



ANNUAL
REVIEWS **Further**

Click here to view this article's
online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Computational Analysis of Behavior

S.E. Roian Egnor and Kristin Branson

Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147;
email: egnorr@janelia.hhmi.org, bransonk@janelia.hhmi.org

Annu. Rev. Neurosci. 2016. 39:217–36

First published online as a Review in Advance on
April 18, 2016

The *Annual Review of Neuroscience* is online at
neuro.annualreviews.org

This article's doi:
10.1146/annurev-neuro-070815-013845

Copyright © 2016 by Annual Reviews.
All rights reserved

Keywords

automated behavioral analysis, tracking, animal behavior, computer vision,
machine learning

Abstract

In this review, we discuss the emerging field of computational behavioral analysis—the use of modern methods from computer science and engineering to quantitatively measure animal behavior. We discuss aspects of experiment design important to both obtaining biologically relevant behavioral data and enabling the use of machine vision and learning techniques for automation. These two goals are often in conflict. Restraining or restricting the environment of the animal can simplify automatic behavior quantification, but it can also degrade the quality or alter important aspects of behavior. To enable biologists to design experiments to obtain better behavioral measurements, and computer scientists to pinpoint fruitful directions for algorithm improvement, we review known effects of artificial manipulation of the animal on behavior. We also review machine vision and learning techniques for tracking, feature extraction, automated behavior classification, and automated behavior discovery, the assumptions they make, and the types of data they work best with.

Contents

INTRODUCTION	218
WHAT IS BEHAVIOR?	218
TRADITIONAL BEHAVIORAL ANALYSIS	220
RECORDING BEHAVIOR	221
TRACKING ANIMAL POSITION AND POSE	221
AUTOMATIC BEHAVIORAL ANALYSIS	223
Low-Level Representations of Behavior for Automated Analysis	223
Automatic Behavior Classification	226
Behavioral Building Blocks	227
Automatic Behavioral Analysis from Audio	228
BEHAVIOR IN THE WILD	228
Why Is Recording High-Quality Behavior Important?	228
Why Is Recording High-Quality Behavior Hard?	229
CONCLUSIONS	230

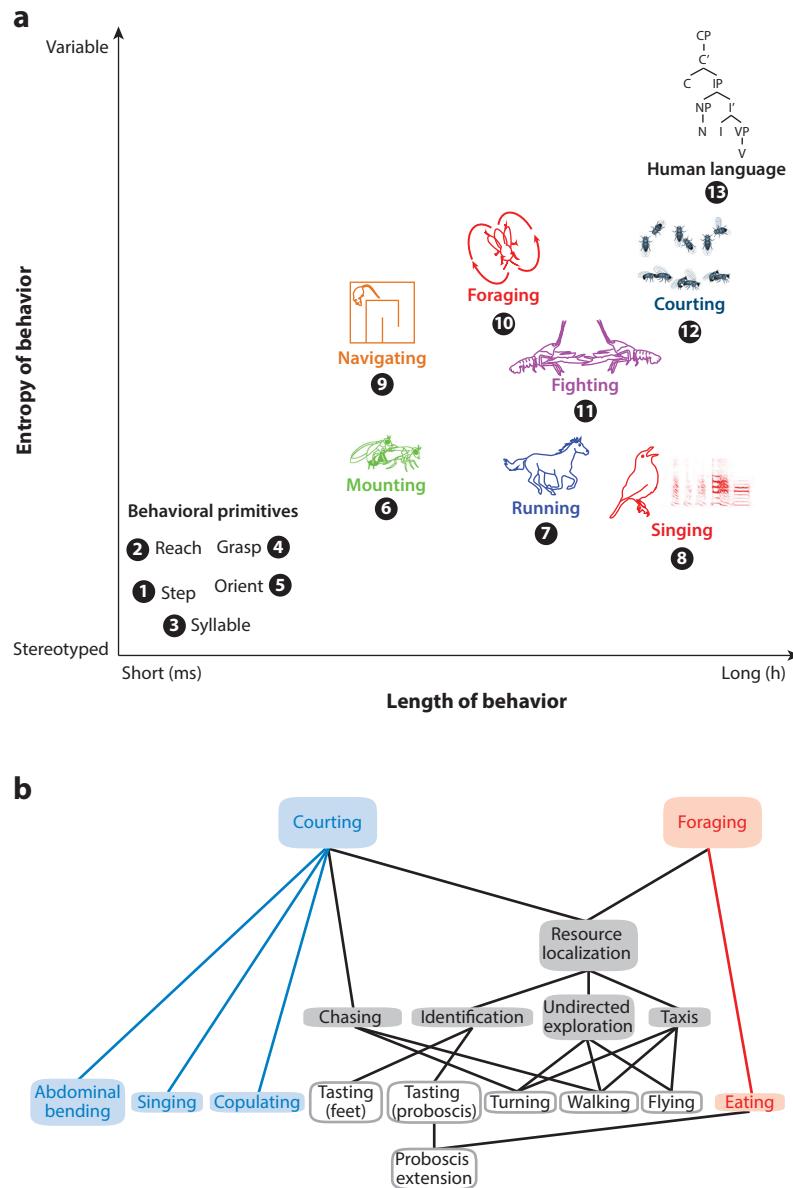
INTRODUCTION

Quantitative measurement of animal behavior is an important tool in a variety of fields within biology. Whether the object under study is the structure of the behavior itself or a means of revealing underlying neural or genetic mechanisms, researchers are interested in useful, quantifiable, and robust descriptions of behavior that allow us to describe how behavior differs in various conditions. For example, when we optogenetically activate a particular set of neurons, which behaviors does the animal perform more or less? Which behaviors does it perform differently, and how? In this review, we discuss automated approaches from machine vision and learning to behavioral analysis, which have the ability to analyze orders of magnitude larger data sets, discover features that humans cannot, and provide a vocabulary for discussing and describing behavior that is consistent across labs (and even across organisms).

We provide an overview of issues associated with the computational analysis of behavior, describing methods for eliciting naturalistic or controlled behavior as well as the latest methods for automatically recording, tracking, and classifying different behaviors. The intended audience includes two broad groups of scientists: (*a*) researchers in computer vision who are interested in what methods and features would be of interest to biologists and (*b*) biologists who are interested in using computational methods to improve their behavioral analysis or who want to expand their work into the analysis of behavior.

WHAT IS BEHAVIOR?

A behavior is a sequence of movements performed by an animal (or person). But which movements should we consider? And how should we group them together? A behavior can be defined in terms of a goal: navigating to a burrow (Wehner 2003), escaping a predator (Roeder 1962), courting a mate (Bradbury & Vehrencamp 2011), or defeating a rival (Kravitz 2000, Naguib & Kipper 2006). Behaviors can be stereotyped or variable, and can vary in duration by orders of magnitude (**Figure 1a**). A behavior can often be broken down into smaller behaviors (Fentress & Stilwell 1973, Seeds et al. 2014). Sequences of movements by a male fruit fly such as orienting, tapping, and singing to a female can be grouped together in time as a bout of courtship (Sokolowski

**Figure 1**

Behavioral structure across time and variability scales. (a) Behaviors vary in duration and variability, from short, stereotyped behavioral primitives to long and variable courtship displays. (1) Step (Mendes et al. 2013), (2) reach (Whishaw & Pellis 1990), (3) syllable (Clark et al. 1987, Holy & Guo 2005, Riede 2014, Tchernichovski et al. 2000), (4) grasp (Sacrey et al. 2009), (5) orient (Knudsen et al. 1979, Mays & Sparks 1980), (6) mounting (McGill 1962), (7) running (Bender et al. 2011), (8) singing (Marler 1970), (9) navigating (Frisch 1967, O'Keefe & Dostrovsky 1971, Wehner 2003), (10) foraging (Frisch 1967, Huang et al. 2008), (11) fighting (Chen et al. 2002, Dankert et al. 2009, Huber & Kravitz 1995), (12) courting (Spieth 1974; Sokolowski 2001; image adapted with permission from Sokolowski 2001), and (13) human language (Blevins 1995, Chomsky 1956). (b) Specific behaviors can be made up of subbehaviors, some of which may be shared between different high-level behaviors.

2001; **Figure 1a**, courting), and different sequences of these movements, all with the same goal of courting a mate, can be classed together as courtship behaviors. Each of these simpler actions (orienting, tapping, and singing) is itself also a behavior, which in turn can be broken down into even smaller behaviors strung together with rules. Subbehaviors can be specific to a particular behavior or shared between different behaviors (**Figure 1b**).

The precise definition of behavior depends on the goals of the study. A researcher interested in foraging will focus on the location of animals within the environment relative to sources of food, whereas a researcher interested in social behavior will focus on movements and changes in orientation relative to other animals, regardless of their position in the environment. For the purposes of this review, we define a behavior as any movement an animal makes.

TRADITIONAL BEHAVIORAL ANALYSIS

Historically, researchers have analyzed behavior using two distinct methods—either detailed manual descriptions of behavior by human observers or simple assays designed to capture some aspect of a complex behavior. An example of the former is focal sampling (Altmann 1974, Arrington 1943), in which all behaviors of a target animal are recorded. An example of the latter is the tube test in mice, in which dominance is assessed by priority of right-of-way in a narrow tube (Lindzey et al. 1961). Detailed manual behavior descriptions are time-consuming, laborious, and difficult to standardize across labs, but they provide a wealth of useful mechanistic detail and ensure the behavior of interest is captured. Conversely, behavioral assays are easy to implement, fast, and more easily standardized across labs but may not measure the behavior of interest and may obscure useful mechanistic details.

An example of the trade-off inherent in behavioral assays can be seen in the tube test. This simple-to-implement and widely used assay was developed initially by Lindzey and colleagues (1961) to test dominance behavior in mice. Five years later, the same researchers compared the result of the tube test with two other, more biologically direct measures of dominance: priority access to food and success in fights. These experiments showed that the tube test not only fails to predict dominance relationships between mice as measured by access to food and success in fights (both of which have direct fitness consequences) but is actually anticorrelated (Lindzey et al. 1966), suggesting that different aspects of mouse aggression are being measured by the different methods. Thus, although simpler to implement and standardize across labs, behavioral assays may miss aspects of behavior that are of interest to the researcher.

However, detailed manual annotation is not without its pitfalls, in addition to the time and labor required. Most notable is the problem of classification variability—either across observers or within a single observer over time (Arrington 1943, Paylor 2008)—making it difficult to compare results across laboratories (Anderson & Perona 2014). Because of this difficulty, researchers often compress human observations into a numeric index that attempts to capture some aspect of behavior (e.g., Ludvig et al. 2003, Michael & Saayman 1967). However, this compression may obscure mechanistic detail about the underlying structure of a behavior (Anderson & Perona 2014). For a discussion of the pitfalls of behavioral assays, see Fonio et al. (2012).

As we show in this review, with modern computational analysis methods, it is now often possible to observe and quantify the behavior of interest directly, rather than relying on a correlated behavioral test or laborious manual annotations. The subsequent sections describe how more complete, fine-grained, and nuanced descriptions of behavior can be acquired automatically, allowing the researcher to take advantage of the mechanistic structure offered by detailed descriptions without sacrificing the speed of traditional behavior assays, thus allowing high-quality, high-throughput behavioral analysis.

RECORDING BEHAVIOR

Behaviors range between short and simple—such as a saccade to an auditory target (Knudsen et al. 1979)—and long and complex—such as a courtship sequence (Spieth 1974) (**Figure 1a**). Once the behavior of interest has been identified, the next step is to figure out how to record it. This requires first deciding what features of the behavior are important—for example, is it enough to approximate the position of a single animal by a point [as in place preference assays (Dolen et al. 2013)], or is it necessary to measure the position of multiple body parts of multiple animals [as in studies of courtship or aggression (Chen et al. 2002, Dankert et al. 2009)]? Properties of the behavior to be measured determine which recording method is best. Biologists have deployed a wide array of sensors to capture animal behavior, including infrared sensors (Friedman et al. 1997), microphone arrays (Neunuebel et al. 2015), laser Doppler vibrometry (Elias et al. 2003), accelerometers (Spence et al. 2010), touch screens (Bermejo et al. 1994), search coils for monitoring eye or head position (Knudsen et al. 1979), and frustrated Total Internal Reflection (Mendes et al. 2013). Because it is relatively noninvasive, inexpensive, and general-purpose, it is common to record behavior on video and then extract features relevant to the behavior using computer-vision techniques. This review focuses only on methods associated with extraction of features from video. However, the issues that we raise—how to define the behavior, record it, extract descriptive features, and combine features to identify behaviors—hold true conceptually, regardless of the sensor used to capture behavior.

TRACKING ANIMAL POSITION AND POSE

Many behavioral studies are based on analyses of the position and pose of the animals over time. Animal tracking from video was initially performed manually [e.g., analysis of *Drosophila melanogaster* flight (Buelthoff et al. 1980), wolf social behavior (Moran et al. 1981), and human social behavior (Kendon 1975)]. However, this is extremely time-consuming, and in recent years researchers have developed many automatic tracking programs (see Dell et al. 2014 for an overview of available systems and their strengths and limitations, and the Related Resources for a list of freely available tracking software packages).

One way to think about tracking is as a data-association problem (Forsyth et al. 2005, Poppe 2007). We would like to associate each pixel in each frame of a video (**Figure 2a**) with either an animal or the background (**Figure 2b**). When tracking multiple animals, we also need to further label each pixel within an animal with the animal identity (**Figure 2c**), and if tracking body parts, we need to further label each pixel with the body part identity. Given this segmentation of pixels in each frame into background and individual animals and their body parts, one can then fit a parametric model of the pose of the animal (**Figure 2d**).

Finding the optimal assignment for all pixels in the entire video is an extremely hard problem, and all tracking algorithms decompose it into smaller subproblems. For example, many animal tracking algorithms (e.g., Branson et al. 2009, Noldus et al. 2002, Ohayon et al. 2013, Swierczek et al. 2011) begin by solving the animal-background segmentation problem in each frame independently. They then assume that this segmentation is correct and further cluster the pixels labeled as animal in a single frame into groups corresponding to different animals. Finally, they assign individual animal identities to these clusters of pixels across frames.

Within a controlled laboratory setting, we can design our experiments to make these data-association problems easier. We can make the background a distinct color from the animals to simplify animal-background segmentation (e.g., Branson et al. 2009, Ohayon et al. 2013, Swierczek et al. 2011). We can design our recording apparatus so that the background is static (i.e., neither

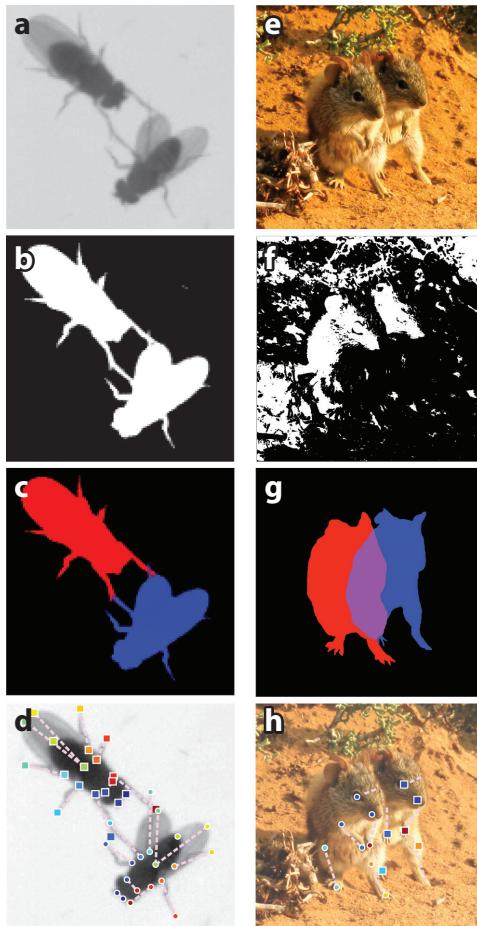


Figure 2

Identifying animals in video images. (a) Raw video image of two unmarked flies. (b) Background-animal segmentation of image in panel a. (c) Assigning individual identity based on position and movement. (d) Estimating the pose of the flies, represented as the locations of their parts. (e) Two adult striped mice (*Rhabdomys pumilio*) basking outside of their nest (photo credit: Susanne Guldener). (f) A simple threshold does not produce useable animal-background segmentation because color and texture are similar between the mice and the environment. (g) Even if perfect foreground-background segmentation could be achieved, the fact that the mice overlap (*purple region*) makes identifying the body parts (b) of individual mice difficult.

the background nor the camera move), allowing us to model the background appearance at every pixel location, then use background subtraction to segment each frame (Piccardi 2004). Biologists have long used bleach or dye to mark animals for identification in the field (Stonehouse 1978), and such markings can allow automatic tracking systems to more easily associate pixels with the identities of individuals over long timescales (Ohayon et al. 2013, Shemesh et al. 2013). In carefully controlled settings, even unmarked animals that appear identical to the human observer may be distinguishable using sensitive computer-vision techniques (Pérez-Escudero et al. 2014).

Tracking algorithms that do not use appearance to distinguish individuals rely on the motion of the animals being small and smooth to match identities between frames. A common approach is to assign identities in each frame in order. For example, Branson et al. (2009) arbitrarily assign

identities in the first frame. Then, for each consecutive frame t , they assume that the pose estimates in the previous frames are correct, and they assign identities in frame t so that the implied motion between frames $t-1$ and t matches an assumed motion model well. A disadvantage of such an approach is that if a mistake is made in one frame, it is propagated forward to all later frames. Methods such as particle filtering (Murphy 2012) instead estimate a distribution of possible pose fits at each frame, allowing the algorithm to recover from some mistakes. Recent methods estimate the poses of all targets in all frames simultaneously, and current research is focused on improving the efficiency and applicability of these methods (Berclaz et al. 2011, Pirsavash et al. 2011, Schiegg et al. 2013). Unfortunately, all tracking algorithms that use motion alone to track identities will fail in certain cases. In particular, these algorithms might fail if one or more animals are occluded (hidden behind each other or another object relative to the camera; **Figure 2e**) or if the animals move quickly and unexpectedly (relative to the video frame rate).

Up to this point, we have discussed how we might find the optimal assignment of labels to pixels, assuming we have some mathematical criterion that allows us to evaluate how well a labeling fits the data. But how do we come up with this criterion? We can manually design it—for example, by assuming that animals move with a constant velocity and have certain shapes, sizes, and colors (e.g., Branson et al. 2009, Ohayon et al. 2013, Swierczek et al. 2011). Such approaches usually have many parameters that must be set carefully. Alternatively, we can use machine learning approaches to learn the parameters of this criterion automatically from small sets of manually labeled data (Li et al. 2013, Lou & Hamprecht 2012, Yang et al. 2011). We note that, for both such approaches, the resulting criterion will be finely tuned to the type of video used when setting the parameters to the criterion. Thus, if one later changes some properties of the video (for example, the lighting in the scene), the parameters of the criterion may then need to be reset. For this reason, for large-scale experiments, it is important to keep the video collection rig as constant as possible.

Existing animal tracking software was developed for experiments in the lab and relies on assumptions that can be made only in artificial environments (e.g., that the animal-background segmentation problem is easy). They are thus not applicable to tracking multiple, unmarked animals in more natural scenes (**Figure 2e–h**) with occlusion, variation in background, and variation in lighting. Tracking animals in the wild is still a difficult problem under active investigation (see the section titled Behavior in the Wild).

AUTOMATIC BEHAVIORAL ANALYSIS

Trajectories (the positions over time) of the animals and/or their parts (such as limbs or facial features), in their raw form, are rarely a useful representation of behavior. A common goal in automatic behavioral analysis is to categorize behavior at each time point into distinct classes, such as walking, grooming, or chasing. By segmenting these time series into distinct behaviors, we can then measure changes in the frequency and manner with which animals perform given behaviors.

Low-Level Representations of Behavior for Automated Analysis

An important component to all these methods is how we represent the video sequence input to the classification algorithm. In this section, we describe common low-level representations of behavior that can be extracted from video. First, we discuss trajectory-based features (**Figure 3a,c**)—features extracted from the tracked positions of the animals and/or their parts over time (Bulling et al. 2014, Burgos-Artizzu et al. 2012, Eyjolfsdottir et al. 2014, Jhuang et al. 2010, Kabra et al. 2013, Mirat et al. 2013, Ramanan & Forsyth 2003, Rao et al. 2002, Rohrbach

et al. 2012). Beyond being useful as inputs to behavior classification, trajectory-based features are also a useful, interpretable representation of more continuous properties of behavior.

Often, the video frame rates and pixel counts necessary to capture a behavior via part tracking are enormous, and tracking these parts may be more difficult than the end goal of behavior classification. Thus, in addition to, or in lieu of, tracking, many computer-vision algorithms extract low-level representations of behavior directly from the raw pixel values of the video (Aggarwal & Ryoo 2011; Dollár et al. 2005; Efros et al. 2003; Jhuang et al. 2010; Kläser et al. 2008; Laptev & Lindeberg 2003; Laptev et al. 2008; Ryoo & Aggarwal 2009; Wang et al. 2009, 2013; Willems et al. 2008), which we term pixel-based features (**Figure 3b,d**).

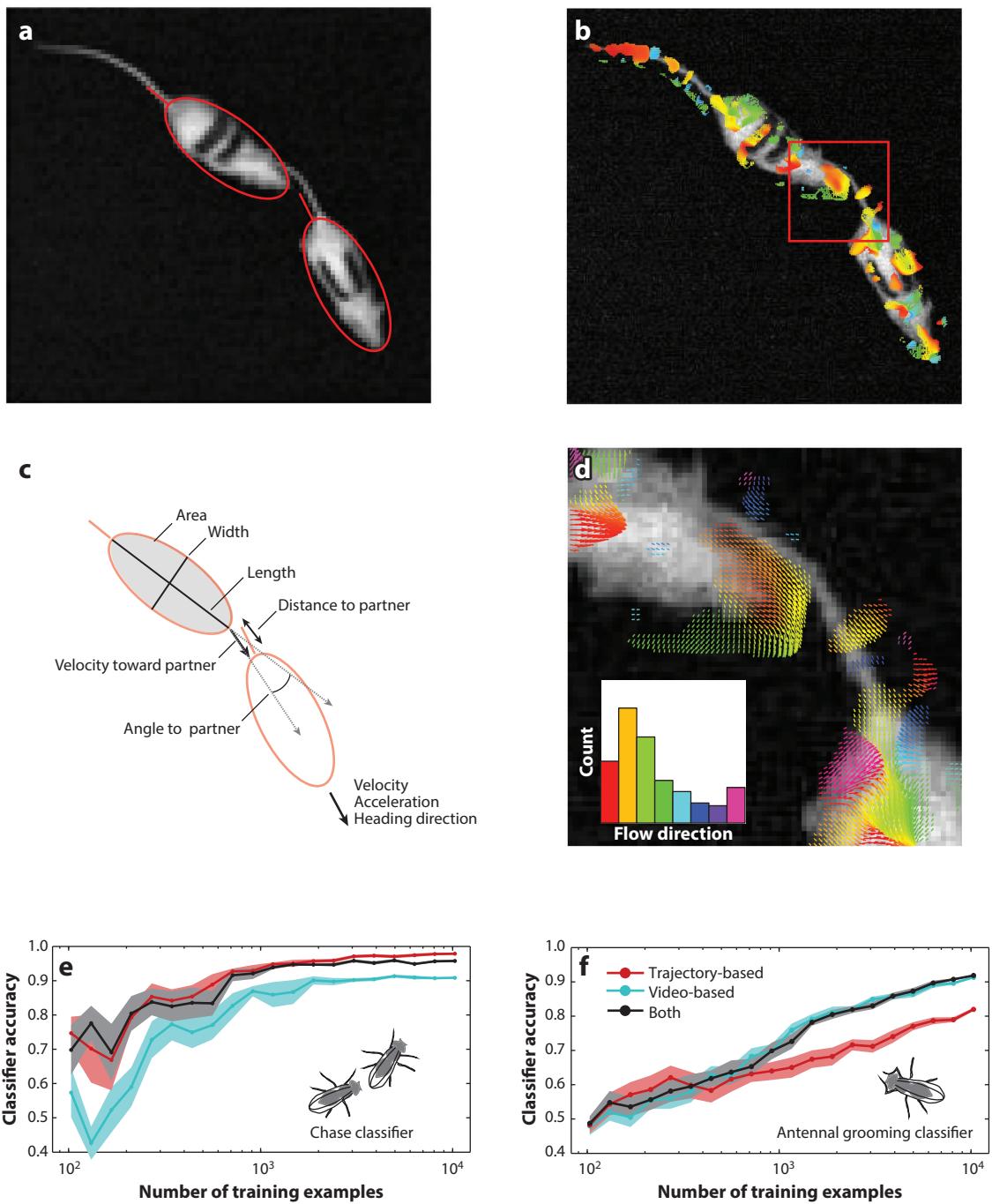
Trajectory-based features. The change in the position of an animal as a function of time can carry an enormous amount of information about behavioral state (see **Figure 3c** for examples of trajectory-based features). This can be seen in the ability of humans to identify actions (walking, running, dancing) from displays consisting of only dots on body joints (Johansson 1973, Song et al. 2003). For single animals, simple measurements derived from the trajectory, such as the first time derivative (speed) can identify bouts of rest, walking, and running (Martin 2004, Valente et al. 2007). Features derived from the relative positions of multiple body parts (de Chaumont et al. 2012) or multiple animals (Branson et al. 2009, Burgos-Artizzu et al. 2012) can identify more complex behaviors such as chasing. Even in the absence of explicit part-based tracking, stereotypic motions of the body, captured by ellipse-fit tracking, can be used to train body part–specific classifiers. For example, a *Drosophila* wing-grooming classifier can be trained from an ellipse-fit trajectory that does not track the wings explicitly (Kabra et al. 2013). However, the quality of trajectory-based features depends on the accuracy of the trajectories from which they are computed (Aggarwal & Ryoo 2011).

Pixel-based features. Pixel-based features are extracted from local patches of video (x - y - t subcuboids; see **Figure 3b, d**). There is a long history of designing feature descriptors to represent the useful information in a static x - y image patch (a subrectangle of a single frame), such as the histogram of oriented gradients (HOG) (Dalal & Triggs 2005). These features have been designed to retain information relevant for object detection (e.g., edge orientations) and discard potentially distracting information (e.g., lighting-dependent information). These feature descriptors have been extended to describe motion in video patches—for example, histogram of optical flow (HOF) (Laptev et al. 2008) (**Figure 3d**). To represent a video patch, researchers commonly combine both static-image and motion feature descriptors.

Often, to represent a video sequence, pixel-based features are computed at many patches within the sequence, chosen either using interest-point detectors (Dollár et al. 2005, Laptev & Lindeberg

Figure 3

Low-dimensional features to describe behavior—trajectory-based and pixel-based features. (a) Ellipse fits to measure body positions of two mice involved in a chase. (b) Local optical flow vectors for the same two mice. (c) Examples of features that can be calculated from trajectories of body position over time. (d) Example of histogram of optical flow (HOF) features calculated from local optical flow vectors (arrows indicate the magnitude and direction). Accuracy of a fly chase classifier (e) and an antennal grooming classifier (f) trained with trajectory-based features (red), pixel-based features (cyan), or both (black) as a function of the number of training examples. Note that for the chase classifier, trajectory-based features produce more accurate classification, with fewer training examples, than pixel-based features, whereas pixel-based features perform better for antennal grooming. The trajectory-based features are based on an ellipse fit to the animal's body and thus do not contain explicit information about the location and motion of the legs or antennae; pixel-based features, by contrast, do contain information about the position and motion of the antennae and limbs, leading to more accurate classification using these features. The definition of a chase is well described by comovement of the centroids of the two animals, leading to accurate classification using trajectory-based features.



2003) or exhaustively (Wang et al. 2009). These patches are quantized into spatiotemporal prototypes, and the count of occurrences of each prototype in a dictionary is used to represent the video sequence. For a discussion and comparison of pixel-based features, see Wang et al. (2009).

Researchers designed both the trajectory- and pixel-based features described above to encode relevant information in a format amenable to behavior classification. Comparisons of trajectory- and video-based features have had mixed results (Burgos-Artizzu et al. 2012), and which type of feature performs best will depend on the behavior to be detected and the tracking quality (**Figure 3e**). Better classifiers can be trained using trajectory-based features if the relevant body parts are tracked explicitly and accurately. However, if tracking parts explicitly is not possible, pixel-based features can be more effective. The current trend in computer vision is to learn generally useful feature descriptors directly from extremely large image (Krizhevsky et al. 2012) and video (Karpathy et al. 2014) data sets.

Automatic Behavior Classification

In this section, we discuss three approaches to automatic behavior classification. Rule-based classifiers (Dankert et al. 2009, de Chaumont et al. 2012) are sets of manually defined rules that describe whether the animal is or is not performing a given behavior. Supervised machine-learning methods train classifiers automatically from small sets of manually labeled video sequences (Burgos-Artizzu et al. 2012, Kabra et al. 2013). And finally, unsupervised machine-learning methods can automatically cluster groups of video sequences into classes (Berman et al. 2014, Vogelstein et al. 2014).

Rule-based classifiers. The simplest way to create a behavior classifier from trajectory-based features is to choose features to threshold manually (de Chaumont et al. 2012, Gomez-Marin et al. 2012, Martin 2004, Robie et al. 2010, Swierczek et al. 2011). For example, we can threshold speed to define bouts of activity and inactivity (Martin 2004, Robie et al. 2010) or threshold several measures of the relative positions and movements of the animals to define chasing or following (Dankert et al. 2009, de Chaumont et al. 2012). However, this type of behavior detector is difficult to tune; may depend only on a handful of features; cannot be used for more complicated behaviors; and rarely generalizes well across large, behaviorally diverse populations of animals.

Training classifiers with supervised machine learning. More complex classifiers can be trained automatically using supervised machine learning techniques (Aggarwal & Ryoo 2011, Branson et al. 2009, Burgos-Artizzu et al. 2012, Dankert et al. 2009, Dollár et al. 2005, Jhuang et al. 2010, Kabra et al. 2013, Kain et al. 2012, Mirat et al. 2013, Wang et al. 2009). Instead of setting the parameters of the classifier manually, we label the behavior (the desired output of the behavior classifier) manually for a set of frames. We then use machine learning to find the parameters of the classifier function that can replicate these annotations using pixel- or trajectory-based features. Many machine learning algorithms can be used to train behavior classifiers. Searching a more complex family of classifiers allows one to learn a more complex behavior classifier, possibly from rawer forms of input, but requires more training data and a longer training time.

Machine learning algorithms (both supervised and unsupervised) are based on several assumptions that might not hold for animal behavioral analysis. It is important to know these assumptions to understand how to use them properly.

Most importantly, (standard) machine learning algorithms assume that each example, both in the training and future test data, is an independent draw from the exact same distribution. However, in many behavior experiments, the training data set has systematic differences from the future test data. For example, this is the case if we train our classifier on data from wild-type

animals and then use it on a mutant population. It is therefore best to include labeled examples from both populations in the training set. Alternatively, transfer learning algorithms address this exact problem and try to adapt classifiers to new types of test data (Raina et al. 2007).

Another potential hazard in training behavioral classifiers is noise in the input labels. Although some machine learning algorithms are robust to some kinds of label noise (Frénay & Verleysen 2014), it can have calamitous effects on many standard algorithms. Kabra et al. (2013) urge the user to label only frames for which they are confident of the behavior label.

Finally, in behavior classification, adjacent frames are highly related to each other, a relationship often modeled using hidden Markov models (HMMs) (Aggarwal & Ryoo 2011, Yamato et al. 1992), grammars (Aggarwal & Ryoo 2011, Ivanov & Bobick 2000, Pirsavash & Ramanan 2014), or other sparse models of conditional dependence. Such behavior classifiers can be trained using structured machine learning algorithms (Eyjolfsdottir et al. 2014, Jhuang et al. 2010, Tsochantaridis et al. 2005). These approaches are more demanding in terms of amount and type of training data and time to train.

Discovering behavior structure with unsupervised machine learning. As yet, we have assumed that we know the behaviors of interest and need only to automate their detection. How do we know that our definitions of behavior correspond to behavioral distinctions that are meaningful to the animal or relevant for neurobiological control? We can use unsupervised machine learning techniques to try to mine the inherent structure in behavior from large video or trajectory data sets.

We can cluster video sequences into behavior categories such that members of different clusters are more different than members of the same clusters (Berman et al. 2014, Brown et al. 2013, Niebles et al. 2008, Schwarz et al. 2015, Vogelstein et al. 2014, Zhong et al. 2004). To do this, we must define mathematically how different two video segments are, for example, by measuring the Euclidean distance between the video segments in one of the low-dimensional behavior representations described previously. Alternatively, we can search for a low-dimensional representation of a video segment such that we can reconstruct the original video segment well (Berman et al. 2014, Stephens et al. 2008, Zelnik-Manor & Irani 2001). Here, we again rely on a mathematical specification of behavior difference to measure the error between the reconstructed and original video segments. Thus, in unsupervised learning, we trade off dependence on human-produced behavior labels for a stronger dependence on the choice of feature representation of the video segment and the modeling assumptions of our algorithm.

Between supervised and unsupervised learning live two other classes of algorithms: semi- and weakly supervised learning. In semi-supervised learning, we try to combine a smaller set of labeled examples with the structure implicit in a large set of unlabeled examples (Chapelle et al. 2006; Stikic et al. 2008, 2011). In weakly supervised learning, we replace labels of the exact desired classifier output with easier-to-obtain or more trustworthy signals (Ikizler-Cinbis & Sclaroff 2010, Laptev et al. 2008, Stikic et al. 2011). Such approaches have the potential to capitalize on the strengths of both unsupervised and supervised learning.

Behavioral Building Blocks

Researchers in both computer science and biology are interested in finding which features extracted from video are most efficient for training behavioral classifiers or for the identification of intrinsic behavioral structure. Even before it was possible to record behaviors at high resolution, biologists had asked whether complex behaviors are composed of fundamental units [motor primitives (Overduin et al. 2012), modules (Riede 2014), or components (Fentress & Stilwell 1973)]

defined by constraints imposed by biophysics, neural control, or both (**Figure 1a,b**). The discovery of such fundamental units would provide a compact and efficient description of behavior. Although there is consensus that complex behaviors can often be composed of sequences of simpler behaviors, the identity of these behavioral primitives and whether behaviorally defined subunits correspond to building blocks on the neural level are areas of active investigation.

An elegant example of the utility of identifying a basis set can be seen in the description of the behavior and underlying genetics of *Caenorhabditis elegans*. Stephens and colleagues (2008) showed that the locomotory behavior of worms was well described by four basic shapes, or eigenworms. Using this representation, Brown and colleagues (2013) were also able to describe behavior in mutant worms, discover clusters of related genes by clustering mutants based on shape, and reveal previously undetected phenotypes. A related set of descriptive primitives were developed by Berman and colleagues (2014), which they called fly postural modes, to describe behavior in several fly species.

Computer-vision researchers have also discussed the notion of behavioral building blocks [movemes (Goncalves et al. 2004), gestures (Aggarwal & Ryoo 2011), and dynamic instants (Packer et al. 2012)]. However, the majority of related research in computer vision has focused instead on designing (Wang et al. 2009) or learning (Karpathy et al. 2014) video-based representations that are useful for training behavior classifiers. Because supervised machine learning algorithms are designed to be robust to redundant and irrelevant features, a useful representation for behavior classification does not necessarily correspond to true building blocks of behavior. Nonetheless, the practical search for a useful set of features (or a useful algorithm for finding them) may also provide insight into more theoretical questions about the underlying structure of behavior itself.

Automatic Behavioral Analysis from Audio

We focus on the automatic extraction of movement-based behavior from video in this review. However, vocal behavior, a significant component of social and foraging behavior across animal taxa, can also be analyzed using automated methods.

Birdsong is a particularly attractive behavior to investigate because of its stereotypy and biological relevance: Male songbirds produce song to attract mates and defend territory (Konishi 1989). Initial descriptions of songbird song were performed manually (Thorpe 1954). However, more automated methods have been developed that have accelerated work in the field. For example, Tchernichovski and colleagues (2000) published a method for describing song using only four extracted features: Wiener entropy, spectral continuity, pitch, and frequency modulation. They used this system to make fundamental advances in our understanding of birdsong, for example, that the structure of birdsong can be developed in the absence of a tutor (Feher et al. 2009) and that both song and language learning proceed in discrete steps (Lipkind et al. 2013). The authors made their software easy to use and freely available (<http://soundanalysispro.com>). Since Song Analysis Pro was published, researchers have used it in a wide variety of studies, from the effect of basal ganglia input on vocal plasticity in zebra finches (Kojima et al. 2013) to a description of the vocal development of marmosets (Takahashi et al. 2015).

BEHAVIOR IN THE WILD

Why Is Recording High-Quality Behavior Important?

Recording behavior involves a balance between the demands of the behavior and the demands of the recording. Behaviors are often modified for practical purposes, based on recording constraints.

Arenas are made small (to keep the animals within recording range) or short (to prevent occlusions). Animals are restrained (head fixation), hormonally or surgically manipulated (estrus induction in mice, wing clipping in flies), or, perhaps most dramatically of all, inbred for hundreds of generations. All these manipulations have been shown to modify the expression of behavior (Agrawal et al. 2014, Benjamini et al. 2010, Chalfin et al. 2014). For example, Kimchi, Dulac, and colleagues show that many critical social behaviors, such as those associated with female mate choice, are lost in inbred mice (Chalfin et al. 2014). In many cases, such modifications are necessary to observe the behavior at all. However, it is important to recognize that modifications to improve recording quality may degrade the quality of the behavior under study.

Some of these trade-offs can be mitigated by providing appropriate feedback. For example, songbirds (Thorpe 1961) and mice (Sewell 1967) produce vocalizations during social interactions but become silent when they are head-fixed, restricting the methods that can be used to study the neural basis of this behavior. However, providing a treadmill for the head-fixed animal helps to restore normal vocal behavior in both [mice: R. Egnor & M. London, personal communication; songbirds: Picardo et al. 2016]. The use of virtual reality (Dombeck & Reiser 2012), in combination with spherical treadmills, in rodents (Harvey et al. 2009, Holscher et al. 2005, Schnee 2008) and insects (Buchner 1976, Kerfoot 1968, Seelig et al. 2010) can allow relatively natural behavior, tight stimulus control, and stable neural recording.

At short timescales, studies have shown that the common practice of transferring an animal from its home environment to a testing apparatus can produce dramatic changes in behavior (Coe et al. 1978, Ferland & Schrader 2011, Hennessy & Moorman 1989). In response to this realization, many investigators are designing experiments to be run in the home cage (Goulding et al. 2008, Jhuang et al. 2010). The developmental experience and long-term housing conditions of behavior subjects can also have dramatic effects on behavior. At the extreme, deprivation of particular sensory inputs during development can prevent even basic behaviors from being expressed—for example, visual deprivation during development causes blindness in cats (Wiesel & Hubel 1963). Even seemingly innocuous decisions, such as filtering air, can produce mice with marked sensory deficits in adulthood (Oliva et al. 2010). Researchers have also documented effects of early experience on the behavior of shorter-lived animals, such as *D. melanogaster* and *C. elegans* (Burns et al. 2012, Ebrahimi & Rankin 2007, Hirsch & Tompkins 1994, Hirsch et al. 1995, Kim & Ehrman 1999). For a good review of these issues and some suggestions for mitigating them, see Lahvis (2016).

Why Is Recording High-Quality Behavior Hard?

In natural video images (**Figure 2e**), the background and scene lighting are neither uniform nor static. Because the subjects to be tracked can change their distance and orientation to the camera, they can change size and shape. Subjects can occlude parts of themselves and each other and can be occluded by objects in the environment. Subjects can be anywhere in the image and can appear and disappear entirely from the recording frame. Tracking people or animals in natural scenes is therefore difficult because few assumptions can be made about the position, shape, size, or color of subjects, and there are also other objects in the scene and variation in lighting. Modern approaches to tracking targets in more natural scenes must decompose the tracking problem into different, larger subproblems and exploit advanced machine learning techniques (e.g., Tompson et al. 2015).

In addition to using better algorithms, ambiguities in tracked video can be resolved by using additional information. Although this is useful in a controlled laboratory setting, it becomes critical with natural video. This information can come from the use of multiple cameras to capture a scene from multiple angles (Mischiati et al. 2015, Straw et al. 2011). Or this can come from

independent channels of information about the position of objects—for example, the Kinect, which provides depth information associated with video (Packer et al. 2012)—or from radiofrequency identification (RFID) tags (Weissbrod et al. 2013).

CONCLUSIONS

Neuroscience has seen an explosion of powerful new tools for recording and manipulating neural activity. In parallel, researchers have developed more accurate computer-vision methods for identifying people automatically in images taken in more realistic contexts. Although the promise of combining these advances is clear, these methods are not yet plug-and-play, and a deep understanding of the strengths and weaknesses of both the quality of the data to be analyzed and the methods of analysis is needed. Therefore, there is a critical need for tight and ongoing collaborations between biologists and computer scientists. Such collaborations, which allow the development of algorithms to take advantage of new data set sizes and new sources of information about animal behavior, have the potential to reveal the underlying neurobiological structure of behavior in extraordinary detail.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Julie Simpson, Galit Shohat-Ophir, and Ben Cocanougher for comments on the manuscript and Susanne Guldener and the Succulent Karoo Research Station for **Figure 2e**. We also thank Stefanie Hampel for the data used to generate **Figure 3e** and Alice Robie and Marta Rivera-Alba for data analysis and discussions.

LITERATURE CITED

- Aggarwal JK, Ryoo MS. 2011. Human activity analysis: a review. *ACM Comput. Surv.* 43:16
- Agrawal S, Safarik S, Dickinson M. 2014. The relative roles of vision and chemosensation in mate recognition of *Drosophila melanogaster*. *J. Exp. Biol.* 217:2796–805
- Altmann J. 1974. Observational study of behavior: sampling methods. *Behaviour* 49:227–67
- Anderson DJ, Perona P. 2014. Toward a science of computational ethology. *Neuron* 84:18–31
- Arrington RE. 1943. Time sampling in studies of social behavior: a critical review of techniques and results with research suggestions. *Psychol. Bull.* 40:81–124
- Bender JA, Simpson EM, Tietz BR, Daltorio KA, Quinn RD, Ritzmann RE. 2011. Kinematic and behavioral evidence for a distinction between trotting and ambling gaits in the cockroach *Blaberus discoidalis*. *J. Exp. Biol.* 214:2057–64
- Benjamini Y, Lipkind D, Horev G, Fonio E, Kafkafi N, Golani I. 2010. Ten ways to improve the quality of descriptions of whole-animal movement. *Neurosci. Biobehav. Rev.* 34:1351–65
- Berclaz J, Fleuret F, Turetken E, Fua P. 2011. Multiple object tracking using k-shortest paths optimization. *IEEE Trans. Pattern Anal. Mach. Intell.* 33:1806–19
- Berman GJ, Choi DM, Bialek W, Shaevitz JW. 2014. Mapping the stereotyped behaviour of freely moving fruit flies. *J. R. Soc. Interface* 11:20140672
- Bermejo R, Houben D, Zeigler HP. 1994. Dissecting the conditioned pecking response: an integrated system for the analysis of pecking response parameters. *J. Exp. Anal. Behav.* 61:517–27

- Blevins J. 1995. The syllable in phonological theory. In *The Handbook of Phonological Theory*, ed. JA Goldsmith, pp. 206–44. London: Blackwell
- Bradbury JW, Vehren camp SL. 2011. *Principles of Animal Communication*. Sunderland, MA: Sinauer
- Branson K, Robie AA, Bender J, Perona P, Dickinson MH. 2009. High-throughput ethomics in large groups of *Drosophila*. *Nat. Methods* 6:451–57
- Brown AEX, Yemini EI, Grundy LJ, Jucikas T, Schafer WR. 2013. A dictionary of behavioral motifs reveals clusters of genes affecting *Caenorhabditis elegans* locomotion. *PNAS* 110:791–96
- Buchner E. 1976. Elementary movement detectors in an insect visual system. *Biol. Cybern.* 24:85–101
- Buelthoff H, Poggio T, Wehrhahn C. 1980. 3-D analysis of the flight trajectories of flies (*Drosophila melanogaster*). *Z. Naturforschung C* 35:811–15
- Bulling A, Blanke U, Schiele B. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.* 46:33
- Burgos-Artizzu XP, Dollár P, Lin D, Anderson DJ, Perona P. 2012. *Social behavior recognition in continuous video*. Presented at 2012 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Providence, RI
- Burns JG, Svetec N, Rowe L, Mery F, Dolan MJ, et al. 2012. Gene-environment interplay in *Drosophila melanogaster*: Chronic food deprivation in early life affects adult exploratory and fitness traits. *PNAS* 109(Suppl. 2):17239–44
- Chalfin L, Dayan M, Levy DR, Austad SN, Miller RA, et al. 2014. Mapping ecologically relevant social behaviours by gene knockout in wild mice. *Nat. Commun.* 5:4569
- Chapelle O, Schölkopf B, Zien A, eds. 2006. *Semi-Supervised Learning*. Cambridge, MA: MIT Press
- Chen S, Lee AY, Bowens NM, Huber R, Kravitz EA. 2002. Fighting fruit flies: a model system for the study of aggression. *PNAS* 99:5664–68
- Chomsky N. 1956. Three models for the description of language. *IRE Trans. Inf. Theory* 2:113–34
- Clark CW, Marler P, Beeman K. 1987. Quantitative-analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology* 76:101–15
- Coe CL, Mendoza SP, Smotherman WP, Levine S. 1978. Mother-infant attachment in the squirrel monkey: adrenal response to separation. *Behav. Biol.* 22:256–63
- Dalal N, Triggs B. 2005. *Histograms of oriented gradients for human detection*. Presented at 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., San Diego
- Dankert H, Wang L, Hooper ED, Anderson DJ, Perona P. 2009. Automated monitoring and analysis of social behavior in *Drosophila*. *Nat. Methods* 6:297–303
- de Chaumont F, Coura RD, Serreau P, Cressant A, Chabout J, et al. 2012. Computerized video analysis of social interactions in mice. *Nat. Methods* 9:410–17
- Dell AI, Bender JA, Branson K, Couzin ID, de Polavieja GG, et al. 2014. Automated image-based tracking and its application in ecology. *Trends Ecol. Evol.* 29:417–28
- Dolen G, Darvishzadeh A, Huang KW, Malenka RC. 2013. Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature* 501:179–84
- Dollár P, Rabaud V, Cottrell G, Belongie S. 2005. *Behavior recognition via sparse spatio-temporal features*. Presented at 2nd Joint IEEE Int. Workshop Vis. Surveill. Perform. Eval. Track. Surveill., Beijing
- Dombeck DA, Reiser MB. 2012. Real neuroscience in virtual worlds. *Curr. Opin. Neurobiol.* 22:3–10
- Ebrahimi CM, Rankin CH. 2007. Early patterned stimulation leads to changes in adult behavior and gene expression in *C. elegans*. *Genes Brain Behav.* 6:517–28
- Efros AA, Berg AC, Mori G, Malik J. 2003. *Recognizing action at a distance*. Presented at 9th IEEE Int. Conf. Comput. Vis., Nice, France
- Elias DO, Mason AC, Maddison WP, Hoy RR. 2003. Seismic signals in a courting male jumping spider (Araneae: Salticidae). *J. Exp. Biol.* 206:4029–39
- Eyjolfsson E, Branson S, Burgos-Artizzu XP, Hooper ED, Schor J, et al. 2014. *Detecting social actions of fruit flies*. Presented at 13th Eur. Conf. Comput. Vis. (ECCV), Zurich
- Feher O, Wang H, Saar S, Mitra PP, Tchernichovski O. 2009. De novo establishment of wild-type song culture in the zebra finch. *Nature* 459:564–68
- Fentress JC, Stilwell FP. 1973. Letter: grammar of a movement sequence in inbred mice. *Nature* 244:52–53
- Ferland CL, Schrader LA. 2011. Cage mate separation in pair-housed male rats evokes an acute stress corticosterone response. *Neurosci. Lett.* 489:154–58

- Fonio E, Golani I, Benjamin Y. 2012. Measuring behavior of animal models: faults and remedies. *Nat. Methods* 9:1167–70
- Forsyth DA, Arikan O, Ikemoto L, O'Brien J, Ramanan D. 2005. Computational studies of human motion: part 1, tracking and motion synthesis. *Found. Trends Comput. Graph. Vis.* 1:77–254
- Frénay B, Verleysen M. 2014. Classification in the presence of label noise: a survey. *IEEE Trans. Neural Netw. Learn. Syst.* 25:845–69
- Friedman D, Haim A, Zisapel N. 1997. Temporal segregation in coexisting spiny mice (genus *Acomys*): role of photoperiod and heterospecific odor. *Physiol. Behav.* 62:407–11
- Frisch KV. 1967. *The Dance Language and Orientation of Bees*. Cambridge, MA: Belknap
- Gomez-Marin A, Partoune N, Stephens GJ, Louis M, Brembs B. 2012. Automated tracking of animal posture and movement during exploration and sensory orientation behaviors. *PLOS ONE* 7:e41642
- Goncalves L, di Bernardo E, Perona P. 2004. Movemes for modeling biological motion perception. In *Seeing, Thinking and Knowing: Meaning and Self-Organisation in Visual Cognition and Thought*, ed. A Carsetti, pp. 143–70. Ser. A: Philos. Methodol. Soc. Sci. New York: Kluwer Acad.
- Goulding EH, Schenk AK, Juneja P, MacKay AW, Wade JM, Tecott LH. 2008. A robust automated system elucidates mouse home cage behavioral structure. *PNAS* 105:20575–82
- Harvey CD, Collman F, Dombeck DA, Tank DW. 2009. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* 461:941–46
- Hennessy MB, Moorman L. 1989. Factors influencing cortisol and behavioral responses to maternal separation in guinea pigs. *Behav. Neurosci.* 103:378–85
- Hirsch HVB, Barth M, Luo S, Sambaziotis H, Huber M, et al. 1995. Early visual experience affects mate choice of *Drosophila melanogaster*. *Anim. Behav.* 50:1211–17
- Hirsch HVB, Tompkins L. 1994. The flexible fly: experience-dependent development of complex behaviors in *Drosophila melanogaster*. *J. Exp. Biol.* 195:1–18
- Holscher C, Schnee A, Dahmen H, Setia L, Mallot HA. 2005. Rats are able to navigate in virtual environments. *J. Exp. Biol.* 208:561–69
- Holy TE, Guo Z. 2005. Ultrasonic songs of male mice. *PLOS Biol.* 3:e386
- Huang KM, Cosman P, Schafer WR. 2008. Automated detection and analysis of foraging behavior in *Caenorhabditis elegans*. *J. Neurosci. Methods* 171:153–64
- Huber R, Kravitz EA. 1995. A quantitative-analysis of agonistic behavior in juvenile American lobsters (*Homarus americanus* L.). *Brain Behav. Evol.* 46:72–83
- Ikizler-Cinbis N, Sclaroff S. 2010. *Object, scene and actions: combining multiple features for human action recognition*. Presented at 11th Eur. Conf. Comput. Vis. (ECCV), Crete, Greece
- Ivanov YA, Bobick AF. 2000. Recognition of visual activities and interactions by stochastic parsing. *IEEE Trans. Pattern Anal. Mach. Intell.* 22:852–72
- Jhuang H, Garrote E, Yu XL, Khilnani V, Poggio T, et al. 2010. Automated home-cage behavioural phenotyping of mice. *Nat. Commun.* 1:68
- Johansson G. 1973. Visual-perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14:201–11
- Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson K. 2013. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat. Methods* 10:64–67
- Kain JS, Stokes C, de Bivort BL. 2012. Phototactic personality in fruit flies and its suppression by serotonin and white. *PNAS* 109:19834–39
- Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. 2014. *Large-scale video classification with convolution neural networks*. Presented at 2014 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Columbus, OH
- Kendon A. 1975. Some functions of the face in a kissing round. *Semiotica* 15:99–334
- Kerfoot W. 1968. Orientometer for study of insect behavior. *Science* 162:477
- Kim YK, Ehrman L. 1999. Developmental isolation and subsequent adult behavior of *Drosophila paulistorum*. V. Survey of six sibling species. *Behav. Genet.* 29:65–73
- Kläser A, Marszałek M, Schmid C. 2008. *A spatio-temporal descriptor based on 3D-gradients*. Presented at Br. Mach. Vis. Conf., Leeds, UK

- Knudsen EI, Blasdel GG, Konishi M. 1979. Sound localization by the barn owl (*Tyto alba*) measured with the search coil technique. *J. Comp. Physiol. A* 133:1–11
- Kojima S, Kao MH, Doupe AJ. 2013. Task-related “cortical” bursting depends critically on basal ganglia input and is linked to vocal plasticity. *PNAS* 110:4756–61
- Konishi M. 1989. Birdsong for neurobiologists. *Neuron* 3:541–49
- Kravitz EA. 2000. Serotonin and aggression: insights gained from a lobster model system and speculations on the role of amine neurons in a complex behavior. *J. Comp. Physiol. A* 186:221–38
- Krizhevsky A, Sutskever I, Hinton GE. 2012. *ImageNet classification with deep convolutional neural networks*. Presented at NIPS 2012 Neural Inf. Process. Syst., Lake Tahoe, NV
- Lahvis G. 2016. Animal models of autism, epigenetics, and the inescapable problem of animal constraint. In *Animal Models of Behavior Genetics Research*, ed. JC Gewirtz, Y-K Kim. New York: Springer. In press
- Laptev I, Lindeberg T. 2003. *Space-time interest points*. Presented at 9th IEEE Int. Conf. Comput. Vis., Nice, France
- Laptev I, Marszalek M, Schmid C, Rozenfeld B. 2008. *Learning realistic human actions from movies*. Presented at IEEE Conf. Comput. Vis. Pattern Recognit., Anchorage, AK
- Li X, Hu W, Shen C, Zhang Z, Dick A, Van Den Hengel A. 2013. A survey of appearance models in visual object tracking. *ACM Trans. Intell. Syst. Technol.* 4:58
- Lindzey G, Manosevitz M, Winston H. 1966. Social dominance in the mouse. *Psychon. Sci.* 5:451–52
- Lindzey G, Winston H, Manosevitz M. 1961. Social dominance in inbred mouse strains. *Nature* 191:474–76
- Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, et al. 2013. Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* 498:104–8
- Lou X, Hamprecht FA. 2012. *Structured learning from partial annotations*. Presented at Int. Conf. Mach. Learn., Edinburgh, Scotl.
- Ludwig N, Tang HM, Eichenbaum H, Gohil BC. 2003. Spatial memory performance of freely-moving squirrel monkeys. *Behav. Brain Res.* 140:175–83
- Marler P. 1970. A comparative approach to vocal learning—song development in white-crowned sparrows. *J. Comp. Physiol. Psych.* 71:1–25
- Martin J-R. 2004. A portrait of locomotor behaviour in *Drosophila* determined by a video-tracking paradigm. *Behav. Processes* 67:207–19
- Mays LE, Sparks DL. 1980. Saccades are spatially, not retinocentrically, coded. *Science* 208:1163–65
- McGill TE. 1962. Sexual behavior in three inbred strains of mice. *Behaviour* 19:341–50
- Mendes CS, Bartos I, Akay T, Marka S, Mann RS. 2013. Quantification of gait parameters in freely walking wild type and sensory deprived *Drosophila melanogaster*. *eLife* 2:e00231
- Michael RP, Saayman G. 1967. Sexual performance index of male rhesus monkeys. *Nature* 214:425
- Mirat O, Sternberg JR, Severi KE, Wyart C. 2013. ZebraZoom: an automated program for high-throughput behavioral analysis and categorization. *Front. Neural Circuits* 7:107
- Mischiati M, Lin HT, Herold P, Immler E, Olberg R, Leonardo A. 2015. Internal models direct dragonfly interception steering. *Nature* 517:333–38
- Moran G, Fentress JC, Golani I. 1981. A description of relational patterns of movement during ritualized fighting in wolves. *Anim. Behav.* 29:1146–65
- Murphy KP. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press
- Naguib M, Kipper S. 2006. Effects of different levels of song overlapping on singing behaviour in male territorial nightingales (*Luscinia megarhynchos*). *Behav. Ecol. Sociobiol.* 59:419–26
- Neunuebel JP, Taylor AL, Arthur BJ, Egnor SER. 2015. Female mice ultrasonically interact with males during courtship displays. *eLife* 4:e06203
- Niebles JC, Wang H, Fei-Fei L. 2008. Unsupervised learning of human action categories using spatial-temporal words. *Int. J. Comput. Vis.* 79:299–318
- Noldus LPJJ, Spink AJ, Tegelenbosch RAJ. 2002. Computerised video tracking, movement analysis and behaviour recognition in insects. *Comput. Electron. Agric.* 35:201–27
- Ohayon S, Avni O, Taylor AL, Perona P, Egnor SER. 2013. Automated multi-day tracking of marked mice for the analysis of social behaviour. *J. Neurosci. Methods* 219:10–19
- O’Keefe J, Dostrovsky J. 1971. Hippocampus as a spatial map. Preliminary evidence from unit activity in freely-moving rat. *Brain Res.* 34:171–75

- Oliva AM, Salcedo E, Hellier JL, Ly X, Koka K, et al. 2010. Toward a mouse neuroethology in the laboratory environment. *PLOS ONE* 5:e11359
- Overduin SA, d'Avella A, Carmena JM, Bizzi E. 2012. Microstimulation activates a handful of muscle synergies. *Neuron* 76:1071–77
- Packer B, Saenko K, Koller D. 2012. *A combined pose, object, and feature model for action understanding*. Presented at 2012 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Providence, RI
- Paylor R. 2008. Simultaneous behavioral characterizations: embracing complexity. *PNAS* 105:20563–64
- Pérez-Escudero A, Vicente-Page J, Hinz RC, Arganda S, de Polavieja GG. 2014. idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nat. Methods* 11:743–48
- Picardo M, Merel J, Katlowitz K, Vallentin D, Okobi D, et al. 2016. Population-level representation of a temporal sequence underlying skilled behavior. *Neuron*. In press
- Piccardi M. 2004. *Background subtraction techniques: a review*. Presented at IEEE Int. Conf. Syst. Man Cybern., The Hague, Neth.
- Pirsavash H, Ramanan D. 2014. *Parsing videos of actions with segmental grammars*. Presented at IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Columbus, OH
- Pirsavash H, Ramanan D, Fowlkes CC. 2011. *Globally-optimal greedy algorithms for tracking a variable number of objects*. Presented at IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Colorado Springs, CO
- Poppe R. 2007. Vision-based human motion analysis: an overview. *Comput. Vis. Image Underst.* 108:4–18
- Raina R, Battle A, Lee H, Packer B, Ng AY. 2007. *Self-taught learning: transfer learning from unlabeled data*. Presented at 24th Annu. Int. Conf. Mach. Learn., Corvallis, OR
- Ramanan D, Forsyth DA. 2003. *Automatic annotation of everyday movements*. Presented at 24th Annu. Conf. Neural Inf. Process. Syst. (NIPS), Vancouver, Canada
- Rao C, Yilmaz A, Shah M. 2002. View-invariant representation and recognition of actions. *Int. J. Comput. Vis.* 50:203–26
- Riede T. 2014. Rat ultrasonic vocalization shows features of a modular behavior. *J. Neurosci.* 34:6874–78
- Robie AA, Straw AD, Dickinson MH. 2010. Object preference by walking fruit flies, *Drosophila melanogaster*, is mediated by vision and graviperception. *J. Exp. Biol.* 213:2494–506
- Roeder KD. 1962. The behaviour of free flying moths in the presence of artificial ultrasonic pulses. *Anim. Behav.* 10:300–4
- Rohrbach M, Amin S, Andriluka M, Schiele B. 2012. *A database for fine grained activity detection of cooking activities*. Presented at: 2012 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Providence, RI
- Ryoo MS, Aggarwal JK. 2009. *Spatio-temporal relationship match: video structure comparison for recognition of complex human activities*. Presented at Int. Conf. Comput. Vis. (ICCV), Kyoto, Japan
- Sacrey LA, Alaverdashvili M, Whishaw IQ. 2009. Similar hand shaping in reaching-for-food (skilled reaching) in rats and humans provides evidence of homology in release, collection, and manipulation movements. *Behav. Brain Res.* 204:153–61
- Schlieg M, Hanslovsky P, Kausler BX, Hufnagel L, Hamprecht F. 2013. *Conservation tracking*. Presented at IEEE Int. Conf. Comput. Vis. (ICCV 2013), Sydney
- Schnee A. 2008. *Rats in virtual reality: the development of an advanced method to study animal behaviour*. PhD Thesis, Eberhard Karls Univ., Tübingen, Ger.
- Schwarz RF, Branicky R, Grundy LJ, Schafer WR, Brown AE. 2015. Changes in postural syntax characterize sensory modulation and natural variation of *C. elegans* locomotion. *PLOS Comput. Biol.* 11:e1004322
- Seeds AM, Ravbar P, Chung P, Hampel S, Midgley FM Jr., et al. 2014. A suppression hierarchy among competing motor programs drives sequential grooming in *Drosophila*. *eLife* 3:e02951
- Seelig JD, Chiappe ME, Lott GK, Dutta A, Osborne JE, et al. 2010. Two-photon calcium imaging from head-fixed *Drosophila* during optomotor walking behavior. *Nat. Methods* 7:535–40
- Sewell GD. 1967. Ultrasound in adult rodents. *Nature* 215:512
- Shemesh Y, Sztainberg Y, Forkosh O, Shlapobersky T, Chen A, Schneidman E. 2013. High-order social interactions in groups of mice. *eLife* 2:e00759
- Sokolowski MB. 2001. *Drosophila: genetics meets behaviour*. *Nat. Genet.* 2:879–90
- Song Y, Goncalves L, Perona P. 2003. Unsupervised learning of human motion. *IEEE Trans. Pattern Anal. Mach. Intell.* 25:814–27

- Spence AJ, Revzen S, Seipel J, Mullens C, Full RJ. 2010. Insects running on elastic surfaces. *J. Exp. Biol.* 213:1907–20
- Spieth HT. 1974. Courtship behavior in *Drosophila*. *Annu. Rev. Entomol.* 19:385–405
- Stephens GJ, Johnson-Kerner B, Bialek W, Ryu WS. 2008. Dimensionality and dynamics in the behavior of *C. elegans*. *PLOS Comput. Biol.* 4:e1000028
- Stikic M, Laerhoven KV, Schiele B. 2008. *Exploring semi-supervised and active learning for activity recognition*. Presented at 12th IEEE Int. Symp. Wearable Comput., Newcastle, UK
- Stikic M, Larlus D, Ebert S. 2011. Weakly supervised recognition of daily life activities with wearable sensors. *IEEE Trans. Pattern Anal. Mach. Intell.* 3:2521–37
- Stonehouse B, ed. 1978. *Animal Marking: Recognition Marking of Animals in Research*. Baltimore, MD: Univ. Park
- Straw AD, Branson K, Neumann TR, Dickinson MH. 2011. Multi-camera real-time three-dimensional tracking of multiple flying animals. *J. R. Soc. Interface* 8(56):395–409
- Swierczek NA, Giles AC, Rankin CH, Kerr RA. 2011. High-throughput behavioral analysis in *C. elegans*. *Nat. Methods* 8:592–98
- Takahashi DY, Fenley AR, Teramoto Y, Narayanan DZ, Borjon JI, et al. 2015. The developmental dynamics of marmoset monkey vocal production. *Science* 349:734–38
- Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. 2000. A procedure for an automated measurement of song similarity. *Anim. Behav.* 59:1167–76
- Thorpe WH. 1954. The process of song-learning in the chaffinch as studied by means of the sound spectrograph. *Nature* 173:465–69
- Thorpe WH. 1961. *Bird Song*. Cambridge, UK: Cambridge Univ. Press
- Tompson J, Goroshin R, Jain A, LeCun Y, Bregler C. 2015. Efficient object localization using convolutional networks. arXiv:1411.4280 [cs.CV]
- Tsochantaridis I, Joachims T, Hofmann T, Altun Y. 2005. Large margin methods for structured and interdependent output variables. *J. Mach. Learn. Res.* 6:1453–84
- Valente D, Golani I, Mitra PP. 2007. Analysis of the trajectory of *Drosophila melanogaster* in a circular open field arena. *PLOS ONE* 2:e1083
- Vogelstein JT, Park Y, Ohyama T, Kerr RA, Truman JW, et al. 2014. Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science* 344:386–92
- Wang H, Kläser A, Schmid C, Liu C-L. 2013. Dense trajectories and motion boundary descriptors for action recognition. *Int. J. Comput. Vis.* 103:60–79
- Wang H, Ullah MM, Kläser A, Laptev I, Schmid C. 2009. *Evaluation of local spatio-temporal features for action recognition*. Presented at Br. Mach. Vis. Conf., London
- Wehner R. 2003. Desert ant navigation: how miniature brains solve complex tasks. *J. Comp. Physiol. A* 189:579–88
- Weissbrod A, Shapiro A, Vasserman G, Edry L, Dayan M, et al. 2013. Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nat. Commun.* 4:2018
- Whishaw IQ, Pellis SM. 1990. The structure of skilled forelimb reaching in the rat: a proximally driven movement with a single distal rotatory component. *Behav. Brain Res.* 41:49–59
- Wiesel TN, Hubel DH. 1963. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol.* 26:1003–17
- Willem G, Tuytelaars T, Van Gool L. 2008. *An efficient dense and scale-invariant spatio-temporal interest point detector*. Presented at Eur. Conf. Comput. Vis., Marseille, France
- Yamato J, Ohya J, Ishii K. 1992. *Recognizing human action in time-sequential images using hidden Markov model*. Presented at IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Champaign, IL
- Yang B, Huang C, Nevatia R. 2011. *Learning affinities and dependencies for multi-target tracking using a CRF model*. Presented at IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Colorado Springs, CO
- Zelnik-Manor L, Irani M. 2001. *Event-based analysis of video*. Presented at IEEE Comput. Vis. Pattern Recognit. (CVPR 2001), Kauai, HI
- Zhong H, Shi J, Visontai M. 2004. *Detecting unusual activity in video*. Presented at IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR 2004), Washington, DC

RELATED RESOURCES

Below are links to freely available software for implementing the methods discussed in this review. These websites also include information about hardware selection and recording rig design.

Tracking Software

Ctrax: The Caltech Multiple Walking Fly Tracker: <http://ctrax.sourceforge.net/>

Motr: MOuse TRacker: <http://motr.janelia.org/>

idTracker: <http://www.idtracker.es/>

MiceProfiler: http://icy.bioimageanalysis.org/plugin/Mice_Profiler_Tracker

Multi-Worm Tracker: <http://sourceforge.net/projects/mwt/>

SOS-track: Sensory Orientation Software to track animal behavior: <http://sourceforge.net/projects/sos-track/>

Behavioral Analysis Software

JAABA: Janelia Automated Animal Behavior Annotator: <http://jaaba.sourceforge.net/>

Acoustic SegmenteR (Ax): <https://github.com/JaneliaSciComp/Ax>

Sound Analysis Pro: <http://soundanalysispro.com/>

MotionMapper: <https://github.com/gordonberman/MotionMapper>