging from “synthetic” to “real”. The numerical values associated with the subjective descriptors range from zero for “synthetic” to one for “real”.

The evaluators attributed a mean naturalness score of 0.72 to the original database sequences and a mean score of 0.55 for the whole set of synthesized sequences. These scores are not evenly distributed across the four database subjects nor across the different interpolation factors. Fig. 5 illustrates the variability of the perceived naturalness according to the four subjects for original database sequences, synthesized sequences corresponding to the three original speeds and synthesized sequences corresponding to any interpolation factor. It can be observed that major differences are perceived between the subjects, even for the original database walk sequences. In particular, the naturalness score of subject 3 is much lower than the others while the score is slightly higher for subject 4. Subject 3 has hence walked in a way that did not seem very natural to the evaluators, even in the original recordings. Original sequences from the database and synthesized sequences corresponding to original speeds are perceived as equally natural looking. For subjects 2 and 3, synthesized sequences were even better perceived that the original database ones. The interpolated walk sequences follow the trend across the subjects, with lower values for subject 3 and higher ones for subject 4. However they always appear slightly less natural to the evaluator. This is especially due to the extrapolation factors (interp = -1 or 2) as illustrated in Fig. 6.

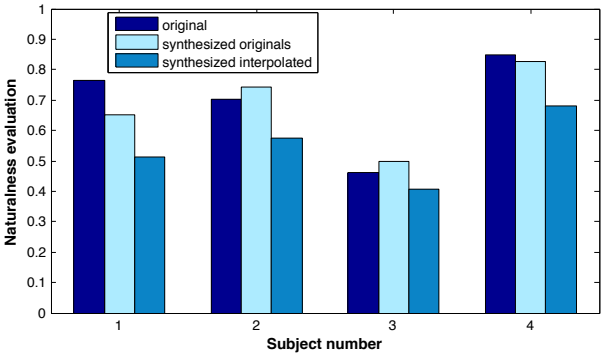

Fig. 5. Naturalness evaluation of original database sequences, synthesized sequences of original speed and synthesized sequences of interpolated speeds, for the four database subjects modeled and evaluated in these tests

Fig. 6 illustrates the naturalness score obtained for the synthesized sequences, according to the interpolation factor. We observe that the higher naturalness corresponds to the original speeds (interp = 0 or 1). However, the evaluators still perceive the interpolation values close to the original styles as more natural than synthetic, as the naturalness score remains higher than 0.5 for the -0.5, 0.5 and 1.5 and values of the interpolation factor. For extrapolations farther from the original speeds, motion artifacts appear under the form of physically impossible motions and the walk sequence lose its naturalness.

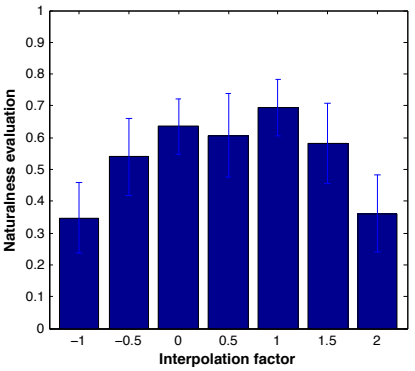

Fig. 6. Evaluation score of the naturalness corresponding to the seven interpolation factors. The error bars describe the .95 confidence interval.

8 Conclusion

In this work, a set of three different walk speed models based on Trajectory Hidden Semi Markov Models was trained for 41 subjects and used as a basis for speed interpolation and extrapolation. Through this approach, the user is given continuous control on the speed of the synthesized motion while preserving its naturalness, and is hence able to synthesize walk speeds not present in the original database. The speed values could be controlled between the original speeds present in the database (values of *interp* between zero and one), but also exaggerated beyond these values (values of *interp* greater than one), or at the opposite of the controlled speed (values of *interp* lower than zero). Our model also enables us to control the individual subject-related styles applying the same interpolation and extrapolation approach on the different subject models . Some examples of walk sequences synthesized with our method can be found on the author's website (www.tcts.fpms.ac.be/˜tilmanne). Qualitative user evaluation assessed that the THSMM approach models accurately the different speeds of the training data, that the trend of the interpolation factor was perceived by the evaluators, and that the naturalness of the motion was preserved for speeds close to the original speeds.

Future works should study the limits that could be applied to the different joints of the skeleton, so that the user is allowed to extrapolate the motion speeds without breaking the laws of biomechanics. Part of the naturalness loss that evaluators could experience when comparing the synthesized and original sequences is due to the repetitive pattern that the synthesized sequences display, as the successive steps present small variations compared to the higher variability of a natural sequence of steps. We intend to take advantage of the statistical modeling of our model in future studies in order to add more variability to the synthesized sequences, for motions that will look more alive. We will also study how the speed characteristics could be combined to other style transformations in one single walk model.

Acknowledgment. This project was partly funded by the Ministry of Région Wallonne under the Numediart research program (grant N0716631). We gratefully acknowledge A. Moinet and J. Urbain for their help in designing the online evaluation procedure

References

1. Tilmanne, J., Dutoit, T.: Continuous control of style through linear interpolation in HMM based stylistic walk synthesis. In: Proc. CW 2011, pp. 232–236 (2011)
2. Troje, N.F.: Retrieving information from human movement patterns. In: Understanding Events: How Humans See, Represent, and Act on Events, pp. 308–334 (2008)
3. Tilmanne, J., Dutoit, T.: Expressive Gait Synthesis Using PCA and Gaussian Modeling. In: Boulic, R., Chrysanthou, Y., Komura, T. (eds.) MIG 2010. LNCS, vol. 6459, pp. 363–374. Springer, Heidelberg (2010)
4. Uratasun, R., Glardon, P., Boulic, R., Thalmann, D., Fua, P.: Style-based Motion Synthesis. Computer Graphics Forum 23(4), 799–812 (2004)
5. Shapiro, A., Cao, Y., Faloutsos, P.: Style components. In: Proc. of Graphics Interface, pp. 33–39 (2006)
6. Min, J., Liu, H., Chai, J.: Synthesis and editing of personalized stylistic human motion. In: Proceedings of SI3D, pp. 39–46 (2010)
7. Hsu, E., Pulli, K., Popovic, J.: Style Translation for Human Motion. In: Proc. SIGGRAPH 2005, pp. 1082–1089 (2005)
8. Wang, Y., Xie, L., Liu, Z., Zhou, L.: The SOMN HMM model and its application to automatic synthesis of 3d character animation. In: IEEE Conference on Systems, Man, and Cybernetics, pp. 4948–4952 (2006)
9. Brand, M., Hertzmann, A.: Style machines. In: Proc. SIGGRAPH 2000, pp. 183–192 (2000)
10. Yamazaki, T., Niwase, N., Yamagishi, J., Kobayashi, T.: Human Walking Motion Synthesis Based on Multiple Regression Hidden Semi-Markov Model. In: Proc. Conference on Cyberworlds, CW 2005, pp. 445–452 (2005)
11. Tilmanne, J., Sebbe, R., Dutoit, T.: A Database for Stylistic Human Gait Modeling and Synthesis. In: Proceedings of the eNTERFACE 2008 Workshop on Multimodal Interfaces, Paris, France, pp. 91–94 (August 2008)
12. IGS-190. Animazoo website, `http://www.animazoo.com`
13. Grassia, F.S.: Practical parameterization of rotations using the exponential map. Journal of Graphics Tools 3, 29–48 (1998)
14. HTS working group: The HMM-based speech synthesis system (HTS) Version 2.1 (2010), `http://hts.sp.nitech.ac.jp/` (accessed)
15. Tilmanne, J., Moinet, A., Dutoit, T.: Hidden Markov Model based stylistic gait synthesis. Eurasip Journal on Advances in Signal Processing (2012)
16. Tokuda, K., Yoshimura, T., Masuko, T., Kobayashi, T., Kitamura, T.: Speech parameter generation algorithms for hmm-based speech synthesis. In: Proc. of ICASSP, pp. 1315–1318 (June 2000)
17. Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T.: Duration modeling for HMM based speech synthesis. In: Proc. of ICSLP, pp. 29–32 (1998)
18. Yamagishi, J., Kobayashi, T., Nakano, Y., Ogata, K., Isogai, J.: Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm. IEEE TASLP 17(1), 66–83 (2009)