Looking for a Simplicity Principle in the Perception of Human Walking Motion

by

Giles Herbert James Holland

A thesis submitted to the Centre for Neuroscience Studies

in conformity with the requirements for

the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

October, 2010

ABSTRACT

The simplicity principle posits that we interpret sense data as the simplest consistent distal cause, or that our high level perceptual representations of stimuli are optimized for simplicity. The traditional paradigm used to test this principle is coding theory, where alternate representations of stimuli are constructed, simplicity is measured as shortness of representation length, and behavioural experiments attempt to show that the shortest representations correspond best to perception. In this study we apply coding theory to marker-based human walking motion. We compare two representation schemes. The first is based on marker coordinates in a body-centred Cartesian coordinate system. The second is based on a model of 15 rigid body segments with Euler angles and a Cartesian translation for each. Both of our schemes are principal component (PC)-based implementations of a norm-based multidimensional object space – a type of model for high level perceptual schemes that has received attention in the literature over the past two decades. Representation length is quantified as number of retained PC's, with error increasing with discarded PC's. We generalize simplicity to efficiency measured as error across all possible lengths, where more efficient schemes admit less error across lengths. We find that the Cartesian coordinates-based scheme is more efficient than the Euler angles and translations-based scheme across a database of 100 walkers. In order to link this finding to perception we turn to the caricature effect that subjects can identify caricatures of familiar stimuli more accurately than veridicals. Our design was to compare walker caricatures generated in our two schemes in the hope of finding that one gives caricatures that benefit identification more than the other, from which we would conclude the former to be a better model of the true perceptual scheme. However, we find

that analogous caricatures between the two schemes are only distinguishable at caricature levels so extreme that identification performance breaks down, so our design became infeasible and no conclusion for a simplicity principle in walker perception is reached. We also measure a curve of increasing then decreasing identification performance with caricature level and an optimal level at approximately double the distinctiveness of a typical walker.

ACKNOWLEDGEMENTS

Thanks to NSERC as well as to the Centre for Neuroscience at Queen's University and my supervisor Dr. Nikolaus Troje for funding this thesis. Thanks to the many members of the Centre for creating and fostering a welcoming and creative academic environment. Also sincere thanks to Dr. Troje in particular for many things, but on the short list comes patience, understanding, and open-mindedness. And thanks of course to my parents for everything.

TABLE OF CONTENTS

TABLE OF TERMS

General Acronyms

PLD	point light display
DFT	discrete Fourier transform
PCA	principal components analysis
PC	principal component

General Terms

representation	a realized surrogate for a specific object – e.g. a JPEG or bitmap of a specific picture
(representation) scheme	for a class of objects specifies everything necessary to translate those objects into representations of a particular kind as well as to interpret those representations – e.g. the JPEG and bitmap formats in general
linearize	To represent a class of objects along a finite number of linear dimensions in a morphable fashion, accomplished in our pre-PCA schemes. Better quality linearization captures more variance along fewer dimensions and can be gauged by PC decomposition efficiency.
Fourier decomposition	Transformation of data into a series of Fourier components. Higher components capturing less variance may be discarded and typically error is acquired.
PC decomposition	Transformation of data into the mean vector plus a series of PC's. Higher PC's capturing less variance may be discarded and typically error is acquired.
representation efficiency	total error in a representation of an object across representation lengths, with less error across lengths corresponding to greater efficiency
scheme efficiency	representation efficiency averaged across representations under the scheme
decomposition efficiency	Error due to Fourier or PC decomposition across number of retained harmonics or PC's, with less error across numbers corresponding to greater efficiency. Usually averaged across objects.

Walker Representation Schemes

basic coordinate	Walkers are represented as time series of low-level coordinates.
pre-PCA	Basic coordinate data are linearized by Fourier decomposition.
PC-based	Linearized data are optimized for scheme efficiency by PCA.

cm	time series of Cartesian coordinates of markers (basic coordinate)
cmF	cm Fourier-based (pre-PCA)
cmPC	cm PC-based
ets	time series of Euler angles and translations of segments, with markers represented by constant coordinates in the coordinate systems of the respective segments to which they are attached (basic coordinate)
etsF	ets Fourier-based (pre-PCA)
etsPC	ets PC-based

Walker Marker Sets

61	61-marker set – all markers
15	15-marker set – one marker per segment, used for PLD's
15clav	15-marker set with mid-clavicle marker substituted on thorax – used for stick figures
45	45-marker set – origin + two other markers per segment. The minimum number of markers for which cm- and ets-based representations encode identical information, used in the computational experiment.

CHAPTER 1. INTRODUCTION

1.1. Overview

Marker-based models of the human body see extensive use in a range of fields, including computer character animation, biomechanics, machine learning and gait analysis, and perceptual psychology (Johansson 1973, Cutting 1978, Troje 2008). In these models the body is depicted by markers rigidly attached at key locations – typically the major joints and a judicious selection of other locations such as the head and chest. Marker-based walking ("walker") data can be represented in numerous different ways. In this study we work with two basic ways that are in common usage. The first is a standard for motion capture data representing motion in terms of markers with coordinates in a common body-centred 3D Cartesian coordinate system as a function of time. The second is a standard used in computer character animation representing motion in terms of an articulated system of 15 rigid body segments – mostly limb segments such as upper arms, forearms, thighs, shanks, etc. – with three Euler angles and three Cartesian translation components encoding pose as a function of time for each, and with markers represented only by constant coordinates in the moving Cartesian coordinate systems of the respective segments to which they are attached.

Turning for the moment from representation schemes for data in the lab to schemes for sense data plausibly implemented by the neural architecture of the human perceptual system, the norm-based multidimensional object space model has enjoyed success in the study of perception of several classes of objects over the past two decades (Valentine 1991, Rhodes et al. 2005, Troje 2002). The model conceptualizes the high level perceptual representations of stimuli in a given class (in this study walkers) as

vectors in a space whose origin corresponds to the mean stimulus or "norm" from the subject's experience of the class and whose dimensions capture features along which stimuli in the class statistically vary. The dimensions can be numerous and the features and class are high level. Principal components analysis (PCA) applied to calculate a low dimensional high level vector space for stimulus data in the lab has been one especially fruitful way of implementing object space models of perceptual schemes (Turk & Pentland 1991, Vetter & Troje 1997, Troje 2002).

The simplicity principle in perceptual psychology posits that we interpret ambiguous sense data as the simplest distal cause consistent with it, or that our high level perceptual representations of stimuli are optimized for simplicity (Attneave 1954, Hatfield & Epstein 1985, Chatter 1997, 1999, Van der Helm 2000). The traditional paradigm used to test this principle is coding theory, where alternate representations of stimulus data in the lab are constructed, simplicity is measured as shortness of representation length, and behavioural experiments attempt to show that for a given stimulus the shortest representation corresponds best to perception (Hochberg & McAlister 1953, Leeuwenberg 1971, Restle 1979). In this study we apply coding theory to ask how our two kinds of schemes in the lab relate to the true high level perceptual scheme for walkers. We test two principal component (PC)-based schemes, one based on Cartesian coordinates of markers, the other on Euler angles and translations of segments. Both are analogous implementations of the norm-based multidimensional walker space model. In a computational experiment we compare how efficiently these two schemes encode walker data from an informational point of view. In behavioural experiments our original design was to then ask whether the scheme found to be the more efficient is also

closer to the true perceptual scheme for walkers as a simplicity principle would predict. Specifically, we turned to the caricature effect established in the literature that subjects can identify caricatures of familiar stimuli more accurately than corresponding veridicals (Rhodes et al. 1987, Pollick et al. 2001). Both of our PC-based schemes can be used to generate caricatures from veridical walker data. Our design was to compare caricatures generated in the two schemes in the hope of finding that caricatures generated in one benefit identification more than caricatures generated in the other, from which we would conclude the former to be a better model of the true perceptual scheme. As will be seen, the caricature approach proves infeasible.

1.2. The likelihood principle, the simplicity principle, and coding theory

Consider three stimuli depicted in Figure 1 below. Something noteworthy about each of these stimuli, and in general about all of our sense data, is that we perceive not the raw sense data per se but rather an interpretation of its distal cause. Moreover, we do so under incomplete information; for each stimulus an endless number of causes are physically consistent with it but our perceptual system usually infers only one (and usually the correct one). Figure 1a is very straightforwardly compelling: both views could have been generated by a two-dimensional projection of the same transparent cube at different orientations but while in the right view we often perceive a cube, in the left, which would correspond to a much more accidental orientation, we usually perceive a two-dimensional pattern (Hochberg & McAlister 1953). In Figure 1b the motion of either dot in isolation may be perceived as moving in a cycloidal pattern as depicted by the solid lines but typically we perceptually organize both dots into a vivid impression of the rolling wheel

that was used to generate the display (Cutting & Proffitt 1982). And finally in Figure 1c we typically see a skipping rope rather than some independent, coincidentally related motion of six dots (Restle 1979).

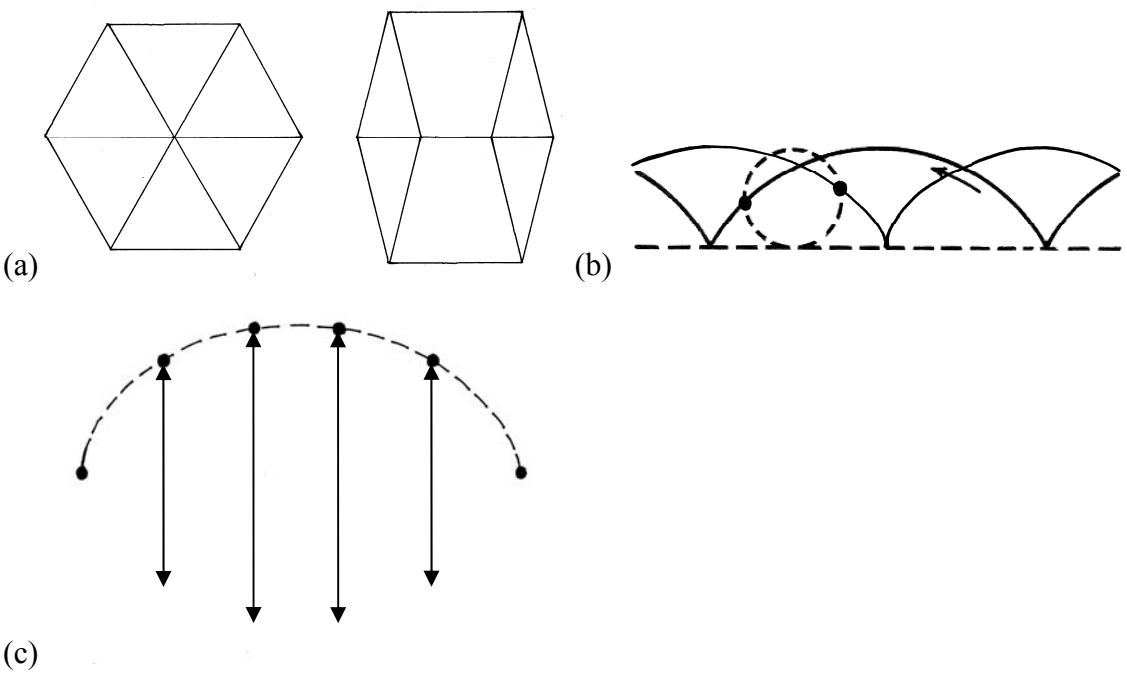

Figure 1. (a) 2D figures or a 3D cube from accidental orientations. (b) Lines depict the paths of two lights attached to opposite edges of a wheel rolling at constant speed in the dark – the wheel is depicted as the dashed circle but would be invisible. (c) Lines depict the paths of lights attached to a swinging skipping rope in the dark.

If any number of causes could give rise to any given phenomenon, how does our perceptual system so quickly and effortlessly abstract one to the exclusion of others? This is the problem of perceptual inference. A first guess as to the criteria the perceptual system uses could follow from consideration of the basic fact that it goes beyond sense data to perceive causes at all: the perceptual system is specialized by evolution to allow us to interact optimally as agents with the exterior world, so perhaps interpretations are decided such that they are the most likely. Such a fundamental operational principle in

perceptual psychology is referred to as the likelihood principle. In its modern formulation it is often modeled using Bayesian probability:

$$P\left(I|D\right) \propto P\left(D|I\right) P\left(I\right) \tag{1}$$

This equation is read as the probability of a potential interpretation *I* given a set of sense data *D* is proportional to the probability of having received that sense data assuming the interpretation is true multiplied by the prior probability of encountering that interpretation at all. The likelihood principle states that the perceptual system picks interpretations that maximize $P\left(I|D\right)$. The latter two terms are most easily seen as known through prior experience and to a certain extent through innate programming (Yuille & Kersten 2006). For example, from the point of view of the likelihood principle we perceptually organize the two dots in Figure 1b into a wheel rather than perceive them as independent cycloids because in nature, or at least in our typical environments, a wheel is much more likely than independent dots moving with such a relationship only by coincidence.

Given the purpose of the perceptual system to facilitate interaction with the exterior world, likelihood may be taken prima facie as a principle of its operation. Nonetheless, this does not preclude other principles from playing roles that are either counterbalanced or commensurate with likelihood. One such principle that has been a focus in perception research for at least a century is the simplicity principle. While the likelihood principle states that we infer the most likely causes of our stimuli, the simplicity principle states that we infer the *simplest* causes. The simplicity principle is perhaps more subtle than the likelihood principle in that it is not immediately clear what precisely is meant by simplest. It is perhaps intuitive that a rolling wheel with two lights

on it is simpler than two independent but coincidentally moving dots in Figure 1b. But what exactly makes this interpretation simpler, or said differently, how exactly can we quantify how simple competing interpretations are? Indeed, arriving at more appropriate and precise definitions of simplicity in the context of perception has been a focus throughout much of the history of research into the simplicity principle.

In the mainstream history of perceptual psychology the simplicity principle was first elaborated in Gestalt psychology. Here researchers noted that we tend to interpret stimuli globally at a primary level, rather than secondary to a reductionist analysis of their local features, and that we tend to interpret them in ways that are regular, symmetric, redundant… or just "simple". Gestaltists referred to a tendency of the perceptual system to draw interpretations that maximize Prägnanz, which is sometimes translated to English as "goodness" but is probably closer to the word "prominence". They outlined a number of "laws", each according to a different aspect of Prägnanz. These included, not exhaustively: the law of closure, that we interpret features not actually present in a stimulus in order to complete regular figures such as occluded shapes; the law of similarity, that we perceptually organize similar features into coherent wholes, where similarity can be based on many different kinds of attributes such as form, colour, etc.; the law of symmetry, that we organize features into coherent wholes so as to maximize symmetry in those wholes; the law of continuity, that we organize and complete patterns and trends; etc. Gestalt theory is fairly rudimentary: many of its laws are not mutually exclusive, how to combine them when they are not or counterbalance them when they are is not always clear, and neither is how to quantify them (Buffart, Leeuwenberg, & Restle 1981, Chater 1997). Prägnanz is a unifying label but not so much

a unifying concept – it is defined only implicitly through the many laws. Nonetheless, the usefulness of Gestalt theory lies in its elucidation of at least some trends that seem to be at play in perception, all of which do generally point toward some gist of simplicity. It was a useful stepping stone and remains useful today at least as a pedagogical tool.

Later studies looked at organization of moving dot patterns – an example of which is our two dots on a rolling wheel – and attempted to identify simplicity with such notions as minimizing the relative motion of the dots leaving common motion as a residual, or vice versa, or various other geometric quantities (Cutting & Proffitt 1982). Results tended to be long and complicated and principles continued to be overly specific, and in turn only applicable to overly specific kinds of stimuli.

A hint of a more sophisticated definition of simplicity can be found in the understanding of a simplicity principle in a different context altogether: that of the scientific method itself. Here a deeply intuitive simplicity principle is well known as Occam's Razor. We can imagine presenting a scientific explanation for a phenomenon, just as our perceptual system presents an inferred explanation for a stimulus. More complex explanations will be longer and simpler explanations will be shorter, and moreover if we give the simplest explanation we can then the simpler the phenomenon the shorter the explanation. And of course Occam's Razor characterizes such a minimal explanation (or theory) as optimal and preferred. The Razor is a powerful maxim that (ironically) has a propensity for being captured in strikingly simple one-line quotations – for example Einstein's advice to make things "as simple as possible, but not simpler" (Chater 1999).

If we think along the lines of Occam's Razor, the notions of complexity and simplicity seem to involve the lengths of representations of items of information. A precise mathematical theory of information and complexity in which representation and representation length play central roles arose in Shannon's information theory in the late 1940's (Shannon 1948). Information theory was designed initially as a theory of signal transmission in engineering. It deals with items of information drawn from a finite class and representation schemes that assign to each item in the class an unambiguous representation composed of a finite sequence of symbols drawn from a finite alphabet (such as digits). Representation length is measured as number of symbols – e.g. for binary codes where symbols are either 0 or 1 length is measured in bits. Representation "simplicity" is simply shortness of length. Because of the constraints that representations as well as the alphabet of symbols are finite and that no two items can have the same representation under a scheme, assigning shorter representations to some items comes at a cost of being left with only longer ones available for others. The main problem addressed in information theory is how to devise a maximally efficient scheme where representation length averaged across the class is minimized. Information theory famously identified the "amount of information" in any given item as the length of its representation under such an optimal scheme.

In the 1950's, in an effort to unify the Gestalt laws of simplicity and inspired by information theory, psychologists introduced coding theory methodology. Coding theory measures the simplicity of an interpretation of a stimulus as the shortness of the shortest representation of that interpretation under a "coding language" specified at the beginning of each study. Representations are sequences of elements and length is typically

measured as number of elements. A coding language defines the types of elements that are available and the instructions by which they are interpreted with respect to features in information to be represented. Depending on the kinds of stimuli, interpretations, and representations considered, in some studies elements have been symbols or digits, in others parameters ranging from single numbers to entire vectors, while in others elements from a formal grammar or instructions. For each interpretation of each stimulus, one or more representations of the stimulus data in the lab are constructed, meaning representations that compress informational patterns, redundancies, and constraints in the data definitive of that interpretation. The simplicity principle holds that for a given stimulus the perceptually favoured interpretation will be the one whose representation has the shortest length. The greater the difference in length between the shortest representation of each competing interpretation, the greater the degree of perceptual prominence the simplest interpretation is hypothesized to have (and if two or more interpretations have equal lengths those interpretations are hypothesized to be bistable or multistable). Researchers test the simplicity principle by testing if such hypotheses hold in behavioural experiments.

By way of example, consider the 2D projections of cubes in Figure 1a, which were investigated in the earliest coding theory study (Hochberg & McAlister 1953). The authors used a coding language where one element is used to represent each uninterrupted line segment in a stimulus. The first stimulus represented as a 2D interpretation requires a total of 12 line segments to be specified while the second requires 16. Meanwhile, both stimuli represented as the same 3D interpretation – the cube that generated the projections – require 12 line segments (the number of edges in a cube).

Hochberg & McAlister noted that for the second stimulus the representation of the 3D interpretation is shorter and thus that interpretation simpler than the 2D interpretation, and correspondingly a cube is perceived in 99% of the subjects tested. For the first stimulus, representation lengths are the same; correspondingly, 60% of subjects perceived it as a 2D pattern and 40% as 3D. Over these and two other projections of the cube, the authors noted a correlation between simplicity of the 2D interpretation and percentage of subjects perceiving that interpretation.

Due to its preliminary nature the above example constitutes a fairly basic coding theory study whose coding language could accommodate only a very limited class of stimuli and interpretations. The skipping rope example in Figure 1c was treated in a later and more advanced study (Restle 1979). Here the coding language was able to handle any 2D projected display of circular motion of dots, including hierarchies of nested motion. Representation elements were dot parameters specifying amplitude, phase, frequency, angle of tilt, angle between the plane of rotation and the plane of display, and centre of motion. The language could also accommodate parameters whose values are specified once and interpreted as common across multiple dots. Representing the skipping rope stimulus as an independent dots interpretation requires six elements to represent position and $5 \times 6 = 30$ elements to represent motion, for a total of 36. Representing it as dots on a skipping rope requires only six elements for position and four motion elements common to all dots plus three amplitudes, each of which is common to two dots, for a total of 13. Again, the skipping rope interpretation, with the shorter representation length, was the perceived alternative for tested subjects. The study found similar results for numerous other circular dot motion stimuli.

Buffart, Leeuwenberg, & Restle (1981) used a coding language capable of representing any 2D line drawing. The language used straight line segments of variable lengths deviating at variable angles to one another, and could represent curves as an arbitrarily large number of arbitrarily short line segments. The coding language contained provisions for repeating sequences of line segments in representations, iterating sequences, continuing sequences until line segments already represented are encountered, generating reversals and symmetries of sequences, etc. The authors applied their coding language to investigate 25 2D figures, all of which could be naturally interpreted as one shape occluding a second shape. For each, representations were constructed for a tiled interpretation (no occlusion) and one or more occluded interpretations differing by which shape was hidden behind the foreground shape. Remarkably, for all 25 shapes the simplest representation corresponded to the perceptually favoured interpretation for tested subjects. The authors also noted that of a variety of more specific notions of simplicity, such as Gestalt laws of symmetry or continuity, all incorrectly predicted interpretations for at least one figure. It is interesting to note that the coding language used in this study was actually a simplified version of one capable of handling 3D line drawings as proposed in Leeuwenberg (1971). In that study, representations were constructed for many extravagant example figures. These aside, not all coding theory studies have faired as well: see Boselie (1988), and Boselie & Wouterlood (1989), which used coding languages identical or nearly identical to that of Buffart, Leeuwenberg, & Restle (1981) but returned more mixed results.

If we accept that the perceptual system operates by a likelihood principle at least to a large extent, how does a simplicity principle fit in? If simplicity is a counterbalancing

principle to likelihood then new theoretical difficulties arise in answering questions such as to what degree we should expect to observe it and to what degree it should defer to likelihood, how it is counterbalanced, and indeed why it should be a counterbalancing principle at all. Recent arguments drawing from information theory, algorithmic information theory, and methods from statistics and computer science such as minimum description length inference are suggestive that the simplicity principle is commensurate or perhaps even synonymous with the likelihood principle – specifically, that interpretations of stimuli that are maximally simple also tend to be maximally likely in the Bayesian sense of equation 1 (Chater 1996, Van der Helm 2000, Chater & Vitanyi 2003). Indeed, an open question in the study of the likelihood principle is how the perceptual system maximizes likelihood from a practical point of view – for example, how it might possess or acquire the probabilistic components of equation 1 for the wide variety of data and interpretations it must deal with. Maximizing simplicity may be a more straightforward problem to operationalize. Thus, if Chater's proposition is correct, the likelihood principle might be what the visual system is doing and the simplicity principle how it is doing it. This is an important line of thought to bear in mind because it lets our prior confidence in the likelihood principle lend legitimacy to the search for a simplicity principle, or at least not cast doubt on its appropriateness.

1.3. Coding theory applied to walking motion

Research into the perception of walking motion and other forms of biological motion took off when Johansson (1973) introduced the point light display (PLD) stimulus, a marker-based model of the human body visualized solely as white dots or "point lights"

for markers on an otherwise black background. Typical PLD stimuli have 15 or fewer point lights, mostly at major joints. Johansson's early PLD's were constructed from subjects wearing retroreflective markers video-taped in a dark-lit room (Johansson 1973). Since then methods have evolved to encompass the use of digital motion capture data where 3D coordinates of markers are tracked by an array of infrared cameras and recorded.

Research has shown that our perceptual system is able to effortlessly organize PLD's into a vivid interpretation of human form; in fact, this is possible even with fewer than 10 dots and 200 ms display times (Johansson 1976). In addition to perceptual organization, many lines of research into the perception of biological motion have grown since the 70's, a small sample of which include its robustness under various distracters and masks, the ability to extract specific information such as gender, emotion, and identity, various developmental and animal behavioural aspects, the contribution of basic structural versus kinematic and local versus global cues, the role of attention and peripheral versus foveal perception, the inversion effect in walking motion, fMRI and other imaging studies addressing neural substrates of biological motion perception, etc. – see Troje (2008) for a partial review. In psychology marker-based stimuli can be a way of representing basic body structure and kinematics in isolation from other body information such as texture, sound, surface shape[1], etc. Thus, they allow researchers to investigate the perceptual response to biological motion information in precise and controlled ways.

Coding theory, or at least a variant of it, has been applied to PLD walkers in one previous study in the literature (Cutting 1981). This study used PLD's at a side

[1] Surface shape information is minimized but perhaps not completely eliminated since markers are usually located on the surface of the skin.

orientation generated artificially using an algorithm based on sinusoids and basic anatomical length parameters to calculate point light trajectories – the so-called "Cutting walker" (Cutting 1978). Such displays are imperfect but still convincing and sufficient for many purposes. The Cutting walker is an articulated system of joints, which may or may not coincide with markers, each moving with the positional range of a pendulum about another joint lower in the hierarchy. For example, a wrist joint has the positional range of a pendulum moving about the corresponding elbow joint, the elbow joint of a pendulum about the shoulder joint, etc. At the root of the hierarchy is an abstract "joint" that is the walker centre of motion (not denoted with a marker).

Three kinds of stimuli were used in the study. (1) was an "on-joint" stimulus consisting of markers primarily at a selection of the walker joints: shoulders, hips, wrists, ankles, and one other marker mid-head. (2) was an "off-joint" stimulus consisting of those markers but with hip markers migrated down to mid-thighs, shoulder markers down to mid-upper arms, wrist markers up to mid-forearms, and ankle markers up to mid-shanks. Finally, (3) was a scrambled stimulus consisting of the first stimulus but with the centres of motion of all markers scrambled across the screen. Cutting used a coding language where each element was a vector representing pendulum-like motion of a point relative to a centre, where that centre could in turn exhibit relative motion represented by another element, etc. Using this coding language, stimuli can be represented as, for each marker, the vector sum of its element and those of a sequence of joints down the hierarchy from it to the walker centre of motion. Thus, the overall motion of a wrist marker in an on-joint stimulus can be represented as $M(W)+M(E)+M(S)+M(C)$ or of a shoulder marker as $M(S)+M(C)$, where $M(W)$ is the element representing the

relative motion of the wrist joint (with marker), $M(E)$ of the elbow joint (without marker), $M(S)$ of the shoulder joint (with marker), and $M(C)$ of the walker centre of motion (without marker). Similarly, the overall motion of a forearm marker in an off-joint stimulus can be represented as $\frac{M(W)}{2} + M(E) + M(S) + M(C)$ or of an upper arm marker as $\frac{M(E)}{2} + M(S) + M(C)$. The coding language allowed representations to be shortened by taking advantage of sequences of elements common between markers. Thus, the overall motion of the wrist marker in the on-joint stimulus above can be more simply represented as $M(W) + M(E) + overall\,motion\,shoulder\,marker$. A simpler representation of the overall motion of the forearm marker in the off-joint stimulus is likewise possible but not as short as for the wrist marker because there are fewer elements that are common between it and the upper arm marker representation. The simplest representation that can be constructed for an analogous scrambled stimulus marker is longer still. Altogether, representations of stimuli are simplest for on-joint stimuli, less simple for off-joint, and least simple for scrambled. On the behavioural side of the coin, subjects were less successful at a gender discrimination task with off-joint walkers than with on-joint walkers, and of course even less successful still with scrambled walkers (which they could not identify as walkers at all).

Note the subtle difference between this and traditional coding theory studies: whereas traditional studies compare the simplicities of alternate perceptual interpretations within stimuli, Cutting's coding theory study compared the simplicities of alternate stimuli themselves, showing that the simplest stimuli afforded better performance in a perceptual task. He took this approach perhaps due to the difficulty of clearly associating

different well-defined interpretations with different representations of walkers due to their complexity and perhaps also the special social salience of human biological motion. In any event, technically this study did not search for a simplicity principle per se.

1.4. Computational experiment: Cartesian coordinates of markers-based vs. Euler angles and translations of segments-based representation schemes

In the present study we apply coding theory to marker-based human walking motion. Our stimuli derive from digital motion capture. We compare two alternate representation schemes for the class of walkers (in coding theory parlance these comprise our coding language). The first is based on Cartesian coordinates of markers (*cm-based*), the second on Euler angles and translations of segments (*ets-based*). In keeping with standard coding theory practice, our study is comprised of a computational part where we compare how efficiently these two schemes encode walker data from an informational point of view, followed by a behavioural part where subjects' perception of walkers is tested to see if it corresponds to the scheme found to be the most efficient as per a simplicity principle.

Before proceeding further, it's worth taking a moment to underline the distinction between representation and representation scheme as used throughout this study. A *representation* is a realized surrogate for a specific object or item of information – in this study a specific walker. A *representation scheme* for a class of objects specifies everything necessary to translate those objects into representations of a specific kind as well as to interpret those representations. To give an everyday example, a JPEG or bitmap of a specific picture would both be representations of that picture, while the JPEG

and bitmap formats in general would be representation schemes. In this study the single word "scheme" will always refer to "representation scheme".

Our two main schemes of interest represent walkers as vectors and are analogously constructed by principal components analysis (PCA). PCA is a standard method for optimally reducing the dimensionality (representation length) of a set of data vectors relative to error acquired by compression across the set. It facilitates a change of origin to the mean vector of the data set followed by a change of orthonormal basis such that the new basis vectors, or principal components (PC's), successively capture as much of the variance of the mean-centred data as possible. PCA calculates the set of PC's as eigenvectors of the covariance matrix of the pre-PCA vectors, where ordering by eigenvalue corresponds to ordering by variance captured. They are in the same basis as the mean-centred pre-PCA vectors and thus are similarly interpretable – for example, PC's of a set of vectors representing faces resemble faces when visualized and have been referred to as "eigenfaces" (Turk & Pentland 1991, Vetter & Troje 1997), and PC's of a set of vectors representing walkers resemble walkers and have been referred to as "eigenwalkers" (Troje 2002). The number of PC's returned is the minimum necessary to span the mean-centred data, which is the lesser of its pre-PCA dimensionality and the number of data vectors -1 (the one dimension apparently lost corresponds to the component of the subtracted mean vector orthogonal to the hyperplane containing the vectors, and is retained in the mean vector). If those data span only a subspace of the pre-PCA space then dimensionality can be straightforwardly reduced with no error. Dimensionality can be reduced beyond this point by retaining only a subset of PC's from the first upwards, with the fewer the PC's the less the representation length but typically

the greater the error acquired. That said, this error can be small if the variance of the vectors is clustered mainly along the retained subset. Vectors not in the pre-PCA data set can benefit from the same dimensionality reduction by mean-centring using the mean of the data set, projection onto the space spanned by the retained PC's, and a change to the PC basis. Here the pre-PCA data can be thought of as a training set for the PC-based representation of novel vectors. If novel mean-centred vectors are already in the space spanned by the retained PC's, there will be no error in such a projection; otherwise, the amount of error acquired will depend on how appropriate the training set is relative to the new vectors – i.e. how well the training vectors span the class to which the new vectors belong. Refer to most textbooks on multivariate statistics or the standard book by Jolliffe (2002) for further reading. Applied in our study, our two PC-based schemes express walker vectors in a PC basis calculated from a training set of vectors representing 100 motion captured subjects. The difference between the two lies in a cm- versus ets-based pre-PCA scheme.

Any vectors in a PC-based scheme lie in the span of the mean-centred training vectors but may be projections onto and/or within that space due to exclusion from the training set and/or truncation of the PC basis respectively, and if so have acquired some error. One question that can be asked is do such displaced vectors at least still represent legitimate members of the class the scheme is representing? In other words, is the acquired error always within the class, or could it be outside? For example, does a displaced walker vector at least visualize as *some* walker, or not as a walker at all? The former seems more desirable and indeed has been found to be associated with lower error vs. number of retained PC's – e.g. for face images in Vetter & Troje (1997). A sufficient

condition for it is that the scheme be morphable, meaning all vectors in the space and not just those that are direct exemplars of the class correspond to members of the class, at least within a convex statistical boundary. One way to think of a morphable scheme is a "compact" mapping to the space where no points fail to be valid representations. Morphability not only makes for flexible PC-based schemes but also allows for meaningful interpolation, extrapolation, and averaging of representations by arbitrary linear combination (hence the term morphable), as well as generation of artificial objects in the class as arbitrary points.

A striking visual example of the impact of a morphable versus non-morphable scheme was shown in the context of faces in Vetter & Troje (1997). PC-based representations of arbitrary faces in a non-morphable scheme visualized as just pixel-by-pixel superpositions of eigenface basis vectors and appeared blurry and ghost-like, whereas representations in a morphable scheme visualized as eigenfaces morphed together at the level of coherent features and, while typically still in error with respect to original information, at least looked like faces.

A morphable PC-based scheme follows directly as a result of a morphable pre-PCA scheme since the relevant topology of the space remains unaltered by PCA. Essentially, the critical work of establishing a mapping between objects and vectors that *linearizes* the class, meaning representing them along a finite number of linear dimensions in a morphable fashion, occurs at the pre-PCA step. PCA then serves only to optimize representation by allowing length to be pruned at minimal cost in error. That said, even granted morphability, not all pre-PCA schemes are created equal. Better ones give representations such that as much variance is captured along as few dimensions as

possible. The quality of linearization achieved in a pre-PCA scheme is reflected in how efficiently the representations under it decompose in the subsequent PCA, where more efficient decompositions capture more variance in lower PC's such that more high PC's can be discarded relative to error acquired. This is depicted in Figure 2 below.

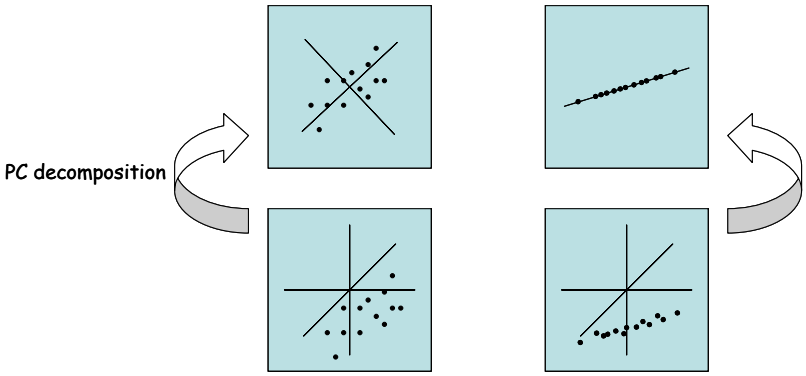

Figure 2. On the left is a depiction of a three-dimensional pre-PCA scheme and its subsequent PC-based scheme where two PC's must be retained to capture the data. On the right is a pre-PCA scheme that better linearizes the same class, mapping most of its variance to one dimension. One PC can be retained in its PC-based scheme at minimal cost in error.

In the present study our two pre-PCA schemes are both linearizing schemes and both accomplish linearization by Fourier decomposition of original time series of basic walker coordinates, developed in Troje (2002) for cm-based and extended to ets-based here. This process comes at its own cost in error with respect to basic coordinate data. With Fourier and PC decomposition applied identically across the board, any difference at our two PC-based schemes arises solely from the difference between basic coordinate schemes. On the cm-based side we have coordinates of markers in a common body-centred 3D Cartesian coordinate system as a function of time (denoted just as *cm*). On the ets-based side basic coordinates take advantage of a model of the body as an articulated system of 15 rigid segments – mostly limb segments such as upper arms, forearms,

thighs, shanks, etc. – whose motion is represented with three Euler angles and three Cartesian translation components as a function of time for each, with markers represented only by constant coordinates in the moving Cartesian coordinate systems of the respective segments to which they are attached (denoted just as *ets*). Following from the cm basic coordinate scheme, our cm-based pre-PCA scheme is denoted *cmF* (cm Fourier-based) and PC-based scheme *cmPC* (cm PC-based). The analogous schemes following from the ets basic coordinate scheme are denoted *etsF* and *etsPC*. Altogether the six schemes and their relationships to each other are depicted graphically in Figure 3 below.

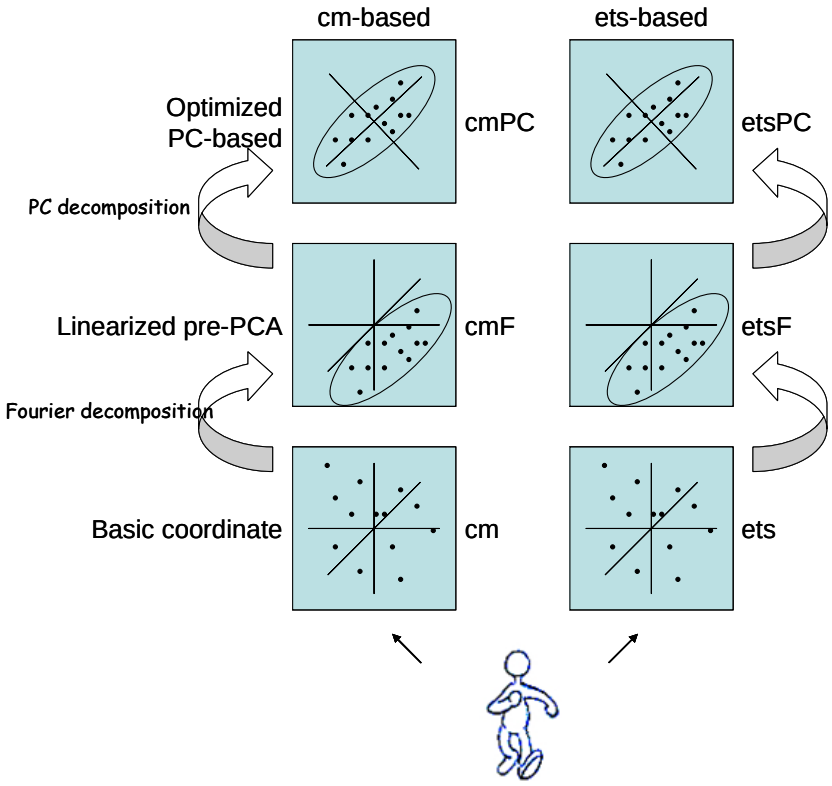

Figure 3. The six representation schemes that play roles in this study. cm- versus ets-based schemes are organized in columns. Analogous schemes for each are organized in rows. At the bottom are basic coordinate schemes cm and ets – dots depict points that represent walkers. These schemes are linearized in pre-PCA schemes cmF and etsF – ovals indicate regions of space where the schemes are morphable and all points represent walkers. Finally these schemes are optimized in PC-based schemes cmPC and etsPC – the origin is shifted to the mean of the data and redundant dimensions are discarded. It is these PC-based schemes that are of direct interest from a coding theory perspective. Typically error is acquired at both Fourier decomposition and PC decomposition steps. Three-dimensional spaces are depicted here for simplicity but actual dimensionalities are much higher.

While traditional coding theory studies look at lengths of perfect representations, in this study we allow for imperfect representations as a trade-off for lower lengths. In a generalization of representation length as well as of scheme efficiency in the information theoretic sense of the word (Section 1.2), we take scheme efficiency to be total error across lengths averaged across objects, where more efficient schemes admit less error across lengths. Here total error is the aggregate of error at Fourier decomposition and error at PC decomposition with respect to basic coordinate data and lengths range from

zero retained PC's to the full complement. Thus, in our computational experiment we come to compare not PC-based representation simplicities per se but rather PC-based scheme efficiencies. It is how the two basic coordinate schemes react to Fourier decomposition and PC decomposition that decides these efficiencies.

1.5. Behavioural experiments: the norm-based multidimensional walker space model and the caricature effect

A common criticism of coding theory has been that representations compared in studies may bear little or no resemblance to representations actually employed by the perceptual system (see for example Chater 1996, Hatfield & Epstein 1985). For example, the coding language used in Buffart, Leeuwenberg, & Restle (1981) (Section 1.2) has a certain practical appeal for manual construction of 2D line drawings by humans, even resembling the turtle graphics functionality of the classic Logo programming language, but its sequential, local element-based nature makes it a highly unlikely relative of any perceptual scheme implemented in neural architecture. If a coding language does not have enough in common with the perceptual scheme(s) it is supposed to be modeling then how representations constructed under it compare in terms of simplicity may likewise have nothing to do with the actual functioning of the perceptual system since entirely different informational redundancies may have been compressed than those compressed in the perceptual system. In fact, Hatfield & Epstein (1985) point out that logically any coding theory experiment must either presume the appropriateness of its coding language and test for the simplicity principle *or* presume the simplicity principle and test the appropriateness of its coding language, but not both. The fact that coding theorists have

returned a reasonable trend of positive results for the simplicity principle is suggestive that their choices of coding language have not been too naïve. Still, perhaps they could have been better, and perhaps negative results such as in Boselie & Wouterlood (1989) could have been avoided if they had.

With this concern in mind, the PC-based schemes whose efficiencies we compare were chosen partly because they are alternate implementations of the norm-based multidimensional object space model. This is a model of object perception that has been extensively studied in contemporary psychological literature and has proven itself robust in terms of both experimental verification and predictions. It models the high level perceptual representations of stimuli in a given class as morphable vectors in a space whose origin corresponds to the mean stimulus or "norm" from the subject's experience of the class and whose dimensions capture features along which stimuli in the class statistically vary. The dimensions can be numerous and the features and class are high level.

The object space model was first proposed in a paper by Valentine (1991) in the context of the class of faces. The paper presented much detailed theoretical reasoning supported by five experiments and a connection to some results from the field of neural networks. Valentine proposed a face space model where the origin was the mean face and dimensions captured features as broad as sex, race, emotion, facial hair, weight, etc. (though not those features in particular and not necessarily conforming to such single intuitive descriptions). The author further proposed that face space explained a number of experimental findings concerning the effects of face distinctiveness, inversion, and race

on face recognition and classification, including findings that older models such as those based on configural processing of local features could not.

The object space model has seen many applications but one that is particularly striking and will be of direct consequence in our study is the distinctiveness effect or caricature effect. The perceived identity of a stimulus, such as the identity by name of a person from her face, is modeled by the angle of its vector in the object space. In other words, any vector in a given direction corresponds to the same perceived identity, even though the physical features of the stimuli will certainly differ with vector length and thus overall coordinates. When an object is perceived the coordinates of its vector are modeled as having some margin of error. The margin of error in coordinates translates by trigonometry to a margin of error in angle. Assuming margin of error in coordinates is approximately constant with distance from the origin, margin of error in angle must decrease with distance (Figure 4). Putting this together, vectors of increasing length model stimuli that are perceived to have constant identity but whose identity is perceived more precisely. This is the distinctiveness effect and it has been shown in many recognition and identification studies, including Valentine's original paper (1991) where subjects were required to recognize familiar faces and performance was better for faces further from the norm.

The caricature effect is the distinctiveness effect applied: the perceived identity of a given stimulus can be strengthened by using an alternate stimulus whose vector has the same direction but greater length. This is a stimulus where all high level features of the original (the veridical) are exaggerated relative to the norm. Such a stimulus has been known for centuries as a caricature (from the Italian *caricare* meaning "to load" – i.e. a

caricature is a loaded depiction). Examples of caricatures are perhaps most familiar to us in the form of artists' caricatures in popular and political culture such as in Figure 5 below, but caricatures can also be computationally generated in implementations of object spaces (such as our cmPC and etsPC schemes) simply by scaling out vectors containing veridical stimulus data in the lab, where multiplier is taken as *caricature level*. "Anticaricatures" can also be generated as vectors of decreased length and are correspondingly less precisely identifiable.

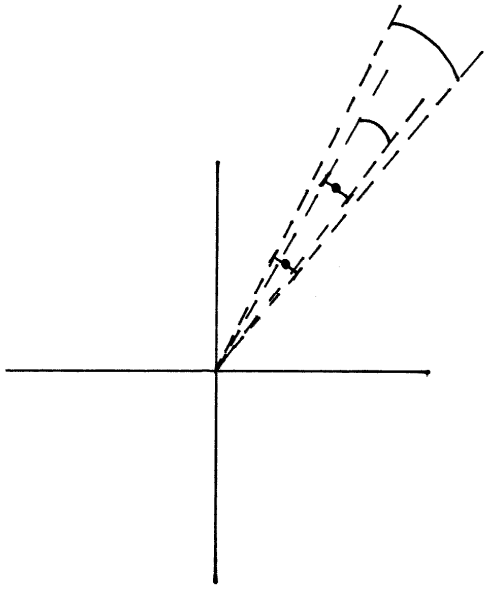

Figure 4. A veridical stimulus and a caricature of that stimulus as vectors in a 2D depiction of an object space. Margin of error in coordinates is denoted by error bars and resultant margin of error in angle, which corresponds to margin of error in perceived identity, is denoted by dashed lines. The greater distance of the caricature from the norm results in greater precision in perceived identity.

Figure 5. A veridical and an artist's caricature from popular culture (caricature reproduced with permission)

The caricature effect has been shown in numerous caricature identification studies where subjects are asked to identify caricatures of familiar stimuli and performance is typically found to be better than for veridicals. Refer for example to the original Rhodes et al. (1987), Frowd et al. (2007) that employed the caricature effect to increase identification performance in the context of facial composites in law enforcement, and Byatt & Rhodes (1998) that also investigated the effect of generating caricatures relative to different face norms appropriate for different races.

The object space model has been most extensively study for the class of faces but a body of literature also exists wherein it has been applied to walkers. Here we have a "walker space" whose origin is the mean walker and whose dimensions capture high level features that are both structural and kinematic along which walkers statistically vary. Analogous results have emerged. For example, Giese et al. (2008) showed an isomorphism between true perceptual walker space and an implemented space similar to our PC-based spaces as measured by subjects' judgments of similarity under a multi-dimensional scaling analysis (MDS). Troje (2002) used our cmPC scheme along with

subjects' ratings of walker gender to create a linear discriminant classifier that calculates a single "gender axis" in that walker space upon which increasing projections of walker vectors on one side of the origin maximally correspond to increasing perceived masculinity and on the other side to femininity, with the mean walker perceived as gender-neutral (and of course neutral in most other attributes). Troje (2008) generalizes this technique to numerous other features such as affect, weight, and even features as creative as voting behaviour and socioeconomic class. And although not using walking motion, Hill & Pollick (2000) showed a caricature effect in biological motion in the domain of PLD's of arms engaged in drinking action, and Pollick et al. (2000) showed a caricature effect in the domain of textured body depictions of tennis serve actions.

In the behavioural part of our study subjects' perception of walkers is tested to see if it corresponds to the most efficient scheme measured in the computational part as per a simplicity principle. As in Cutting (1981), it is difficult if not impossible to associate different well-defined perceptual interpretations with different representations of walkers as standard coding theory practice would have. An alternative approach was suggested by the connection between our schemes and the walker space model, which opened the door to all of the experimental paradigms that object space models have become established for. Our design became to use the caricature effect, comparing caricatures generated in the cmPC versus etsPC scheme at the same caricature level in the hope of finding that caricatures generated in one benefit identification more than caricatures generated in the other. The basic idea is that, as alternate models of the true high level perceptual scheme for walkers, one of our two schemes is probably a more accurate model than the other. A caricature generated in the more accurate model is expected to map to a point in the true

perceptual space whose angle is closer to the angle of its veridical than a caricature from the same veridical but generated in the less accurate model, thus more accurately retaining identity and affording greater identification performance (Figure 6).

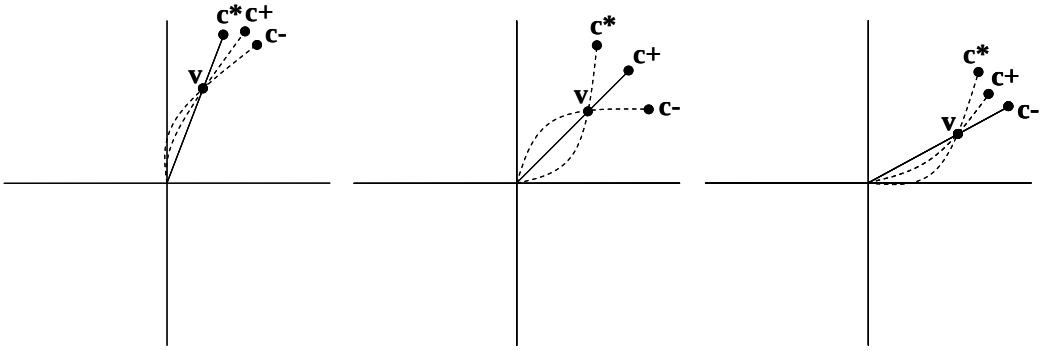

Figure 6. On the left is a 2D depiction of the true high level perceptual walker space (*), in the middle a depiction of whichever is the more accurate of our two models (+), and on the right the less accurate of our two models (-). **v** ***marks the point in each space corresponding to a common veridical walker.*** **c*** ***marks points corresponding to an optimal caricature arising from linear scaling in the true space,*** **c+** ***to a caricature generated in the more accurate model, and*** **c-** ***to a caricature generated in the less accurate model.*** **c+** ***should afford better identification performance than*** **c-** ***since its direction in the true space is closer to the direction of the veridical. Dotted lines denote the non-linear mappings of straight lines from one of the other spaces.***

If a simplicity principle holds for walker perception then we expect the scheme with greater perceptual affordance and thus the closer model of the true perceptual scheme to also be the more efficient as measured in the computational part of our study.

Our work consists specifically of one computational experiment and three behavioural experiments. The computational experiment compares cmPC versus etsPC scheme efficiencies. The first two behavioural experiments are preparatory for our originally planned comparison of caricatures. In the first behavioural experiment we measure just noticeable differences (JND's) along each of ten retained dimensions in cmPC space for use in equalizing distinctiveness of veridical walkers and reducing

variance when averaging identification performance across different caricatures in following experiments. In the second experiment we measure identification performance for cmPC caricatures across a range of caricature levels to determine the level that gives optimal performance, with other levels being either not enough or too extreme. Unfortunately, we then find that before any difference can be perceived between analogous caricatures in cmPC versus etsPC schemes it is necessary to use caricature levels far in excess of the measured optimal level and so extreme as to be counterproductive. Our third behavioural experiment verifies that subjects are unable to distinguish between cmPC and etsPC caricatures when generated from common veridicals at the optimal level. As such, it was not possible to run our originally planned main behavioural experiment seeking a difference in identification benefit from cmPC versus etsPC caricatures.

CHAPTER 2. GENERAL METHODS

2.1. Motion capturing the RUB100 walker data set

Walker data were derived from a database of 100 subjects – 50 males and 50 females – motion captured at Ruhr University Bochum, Germany. The database is denoted *RUB100*. 41 retroreflective markers were attached to the body using a modified version of the Helen Hayes marker set (Davis III et al. 1991) and positions were recorded in real time in the world coordinate system – a 3D Cartesian coordinate system rigidly attached to the space subjects walked in – using a surrounding array of 120 Hz cameras (Vicon 512, Oxford Metrics, Oxford, UK). x and y axes were set along the floor with z axis pointing up. Most markers were taped directly to the skin. Markers on the head, wrists, and ankles were fixed to elastic bracelets, and markers on the feet were taped to tight-fitting dancer's shoes. A number (typically two to four) of separate walking instances were recorded for each subject on the open floor, each typically consisting of two to four gait cycles.

Each marker is modeled as attached to one of 15 rigid body segments. Positions at each frame of animation for 25 further "virtual markers" were calculated using the data of the 36 physical markers and auxiliary body measurements (width of the knees, elbows, ankles, etc.). Unlike physical markers, virtual markers can be located inside the body of the walking subject. Some virtual markers simply fill space between given physical markers and are calculated straightforwardly using the coordinates of those physical markers. Fifteen special virtual markers represent body segment origins, most of which model joint centres inside the body such as at the shoulders, elbows, hips, etc. Body segments and their origins will be explained in detail in Section 2.54 on the ets scheme.

All virtual markers were calculated by the Visual3D software package (C-Motion Inc., Germantown, MD, USA). Once virtual markers are calculated, there is no practical difference between them and physical markers in walker data. For this study, after virtual markers had been established, 61 physical and virtual markers were retained.

2.2. Marker sets

Four marker sets were used in this study for various purposes (Figure 7). The *61-marker set* consisted of all 61 markers. The remaining three sets were subsets of the 61-marker set. The *45-marker set* consisted of the respective origin marker and two other markers per segment. This set contains the minimum three markers per segment needed to calculate an ets representation from a cm (Section 2.54) and thus the minimum for which cm-based and ets-based representations carry identical information. Accordingly, it was used for comparison of scheme efficiencies in the computational experiment. The *15-marker set* consisted of 15 virtual markers, one per segment, where twelve of those 15 were the origin markers at the shoulder, elbow, wrist, hip, knee, and ankle joints, and the remaining three were calculated from the remaining head, thorax, and pelvis origin markers by shifting for more intuitive visualization of those segments. The 15-marker set is designed to give good minimal PLD's for behavioural work and is the standard set used in our lab for most studies. Finally, the *15clav-marker set* was a slight modification of the 15-marker set consisting of substituting a mid-clavicle marker for the one in the vicinity of the sternum. In our behavioural experiments we used both PLD and stick figure marker-based stimuli. In the latter lines are drawn to connect certain markers reminiscent of body segments. For stick figures we used the 15clav-marker set because it

gives a more intuitive visualization in that format – the lines connecting the new marker to the shoulder markers mimic the natural slopes of the clavicle bones as opposed to the upward slopes that would be seen from the sternum marker in the 15-marker set.

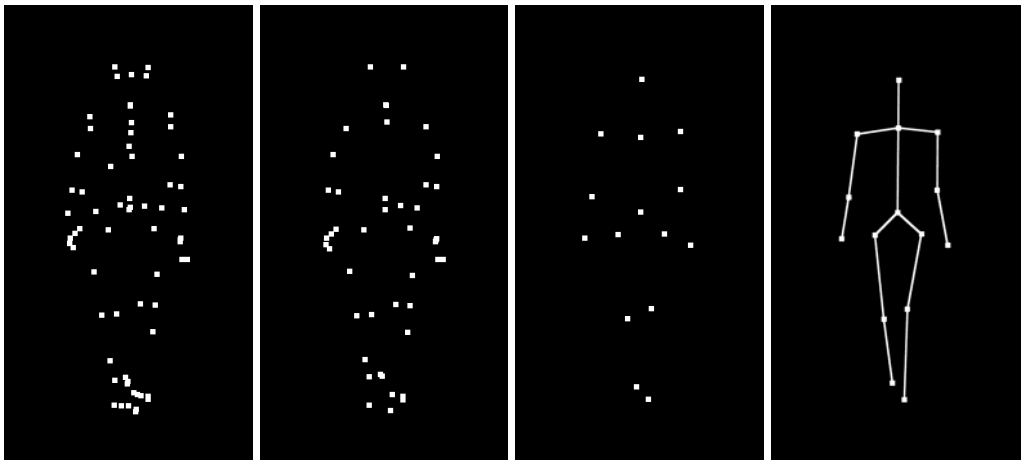

Figure 7. Example female walker as a 61-marker PLD, 45-marker PLD, 15-marker PLD, and 15clav-marker stick figure. Each is at a different random point in the gait cycle.

2.3. Additional pre-processing

After 61-marker data was established for walkers several further processing tasks were run to finalize the raw data. Walkers were rotated about the *z* axis such that their walking direction was the positive *x* axis. Specifically, the angle of rotation was calculated as the angle between the line fitted to the smoothed frames of walker positions in the *xy* plane and the *x* axis. Walker translation was removed by subtracting from the *x,y* position of each marker at each frame the mean *x,y* position across markers at that frame, implementing a virtual treadmill effect and centring the walker at $x = 0$. Walkers were also centred vertically at $z = 0$ by subtracting from the *z* position of each marker at each frame the mean *z* position across markers and frames. (Note the subtle difference between these two operations.) Altogether these were equivalent to a transformation from a lab

coordinate system to a body-centred coordinate system rigidly attached to the walker. Each walker was then size-normalized such that the mean across coordinates of the mean coordinates across frames, divided by the corresponding mean further averaged across walkers, equaled one. Unless otherwise stated for the vertical centring and size normalization operations the mean across the largest available integer number of gait cycles beginning at the first frame was used, where the precise number of frames was determined as for the cmF scheme explained later in Section 2.52. For all three operations, the 15-marker set was used for calculations to prevent the fact that some body segments have a higher density of markers than others in the 61-marker set from skewing the means. However, the results were always applied to all 61 markers.

2.4. Representation length

Our walker representations are all matrices or vectors and we measure length as number of elements (in the mathematical sense). As per the standard approach in coding theory, we include in a representation all information that is necessary to encode the specific walker being represented – i.e. all information distinctive of that walker. Information that is common across all walkers is considered part of an encoder/decoder for the scheme and not part of a representation itself. Note that PLD and stick figure stimuli therefore have identical representations since the only information needed to encode sticks are which pairs of markers to be joined, and this is common across walkers.

2.5. Walker representation schemes

cm-based

2.51. cm (Cartesian coordinates of markers)

cm is our first basic coordinate scheme and also our basic motion capture data format. It encodes a subject's walk over a finite period of time as *x,y,z* coordinates (mm) for each marker for each frame of animation in the body-centred coordinate system. Representation length is thus $3 \times n_m \times n_f$ where n_m is the number of markers and n_f is the number of frames. For purposes of the cmF scheme to follow it will be convenient to express a cm walker as a matrix **w_cm** where each column or frame vector contains all coordinates of all markers at a given frame and each row or coordinate time series vector contains a given coordinate of a given marker across frames.

2.52. cmF (cm Fourier-based)

All Cartesian marker coordinates in a cm representation change smoothly and, since the motion is repetitive walking, approximately periodically with time. This suggests that the relationship of marker coordinates to time might be efficiently captured in a Fourier series. A cmF representation of a walker is the expansion of a cm representation as a Fourier series of sines and cosines for each cm coordinate time series, with harmonic frequencies common across coordinates. The exact method used is a change for the cm coordinate time series vectors from the discrete time basis (time measured in frames) to a basis of sine and cosine time series vectors at frequencies consisting of the gait cycle frequency and its harmonics:

$$\mathbf{w\_cm}_F = \mathbf{w\_cm} \cdot \left[\mathbf{1}, \sin\left(\frac{2\pi}{T}\mathbf{t}\right), \cos\left(\frac{2\pi}{T}\mathbf{t}\right), \sin\left(2\frac{2\pi}{T}\mathbf{t}\right), \cos\left(2\frac{2\pi}{T}\mathbf{t}\right), \ldots \right]^{\mathrm{T}\times} \quad (2)$$

where the F subscript denotes Fourier basis, $\mathbf{1}$ is a column vector of n_f ones, the fundamental period T is the gait period (in frames), $\mathbf{t} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ n_f \end{bmatrix}$, a function of a vector argument denotes a vector whose components are that function at the respective elements of the argument, and the * superscript denotes the Moore-Penrose pseudoinverse giving a least squares/minimum norm solution. Components along the first basis vector give the zeroth harmonic for each cm coordinate (these correspond to mean coordinates), along the second and third the sine and cosine components respectively of the first harmonics (fundamentals), along the fourth and fifth the sine and cosine components of the second, etc. We refer to the vector of zeroth harmonics (the first column of $\mathbf{w\_cm}_F$) simply as the zeroth harmonic of the cmF walker, the vectors of sine and cosine components of the first harmonics (the second and third columns of $\mathbf{w\_cm}_F$) simply as the first harmonic, etc.

Unless otherwise stated representations are calculated across all available frames and T is taken as the average period of the harmonics with maximum amplitude from the discrete Fourier transforms (DFT's) of a subset of cm coordinates with particularly large and reliable periodic behaviour zero-padded to a window size of 8192. For these we use the x coordinates of the two foot origin markers (capturing their front-to-back motion). This gives a resolution for T of about 2 frames. The non-integer number of gait cycles comprised by the frames and the estimated T both result in Fourier basis vectors that are not exactly orthogonal and do not exactly span the space of time series vectors,

necessitating the pseudoinverse. Optionally, if exactly one gait cycle of cm data is taken and $T = n_f$, then equation 2 reduces to

$$\mathbf{w\_cm}_F = \mathbf{w\_cm} \cdot \left[\frac{\mathbf{1}}{n_f}, \frac{\sin\left(\frac{2\pi}{n_f}\mathbf{t}\right)}{n_f \div 2}, \frac{\cos\left(\frac{2\pi}{n_f}\mathbf{t}\right)}{n_f \div 2}, \frac{\sin\left(2\frac{2\pi}{n_f}\mathbf{t}\right)}{n_f \div 2}, \frac{\cos\left(2\frac{2\pi}{n_f}\mathbf{t}\right)}{n_f \div 2}, \ldots \right] \quad (3)$$

Here the Fourier basis vectors are orthonormal, giving a lossless transformation of the walker with $\frac{n_f - 1}{2}$ (n_f odd) harmonics or $\frac{n_f}{2}$ (n_f even) harmonics (in the latter case the sine component of the final harmonic vanishes and the basis vector for the cosine component of the final harmonic is normalized by n_f rather than $n_f \div 2$). The process is then equivalent to a DFT of each time series vector.

Because the courses of the cm coordinates are smooth and approximately periodic, the majority of power is concentrated in the first few harmonics. A subset of harmonics from $1 \ldots n_h$ can be retained, with the fewer harmonics typically the greater the error acquired at Fourier decomposition. A cmF representation thus consists of the $3 \times n_m \times \left(2 \times n_h + 1\right)$ elements in the $\mathbf{w\_cm}_F$ matrix plus the single element *T*. Unless otherwise stated two harmonics are retained, which is sufficient for a reasonably low error reconstruction of an original cm walker and is the optimal number for cmPC scheme efficiency (Section 3).

A cm walker can be reconstructed without further loss of information by the reverse transformation:

$$\mathbf{w\_cm}' = \mathbf{w\_cm}_F \cdot \left[\mathbf{1}, \sin\left(\frac{2\pi}{T}\mathbf{t}\right), \cos\left(\frac{2\pi}{T}\mathbf{t}\right), \sin\left(2\frac{2\pi}{T}\mathbf{t}\right), \cos\left(2\frac{2\pi}{T}\mathbf{t}\right), \ldots \right]^{\mathrm{T}} \quad (4)$$

where prime is used to indicate the reconstruction because it will typically be different from the original **w_cm**. This error is partly due to discarded harmonics. However, there is another contributing factor. Because the fundamental period is taken to be the gait period, the cmF scheme encodes a kind of average gait cycle across cycles in the original cm data. Whereas every gait cycle in cm data will typically be slightly different, every gait cycle reconstructed from a cmF representation will be identical (and slightly different from each of the original cm gait cycles). This contribution to error can actually be seen as beneficial in some circumstances: a cmF representation can be thought of as a representation of a walker in the ideal sense, obtained from a sample of real "imperfect" gait cycles in original cm data. Moreover, the ideal gait cycle can be displayed at any point in time, including between original frame times, simply by using any value for *t* in equation 4. The gait cycle can also be displayed indefinitely and is always smoothly repeating, in contrast to the discontinuity that generally results from looping cm data.

The purpose of the cmF scheme is to linearize cm data. The cm scheme is not morphable because any coordinate in **w_cm** matrices does not have a coherent interpretation across walkers because its frame vectors do not correspond to the same point in the gait cycle. This is due both to the fact that the initial frames may correspond to different points in the gait cycle and due to the fact that different walkers walk at different speeds. In a Fourier-based representation the latter is entirely accounted for by the single coordinate capturing fundamental period, which *is* coherently interpreted across walkers. The former is entirely accounted for by the phases of the fundamental harmonic of the walker, implicit in its sine and cosine elements. Unless otherwise stated we shift the phases of cmF walkers so that at time zero all walkers are aligned at the same

point in the gait cycle. That is, if we express the Fourier series for any given cm coordinate *c* of a given walker in terms of sine amplitudes/phases,

$$\mathbf{w\_cm}'_{(c)} = a_{0(c)} + a_{1(c)} \sin\left(\frac{2\pi}{T}\,\mathbf{t} + p_{1(c)}\right) + a_{2(c)} \sin\left(2\,\frac{2\pi}{T}\,\mathbf{t} + p_{2(c)}\right) + \ldots \qquad (5)$$

where coefficients *a* and phases *p* are both specific to cm coordinate, then we shift the representation of the walker by some phase angle φ constant across harmonics and cm coordinates to bring the walker into alignment:

$$\mathbf{w\_cm}'_{(c)} \rightarrow a_{0(c)} + a_{1(c)} \sin\left(\frac{2\pi}{T}\,\mathbf{t} + p_{1(c)} - \varphi\right) + a_{2(c)} \sin\left(2\,\frac{2\pi}{T}\,\mathbf{t} + p_{2(c)} - 2\varphi\right) + \ldots \quad (6)$$

where the addition of a scalar and a vector denotes the addition of the scalar to each element of the vector. The equivalent transformation for the sine/cosine amplitudes that actually constitute the cmF representation follows from basic trigonometric identities. We calculate φ for each walker such that the circular mean of the fundamental phase angle of the *x* coordinate of the left foot origin marker and the fundamental phase angle of the right foot origin marker $+\pi$ is shifted to $\frac{\pi}{2}$, giving the left foot at about $\frac{\pi}{2}$ (all the way forward) and the right at about $-\frac{\pi}{2}$ (all the way back). Refer to Mardia (1972) on the circular mean for angular data. Once walkers are phase-aligned, all cmF coordinates become coherently interpreted across them – namely, describing the same coordinate of the same marker in the same sine or cosine component of a harmonic that models the same temporal aspect of gait cycle. The scheme thus becomes morphable, allowing it to function as the pre-PCA scheme for cmPC, outlined next. Note that for a cm

reconstruction from a cmF walker to be as close to the original cm data as possible phase must be unshifted, or at least phase shift reversed at reconstruction.

Unless otherwise stated, cmF RUB100 walkers were each the mean of the cmF representations of the several motion captured walks for the respective subject. For purposes of the cmPC scheme it will be convenient to express a cmF walker as a column vector $\mathbf{w\_cmF}$ consisting of the elements of the $\mathbf{w\_cm}_F$ matrix along with the gait cycle period *T*.

2.53. cmPC (cm principal component-based)

cmPC is our first PC-based scheme. PCA is computed across training walkers in the morphable cmF scheme to optimize scheme efficiency. Our PCA process includes a whitening step between mean-centring and computation of PC's where each coordinate is divided by its standard deviation across training walkers (provided it is non-zero). This is to accommodate for the fact that not all coordinates have a common unit of measurement. The PC's, or eigenwalkers, are equivalent to mean-centred and whitened cmF walker vectors. The number of PC's returned is the lesser of the cmF dimensionality $= 3 \times n_m \times \left(2 \times n_h + 1\right) + 1$ and the number of training walkers $- 1$. Any cmF walker vector (with the same numbers of markers and harmonics) can then be mean-centred, whitened, projected onto the walker space spanned by the PC's, and expressed using them as a basis:

$$\mathbf{w\_cmPC} = \mathbf{cmPC}^T \cdot \mathbf{S}^T \left(\mathbf{w\_cmF} - \overline{\mathbf{w\_cmF}} \right) \tag{7}$$

where **cmPC** is a matrix containing the PC's in whitened cmF coordinate basis in columns, $\overline{\mathbf{w\_cmF}}$ is the mean cmF training walker vector, **S** is a diagonal matrix containing the standard deviations of the cmF coordinates across training walkers in successive elements, and **w_cmPC** is the resulting cmPC walker representation.

A subset of PC's can be retained from the first upwards, with the fewer PC's the less the representation length but typically the greater the error acquired. Because the **cmPC** and **S** matrices and the $\overline{\mathbf{w\_cmF}}$ vector are general to all walkers, a cmPC representation consists only of the **w_cmPC** vector. Representation length is thus equal to the number of retained PC's and is typically tiny compared to cm length. Representation quality critically relies upon a large or at least well-chosen training set relative to the walker being represented and an adequate number of retained PC's, but if quality is good then representation efficiency is dramatically improved.

A reconstructed cmF representation **w_cmF'** can be obtained from a cmPC representation without further loss of information starting with a change from PC basis back to whitened cmF coordinate basis followed by unwhitening and adding back the mean cmF training walker vector:

$$\mathbf{w\_cmF'} = \mathbf{S} \cdot \mathbf{cmPC} \cdot \mathbf{w\_cmPC} + \overline{\mathbf{w\_cmF}} \tag{8}$$

Unless otherwise stated the cmF RUB100 walkers comprised the training population and we retained 10 PC's, which is sufficient for reasonable reconstructions (Section 3).

ets-based

2.54. ets (Euler angles and translations of segments)

ets is our second basic coordinate scheme. It is analogous to the cm scheme in that it consists of a set of vectors for each frame of animation. However, it is a departure from the cm framework computed by a non-linear transformation of cm range (time-dependent) data. The ets scheme is designed to more efficiently represent body motion than the cm by taking advantage of a model of the body as an approximately articulated system of 15 rigid segments. Each segment is represented by a 3D Cartesian coordinate system that moves with time. "Approximately articulated" means that each segment is rigidly attached to the next segment deeper in the hierarchy with only small variance or play at the attachment, or equivalently that the origin of a segment coordinate system can be expressed with almost constant coordinates over time in the coordinate system of the segment it is attached to. The latter segment is the "parent" of the former and the former the "child" of the latter. Each segment has one parent, with the body-centred/cm space acting as parent to the root segment of the hierarchy. Each segment with the exceptions of the segments at the ends of the hierarchy has one or more children. In our model we conform to the standard practice of using the pelvis segment as the root and the head, hands, and feet as the ends.

Each marker is attached to one of the segments and is represented only by constant coordinates in the moving coordinate system of that segment. ets segments and their relationship to markers are depicted in Figure 8 below. There is some degree of approximation to rigid marker attachment, and this is the only potential source of error in an ets representation relative to a cm. It is a very small one – we measured it to be about

3% relative RMS error for ets walkers in the 61-marker set as measured by Section 3 equation 22. Furthermore, while a legitimate part of the error is a product of non-rigidity of segments (mostly the thorax segment) and play in marker attachment due to movement of flesh about bones, another part is only a consequence of movement of clothing to which markers are attached, which amounts to error in measurement and not in the model itself. Thus, cm and ets schemes can be approximately regarded as equivalent basic coordinate schemes for walkers.

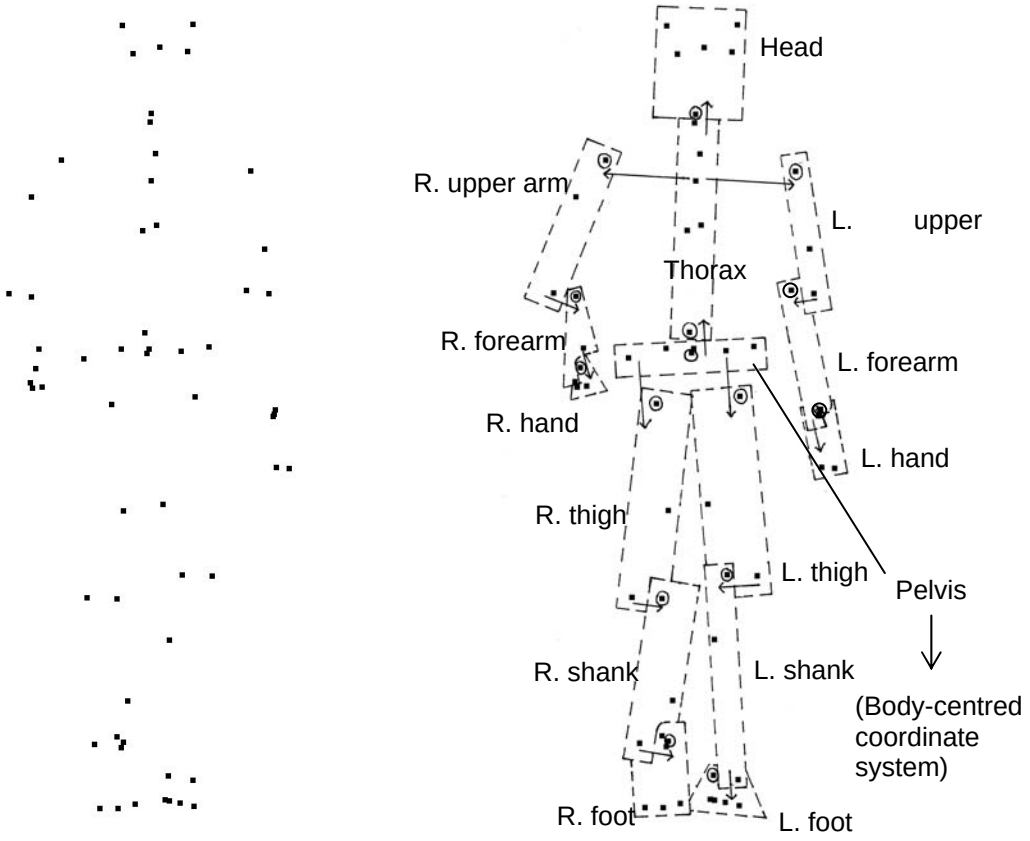

Figure 8. On the right is a 61-marker PLD at a random point in the gait cycle and on the left the same PLD with body segments outlined according to the markers on each segment. Arrows depict the segment hierarchy rooted at the pelvis. Markers denoted in circles are segment origin markers (explained below).

As 3D coordinate systems, segments must be specified by both 3D position (origin) and 3D orientation (basis vectors) – altogether pose – in a six degrees of freedom (6DOF) model. The position and orientation of each segment at each frame are represented respectively as the Cartesian translation from segment origin to parent origin and the rotation from segment basis to parent basis necessary to transform any coordinates in the segment coordinate system to coordinates in its parent coordinate system. More detail will be explained momentarily, but for now consider the example of a forearm segment with origin at the elbow joint, with upper arm parent segment with origin at the shoulder joint. Translation physically corresponds to the displacement from shoulder to elbow and, insofar as it varies with time, play in the elbow joint, and rotation corresponds to elbow joint orientation. Length of the upper arm can be inferred directly from translation.

A cm representation is the starting point for the mathematical details of an ets representation. A minimum of three markers per segment are required; in this study the 45-marker set is used for this purpose. One marker **o** defines the origin of a segment coordinate system; in this study the 15 origin markers are virtual markers calculated by Visual3D (Section 2.1) to optimize the rigid marker attachment approximation. The segment coordinate system basis vectors are taken such that the mean orientation of the segment across frames is equal to the mean orientation of its parent, meaning on average the segment coordinate system is aligned with the parent coordinate system. Bases are calculated beginning with the root pelvis segment, which we mean-centre to the body-centred/cm basis vectors chosen by convention with z up and walking motion in the x direction as per Sections 2.1 and 2.3 – thus, parent basis vectors are always

unambiguously established. The second marker "reference marker 1" **ref1** is used to define a preliminary *x* basis vector, the third "reference marker 2" **ref2** is used to define a preliminary *y* basis vector, and a preliminary *z* basis vector is defined using the preliminary *x* and *y* basis vectors. A constant rotation is then applied to these basis vectors at all frames to rotationally mean-centre them with respect to the parent basis. For a segment *s* at a given frame of animation:

$$\begin{cases} \hat{\mathbf{x}}_{s,b} = \dfrac{\mathbf{o}_{s,b} - \mathbf{ref1}_{s,b}}{\left\| \mathbf{o}_{s,b} - \mathbf{ref1}_{s,b} \right\|} \\ \hat{\mathbf{y}}_{s,b} = \dfrac{\left(\mathbf{o}_{s,b} - \mathbf{ref2}_{s,b}\right) - \left(\mathbf{o}_{s,b} - \mathbf{ref2}_{s,b}\right) \cdot \hat{\mathbf{x}}_{s,b}}{\left\| \left(\mathbf{o}_{s,b} - \mathbf{ref2}_{s,b}\right) - \left(\mathbf{o}_{s,b} - \mathbf{ref2}_{s,b}\right) \cdot \hat{\mathbf{x}}_{s,b} \right\|} \\ \hat{\mathbf{z}}_{s,b} = \hat{\mathbf{x}}_{s,b} \times \hat{\mathbf{y}}_{s,b} \end{cases} \tag{9}$$

$$\left[\hat{\mathbf{x}}_{s,b} \;\; \hat{\mathbf{y}}_{s,b} \;\; \hat{\mathbf{z}}_{s,b} \right] \rightarrow \overline{\mathbf{R}_{s \rightarrow p,b}} \cdot \left[\hat{\mathbf{x}}_{s,b} \;\; \hat{\mathbf{y}}_{s,b} \;\; \hat{\mathbf{z}}_{s,b} \right] \tag{10}$$

The first subscripts before commas indicate that the vectors pertain to the segment *s* under discussion (here this is redundant but in later equations will contrast with parent *p*, so it's used here for consistency). The second subscripts after commas indicate that the vectors are in body-centred basis (in later equations this will contrast with segment *s* basis or parent *p* basis). $\overline{\mathbf{R}_{s \rightarrow p,w}}$ is the mean across frames of the matrices that rotate segment basis to parent basis, all in the body-centred basis, which are calculated at each frame as:

$$\mathbf{R}_{s \rightarrow p,b} \cdot \left[\hat{\mathbf{x}}_{s,b} \;\; \hat{\mathbf{y}}_{s,b} \;\; \hat{\mathbf{z}}_{s,b} \right] = \left[\hat{\mathbf{x}}_{p,b} \;\; \hat{\mathbf{y}}_{p,b} \;\; \hat{\mathbf{z}}_{p,b} \right]$$

$$\mathbf{R}_{s\to p,b}=\left[\hat{\mathbf{x}}_{p,b}\ \ \hat{\mathbf{y}}_{p,b}\ \ \hat{\mathbf{z}}_{p,b}\right]\cdot\begin{bmatrix}\hat{\mathbf{x}}_{s,b}^{T}\\ \hat{\mathbf{y}}_{s,b}^{T}\\ \hat{\mathbf{z}}_{s,b}^{T}\end{bmatrix} \tag{11}$$

Unless otherwise stated the mean is computed across the largest available integer number of gait cycles beginning at the first frame, determined as in the cmF scheme outlined previously. Calculation of the mean rotation matrix is by singular value decomposition as described in Gramkow (2001). It should be noted that more than one sensible definition of a mean rotation matrix exists, and thus more than one method for calculating such a matrix. This particular method is relatively simple, returns intuitive results, and is widely used. The choice of defining segment orientations by rotational mean-centring was made solely to facilitate the etsF scheme and will be clarified in the next section. Note that this choice determines segment orientations independent of how reference markers are chosen, with the caveat that for each segment they and the origin marker must not be collinear.

Once a segment's coordinate system has been computed at all frames, the position of each marker **m** on that segment at each frame is computed as:

$$\mathbf{m}_{\_,s}=\begin{bmatrix}\hat{\mathbf{x}}_{s,b}^{T}\\ \hat{\mathbf{y}}_{s,b}^{T}\\ \hat{\mathbf{z}}_{s,b}^{T}\end{bmatrix}\left(\mathbf{m}_{\_,b}-\mathbf{o}_{s,b}\right) \tag{12}$$

The mean of the positions across frames $\overline{\mathbf{m}_{\_,s}}$ is then taken as its constant position in that coordinate system. Here positions for any number of markers can be calculated. Thus,

while 45 markers are needed to calculate the 15 segments, once that is done any number of them can be dropped or additional ones encoded as desired.

The pose of each segment at each frame is represented as the rotation and translation needed to transform any coordinates in the segment coordinate system to coordinates in the parent coordinate system. Mathematically:

$$\mathbf{m}_{\_,p} = \mathbf{R}_{s|p} \cdot \mathbf{m}_{\_,s} + \mathbf{t}_{s-p,p} \tag{13}$$

where $\mathbf{R}_{s|p}$ is the rotation matrix effecting the change from segment *s* basis to parent *p* basis and $\mathbf{t}_{s-p,p}$ the translation vector effecting the change from segment *s* origin to parent *p* origin in parent *p* basis (Figure 9). Here,

$$\mathbf{t}_{s-p,p} = \begin{bmatrix} \hat{\mathbf{x}}_{p,b}^{T} \\ \hat{\mathbf{y}}_{p,b}^{T} \\ \hat{\mathbf{z}}_{p,b}^{T} \end{bmatrix} \cdot \left(\mathbf{o}_{s,b} - \mathbf{o}_{p,b} \right) \tag{14}$$

and

$$\mathbf{R}_{s|p} = \begin{bmatrix} \hat{\mathbf{x}}_{s,p} & \hat{\mathbf{y}}_{s,p} & \hat{\mathbf{z}}_{s,p} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}_{p,b}^{T} \\ \hat{\mathbf{y}}_{p,b}^{T} \\ \hat{\mathbf{z}}_{p,b}^{T} \end{bmatrix} \cdot \begin{bmatrix} \hat{\mathbf{x}}_{s,b} & \hat{\mathbf{y}}_{s,b} & \hat{\mathbf{z}}_{s,b} \end{bmatrix} \tag{15}$$

Note the subtle but important difference between $\mathbf{R}_{s \to p,b}$ used earlier and $\mathbf{R}_{s|p}$ used here. The former rotates vectors in the body-centred basis such that when applied to segment basis vectors they are transformed to parent basis vectors. The latter effects a change of basis from segment basis to parent basis.

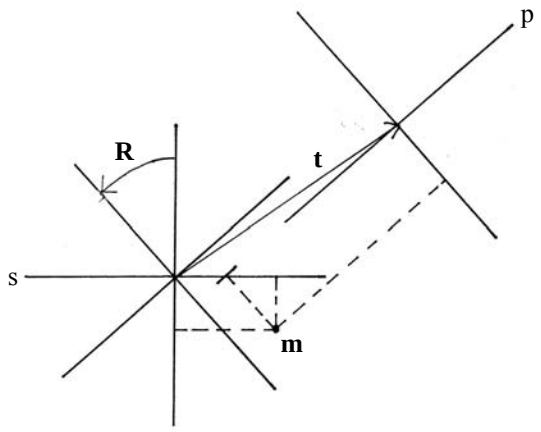

Figure 9. Rotation* R *and translation* t *taking the coordinates of a marker* m *from a segment coordinate system* s *to parent coordinate system* p*. Dashed lines denote coordinates in either system.

Segment orientations have typically three degrees of freedom and are not encoded as nine-element rotation matrices but rather as three Euler angles φ, θ, ψ. Using Euler sequence *x,y,z*, Euler angles can be calculated from their rotation matrix and a rotation matrix from its Euler angles as follows:

$$\begin{cases} \varphi = \arctan2\left(R_{23}, R_{33}\right) \\ \theta = -\arcsin\left(R_{13}\right) \\ \psi = \arctan2\left(R_{12}, R_{11}\right) \end{cases} \tag{16}$$

$$\mathbf{R} = \begin{bmatrix} c(\theta)c(\psi) & c(\theta)s(\psi) & -s(\theta) \\ s(\varphi)s(\theta)c(\psi) - c(\varphi)s(\psi) & s(\varphi)s(\theta)s(\psi) + c(\varphi)c(\psi) & c(\theta)s(\varphi) \\ c(\varphi)s(\theta)c(\psi) + s(\varphi)s(\psi) & c(\varphi)s(\theta)s(\psi) - s(\varphi)c(\psi) & c(\theta)c(\varphi) \end{bmatrix} \tag{17}$$

where *s* is short for sine and *c* for cosine. Refer to Diebel (2006) for an excellent compendium explanation of Euler angles and their relationship to rotation matrices. The choice of Euler sequence *x,y,z* out of the twelve possible sequences was arbitrary since all

non-repeating sequences were found to afford the same quality of ets reconstructions from etsF data (next section). In summary, the ets scheme encodes each segment's pose at each frame as three Euler angles φ, θ, ψ (rad) for rotation and three Cartesian components (mm) in the vector $\mathbf{t}_{s-p,p}$ for translation.

Total ets representation length is $6 \times n_s \times n_f$, where number of segments $n_s = 15$, to encode segment poses, plus $3 \times n_m$ to encode marker positions. How this compares to cm length of $3 \times n_m \times n_f$ depends on n_m and n_f. Three markers per segment for a total of 45 is the minimum needed to carry identical information between cm and ets representations. At 45 markers and typical numbers of frames ets length is significantly smaller (e.g. for 500 frames 45,135 elements versus 67,500 in a cm representation) and, with negligible error, an ets representation is significantly more efficient. This reflects the benefit of reducing from the position of each marker at each frame to the poses of only 15 segments at each frame.

An ets representation can be converted without further loss of information to a cm by calculating rotation matrices from the Euler angles for each segment at each frame and applying to each marker at each frame the transformation in equation 13 sequentially down the hierarchy of segments, beginning with the marker's constant coordinates in the coordinate system of the segment to which it is attached and ending with the marker's varying coordinates in the body-centred coordinate system following the last transformation from the pelvis segment:

$$\mathbf{m}_{\_,b} = \mathbf{R}_{pn|b} \cdot \left(\ldots \mathbf{R}_{p1|p2} \cdot \left(\left(\mathbf{R}_{s|p1} \cdot \mathbf{m}_{\_,s} + \mathbf{t}_{s-p1,p1} \right) \right) + \mathbf{t}_{p1-p2,p2} \ldots \right) + \mathbf{t}_{pn-b,b} \qquad (18)$$

where *pn*'s in subscripts identify parent segments proceeding down the hierarchy. Following convention in computer graphics, each affine rotation + translation transformation can be applied as a single linear transformation in an augmented matrix $\mathbf{Rt}_{s/p}$ acting on homogeneous coordinates:

$$\begin{bmatrix} \vdots \\ \mathbf{m}_{\_,p} \\ \vdots \\ 1 \end{bmatrix} = \mathbf{Rt}_{s/p} \cdot \begin{bmatrix} \vdots \\ \mathbf{m}_{\_,s} \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \ddots & & & \vdots \\ & \mathbf{R}_{s\|p} & & \mathbf{t}_{s/p,p} \\ & & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \vdots \\ \mathbf{m}_{\_,s} \\ \vdots \\ 1 \end{bmatrix} \tag{19}$$

giving the easily composed sequence of transformations:

$$\begin{bmatrix} \vdots \\ \mathbf{m}_{\_,b} \\ \vdots \\ 1 \end{bmatrix} = \mathbf{Rt}_{pn/b} \cdot \ldots \cdot \mathbf{Rt}_{p1/p2} \cdot \mathbf{Rt}_{s/p1} \cdot \begin{bmatrix} \vdots \\ \mathbf{m}_{\_,s} \\ \vdots \\ 1 \end{bmatrix} \tag{20}$$

For purposes of the etsF scheme to follow it will be convenient to express the segment pose part of an ets walker as a matrix $\mathbf{w\_ets}$ where each column or frame vector contains the φ, θ, ψ Euler angles and the three components of the $\mathbf{t}_{s-p,p}$ translation vectors for all segments at a given frame and each row or coordinate time series vector contains a given coordinate of a given segment across frames.

2.55. etsF (ets Fourier-based)

An etsF representation is analogous to a cmF representation based on cm data as outlined in Section 2.52 and is computed in the same way on the segment pose part of an ets representation. However, fundamental period and phase shift are both taken from corresponding cmF data. Also, the marker position part of an ets representation, being unrelated to time, remains unchanged in an etsF. A representation thus consists of the

$6 \times n_s \times \left(2 \times n_h + 1\right)$ elements in a $\mathbf{w\_ets}_F$ matrix plus the single number T plus the $3 \times n_m$ elements in the unchanged marker position part. As with cmF data, unless otherwise stated representations are calculated across all available ets frames, two harmonics are retained, and walkers are phase-shifted, making the etsF scheme morphable and allowing it to function as the pre-PCA scheme for etsPC, outlined next. Note that the coordinates of the marker position part are already coherently interpreted across walkers and thus already morphable.

One further point peculiar to calculating etsF data should be noted. The first Fourier basis vector (the column of 1's in the analogous equation 2) has the effect of isolating the linear arithmetic mean of each ets coordinate across frames (at least for an integer number of gait cycles in the ets data). This is appropriate when this mean is the centre of smooth periodic variance of a coordinate, allowing the residual data to be efficiently decomposed by the remaining sinusoidal basis vectors. However, this is not necessarily the case for angular data, specifically the three Euler angle coordinates for each segment, due to their modular nature (the fact that any angle is equivalent to that same angle $\bmod 360\,°$). Specifically, it will fail to be the case if the time course of an Euler angle coordinate overlaps the wrapping point for that coordinate – e.g. $180\,°$ for Euler angles φ and ψ defined on $\left(-180, 180\right]°$. For example, if a time course for a φ coordinate were hypothetically a pure sinusoid varying between $-175\,°$ and $175\,°$ the linear arithmetic mean would be zero and the sinusoidal basis vectors would be forced to decompose data that would be highly discontinuous, repeatedly jumping from the interval $\left(-180, -175\right]°$ to $\left[175, 180\right]°$. The resulting decomposition would be very inefficient despite the time course being a simple sinusoid.

The appropriate way to handle such cases would be to use a circular mean rather than a linear mean and incorporate wrapping when isolating and adding the first component in the Fourier series. The standard circular mean used for angular data in most contexts is the circular directional mean (often just referred to as the circular mean or the directional mean), computed as the angle of the mean of vectors on the unit circle representing the angles in the data – see Mardia (1972). In the example above, the circular directional mean of the sinusoid varying between $-175°$ and $175°$ would give the more sensible answer of $180°$ rather than zero, which when isolated with wrapping would leave residual data consisting of the sinusoid centred at zero to be decomposed nicely with a single basis vector.

The only problem is that while the circular directional mean works quite well in most contexts, because the sinusoidal basis vectors all have arithmetic mean zero and will therefore be less able to decompose residual data with any non-zero arithmetic mean, the mean isolated by the first basis vector must be an arithmetic one or error at Fourier decomposition will be increased. The ideal approach would be to use the circular arithmetic mean (in contrast to the regular linear arithmetic mean). The circular arithmetic mean is the quantity μ that minimizes the error measure $\sum_n \left|\theta_n - \mu\right|$ across angles, whereas the circular directional mean only minimizes $\sum_n \left(-\cos\left(\theta_n - \mu\right)\right)$ (Mardia 1972). Intuitively, one can think of the circular arithmetic mean as the mean that gives exactly the same answer as the linear arithmetic mean for any spread of angles that does not overlap the wrapping point but gives an answer akin to the circular directional mean otherwise. Unfortunately, the circular directional mean is approximately but not in

general exactly equal to the circular arithmetic mean and a straightforward method for calculating the circular arithmetic mean for arbitrary data is not known.

That said, since the linear arithmetic and circular arithmetic means are identical for angular data that do not overlap the wrapping point, another solution is to ensure that this is always the case and rely on the linear arithmetic mean after all – essentially, to ensure that our angular data can always be treated as if they were linear. This then became a criterion when designing our ets scheme: ideally, to have angular data centred as far from their respective wrapping points as possible. This, finally, was the rationale behind our definition of segment coordinate system orientations by rotational mean-centring across frames. While in general mean-centring by subtracting mean rotation matrices as outlined in Section 2.54 does not exactly centre all corresponding Euler angles to zero, it does so to a good approximation, and adequately for our purpose.

Unless otherwise stated, etsF RUB100 walkers were each the mean of the etsF representations of the several motion captured walks for the respective subject. For purposes of the etsPC scheme to follow it is convenient to express an etsF walker as a column vector **w_etsF** consisting of the elements of $\mathbf{w\_ets}_F$ along with the gait period T and the unchanged marker positions elements.

2.56. etsPC (ets principal component-based)

etsPC is our second PC-based scheme. It is calculated from etsF data exactly the same as the cmPC scheme is from cmF data as outlined in Section 2.53. Unless otherwise stated the etsF RUB100 walkers comprised the training population and we retained 10 PC's.

Each of the six schemes outlined above convert directly to up to two of the other schemes and a representation in any scheme can be converted to any other scheme by taking the appropriate steps through the conversion chain depicted in Figure 10, though error may be acquired in the process.

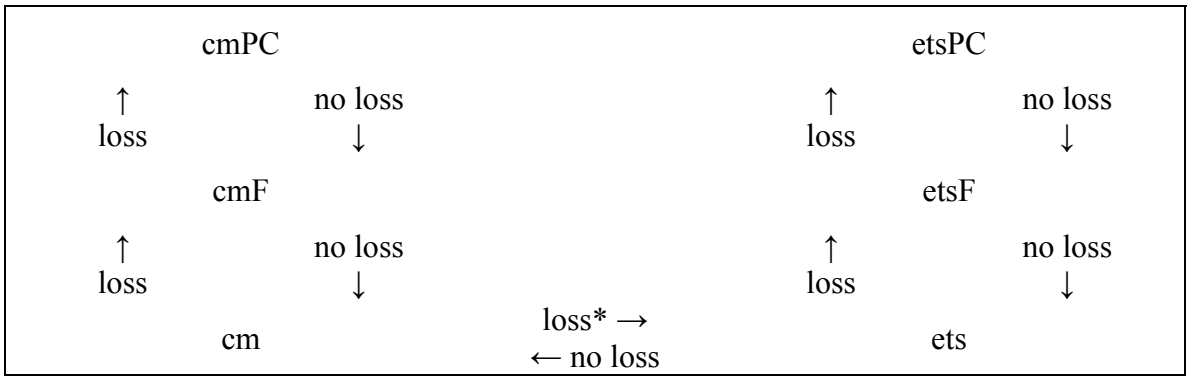

Figure 10. The conversion chain across our six schemes. "loss" indicates that typically error is acquired in a conversion; "no loss" indicates that no error is acquired. *Loss at cm → ets conversion due to non-rigid marker attachment is for many purposes negligible. In the computational experiment original cm data is preprocessed to include this error such that it is exactly equivalent to original ets data (Section 3.1).

2.6. Walker visualization

Regardless of representation, for visual display any walker was converted without further loss of information to cm frame vectors at each frame time, which were in turn orthographically projected onto 2D screen coordinates. For stick figure stimuli sticks were displayed between appropriate point lights in those projections. PLDTools software (Biomotion Lab, Queen's University, Kingston, Canada), a MATLAB toolbox based on Psychophysics Toolbox (Brainard 1997, Pelli 1997), was used to display all stimuli and run all behavioural experiments. Walkers were displayed with 3×3 pixel white point lights and, for stick figures, 1 pixel-wide white lines, with random phase, no simulated occlusion, and on a black background on a 19" LCD TFT monitor at 1024×768

resolution and 75 Hz refresh rate. For all behavioural experiments walkers subtended approximately 6 ° of vertical visual angle.

CHAPTER 3. COMPUTATIONAL EXPERIMENT

COMPARISON OF CMPC AND ETSPC REPRESENTATION SCHEME EFFICIENCIES

3.1. Introduction

Here we compare the efficiencies of PC decomposition of walker data in the cmPC and etsPC schemes, as well as the overall efficiencies of those schemes. Our basic approach is to start with identical walker data in cm and ets representations, compute cmPC and etsPC representations respectively at each possible number of retained PC's (representation length), and then from each reconstruct a lower level representation and compute error in that reconstruction. The schemes that reconstructions are computed in, referred to as *reconstruction schemes*, are chosen such that the downward reconstruction process is always lossless. Thus, error in reconstructions is entirely error that was inherent in the high level cmPC and etsPC representations. Error across numbers of PC's is compared for reconstructions from cmPC versus etsPC data, with less error across numbers of PC's corresponding to greater efficiency. Error curves are averaged across the 100 walkers in the RUB100 set to address efficiency at the level of schemes rather than individual representations.

Whether a comparison is of PC decomposition efficiency or of overall PC-based scheme efficiency is decided by reconstruction scheme(s):

1. *Comparison of PC Decomposition Efficiency*

   Comparison in cmF and etsF reconstruction schemes for cmPC versus etsPC data respectively

   original cm → cmF → cmPC → cmF reconstruction

   original ets → etsF → etsPC → etsF reconstruction

   Error here is measured as deviation from pre-PCA cmF or etsF data. This error is due to discarded PC's only. This gives a comparison of efficiencies of PC decomposition, or equivalently of quality of linearization of cm versus ets data in their respective pre-PCA schemes. Comparison of errors in these "home" reconstruction schemes is sensible even though the schemes are different because we employ a common relative measure of error (equation 21 below).

2. *Comparison of Overall PC-Based Scheme Efficiency*

   Comparison in common cm reconstruction scheme for cmPC versus etsPC data

   original cm → cmF → cmPC → cmF → cm reconstruction

   original ets → etsF → etsPC → etsF → ets → cm reconstruction

   Error here is measured as deviation from original cm data. This error is a combination of error due to discarded harmonics at Fourier decomposition and discarded PC's at PC decomposition, and is thus the total cost in information of representation in a PC-based scheme. This gives a comparison of overall efficiencies of the PC-based schemes. Comparison of errors in a common cm representation is fair because original cm data are preprocessed by conversion to ets then back to cm to apply the rigid marker attachment approximation (Section 2.54) such that original cm and ets data contain identical information.

Efficiency is affected by the number of harmonics retained pre-PCA – PC decomposition efficiency indirectly and overall PC-based scheme efficiency likewise and also directly since error at Fourier decomposition contributes to overall error. We are interested in cmPC and etsPC schemes at configurations where their overall efficiencies are maximal – i.e. where optimal numbers of harmonics have been retained at cmF and etsF stages – such that they are functioning at their respective bests. Thus, the first part of our investigation is to map the impact of number of harmonics on PC-based scheme efficiency and find these optimal numbers. We then conclude with our comparison of these optimized cmPC and etsPC schemes.

3.2. Methods

Walkers in the 45-marker set were used since this contains the minimum number of markers needed to carry identical information in cm-based and ets-based representations (Section 2.54). Gait period for cmF and etsF representations was rounded to an integer number of frames and that number of frames beginning with the first (giving the first gait cycle) was used for cm → cmF and ets → etsF as well as for cmF → cm and etsF → ets steps. This was done to facilitate maximally efficient Fourier decompositions equivalent to DFT's – i.e. at a full complement of harmonics error at Fourier decomposition falls to exactly zero on a walker-by-walker basis. For consistency, the same first gait cycle was also used for vertically centring and size normalizing walkers (Section 2.3), and for mean-centring segment basis vectors when calculating original ets data. Phase shifts needed to make the Fourier-based schemes morphable were recorded for each walker at

cm → cmF and ets → etsF steps and reversed at cmF → cm and etsF → ets steps so that reconstructed cm representations were aligned in time with original data for comparison. When constructing PC-based schemes a leave-one-out methodology was employed: for each of the RUB100 walkers tested a different training population consisting of the 99 remaining walkers was used such that the test walker was not included. In every case a full set of PC's consisted of the number of training walkers $-1 = 98$ vectors.

Measured error was relative root mean squared (RMS) error across the RUB100 data set. Relative RMS error in the cmF or etsF reconstruction scheme was calculated for each number of retained PC's as follows. The basic object of reconstruction is a cmF or etsF walker vector. The squared relative error of a reconstructed test walker is the mean across coordinates of the squared error of each coordinate *c* divided by the mean of the variances across the 99 training walkers of all coordinates with the same unit of measurement. For example, the error of an Euler angle-related coordinate measured in rad in an etsF representation is scaled by the mean variance of all Euler angle-related coordinates only. This is to accommodate for the fact that not all coordinates have a common unit of measurement. Error is measured with respect to pre-PCA cmF or etsF data. We then take the mean across reconstructed walkers, followed by the square root, to give relative RMS error:

$$\mathrm{RMSRE} = \sqrt{\frac{1}{n_w}\sum_{w}\frac{1}{n_c}\sum_{c}\frac{\left(w_{w(c)} - w'_{w(c)}\right)^2}{\frac{1}{n_\gamma}\sum_{\gamma}\frac{1}{n_\upsilon}\sum_{\upsilon}\left(w_{\upsilon(\gamma)} - \overline{w_{w(\gamma)}}\right)^2}} \tag{21}$$

where $w_{w(c)}$ is coordinate *c* in pre-PCA cmF/etsF walker vector indexed *w*, prime denotes reconstruction, γ is used to represent coordinates with the same unit of measurement as a

given coordinate c, υ is used to represent training walkers for a given test walker, $\overline{w_{w(\gamma)}}$ is the mean coordinate γ across training walkers for a given test walker indexed w, and $n_w = 100$, $n_\upsilon = 99$, n_c, and n_γ are the total number of walkers, training walkers for a given test walker, coordinates, and coordinates in a given unit of measurement respectively. Since the result of a reconstruction at zero PC's is the mean training walker, relative RMS error would reduce to one at that point if test walkers were included in training walkers. For perfect reconstructions the numerator vanishes and relative RMS error reduces to zero. This makes equation 21 a suitably relative definition of error. Note that in etsF representations the marker position elements for origin markers are uniformly zero across walkers and thus do not affect PCA and would contribute zero error, which would also be scaled by zero variance. Thus, we do not include those elements in calculation of that error.

Relative RMS error in the cm reconstruction scheme was calculated analogously. Here the basic object of reconstruction is a cm frame vector for a walker. The squared relative error of a reconstructed frame is the mean across coordinates of the squared error of each coordinate c divided by the mean of the variances across frames for the original walker of all coordinates with the same unit of measurement (noting that in the cm scheme all coordinates have the common unit of measurement mm). Error is measured with respect to original cm data. We then take the mean across frames for each reconstructed walker and then across walkers, followed by the square root, to give relative RMS error:

$$\mathrm{RMSRE} = \sqrt{\frac{1}{n_w}\sum_{w}\frac{1}{n_f}\sum_{f}\frac{1}{n_c}\sum_{c}\frac{\left(w_{w(cf)} - w'_{w(cf)}\right)^2}{\frac{1}{n_\gamma}\sum_{\gamma}\frac{1}{n_f}\sum_{fl}\left(w_{w(\gamma fl)} - \overline{w_{w(\gamma)}}\right)^2}} \tag{22}$$

where $w_{w(cf)}$ is coordinate c in frame vector f for original cm walker indexed w, prime denotes reconstruction, γ is used to represent coordinates with the same unit of measurement as a given coordinate c, $\overline{w_{w(cl)}}$ is the mean coordinate cl across frames for an original walker indexed w, and $n_w = 100$, n_f, n_c, and n_γ are the total number of walkers, frames for a given walker, coordinates, and coordinates in a given unit of measurement respectively. Since the result of a reconstruction at zero harmonics but with no further error acquired at PC decomposition (essentially impossible due to the leave-one-out methodology) is the mean frame, relative RMS error reduces to one under those circumstances. For perfect reconstructions the numerator vanishes and relative RMS error reduces to zero. This makes equation 22 also a suitably relative definition of error.

3.3. Results and discussion

Figure 11 and Figure 12 below show curves of relative RMS error from cmPC data versus number of PC's at several sample numbers of retained harmonics at the pre-PCA cmF stage. In Figure 11 it can be seen that as number of harmonics increases error at PC decomposition as measured in cmF reconstructions also increases. Note that any error at 98 PC's is entirely due to the leave-one-out methodology. For total error as measured in cm reconstructions in Figure 12 the increase in error at PC decomposition with number of harmonics couples with the straightforward decrease in error at Fourier decomposition to

give an optimal number of two retained harmonics. With fewer harmonics than two, error acquired at Fourier decomposition outweighs error avoided at PC decomposition such that total error increases, while with more harmonics than two, error acquired at PC decomposition outweighs error avoided at Fourier decomposition such that total error still increases. In fact, trying to improve the quality of cmPC representations by retaining four harmonics pre-PCA is about as bad as retaining only one harmonic.

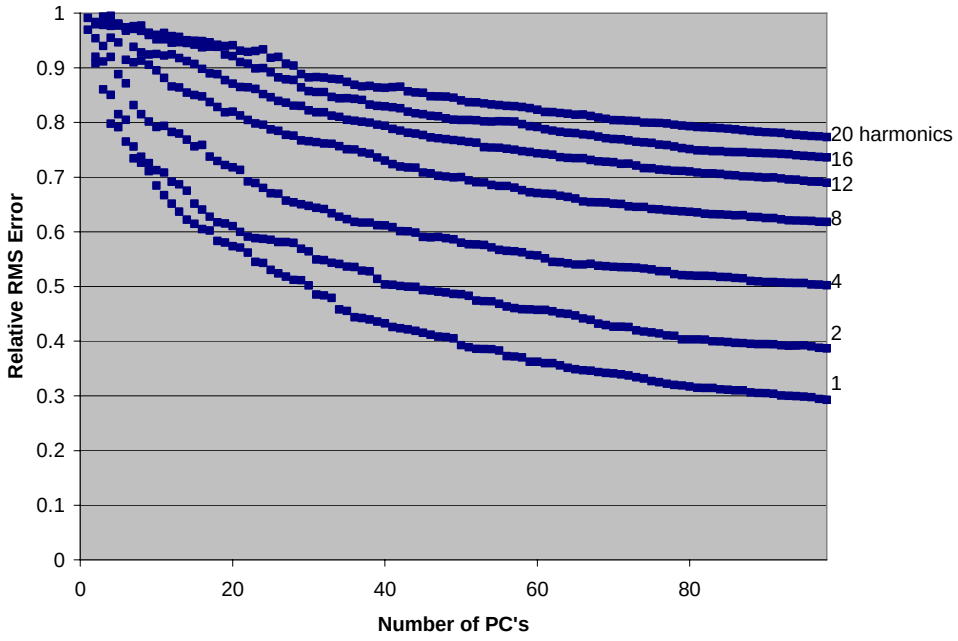

Figure 11. cmPC relative RMS error (cmF reconstructions) vs. number of PC's for 1, 2, 4, 8, 12, 16, and 20 retained harmonics at the pre-PCA cmF stage. Error bars in this and the following figures representing standard error are generally the same size as or smaller than data points and are omitted for readability. Error bars can be somewhat larger for points in isolated regions where a curve fluctuates far from smoothness, but in general smoothness is the most effective visualization of uncertainty.

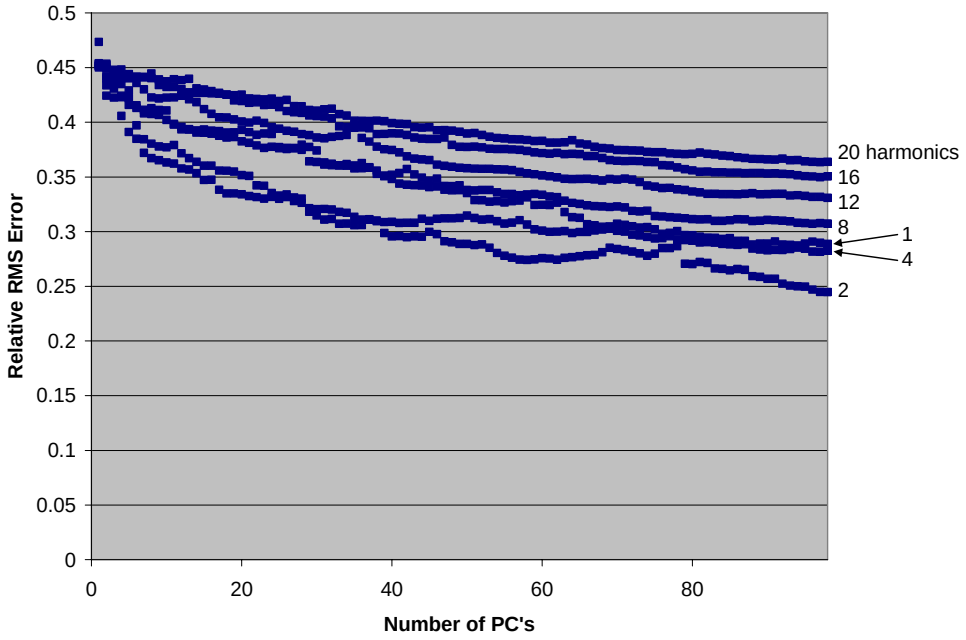

Figure 12. cmPC relative RMS error (cm reconstructions) vs. number of PC's for 1, 2, 4, 8, 12, 16, and 20 retained harmonics at the pre-PCA cmF stage

Results at all other numbers of harmonics fall in line with these trends and similar results are seen for etsPC errors. The big picture of relationship of number of retained harmonics to PC decomposition efficiency and overall PC-based scheme efficiency is captured in Figure 13 and Figure 14 below. For these plots we have calculated areas under relative RMS error curves (which reduce to the sum of relative RMS errors across numbers of PC's). Since we equate error across numbers of PC's to efficiency, we take area under the curve as a measure that is inversely related to efficiency. One curve, and thus one area under a curve, exists for each of the cmPC and etsPC schemes at each number of harmonics retained in the pre-PCA cmF or etsF stage and as measured in a given reconstruction scheme. In Figure 13 (relates to Figure 11 above) we see that both cmPC and etsPC PC decomposition efficiency as based on error measured in cmF or etsF reconstructions respectively monotonically decreases with number of harmonics. In

Figure 14 (relates to Figure 12 above) the focus is on overall cmPC and etsPC scheme efficiencies as based on error measured in cm reconstructions. Here we see the balance between the benefit at Fourier decomposition and the cost at PC decomposition of retaining harmonics giving rise to an optimal numbers of harmonics for PC-based scheme efficiency. For both cmPC and etsPC this optimum is two. In both Figure 13 and Figure 14 results are only plotted to 20 harmonics but trends continue across the full complement of harmonics. It should also be noted that while we have focused on efficiencies as areas under error curves, similar results are obtained when focused on errors alone at a given number of PC's, including the standard 10 PC's for the cmPC and etsPC schemes in applications.

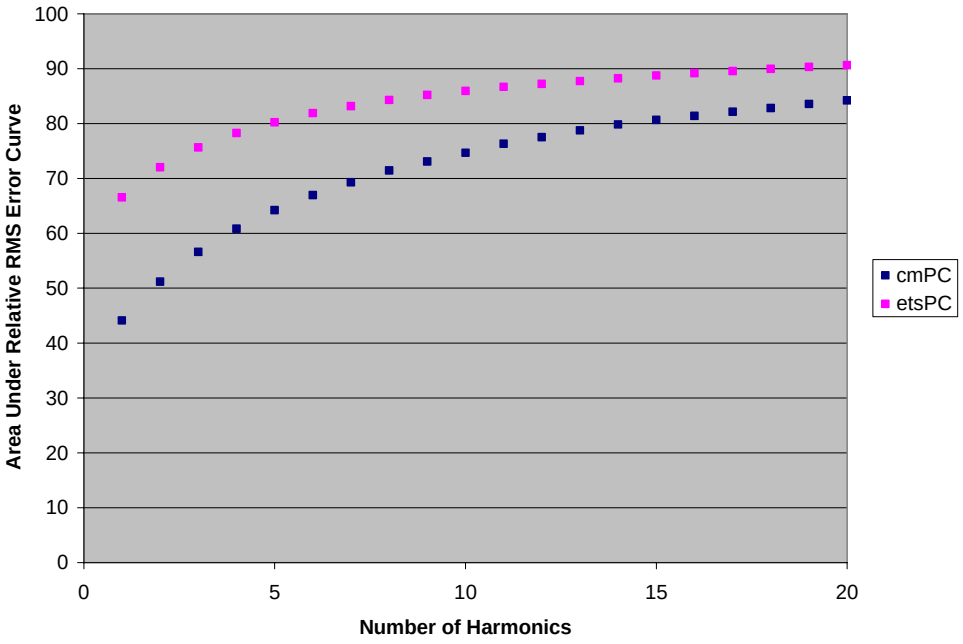

Figure 13. Area under relative RMS error curve (cmF/etsF reconstructions) vs. number of harmonics retained at the pre-PCA cmF/etsF stage for cmPC and etsPC schemes. Area under the curve here is taken as a measure that is inversely related to PC decomposition efficiency.

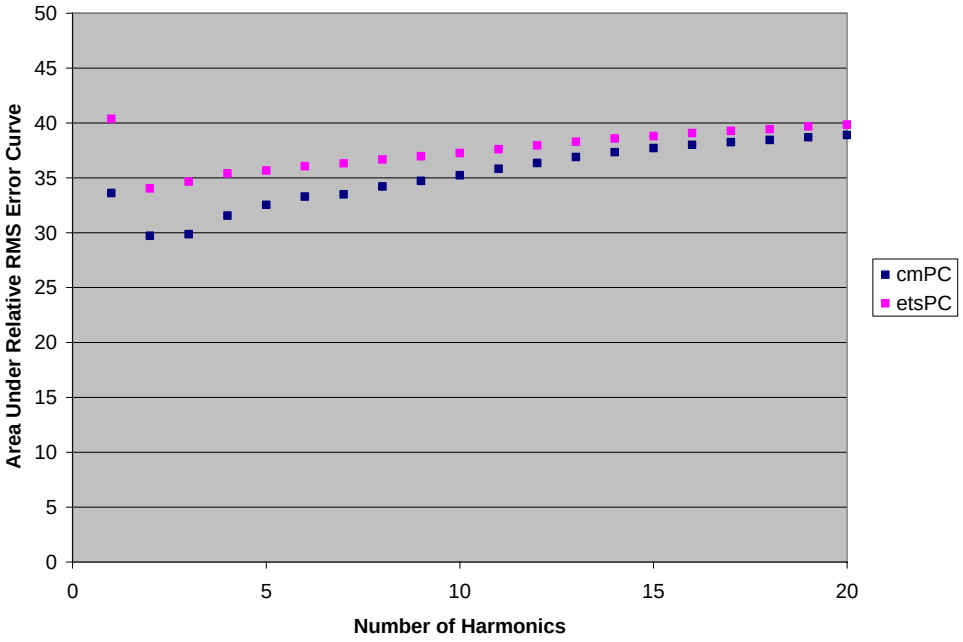

Figure 14. Area under relative RMS error curve (cm reconstructions) vs. number of harmonics retained at the pre-PCA cmF/etsF stage for cmPC and etsPC schemes. Area under the curve here is taken as a measure that is inversely related to PC-based scheme efficiency.

With the optimal number of harmonics found to be two for both cmPC and etsPC scheme efficiency, we compare these schemes specifically at this number of harmonics retained pre-PCA. Figure 15 shows relative RMS error versus number of PC's in cmF reconstructions from cmPC and etsF reconstructions from etsPC. The significantly lower error across numbers of PC's for the cmPC scheme indicates greater PC decomposition efficiency, meaning data are better linearized in the cmF pre-PCA scheme than in the etsF. Figure 16 focuses on reconstructions in a common cm scheme. Error here is also lower across numbers of PC's for cmPC than etsPC, indicating greater overall cmPC scheme efficiency than etsPC. It can be noted that error due to Fourier decomposition alone at two harmonics as measured in cm reconstructions from pre-PCA cmF and etsF data is about the same between the two at 8-9% relative RMS error (data not shown).

Thus, not only is more error acquired at PC decomposition, but the difference in total error between cmPC and etsPC is mostly the difference at PC decomposition. One additional finding of practical value that can be taken from Figure 16 is that error for both cmPC and etsPC begins to decrease noticeably more gradually around the standard 10 PC's used in applications. The difference between cmPC and etsPC error at this point is less than 5% relative RMS error.

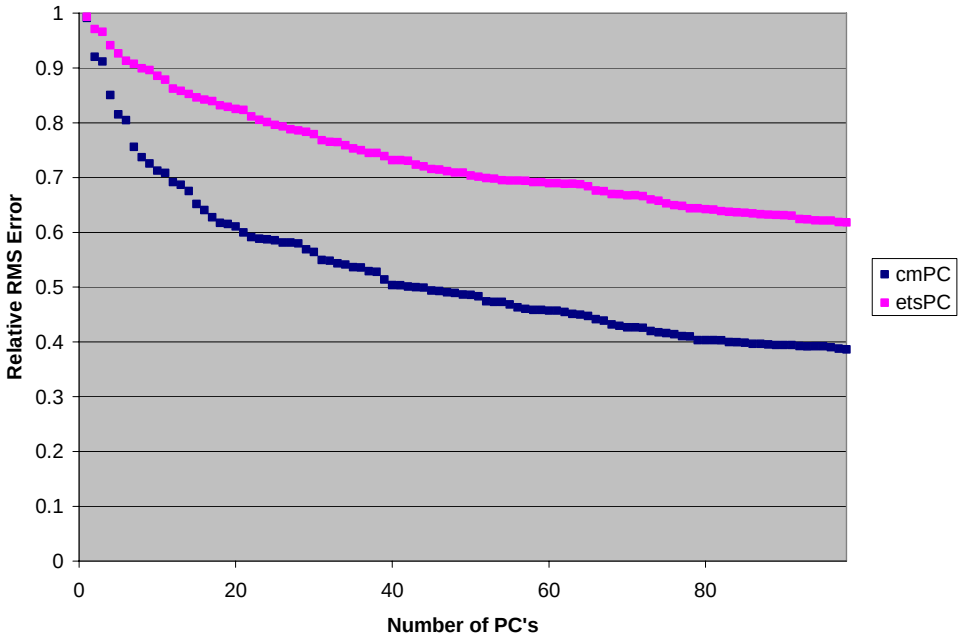

Figure 15. Relative RMS error (cmF/etsF reconstructions) vs. number of PC's for cmPC and etsPC schemes

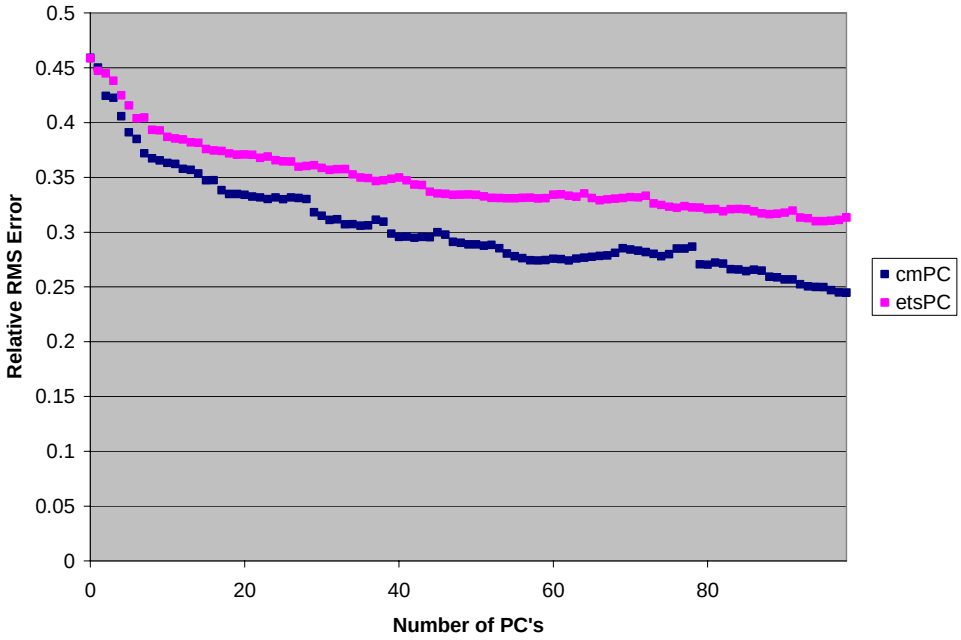

Figure 16. Relative RMS error (cm reconstructions) vs. number of PC's for cmPC and etsPC schemes

Notes on Choice of Reconstruction Schemes

The reconstruction schemes we have used cover all of the obvious venues for comparing error without acquiring further error in reconstruction. A comparison in a common ets reconstruction scheme could entail a further loss of information at the cm → ets conversion for reconstructions from cmPC since the rigid marker attachment approximation that was implemented for original cm data may no longer hold due to deviation acquired at the upward cm → cmF → cmPC steps. Another comparison that may seem conspicuously absent is in cm and ets reconstruction schemes for cmPC and etsPC data respectively. The reason we did not attempt this comparison is that error measured in the ets scheme using equation 22 would have to ignore error acquired in the marker position part of ets representations since such error would be scaled by zero

variance across frames. Meanwhile, this error fully contributes (implicitly) to cm error. Thus, it would not be an even and fair comparison.

CHAPTER 4. BEHAVIOURAL EXPERIMENT 1

MEASUREMENT OF JUST NOTICEABLE DIFFERENCES IN CMPC WALKER SPACE FOR PURPOSES OF EQUALIZING WALKER DISTINCTIVENESS

4.1. Introduction

Caricatures are generated by multiplying veridical walker vectors by some constant > 1, thus increasing distance from the mean walker (the origin). With identity modeled by angle in walker space, a caricature is modeled as affording improved identification performance because margin of error in angle of its vector is decreased from margin of error in angle of the veridical vector. This is so because margin of error in angle is an increasing function of margin of error in coordinates and a decreasing function of distance from the origin, so increasing walker vector length decreases margin of error in angle assuming constant margin of error in coordinates with distance (Section 1.5 Figure 4).

In Section 5 we will test subjects' identification performance of caricatures in cmPC walker space at different caricature levels in order to find the level that affords the best performance in preparation for our originally planned comparison of caricatures in cmPC versus etsPC schemes. Our measurements of identification performance will be averaged across different caricatures and in order to reduce variance we will want to equalize the distinctiveness of their veridical walkers. This is not necessarily accomplished simply by equalizing veridical walker vector lengths in cmPC units. While we assume that margin of error in coordinates is approximately constant with distance from the origin, we do not assume that it is constant across angles – i.e. vectors of the same length but in different directions may have different margins of error in coordinates

and thus different margins of error in angle, corresponding to walkers of different distinctiveness.

We attempt to solve the problem of equalizing veridical walker distinctiveness as follows. We assume that margin of error in coordinates is related to just noticeable difference (JND) between walkers – the minimum Euclidian distances between vectors at which subjects are able to perceive the corresponding walkers as different. The purpose of the present experiment is to measure JND in cmPC units along each dimension of the walker space. We can then change the metric of the space from cmPC units to units of those JND's. In this metric JND's (measured in JND's) straightforwardly become equal to one in every direction. Vectors in different directions will then have equal margins of error in coordinates, and if we equalize their lengths then equal margins of error in angle, corresponding to walkers of equal distinctiveness.

The stimuli used in the next experiment will begin as walkers in the cmPC scheme with our standard 10 retained PC's – i.e. in a 10-dimensional space. Accordingly, here we measure JND along each of those 10 dimensions. We use a consecutive match to sample paradigm in combination with adaptive staircase methodology. In each trial subjects are shown a sample walker followed by a pair of potential match walkers and are asked to respond with which of the potential match walkers is the same as the sample, thus testing their ability to distinguish the potential match walkers. The two potential match walkers are positioned along one of the PC dimensions, centred at the origin, at a certain distance from the origin. For each PC we run an independent staircase of trials where this distance is increased with subject's correct responses, making subsequent trials easier, and decreased with incorrect responses, making subsequent trials harder. In

this way distance is adapted to the value at which performance is at a certain threshold above chance. That distance gives JND along that PC.

4.2. Methods

4.21. Subjects

Six subjects participated in this experiment on a voluntary basis. All were members of our lab. There were three females and three males. All had normal or corrected to normal vision.

4.22. Stimuli

Stimuli consisted of artificially generated 15-marker walkers. For each trial a pair of walkers was generated in the cmPC scheme where one walker equaled $\alpha_n \mathbf{PC}_n$ and the other $-\alpha_n \mathbf{PC}_n$, where $\mathbf{PC}_n$ is one of the 10 PC basis vectors. Using an adaptive staircase procedure for each of the 10 α's, their values changed as trials were logged. Orientation was either $5°$ to the left or $10°$ to the right of frontal (see below).

4.23. Procedure

In each trial one of the pair of walkers for the trial – the sample walker – was shown for 1.5 s, followed by an inter-stimulus interval of 0.5 s, followed by both walkers of the pair – the potential match walkers – with approximately $8°$ of visual angle between them horizontally. A left/right arrow prompt was displayed at the bottom of the screen during the potential match walkers and subjects had to respond with the left or right arrow key indicating which was the same as the sample. Potential match walkers were both displayed at an orientation of either $5°$ to the left or $10°$ to the right of frontal and the

sample walker to the other orientation such that the task required matching of underlying walker information rather than just of 2D stimuli at a superficial level. Which of the potential match walkers was on the right or the left, which was the sample walker, and orientation was randomized and counterbalanced across trials. Each trial ended when input was received. Subjects were instructed to use any features except orientation and to make their choices as fast as possible without unreasonably sacrificing accuracy. Inter-trial interval was 0.8 s.

The experiment was conducted as 10 randomly interleaved 3-down/1-up adaptive staircases with α for the respective PC's as the staircase variables. In such staircases, the first trial of the staircase begins at a starting value for the staircase variable and the variable is increased by a step up size every time an incorrect response is received, making subsequent trials easier, or decreased by a step down size every time three correct responses are received in a row, making subsequent trials harder. The staircase methodology was first introduced in Dixon & Mood (1948) and has been widely used, refined, and varied since then. Two papers in particular by Garcia-Perez in 1998 and 2000 respectively carefully investigated the optimal setup of forced choice staircases with fixed step sizes in terms of the statistical properties of their results. Following their recommendations, we set the step up size for each staircase to lie in the interval between $\frac{2}{3}$ and $\frac{3}{3}$ of the previously estimated spread of the psychometric function of the staircase variable and the step down size $= 0.74 \times$ step up size. The spread of the psychometric function is, loosely, the length of the interval of the staircase variable on which the function is non-asymptotic (a precise definition is given in Garcia-Perez 1998). Under this configuration the average across turning points of the staircase variable convergently

corresponds to its threshold at approximately 83% correct, with a staircase long enough to provide at least 20 and ideally 30 or more turning points trustworthy to give good convergence. Staircase variable values should fluctuate quickly across trials about the true value, usually wandering within one step up above and two steps down below (Figure 17). Also following the recommendations in these papers, we set the starting value for each staircase variable several step down sizes above its previously estimated value and used a 1 down/1 up rule until the first turning point to facilitate as quick a convergence as possible to that neighbourhood. The first two turning points were omitted from calculation of threshold. Final JND's equaled 2α at the respective thresholds.

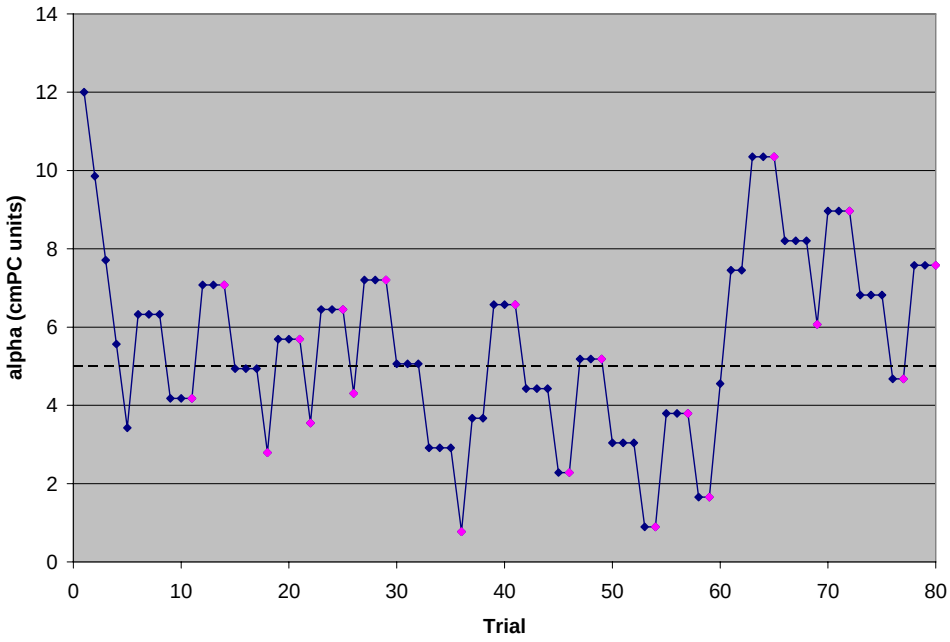

Figure 17. Results of a typical well-configured staircase. Turning points are marked by light points. The dotted line shows the final calculated threshold.

Each of our staircases ran for 80 trials or 23 included turning points (25 total), whichever came first, for a maximum total of 80 trials per staircase or 800 total. Subjects

were allowed to participate in blocks of any duration they wanted with any amount of time between blocks in order to prevent fatigue.

4.3. Results and discussion

Refer to Figure 18 below. The main results to draw from this experiment are JND along each PC in cmPC units to be used to rescale the walker space in JND's in the next experiment. However, it can be noted as an aside that, particularly when JND's are scaled in more conceptually meaningful units of standard deviations of the cmPC RUB100 walker set along the respective PC's, those along the first few PC's were notably smaller than those along higher PC's. This is even more clearly depicted in Figure 19 where JND scaled in standard deviations is plotted versus those standard deviations. The interpretation to be drawn is that along dimensions in walker space that capture more walker variance JND's are smaller in proportion to those variances. This may reflect the greater perceptual saliency of such dimensions.

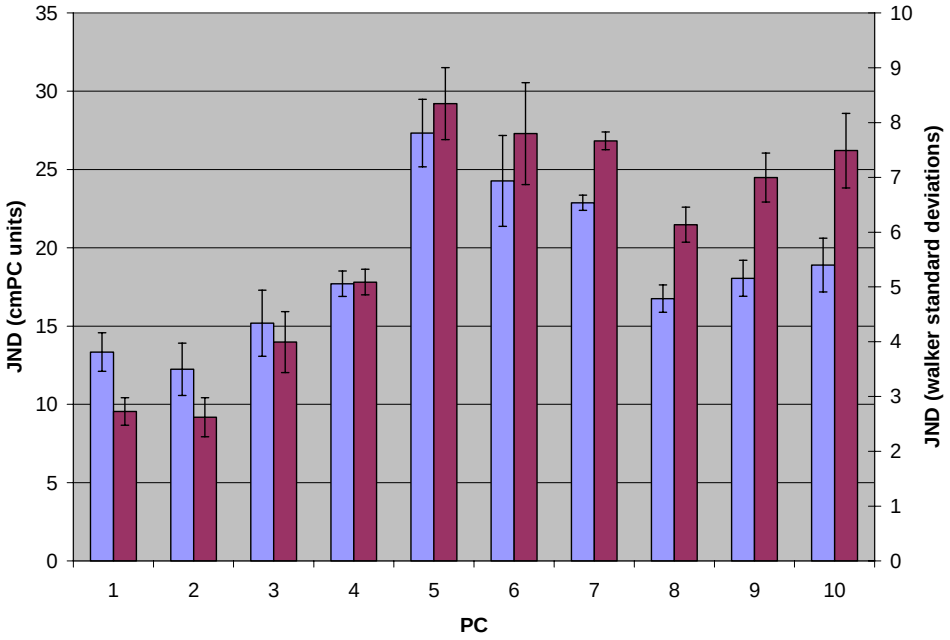

Figure 18. JND (83% correct) vs. PC. Light bars are in cmPC units and dark bars are in standard deviations of the cmPC RUB100 data set along the PC. Error bars are standard error of the mean with between-subjects variance removed.

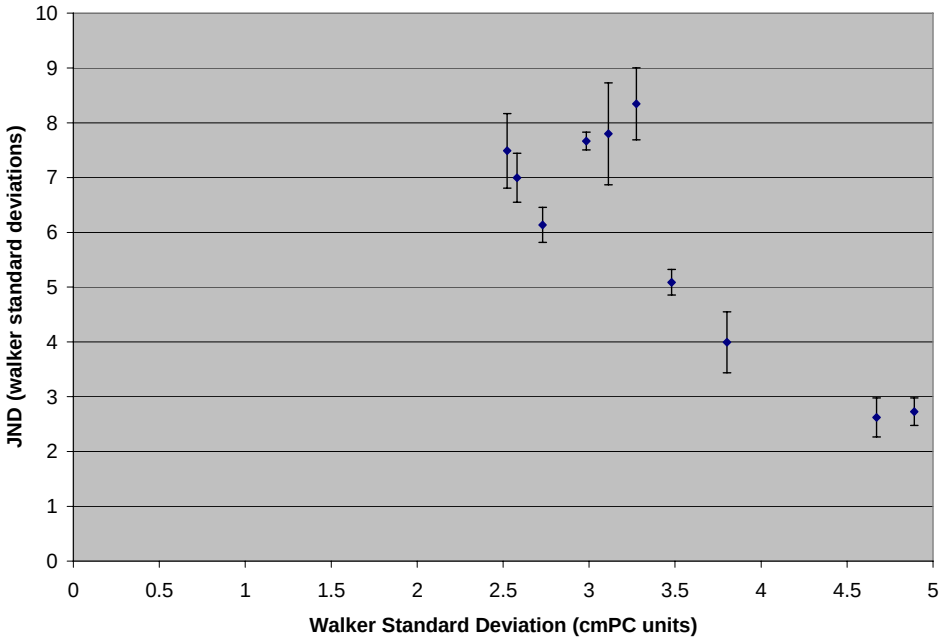

Figure 19. JND (83% correct) vs. standard deviation of the cmPC RUB100 data set along the PC. Error bars are standard error of the mean with between-subjects variance removed.

When read in units of walker standard deviations, our measured JND's may seem excessive. Partly this is due to the fact that the standard deviation of projections of typical walker vectors throughout the space onto an axis is not the same as (and is generally much smaller than) lengths of typical walker vectors along that axis. However, the nature of the task should also be born in mind: that sample walkers were displayed for only 1.5 s and then disappeared such that no side-by-side comparison with potential match walkers was possible, and also that staircases were interleaved such that subjects were prevented from learning where exactly to look for the differences between potential match walkers characteristic of variance along each PC. The kinds of walkers that were required for success in such a difficult task were often well into the realm of cartoonish caricatures. The significance of our results here, and their usefulness for equalizing veridical walker distinctiveness in the next experiment, lies in their relative magnitudes.

CHAPTER 5. BEHAVIOURAL EXPERIMENT 2

MEASUREMENT OF WALKER IDENTIFICATION PERFORMANCE VERSUS CARICATURE LEVEL

5.1. Introduction

Our originally planned main behavioural experiment was a comparison of caricatures generated in the cmPC versus etsPC scheme at the same caricature level in the hope of finding that caricatures generated in one benefit identification more than caricatures generated in the other. Caricatures are generated by multiplying veridical walker vectors by some constant > 1, thus increasing distance from the mean walker (the origin), where multiplier corresponds to caricature level. In our preliminary work we noticed that typical caricatures which we would expect to improve identification performance looked identical between the two schemes when generated from common veridicals. In order to generate analogous caricatures that were perceptibly different and so could plausibly give rise to different identification performance, it was necessary to use such large multipliers that the caricatures looked so extreme as to actually be more difficult to identify than their veridicals. In this way it became important to be clear about how caricature level relates to identification performance, in particular if levels that are too great hinder rather than help identification, and if so what level is optimal. In other words, is too much caricature a bad thing, how much is too much, and how much is just right?

In walker space, the walkers that we encounter in life correspond to a roughly normally distributed population centred around the mean walker. Caricatures at distances significantly beyond this population are known to improve identification performance. Rhodes et al. (1987), Mauro & Kubovy (1992), Byatt & Rhodes (1998), Hill & Pollick

(2000), Rhodes et al. (2000), Pollick et al. (2001), Frowd et al. (2007), and Rodriguez et al. (2008) all investigated caricatures of faces or biological motion at one or a small number of caricature levels that were beneficial, measuring identification performances that were increased or monotonically increasing across levels. We expected there to come a point where caricatures are so far outside the population that they begin to impede performance by distorting distinctive features too much and even, at great enough distances, become difficult to recognize as related to the object class at all. In this experiment we test cmPC caricatures at six levels where at least the highest is chosen deliberately to be counterproductive. We employ the caricature identification paradigm where subjects are asked to learn a number of walkers by name in a training phase and then to identify caricatures of those walkers in a test phase. We map a function of identification performance in the test phase versus caricature level. Any level practicable for an experiment seeking a difference in identification benefit from cmPC versus etsPC caricatures had to be sufficient to give perceptibly different analogous caricatures between the two schemes while also at or not too far beyond the optimal level measured in this experiment.

5.2. Methods

5.21. Subjects

Fifteen subjects participated in this experiment. All were students in psychology or neuroscience at Queen's University. Subjects were reimbursed $10 plus a yummy chocolate bar for their participation. There were nine females and six males. All had normal or corrected to normal vision.

5.22. Stimuli

Veridical stimuli consisted of randomly generated 15clav-marker walkers. For each subject eight walkers – four males and four females – were originally generated in the 15-marker set in the cmPC scheme. The projections of males onto the gender axis were equal and parameterized by a number α and the projections of females were equal and parameterized by $-\alpha$. The gender axis is represented by a unit walker vector **g** and is the axis calculated by a linear discriminant classifying cmPC RUB100 walkers according to the binary male/female feature, with increasing projections of walker vectors onto the axis on one side of the origin optimally corresponding to increasing perceived masculinity and on the other side to femininity. Refer to Troje (2002) for details on calculation of the gender axis. The perpendicular distance from the gender axis to all walkers – essentially the component of walkers capturing all features other than gender – was parameterized by a number β. This implies that all walkers had an angle $\tan^{-1}\frac{\beta}{\alpha}$ (male) or $180° - \tan^{-1}\frac{\beta}{\alpha}$ (female) to the gender axis. Figure 20 below depicts this geometry.

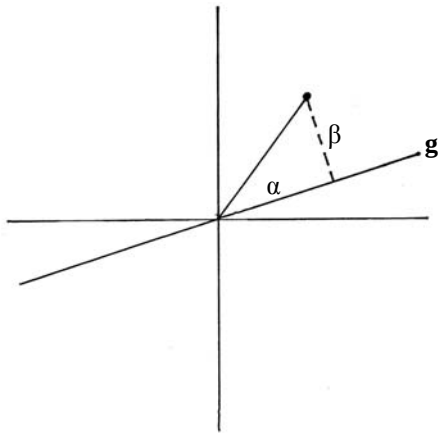

Figure 20. An artificial walker vector in a 2D depiction of the walker space.* g *denotes the gender axis.

The four walker vectors of each gender were generated as close to orthogonal as possible in order to test the caricature effect across as much of walker space as possible. Whether angles of $90°$ are geometrically possible or whether some angle $< 90°$ is the largest that can be achieved depends on the chosen values of α and β. $\alpha = 1.3$ in units of standard deviations of the cmPC RUB100 walkers along the gender axis was chosen to give walkers that were distinctly but not excessively male or female (the mean male walker is 0.78 standard deviations and the mean female -0.78 standard deviations on the gender axis). $\beta = 1.8$ in the same units (see Discussion for the reasoning here). At these values the maximum separating angle between walker vectors within genders that could be achieved was approximately $82°$. Aside from these constraints, walker vectors were generated randomly. The fact that the cmPC scheme is morphable ensures that random points in the space, at least within the statistical distribution of walkers, all represent legitimate walkers. Mathematical details of the generation process are supplied in Appendix 1.

Once the eight veridical walkers were generated, we standardized the lengths of their vectors to be equal when the space is scaled in JND's, using the 10 JND's along the respective PC basis vectors measured in the previous experiment. As outlined there, we expected this process to equalize distinctiveness. The eight vectors were each scaled either up or down to equalize their lengths in JND's under the constraint that the mean of their lengths in cmPC units remained unchanged. After equalization, the mean length of veridical walker vectors in units of standard deviations of projections of cmPC RUB100 walkers was 2.6 with a standard deviation of 0.5, or in units of mean cmPC RUB100 walker vector length was 0.87 with a standard deviation of 0.08, measured across 12 batches of eight walkers.

Caricatures were then calculated from these veridical walkers by multiplying them by some caricature level γ, where a γ of 0 corresponds to the mean walker, 0.5 to an anticaricature at half the distance from the mean to the veridical walker, 1 to the veridical, 2 to a caricature at double the distance from the mean to the veridical, etc. For each walker caricatures at γ = 0.5, 1, 1.7, 2.4, 3.1, and 3.8 were generated. This yielded eight walkers at six caricature levels each for a total of 48 stimuli.

PLD's are the most common stimuli in biological motion research and the 15-marker PLD is the standard stimulus used in our lab. Accordingly, we defaulted to using them in the previous experiment and in pilot runs for this experiment. However, we found in those pilot runs that at least some subjects have a tendency to focus on only one or two of the markers (e.g. at the hips) when performing identification. Since we wanted to measure a caricature effect and caricaturization affects whole stimuli, we wanted to encourage subjects to perceive the walkers as globally as possible. To this end, for this

experiment as well as the next we amended our approach to displaying stick figures. To give the most natural looking stick figures we also amended our stimuli from the 15-marker set to the 15clav-marker set at display (see Section 2.2 for reasoning on this point). This conversion process operates at the step of cmF data in the display algorithm (Section 2.6). It extrapolates 61-marker data from 15-marker data using the 15-marker and corresponding 61-marker RUB100 data sets as exemplars, then simply takes 15clav-marker data as a subset of the 61-marker data. It works by the assumption that the same linear combination of 15-marker exemplars that gives original 15-marker data gives corresponding 61-marker data when applied to 61-marker exemplars. Data for those of the original 15 markers that overlap as members of the final 15clav-marker set (which are all but the one mid-clavicle marker) are unchanged. Mathematical details can be found in Appendix 2. Frontal orientation was used for all stimuli at display. A sample from an example sequence of caricatures is shown in Figure 21 below.

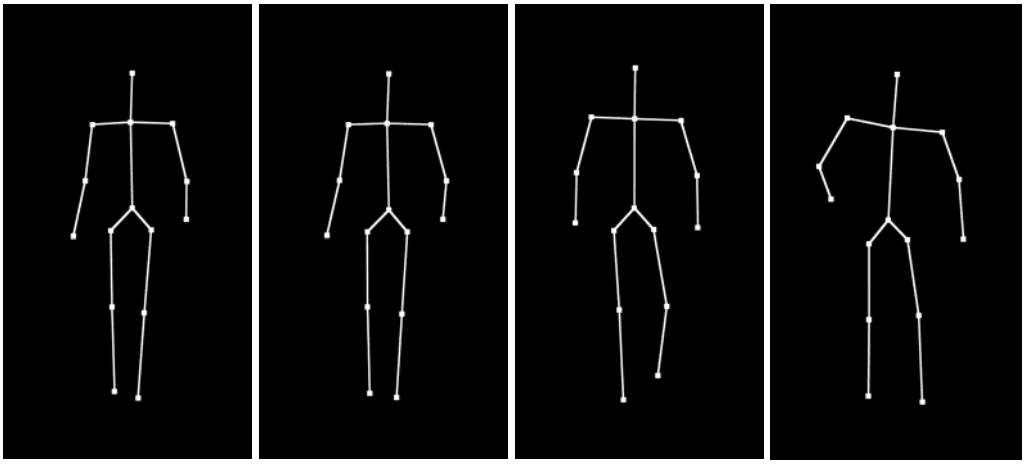

Figure 21. Example stimuli for a male at $\gamma = 0.5$ (anticaricature), 1 (veridical), 2.4, and 3.8 (most extreme). Each is at a different random point in the gait cycle. The kinematic aspect of the caricatures which cannot be shown here is probably even more vivid than the structural.

5.23. Procedure

Before the experiment subjects were shown an example of a male walker and a female walker. The concept of a caricature as an exaggeration of distinctive features was explained with the help of a couple of drawings of artists' caricatures of celebrities. We then explained that caricaturization can also be applied to walkers as an exaggeration of features relative to the norm and we showed an example sequence of walker caricatures across the range of caricature levels tested in this experiment excluding $\gamma = 0.5$. Subjects were told that features open to exaggeration included any aspect of shape or motion of any part of the body, including gait speed.

The first part of the experiment was the training phase. In each trial subjects were shown a random one of four veridical stimuli – two of the generated males and two of the females. On the screen below the display arranged horizontally were four blue buttons containing fictional names in white text associated with the four possible walkers: Joe, Bob, Becky, and Jane. Subjects had to identify the walker in each trial by clicking on the appropriate button with the mouse. Feedback was provided by turning the clicked button green if correct or else turning the button that would have been correct red for 1.5 s, thus enabling subjects to learn the identities of the walkers. Subjects were instructed that trials always contained one of the same two males and two females and that no caricatures were involved in this part of the experiment. Subjects were instructed to use any features and to make their choices as fast as possible without unreasonably sacrificing accuracy. Each trial ended when input was received. Inter-trial interval was 0.5 s. A total of 180 trials consisting of 45 of each walker were performed in random order. 10 random discounted practice trials were also performed at the beginning with supervisor present

and coaching as needed to ensure that subjects understood the paradigm. Subjects typically took between 15 and 20 minutes to complete the 180 training trials, reaching ceiling (90%+) identification performance around the end of the training phase.

The second part of the experiment was the test phase. In each trial subjects were shown a random one of the 48 generated stimuli – thus, possible identities included the familiar two males and two females from the training phase plus the two males and two females not present in the training phase, and all caricatures were included in addition to veridicals. The novel walkers were introduced as distracters so that identification performance for familiar veridicals was reduced from ceiling, allowing room to observe a benefit due to caricature. Below the display the four familiar buttons containing names of the learned walkers remained. Directly below these was one new button marked Other. This button was as large vertically as the four above it but was as large horizontally as all four put together as an indication that in each trial the stimulus was equally likely to have a novel identity as a familiar identity. Subjects had to identify the walker in each trial by clicking on the name associated with its identity or on Other if its veridical was not in the training phase. No feedback was provided. Subjects were told that walker caricatures are defined to have the same identities as their veridicals – i.e. that a walker was to be identified if a subject recognized its distinctive features whether exaggerated or not. Aside from these instructions the test phase began immediately after the training phase. Each trial ended when input was received. Inter-trial interval was 1 s. A total of 288 test trials were performed in random order within six blocks containing one of each of the 48 stimuli. 10 random discounted practice trials were also performed at the beginning with supervisor present and coaching as needed. The 288 test trials were preceded and then

punctuated every 72 trials with a block of eight refresher training trials identical to those that comprised the training phase. Two repetitions of each familiar walker were included in each refresher training block, with trial order randomized within the block. Each block began with a screen prompting the subject that some refresher trials were about to begin and to press the spacebar to proceed and ended with a similar screen prompting the subject that further test trials were about to begin. From beginning to end subjects typically took about 25 minutes to complete the test phase. Measured dependent variable was proportion correct in the test trials where a response was correct if it was the appropriate name for a familiar identity and the Other button for a novel identity.

5.3. Results and discussion

Refer to Figure 22 below. The addition of the four distracter walkers reduced performance for familiar veridical walkers from 90%+ in the training phase (data not shown) to 55-60% in the test phase. A repeated measures one-way ANOVA showed that there was a significant difference in performance across caricature levels, $F(5) = 57.92,\ p < 0.001$. To quantitatively confirm the increasing/decreasing nature of the curve, three pairwise comparisons with Bonferroni correction showed that $\gamma = 0.5$ anticaricatures afforded worse performance than veridical walkers, $\gamma = 2.4$ afforded better performance than veridicals, and $\gamma = 3.8$ worse than 2.4, $p < 0.02$. Thus, we can confirm an increasing benefit of caricature up to about the $\gamma = 2.4$ level. This corresponds to a mean caricature vector length of 6.2 standard deviations of projections of cmPC RUB100 walkers or 2.1 mean cmPC RUB100 walker vector lengths – i.e. about double the distinctiveness of a typical walker. A declining benefit was observed thereafter as

caricatures became too extreme up to the highest measured $\gamma = 3.8$. Judging from the apparent consistent degradation of walker characteristics as γ increases beyond this point, the decreasing trend would continue.

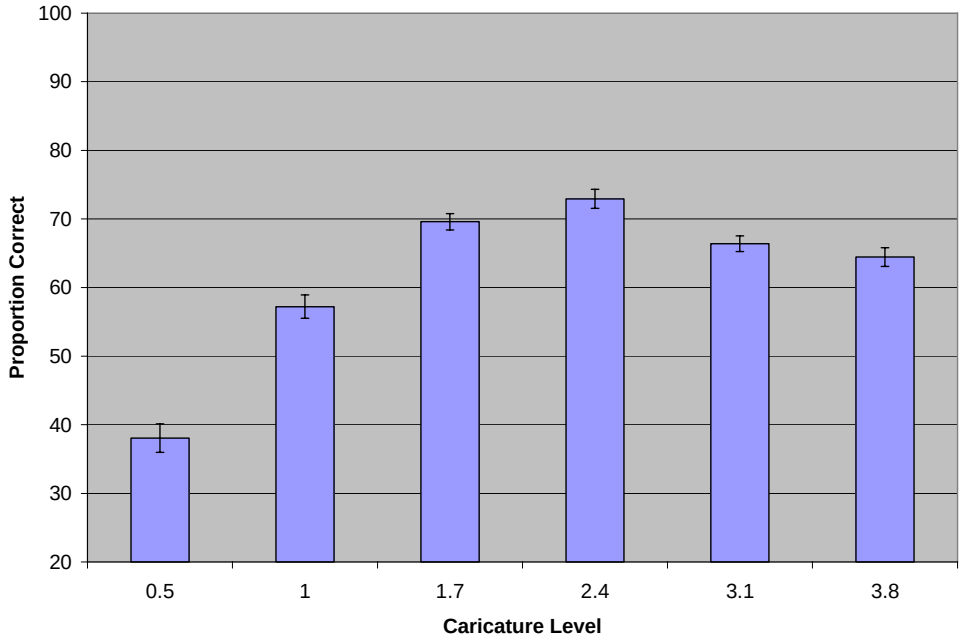

Figure 22. Identification proportion correct vs. caricature level. Error bars are standard error of the mean with between-subjects variance removed. Chance performance was 20%

We also compiled results in the old/new sense. Here measured dependent variable was proportion correct in the test trials where a response was correct if it was any of the four name buttons for a familiar identity or the Other button for a novel identity. These results were essentially the same as identification results, the only exception being that chance performance was 50% as opposed to 20% and correspondingly performance at $\gamma = 0.5$ was about 55%. Accordingly, old/new results are not illustrated.

Issues with Perceptual Distinctiveness of Veridical Walkers

Our attempt to equalize the perceptual distinctiveness of veridical walkers by equalizing their vector lengths with the space scaled in JND's appeared not to be entirely successful. Subjectively, some walkers in each batch of eight were clearly easier to identify than others – a fact noted both by us and by some of our subjects without prompting. There are many possible reasons for this, including inappropriateness of the assumptions that JND's are approximately constant with distance from the origin or that the metric of the walker space is a Euclidian one, or a flaw in our understanding of how the walker space model links to the distinctiveness and caricature effect. Also, our JND's were measured for PLD's in the 15-marker set and walker vector lengths were equalized in that marker set but our final stimuli were stick figures in the 15clav-marker set. It is unlikely that the perceptual difference in marker set is significant since it amounts to only a small shift of one marker, particularly without sticks accentuating that difference, but the difference between PLD's and stick figures is considerable (despite the fact that their representations are identical). Future experiments would be needed to learn more. In any event, the results of this experiment are not discredited because in the worst case scenario unequal veridical distinctiveness just added noise to our results that fell out in the statistics.

Notes on Experiment Design: Distinctiveness, Difference, and Number of Veridical Walkers

A number of considerations that went into the design of this experiment deserve mention. Three main factors decide the level of difficulty for identification of veridical walkers in the training phase. First there is distinctiveness of the walkers as modeled by length of

their vectors, with the greater the distinctiveness the easier identification. Second there is identity difference between walkers as modeled by angle between vectors, with the greater the identity difference between a walker and its closest competitors in walker space the easier identification. Third there is the demand on verbal memory to associate names to walkers once they are recognized by the visual system, which is mainly a function of number of identities that must be juggled. While there may be some interaction between these factors, it is at least true that verbal memory is implicated only in the last. The caricature effect operates at the level of visual recognition and not verbal memory, so in order to key in on it we wanted subjects' performance bottleneck to be the first and second factors. Thus, we used the smallest number of training walkers we realistically could and calibrated difficulty through distinctiveness and identity difference.

To calibrate difficulty we were guided by the following criteria. We wanted subjects to learn the walkers in the training phase thoroughly since activation of a caricature effect requires preexisting familiarity. We wanted the training phase to be no longer than 15 minutes so that as much time could be spent in the test phase as possible. Lastly, we wanted the training phase to be minimally tiring so that subjects would be as fresh and motivated for the test phase as possible. Thus, we calibrated training difficulty such that subjects typically reached ceiling (90%+) performance around the end of the training phase, and the four distracter walkers were then included in the test phase so that identification performance for familiar veridicals was reduced, allowing room to observe a benefit due to caricature.

With walker parameter α preset in order to give an appropriate level of gender and walker vectors generated with separating angles within genders as close to $90°$ as possible in order to canvas as much of walker space as possible, distinctiveness and identity difference of veridical stimuli came down to parameter β at walker generation (Section 5.22). α and β together decide the length of walker vectors $\sqrt{\alpha^2+\beta^2}$ modeling distinctiveness, with greater β giving greater distinctiveness. α and β together also decide the maximum angle $\leq 90°$ geometrically achievable between walker vectors within genders modeling identity difference. For example, $\beta = 0$ would give a maximum angle of $0°$ since all male walker vectors would equal $\alpha\mathbf{g}$ and all female walker vectors $-\alpha\mathbf{g}$, giving no identity difference. The relationship between maximum angle and β for a given α is complex but what is important is that it increases with increasing β up to the $90°$ limit. Thus, greater β gives greater identity difference. Since by both distinctiveness and identity difference greater β gives easier identification, it was possible to calibrate β in pilot runs by increasing it subject-by-subject until subjects reached 90%+ identification performance around the end of the training phase. In this way $\beta = 1.8$ in units of standard deviations of the cmPC RUB100 walkers along the gender axis was chosen.

Notes on Experiment Design: Learning Effects

A limitation of our design may be the propensity for a variety of learning effects to impact the results due to the fact that caricatures of the same walker, and even the same particular caricatures, have to be shown multiple times throughout the test phase to accumulate an adequate number of trials. For example, while a caricature that is too extreme will be difficult to identify, it could at least be easy to recognize as the same

identity as an intermediate caricature (closer to it than the veridical). If that intermediate caricature is in turn easy to identify, subjects could deduce the identity of the extreme caricature via the intermediate after having seen both caricatures and the veridical. By this logic we would expect performance at extreme caricatures, while still probably measured as diminished, to be overestimated. This is only one example of unwelcome learning effects, some of which may be subtle and unforeseen.

Ultimately in designing a caricature identification experiment with biological motion stimuli there is a trade-off between practicality and thoroughness in the training phase and learning effects in the test phase, with the balance decided by the number of walkers trained on and subsequently tested. Ideally one wants enough stimuli that each caricature, preferably even each identity, need never be repeated in the test phase, but while still accomplishing effective training. The closest published experiment to ours is Hill & Pollick (2000) which investigated caricatures of point light arms in drinking motion – it tested six identities and so would have contended with the same learning effects.

Experiments testing face caricatures, on the other hand, tend not to have to contend with the trade-off because face stimuli are much easier to train on (probably due to a combination of rich textural information and greater perceptual salience), allowing a much greater number to be used. For example, Byatt & Rhodes (1998) were able to test 10 different caricatures each of 20 faces with no repetition, Rodriguez et al. (2008) 20 caricature faces plus 20 distracters in an old/new task, and Mauro & Kubovy (1992) even tested 100 faces plus 100 distracters with each trained face shown once veridical and once caricaturized. Face caricature experiments can also take advantage of faces that are

already familiar, such as celebrity faces or faces of people within a workplace where subjects are drawn from, to eliminate the training phase altogether while retaining a large number of test faces – see for example Rhodes et al. (1987) and Lee et al. (2000).

CHAPTER 6. BEHAVIOURAL EXPERIMENT 3

MEASUREMENT OF DISCRIMINATION OF OPTIMAL CARICATURES IN CMPC VERSUS ETSPC REPRESENTATION SCHEMES

6.1. Introduction

At $\gamma = 2.4$, found to be the optimal caricature level in the previous experiment, caricatures in cmPC versus etsPC schemes when generated from common veridicals were subjectively identical to us. Even at the highest tested $\gamma = 3.8$ we felt unable to see any difference. In fact, we had to push γ to about double that before we could, and even then the differences were subtle and only clear in the extremities of the limbs. Such caricatures are far beyond the optimal level and, if the downward trend of identification performance with level measured in the previous experiment continues, would not improve and perhaps would even be detrimental to identification. Thus, it seemed that our originally planned experiment seeking a difference in identification benefit from cmPC versus etsPC caricatures was not feasible. The purpose of this experiment is to verify this by showing in a controlled manner that $\gamma = 2.4$ cmPC and etsPC caricatures are indistinguishable to subjects when generated from common veridicals.

The basis of the experiment is a concurrent match to sample paradigm. In each trial subjects are shown a sample walker and a pair of potential match walkers and are asked to respond with which of the potential match walkers is the same as the sample, thus testing their ability to distinguish the potential match walkers. One of the potential match walkers is an etsPC caricature while the other is a cmPC caricature. We expected the task of distinguishing analogous (i.e. common level and veridical) cmPC and etsPC

caricatures to be impossible and that if these comprised all of the pairs of potential match walkers subjects might lose motivation after a few trials and stop making a genuine effort. Accordingly, the partner of each etsPC caricature is drawn from a set of cmPC caricatures ranging from its analogous caricature to a caricature generated by rotation of the analogous caricature vector 70° in cmPC space, and including caricatures represented by vectors at increments of 10° between. Each rotation of 10° gives a stimulus that is successively more different from the analogous caricature. All possible cmPC caricatures are counterbalanced and randomized across trials. Thus, trial difficulty samples a continuum from what we predicted to be practically impossible to very easy. The trials primarily of interest are the ones with analogous cmPC and etsPC caricatures, but having the randomized range of difficulties encourages subjects to make a genuine effort at every trial. The concurrent match to sample paradigm and a frontal orientation for all stimuli are used in order to make the task as easy as possible. The idea is to show that even under such ideal conditions subjects cannot distinguish between analogous cmPC and etsPC caricatures.

6.2. Methods

6.21. Subjects

Twelve subjects participated in this experiment. All were students in psychology or neuroscience at Queen's University. Subjects not in our lab received course credit for their participation. There were six females and six males. All had normal or corrected to normal vision.

6.22. Stimuli

Stimuli were based on four veridical walkers – two males and two females. As in the previous experiment, for each subject these walkers were randomly generated in the 15-marker set in the cmPC scheme and equalized for distinctiveness. Caricatures were then created by multiplying by $\gamma = 2.4$. Rotated versions of each of these caricature walker vectors were then also generated in increments of $10°$ up to $70°$ in a random plane in the 10-dimensional walker space. For each walker the plane of rotation was defined by its vector and the orthogonal component of a secondary randomly generated vector in the space. The method for rotation is exactly as for conventional 2D or 3D rotations – see Appendix 3. A total of 32 distinct cmPC caricatures resulted.

etsPC caricatures were created from the four original veridical walkers. To do this, the four walkers were converted to etsPC representations (cmPC → cmF → cm → ets → etsF → etsPC). The cm → ets step requires at least three markers per segment, so at the cmF stage walkers were extrapolated to 61 markers. The same technique was used in the previous experiment as part of conversion to the 15clav-marker set (Section 5.22); mathematical details can be found in Appendix 2. etsPC caricatures were then created by multiplying by $\gamma = 2.4$ in etsPC space.[2]

Note that a difference between an analogous cmPC and etsPC caricature could potentially arise in three ways: (1) a difference in the mean walker relative to which

[2] With no loss of information in the etsF → etsPC conversion, an exactly equivalent way to produce these caricatures would be to create them in the etsF scheme as $\mathbf{w\_etsF}_\gamma = \gamma\left(\mathbf{w\_etsF} - \overline{\mathbf{w\_etsF}}\right) + \overline{\mathbf{w\_etsF}}$. Indeed, the topology of etsPC space is already present in the etsF scheme short of the whitening operation in the etsF → etsPC step (Section 2.53 equation 7) which is linear and cancels out, and the only dimensions of etsPC space that are needed for these caricatures are along the four walker vectors themselves. Thus, conversion all the way to etsPC is extraneous. However, we explain our procedure in terms of the etsPC scheme for conceptual coherence with the procedure for cmPC caricatures.

caricatures are generated resulting from averaging the RUB100 walkers in the cmF versus etsF scheme, (2) a difference in how the caricaturization process of increasing the length of the walker vector in the cmPC versus etsPC space changes the physical characteristics of the walker, and (3) a difference in the veridical walker due to loss of information in conversion from cmPC to etsPC. The first two, which are due to how the topology of cmPC versus etsPC spaces impacts linear operations such as averaging and scaling of vectors, are of interest here. The second is particularly of interest since the first turns out to be negligible (data not shown). The third is not of interest here and to minimize it two steps were taken: First, at the ets → etsF step a full complement of harmonics as opposed to the standard two were retained to bring error at Fourier decomposition to near zero. Second, when computing the etsPC scheme we explicitly included our four veridical walkers in the training set in case they fell slightly outside the space spanned by the etsF RUB100 training set, and we also retained all PC's, thus ensuring that no error was acquired at PC decomposition.

In total we arrived at 36 distinct stimuli: four etsPC caricatures, four analogous cmPC caricatures, and for each analogous cmPC caricature a cmPC caricature represented by a vector rotated 10, 20, 30, 40, 50, 60, and 70° in the space. As in the previous experiment, all stimuli were shown as stick figures in the 15clav-marker set. For cmPC stimuli 15clav-marker displays were extrapolated as described in Section 5.22. etsPC stimuli were already in the 61-marker set, so for them we simply took the 15clav-marker subset for display. Frontal orientation was used for all stimuli at display.

6.23. Procedure

In each trial an etsPC caricature and one of the corresponding set of cmPC caricatures – together the pair of potential match walkers – were shown and one of the pair – the sample walker – was simultaneously shown above. There were approximately 8° of horizontal visual angle between the potential match walkers and 6° of vertical angle between the sample and the potential match walkers. A left/right arrow prompt was displayed at the bottom of the screen and subjects had to respond with the left or right arrow key indicating which potential match walker was the same as the sample. Which of the potential match walkers was on the right or the left and which was the sample walker was randomized and counterbalanced across trials. In order to prevent some subjects from studying the walkers for much longer than others, the sample walker disappeared after 10 s; however, each trial ended only when input was received. Subjects were instructed to use any features except gait phase and to make their choices as fast as possible without unreasonably sacrificing accuracy. Inter-trial interval was 0.5 s. A total of 128 trials consisting of four of each cmPC caricature were performed in random order. 10 random discounted practice trials were also performed at the beginning with supervisor present and coaching as needed to ensure that subjects understood the paradigm. Subjects typically took between 15 and 20 minutes to complete the 128 trials.

6.3. Results and discussion

Refer to Figure 23 below. Although it is encouraging to see a good psychometric trend emerge across cmPC caricature vector rotations, the principle point of interest was the unrotated 0° point at which subjects had to distinguish between analogous cmPC and

etsPC caricatures. Performance at that point was $51 \pm 3\%$. A *t*-test of the $0°$ data was unable to reject the null hypothesis of a mean performance of 50% (chance), $t(11) = 0.484,\ P = 0.319$, meaning we were unable to show that subjects can distinguish analogous cmPC and etsPC caricatures at the optimal $\gamma = 2.4$ level. Note that we rigorously tested only 15clav-marker stick figures here, which as per the previous experiment seemed the best stimuli for studying the caricature effect, but subjectively caricatures were similarly indistinguishable as PLD's in all marker sets.

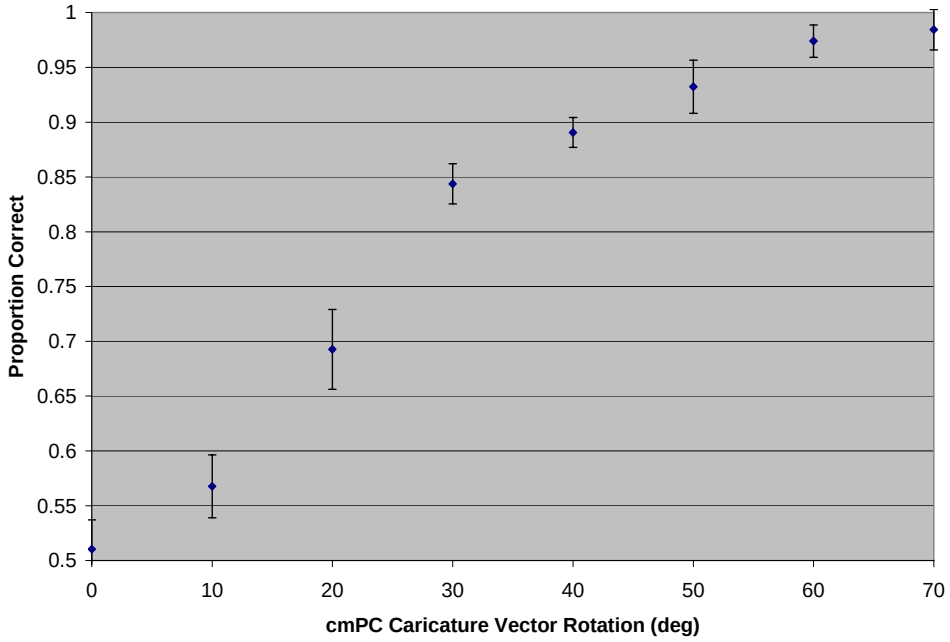

Figure 23. Proportion correct vs. cmPC caricature vector rotation. Error bars are standard error of the mean with between-subjects variance removed.

CHAPTER 7. GENERAL DISCUSSION

Summary

In our computational experiment the Cartesian coordinates of markers Fourier-based scheme was found to linearize walker data better than the Euler angles and translations of segments Fourier-based scheme as measured by efficiency of PC decomposition, and the subsequent cmPC scheme was found to be more efficient overall than the etsPC. In our behavioural experiments an increase in identification performance for cmPC caricatures was observed up to an optimal caricature level of 2.4. This corresponds to a mean caricature vector length of 6.2 standard deviations of projections of RUB100 walkers or 2.1 mean RUB100 walker vector lengths – i.e. about double the distinctiveness of a typical walker. A declining benefit was observed thereafter up to the highest measured level of 3.8. Our originally planned main behavioural experiment was a comparison of caricatures generated in the cmPC versus etsPC scheme at a common level in the hope of finding that caricatures generated in one benefit identification more than caricatures generated in the other, from which we would conclude the former to be a better model of the true high level perceptual scheme for walkers. However, we found that caricatures between the two schemes when generated from common veridicals were distinguishable only at levels three or four times higher than optimal, which judging from the observed downward trend of identification performance with level would not be beneficial. Specifically, in a concurrent match to sample experiment we showed that subjects are unable to distinguish caricatures between the two schemes when generated from common veridicals at the optimal level. As such, our originally planned comparison of caricatures became infeasible and we were unable to relate our two schemes for data in the lab to the

high level scheme for walker sense data in the human perceptual system, unable to conclude whether the more efficient of the two schemes is the better model of the true perceptual scheme, and unable to conclude the presence or absence of a simplicity principle in the perception of human walking motion.

Appropriate Choice of Segment-Based Representation Scheme

One potential issue to mention is the convention used for establishing segment basis vectors in the ets scheme – i.e. under what orientation the coordinate system for each segment is rigidly attached to the physical body segment it is modeling. Our current convention is to choose basis vectors such that the mean orientation of each segment across frames is equal to the mean orientation of its parent (Section 2.54 equations 9 and 10), with the root segment mean-centred to the body-centred/cm basis vectors chosen with the z axis up and walking motion in the direction of the x axis. But the choice of convention is open and can have a dramatic effect on ets data, since for each segment the basis vectors are the axes of rotation about which the Euler angles in its segment pose part are calculated, as well as the orientation of the coordinate system in which the translations of any child segments are calculated. This can in turn affect Fourier decomposition of ets data in the etsF scheme and PC decomposition in the etsPC scheme. In fact, we initially tried several other arbitrary conventions and for some efficient Fourier decomposition became all but impossible due to fast fluctuations and even discontinuities in ets Euler angle data over time. We settled on our current convention purely on the grounds of centring Euler angle variance about zero to facilitate taking arithmetic means when calculating etsF representations (Section 2.55), and it did lead to

fairly efficient Fourier and PC decomposition. But there is no guarantee that this choice is optimal.

Indeed, one reason to suspect that it is not is the fact that our ets data does not reflect basic motion constraints at certain joints in the body. To a very good approximation many joints have fewer than three rotational degrees of freedom. For example, the hinge and pivot elbow joints can be modeled with just two degrees of rotational freedom, as can the glide joint of a wrist, and the hinge joint of a knee can be modeled with one. It should be possible to represent the orientation over time of, say, a forearm with respect to its upper arm with only two angles with very little error. This would directly decrease ets and etsF dimensionality and may thereby improve etsPC efficiency. But under our convention there is significant variance in all three Euler angles for all segments. This is a consequence of the fact that, while our sets of three Euler angles model joint orientations, our axes of rotation for individual Euler angles do not model the physical axes of rotation at joints.

Another issue in the context of perception in particular is that our Euler angles are counterintuitive to interpret or visualize. For example, even explicitly given the basis vectors for a segment and its parent and given three Euler angles for the segment from some arbitrary frame of animation, it would be impossible to visualize the relative orientation of the two segments without working through the equations of Section 2.54. In fact this is not an uncommon characteristic of Euler angles in general; they are in industry standard for representing rotations more because they are time-tested, well understood, widely used in many practical contexts including robotics and aerospace engineering, and above all mathematically easy to work with (relatively speaking). In the

context of perception these credits may not be very convincing. The fact that we are not able to intuitively interpret our Euler angles may cast doubt on their perceptual saliency. While using some alternate scheme for representing segment orientation that gives readily interpretable angles would be no guarantee that that scheme is more promising as a model of the true perceptual scheme for walkers, the obtuseness of our ets scheme is troublesome. Altogether, then, we may be better to test a scheme that gives appropriate rotational degrees of freedom by better corresponding to the physics of body joints, and that has some intuitive appropriateness. This might be an Euler angle scheme under some other convention for attaching segment basis vectors or some other kind of segment-based scheme altogether.

Constraining Segment Translations

The issue of rotational degrees of freedom mentioned above raises a corresponding issue: that of translational degrees of freedom. Many of the segments in the ets scheme have very little variance in translation components because there is little play in the joints connecting them to their parent segments. For example, translation for a forearm segment is approximately constant across frames, with the corresponding constant distance interpreted as the length of the upper arm. Unlike constraints on rotational degrees of freedom, constraints on translational degrees of freedom are typically absolute in the sense that almost no change in translation occurs along any degree. As such, our choice of segment basis vectors is irrelevant to them since translation along every one is close to constant regardless of their orientation. Still, ets and etsF dimensionality and thus potentially etsPC efficiency is expended on all three translation components for low-translational variance segments. Future work could look into discarding such elements.

This practice, as well as accurately modeling joints with less than three degrees of rotational freedom using less than three angles, would amount to one more way of trading off dimensionality for error. And since the error acquired may be very small, it is something worth looking into. Constraining translations at least is already standard practice in computer character animation, typically accompanied by compensatory recalculation of Euler angles to minimize error acquired as measured in cm reconstructions.

Effect of Reducing Pre-PCA Degrees of Freedom on PC-Based Representation Schemes, and Alternate Methods of Dimensionality Reduction

If the information inherent in the structure and motion of an elbow joint can be modeled well by only one rotational and no translational degrees of freedom but three rotational and three translational degrees of freedom are redundantly used as basic coordinates a subsequent process of dimensionality reduction should *ideally* isolate and eliminate that redundancy. The end result would be no difference in efficiency between a PC-based scheme coming from the clever 1DOF basic scheme and one from the wasteful 6DOF basic scheme. Final efficiency would be a function of the degrees of freedom in the underlying information, not the superficial degrees of freedom in whatever representations for that information we happen to feed into the dimensionality reduction. After all, the fact that ets representations are typically shorter than cm for 45-marker walkers was no prior guarantee that the etsPC scheme would be more efficient than the cmPC (and it turned out not to be).

But this is only in the ideal sense. PCA can detect redundancy only when it is couched in certain ways – namely in ways that can be removed by the affine combination

of change of vector space origin and orthogonal basis. Non-linear redundancies may remain following PCA. Indeed, if PCA did work at the level of true degrees of freedom in underlying information then choice of any basic coordinate scheme – cm-based, ets-based, etc. – would have no impact on PC-based scheme efficiency whatsoever, at least provided they all capture identical information. This is not the case, which is why our comparison of cmPC versus etsPC scheme efficiencies was a reasonable idea. Thus, using a more effective version of or alternative to the ets scheme in which some segments can be represented with fewer than six degrees of freedom does have the potential to benefit subsequent PC-based scheme efficiency by further pre-reducing redundancy that PCA is unable to find.

In a similar vein, another possibility to explore is the use of methods of dimensionality reduction other than PCA in the implementation of the norm-based walker space model of walker perception. Some non-linear dimensionality reduction methods may be more efficient and even facilitate a different balance of efficiencies between cm- and ets-based schemes. Even sticking to PCA, the whitening step (Section 2.53 equation 7) can be implemented in many different ways to have an effect. For one example, rather than whitening by dividing by the standard deviations of the pre-PCA coordinates across training vectors we could divide by the means of standard deviations with the same unit of measurement (akin to how we scaled relative RMS errors in our computational experiment). Pilot data from such group whitening show improvement in efficiency for both cmPC and etsPC schemes.

Problem of Sample Size

There is little difference in representation efficiency between specific walkers, as judged from the vanishingly small error bars in all of the plots which are averaged across walkers in the computational experiment. While this adds credence to representing walkers under unified schemes specialized for the class and averaging errors across walkers to obtain meaningful errors for those schemes, it leaves us with a sample size problem for purposes of inferring a conclusion for a simplicity principle. For example, even if there were no relationship between scheme efficiency and perception, we would still have a 50/50 chance of obtaining the false positive result that the more efficient of our two schemes is more perceptually salient. Still, it should be noted that if there were a relationship then we would expect the computational results to match the behavioural. And since we expect a simplicity principle to hold, any negative finding – whether true or false – would therefore still be of legitimate interest. Since our caricature approach did not work, this consideration is academic for now.

Impact of Reconstruction Scheme on Measured Error – Potential Alternative to the Caricature Approach

By way of looking for a potential alternative to our defunct caricature approach, consider Figure 24 below. This is a graph of error at Fourier decomposition alone in etsF walkers as a function of number of harmonics retained. Two types of errors are plotted: relative RMS error in ets reconstructions (original ets → etsF → ets reconstruction)[3] and relative RMS error in cm reconstructions (original ets → etsF → ets → cm reconstruction),

[3] In calculating etsF representations the marker position part of ets data does not acquire error, so the fact that it cannot be included in calculation of relative RMS error in ets reconstructions does not preclude the use of that reconstruction scheme as it did in the computational experiment for etsPC representations.

calculated using methods analogous to those in our computational experiment, including Section 3.2 equation 22 used for calculation of relative RMS error in both reconstruction schemes. Since the ets → cm conversion is lossless, reconstructed information is identical between the two schemes.

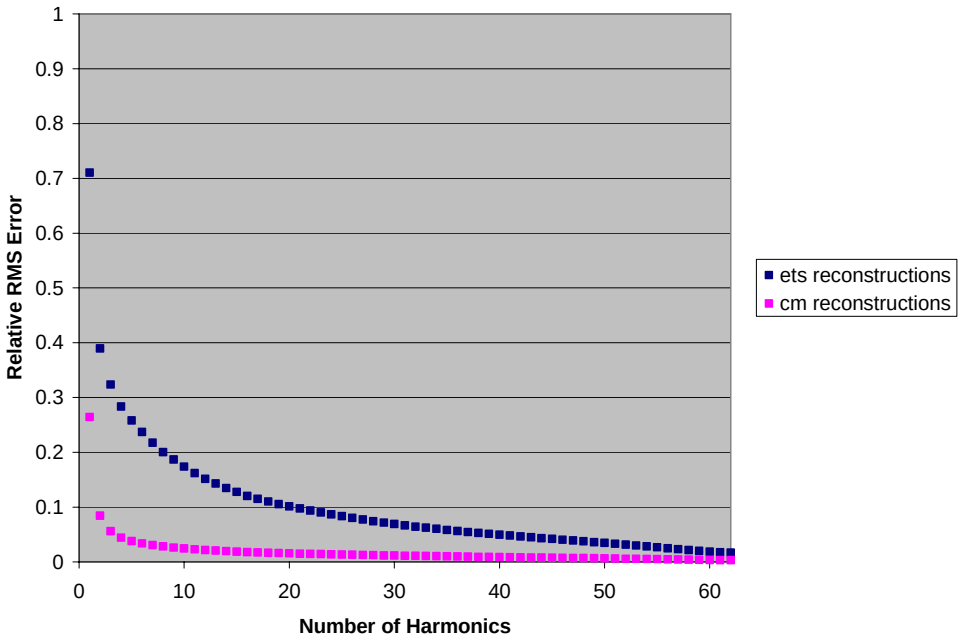

Figure 24. Relative RMS error (ets and cm reconstructions) vs. number of harmonics for the etsF scheme

We focus here only on the fact that error measured in cm reconstructions is considerably different from in ets reconstructions. With this in mind, it might be possible to construct in a controlled manner walkers that have high error as measured in the cm scheme while simultaneously low error as measured in the ets, and in parallel other walkers that have low and high measured error oppositely. If so, we could test subjects in a psychophysical experiment to see which kind of error they are better able to detect, and deduce from that which basic coordinate scheme the perceptual system is more sensitive

to. For now the idea is just a sketch, but it may be a seed for a legitimate alternative to the caricature approach.

REFERENCES

Abramson, N. (1963). *Information Theory and Coding.* New York, USA: McGraw-Hill Book Company

Attneave, F. (1954). *Some informational aspects of visual perception.* Psychological Review 61(3), 183-193

Aguilera, A. & Perez-Aguila, R. (2004). General n-dimensional rotations. *WSCG SHORT Communication papers proceedings*

Boselie, F. (1988). Local versus global minima in visual pattern completion. *Perception and Psychophysics* 43, 431-445

Boselie, F. & Wouterlood, D. (1989). The minimum principle and visual pattern completion. *Psychological Research* 51, 93-101

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision* 10, 433-436

Buffart, H., Leeuwenberg, E., & Restle, F. (1981). Coding theory and visual pattern completion. *Journal of Experimental Psychology: Human perception and Performance* 7(2), 241-274

Byatt, G. & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: implications for models of face recognition. *Vision Research* 38, 2455-2468

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review* 103(3), 566-581

Chater, N. (1997). Simplicity and the mind. *The Psychologist* 10(11), 495-498

Chater, N. (1999). The search for simplicity: a fundamental cognitive principle? *The Quarterly Journal of Experimental Psychology* 52A(2), 273-302

Chater, N. & Vitanyi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences* 7(1), 19-22

Cutting, J.E. (1978). A program to generate synthetic walkers as dynamic point-light displays. *Behaviour Research Methods and Instrumentation* 10(1), 91-94

Cutting, J.E. (1981). Coding theory adapted to gait perception. *Journal of Experimental Psychology: Human Perception and Performance* 7(1), 71-87

Cutting, J.E. & Proffitt, D.R. (1982). The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology* 14, 211-246

Davis III, R.B., Ounpuu, S., Tyburski, D., & Gage, J.R. (1991). A gait analysis data collection and reduction technique. *Human Movement Science* 10, 575-587

Diebel, J. (2006). Representing attitude: Euler angles, unit quaternions, and rotation vectors. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.5134&rep=rep1&type=pdf

Dixon, W.J. & Mood, A.M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association* 43, 109-126

Frowd, C., Bruce, V., Ross, D., McIntyre, A., & Hancock, P.J.B. (2007). An application of caricature: how to improve the recognition of facial composites. *Visual Cognition* 15(8), 954-984

Garcia-Perez, M.A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research* 38, 1861-1881

Garcia-Perez, M.A. (2000). Optimal setups for forced-choice staircases with fixed step sizes. *Spatial Vision* 13(4), 431-448

Giese, M.A., Thornton, I., & Edelman, S. (2008). Metrics of the perception of body movement. *Journal of Vision* 8(9):13, 1-18

Gramkow, C. (2001). On averaging rotations. *Journal of Mathematical Imaging and Vision* 15, 7-16

Hatfield, G. & Epstein, W. (1985). The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin* 97(2), 155-186

Hill, H. & Pollick, Frank E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science* 11(3), 223-228

Hochberg, J. & McAlister, E. (1953). A quantitative approach to figural "goodness". *Journal of Experimental Psychology* 46(5), 361-364

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percetion and Psychophysics* 14(2), 201-211

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research* 38, 379-393

Jolliffe, I.T. (2002). *Principal Component Analysis.* New York, USA: Springer-Verlag, Inc.

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification. *Psychological Science* 11(5), 379-385

Leeuwenberg, E.L.J. (1971). A perceptual coding language for visual and auditory patterns. *American Journal of Psychology* 84(3), 307-349

Mardia, K.V. (1972). *Statistics of Directional Data.* London, UK: Academic Press Inc. (London) Ltd.

Mauro, R. & Kubovy, M. (1992). Caricature and face recognition. *Memory & Cognition* 20(4), 433-440

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* 10, 437-442

Pollick, F.E., Fidopiastis, C., & Braden, V. (2001). Recognising the style of spatially exaggerated tennis serves. *Perception* 30, 323-338

Restle, F. (1979). Coding theory of the perception of motion configurations. *Psychological Review* 84(1), 1-24

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: implications for mental representations of faces. *Cognitive Psychology* 19, 473-497

Rhodes, G., Robbins, R., Jaquet, E., McKone, E., Jeffery, L., & Clifford, C.W.G. (2005). Adaptation and face perception – how aftereffects implicate norm-based coding of faces. In Clifford, C.W.G & Rhodes, G. (eds.) *Fitting the Mind to the World*. Oxford, UK: Oxford University Press, 213-240

Rodriguez, J., Bortfield, H., & Gutierrez-Osuna, R. (2008). Reducing the other-race effect through caricatures. *Proc. 8th IEEE International Conference on Automatic Face and Gesture Recognition,* 17-19

Shannon, C.E. (1948). The mathematical theory of communication. *Bell System Technical Journal* 27, 379-423, 623-656

Troje, N.F. (2002). Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *Journal of Vision* 2, 371-387

Troje, N.F. (2008). Retrieving information from human movement patterns. In Shipley, T.F. & Zacks (eds.) *Understanding Events: From Perception to Action*. Oxford, UK: Oxford University Press, 308-334

Turk, M. & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1), 71-86

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology* 43A(2), 161-204

Van der Helm, P.A. (2000). *Simplicity versus likelihood in visual perception: from surprisals to precisals.* Psychological Bulletin 126(5), 770-800

Vetter, T. & Troje, N.F. (1997). Separation of texture and shape in images of faces for image coding and synthesis. *Journal of the Optical Society of America A* 14(9), 2152-2161

Yuille, A. & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences* 10(7), 301-308

APPENDICES

1. Generating random walkers

The problem is to randomly generate a number of 15-marker walkers in the cmPC scheme. The projections of males onto the gender axis are to be equal and parameterized by a number α and the projections of females are to be equal and parameterized by $-\alpha$. The gender axis is represented by a unit walker vector **g** and is the axis calculated by a linear discriminant classifying cmPC RUB100 walkers according to the binary male/female feature, with increasing projections of walker vectors onto the axis on one side of the origin optimally corresponding to increasing perceived masculinity and on the other side to femininity. Refer to Troje (2002) for details on calculation of the gender axis. The perpendicular distance from the gender axis to all walkers – essentially the component of walkers capturing all features other than gender – is to be parameterized by a number β. The four walker vectors of each gender are to be generated as close to orthogonal as possible. Whether angles of $90°$ are geometrically possible or whether some angle $< 90°$ is the largest that can be achieved depends on the chosen values of α and β. Aside from these constraints, walker vectors are to be generated randomly.

We generate males independently of females. Beginning with the males, we work in an alternate orthonormal basis **Am** for the 10-dimensional cmPC space whose first basis vector $\mathbf{am}_1$ = the gender axis **g** and whose remaining vectors $\{\mathbf{am}_2, \mathbf{am}_3, \ldots, \mathbf{am}_{10}\}$ are random under the constraint of orthonormality of the set. The set is generated by the Gram Schmidt process applied to **g** and nine random vectors. Males $\{\mathbf{m}_j\}$ are then defined as follows:

$$\begin{cases} \mathbf{m}_1 = \alpha\mathbf{g} + \beta\mathbf{am}_2 \\ \mathbf{m}_2 = \alpha\mathbf{g} + k_{22}\,\mathbf{am}_2 + k_{32}\,\mathbf{am}_3 \\ \mathbf{m}_3 = \alpha\mathbf{g} + k_{23}\,\mathbf{am}_2 + k_{33}\,\mathbf{am}_3 + k_{43}\,\mathbf{am}_4 \\ \mathbf{m}_4 = \alpha\mathbf{g} + k_{24}\,\mathbf{am}_2 + k_{34}\,\mathbf{am}_3 + k_{44}\,\mathbf{am}_4 + k_{54}\,\mathbf{am}_5 \end{cases} \tag{23}$$

Coefficients k_{ij} are solved for each male j consecutively beginning with male 2. At each male, the following constraints are used:

$$\begin{cases} m_{j-1} \cdot m_j = \sqrt{\alpha^2 + \beta^2}\,\cos\theta \\ m_{j-2} \cdot m_j = \sqrt{\alpha^2 + \beta^2}\,\cos\theta \\ \vdots \\ m_1 \cdot m_j = \sqrt{\alpha^2 + \beta^2}\,\cos\theta \end{cases} \tag{24}$$

where θ is the angle between every pair of male walker vectors. These equations are

$$\begin{cases} \alpha^2 + k_{2,j-1}\,k_{2j} + k_{3,j-1}\,k_{3j} + \ldots + k_{j-1,j-1}\,k_{j-1,j} + k_{j,j-1}\,k_{jj} = \sqrt{\alpha^2 + \beta^2}\,\cos\theta \\ \alpha^2 + k_{2,j-2}\,k_{2j} + k_{3,j-2}\,k_{3j} + \ldots + k_{j-1,j-2}\,k_{j-1,j} = \sqrt{\alpha^2 + \beta^2}\,\cos\theta \\ \vdots \\ \alpha^2 + k_{21}\,k_{2j} = \sqrt{\alpha^2 + \beta^2}\,\cos\theta \end{cases}$$

or

$$\begin{cases} k_{2,j-1}\,k_{2j} + k_{3,j-1}\,k_{3j} + \ldots + k_{j-1,j-1}\,k_{j-1,j} + k_{j,j-1}\,k_{jj} = \sqrt{\alpha^2 + \beta^2}\,\cos\theta - \alpha^2 \\ k_{2,j-2}\,k_{2j} + k_{3,j-2}\,k_{3j} + \ldots + k_{j-1,j-2}\,k_{j-1,j} = \sqrt{\alpha^2 + \beta^2}\,\cos\theta - \alpha^2 \\ \vdots \\ k_{21}\,k_{2j} = \sqrt{\alpha^2 + \beta^2}\,\cos\theta - \alpha^2 \end{cases} \tag{25}$$

which is a system of $j-1$ linear equations which can be solved for the $j-1$ unknowns $\{k_{2j}, k_{3j}, \ldots, k_{jj}\}$ (all k_{ij}'s for preceding males have already been solved and thus play

the role of coefficients). The remaining $k_{j+1,j}$ can then be solved using the remaining constraint:

$$\sqrt{k_{2j}^2 + k_{3j}^2 + \ldots + k_{j+1,j}^2} = \beta \tag{26}$$

All that is missing in the above is to know the angle θ closest to $90°$ that is possible under the chosen α and β. Rather than solve this maximization problem, it is adequate for our purposes to iterate the above procedure at values of θ beginning at $90°$ and decreasing by $1°$ until the largest workable angle is found. At angles larger than this the final constraint in equation 26 does not provide a real solution for $k_{j+1,j}$. Note that the random aspect of the walkers is entirely due to the randomness of the underlying basis **Am**.

To generate the females the same procedure is repeated beginning with a new random basis **Af** – the only difference being that for females the coefficient of **g** is $-\alpha$ instead of α (and this difference is squared out in equations 24 through 25). Due to the different random bases, angles between any given pair of male and female walkers are random. Once generated, a basis change from **Am** and **Af** to **cmPC** gives the generated walkers in the latter basis as desired – **Am** and **Af** bases need not be retained.

2. Extrapolating 61-marker from 15-marker walkers

The problem is, given a set of 15-marker walkers in the cmF scheme, to extrapolate corresponding 61-marker walkers using the cmF 15-marker and corresponding 61-marker RUB100 walkers as exemplars.

In our application each of the original walkers came from the cmPC scheme. The cmPC basis vectors are the first 10 PC's of the cmF 15-marker RUB100 walkers and therefore span a subspace of the space spanned by those walkers. So, all of the original walkers can be expressed exactly as a linear combination of the RUB100 walkers:

$$\mathbf{W15} = \mathbf{RUB10015} \cdot \mathbf{W15}_{RUB10015} \tag{27}$$

where $\mathbf{W15}$ is a matrix containing the original 15-marker walkers as column vectors, $\mathbf{RUB10015}$ is a matrix containing the 15-marker RUB100 walkers as column vectors, and $\mathbf{W15}_{RUB10015}$ is a matrix to be determined containing the original walkers in the $\mathbf{RUB10015}$ basis as column vectors. This is several (the number of original walkers) systems of 226 linear equations in 100 unknowns each. $\text{rank}(\mathbf{RUB10015}) = 100$, so each system of linear equations can be solved to give a unique solution for $\mathbf{W15}_{RUB10015}$. Once we have this, we extrapolate to $\mathbf{W61}$ by assuming that the corresponding 61-marker walkers are expressed as the same linear combination of the 61-marker RUB100 walkers as the original 15-marker walkers are of the 15-marker RUB100 walkers – i.e. that $\mathbf{W61}_{RUB10061} = \mathbf{W15}_{RUB10015}$. Then,

$$\mathbf{W61} = \mathbf{RUB10061} \cdot \mathbf{W61}_{RUB10061} = \mathbf{RUB10061} \cdot \mathbf{W15}_{RUB10015} \tag{28}$$

One positive quality of these extrapolations is that, because the original walkers are in the space of the RUB100 walkers and the systems of linear equations (27) are therefore consistent, the elements of $\mathbf{W61}$ corresponding to the original 15 markers (which are a subset of the 61-marker set) are identical to those elements in $\mathbf{W15}$,

meaning the data for those 15 markers remain unchanged in the 61-marker extrapolations.

To be fair, the first 10 PC's underlying the cmPC scheme from which the original walkers came actually span a subspace of the cmF 15-marker RUB100 walkers *mean-centred and whitened,* and so the conclusion to be drawn at first glance is only that every one of the mean-centred and whitened original cmF 15-marker walkers can be expressed as a linear combination of the mean-centred and whitened cmF 15-marker RUB100 walkers. However, our method of whitening is a linear transformation (Section 2.53 equation 7) and so does not affect linear combinations. Mean-centring is not a linear operation and so can affect linear combinations. If we had worked with all mean-centred walkers in both equations above, the rank of the matrix of exemplar walker vectors would have equaled 99 rather than 100 and the systems of equations (27) would have been underdetermined, yielding infinite solutions. But it can be shown that if the Moore-Penrose pseudoinverse is used to isolate minimum norm solutions and the mean walker is added to the result of the extrapolation in equation 28 to arrive at a non-mean-centred $\mathbf{W61}$, the result is identical to above.

Also, note that scaling mean-centred **W15** walkers (such as by caricaturizing) simply scales the mean-centred $\mathbf{W15}_{RUB10015}$ by the same amount and thereby also the mean-centred $\mathbf{W61}$. The upshot is that this method works just the same for scaled data, meaning it extrapolates markers that are scaled equivalently to the existing 15.

It should be noted that this method could be used for 15-marker walkers that are not exactly in the space spanned by the RUB100 walkers. The systems of equations (27) would become inconsistent but a least squares solution could still be computed. The

quality of the extrapolation would depend on how appropriate the exemplar walkers are for the walkers being extrapolated, and may be improved by prior dimensionality reduction (such as by PCA) to avoid overfitting.

3. Rotating a vector in n-dimensional space

Rotation of a vector can be generalized from 2D or 3D space to *n*-D space by defining a rotation as a linear transformation that preserves the norm of a vector and leaves the resultant at an angle of rotation θ to the original where both vectors share a (2D) plane of rotation whose normal is the axis of rotation. A number of methods for calculation of rotation matrices in arbitrary dimensions are published – see for example Aguilera & Perez-Aguila (2004). Since our only requirement was to calculate resultants of rotations of specific vectors, we could take a more direct geometric approach. Given an original vector **v** with norm *v*, a vector **d** orthogonal to **v** that together with **v** defines the plane of rotation, and an angle of rotation θ in the initial direction of **d**, the resultant $\mathbf{v}'$ can be calculated just as in 2D or 3D contexts:

$$\mathbf{v}' = v \sin\theta \hat{\mathbf{d}} + v \cos\theta \hat{\mathbf{v}} \tag{29}$$