To appear in *Proc. Comp. Vis. and Pat. Rec.* **June, 2000.**

Categorical Representation and Recognition of Oscillatory Motion Patterns

James Davis
Media Laboratory
Mass. Inst. of Tech.
Cambridge, MA 02139
jdavis@media.mit.edu

Aaron Bobick
College of Computing
Georgia Inst. of Tech.
Atlanta, GA 30332
afb@cc.gatech.edu

Whitman Richards
AI Lab, Brain and Cog. Sci.
Mass. Inst. of Tech.
Cambridge, MA 02139
wrichards@mit.edu

Abstract

Many communicative behaviors in the animal kingdom consist of performing and recognizing specialized patterns of oscillatory motion. Here we present an approach to the representation and recognition of these oscillatory motions based on the categorical organization of a simple sinusoidal model having very specific and limited parameter values. This characterization is used to specify the types and layout of computation for recognizing the patterns. Results of the method are demonstrated with real oscillatory motions showing the viability of a structured categorical framework.

1. Introduction

Many communicative behaviors throughout the animal kingdom consist of performing and recognizing specialized patterns of motion. In this domain, oscillatory movements are quite prevalent. For example (as shown in Fig. 1), Mallard ducks bob their head *up-and-down* to a female during courtship [22], honey bees dance in a *circle* or *figure-8* to signal a food source to the hive [24], and people *nod* or *wag* their head as a subtle responsive gesture. We present a simple sinusoidal model with regularized structure that categorically describes a set of common oscillatory motions. A simple-to-complex ordering for the motions is created using the number of model parameters per motion as a measure of complexity. This organization is reinforced by the general ubiquitousness of the simpler motions across species, but not of the complex motions. We also show that human generation of oscillatory motions suggest a particular form for the data representation. Results are demonstrated using motions from a special-purpose tracking system and real video. The categorical approach is offered as a means to organize and identify related motion patterns without the necessity of multiple training examples or individual ad hoc models.

1.1. Periodic motion research

In related work on recognizing periodic motion, much reliance on Fourier analysis of trajectory information is used. For example, Polana and Nelson [18] employ a template-based method registering periodicity values distributed across a tracked region. Similarly, Liu and Picard [13] use periodicity templates as a representation to localize and detect repetitive motion along temporal lines for each pixel region. Cutler and Davis [5] check a self-similarity measure between motion frames for detecting periodic motion. Tsai et al. [23] examine the Fourier transform of a curvature trajectory to recognize cyclic walking motion. Other research targeting periodic walking patterns includes Little and Boyd [12] who examine the phase relationships of periodic elements derived from optical flow for recognition and person identification. Non-Fourier approaches to cyclic motion analysis include Niyogi and Adelson's [17] approach of detecting braided walking patterns in spatio-temporal slices, Seitz and Dyer's method [21] for view-invariant analysis using a period trace, and Cohen et al.'s technique [3] using linear dynamic models. We continue the development of approaches for the analysis and recognition of repetitive motion, but instead offer a categorical method based on strong structural regularities found within a set of naturally occurring, communicative oscillatory motions.

2. Categorical motion recognition

The success of perception is intimately coupled with the ability to construct internal model representations whose assumptions and constraints reflect the proper structure and regularities present in the world [19]. In other words, fundamental to perception is the notion that there is indeed structure in the world that transfers to the visual image (similar to ecological optics of [8]). This suggests that if perceptible regularities exist for a class of oscillatory movements, they would have the effect of making sensory information of the motions highly reliable for the purpose of categorical perception. The approach here for recognizing oscillatory

Figure 1. Communicative oscillatory motions in nature. (a) Mallard duck head-bobbing during courtship. (b) Honey bee dancing in a circle and figure-8 to signal food. (c) Person gesturing his head 'No'.

motions develops an underlying categorical description for a set of related motions, where the deep structural form of the category is presumed to be a parameterized motion description.

In opposition to a *structural* categorical approach is analysis and recognition from prototypes or training examples of the motions within generic high-dimensional parameter spaces. For example, simply template-matching trajectories to a prototype trajectory is an extremely context sensitive approach (even when using DTW). Other prototypical methods such as HMMs collect examples for a particular motion to construct a data-specific model for recognition. Though less sensitive than straight template matching, the model does not extend to a larger category of patterns. In methods such as PCA, exhaustive training examples for the entire *class* are collected and decomposed into a set of basis motions capturing the *appearance* of the data. In the case of oscillations, if the input primitives are the functional sinusoids, then linear decomposition methods do not capture the true underlying structural parameters (non-linear). Designing such specialized systems as mentioned here results in having only a set of mimicry devices with no understanding of the underlying processes and constraints [15].

Instead, with the notion of a category, the structural constraints for the class of motion are made explicit, and model generation for various categorical motions is made simple by choosing from a set of fixed parameter values and constraints within the category. Thus creating a new pattern model requires no training, unlike with training-based methods that demand multiple examples for each new model, and may be considered a form of one-shot learning [6]. This approach is also less sensitive to slight deviations in the signal (e.g. from viewing condition or stylistic performance) for it has a notion of parameterized structure that can be used to aid recognition. For example, the analysis can choose to possibly ignore some parameters and compare the relations (rather than absolute values) of the remaining parameters. Training-based approaches using the raw trajectory information, on the other hand, will have difficulty if they encounter a new version of the expected pattern not accounted for during training. Thus the training class would need to encompass all possible structural and stylistic properties that are capable of being produced.

In this research, we construct a recognition system relying on the structural model parameters within a single motion category. In Table 1 (col 1), we list a set of oscillatory motion patterns that are used for communication in the animal kingdom and that are commonly found in bird display behavior (determined after examining the behavioral literature on several hundred bird species). This motion set includes 1-D, 2-D, and translation oscillations.

3. Oscillatory motion regularities

Oscillations in general are quite common in biological systems, especially in animal locomotion [4] and even handwriting [9]. Such oscillations are frequently modeled with sinusoids. We follow suit and model all of the motions in Table 1 by sinusoids with constraints on form (e.g. a "figure-8") and variations in style (e.g. "fast" or "slow"). Also, an ordering established from the number of model parameters required to create each pattern (and from biological observations) can be used to rank the motions from simple to complex.

3.1. Sinusoidal model

We describe the oscillatory motions with three simple parameterized sine-wave generators of the form:

$$
\begin{aligned}
x(t) &= A_x \sin(2\pi f_x t + \phi_x) + B_x t \\
y(t) &= A_y \sin(2\pi f_y t + \phi_y) + B_y t \\
z(t) &= A_z \sin(2\pi f_z t + \phi_z) + B_z t
\end{aligned}
$$

where (x, y, z) is the 3-D Cartesian location of some point feature over time (perhaps the body in motion or a color patch on a moving body part). The above description dictates the shape or path of the motion over time, with no influence from dynamics. Obviously, other more complex and

Motion	Amplitude			Frequency			Phase			Translation			# Species
	A_x	A_y	A_z	f_x	f_y	f_z	ϕ_x	ϕ_y	ϕ_z	B_x	B_y	B_z	
Up-and-down	0	α_y	0	-	f	-	-	0	-	0	0	0	14/14
Side-to-side	α_x	0	0	f	-	-	0	-	-	0	0	0	12/14
Circle	α_x	0	α_z	f	-	f	0	-	$\pm\pi/2$	0	0	0	14/14
Spiral	α_x	0	α_z	f	-	f	0	-	$\pm\pi/2$	0	$(0\ldots]$	0	11/14
Undulate	0	α_y	0	-	f	-	-	0	-	$(0\ldots]$	0	0	5/14
Loop	α_x	α_y	0	f	f	-	0	$\pm\pi/2$	-	$[0-\alpha_x 2\pi f)$	0	0	5/14
Figure-8	α_x	α_y	0	$f/2$	f	-	0	$0, \pi$	-	0	0	0	6/14
U-shuttle	α_x	α_y	0	$f/2$	f	-	0	$-\pi/2$	-	0	0	0	2/14

Table 1. Oscillatory motions and sinusoidal model parameters. Model parameters are shown for the oscillations generated by sinusoidal functions $X(t) = A\sin(2\pi ft + \phi) + Bt$. Values α and f correspond to variable amplitude and frequency values, respectively. Slots with - are non-applicable parameters due to corresponding zero amplitudes. The last column lists the number of different bird species (total of 14 examined) that exhibit the motion patterns.

dynamic models could be explored, but they too must necessarily follow this fundamental oscillatory behavior. Table 1 shows the parameter settings for this sinusoidal model needed to characterize our set of oscillatory motions (in their purest, idealized form).

3.2. Parameter constraints

Looking closely at these parameter values in the table, we see that the only frequency ratios are $\{1:1, 1:2\}$ and the only relative phases (locking ϕ_x for reference) are $\{0, \pm\frac{\pi}{2}, \pi\}$. In particular, the phase relation for circular motions (circle, spiral, loop) must obey $\phi_x - \phi_y = \pm\frac{\pi}{2}$, for figure-8 ($\infty$) the relation must be $2\phi_x - \phi_y = \{0, \pi\}$, and for U-shuttle ($\cup$) the constraint is $2\phi_x - \phi_y = \frac{\pi}{2}$. For looping, the amount of translation B_x is constrained by the product of its corresponding amplitude and frequency ($\alpha_x 2\pi f$), otherwise swinging occurs. Although many other distinctive values *could* exist for the sinusoids (e.g. $\frac{\pi}{3}$, $4f$, etc.), they are not seen in these oscillatory motions. Such special values for this model suggest that structural or generative regularities [25, 14] underlie this class of movement.

3.3. Stylistic parameters

In addition to the structural parameter relations described above for the qualitative pattern, certain parameter values can be used for stylistic variation in the motions while not disrupting the overall form of the pattern. In these motions, special performance styles can be encoded into the frequency (fast, slow) and amplitude (shallow, high) parameters resulting in selective recognition (e.g. as used in species identity [16, 20]).

3.4. Motion ordering

We can group and order the motions in the category by relating the common structures and using the number of active parameters (default values are zero). Clearly, the up-and-down and side-to-side motions form the simplest group of “1-D” motion. Noticing a common translation component, we can group together spiral, undulate, and loop into “Translation” motion. With a special frequency ratio of 1:2, we can group together figure-8 and U-shuttle into “Frequency-doubling” motion. Lastly, this leaves the circle pattern to be “2-D” motion. We can then rank these groups in complexity using the number of parameters and constraints required for each group. The simple 1-D motions can be combined with an additional phase-difference constraint to generate 2-D motion, which can be further specialized into the complex groups of Translation (2-D + translation, except for undulate) and Frequency-doubling (2-D with frequency-doubling). These ranked *sub-categories* are marked in the horizontal divisions from top to bottom in Table 1.

The values listed in the last column of Table 1 are the number of different bird species (total of 14 examined, e.g. ducks, hawks, woodpeckers) which exhibit those motion patterns. The larger values reflect the commonality of those motions across the species. The less frequent the type of motion, the more likely that motion is seen within a species, but not across many species. Thus we see that the simpler sub-categories (1-D and 2-D) are more ubiquitous than the complex sub-categories (Translation, Frequency-doubling). This is not surprising if we believe that the patterns we deemed simple require less cognitive and motor control, and hence should be more universal. We later use these ordered sub-categories to specify the computations for recognition.

4. Data representation

Before constructing the motion recognition system, a representation for the data must be chosen. We require the representation to be time-based, rather than purely spatial, due to the temporal nature of the motion signal and the speed and frequency tunings involved in perception. However, it is not obvious whether the patterns should be characterized using position, velocity, acceleration, etc. Given the choice between representations, we select the one that best retains the sinusoidal shape over varying performance conditions. We examine specifically the trajectories of human arm motion (also used in later experiments).

Previous studies characterizing human arm movement have promoted using minimum-jerk profiles (e.g. [7]). In [1], unrestrained human arm trajectories between *point targets* (under varying conditions of speed and hand-held load) were investigated and shown to have an invariant velocity profile when normalized for both speed and distance. We examined *oscillatory* motions using the arm with alterations in user, speed, amplitude, and performing limb. The variations allowed us to measure and compare the stability and invariance for position and velocity.

To conduct the experiments, we developed a system employing infrared-light that extracts the 2-D position of a reflective ball manipulated by the user (See Fig. 2.a). A computer digitizes the video input, thresholds the images (See Fig. 2.b), and extracts the centroid of the 2-D ball region in each image. Example data trajectories generated by the system are shown in Fig. 2.c,d. The design and output of this system resemble the optical motion capture systems used for motion analysis (similar to [11]).

The first of four experiments was comprised of having three individuals each perform multiple repetitions of a sideways figure-8 pattern (∞), requiring sinusoidal movement in both x and y. Figure 3 shows the layered normalized plots (in both time and amplitude) of half-cycles for the mean-shifted position and velocity trajectories (first low-pass filtered). In this experiment, the velocity profiles show considerably more variance than the position trajectories, (especially for $x'(t)$). The remaining experiments varied the performance with speed (slow and fast), amplitude height (vertically enlarging the loops), and limb (arm, forearm, and torso of one person). Throughout the tests, the position and velocity data retained the basic characteristic shapes as shown in Fig. 3.

The variance increase in velocity can be clearly understood in the frequency domain as a non-linear multiplier of the magnitude (increasing to the Nyquist frequency) resulting from the derivative calculation:

$$x'(t) = \frac{x(t) - x(t - \Delta t)}{\Delta t} \rightarrow \sqrt{\frac{2(1 - \cos(2\pi f \Delta t))}{\Delta t^2}}|X(f)|$$

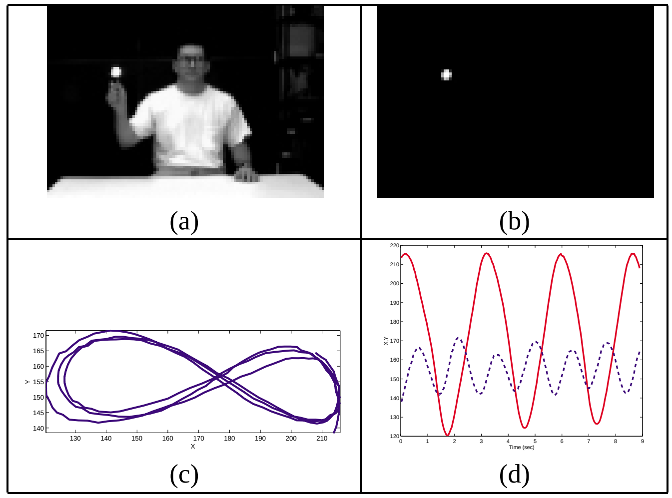

Figure 2. Infrared tracking system. (a) Person holding a wand attached to a reflective ball. (b) Ball region extracted by vision system. (c) Spatial pattern. (d) Temporal position trajectories.

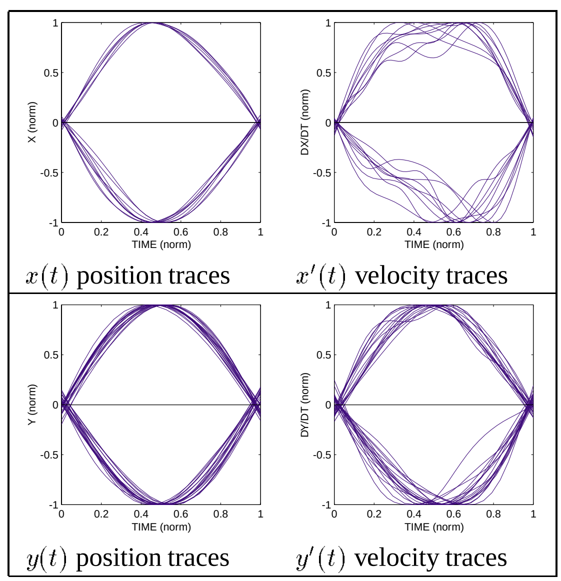

Figure 3. Composite half-cycles for three users performing a sideways figure-8 motion (∞). Top row shows normalized $x(t)$ (left) and $x'(t)$ (right). Bottom row shows normalized $y(t)$ (left) and $y'(t)$ (right).

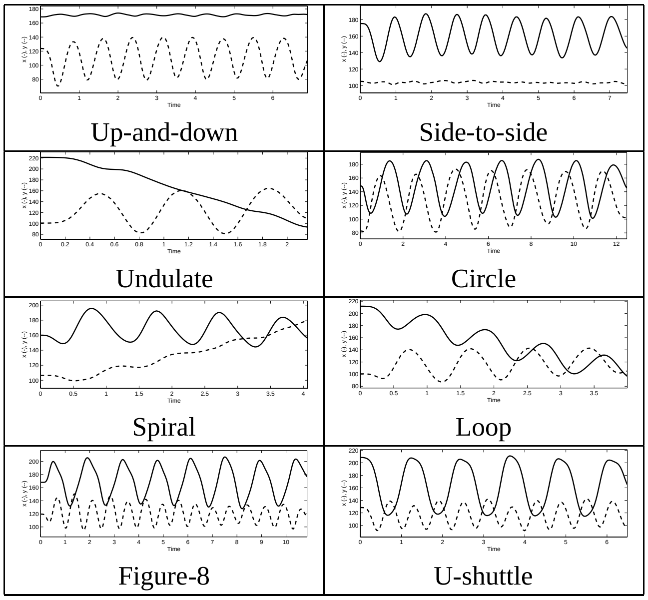

Figure 4. User position trajectories for the motions using the infrared tracking system. The $x(t)$ and $y(t)$ trajectories correspond to the solid and dashed lines, respectively.

Naturally generated sinusoidal motions typically have more variability with harmonic components, and thus the resultant derivative contains significantly amplified harmonic noise. The increase in horizontal versus vertical variance in Fig. 3 may be partially accounted for by the higher degrees of freedom in the arm for side-to-side motion (elbow and shoulder) than for up-and-down motion (hinged elbow).

The results of this experiment suggest that the position information, rather than velocity, is more representative of the sinusoidal motions. Therefore, we chose to use the position trajectories to represent the data. Figure 4 shows the position trajectories for the set of oscillatory motions as generated by a person using the infrared tracking system.

5. Recognition approach

Our categorical approach to the recognition of the oscillatory motions is to estimate the required oscillatory motion parameters from the input position trajectories, and compare them with parameterized category models. We interpret the constrained parameter relations (Table 1) and the ordered sub-categories (Sect. 3.4) of the motion patterns to reflect a series of localized computations assigned to the oscillatory specializations. The sub-category partitioning is thus used to specify and organize the types of computation required for recognition. A initial substrate is first developed to recognize the simplest 1-D motion patterns (up-and-down, side-to-side), and is then evolved to full categorical recognition by successively adding computation to accommodate the sub-categories of 2-D (circle), Translation (spiral, undulate, loop), and Frequency-doubling (figure-8, U-shuttle). This "layering" of computation is similar to the robotics Subsumption architecture championed by Brooks [2] for creating complex behaviors from successive levels of computation. The result is that simple motions require only basic computation, and more complex motions require additional or augmented processing beyond the shared simpler computations.

Each model from the category consists simply of specific values for particular sinusoidal model parameters. The job of recognition is therefore to extract and match only the necessary parameters from the input data to the model using only the required computation. The amount and type of processing depends on the sub-category (complexity) of the motion pattern to recognize. A structural test confirms the necessary parameter relationships as given in Table 1, and if required, a stylistic test matches the parameter values to specific quantities. For recognition we can perform either exhaustive recognition (seeking through all the models of interest) or context-specific recognition (examining only certain models of interest).

5.1. Implementation

The approach initially extracts a single pattern cycle from the $x(t), y(t)$ trajectories and estimates the required model parameters. All trajectories are initially lowpass filtered using a compact IIR design. Due to the non-linear phase response associated with IIR filters, we calculated a single phase difference threshold of 18 degrees for use with the entire set of motions (determined by the largest disparity from Frequency-doubling motions).

We designed the system to require only a single cycle of the pattern and to use the sign changes (i.e. zero-crossings) in velocity (See Fig. 5.a) to segment the pattern cycles. After the pattern is extracted, the motion is shifted by the pattern center to remove the viewing condition. Recognition therefore can be attempted at these zero-crossing locations during repetitions, rather than only at the end of each pattern.

To calculate the amplitude, frequency, and phase parameters (when applicable) for the pattern cycle, Fourier analysis (FFT) is applied individually to the $x(t)$, $y(t)$ trajectories. An amplitude threshold (3 pixels) is then applied to "flatten" trajectories not strongly sinusoidal, and a relative-amplitude scale threshold (10%) is applied to remove a small-amplitude trajectory overshadowed by a reference sinusoidal trajectory (specified by the model). The computations for matching are locally arranged to reflect the sub-categorical organization for the motions:

• *1-D motion: up-and-down, side-to-side*

To extract a single pattern cycle, the position data for both trajectories are extracted between the first and third velocity zero-crossing in the sinusoidal trajectory ($x'(t)$ for side-to-side, $y'(t)$ for up-and-down). The structural test for these motions then confirms that only the sinusoidal trajectory has a non-zero amplitude. The stylistic test matches the actual frequency and amplitude measurements to model specifics.

• *2-D motion: circle*

In this addition, the pattern is extracted between three zero-crossings, all in the $x'(t)$ or $y'(t)$ trajectory. The analysis follows the same pathway as with 1-D motion, but additionally verifies non-zero amplitudes ($\alpha_x > 0, \alpha_y > 0$), a frequency ratio of 1 ($f_x = f_y$), and a phase difference ($\phi_x - \phi_y$) of $\pm\frac{\pi}{2}$. The stylistic test additionally matches the amplitudes for *both* trajectories to the model specifics.

• *Translation motion: spiral, undulate, loop*

Movements that have a translation component are next incorporated. The velocity zero-crossings in the non-translating trajectory ($y'(t)$ for loop and undulate, $x'(t)$ for spiral) are used to extract the pattern cycle. To remove the translation, the linear component between the start and stop points of the translated trajectory is subtracted. The structural test includes an additional check for a translation above (5%) a relative minimum (relative to the translating trajectory amplitude for spiraling and looping, or relative to the non-translating trajectory amplitude for undulating). The frequency and phase checks are the same as in 2-D motion (suppressed for undulate). The stylistic test may additionally look for a specific translation quantity according to the model specifics (e.g. “tight” vs. “spread-out” loops).

• *Frequency-doubling motion: figure-8, U-shuttle*

For these motions, there is a change in the frequency ratio (having one trajectory moving twice as fast as the other) with no translation. Five velocity zero-crossings are found in the faster $y'(t)$ trajectory to extract a single pattern cycle (earlier experiments show this trajectory to be more stable). In the structural test, the frequency ratio is checked to be $1/2$ ($f_x = \frac{1}{2}f_y$), and the relative phase difference ($2\phi_x - \phi_y$) is verified to be 0 or π for figure-8 motion, or $\frac{\pi}{2}$ for U-shuttle motion. The stylistic tests are the same as for 2-D motion.

An associated periodicity value is also calculated for each data trajectory to indicate the quality of the sinusoidal fit to the data as expected by the model. The measure for a trajectory segment $x[t]$ is a product of a discrete frequency magnitude ratio and a time-based tail alignment:

$$P_x = \left(1 - \frac{\max(|X[f \neq f_0]|)}{|X[f_0]|}\right) \times \left(1 - \frac{|x[t_a] - x[t_b]|}{|X[f_0]|}\right)$$

where f_0 is the frequency containing the maximum energy, $|X[f]|$ is the magnitude at frequency f, and $(x[t_a], x[t_b])$ are the start and end locations of the cycle. The term

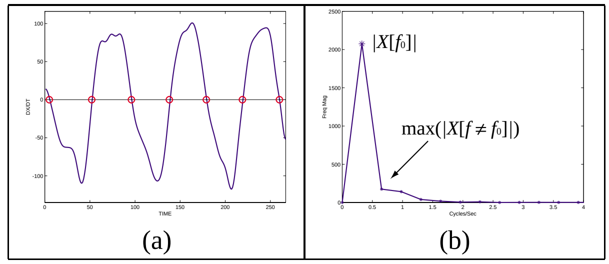

Figure 5. Implementation figures. (a) Velocity zero-crossings of mean-shifted $x(t)$ trajectory for figure-8 pattern. (b) Frequency magnitudes for a single cycle. Periodicity components are noted.

$|X[f \neq f_0]|$ is the maximum magnitude value in the discretely sampled spectrum other than at the maximum $|X[f_0]|$ (See Fig. 5.b). In the target case of sinusoidal motions, this residual maximum should be small (even as a harmonic peak). The first component in P_x looks for a single sinusoid, and the second component determines how well the pattern is “closed” (the translation, if any, is removed before this measure). If there is no motion (i.e. $|X[f_0]| = 0$), the resultant periodicity defaults to 1. The periodicity values for both trajectories (P_x, P_y) are multiplied together to yield a single periodicity measurement P_{xy} for the entire motion pattern. This value is returned as a confidence measure when a correct structural match is found.

6. Recognition analysis

To examine the recognition capability of this categorical implementation, we analyzed the model confusion using infrared tracking data to identify the nature of the misclassifications. Next, we tested the recognition system with trajectories extracted from real video sequences of simple human and animal oscillations.

6.1. Model confusion

In this framework, the computations allocated to recognize the simple motions are unaware of the additional components (or specifications) that exist in the more complex motions. Thus there may be unexpected results when models of one complexity (sub-category) are applied to motions of another complexity. Here we examine the confusion results of testing each model with an example of each motion using infrared tracking trajectories produced by two individuals (none of the authors). For this analysis, only the structural components in the motions were examined. The

	UD	SS	C	S	U	L	F8	US
UD	**.91**	-	-	-	-	-	-	-
SS	-	**.96**	-	-	-	-	-	-
C	-	-	**.83**	.02	-	.59	.16	.13
S	-	-	.20	(**.82**)	-	.39	-	.04
U	-	-	-	-	**.88**	-	-	-
L	-	-	.65	-	-	**.90**	-	.83
F8	-	-	-	-	-	-	**.74**	-
US	-	-	-	-	-	-	-	**.71**

Table 2. Confusion matrix. Periodicity values P_{xy} of structural matches for models (row) applied to data (col). Slots with - represent no structural match. Labels correspond to the ordering in Table 1.

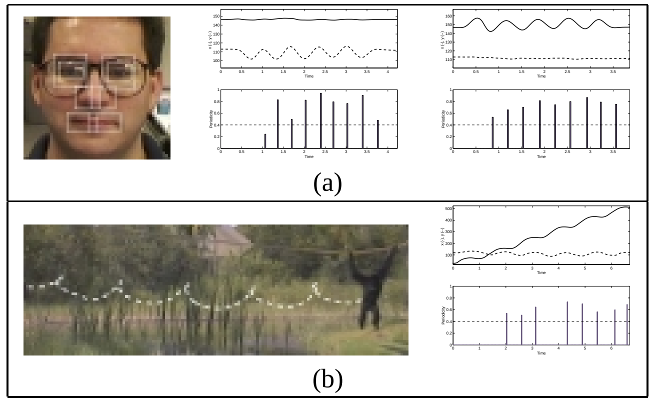

Figure 6. Recognition of (a) up-and-down 'Yes' and side-to-side 'No' head gesture, and (b) swinging pattern of a gibbon.

results are placed in the confusion matrix in Table 2. Values shown in the table are the maximum (between the two people) periodicity values P_{xy} returned when a structural match was found. The bold values along the diagonal represent the correct matches.

All the correct matches were found with high confidence for both people except for the spiral motion. This 3-D motion was difficult to perform and recognize in 2-D. Only if we open the phase-difference threshold from 18 to 46 degrees would the spiral model recognize both spiraling sequences. After opening the threshold for this model, only the U-shuttle additionally triggered a response (P_{xy}=0.07) by one subject. In the remaining false matches, errors occurred for circle, spiral, and loop models. For the circle model, the circular translation motions were similar and a loop of the figure-8 and a swing of the U-shuttle just barely registered a circular form. The results were similar for the spiral and looping models.

6.2. Real video examples

One of the simplest and most basic of human gestures is nodding or wagging the head to indicate a response (e.g. 'Yes' or 'No'). We applied an automatic face detector and real-time face tracker [10] to extract the motion of the head performing these simple gestures. The face model and results for up-and-down 'Yes' and side-to-side 'No' gestures of the head are shown in Fig. 6.a. The top plots in the row show the $x(t), y(t)$ position traces and the bottom plots show the recognition results. The spike magnitudes reflect the periodicity measures P_{xy} at the velocity zero-crossings used to segment the cycles. None of the other models matched these motion trajectories.

Another main advantage of using a categorical representation is that it is easy to create new models to extend the class of motions. With a single descriptive model (sinusoids), we need only change parameter values or constraints to create a new model. For example, increasing the translation for looping above a certain value ($\alpha_x 2\pi f$) transforms it into swinging. We therefore can easily create a swinging model to add to the class of motions by altering this parameter in looping. We applied this swinging model to a video sequence of a gibbon swinging (brachiating) across a rope (See Fig. 6.b). The gibbon trajectory was automatically extracted by thresholding the intensity images and collecting the centroid locations in the sequence of the dark region (the gibbon). Results in Fig. 6.b show recognition of all but one portion of a swing, which was due to a phase difference above tolerance. When all the other models were tested with this sequence, only the circle model had a response (P_{xy}=0.05). Note that partial occlusions have little impact on recognition for translation motions. Since the recognition approach requires only the data between three velocity zero-crossings (one cycle), a stationary occluder in the sequence will only disrupt recognition while the tracked motion is not visible. Once the moving object fully reappears (because of the translation), then only one cycle of the motion pattern is required for recognition to commence.

7. Summary

We presented a categorical method for the representation and recognition of a set of oscillatory biological motions. The approach was motivated by a collection of oscillatory motions used for animal communication that could be characterized by a simple sinusoidal model with specific and limited parameter values. By organizing these motions using the number of model parameters and constraints needed to describe the patterns, we formed sub-categorical specializations within the class that were used to specify the types and arrangement of computation needed to recognize the motions. The task of recognition was then to estimate and match the required model parameters and constraints to the corresponding categorical models. It was determined by experiment with human arm movements that the position trajectories were better suited than velocity to represent the sinusoidal data. Results using real oscillatory motions showed that the approach offers a simple and viable means for the recognition of categories of movements, without the need for statistical training on large sets of examples.

References

[1] C. Atkeson and J. Hollerbach. Kinematic features of unrestrained vertical arm movements. *J. Neurosci.*, 5(9):2318–2330, 1985.

[2] R. Brooks. A robust layered control system for a mobile robot. *IEEE J. Robotics and Automation*, RA-2(1):14–23, 1986.

[3] C. Cohen, L. Conway, and D. Koditschek. Dynamical system representation, generation, and recognition of basic oscillatory motion gestures. In *Proc. Int. Conf. Auto. Face and Gesture Recog.*, pages 60–65. IEEE, 1996.

[4] J. Collins and I. Stewart. Coupled nonlinear oscillators and the symmetries of animal gaits. *J. Nonlin. Sci.*, 3:349–392, 1993.

[5] R. Cutler and L. Davis. Real-time periodic motion detection, analysis, and applications. In *Proc. Comp. Vis. and Pattern Rec.*, pages 326–332. IEEE, 1999.

[6] J. Feldman. The structure of perceptual categories. *J. Math. Psych.*, 41:145–170, 1997.

[7] T. Flash and N. Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. *J. Neurosci.*, 5(7):1688–1703, 1985.

[8] J. J. Gibson. Ecological optics. *Vis. Res.*, 1:253–262, 1961.

[9] J. Hollerbach. An oscillation theory of handwriting. *Biological Cybernetics*, 39:139–156, 1981.

[10] T. Jebara and A. Pentland. Parameterized structure from motion for 3D adaptive feedback tracking of faces. In *Proc. Comp. Vis. and Pattern Rec.*, pages 144–150. IEEE, 1997.

[11] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2):201–211, 1973.

[12] J. Little and J. Boyd. Describing motion for recognition. In *Proc. Symp. Comp. Vis.*, pages 235–240. IEEE, 1995.

[13] F. Liu and R. Picard. Finding periodicity in space and time. In *Proc. Int. Conf. Comp. Vis.*, pages 376–383. IEEE, 1998.

[14] D. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Press, 1985.

[15] D. Marr. *Vision*. W.H. Freeman and Company, 1982.

[16] E. Martins. A comparative study of the evolution of Sceloporus push-up displays. *Am. Nat.*, 142(6):994–1018, 1993.

[17] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in XYT. In *Proc. Comp. Vis. and Pattern Rec.*, pages 469–474. IEEE, 1994.

[18] R. Polana and R. Nelson. Detection and recognition of periodic, nonrigid motion. *Int. J. Comp. Vis.*, 23(3):261–282, 1997.

[19] W. Richards and A. Bobick. *Computational Processes in Human Vision: An Interdisciplinary Perspective*, chapter Playing twenty questions with nature, pages 3–26. Ablex, 1988.

[20] M. Salmon. Waving display and sound production by Florida fiddler crabs (genus Uca). *Anim. Behav.*, 15:449–459, 1967.

[21] S. Seitz and C. Dyer. View-invariant analysis of cyclic motion. *Int. J. Comp. Vis.*, 25(3):1–23, 1997.

[22] D. Stokes. *A Guide to Bird Behavior*, volume 1. Little, Brown and Company, 1979.

[23] P.-S. Tsai, K. Keiter, T. Kasparis, and M. Shah. Cyclic motion detection. *Pat. Rec.*, 27(12):1591–1603, 1994.

[24] K. von Frisch. *The Dance Language and Orientation of Bees*. Harvard Univ. Press, 1967.

[25] A. Witkin and J. Tenenbaum. *Human and Machine Vision*, chapter On the role of structure in vision, pages 481–543. Academic Press, 1983.