Dynamic dance warping: Using dynamic time warping to compare dance movement performed under different conditions

Sam Ferguson
Creativity & Cognition Studios
School of Software
Faculty of Engineering & IT
University of Technology, Sydney
Samuel.Ferguson@uts.edu.au

Emery Schubert
Empirical Musicology Laboratory
School of Arts and Letters
Faculty of Arts, UNSW
e.schubert@unsw.edu.au

Catherine J. Stevens
MARCS Institute and
School of Psychology
University of Western Sydney
kj.stevens@uws.edu.au

ABSTRACT

Dynamic time warping (DTW) is proposed as a technique to assess the difference between two dance performances in terms of timing and to provide further insight into dancer cognition. The DTW method is validated for use with dance performance motion tracking data by comparing its results with 'ground truth' results obtained from a comparison between videos of two motion tracked performances. The technique was extended to investigate two hypothesised processes that affect movement timing-scaling (a fixed ratio alteration) and lapsing (caused by insertion or deletion of movement material). As an example of the use of the technique, an ensemble contemporary dance work was performed with the motion of one of three dancers captured in two conditions - with no music (NM) and with music (WM) - with one repeat of the two conditions. The application of the DTW-based algorithm demonstrates that lapses explained much of the timing mismatch (9.6 out of 14 seconds), with a small proportion explained by scaling (a ratio of 0.976) consistent with previous research. However, after again performing the dance under NM and WM conditions the DTW technique demonstrated a non-trivial contribution of scaling in explaining time differences across the various combinations of conditions. In these comparisons, scaling cannot be eliminated as a possible underlying factor of timing error, and it may be that correct scaling (aiming for a ratio of 1) must be learned via practice.

Categories and Subject Descriptors

J.5 [**Arts And Humanities**]: Performing arts (e.g., dance, music); I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis-Motion

Keywords

dance, dynamic time warping, lapsing, scaling

MOCO'14, June 16-17 2014, Paris, France.

ACM 978-1-4503-2814-2/14/06 ...$15.00.
http://dx.doi.org/10.1145/2617995.2618012.

1. INTRODUCTION

Analysis of dance movement has resisted systematic scientific investigation and quantification until recently. The last decade has seen growing interest in understanding the mental processes involved in dance creation, performance and perception [1]. One of the reasons for the growth of this field of research is because it provides a new lens on temporal cognition and non-verbal communication. The endeavour has been supported by the development of sophisticated motion capture devices, essential tools for many of the pertinent research questions, and through the development of software libraries to analyse this new stream of data [2].

Automation of lapsing and scaling detection has broad application to investigation of time-series data from music and dance performance and perception [3]. The technique is applicable to interpersonal coordination and entrainment [4] at a global or whole-body level. One example is for a student's musical or dance performance to be compared with an expert's performance, to assess the degree to which the student synchronises with the teacher along the time dimension. A dance could also be compared in two conditions providing insights into cognition about how the dancer changes or adapts in the test condition relative to the other condition. The latter comparative approach is the one adopted here.

This paper builds on and extended a recent study [5] which investigated a small dance ensemble in 'with music' and 'no music' dance conditions. In that paper it was asserted that there were two cognitive time related processes, lapsing and scaling, that explained differences between two performances under different conditions (with music [WM] versus no music [NM] conditions). If the change in dancer time adjustment is due to an 'internal clock' that is not in perfect time alignment across conditions, then any change from the ideal would be seen as a drift in time of one performance with respect to the other- a 'scaling' of large portions of the sequence; alternatively, if time difference is due to a memory 'lapse' then specific sections would be inserted or omitted. For the dance movement sequence investigated in that paper, both scaling and lapsing mechanisms were identified, with lapses dominating. This finding was consistent with other areas of research such as performance of memorised music, where there is little evidence of scaling (that is, scaling tends to remain close to 1) [6, 7].

Nevertheless this earlier study had some possible limitations. First, only one performance in each condition was examined. The present paper reports new results wherein the two conditions, NM and WM, are repeated. Further, in the earlier study we explored a time domain technique for identifying when movements were different across conditions. It used cross-correlation functions to estimate the similarity between appropriately chosen windows of motion data from each condition. To test whether scaling or lapsing was occurring, the dependent motion data were resampled (for scaling) or offset (for lapsing) and the magnitude and offset of the correlation peak was used to diagnose whether these processes yielded an improvement in timing.

We propose the method of Dynamic Time Warping as an alternative approach for the problem of comparing similar time series datasets. However, several adaptations to the technique were required before they could be appropriately applied. First, we introduce dynamic time warping, describe how it has been adapted to scrutinise the hypothesised processes (lapsing and scaling), and then apply it to an example of dancer motion captured under two presentations of two different conditions. The goal is to develop methods for comparing pairs of identical choreographic intentions performed under different environmental (in this case, auditory) conditions. The subsequent adjusted data are then used to interpret the kinds of memory and time-keeping processes that underlie the dance actions and performance.

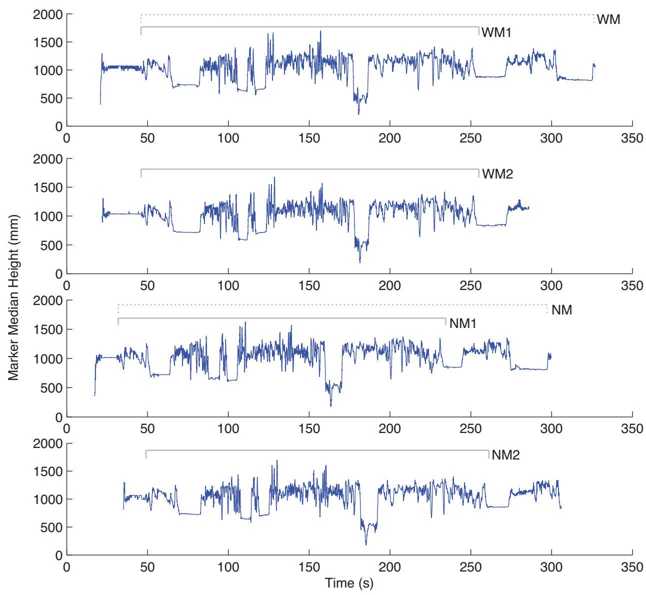

Figure 1: Time series of median $z-axis$ (height) of markers for each dance condition. Selected portions of the performances for each condition are shown with a bracket on top of each series. WM and NM (dotted) are used for validating the time adjustment method, while WM1, WM2, NM1 and NM2 (solid) are used for investigating the dancer's behaviour.

2. METHOD

2.1 Materials

The motion data were captured using a motion capture system, consisting of 10 Vicon cameras: 4 x MX40 cameras positioned 15 feet above the performance area and 6 x MX3 cameras positioned on 8 foot high tripods around the work performed live to an audience of 40 people. This Vicon system had a temporal resolution of 100 Hz. Using a digital video camera (Sony HandyCam HCR-30E) a video of each performance was made from the front right of the audience area. One of the three dancers (also a choreographer of the piece) wore a black lycra suit onto which 24 reflective markers were sewn. The positions of the left and right markers chosen were the ear (2), shoulder blade (4), top of shoulder (2), top of femur (2), elbow (2), wrist (2), hip (2), knee (2), ankle (2), and foot (2); the collar bone and base of the sternum served as two reference points. The position of each of the markers was recorded in 3 dimensions and the video recording was synchronised with the motion capture recording, by placing a timecode display in view.

2.2 Procedure

The dance performed was *Reactional Movement*, choreographed by Emma Batchelor and James Batchelor, which had the duration 4 minutes and 40 seconds. The accompanying musical piece was *Mysta-Lilli Pilli Drive* by Fourplay (from the album Digital Manipulation). The dance was performed in front of an audience, and was undertaken four times in two different conditions with one repeat of each condition. The two conditions consisted of a performance with the audio soundtrack playing from loudspeakers, and a performance where this soundtrack was omitted and the dancers performed without an audible soundtrack (although there was still sound produced during the dance, primarily from the dancers' foot movements and their breathing). These two conditions were repeated immediately for comparison purposes. These conditions are called With Music 1 and 2 (WM1 & WM2) and No Music 1 and 2 (NM1 & NM2) and were performed in the order NM1, WM1, NM2, WM2. There were some data points missing from the ends of the performances in conditions WM2 and NM2, and therefore all the conditions were shortened so as to consider the same portion of the dance for each condition. The start point of the data selections was selected by aligning the data sequences and checking the times against a simultaneously recorded video. For the initial investigation of the data analysis technique (Section 2) the entire performance is used in conditions With Music (WM) and No Music (NM). The term 'with music' only refers to the playback of a musical recording during the dance performance, and does not refer to any defined method of choreography to music, and similarly 'no music' is the absence of this playback. All six combination of conditions NM1, NM2, WM1 and WM2 are then compared against each other to test the effect of the musical soundtrack.

We selected one axis (the vertical, z-axis) and took the median of all markers at each sample, thus creating a single univariate time-series of movement. The z-axis was chosen intuitively, based on the dancer's predominant jumping and bending actions, and the median is chosen as it is generally more robust than the mean [8], as errors in one outlying marker are unlikely to cause a large change in the value of the median, but may make more notable changes to the mean. Figure 1 shows the entire median z-axis (height) time-series data and the boundaries used to delineate each of the conditions under investigation.

2.3 Dynamic Time Warping

Dynamic Time Warping (DTW) is a process used to align sequences and compare their similarity [9]. There are three stages to the process: The signals are windowed and the Short Term Fourier Transform (STFT) is applied, (window size of 32 samples; sampling rate of 100 Hz; step size between windows of 4 samples); A similarity matrix is then calculated (see Figure 2), which compares the two signals in the frequency domain. The size of the matrix is determined by the number of frames in the STFT representations of both signals; Finally, dynamic programming [10] is used to chart a path through the matrix that has the maximum similarity (Figure 2), within the constraints that ensure that the path does not turn back on itself: This process ensures that no repeating paths are found. This stage results in a set of Cartesian matrix coordinates that represent

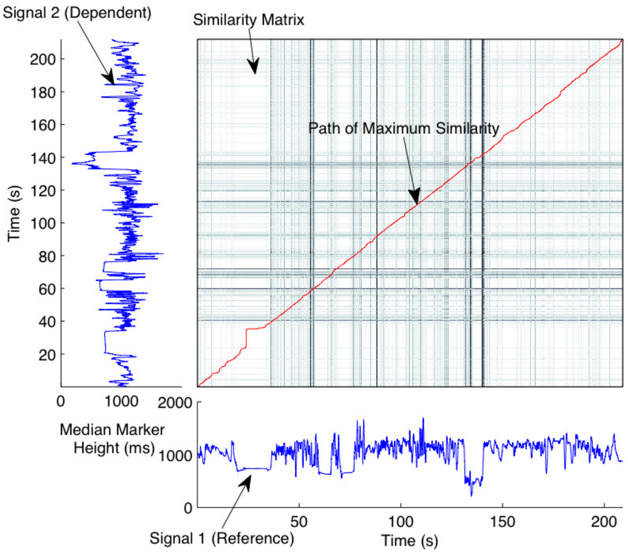

Figure 2: Finding the path of maximum similarity.

the portion of each signal that are maximally similar. In this work we use the implementation programmed by Turetsky and Ellis [11].

The path-finding algorithm is altered slightly for our particular purpose. In typical DTW, as described above, the dependent sequence is matched to a reference sequence by deleting or repeating sections of the dependent sequence. However, in our implementation we are not interested in matching alone, but in a two-way comparison. Thus we employ an extra phase in the algorithm: we collect successive values, or values that are separated by a single sample, together into a single value. This means that when the dependent signal is stalled, waiting for the reference to 'catch up', we represent this with a single number representing how long the dependent signal was stalled. This contrasts to the usual path-finding approach, which requires matching parts of each signal for each step forward through the reference signal. This requirement is unnecessary here as we are only interested in the relationship between the two signals in terms of time.

Once the DTW is processed the match is transformed to a time adjustment profile graph (Figure 3), describing the relationship of the dependent signal to the reference signal in terms of time. The coordinates described by the path of maximum similarity correspond to a time value in both of the signals (see Figure 2), and the 'time adjustment profile' is the difference between these time points as a function of the time axis of the reference signal. Where the path of maximum similarity corresponds to identical time values in both signals the time adjustment profile will read zero across the entire time axis.

2.4 Lapses and Scaling

To identify possible cognitive mechanisms contributing to non-veridical performances across dance conditions we quantify the different components of the time adjustment data that are gained through DTW. The aim is to separate long-term scaling from short-term lapses. A simple heuristic to separates the lapsing and scaling components is to examine the rate of change of the DTW adjustments required to optimally match the two series. A high rate of change indicates lapsing, and conversely a gradual rate of change indicates scaling. Therefore, a gradient threshold of the rate of change for identifying a lapse (gradient above threshold) or scaling (gradient below threshold) is used. We used a step of 0.75 s/40

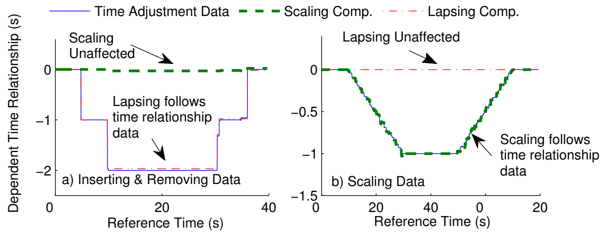

Figure 3: Graphs showing the effect upon synchronisation of a) inserting and subtracting portions of data before comparison and b) stretching or contracting (through resampling) portions of the signal before comparison.

(a) A condition with a poor match between the two videos.

(b) A condition with a good match between the two videos.

Figure 4: Manual video-based time adjustment data capture.

ms window based on an inspection of the distribution of difference values using random sequence validation (the 40 ms value is used due to the 25 Hz sampling rate). This is described in the following sections, followed by a manual time adjustment validation exercise.

2.5 Manual Video Time Adjustment Data Capture

To check that the method functioned adequately we devised a manual time adjustment output. Videos of the dancers in the two conditions WM and NM, recorded at the same time as the motion capture data, were superimposed on each other in a computer program (see Figure 4). The first author stepped frame by frame through the superimposed videos, choosing a number of frames to hold or progress the dependent video, so as to maintain the synchronisation between the reference video and the dependent (controlled) video. These frame step numbers were recorded as a measure of manual time adjustment. At a frame-rate of 25 fps (a time increment of 40 ms), it was possible to align each of the movements accurately - even leaping movements were sampled at many points.

Theoretically, the manual time adjustment data are subject to errors resulting from the sampling frequency of the video (25 fps) meaning that any time adjustment measurement can only be accurate to within $+/-$ 20 ms, half the time step of each video frame. The accuracy is further affected by the single camera angle recorded, as at certain times in the performance the view of the dancer wearing the markers was occluded by other dancers (although this was true only for a very small amount of the time recorded < 0.5%), and there were also a instances where there were disparities between the movements danced (usually of the nature of an omitted or added step or turn too many).

After transforming this data into the manual time adjustment, we tested the DTW algorithm's output time adjustment profile, and its ability to accurately match the manual time adjustment data, based on various parameters to the algorithm. A first parameter tested was whether the data should be differenced (attempting to remove serial correlation) or logarithm transformed. The best results (lowest error) were obtained for differencing only, over raw data, log-transformed only or differenced and log-transformed. The window length was also tested, and with our dataset a 32 sample window (eg. 320 ms at $fs = 100$ Hz) was found to have the lowest error (against 8, 16 or 64 samples) when differencing was applied as per above.

3. RESULTS

Based on the preliminary investigations, we have shown DTW can measure time comparisons between two dance motion sequences, and with empirically determined, optimised parameters would be able to identify lapsing and scaling of one performance signal (dependent) compared against another (reference). This process is now used to compare each of the 2 experimental conditions and 2 repeats (WM1, WM2, NM1 and NM2) with each other, resulting in six comparisons.

3.1 Time Adjustment Profiles

The time adjustment profile comparisons between each of the conditions are presented in Figure 5 and the accumulated time adjustment totals are tabulated in Table 1. Generally, there seems both an effect of lapses and also an overall trend of scaling. In the following section we will discuss expansion/contraction values as a percentage of the total time in the reference motion data sequence - an expansion value of 1% means that the dependent sequence is 1% longer than the reference sequence. As expected, the comparison between the two 'With Music' conditions (WM1/WM2) performed with musical accompaniment show close synchronisation (0.8% expansion) throughout the performance, finishing closely matched. This result is expected because both WM conditions use the same soundtrack, which will therefore provide the dancers with strong temporal reference points. When compared with WM1 (reference) the NM1 (dependent) or NM2 (dependent) sequences show different effects: NM1 is compressed by 2.2% (that is, expansion of -2.2%) and NM2 is expanded by 1.5%. Comparisons between WM2 (reference) and NM1 (dependent) or NM2 (dependent) show a similar effect - NM1 is compressed by 2.9% (i.e., -2.9% expansion) and NM2 is expanded by 1.3%. The poorest time adjustment result is an expansion of 4.3%, exhibited when the two 'No Music' conditions (NM1/NM2) are compared.

In the comparison between WM1 and NM1 there is a sudden change at approximately 30 seconds into the dance (see top-centre profile of Figure). On viewing the performance videos, two of the three dancers were standing still for a period, and in the NM condition, one of the dancers did not wait long enough (with respect to the corresponding section in the WM condition), leading into the next section early. Furthermore, at around 148 elapsed seconds there is a dip in height movement, with a sharp change in direction in the time adjustment profile. If the musical soundtrack is listened to at this point, the downwards direction corresponds to a section of the music where there is a relaxation (in general tone and rhythmic direction, but not in tempo), and the upwards slope following the trough occurs during a section where there is a corresponding musical build-up (entries layering instrument by instrument). It is possible that in this section the dancer is distorting their internal impression of time (tempo) based on an influence from other musical attributes.

The comparison between WM1 and WM2 shows a dramatic change in time adjustment at approximately 70-80 s. However, when the original data sequences are inspected this is revealed to be an error. The feature (a jump forward in time) occurs when no significant physical movement in both conditions, which is clearly incorrect.

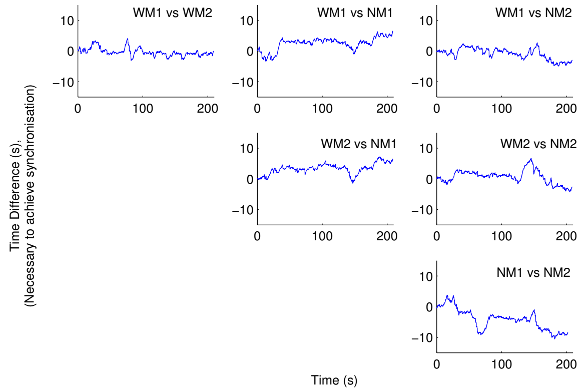

Figure 5: DTW and manual adjusted time adjustment profiles are compared for four window lengths -demonstrating that the 32 sample length results in the best match (minimised Mean Absolute Error and equal best in Goodness of Fit).

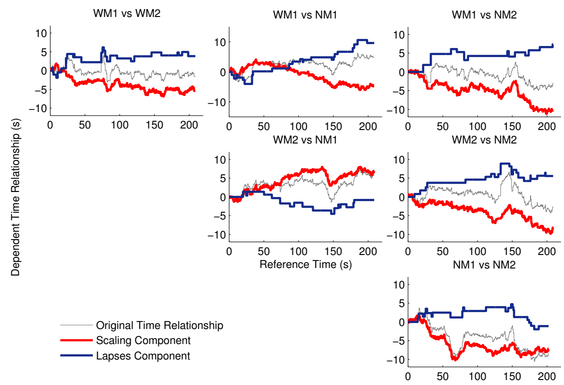

Figure 6: The time adjustment profiles for each of the comparisons, Trends in the profile suggest scaling, while abrupt changes demonstrate the presence of lapses. The plots, therefore, show that there is a mixture of scaling and lapsing occurring. The top row of graphs use WM1 as the reference, the middle row uses WM2, and the bottom row NM1.

3.2 Lapsing and Scaling Components

The scaling and lapsing components of each time adjustment profile were separated through a gradient threshold as mentioned above, and these components are graphed in Figure 6. In the following section we discuss the amount of 'compression' or 'expansion', that scaling or lapsing applies to the performance's length, even though clearly lapsing would actually involve addition or subtraction of time data.

The comparison between WM1 and WM2 shows the lapsing and scaling components separating early and then remaining mostly unchanged for the duration of the profile, contributing complementary amounts of adjustment, 1.8% compression for the lapsing component and 2.7% expansion for the scaling component.

When WM1 was compared against NM1 or NM2 we see sim-

Table 1: The final summary values in the time adjustment profile, the scaling component and the lapsing component. These values represent the accumulated time adjustment applied to the dependent (2nd) signal in seconds, and in parentheses the associated percentage compression (negative) or expansion (positive). Also presented are correlations between the lapsing component and the time adjustment profile, against between the scaling component and the time adjustment profile.

Comparison	Accumulated Time Adjustment	Accumulated Scaling	Accumulated Lapsing	Correlation (Lapsing, Scaling)
WM1 vs WM2	-1.6 (0.8%)	-5.44 (2.7%)	3.84 (-1.8%)	0.04, 0.61
WM1 vs NM1	4.80 (-2.2%)	-4.80 (2.4%)	9.60 (-4.4%)	0.70, -0.23
WM1 vs NM2	-3.00 (1.5%)	-10.20 (5.1%)	7.36 (-3.4%)	-0.21, 0.78
WM2 vs NM1	6.12 (-2.9%)	6.96 (-3.2%)	-0.84 (0.4%)	0.05, 0.74
WM2 vs NM2	-2.64 (1.3%)	-8.20 (4.1%)	5.56 (-2.6%)	0.33, 0.64
NM1 vs NM2	-8.36 (4.3%)	-7.08 (3.6%)	-1.12 (0.6%)	0.46, 0.87

ilar effects occurring - in the WM1 vs NM1 case we see scaling contributing 2.4% expansion and lapsing contributing 4.4% compression, and in the WM1 vs NM2 case scaling contributes 5.1% expansion and lapsing contributes 3.4% compression. In both cases the scaling component exhibited a consistent downward slope indicating consistent expansion, while the lapsing component had a less consistent upward slope. The comparisons between WM2 and NM1 or NM2 were less straightforward. Comparing WM2 with NM1 shows scaling contributing 3.2% compression and lapsing contributing almost no expansion (0.4%). However, comparing WM2 against NM2 shows lapsing contributing compression (2.6%) and scaling contributing expansion (4.1%). Finally, comparing NM1 against NM2 shows a small expansion from lapsing (0.6%), and a larger expansion from scaling (3.6%). These results are summarised in Table 1.

In several comparisons a consistent scaling effect was seen - in the WM1 vs NM1 and the WM2 vs NM1 comparisons the scaling component had a consistent slope, except for a short-term change diversion in WM2 vs NM1. WM2 and NM2 also showed a consistent downward slope with the exception of a single diversion. This slope would be expected where there was a general disparity between the base tempo of the two performances, or where a consistent dragging or speeding effect was occurring in one or other of the performances.

Finally, Pearson correlations were used to compare the similarity between both the lapsing and scaling components, against the original time adjustment profile (Table 1). In all comparisons except for WM1 vs NM1, the scaling component exhibits higher correlation to the time adjustment profile than the lapsing component does (see Table 1). The WM1 vs NM1 is an interesting case, as it seems the scaling component and the lapsing components have opposite trend directions, implying that perhaps the lapsing and scaling are accounting for each other. This pattern is not borne out in other similar comparisons, such as the repeated comparison, WM2 and NM2 (which may have been corrected or improved by the presence of music cues in the preceding condition, WM1).

Generally, in each comparison we can observe effects of both the scaling and lapsing components. Overall, the ‘with-music’ conditions were very similar in their repeated performances, and the ‘no-music’ conditions were relatively different from one another.

4. DISCUSSION

This paper reported a new analytic approach to examining the underlying processes responsible for a mismatch in two performances of the same dance work. A dance work was recorded using motion capture, and height information from one of the dancers was used for analysis. The dance work was performed first in ‘silence’ (without a music soundtrack) and then in the presence of a music soundtrack, and the two conditions were repeated. This provided an opportunity to see if consistencies emerged when comparing a dance work performed with music versus without. A previous study which analysed the comparison but without a repeat provided us with the prediction that lapsing would account for the largest amount of any mismatch between the music and no-music conditions. The other mechanism assumed to affect the dancer’s ability to perform perfectly synchronised across the conditions was scaling, and it was predicted that this would account for a negligible amount of time adjustment across conditions.

The dynamic time warping statistics applied to different dance conditions have been adapted to identify the kinds of processes that dancers are using under different dance conditions. The results are comparable to the previous study [5] that examined one performance with music and repeated without music, and uses an alternative, time-domain based approach. In that study as with this one, we found that dancers use both scaling and lapsing when making movement memory ‘errors’. As mentioned, that study focussed on a comparison of what we refer to here as the WM1 and NM1 conditions. In that study 10.45 seconds of lapsing were detected, compared with 9.6 seconds in present study, an error between the analyses of 10% which, considering the vastly different techniques applied (in particular time domain versus frequency domain approaches), are considered comparable.

However, the present study suggests greater effects of scaling than those reported in previous research. For example, the WM1 and NM1 conditions produced a scaling ratio of 0.997 (essentially unity) in the previous study [5], whereas the present study detected a compression of 4.4% (ratio of 0.956). In fact the correlation of the scaling component of the DTW was always better correlated with the overall time adjustment profile than the lapsing component, with the exception of NM1 and WM1 (see Table 1). This indicates that scaling not only made a non trivial contribution to the explanation of the underlying process, but that it better explained the difference in timing than did lapsing. The analytic technique based on dynamic time warping presented here therefore identifies some subtle nuances, that we think are improvements to the method of analysis.

The finding that scaling makes a non-trivial contribution to changes in duration may be a controversial one and in need of further investigation. The literature that suggests that scaling does not play a major role in explaining timing mismatch between two similar performances is drawn largely from music performance literature for pieces that have been thoroughly learnt. The present investigation examined a dance work that was not yet as thoroughly prepared. It may be that the deterioration of timing through scaling error is something that diminishes with practice. The second performance of the dance without music (NM2) may also have been influenced by the presence of music in WM1. This may have re-set the timing and would explain the greater differences between timing in the two no-music conditions.

4.1 Limitations

While we argue that the present technique has validity at various levels, given the calibration of the DTW parameters using empirical ‘proof of method’ testing, there were still limitations to this research. First, the motion data were formed from a large number of markers measured in 3 dimensions, which were combined to a single dimension for this study. It is possible accuracy may be improved by using multiple dimensions, but initial efforts have not resulted in significant improvements, and most differences between sets of marker positions will indeed be adequately represented within a median value. However, the naïve use of DTW applied to multiple dimensions is not sufficient -much greater use of heuristics within the algorithm is necessary, as each marker has the potential also to introduce noise which may overwhelm the signal.

Second, this study measured a single dancer in a trio of dancers. Much communication between the dancers, through watching and listening to each other and through eye contact, was observed when the videos were viewed. Much of the changes in time adjustment observed may stem from the three dancers interacting during the performances, rather than from the measured dancer's cognitive processes - the detailed investigation of which would require all three dancers to be measured using the motion capture system. The possibility of repeating such a study but with a solo dancer, for example, is worth considering.

A third limitation comes from the application of the traditional DTW technique itself. In Dynamic Time Warping the length of sequences compared will exponentially affect the size of the similarity matrices used to compare them, and the ensuing calculation time and memory necessary. The method also struggles in points where there is no activity (e.g. the dancers remain still). There exist various modifications to the traditional DTW algorithms that deal with such problems (e.g. [13]), and may result in improvements in accuracy when higher sample rates are used.

Fourth, no clear boundary between lapses and scaling effects was observed in this study, and thus the choice of threshold for distinguishing lapsing and scaling was relatively arbitrary. Improvements in the algorithm's accuracy and noisiness may result in a more detailed picture of the relationship. Alternatively, these definitions may require further classification, as there are probably different cognitive processes that are occurring when a section of a continuous movement is omitted, inserted or scaled, as opposed to when a dancer, after a period of no movement, begins a movement early or late. Schubert et al. [14] discuss this problem using a comparison between a calculus based description and one that employs semantic units. They suggest a solution may rely on the accurate definition of a grammar of dance, possibly involving the definition of a 'danceme' analogous to a phoneme in speech. Future research could include a behavioural experiment where lapsing and scaling are included deliberately to confirm the psychological reality of lapsing as a relatively fast change and scaling as a relatively slow change.

5. CONCLUSION

This research outlines ways of comparing two similar dance performances, using the dynamic time warping algorithm. We discussed the method and validated it by using it to time-adjust random sequences that have been altered, and by comparing them against a manual synchronisation method, finding an optimal arrangement was to use a difference transform, and a 32-sample window length.

Finally, we applied the method to a comparison of dances performed with a musical soundtrack and without a musical soundtrack. The timing in the two 'with-music' conditions was similar whereas timing in the 'no-music' conditions was less similar. We found a good degree of replication with one of the conditions that was analysed using a time-domain based approach, however, other conditions revealed that scaling may make a more important contribution to the changes in performer timing than previously thought. In conclusion, the area of dance cognition is likely to find significant benefits through the development of such analytic techniques, providing tools to examine actual variations in dance as a means of understanding underlying cognitive processing, such as entrainment and memorisation, without reliance on subjective self report techniques and speculation.

6. ACKNOWLEDGEMENTS

This research was supported by the Australian Research Council through its Linkage Project and Infrastructure grant schemes (Grant Numbers: LP0211991,LP0562687, LE0347784, LE0668448), MARCS Institute at the University of Western Sydney, UNSW via its Strategic Investment in Research Scheme, and industry partners The Australia Council for the Arts Dance Board, The Australian Dance Council (Ausdance), and QL2 Dance. Our thanks to the team at QL2–Mark Gordon, Ruth Osborne, Gary Barnes, Quantum Leap Choreographic Youth Ensemble, Emma Batchelor, James Batchelor, Rebecca Frasca, Sam Maxted, research colleagues Christian Kroos, Kim Vincs, Garth Paine, Shaun Halovic and research assistants Clare Howell, Eva Walker, Tim Byron, Tajana Opacic, Kirk Olsen, Julien Tardieu. Further information: http://marcs.uws.edu.au.

7. REFERENCES

[1] B. Calvo-Merino, C. Jola, D. E. Glaser, and P. Haggard, "Towards a sensorimotor aesthetics of performing art," *Consciousness and cognition*, 17(3), 911–22, 2008.

[2] A. Camurri, I. Lagerlof, and G. Volpe, "Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques," *Int. J of Human-Computer Studies*, vol. 59, 213–225, 2003.

[3] A. W.-C. Fu, L. Y. H. Lau, C. A. Ratanamahatana, and R. C.-W. Wong, "Scaling and time warping in time series querying," *The VLDB J*, 17, 899–921, 2008.

[4] J. Phillips-Silver and P. E. Keller, "Searching for roots of entrainment and joint action in early musical interactions," *Frontiers in Human Neuroscience*, 6(26), 2012.

[5] C. J. Stevens, E. Schubert, S. Wang, C. Kroos, and S. Halovic, "Moving with and without music: Scaling and lapsing in time in the performance of contemporary dance," *Music Perc.*, 26(5), 451–464, 2009.

[6] F. J. P. Langheim, J. H. Callicott, V. S. Mattay, J. H. Duyn, and D. R. Weinberger, "Cortical systems associated with covert music rehearsal," *NeuroImage*, 16, 901–908, 2002.

[7] B. H. Repp, "Acoustics, perception, and production of legato articulation on a digital piano," *J of Acoustical Soc. of America*, 97(6), 3862–3874, 1995.

[8] J. Oosterhoff, "Trimmed mean or sample median?," *Statistics & Probability Letters*, 20(5), 401–409, 1994.

[9] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[10] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Tran. on Inf. Theory*, 13(2), 260–269, 1967.

[11] R. Turetsky and D. Ellis, "Ground-truth transcriptions of real music from force-aligned midi syntheses," in *4th Int. Symp. on Music Inf. Retrieval*, (Baltimore), 135–141, 2003.

[12] C. J. Stevens, E. Schubert, R. H. Morris, M. Frear, J. Chen, S. Healey, C. Schoknecht, and S. Hansen, "Cognition and the temporal arts: Investigating audience response to dance using pdas that record continuous data during live performance," *Int J of Human-Computer Studies*, 67, 800–813, 2009.

[13] S. Dixon and G. Widmer, "Match: A music alignment tool chest," in *6th Int. Conference on Music Information Retrieval*, (London, UK), 492–497, 2005.

[14] E. Schubert, C. Stevens, and S. Ferguson, "The relationship between lapsing and scaling: Explaining timing variations in a contemporary dance performance," in *ASCS09: 9th Conf. Aust. Soc. Cog. Sc.* (W. Christensen, E. Schier, and J. Sutton, eds.), 311–315, 2010.