BUILDING A BASIS FOR BUDGIE BEHAVIOR

A PREPRINT

Ariel K. Feldman*
Departments of Computer Science and Cognitive Sciences
Rice University
Houston, TX 77005
Ariel.K.Feldman@rice.edu

Eugene Kim
Department of Neurobiology and Behavior
Cornell University
Ithaca, New York 14853
ek569@cornell.edu

Mert R. Sabuncu
Departments of Electrical and Biomedical Engineering
Cornell University
Ithaca, New York 14853
msabuncu@cornell.edu

Jesse H. Goldberg
Department of Neurobiology and Behavior
Cornell University
Ithaca, New York 14853
jessehgoldberg@gmail.com

August 8, 2019

ABSTRACT

In this work, we discuss methodologies for extracting, identifying and characterizing budgie behavior from video data, while minimizing user interference by removing the need to label training data. Further, we discuss methodologies for creating a three dimensional rendering of budgies from two cameras, which record in only two dimensions each, and stitching them together for three dimensional replication and analysis. Such functionality would facilitate interaction with budgies in novel ways, allowing us to control one aspect of behavioral paradigms and influence social learning.

***K*eywords** Pose Identification · Unsupervised Machine Learning · Budgie

1 Introduction

Language is one of the most critical inventions of mankind, the variety of grammatical structures and vocabulary within a single language facilitating the expression of intangible concepts between otherwise separated minds. Imagine, however, humans could only say one sentence. Just one — the intonation remains the same, the meaning doesn't change across time as one says the sentence. This hypothetical is inherently a learning paradigm, considering that an inability to learn new words and phrases may be considered the only roadblock to developing new sentence patterns. Once children learn this one sentence, that's it — we have a solid milestone for when learning is completed.

Perhaps this phenomenon is not observed in humans, but other species do exhibit such simplistic, structured learning during an easily identifiable interval. Zebra finches, a type of songbird, are one such creature. As they only learn one song throughout their lifespan [4], it is trivial to characterize their learning — they have either learned the song, or they have not. This has made zebra finches an incredibly attractive specimen in the field of vocal motor learning, as it is well documented what they should be learning, and, more importantly, the general time frame over which learning takes place. This enables investigation into whether different neural manipulations (in the form of gene alterations, lesions, electrophysiological stimulation, and more) impair or inhibit finch learning [3]. However, the goal of neural investigations in animal models is to extend those findings to human applications. Since humans are clearly not constrained to a very limited number of vocalization patterns, zebra finches do not provide an ideal model in which to study vocal learning in a more general, human-centric context.

The hunt for such a model has led to budgerigars (budgies), a type of songbird with no notable end to vocal learning. Not only can they continue to learn new songs [6], but they possess the capability to learn from humans

*More information may be found at akf1.web.rice.edu.

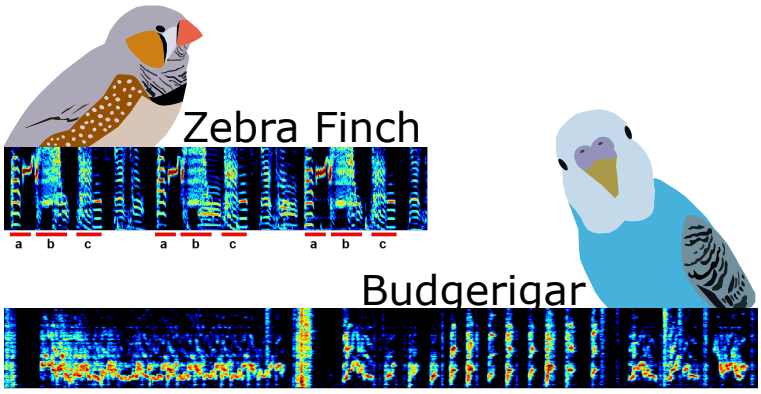

Figure 1: An example of the difference in song of zebra finches versus budgies via spectograms of vocalization, in which audio data is transformed into the frequency domain (x-axis representing time in milliseconds, and y-axis representing frequency in kHz). As is visible above, the zebra finch's song is repeated — the entirety of the finch's vocalization is composed of a single, unaltered song. The budgie, however, has variation in length, pitch, repetition, and more. Thus, we are more concerned with the complex vocalizations of budgies, as they lead one step closer to understanding human speech patterns.

— both in terms of vocalization and behavior. These birds have been known to repeat phrases they have overheard from conversations their owners have, as well as parroting some dance-like motions; funny how language works, as English and Spanish have adopted verbs derived from their respective words for "parrot" to describe humans imitating observed behavior in this very way. This interspecific mimicry is of high interest to motor learning investigators, as this phenomenon indicates a much greater complexity than simply mirroring another animal given the distinct differences in vocal range and body composition. Thus, budgies evoke exciting questions regarding vocalization from the motor learning community, as a step forward towards understanding human behavior.

It is not a far stretch to wonder how broad this constant learning ability extends — beyond vocalization, can budgies learn new *behaviors* throughout their lives? Being the extremely social creatures they are [8], it is natural to inquire whether or not there are individually specific behaviors, and, if so, the timeframe over which budgies learn from each other, the social bond (or lack thereof) required, and more. To thoroughly address these questions, it is imperative to build and maintain a dictionary of budgie behaviors and, on a smaller scale, sub-behaviors that can be rearranged into different, larger timescale behaviors [9]. To do so, we propose the following methodology for capturing and analyzing video data of interacting budgies.

2 Methods

Our methodology for obtaining and processing such data from budgies may be broken into the following steps: video acquisition, keynote identification/tracking, and spatio-temporal data mining.

2.1 Video Acquisition

In order to accurately identify different behaviors, it is necessary the acquisition paradigm be designed in such a way to provide ample room for free, unconstrained behavior, an unobstructed view for multiple high-speed cameras, and the interaction of multiple budgies with the option of individual granularity. Thus, a simple solution is to construct a clear, acrylic box with a similarly clear divider in the middle, allowing for two budgies to be housed separately in the box at once. However, the clarity of this divider facilitates interaction between the two budgies. Not only are we able to document individual behavior, then, but we are capable of capturing social learning behaviors and the difference in behavior distributions in the presence of another budgie. See Figure 2 for an example layout of our budgie pair behavior box, with three 2-dimensional cameras recording different viewpoints.

Budgie movement, however, typically occurs at a high rate of speed — this poses a challenge for video capture, as the camera frame rate must be relatively high. Ensuring clear frames throughout the entirety of the video would require a frame rate at least twice the maximum speed of budgie motion. In our case, we utilized FLIR Blackfly S USB3 cameras, which have a maximum frame rate of about 522 frames per second and can be externally triggered for capture. This latter fact is attractive for us in the case of dimensionality, as external triggers allow for synchronized

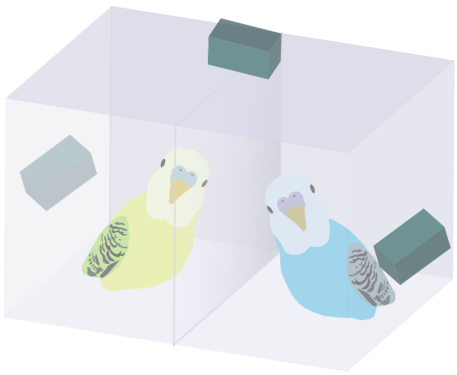

Figure 2: An example of the interacting budgie setup. Being separated by a clear, acrylic screen allows us to individually monitor each budgie's behavior, while they are free to communicate with each other. In this figure, 3 cameras are mounted around the budgie enclosure. These different views should enable multi-dimensional analysis at a later date, if necessary. Otherwise, these cameras provide options in terms of preferred tracking angle.

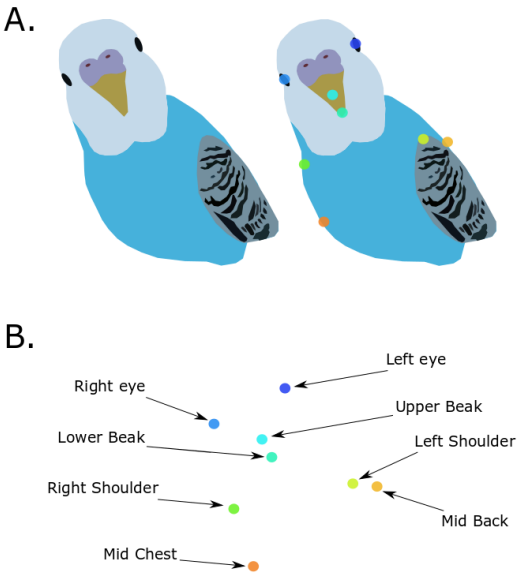

Figure 3: The labeling process of budgies, both as an input and an output using DeepLabCut [7]. A) An example of what an unlabeled budgie versus a labeled budgie would look like, prepared as part of a dataset. As many or few body parts may be labeled as required, so long as the dataset is large enough to encapsulate the variability of the experimental paradigm. B) Example labels of a budgie. The x-y coordinates of these labels are all we care about, as each label will have it's own path through the context.

video acquisition, making it easier to perfectly align videos from separate views. Beyond the sake of synchronization, this external trigger also facilitates implementation of video recording only during audio regions of interest — in other words, we can maintain a small buffer of video data, constantly updating with realtime frame capture, and only write to a file when at least one budgie vocalizes. An additional perk of using these cameras is rooted in the fact that some networks we want to apply to this paradigm have been shown to work well with these specific cameras [2], but that will be discussed later.

Fixing camera positions is a simple yet critical portion of setting up this paradigm. By ensuring there is no camera movement between frames, videos or days, we facilitate alignment of different viewpoints, as well as inherent consistency in the context. Such stability in acquisition additionally permits quick comparison between different days, budgies and pairs of interacting budgies.

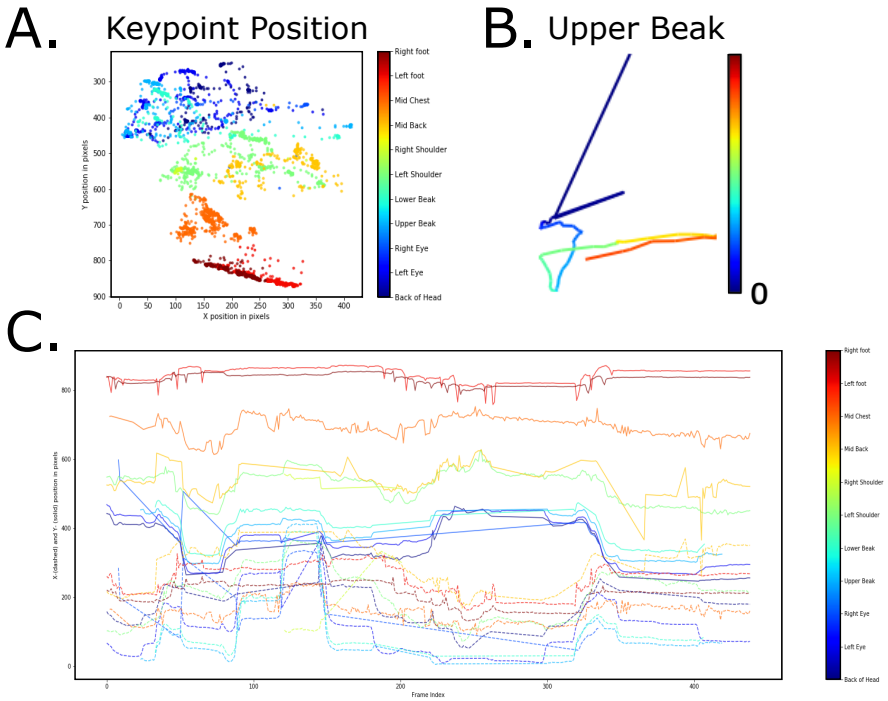

Figure 4: Plots generated by position tracking analysis. A) The x and y pixel locations of different body parts, as denoted by color, per frame of the video. It is easy to see that, despite movement, the same body parts tend to cluster near each other in space. B) Motion tracking of just the upper beak over the first five seconds of recording. The heatmap represents time since the first frame (frame 0, or the darkest blue) and becomes warmer as the frame becomes more recent. C) Different body parts, once again denoted by color, and their x (dashed lines) and y (solid lines) coordinates in each frame.

2.2 Position Tracking

To interpret the above acquired data, it is necessary to translate their behavior from a video format to another, which we can pass into a network architecture. Thus, by choosing specific keynotes, and converting their pixel positions throughout the videos to time-series data of this information, we are able to pre-process the budgie behavior into a format on which we can operate. At the moment, we are using DeepLabCut [7] to label keypoints throughout budgie videos. See Figure 4 for various examples of such a format, which we will motivate as follows.

Visualizing each body part's movement through space without considering time is useful for evaluating general motion through space. For example, in Figure 4A it is clear to see that the budgie is only scooting along it's perch, as the left and right foot markers. In videos in which budgie movement is not extremely varied, but rather in which this overall position plot should be decently predictable, it is useful to generate such a plot to evaluate the accuracy of the network labeling the budgies. Generally, if we can cluster body parts together, and it matches overall plotting we would expect, then we know we have a relatively high performing tracker.

Converting this data into the time domain, we can better visualize how budgie behavior alters over the course of the analyzed video in two dimensions (x and y axes), as seen in Figure 4C. This allows for a finer resolution in terms of behavioral quantification, as perhaps a certain behavior has a much greater change in position along one axis than another. Thus, extracting and identifying said behavior may be more obvious considering mainly the x-axis over the y-axis, or vice versa. In some cases, however, we are only interested in the movement of one body part. Consider tracking only the upper beak — maybe we want to observe how budgies alter beak motion in the presence of one another, or even just quantify the trajectory through space. Figure 4B depicts such a case, in which we differentiate time by color, and plot the x-y coordinates of the upper beak throughout the video. These position tracking techniques allow us to quantify behavior in ways our eyes simply cannot, at a smaller level and over a wider range of data.

2.3 Unsupervised Behavioral Identification

Once position tracking data in the time domain has been obtained, the issue of how to interpret this data arises. But, what do these time-series mean to us? They signify passage through space not only over the entirety of the video, but more importantly over distinct, small timeframes. Analyzing movement on the order of milliseconds thus allows us to investigate sub-behaviors that, when combined, may manifest themselves in different observable behaviors [9]. However, doing so becomes difficult when our objective is to build a basis of all budgie behavior, not just of known budgie behavior, since any human labelers would be biased to label only behaviors with which they are already familiar, or that qualify as their personal definition of similar. Additionally, they would not be able to distinguish sub-behaviors on nearly as small of a scale as a computer would, leading us to believe the most promising solution to be the application of unsupervised machine learning algorithms to this context.

Within the field of spatio-temporal data mining (STDM) lies the sub-field of frequent pattern identification. Essentially, this area of STDM concerns itself with identifying particular modulations of data that typically occur in sequence, and larger scale behaviors that typically accompany each other [1]. By unconvering these atomic portions of behavior, we theoretically should be able to construct any complex budgie behavior given our discovered sub-behaviors. Hence, moving forward, we want these algorithms to be the focal point of our pipeline.

3 Discussion and Future Works

Successful construction of such a set of budgie behaviors would enable eventual, realistic simulations. Beyond simple clustering algorithms, numerous unsupervised methodologies for behavioral identification have been suggested for other animal models [5] and would be worth looking into. As mentioned earlier, such a feat would grant significantly greater control to the songbird motor learning community in terms of experimental design and thus the inquiries they are able to investigate. However, before this may occur, there are several more alterations that must be made to our video data processing pipeline.

It may be beneficial to stitch together the multiple viewpoints into a cohesive, 2.5- or 3-dimensional model of a behaving budgie. Multiple 2-dimensional cameras have already been used to collect synchronous video data of various items in action and build a 3-dimensional model, even tracking occluded keypoints [10]. If we can accomplish a similar feat on budgies, then the task of tracking budgies becomes trivial, as does the task of creating 3-dimensional data — currently, we would need to align all views to generate such a model. However, by building a 3-dimensional schema for what a budgie should look like, applying it to a 2-dimensional video should allow for extrapolation of key point locations in 3-dimensional space.

Another option for position tracking would be to adopt multi-person position tracking networks that are already proven to perform well on humans, even in occlusive settings [2], and apply them to our situation via transfer learning. A huge benefit of doing so would eliminate the need for a barrier between the two interacting budgies, thus permitting them to socialize in a more natural manner, possibly inviting the exhibition of new behaviors we would not have observed otherwise. Overall, this project seems feasible to accomplish within the next year.

4 Acknowledgements

This work was supported by National Science Foundation (DBI-1707312) and the Cornell University Neurotechnology Hub. A huge thank you to Matthew Pool as well for his guidance in lab, around campus and through the summer! His wisdom and musicality were (and still are) much appreciated.

References

[1] Atluri, Gowtham, Anuj Karpatne, and Vipin Kumar. "Spatio-temporal data mining: A survey of problems and methods." In *ACM Computing Surveys (CSUR) 51.4 (2018): 83.*

[2] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2017.

[3] Chakraborty, Mukta, and Erich D. Jarvis. "Brain evolution by brain pathway duplication." In *Philosophical Transactions of the Royal Society B: Biological Sciences 370.1684 (2015): 20150056.*

[4] Hahnloser, Richard HR, Alexay A. Kozhevnikov, and Michale S. Fee. "An ultra-sparse code underliesthe generation of neural sequences in a songbird." In *Nature 419.6902*, (2002): 65.

[5] Klibaite, Ugne, et al. "An unsupervised method for quantifying the behavior of paired animals." In *Physical biology 14.1*, (2017): 015006.

[6] Manabe, Kazuchika, and Robert J. Dooling. "Control of vocal production in budgerigars (Melopsittacus undulatus): Selective reinforcement, call differentiation, and stimulus control." *Behavioural processes 41.2*, (1997): 117-132.

[7] Mathis, Alexander, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Publishing Group*, 2018.

[8] Sewall, Kendra B., Anna M. Young, and Timothy F. Wright.a "Social calls provide novel insights into the evolution of vocal learning." In *Animal behaviour 120*, (2016): 163-172.

[9] Wiltschko, Alexander B., et al. "Mapping sub-second structure in mouse behavior." In *Neuron 88.6*, (2015): 1121-1135.

[10] Wu, Shangzhe, Christian Rupprecht, and Andrea Vedaldi. "Photo-geometric autoencoding to learn 3D objects from unlabelled images." In *arXiv preprint arXiv:1906.01568* , (2019).