On Cross-Modal Perception of Musical Tempo and the Speed of Human Movement

Kathleya Afanador, Ellen Campana, Todd Ingalls, Dilip Swaminathan,
Harvey Thornburg, Jodi James, Jessica Mumford, Gang Qian, and StJepan Rajko

Arts, Media and Engineering Program
Arizona State University
Tempe, AZ 85287-8709
United States

Abstract. Studies in crossmodal perception often use very simplified auditory and visual contexts. While these studies have been theoretically valuable, it is sometimes difficult to see how the findings can be ecologically valid or practically valuable. This study hypothesizes that a musical parameter (tempo) may affect the perception of a human movement quality (speed) and finds that although there are clear limitations, this may be a promising first step towards widening both the contexts in which cross-modal effects are studied and the application areas in which the findings can be used.

1 Introduction

Three intriguing occurrences—a misunderstood word, a talking puppet, and an elusive collision—have propelled the psychological research on cross-modal perception in which audition and vision are inextricably intertwined. These occurrences are now known as the McGurk Effect [16], the ventriloquist effect [1], and the bounce-inducing effect [28] respectively, and all three have proven remarkably robust. Yet the study of cross-modal perception currently relies heavily on behavioral experiments using simple sounds and simple animations. While these studies have been theoretically informative, their contexts are so simplified that it is often difficult to see how the findings can be ecologically valid or practically valuable. The present study examines possible cross-modal effects of a musical parameter (tempo) on the perception of a human movement quality (speed) in hopes that this may be a first step towards widening both the contexts in which cross-modal effects are studied and the application areas in which the findings can be used.

1.1 Cross-Modal Perception

Discussions about cross-modal perception often center around the McGurk effect or the ventriloquist effect and their variants, both of which are situations in which vision dominates the effect. This paper is primarily interested in the opposite situation—when sound affects vision—and thus draws from examples in which sound directly affects perception of spatial/temporal organization and visual movement. The way in

R. Kronland-Martinet, S. Ystad, and K. Jensen (Eds.): CMMR 2007, LNCS 4969, pp. 235–245, 2008.
© Springer-Verlag Berlin Heidelberg 2008

which sound is combined, or not combined, with a visual display can influence perception of object organization and movement within the scene. O'Leary and Rhodes [21] for example, showed that the perceived organization of a sequence of high and low tones could influence the perceived organization of moving dots on a visual display. Auditory information alone may be perceived differently depending on tempo: at slow tempi, alternating high and low tones are perceived as a single stream of sound while at high tempi the high and low tones segregate into separate streams [4] The perception of visual information varies similarly: when dots are displayed moving from left to right in alternating high and low positions at slow rates, a viewer perceives a single dot moving up and down while at faster rates a viewer is more likely to perceive two dots moving horizontally. O'Leary and Rhodes showed that when the high and low tones were heard as two streams, viewers were more likely to see two dots even at rates which would, in a unimodal display, normally result in the perception of one dot. The perceptual organization of objects in the scene therefore also produced a change in the objects' perceived movement pathways. Examining this more directly, Sekuler, Sekuler, and Lau [28] showed that movement pathways can be interpreted differently in the absence or presence of sound through the bounce-inducing effect, where two moving targets are seen to stream through each other in silence but are seen to bounce off of each other when a sound is introduced at the moment of visual coincidence. This effect occurs because the visual stimulus is inherently ambiguous. Sound resolves the ambiguity by biasing a viewer to favor integrating sound and movement into a single event that makes sense [1].Yet Shams, Kamitani, and Shimojo [29] demonstrated that a single flash of light accompanied by multiple beeps is perceived as multiple flashes. Thus even when no ambiguity is present sound can qualitatively alter perception of a visual stimulus. These findings therefore support Vroomen and de Gelder's contention that "cross modal combinations of features not only enhance stimulus processing but can also change the percept." [31]

Perceptual judgment tasks have indicated that audition dominates vision in temporal processing. This is sometimes called auditory capture and it stems from claims that vision and audition are each more sensitive to spatial and temporal processing respectively and from evidence that one modality dominates the other when conflicting spatial and temporal information is presented. One such study by Repp and Penel [26] asked participants to tap their finger in synchrony with auditory and visual sequences containing an event onset shift, with the expectation that this would cause involuntary phase correction responses. Their auditory sequences consisted of identical high pitched piano tones and their visual sequences consisted of black X's on a screen and flashing lights. Within the unimodal conditions, audition produced the smallest variability in taps, larger phase correction responses, and better event onset shift detection. Interestingly, results from the bimodal condition were very similar to those of the unimodal auditory condition indicating that although viewers' attention was aimed at the visual sequences, they depended more upon auditory information to perform the task. If this holds true for more complex stimuli, it suggests the possibility that auditory information also dominates temporal perception when watching human movement. The bounce-inducing effect is additionally an example of congruence—the combination of two media that produces the perception of a relationship between them even when such relationships are coincidental. When two media are presented simultaneously, a viewer assumes relationships between the two media exist and thus looks for them [5][10][18]. Bolivar et al. termed this finding “visual capture” [5]; in other words, visual stimuli influence people to interpret simultaneously presented auditory stimuli as somehow related. Likewise auditory capture may occur from congruence, as when music influences people to perceive simultaneously presented visual material as somehow related. Lipscomb and Kendall [15] for example, paired an abstract film excerpt with a variety of different musical accompaniments and found that viewers perceived several musical choices as a “good fit.” Similarly, Mitchell and Gallaher [18] paired three different dance sequences with three different musical sequences and found that congruence was perceived among several different combinations of dance and music (not only between the dance and its intended musical selection). Although Bolivar et al. used visual images with a strong narrative context in their experiment, the findings of Lipscomb and Kendall as well as those of Mitchell and Gallaher suggest that the simultaneous presentation of abstract sound and movement may be well suited to produce perceptions of similarity which may, in turn, facilitate crossmodal effects.

1.2 Music Perception and Human Motion

Perception of sound with human movement has been studied to some degree in the area of music perception as it relates to dance. Much of this work focuses on establishing congruence between music and dance by focusing specifically on dynamic qualities [9], general emotion or style [18], or section beginnings and endings [12] of both sound and movement. One recent study, however, examined the effects of various sound parameters on imagined motion. Eitan and Granot [8] asked participants in their experiments to visualize an animated human character (cartoon) of their choice. They were presented brief musical selections, and for each selection were asked to visualize their character moving in an imaginary animated film shot with the given melody as its soundtrack. Their purpose was to analyze the relationship between music and motion in imagined space based upon Clarke’s [6] contention that “since sounds in the everyday world specify (among other things) the motional characteristics of their sources, it is inevitable that musical sounds will also specify…the fictional movements and gestures of the virtual environment which they conjure up” [8]. The experiment produced an asymmetrical model of imagined musical space—the fact that a musical stimulus seemed to suggest a particular kinetic quality did not imply that the opposite musical stimulus suggested the opposite kinetic quality. Central to the results of this experiment however, is the finding that by changing sound parameters, participants’ imagined motion would change predictably. This suggests that there may be certain natural affinities between sound parameters and movement parameters, yet the asymmetries discovered suggest that the way these affinities are structured may be somewhat nuanced.

2 Experiment

Among the various sound parameters in Eitan and Granot’s study, inter-onset-intervals (IOI, the interval of time between the onsets of successive sounds) were found to affect imagined motion most strongly and symmetrically. Decreasing and increasing intervals strongly influenced participants to imagine motion speeding up and slowing down respectively. In short, what we hear affects the movement we imagine. Historically and theoretically, this finding is not surprising. The association between tempo and human movement speed is arguably the most apparent sound-motion relationship. This association begins in early infancy, evident in high sensitivity towards "regular synchronization of vocal and kinesthetic patterns" [22] and this sensitivity continues to develop through childhood [19]. Humans seem to have an ingrained penchant for rhythmic synchronicity in their own movements [17], whether it is to synchronize with an auditory pulse or to synchronize with the movements of others around them. Phillips-Silver & Trainor have further established that for both infants and adults, auditory encoding of rhythmic patterns can be directly influenced by the movement of their own bodies [24][25]. In short, the movement we feel affects what we hear. Within the context of music and dance, it is also fairly common for different pieces of music to "bring out" different dynamic qualities in the same dance. Although this particular point has not been studied empirically, it may be supported somewhat by the congruence studies mentioned above and it hints at the possibility that a sonic change could cause a real change in the perception of a dynamic movement quality. In short, it may be suggested that what we hear affects the movement we see. Based on these reasons, the present study hypothesized that a decrease or increase in inter-onset-intervals would cause a change in the perception of visual movement speed. Would viewers be influenced to perceive a pairing of movement with a fast tempo as faster overall than a pairing of the same movement with a slow tempo? And if so, could it be conceptualized as a variation on auditory capture?

2.1 Method

Fourteen undergraduate students participated in this study on a voluntary basis and received one class credit for their time. They were instructed that the experiment was about the perception of human motion but beyond that, all were naïve to the purpose motivating this study. Stimuli consisted of videos showing six movements chosen from Laban Movement Analysis [14]—rising, sinking, advancing, retreating, spreading, and enclosing (Figure 1). A single dancer was recorded doing all six movements at three speeds—fast, medium, and slow—with a camcorder synchronized to a motion capture system resulting in 18 video clips and 18 corresponding motion capture data files.

The motion capture data was fed into a pattern recognition model, which analyzed the movement (100 frames/sec) based on the probability that one of these six movements was occurring (see Appendix for details). These probabilities were then put through a Max/MSP program, which generated sound from the data. The sounds produced were series of clicks, varying in IOI rate according to the speed of movement. Three base rates (550 ms, 500 ms, and 450 ms) and three maximum rates (150 ms, 100 ms, and 50 ms) were used to control the IOI and the probability ratings from the motion analysis determined the transition from base rate to maximum rate. Thus when no movement occurred, the IOI rate was simply the base rate; when fast movement occurred, the recognition model would rate the probability of one of these movements occurring very highly and consequently the IOI would decrease quickly but when slow movement occurred, the probability ratings increased more slowly causing the IOI also to decrease slowly. Each of the 18 data files was put through the Max/MSP

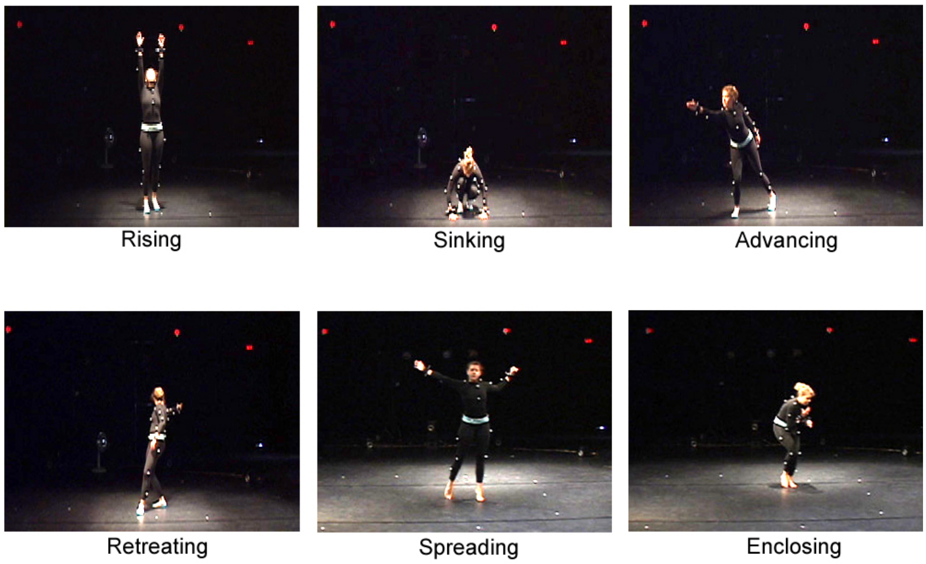

Fig. 1. The dancer demonstrating the shape qualities: rising, sinking, advancing, retreating, spreading, and enclosing

patch 9 times (one for each base rate/max rate pairing) and the resulting sound files were synchronized with their corresponding video clip, producing a total of 162 video clips. Participants watched the videos on a 20 inch wide screen computer and heard the sound through external speakers. They were presented each video individually followed by two statements with which they rated their agreement on a scale of 1 (strongly disagree) to 7 (strongly agree), indicated by numbered buttons. The first statement was always either "the movement was fast" or "the movement was slow." The second statement functioned primarily as a distracter. Participants saw each video exactly twice and responded to both the fast and slow statements for each video. The order in which the videos were presented was randomized and for each video, half of the participants responded to the fast statement first while half responded to the slow statement first.

2.2 Results

Changes in IOI were found to influence viewers' perception of human movement speed for one set of videos in our experiment. For the medium speed retreating movement, participants indicated significantly different levels of agreement with the statement "the movement was fast" as the minimum IOI length decreased, even though the actual movement they saw was identical across the different tempos. This difference was statistically significant ($F(2, 96) = 3.17$, $p < .05$). There were no statistically significant differences across inter-onset intervals for the other videos that we presented, although there was a main effect of actual movement speed (Figure 2), indicating that participants were attentive to movement speed and could accurately distinguish between slow, medium, and fast movements ($F(2, 1779) = 400.16$, $p < .01$).

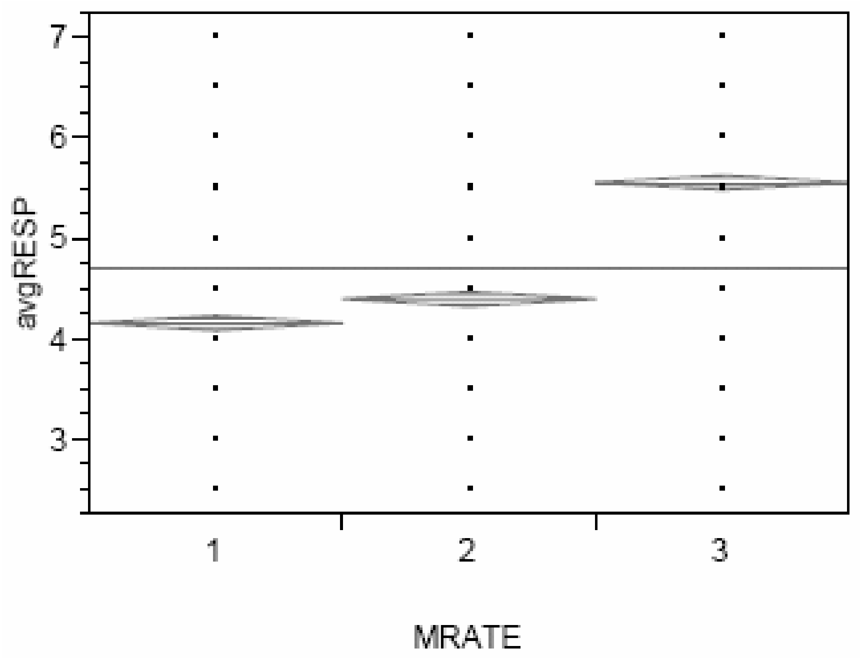

Fig. 2. This graph shows the average rating of movement speed (avgRESP) participants gave for the videos as a function of the actual speed of movement (MRATE). MRATE is defined as the actual movement rate and coded as 1 (slow), 2 (medium), and 3 (fast).

2.3 Discussion

The results provide some preliminary support for the hypothesis that differences in tempo, specifically inter-onset intervals, may be able to affect the perception of observed movement; however, the limitations of this study are clear. It is possible that the rating task used to measure perceived movement speed was either too cognitive or too coarse-grained. More precise measures achieved through a staircasing method may prove fruitful in attaining significant results. Additionally, it may be necessary to define objectively what is meant by slow, medium, and fast movement speeds. If many instances of cross-modal effects occur when there is ambiguity within one modality then it should be the case that ambiguous movement speed—the range within which it can be perceptually be interpreted as either fast or slow—would be most likely to show the effect. We are currently exploring both of these with additional tests.

3 Conclusions

Further studies will solidify the speculations prompted by this preliminary study and it remains to be seen how and where the findings of such studies can be applied. The various ways in which audition and vision have already been shown to interact make this area worth investigating further and furthermore, this area may prove interesting to artists and developers of computer interactive environments that use human movement as input. If sonic feedback is to be used as a response to movement, it will be informative to know what, if any, effects on human motion perception are caused by dynamic sound changes, how these effects may function differently within a given context, and how attention may facilitate or inhibit them.

References

[1] Alais, D., Burr, D.: The ventriloquist effect results from near-optimal bimodal integration. Current Biology 14(3), 257–262 (2004)

[2] Arulampalam, S., Maskell, S., Gordon, N., Clapp, T.: Tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing (2001)

[3] Bishop, C.: Neural Networks for Pattern Recognition. Oxford University Press, Oxford (1995)

[4] Bregman, A.S.: Auditory scene analysis. MIT Press, Cambridge (1990)

[5] Bolivar, V.J., Cohen, A.J., Fentress, J.C.: Semantic and formal congruency in music and motion pictures: Effects the interpretation of visual action. Psychomusicology 2, 38–43 (1994)

[6] Clarke, E.F.: Meaning and the specification of motion in music. Musicae Scientiae 5, 213–234 (2001)

[7] Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley, Chichester (1999)

[8] Eitan, Z., Granot, R.: How music moves. Music Perception 23(3), 221–248 (2006)

[9] Hodgins, P.: Relationships between score and choreography in 20th century music and dance: Music and metaphor. Mellen, London (1992)

[10] Iwamiya, S.: Interactions between auditory and visual processing when listening to music in an audiovisual context. Psychomusicology 13, 133–154 (1994)

[11] Jaynes, E.T.: Probability Theory: Logic of Science. Cambridge (2003)

[12] Krumhansl, C.L., Schenck, D.L.: Can dance reflect the structural and expressive qualities of music?: A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 5. Musicae Scientae 1(1), 63–85 (1997)

[13] Laban, R.V.: The Language of Movement: A Guidebook to Choreutics: Plays, inc. (1974)

[14] Laban, R.V.: Principles of dance and movement notation: With 114 basic movement graphs and their explanation. Macdonald & Evans, London (1956)

[15] Lipscomb, S.D., Kendall, R.A.: Perceptual judgment of the relationship between musical and visual components in film. Psychomusicology 13, 60–98 (1994)

[16] McGurk, H., MacDonald, J.W.: Hearing lips and seeing voices. Nature 264, 746–748 (1976)

[17] McNeill, W.H.: Keeping together in time: dance and drill in human history. Harvard University Press, London (1995)

[18] Mitchell, R.W., Gallaher, M.C.: Embodying music: Matching music and dance in memory. Music Perception 19(1), 65–85 (2001)

[19] Moog, H.: The development of musical experience of children in the pre-school age. Psychology of Music 4, 38–45 (1976)

[20] Murphy, K.: Dynamic Bayesian Networks:Representation, Inference and Learning. PhD thesis, University of California, Berkeley (2002)
[21] O'Leary, A., Rhodes, G.: Cross-modal effects on visual and auditory object perception. Perception & Psychophysics 35, 565–569 (1984)
[22] Papousek, M.: Intuitive parenting: A hidden source of musical stimulation in infancy. In: Deliege, I., Sloboda, J. (eds.) Musical Beginnings: Origins and Development of Musical Competence, pp. 88–112. Oxford University Press, Oxford, New York, and Tokyo (1996)
[23] Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers, San Mateo, Calif. (1988)
[24] Phillips-Silver, J., Trainor, L.J.: Hearing what the body feels: Auditory encoding of rhythmic movement. Cognition (in press, 2007)
[25] Phillips-Silver, J., Trainor, L.J.: Feeling the beat: movement influences infants rhythm perception. Science 308, 1430 (2005)
[26] Repp, B.H., Penel, A.: Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. Journal of Experimental Psychology: Human Perception and Performance 28(5), 1085–1099 (2002)
[27] Savitzky, A., Golay, M.J.E.: Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry, 1627–1639 (1964)
[28] Sekuler, R., Sekuler, A.B., Lau, R.: Motion perception. Nature 385, 308 (1997)
[29] Shams, L., Kamitani, Y., Shimojo, S.: Visual illusion induced by sound. Cognitive Brain Research 14, 147–152 (2002)
[30] Thornburg, H.: Detection and Modeling of Transient Audio Signals with Prior Information, Stanford University, Department of Electrical Engineering (2005)
[31] Vroomen, J., de Gelder, B.: Sound enhances visual perception: Cross-modal effects of auditory organization on vision. Journal of experimental psychology: Human perception and performance 26(5), 1583–1590 (2000)

Appendix

Laban Movement Analysis (LMA) framework is a systematic approach to understand, analyze and notate full body human movement, originated by Rudolf Laban [13]. For the most part, LMA is divided into four categories: Body, Space, Effort and Shape. The Shape, component of LMA in general elicits the form, or forming of the body. One sub-component of Shape, Shape Qualities, concerns itself with how the body changes its shape in a particular direction. Figure3 shows a body-centered coordinate system with horizontal plane side-to-side (across the shoulders), vertical/coronal plane head-to-toe, and sagittal plane back-to-front. Rising/sinking fall on the vertical plane, retreating/advancing fall on the sagittal plane, and enclosing/spreading fall on the horizontal plane as well as reveal the general folding and unfolding of the body. All movement is comprised of one, two or three of these qualities depending on the complexity of the movement itself. Metaphorically, how one embodies Shape Qualities can reveal nuances of one's mood or character. Currently, we have focused most of our attention on Shape Quality (SQ) analysis.

While it may not seem complicated to a casual observer, doing SQ analysis computationally is quite difficult because there is no single, consistent way one can express a particular quality. One may advance, for instance, by walking towards something, by

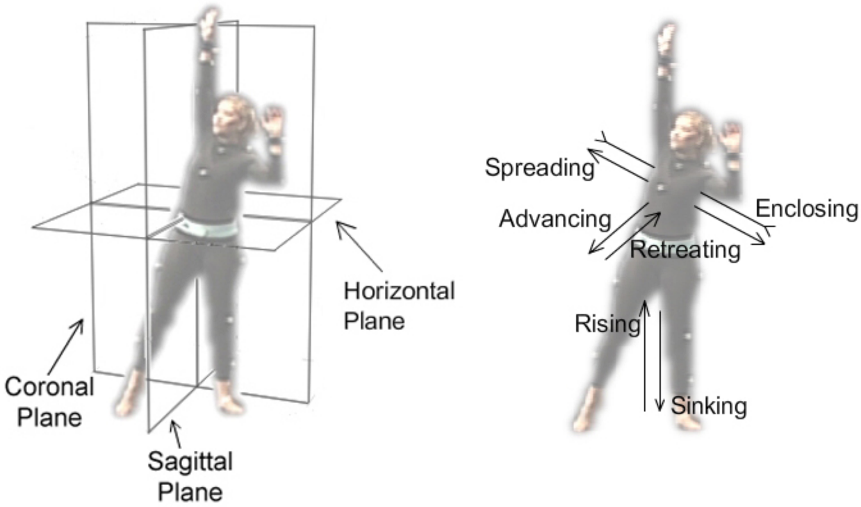

Fig. 3. Body-centric coordinate system and Shape Quality (SQ) illustrations

pointing, or simply by craning one's neck forward in a slight and subtle way. Nevertheless, SQs do imply tendencies on the movement of individual joints and limbs within the context established by a body-centric coordinate system with origin at the navel (Figure 3). For instance, if someone is rising, it is more likely that their torso will rise than sink. Similarly, SQs may imply non-local tendencies, such as an upwards shift of the body's enter of mass with respect to the horizontal plane.

We briefly present the key features of our framework "bottom-up" for inferring Shape Qualities, beginning with raw motion capture data and ending with the SQ hypothesis. First, we extract mid-level features from the raw, labeled marker position data. Second, we model movements of individual body parts and changes in global body characteristics (e.g. are the arms spreading or enclosing, the body centroid rising or sinking?) in terms of feature trajectory dynamics.

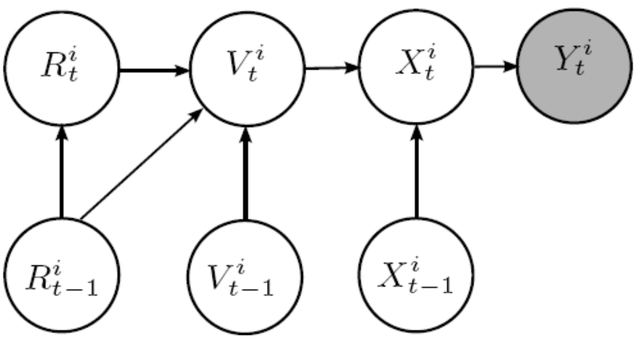

Fig. 4. Single time slice DAG of basic Shape Quality inference model

Figure 4 displays one time slice of the DAG (directed acyclic graph) or probabilistic influence diagram [20, 23, 30] underlying our model, with clear and shaded nodes representing hidden and observed variables respectively. A DAG is a graphical representation of a factorization of a joint probability distribution into conditional distributions. If a DAG consists of nodes $X_{1:N}$ the corresponding factorization of $P(X_{1:N}) = \prod_{i=1:N} P(X_i \mid Pa\{X_i\})$, where $Pa\{X_i\}$ are the parent nodes of X_i. For instance, in Figure.2 we have $P(R^i_{t-1}, R^i_t, V^i_{t-1}, V^i_t, X^i_{t-1}, X^i_t, Y^i_t) = P(R^i_t \mid R^i_{t-1})\, P(V^i_t \mid V^i_{t-1}, R^i_{t-1}, R^i_t)\, P(X^i_t \mid X^i_{t-1}, V^i_t)\, P(Y^i_t \mid X^i_t)$.

In Figure 2 Y^i_t: $i \in 1{:}3$, $t \in 1{:}N$, denotes each of the mid-level feature observations calculated at the current time t (frames; 100fps) using the motion capture data. These features depend on the body-centric coordinate system shown in Figure 1, displaying horizontal, coronal, and sagittal planes. These features are as follows.

- Y^1_t -- Mean marker height describes the global body position along the vertical axis (perpendicular to the horizontal plane). Changes in this feature are highly informative of rising/sinking.
- Y^2_t -- Frontward placement is the running sum of the change in mean marker position as projected onto the front direction (perpendicular to the coronal plane). Changes in this feature are highly informative of advancing/retreating.
- Y^3_t -- Lateral marker variance is the magnitude variance of all marker positions as projected onto the coronal plane. Changes in this feature describe the folding/unfolding of the body about its navel center, and are thus highly indicative of enclosing/spreading.

To reduce the effect of noise in the marker positions, as well as marker mislabeling and occlusion, the underlying state values are partially denoised using a second order Savitsky-Golay filter [27] over 20 frames.

Each mid-level feature observation, Y^i_t, is modeled as a corresponding inherent feature X^i_t, corrupted by noise to sensor inaccuracies. We model this noise as additive zero mean Gaussian noise: $P(Y^i_t \mid X^i_t) \sim N(X^i_t, \sigma^2_{Y,i})$

Also corresponding to each Y^i_t is a discrete Shape-indicator hypothesis $R^i_t \in \{-1,0,1\}$ corresponding to different Shape Qualities depending on i and models X^i_t as *decreasing* (-1), *constant* (0), or *increasing* (1). For example, suppose the inherent mean marker height is increasing. We denote this fact by $R^1_t = 1$, which specifies that X^1_t trends upward and implies that the Shape Quality expressed on the vertical plane is rising. Similarly, $R^1_t = 0$ and $R^1_t = -1$ denote neutral and sinking qualities on the vertical plane.

We observe that the trajectory of X^1_t is smooth, and displays tendencies conditioned upon the Shape Quality expressed at time t. To model the dynamics of X^i_t, we specify V^i_t, as the time derivative of X^i_t i.e., $P(X^i_t \mid X^i_{t-1}, V^i_t)$ concentrates deterministically on $X^i_t = X^i_{t-1} + V^i_t$. The tendencies displayed by the trajectory of V^i_t given R^i_t are as follows:

1. $V^i_t > 0$, $V^i_t \approx V^i_{t-1}$ when $R^i_t = 1$ and $R^i_{t-1} = 1$
2. $V^i_t < 0$, $V^i_t \approx V^i_{t-1}$ when $R^i_t = -1$ and $R^i_{t-1} = -1$
3. $V^i_t > 0$, $V^i_t \neq V^i_{t-1}$ when $R^i_t = 1$ and $R^i_{t-1} \neq 1$
4. $V^i_t < 0$, $V^i_t \neq V^i_{t-1}$ when $R^i_t = -1$ and $R^i_{t-1} \neq -1$
5. $V^i_t \approx 0$, when $R^i_t = 0$

$P(V^i_t \mid V^i_{t-1}, R^i_{t-1}, R^i_t)$ is developed by encoding the above tendencies using Jaynes' principle of maximum entropy, according to which maximum entropy estimate is the least biased estimate possible on the given information which is maximally noncommittal with regard to missing information. Hence to deal with situations involving uncertainty i.e., incomplete knowledge regarding the system to be modeled, the optimal solution is the model that satisfies the knowledge or constraints that we have regarding the system and the one which has the maximum entropy [11]. Let us first consider the first two cases i.e., we have $V^i_t > 0$, when $R^i_t = 1$ and $R^i_{t-1} = 1$ and $V^i_t < 0$ when $R^i_t = -1$ and $R^i_{t-1} = -1$. Furthermore, we expect some continuity of Vit; i.e., $V^i_t \approx V^i_{t-1}$, which can be controlled by $E| V^i_t - V^i_{t-1}| < \sigma^2_{V,i}$. Putting these constraints together and using the methods in [7], we can solve for the maximum entropy dependence in closed form.

$$V^i_t \sim N^+(V^i_{t-1}, \sigma^2_{V,i}),\ R^i_t = 1,\ R^i_{t-1} = 1 \tag{1}$$

$$V^i_t \sim N^-(V^i_{t-1}, \sigma^2_{V,i}),\ R^i_t = -1,\ R^i_{t-1} = -1 \tag{2}$$

where N^+ and N^-, are Gaussian distributions truncated to be positive and negative respectively, sharing the mean V^i_{t-1} and variance $\sigma^2_{V,i}$.

When the Shape Quality changes from $t-1$ to t we do not constrain $V^i_t \approx V^i_{t-1}$, we instead allow for sudden changes in dynamics, weakly constraining $V^i_t \approx V^{1,i}$ (for $R^i_t = 1$ and $R^i_{t-1} \neq 1$) and $V^i_t \approx -V^{2,i}$ (for $R^i_t = -1$ and $R^i_{t-1} \neq -1$) where $V^{1,i}, V^{2,i} > 0$ are nominal values;

$$\mathrm{V}^i_t \sim \mathrm{N}^+(\mathrm{V}^{1,i}, \sigma^2_{\mathrm{V1,i}}),\ \mathrm{R}^i_t = 1,\ \mathrm{R}^i_{t-1} \neq 1 \tag{3}$$

$$V^i_t \sim N^-(-V^{2,i}, \sigma^2_{V2,i}),\ R^i_t = -1,\ R^i_{t-1} \neq -1 \tag{4}$$

For $R^i_t = 0$, we have $V^i_t \approx 0$; using similar arguments as above we specify: $V^i_t \sim N(0, \sigma^2_{z,i})$. To complete the description of our model we need to specify $P(R^i_t \mid R^i_{t-1})$. We define q^+ as the probability of transition out of state 1, meaning the probability that $R^i_t \neq 1$ given $R^i_{t-1} = 1$. The average length of 1 regions is approximately $1/q^+$ and $p^+=1-q^+$. Similarly, q^- is the probability of transition out of state -1, with $p^-=1-q^-$. Now we can proceed to specify $P(R^i_t \mid R^i_{t-1})$ as a first order Markov, state transition distribution (Table.1).

Table 1. *Specification of* $P(R^i_t \mid R^i_{t-1})$

R^i_{t-1}	$R^i_t = 1$	$R^i_t = 0$	$R^i_t = -1$
1	p^+	$1- p^+- q^+p^-$	q^+p^-
0	p^+	$1- p^+- p^-$	p^-
1	q^-p^+	$1- q^-p^+- p^-$	p^-

The overall dynamic Bayesian network is in the form of a non-linear, non-Gaussian switching state space model. Via SIR particle filtering methods \cite{Arulampalam01}, we compute the posterior distribution of the SQ hypothesis given all feature observations up to and including the present time and choose Rit that maximizes this posterior; i.e. the filtered posterior $P(R^i_t \mid Y^i_{1:t})$, $i \in 1{:}3$. It is well known that this choice of R^i_t yields the minimum-error decision [2].