

# Multimedia Signal Processing for Behavioral Quantification in Neuroscience

Peter Andrews<sup>1</sup>, Sigal Saar<sup>2</sup>, Haibin Wang<sup>1</sup>, Dan Valente<sup>1</sup>, Jihène Serkhane<sup>1</sup>,  
Ofar Tchernichovski<sup>2</sup>, Ilan Golani<sup>3</sup>, Partha P. Mitra<sup>1</sup>

<sup>1</sup>Cold Spring Harbor Laboratory  
1 Bungtown Road  
Cold Spring Harbor, NY 11724, USA  
001-516-367-8800

<sup>2</sup>City College of New York  
138<sup>th</sup> Street & Convent Avenue  
New York, NY 10031, USA  
001-212-650-7000

<sup>3</sup>Tel Aviv University  
Ramat Aviv  
Tel Aviv 69978, Israel  
72-3-640-9391

andrewsp@cshl.edu, saarsigal@yahoo.com, haibinw@cshl.edu, valente@cshl.edu,  
serkhane@cshl.edu, ofer@sci.ccnyc.cuny.edu, ilan99@post.tau.ac.il, mitra@cshl.edu

## ABSTRACT

While there have been great advances in quantification of the genotype of organisms, including full genomes for many species, the quantification of phenotype is at a comparatively primitive stage. Part of the reason is technical difficulty: the phenotype covers a wide range of characteristics, ranging from static morphological features, to dynamic behavior. The latter poses challenges that are in the area of multimedia signal processing. Automated analysis of video and audio recordings of animal and human behavior is a growing area of research, ranging from the behavioral phenotyping of genetically modified mice or drosophila to the study of song learning in birds and speech acquisition in human infants. This paper reviews recent advances and identifies key problems for a range of behavior experiments that use audio and video recording. This research area offers both research challenges and an application domain for advanced multimedia signal processing. There are a number of MMSP tools that now exist which are directly relevant for behavioral quantification, such as speech recognition, video analysis and more recently, wired and wireless sensor networks for surveillance. The research challenge is to adapt these tools and to develop new ones required for studying human and animal behavior in a high throughput manner while minimizing human intervention. In contrast with consumer applications, in the research arena there is less of a penalty for computational complexity, so that algorithmic quality can be maximized through the utilization of larger computational resources that are available to the biomedical researcher.

## Categories and Subject Descriptors

J.3 [Life and Medical Sciences]: Biology and genetics – neuroscience

I.4.0 [Image Processing and Computer Vision]: General

J.4 [Social and Behavioral Science]: Neuroscience

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'06, October 23–27, 2006, Santa Barbara, California, USA.  
Copyright 2006 ACM 1-59593-447-2/06/0010...\$5.00.

**General Terms:** Experimentation, Measurement, Algorithms

**Keywords:** audio, behavior, birdsong, human, infant, locomotion, mouse, multimedia, neuroscience, phenotype, signal processing, video, vocal development, zebra finch

## 1. INTRODUCTION

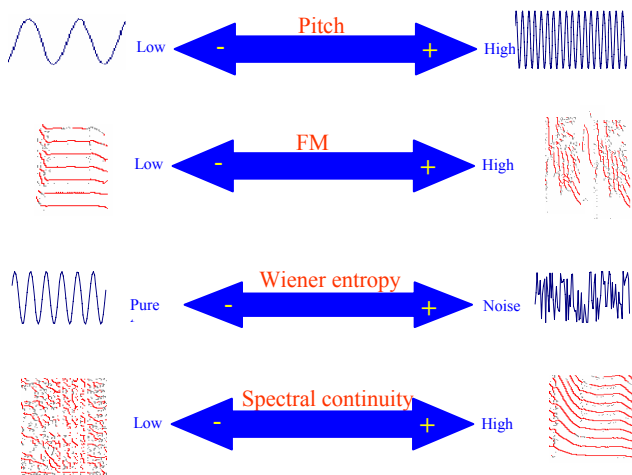
Understanding how brains work is perhaps the “final frontier” for science. Technical advances have played a major role in contemporary progress in neuroscience, as illustrated by the rapid growth in the use of brain imaging techniques. While not quite as prominent in the public gaze, a small technical revolution is currently taking place in the area of automated analysis of animal and human behavior. Quantification of behavior is critical to understanding brains, since behavior is the output of brain function. The scientific study of animal behavior or Ethology is a discipline that was developed in the early and mid-twentieth century. What we are now witnessing is the growth of the field of Quantitative Ethology, aided by computational analyses of digitized recordings of behavior. Multimedia signal processing (MMSP) is central to this field, since behavior is typically digitized by making audio and/or video recordings. In this paper we examine four case studies of Quantitative Ethology using audio and video signal processing. The field is still in its infancy and presents research challenges in MMSP with both practical and scientific interest.

Many neuroscience behavior experiments can be characterized by the sets of time-varying sensory inputs presented to the subject and the behavioral plus physiological output signals which result. The output may be used as feedback during the experiment and/or recorded for later analysis. Audio and video presentation and recording are often an integral part of an experimental paradigm, and appropriate signal processing and general analysis techniques need to be applied in order to properly interpret the results or drive the experiment in real-time.

Audio has long been a staple channel as both input and output, but only relatively recently have advanced signal processing techniques entered into neuroscience for its analysis. The usage of simultaneous video recordings is also increasing as mass-market consumer electronics advances lower the monetary, computing power, bandwidth and storage capacity costs. Video is often an unobtrusive addition to an existing setup and may help to explain a finding or an anomaly later in the analysis. However, the additional complexity

and heterogeneity of video data compared to audio presents greater difficulties for the average neuroscientist.

Our goal is to be illustrative rather than exhaustive in our review. In each case, we present some biological background to motivate the signal and image processing methods, followed by some details of the algorithms used, and biological findings. Some of the techniques that we present in the paper are relatively elementary, but have already produced important advances in the field, thus illustrating the potential for further advances. In all cases, the data volumes are large, and may easily reach into the terabyte range for audio alone. We will not focus on the issues that arise from managing such large volumes of data; these do form an integral part of the data analysis challenge. We will also not attempt to cover any of the literature on adult human speech, although this is perhaps the oldest and most well developed form of behavioral quantification. Finally, we will not cover areas such as annotation of digital television or film, that are already established areas in MMSP that are nevertheless related to the subject matter discussed in this paper. We chose our examples to illustrate the potential for MMSP in neuroscience research.



**Figure 2.1.** (a) Pitch is a measure of the period of the sound and its value is high when the period is short and low when the period is long. (b) Frequency modulation is a measure of the mean slope of frequency contours. (c) Wiener entropy (a measure of randomness) is high when the waveform is random, and low when the waveform is of pure tone. (d) The spectral continuity value is high when the contours are long and low when the contours are short.

## 2. AUDIO ANALYSIS

### 2.1 Birdsong

All true songbirds (in biological terms: order Passeriformes, suborder Oscines) are thought to develop their songs by reference to auditory information [28]. This can take the form of improvisation or imitation [19, 33, 51]. In both cases, vocal development is guided by auditory feedback [24], [39]. The quantification of the similarity between two songs is necessary to study a number of phenomena, such as the choice of model song and the timing of model acquisition by songbirds, the social context in which imitation occurs, and the neural basis for song learning behavior.

Early analysis of birdsong was based on visual inspection of spectrograms [50]. Songs may be partitioned into ‘syllables’ or ‘notes’, defined as continuous sounds preceded and followed by silent intervals or by abrupt changes in frequency. In early studies of vocal learning, notes of the pupil’s song that best matched the tutor’s song were identified, and assigned a numeric similarity score [40, 46] through visual inspection. Although the visual approach to scoring similarity made good use of the human eye and brain to recognize patterns, such measures of similarity were arbitrary and idiosyncratic. It was recognized a while ago that a quantitative, automated scoring of similarity was necessary to improve the quality of the measurements and facilitate comparisons between results obtained by different laboratories. However, it took several attempts by a number of groups before such methods could be used in practice.

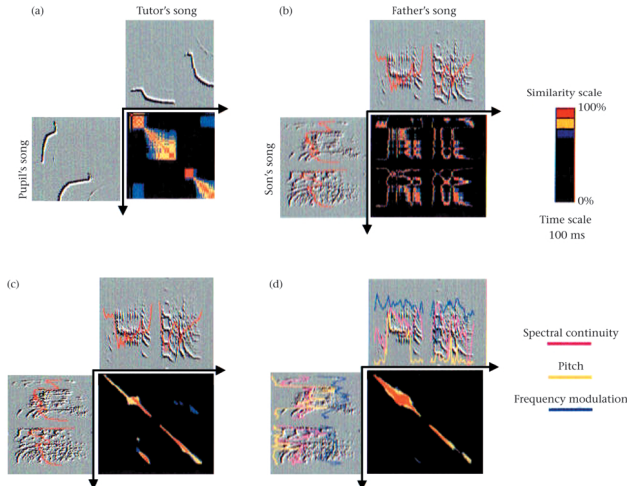
Here an analytical framework is described for the automated characterization of the songbird vocalization, using zebra finch as an example [18, 47]. The approach uses a robust spectral analysis technique that identifies those acoustic features that have good articulatory correlates. The articulatory features are based on in vitro observations and theoretical modeling of sound production in an isolated syrinx [12]. Some results are presented here from the application of this method to a large database of zebra finch songs<sup>1</sup>, illustrating the dynamics of the vocal imitation process [45] and the effects of sleep on the developmental learning of bird song [8]. The methods were developed in collaborative work involving two of the authors of the paper.

#### 2.1.1 Continuous recording of birdsong and automated measurement of song similarity

It is important to have continuous recording of bird song in order to study the vocal learning process. In our recording and analysis system, the training regimen for each bird is automated and song playbacks are delivered in response to key pecks. Once the bird has been placed in the training box, the system records its vocalization continuously. A song recognition procedure [44] detects and records the songs, discarding isolated calls and cage noises.

Rather than using the spectrogram directly to analyze the audio signals, the method uses time and frequency derivatives of the spectrogram in the time-frequency plane. These derivatives can be estimated robustly using Multitaper spectral methods [48], [49] and remove slowly varying noise sources (including gain noise). Zero crossings of the derivatives are used to estimate the time dependent positions of peaks in the spectrum (peak frequency contours), providing a spectral representation that is not blurred by time-frequency uncertainty. This facilitates the accurate detection of events in the time frequency plane. The spectrogram is high dimensional; it is easier to proceed by reducing dimensionality by introducing a small set of features (see Figure 2.1). A feature vector is used to represent each time window of a song. To arrive at an overall score of similarity, the features are normalized and combined using appropriate weights.

<sup>1</sup> Currently at over 100 birds, with every song made during their lifetime recorded at 16 bit 44.1 KHz (100-150 hours total per bird), for a total of over 10 TB of data.



**Figure 2.2. Similarity measure improves as comparisons include longer intervals and more features. (a) Similarity matrix across pitch values of identical, artificial sounds. Because each of these simple sounds has a unique pitch, the similarity matrix shows high similarity values across the diagonal and low values elsewhere. Comparing complex sounds would rarely give such result. As shown in (b), although the songs are similar, high similarity across features (in this case, of Wiener entropy values) are scattered. (c) Ambiguity is reduced when we compare Wiener entropy curves between 70-ms intervals. (d) A combined similarity matrix between 70-ms intervals across features. High similarity values are now restricted to the diagonal, indicating that each of the notes of the father's song was imitated by his son in a sequential order. The curves overlaying the time–frequency derivative in (d) correspond to spectral continuity, pitch and frequency modulation.**

A short timescale similarity matrix is defined by taking pair-wise Euclidean distances between features corresponding to different time slices in the song. The comparison of pitch with an artificial tutor-pupil is shown in Figure 2.2.a. In this case, strong similarity is shown by a bright and concentrated diagonal. In reality, however, different windows often share similar patterns of power spectrum. It follows that high similarity values may occur in off-diagonal locations of the similarity matrix in addition to diagonal locations even for similar songs, as illustrated in Figure 2.2.b (7 ms time window). The solution is to compare intervals consisting of several windows. If such intervals are sufficiently long, they will contain enough information to identify a unique song segment. The combination of multiple features can also improve the comparison (Figure 2.2.c and 2.2.d). The final score of similarity combines the two scales: the ‘large scale’ (70 ms) is used for reducing ambiguity, while the ‘small scale’ (9 ms) is used to obtain a fine-grained quantification of similarity.

### 2.1.2 Dynamics of the vocal imitation process

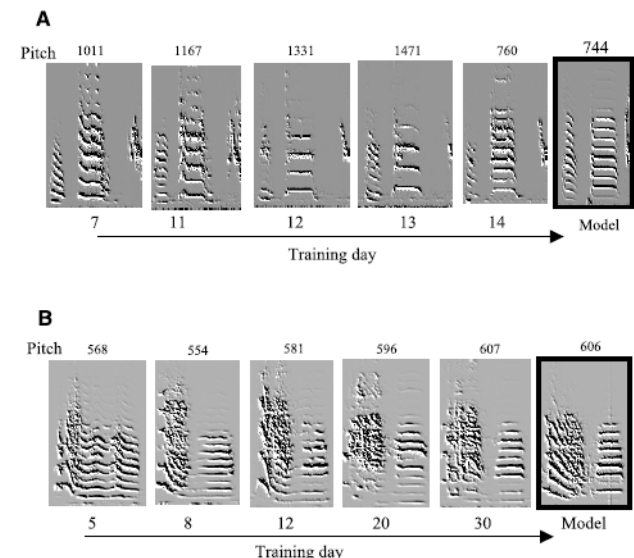
Interesting dynamics may occur during the vocal imitation process of zebra finches [45]. An imitation trajectory could take a path leading directly from the acoustic features of sounds produced before exposure, to those of a target sound present in the model song [5]. Alternatively, as illustrated in Figure 2.3.a, an imitation trajectory might deviate from a direct path due to constraints arising

from neurophysiology or the biophysics of the vocal apparatus. The figure shows a period doubling of the vocal pitch during development. We have also observed other ‘nonlinear’ imitation trajectories during the vocal imitation process, which indicate the presence of interesting dynamics in the development of the underlying neural networks. In particular, we have observed similarities between the developmental trajectories of zebra finch song, and the vocal development of human infants during the babbling stage. More details of this study may be found in the original papers.

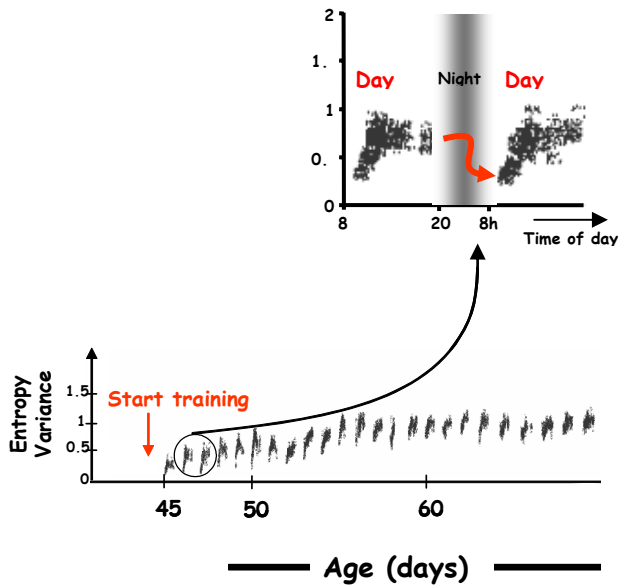
### 2.1.3 Sleep effects on the developmental learning of bird song

The effect of sleep on the developing zebra finch was quantified by comparing the statistical properties of pre-identified syllable classes produced at different times. To establish a baseline for vocal changes and to estimate our measurement errors, we compared differences across two random samples of 100 songs produced during the same day. For similar 100 song samples taken on successive days, we find that the day-to-day vocal changes are higher than within-day changes. Strikingly, the evening-to-next-morning changes are higher than the midday-to-midday changes.

Because vocal changes after 12 h of night-sleep are larger than the overall changes that occur during 24 h, vocal changes must oscillate during a daily cycle, as demonstrated in Figure 2.4. Pooling all the variance features show that during the early syllable development,



**Figure 2.3. Indirect and direct approaches to the imitation of harmonic stacks. (A) Spectral derivatives of a developing harmonic stack in reference to a syllable from a model song (left). The pitch of the harmonic stack is given at the top of each panel. A quantitative examination of the pitch error between this developing harmonic and the model harmonic stack shows a gradual increase of error, followed by an abrupt period doubling that reduced the error in a single step (right). The graph presents the mean pitch values of harmonic stacks produced by this bird across 30-s samples of birdsong recorded on each training day. (B) An example of harmonic stack imitation where pitch error gradually decreased until a match to the model syllable was reached.**



**Figure 2.4.** A song feature that shows sleep-associated oscillations between days 45 and 60.

the diversity in syllable features is low in the morning compared to the previous evening. The decrease in syllable structure in the morning may suggest that song is less structured and more primitive after sleep, but only during days of rapid learning.

The overall magnitude of post-sleep deterioration during development was found to be positively correlated with the eventual similarity to the model song. Thus, birds that showed a large “sleep effect” on the song were also the better imitators. Control experiments show that these oscillations are not a result of “sleep inertia” or lack of practice, indicating the possible involvement of an active process, perhaps neural song-replay during sleep. We suggest that these oscillations correspond to competing demands of plasticity and consolidation during learning, creating repeated opportunities to reshape previously learned motor skills.

### 2.1.4 Discussion

The acoustic features used in our analysis are not necessarily the ‘best’ set for characterization of the bird songs, but bear a close relation to the articulatory variables involved in sound production. New features may be derived not only from deeper understanding of the sound production apparatus, but the statistical analysis of the songs as well.

Although vocal learning remains a highly complex phenomenon with many unknown aspects, the tools that we used simplify the objective study of its dynamics, and enable us to analyze changes in real time. Neural and peripheral recordings are becoming common as well, enabling integrated data analysis at many levels of the song production system. This should provide us with further insights, paving the way for identifying the molecular, cellular, and circuit events that must underlie the moment-to-moment progression toward vocal imitation.

## 2.2 Vocal Development in Human Infants

Adult human vocal communication, including speech, starts its development very early in life, probably even before the first breath (for a review see [41], pp. 7-33 and 93-94). In [29], a follow-up study of two infants’ vocal behavior from birth to one year old was conducted using traditional ethological observation. The segmentation criteria used was based on the breath unit, and the vocalizations were classified according to the type of phonation and the type of articulatory movement<sup>2</sup> performed in each breath unit, with no reference to adult language (see [26] for details). The authors could distinguish 6 stages in early speech development, each of which starts with the use of a new kind of vocal behavior<sup>3</sup>.

At stage I (birth), phonation is continuous, the prosodic patterns are simple and there is no articulatory movement within the breath unit: sound production is mainly driven by the larynx activity.

At stage II (6 weeks), infants start to make use of interrupted phonation: this gives rise to the first rhythmic vocalizations.

Stage III (10 weeks) infants start to produce vocalizations with one articulatory movement, combined with continuous phonation (explored at stage I) or interrupted phonation (experienced in stage II).

Stage IV (20 weeks) is characterized by a diversification of the phonation patterns in terms of intonation, duration and intensity. Furthermore, the frequency of the third stage unitary movements decreases while that of continuous phonation without articulatory movement increases.

Stage V (26 weeks) is marked by the appearance of vocalizations achieved by multiple articulatory movements, be they repeated or varied, combined with continuous or interrupted phonation (combinations coming from stage III) and with a great variety of prosodic patterns (diversity stemming from stage IV). This behavior corresponds to *babbling*, that the authors define as a “repetition or a combination of articulatory movements during one single expiration cycle with interrupted or continued phonation” [52].

At Stage VI (40 weeks), all previous vocalizations start to be used to refer to things, such as events, in specific situations. These are the *first words*.

In sum, the study in [25] shows that early vocal behavior may develop in a hierarchically structured way with temporal and functional links between the abilities of a given stage and those of the previous stages. Such a developmental process has to be investigated quantitatively using detailed acoustic characterization of infant vocalizations, at the population level and with attention to individual variability in the course of development.

<sup>2</sup> Articulatory movement is defined here as “all kinds of constrictions or frictions that can be made by means of articulatory organs”.

<sup>3</sup> The approximate age of the beginning of each stage is displayed between parentheses. At each stage the vocalizations introduced at previous stages are still in use, but lessen progressively in the course of development.



However, research on how vocal communication evolves in infancy mainly focuses on *language* acquisition, with special interest in perception<sup>4</sup> of abstract linguistic categories (*phonemes*). When production<sup>5</sup> is studied, infant vocalizations are identified using phonetic *transcription*<sup>6</sup>.

The phonetic transcription methods in use raise two main issues: (i) the transcribers tend to select and categorize infant vocalizations under the influence of their language background (see [20] for a foreign language example); and (ii) since phonetic transcription is designed to describe adult languages, applying it to infant sounds is making the implicit assumption that the sounds are produced the same way in adulthood and in infancy. Although some speech development studies are based on objective acoustic measurements of infant vocalizations: [22], [29], [36] and [53], they were conducted on small numbers of infants, with a small amount of data per infant collected at relatively few time points in the course of early development.

In contrast with songbirds, *human infants cannot be isolated in experimental boxes*: the recording environments are invariably contaminated by adults talking, sounds of nearby children and all kinds of noise. This introduces privacy concerns on behalf of the infants, family and caregivers, so the participants are often the few children of the investigators and their acquaintances. Moreover, to date, no automated signal processing tools are available in the domain to detect, to separate from adult voices and background noise in the recording environment, and/or to acoustically analyze human infant sounds. As a consequence, researchers tend to manually isolate infant sounds that are (i) free from *any* source of noise, (ii) speech-like only, and (iii) acoustically describable according to features, such as resonance frequencies that can be reliably measured in the case of *adult* speech but not in the case of infant high-pitch sounds (see Figure 2.5). This leads to a lack of statistical significance of the results and impedes accurate investigation into infant vocal behavior over time.

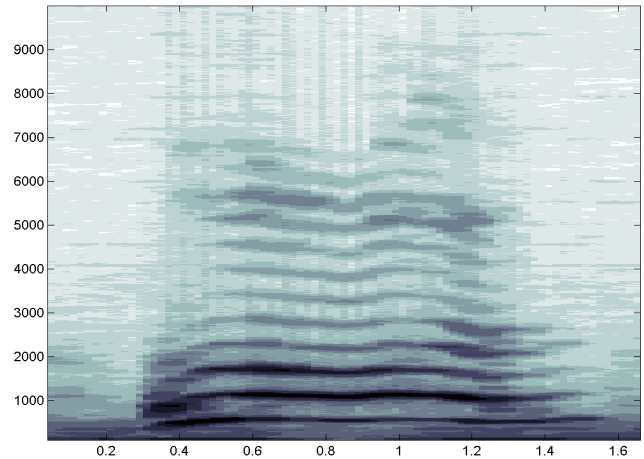
Based on this brief review, it should be clear that the quantitative study of infant vocal development is an area of research that will greatly benefit from the application of MMSP techniques, but there are other challenges to be solved as well for the field to advance. The lack of data-sharing and effective tools may be addressed by bringing together a consortium of researchers with different skills to jointly work on the problem. Analysis software for segmentation, annotation and supervised or unsupervised classification can then be written, which would increase the value of existing data and hopefully encourage more to be taken. If the software could accurately remove adult speech, many privacy and data quality issues could be solved as well. Success in this endeavor would engender trust, which may lead to greater data sharing and perhaps shared repositories or joint large scale studies. While we have only reviewed the analysis of audio data, simultaneous video is often

---

<sup>4</sup> Roughly, perception is the receiver's ability to process incoming signal.

<sup>5</sup> Production is the emitter's ability to make and send outgoing signals.

<sup>6</sup> In phonetic transcription, adults listen to the infant, cut her/his stream into "consonant" and "vowel" pieces and classify each piece as part of a given phonetic category, that is, a consonant or a vowel defined as in the International Phonetic Alphabet.



**Figure 2.5.** A Multitaper spectrogram showing a 6 month old infant vocalization. The horizontal axis is time in seconds, the vertical is frequency in Hz. Colored shading indicates power.

available and could be useful for tasks like sound segmentation and source identification using facial movement and body gestures. Similar considerations as audio apply to the analysis of video data: current efforts are based largely on manual segmentation and classification, and there is much scope for research tools using MMSP.

### 3. VIDEO ANALYSIS

#### 3.1 Rodent Locomotion

The characterization of phenotype in animals often begins with a proper analysis and categorization of locomotor and exploratory behaviors. Not only do these investigations equip researchers with much-needed measures of behavior for use in neurophysiological studies, but they also offer a unique insight into the brain of the animal. When used in conjunction with genetic techniques, a thorough investigation of locomotor behavior can help provide an outside-in description of the animal that connects behavior to neural structure and function to genetics. This description can then help elucidate the mechanisms for complex behaviors such as learning and memory, attention, anxiety, and fear. Genetically modified rodents serve as models for neurological and neuropsychiatric disorders, and behavioral quantification is important in assessing the effect of pharmacological and other therapeutic techniques in these model organisms. Thus, video analysis of rodent behavior plays an important role in research into mental health as well as drug development.

Because of this, locomotor behavior in the Open Field is a widely used test in behavioral neurosciences [1], particularly in the study of genetically engineered mice [3]. In a recent study conducted concurrently in three laboratories, and following an identical protocol, it has been shown, however, that many results were not replicable across laboratories [7]. A subsequent study has further shown that a major source for the replicability problem is the absence of an appropriate measurement methodology [21]. Appropriate tools for data acquisition, data preparation for analysis, isolation of relevant kinematic variables, and segmentation of the data time series into units are thus prerequisites for obtaining results

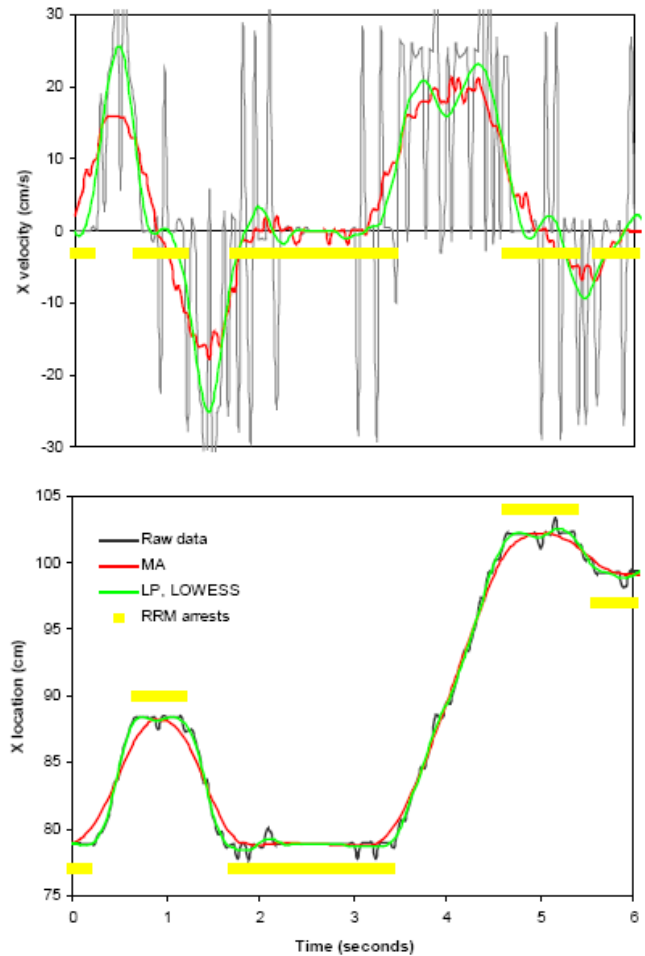
that are replicable across laboratories. In what follows we present several examples of such tools.

The measurement of rodent locomotion from video data is not straightforward. Different analysis techniques need to be used depending upon the resolution, frame rate and desired level of positional characterization. It was found that a high tracking rate is necessary to capture many properties of rodent behavior, especially in the fast-moving mouse [17]. The X and Y coordinates of the center of mass of a rodent in an open arena are perhaps the simplest measures to be made from a video recording, but there are pitfalls even there. The erratic movement of mice presents a challenge for measuring the position and velocity of the mice, which are the raw data from which further phenotypic measures could be derived. This is because the data include several sources of noise, and several modes of movement, each requiring a different method of smoothing. The raw location time-series includes 1. system noise, 2. changes in body shape, and 3. outliers.

Locomotor behavior consists of progression segments and lingering-in-place intervals. A Repeated Running Median (RRM) method is used to identify arrests. Subsequently the method of Locally Weighted Sum of Squares (LOWESS, see [6]) is used to smooth the path between arrests (see Figure 3.1). The RRM method isolates the arrest episodes so they will not be corrupted by the subsequent LOWESS smoothing. LOWESS solves both the precision noise and outliers problems using an iterative procedure similar to the Local Polynomials method (LP, [11]) with added robustness to outliers. As in the weighted LP, the first iteration of LOWESS fits a polynomial to the data in a time-window centered at  $t$ . The resulting polynomial, however, is used only as a first estimation. Each original data point is then assigned a weight according to its difference from its first estimation (residual). A larger residual (indicating a poorer fit) results in a smaller weight for the corresponding data point, implying it will be less relevant for computing the next fitted polynomial. At the extreme, a very large residual indicates that the point is an outlier, and it is assigned a zero weight, implying it will have no effect at all on the next iteration. In the second iteration of LOWESS the raw data in the window is fitted again with weighted LP, but this time using also the weights according to the residuals. In the original algorithm these iterations continue as above until no further change occurs, but practically it has been found that two iterations suffice.

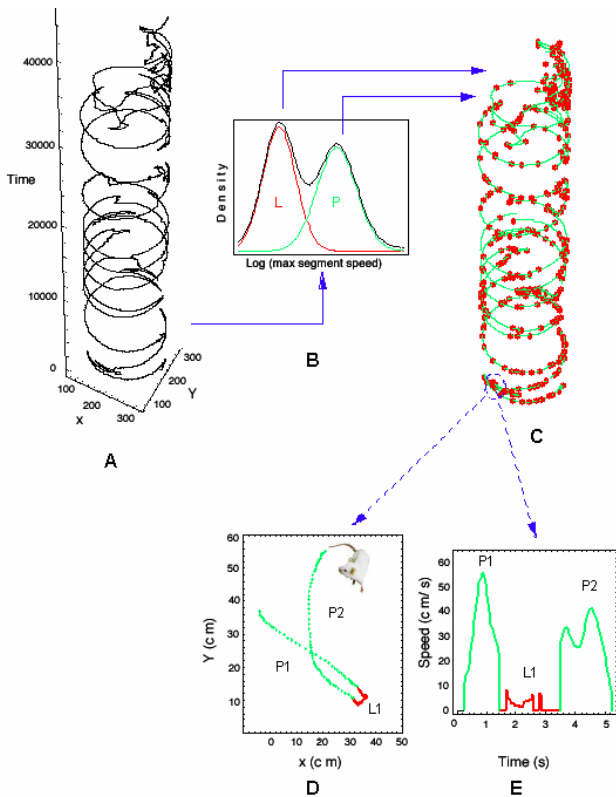
The smoothing algorithm is part of SEE, a publicly available software package (Strategy for the Exploration of Exploratory behavior [2]). The output of the SEE smoothing module includes the smoothed data, data of momentary speeds, and momentary accelerations.

After smoothing, the data time series is segmented into progression segments and lingering, or staying-in-place episodes. The method of deciding which time slices correspond to progression and lingering segments is shown in Figure 3.2.B. A density graph (a sliding window histogram) of the frequency distribution of peak speeds in inter-arrest segments reveals two distinct populations: low speed segments (lingering or stopping episodes) and high speed segments (progression segments). A Gaussian mixture model is used for recognizing the distinct components within the population of inter-arrest segments. The parameters of the mixture model are estimated by using the Expectation–maximization (EM) algorithm. The algorithm estimates the maximum likelihood parameters (proportions, means, and SDs) of a mixture with a given number of



**Figure 3.1. Location (bottom graph) and velocity (top graph) out of 6 s of a mouse’s movement, smoothed with several methods. “Raw” velocity (gray, top graph) is calculated as the differences between consecutive raw locations. MA (Moving Average) velocity (top, red) is calculated as the differences between consecutive MA smoothed locations. LP (Local Polynomial) velocity (top, green) is calculated directly by the LP. LOWESS-smoothed locations and velocities were almost identical to those calculated by LP since there were no outliers. Arrests (time range denoted by yellow stripes) are computed as zero velocity in the RRM (Repeated Running Medians) smoothed series. MA is applied with a window width of 15 frames. LP is applied with a window width of 15 frames and a degree of 2. RRM is applied with four iterations using half windows of 3, 2, 1, and 1.**

components. EM is an iterative algorithm that starts with user-given initial values, and incrementally improves the likelihood function until further iterations yield only a negligible improvement. The actual number of components of the model is determined by comparing the maximum likelihood value of an  $n$ -components mixture with that of a  $(n+1)$ -components mixture until the increased number of components increases the likelihood only marginally. For an exposition see [10]. In Figure 3.2.D the path plot and speed profile of two progression segments (P1 and P2) are separated by



**Figure 3.2.** (A) The path of a rodent using a 3D representation of X, Y, and time (in data points) is shown. (B) The distribution of speed peaks (dark line) is used to parse the data into segments: lingering (L) and progression (P). (C,D,E) The data can be treated as a string of these discrete units.

one lingering episode (L1). The typical properties of these units are used to quantify the behavior. For example, the segment acceleration endpoint (see Figure 3.2.E) is estimated by dividing the segment's speed peak by its duration, i.e., the aspect ratio of a segment's speed profile.

The progression segments are further classified into 4 types of patterns (Figure 3.3): So called "wallcursions", small incursions, median incursions, and large, across center incursions. This is done by using radial speed and maximal distance from wall for further classification of the segments population [30]. The careful smoothing, and the segmentation, which is based on intrinsically defined geometrical and statistical properties, increase the replicability of the behavioral measures.

We have described the video analysis methods that are currently in use to study rodent locomotor behavior. In these current studies, the



**Figure 3.3.** The path traced by a C57BL/6J mouse during a 30 min session in the Open Field, is partitioned by the SEE algorithms into four intrinsically distinct components.

animal is treated as a point object. The challenge for MMSP research would be to develop techniques that will move beyond this and enable the treatment of posture and position of individual limbs. Moreover, recent research shows that rodents vocalize in the ultrasonic range, thus opening up the possibilities of MMSP research into simultaneous video-ultrasonic audio recordings of rodent behavior. The importance of such tools to future advances in neuroscience cannot be over-emphasized.

### 3.2. *Drosophila* Behavior

The fruit fly *Drosophila melanogaster* has become ubiquitous in neurogenetic and neurobehavioral studies for many reasons. Sophisticated mutagenesis techniques have been developed for the fly (e.g. chemical mutagenesis, mobilization of transposable elements, selective breeding), and the short reproductive cycle and ease with which flies are maintained have greatly facilitated genetic engineering of these organisms. In addition, the *Drosophila* genome is known, and over half of *Drosophila* genes have direct human homologues. Add to this the rich repertoire of complex behaviors that fruit flies exhibit, and it becomes apparent that *Drosophila* is an incredibly useful organism for elucidating the genetic and neural mechanisms responsible for the complex behavioral phenotypes seen throughout nature.

Many behavioral assays have been developed to study complex behaviors in *Drosophila* such as attention [54, 55], sleep [16, 42], visual and olfactory learning and memory [31, 32, 37], courtship [15, 38], and aggression [27], [4]. Unfortunately, "fly psychology" is quite difficult; any inference made from these assays can be subject to anthropomorphic bias of the investigator. Therefore, the experimental designs and the resulting behavior require careful analysis if the results are to provide an accurate description of the phenomenon of interest.

The fact that the locomotor behavior of the fly is central to any behavioral assay is often overlooked in these studies; it is difficult to observe behavior without the fly moving some part of its body. In addition, many of these assays require the flies to have normal motor activity; therefore, measurement and characterization of locomotor activity is essential. Naturally, this is no easy task, and much of the difficulty actually lies in defining parameters that characterize such activity. Even so, there has been considerable progress in this area (see [34] for a review). Several parameters that are commonly used for locomotor activity characterization include spatial preference, total distance moved, mean and instantaneous walking speed, and the duration of bouts of activity and inactivity.

The techniques for measuring these parameters are quite varied and range from manually tracing the trajectory of larvae on glass plates [43] to radar motion detection of fly populations in tubes [23]. Unfortunately, many of these techniques suffer from a lack of

sufficient temporal resolution, cannot track individual flies in a population, or are very labor-intensive. Also, the most comprehensive descriptions of behavior come from direct, continuous observation of behaving flies on fine time scales; thus, it is no surprise that automated image acquisition and processing techniques promise to be some of the most useful methods for phenotypic characterization of fly behavior.

Whereas image processing techniques have been instrumental in elucidating the mechanics of flight in *Drosophila* [13, 14], thorough analyses of walking flies has only recently begun [35]. In most cases, the trajectory that the fly takes during a given assay yields the most comprehensive description for this locomotor behavior. From the image data, motion detection algorithms can be used to track the flies and map out their position in the environment<sup>7</sup>, similar to the work that has been performed in mammalian studies [9]. An example of such a procedure is seen in Figure 3.3 for the mouse. Image processing algorithms are also used to compute relevant parameters directly from the raw image data. Using these algorithms, one can then determine a fly's orientation in the environment, its spatial preference, its velocity etc.—many of the parameters that may give insight into the neurophysiological basis of fly behavior.

Many of the difficulties associated with traditional locomotor studies are solved with appropriate video acquisition and image processing. Entire populations of flies can be tracked, as well as individuals; variable frame rates allow for flexibility in temporal resolution of the behavior; and video allows one to easily determine spatial preferences of the flies. With high enough resolution, one would theoretically be able to closely reconstruct the movement of each body part using edge and motion detection. In addition, the fact that video is a relatively unobtrusive method allows a researcher to design an experiment so that the fly is in as close to a natural environment as possible and still gather useful, controlled data. Most important, automated acquisition, tracking, and parameter calculation shifts the human labor from acquisition to analysis.

The current methods of video acquisition, tracking, and image processing are still at a relatively rudimentary level in fly behavioral studies. In most cases, the fly is simply a white dot on a dark background (or vice versa) in a single-camera / single fly scenario, providing a relatively coarse resolution in large open-field experiments even when good cameras are used. In addition, most motion-detection algorithms assume the fly to be a point object, neglect any movement at finer scales, and cannot track more than one fly at a time. For a full characterization of behavior, improvements must be made in both the acquisition and analysis of the image data. For example, one would ideally want an arena equipped with a multi-camera array at different magnification levels to capture each foot-fall of the fly, the position of its thorax in relation to the orientation of the feet, and the position of the head, antennae, and proboscis in relation to the rest of the body. Furthermore, one could envision increasingly complex experimental setups, such as a high magnification camera which physically tracks the subject to allow detailed movies of a large arena at comparatively low data rate, together with a thermal imaging camera for detailed profiling of temperature controlled experiments.

---

<sup>7</sup>Common behavioral analysis software that performs such tracking are SEE, EthoVision, and DIAS.

Microphones can be added to detect song in courtship or aggression experiments and LED displays incorporated to provide the flies with visual stimuli. With these improvements in design, algorithms that determine the fly shape, heading, orientation of body parts, separation of individuals, and correlations between the video and audio data must be employed to extract all of the useful behavioral information. Clearly, sophisticated MMSP techniques must be utilized for analyzing the data from these biological studies.

## 5. CONCLUSIONS

The use of multimedia recordings and the importance of multimedia content analysis in neuroscience behavioral experiments will only increase as consumer technology improves and as researchers design more sophisticated and open-ended experimental setups. This trend will require neuroscientists to become familiar with and embrace MMSP methods for the proper analysis of that data.

As the study of song learning in zebra finch and vocal development in the human infant has shown, *detailed* acoustic analysis combined with robust signal processing are key when characterizing audio recordings of song and language. Automated techniques must segment any sounds into units appropriate for the subject, and robust measures need to be used for proper categorization of sounds. There is a need to replicate the existing zebra finch type studies in a human context by performing large-scale recordings and bringing automated analysis techniques to bear on the database. Simultaneous video recording is already common in both areas, but needs to be integrated more fully into the analysis.

The acquisition of video in neuroscience is not only limited by the state of the art of camera manufacture, but by the ability of the researcher to store and analyze the results. In the rodent case, progress was made by carefully smoothing and segmenting the measured positional data obtained from the video. Locomotion was then characterized according to distinct types of motion, such as lingering and progression along a wall or across a chord. Similar *Drosophila* locomotor studies are in progress, as well as experiments which use tethered flies in a flight simulator for behavioral assays.

By combining video, audio and additional channels, nearly complete characterizations of behavior are possible. Certainly, we will see large advances in characterization of behavioral phenotypes in the zebra finch, human, rodent and *Drosophila* during the next decade. It is likely that advances in MMSP techniques will be critical to these studies. Applications of MMSP techniques applied to behavioral data may even become as important as the application of algorithmic techniques from computer science to genomic data. Only by adding sophisticated phenotypic information extracted from behavior, to genotypic information and electrophysiological and imaging data, can we expect to gain a fuller understanding of brains and of animal behavior. Apart from scientific interest, automated behavioral quantification of animal models is crucial for biomedical research including the development of therapies for mental illness, thus providing MMSP researchers an opportunity to participate in this arena.

## 6. ACKNOWLEDGEMENTS

Work presented here was supported by grants R01MH071744 and R01NS50436 from the NIH to O.T. and P.P.M., and by awards from the DART Neurogenomics Alliance and the Swartz Foundation. Thanks to Anna Dvorkin for her assistance. Some figures in this review are identical with or similar to those previously published by



the authors elsewhere. Figure 2.1 is similar to a figure in [45]. Figure 2.2 is from [8]. Figure 2.3 is from [45]. Figure 2.4 is from part of a figure in [8]. Figure 3.1 is from [17]. Figure 3.2 is similar to a figure in [21]. Figure 3.3 is similar to a figure in [30].

## 7. REFERENCES

- [1] Archer, J., *Tests for emotionality in rats and mice: a review*. Anim Behav, 1973. **21**: p. 205-235.
- [2] Benjamini, Y., et al., *SEE Software Home Page*. 2006.
- [3] Bolivar, V., Cook, M. and Flaherty, L., *List of transgenic and knockout mice: behavioral profiles*. MAMM Genome, 2000. **11**: p. 260-274.
- [4] Chen, S., et al., *Fighting fruit flies: A model system for the study of aggression*. Proceedings Of The National Academy Of Sciences Of The United States Of America, 2002. **99**(8): p. 5664-5668.
- [5] Clark, C.W., Marler, P. and Beeman, K., *Quantitative-Analysis of Animal Vocal Phonology - an Application to Swamp Sparrow Song*. Ethology, 1987. **76**(2): p. 101-115.
- [6] Cleveland, W.S., *Robust locally weighted regression and smoothing scatterplots*. J Am Stat Assoc, 1977. **74**: p. 829-836.
- [7] Crabbe, J.C., Wahlsten, D. and Dudek, B.C., *Genetics of mouse behavior: interactions with laboratory environment*. Science, 1999. **284**: p. 1670-1672.
- [8] Deregnacourt, S., et al., *How sleep affects the developmental learning of bird song*. Nature, 2005. **433**(7027): p. 710-716.
- [9] Draï, D. and Golani, I., *SEE: a tool for the visualization and analysis of rodent exploratory behavior*. Neuroscience And Biobehavioral Reviews, 2001. **25**(5): p. 409-426.
- [10] Everitt, B.S., *Finite Mixture Distributions*. 1981, London: Chapman & Hall.
- [11] Fan, J. and Gijbels, I., *Local polynomial modeling and its applications*. 1996, London: Chapman & Hall.
- [12] Fee, M.S., Shraiman, B., Pesaran, B. and Mitra, P.P., *The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird*. Nature, 1998. **395**(6697): p. 67-71.
- [13] Fry, S.N., Sayaman, R. and Dickinson, M.H., *The aerodynamics of hovering flight in Drosophila*. Journal Of Experimental Biology, 2005. **208**(12): p. 2303-2318.
- [14] Frye, M.A. and Dickinson, M.H., *Closing the loop between neurobiology and flight behavior in Drosophila*. Current Opinion In Neurobiology, 2004. **14**(6): p. 729-736.
- [15] Greenspan, R.J. and Ferveur, J.F., *Courtship in Drosophila*. Annual Review Of Genetics, 2000. **34**: p. 205-232.
- [16] Greenspan, R.J., Tononi, G., Cirelli, C. and Shaw, P.J., *Sleep and the fruit fly*. Trends In Neurosciences, 2001. **24**(3): p. 142-145.
- [17] Hen, I., et al., *The dynamics of spatial behavior: how can robust smoothing techniques help?* Journal of Neuroscience Methods, 2004. **133**: p. 161-172.
- [18] Ho, C.E., Pesaran, B., Fee, M.S. and Mitra, P.P., *Characterization of the structure and variability of zebra finch song elements*. Proceedings of the Joint Symposium on Neural Computation, 1998. **5**: p. 76-83.
- [19] Immelmann, K., *Song development in the zebra finch and in other estrilled finches*. Bird Vocalization (Ed. by R. A. Hinde), 1969: p. 61-74.
- [20] Iverson, P., et al., *A perceptual interference account of acquisition difficulties for non-native phonemes*. Cognition, 2003. **87**: p. B47-B57.
- [21] Kafkafi, N., et al., *Genotype-environment interactions in mouse behavior: A way out of the problem*. Proceedings of the National Academy of Sciences, 2005. **102**(12): p. 4619-4624.
- [22] Kent, R.D. and Murray, A.D., *Acoustic features of infant vocalic utterances at 3, 6 and 9 months*. Journal of the Acoustical Society of America, 1982. **72**: p. 353-365.
- [23] Knoppien, P., van der Pers, J.N.C. and van Delden, W., *Quantification of locomotion and the effect of food deprivation on locomotor activity in Drosophila*. Journal of Insect Behavior, 2000. **13**(1): p. 27-43.
- [24] Konishi, M., *The role of auditory feedback in the control of vocalization in the white-crowned sparrow*. Zeitschrift für Tierpsychologie, 1965. **22**: p. 770-783.
- [25] Koopmans-Van Beinum, F. and Van Der Stelt, J., *Early stages in the development of speech movements*, in *Precursors of Early Speech*, B. Lindblom, e.Z., R., Editor. 1986, Stockton Press: New York. p. 37-49.
- [26] Koopmans-van Beinum, F.J. and van der Stelt, J.M., *Early speech development in child's acquiring Dutch mastering general basic elements*, in *The Acquisition of Dutch*, Gillis, S.e.d.H., A., Editor. 1998, Benjamins: Amsterdam/Philadelphia. p. 101-425.
- [27] Kravitz, E.A. and Huber, R., *Aggression in invertebrates*. Current Opinion In Neurobiology, 2003. **13**(6): p. 736-743.
- [28] Kroodsma, D.E., *Learning and the ontogeny of sound signals in birds*. Acoustic Communication in Birds (Ed. by D. E. Kroodsma & E. H. Miller), 1982: p. 11-23.
- [29] Kuhl, P.K. and Meltzoff, A.N., *Infant vocalizations in response to speech: vocal imitation and developmental change*. Journal of the Acoustical Society of America, 1996. **100**: p. 2425-2438.
- [30] Lipkind, D., et al., *New replicable anxiety-related measures of wall vs. center behavior of mice in the open field*. Jour. Appl. Physiol., 2004. **97**(1): p. 347-359.
- [31] Liu, G., et al., *Distinct memory traces for two visual features in the Drosophila brain*. Nature, 2006. **439**(7076): p. 551-556.
- [32] Margulies, C., Tully, T. and Dubnau, J., *Deconstructing memory in Drosophila*. Current Biology, 2005. **15**(17): p. R700-R713.
- [33] Marler, P. and Tamura, M., *Culturally transmitted patterns of vocal behavior in sparrows*. Science, 1964. **146**: p. 1483-1486.
- [34] Martin, J.R., *Locomotor activity: a complex behavioural trait to unravel*. Behavioural Processes, 2003. **64**(2): p. 145-160.
- [35] Martin, J.R., *A portrait of locomotor behaviour in Drosophila determined by a video-tracking paradigm*. Behavioural Processes, 2004. **67**(2): p. 207-219.

- [36] Matyear, C.L., MacNeilage, P.F. and Davis, B.L., *Nasalization of vowels in nasal environments in babbling: evidence for frame dominance*. *Phonetica*, 1998. **55**: p. 1-17.
- [37] McGuire, S.E., Deshazer, M. and Davis, R.L., *Thirty years of olfactory learning and memory research in Drosophila melanogaster*. *Progress In Neurobiology*, 2005. **76**(5): p. 328-347.
- [38] Mehren, J.E., Ejima, A. and Griffith, L.C., *Unconventional sex: fresh approaches to courtship learning*. *Current Opinion In Neurobiology*, 2004. **14**(6): p. 745-750.
- [39] Nottebohm, F. and Nottebohm, M.E., *Relationship between Song Repertoire and Age in Canary, Serinus-Canarius*. *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology*, 1978. **46**(3): p. 298-305.
- [40] Scharff, C. and Nottebohm, F., *A Comparative-Study of the Behavioral Deficits Following Lesions of Various Parts of the Zebra Finch Song System - Implications for Vocal Learning*. *Journal of Neuroscience*, 1991. **11**(9): p. 2896-2913.
- [41] Serkhane, J.E., *Un bébé androïde vocalisant: Etude et modélisation des mécanismes d'exploration vocale et d'imitation orofaciale dans le développement de la parole*. 2005, Institut Polytechnique de Grenoble: Grenoble (France). p. 263.
- [42] Shaw, P.J., Cirelli, C., Greenspan, R.J. and Tononi, G., *Correlates of sleep and waking in Drosophila melanogaster*. *Science*, 2000. **287**(5459): p. 1834-1837.
- [43] Sokolowski, M.B., *Foraging Strategies Of Drosophila-Melanogaster - A Chromosomal Analysis*. *Behavior Genetics*, 1980. **10**(3): p. 291-302.
- [44] Tchernichovski, O., et al., *Studying the song development process rationale and methods*. *Behavioral Neurobiology of Birdsong*, 2004. **1016**: p. 348-363.
- [45] Tchernichovski, O., Mitra, P.P., Lints, T. and Nottebohm, F., *Dynamics of the vocal imitation process: How a zebra finch learns its song*. *Science*, 2001. **291**(5513): p. 2564-2569.
- [46] Tchernichovski, O. and Nottebohm, F., *Social inhibition of song imitation among sibling male zebra finches*. *Proceedings of the National Academy of Sciences of the United States of America*, 1998. **95**(15): p. 8951-8956.
- [47] Tchernichovski, O., et al., *A procedure for an automated measurement of song similarity*. *Animal Behaviour*, 2000. **59**: p. 1167-1176.
- [48] Thomson, D.J., *Quadratic-Inverse Spectrum Estimates - Applications to Paleoclimatology*. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*, 1990. **332**(1627): p. 539-597.
- [49] Thomson, D.J., *Non-stationary fluctuations in stationary time series*. *Proceedings of the International Society of Optical Engineering*, 1993. **2027**: p. 236-244.
- [50] Thorpe, W.H., *The process of song-learning in the chaffinch as studied by means of the sound spectrograph*. *Nature*, 1954. **173**: p. 465.
- [51] Thorpe, W.H., *The learning of song patterns by birds, with special reference to the song of the chaffinch, Fringella coelebs*. *Nature*, 1958. **100**: p. 535-570.
- [52] van der Stelt, J.M. and Koopmans-van Beinum, F.J., *The onset of babbling related to gross motor development*, in *Precursors of Early Speech*, B. Lindblom, e.Z., R., Editor. 1986, Stockton Press: New York. p. 163-173.
- [53] van der Stelt, J.M., Wempe, T.G. and Pols, L.C.W. *Progression in vowel production: comparing deaf and hearing children*. in *Proceedings*. 2003. Amsterdam (Netherlands): University of Amsterdam.
- [54] van Swinderen, B., *The remote roots of consciousness in fruit-fly selective attention?* *Bioessays*, 2005. **27**(3): p. 321-330.
- [55] van Swinderen, B. and Andretic, R., *Arousal in Drosophila*. *Behavioural Processes*, 2003. **64**(2): p. 133-144.