Stony Brook University

Rat's Orofacial Activity Recognition and Its Applications

A Thesis presented

by

Le Hou

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Computer Science

Stony Brook University

May 2014

Stony Brook University

The Graduate School

Le Hou

We, the thesis committee for the above candidate for the

Master of Science degree, hereby recommend

acceptance of this thesis

Dimitris Samaras
Associate Professor, Department of Computer Science

Alfredo Fontanini
Assistant Professor, Department of Neurobiology and Behavior

Xianfeng David Gu
Associate Professor, Department of Computer Science

This thesis is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Thesis

Rat's Orofacial Activity Recognition and Its Applications

by

Le Hou

Master of Science

in

Computer Science

Stony Brook University

2014

A major problem in neuroscience is determining the perceived value of rewarding and aversive stimuli in animal subjects. Orofacial activity, such as licking and gaping in response to tastes, has been shown to be well correlated with the perceived palatability of tastes sampled by an animal. The current standard for determining these orofacial reactions is frame-to-frame labeling by trained scientists, a very time consuming process. Here we introduce a supervised classifier that can automatically recognize nine distinct yet subtle orofacial activities with an accuracy of 82.00% (chances are 21.16%). The classifier implements data from both videos of rats receiving taste deliveries and concurrent electromyographic recordings of the digastric muscle which is involved in food consumption. We additionally applied our classifier and features to a classical conditioning experiment to determine whether cues predicting different tastes can initiate different orofacial movements prior to an actual taste delivery. By using features extracted following the cue (tone) but before the corresponding taste delivery, we can predict the identity of the cue with an accuracy of 41.39% (chances are 20.11%), showing that the

animals have learned the cue-taste associations. In addition, we can retro-classify the identity of the cue with an accuracy of 65.64% using features extracted after the taste delivery. Based on these results, we claim that our model allows for fast and objective determination of orofacial reactions in rats and for assessing the strength of taste-reinforcer associative learning.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ApEn	Approximate Entropy.
AR	Auto-Regressive coefficients.
CAcid	Citric Acid.
CC	Cepstral Coefficients.
DFA	Detrended Fluctuation Analysis.
DIC	Digital Image Correlation and tracking (features).
DSamp	DownSample.
DSP	DownSampled pixels.
EMG	ElectroMyoGraphy.
GM	Gardner's Measurement.
GUI	Graphical User Interface.
HOA	Histogram of Orofacial Activities.
IAV	Integral Absolute Value.
LBP	Local Binary Patterns.
LDA	Linear Discriminate Analysis.
LTP	Lateral Tongue Protrusion.
MAV1	Modified Mean Absolute Value type 1.
MAV2	Modified Mean Absolute Value type 2.
MFL	Maximum Fractal Length.
MM	Mouth Movement.

MO	Mouth Opening.
NM	No Movement.
OAR	Orofacial Activity Recognition.
Opt-Flow	Optical-Flow.
PCA	Principal Component Analysis.
POA	Predicted Orofacial Activities.
RBF	Radial Basis Function.
RMS	Root Mean Square.
SampEn	Sample Entropy.
SS	Symmetric Score.
SVM	Support Vector Machine.
TP	Tongue Protrusion.
WL	Waveform Length.

Acknowledgments

One year ago before this thesis started, Prof. Samaras, as my adviser, listened to my interests carefully and spent hours looking for topics for me. Without your patience I had not met the wonderful team members of mine. Your support and concern are important for me and much appreciated.

I would like to thank Matthew P.H. Gardner and Alfredo Fontanini also. First, you explained all the background knowledge and experiment settings in great details to me, which helped me familiarized with our research quickly. Second, with the Matthew P.H. Gardner's and Chad Samuelsen's great effort, thousands of video frames were annotated using my broken GUI and feature extraction method was significantly improved. Furthermore, thank you again, Matthew P.H. Gardner, for proofreading this thesis of more than seven thousand words. When I am looking for someone to correct my English at the end of the semester, it is you helped me out when preparing your own Ph.D. thesis defense.

Chapter 1: Introduction

Rat's Orofacial Activity

In the research of neuroscience, experiments including rewarding and unrewarding stimuli play a central role and have been widely studied. In these experiments, liking and disgust of an infant or animal can be reflected in orofacial activities that have a common lexicon (Grill and Norgren, 1978; Berridge, 2000). Examples of orofacial activities are shown in Figure 1. Certain kinds of tongue movements, such as licking, can be interpreted as signs of enjoyment, while gapes can expresses disgust. Therefore, these orofacial activities can be considered as important feedback about the perceived value of stimuli from rats in neurobiology experiments. One such applications, is analyzing the cortical processing of taste (Samuelsen et al., 2012). However, techniques for recognizing orofacial activities are not well developed. In previous neuroscience experiments, either extensive human efforts were applied annotating orofacial activities or immature automatic recognition techniques were applied. One of the previous methods is focused on measuring the change of pixel intensities through video frames (Samuelsen et al., 2012) and cannot distinguish different orofacial activities. Drawbacks of these works are obvious. They suffer from excessive load of work, subjectiveness, lack of quantitative evaluation, and coarse analytic result. In this paper, we introduce a classifier that can automatically classify nine subtle activities and evaluate its performance by cross-validation.

Related Work

A potential way to recognize orofacial activities is to apply recognition especially expression recognition techniques to video frames of facial movements. Human facial expression recognition is a well-studied topic and its technics can be applied to our problem. Here we first discuss feature extraction methods and then modeling algorithms. There are 3 main types of fully automatic feature extraction methods. First, direct dimensionality reduction on facial images was shown to be effective. For example, Eigenfaces and Fisherfaces are methods that can represent faces with much lower dimensionality and perform well in practice (Belhumeur et al., 1997). Second,

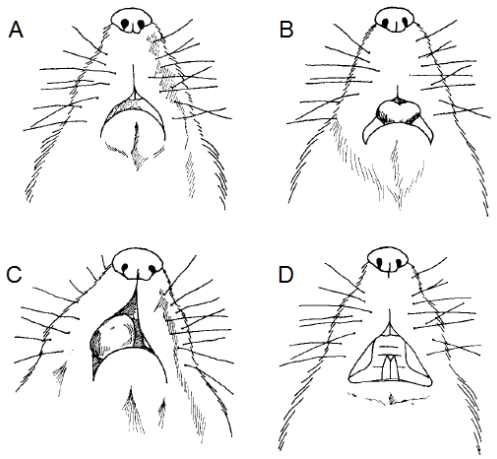

Figure 1: Example of orofacial activities illustrated by Harvey J. Grill (Grill and Norgren, 1978). A: mouth movement. B: tongue protrusion. C: lateral tongue protrusion. D: gape.

statistical methods, including applying linear or nonlinear filters to facial images, capture information of texture, edges, and corners which were proven to be helpful methods. An example of these linear filters is the Gabor filter bank (Deng et al., 2005). Local Binary Patterns (LBP) extracted by nonlinear filters, achieve better results compared to Gabor and are much faster to compute (Shan et al., 2009). Finally, facial movement measurements, such as optical-flow based methods, capture the changes of a face (Hsieh et al., 2010; Zhang and Tjondronegoro, 2011). The advantage of these methods is that they utilize a video segment instead of a single frame. In order to test all 3 kinds of features, we implemented seven feature extraction methods: Eigenfaces, Fisherfaces, and downsampled pixels (dimensionality reduction based); LBP (nonlinear filter based); symmetric score, Digital Image Correlation and tracking(DIC), and optical-flow (measures the facial movement). We tested combinations of these features and found that LBP performs the best. Furthermore, adding other features yielded no significant improvement. Detailed results are listed in later Chapters.

The modeling algorithms' success of facial expression recognition is related to its input, in other words, features. A semi-supervised Bayesian network classifier models the connections between parts of a face and achieves better performance, when applied on continuous video input (Cohen et al.,

2003). Multilevel hidden Markov model was also tested for facial expression recognition (Cohen et al., 2000). These models are based on an assumption that there are multiple hidden states corresponding to a video. In our dataset however, an orofacial activity usually lasts less than 3 frames and we are only interested in the mouths of rats. Thus, the temporal information and correlations between parts are not rich. On the other hand, simple classifiers like Support Vector Machine (SVM) were applied on a single image and achieved over 0.9 accuracy (Boser et al., 1992; Shan et al., 2009). Based on these results, we focused on SVM with Radial Basis Function (RBF) kernels which give better results compared to other kernels (Shan et al., 2009).

In addition to features extracted from video frames, electromyography (EMG) measures the contraction of muscles and can also be utilized for assessing mouth movements (Reaz et al., 2006). In our dataset, EMG was collected from a single muscle and represented as a one dimensional temporal signal. Various features in the time domain and frequency domain were tested in gesture recognition (Ahsan et al., 2009; Phinyomark et al., 2012a, 2013). Within these approaches, features are extracted using a sliding window with fixed size and increment. In this paper, we have implemented and tested eleven features proven to be the most robust in EMG signals (Phinyomark et al., 2013).

Classical Conditioning

In order to test our features and models in practice, we have applied our features and models in a classical conditioning experiment, specifically forward conditioning (Pavlov and Anrep, 1927). A conditioned stimulus, a tone, is given before the unconditioned stimulus, a taste delivery. The conditioned response and unconditioned response are the behavioral responses to the conditioned stimulus and unconditioned stimulus respectively. We have focused on analyzing the conditioned response and unconditioned response in terms of orofacial activities. In previous research related to this experiment, the orofacial activities were measured by human labeling or non-discriminative statistical approaches (Samuelsen et al., 2012). Utilizing the features and models presented in this paper, we found significant differences of orofacial activities in response to different cues and tastes. Our approach has the advantage of relating specific orofacial activities to different cues, a problem which has had no systematic and automatic solution before.

Chapter 2: Dataset

Experiment Setup

We first introduce the setup of our neuroscience experiment and then explain how the dataset is created from it.

The experiment is based on the classical conditioning methods described above (Pavlov and Anrep, 1927). In the experiment, there are five different cues (tones) and four tastes, with four cues each predicting a specific taste delivery, and the fifth cue being predictive of no delivery. In each trial, a cue is given to the rat first. After 3 seconds, the corresponding taste is delivered directly into the mouth of the rat. Figure 2 illustrates the timeline of a trial. In these experiments, the head of the rat is held fixed. There are around 70 trials per session and 33 sessions are used, with each session lasting more than an hour. The mouth of the rats are recorded by a camera, and the EMG of the digastric muscle are also recorded throughout the entire session. The types of taste are: citric acid, NaCl, quinine, sucrose, and null (no taste delivery). The numbers of trials with each taste are the same in each of the 33 sessions.

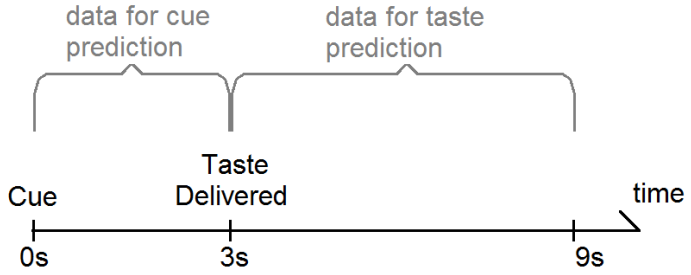

Figure 2: Timeline of a trial in our forward conditioning experiment. There are five cues and four tastes. Each taste corresponds to a cue.

Data Collection

As a supervised learning algorithm, thousands of frames with orofacial activities should be provided before the model can be trained. We recorded

the video and EMG of 3 rats and defined the orofacial activities by observing the videos. Figure 3 shows the GUI based tools used for labeling frames. Each video is first, manually cropped to make sure all frames viewed and used throughout the analysis share roughly the same scale and position of each rat visualized. Next, frames of 9 classes of orofacial activities are labeled by a domain expert using another tool. Finally an instance of orofacial activity is represented by 3 video frames (the previous and next frame of the labeled frame are also extracted) as well as 0.4 seconds of the EMG.

For the application of cue prediction, the labels have already been given as the cue delivered on a particular trial. We trained each rat on the cue-taste association paradigm for 10 days and then subsequently recorded the 33 sessions used in the analysis. An instance of cue reaction is represented by the 75 video frames and 2.8 seconds of EMG following the cue.

Table 1 lists the numbers of instances for each labeled orofacial activity and cues. Figure 4 illustrates examples of labeled frames that are randomly selected. Figure 5 shows examples of the EMG.

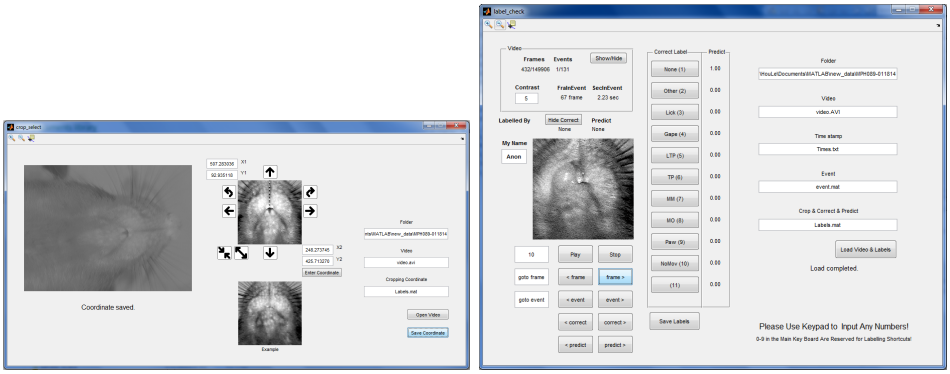

Figure 3: GUI tools written in Matlab for viewing and labeling orofacial activities. Left: the cropping tool. Right: the video player and orofacial activity labeling & viewing tool. Predicted orofacial activities are also showed in this tool.

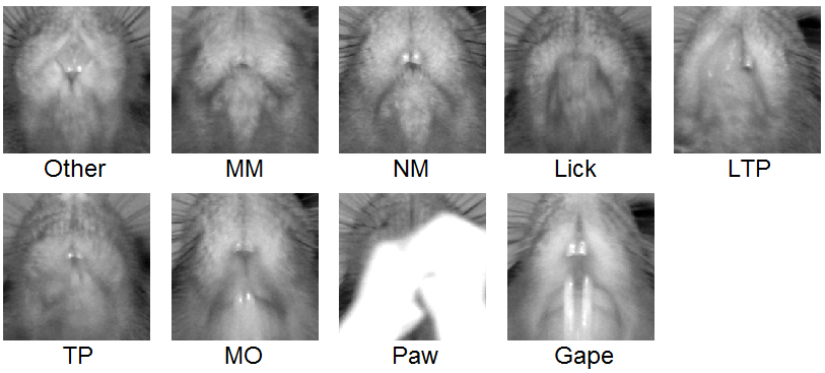

Figure 4: Examples of 9 orofacial activities in our dataset. The resolution of the images is 200 by 200 in gray scale.

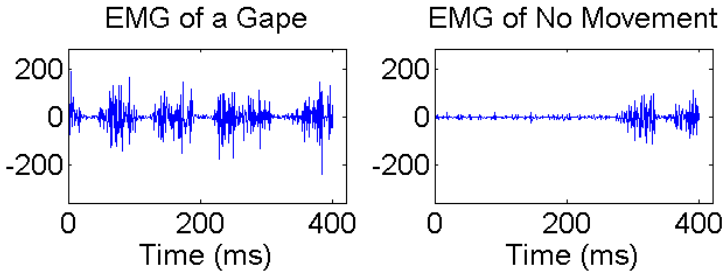

Figure 5: Examples of EMGs in Our Dataset. EMG recordings during two orofacial activities.

Classes	Other	Lick	Gape	LTP	TP	MM	MO	Paw	NM
Numbers	103	1114	265	966	8	1200	342	548	1124

Cues	CAcid	NaCl	Sucrose	Quinine	Null
Numbers	451	453	451	455	452

Table 1: Distribution of classes in our dataset. Top: classes of orofacial activities. The labeled frames are distributed in 12 sessions (each session contains a video and EMG series) of 3 unique rats. The best guessing accuracy is 21.16%. Bottom: classes (cues/taste) in cue prediction. There are 33 sessions each containing 54 to 70 trials. The best guessing accuracy is 20.11%.

Chapter 3: Feature Extraction and Models

Video Features

Although orofacial activities are defined using a single frame, we extract features from 3 adjacent frames centered at the labeled frame, because most orofacial activities last less than 3 frames. The same feature extraction methods are applied to all adjacent frames. It is possible that two adjacent frames might be labeled differently, which means its features are contained in more than one instance. Also notice that although 3 frames are used for feature extraction, the orofacial activity label only describes the center frame.

There are seven clusters of features that can be extracted from each frame:

1. Eigenfaces

   This was introduced by (Turk and Pentland, 1991) as a method based on PCA. An Eigenface is an eigenvector of the covariance matrix of all raw pixels acquired by applying singular value decomposition on all facial images. We use the top 15 Eigenfaces because the 15th eigenvalue is considered as a knee point on all eigenvalues. Figure 6 shows some Eigenfaces of our dataset.

2. Fisherfaces

   This is a supervised dimensionality reduction method based on Fisher's LDA (Welling, unknown). A fisherface is one of the vectors that maximizes external variance between classes and minimizes internal variance within classes. In order to apply Fisherfaces, we apply PCA first to make sure the dimensionality of input instances is smaller than the number of instances. For the dimensionality of the final output, we simply follow the number of Eigenfaces stated above. Therefore the dimensionality of an image is reduced to 15. Examples of Fisherfaces of our dataset are shown in Figure 6. The implementation of LDA is obtained from open source (Alzahrani, 2014).

3. Downsampled pixels (DSP)

   Being the most direct way to reduce dimensionality, downsampling the image proved to be robust for our dataset. A two dimensional Gaussian filter is applied to the image before downsampling. Figure 7 il-

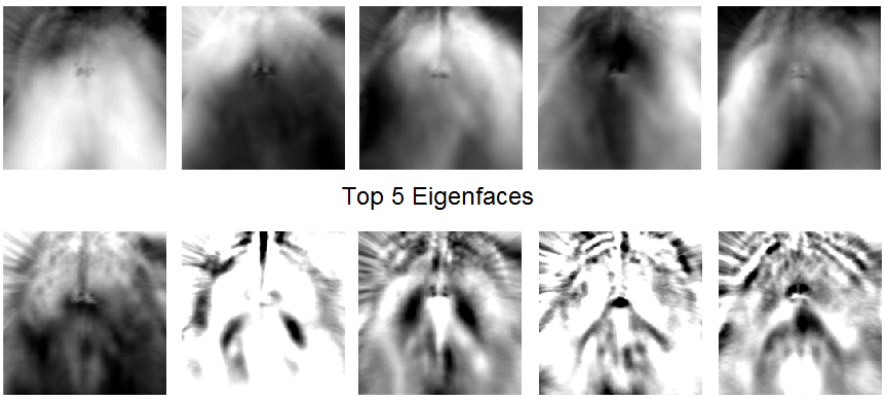

Figure 6: Eigenfaces and Fisherfaces of our dataset shown as gray scale images.

lustrates the process. The downsampling sampling rate is chosen by cross-validation using this feature only. The 6 by 6 downsampled images yield the best classification accuracy.

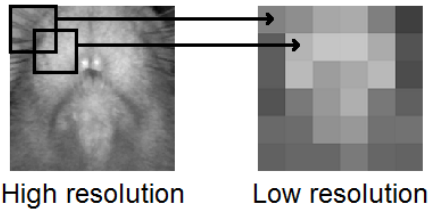

Figure 7: A two dimensional Gaussian filter is applied to the image before downsampling. Then the image is downsampled into a 6 by 6 feature vector.

4. Symmetric Score (SS)

   Because images of the rats are symmetric when there is no movement, asymmetry can be used to assess movement. Symmetric score measures the intensity of mouth movements and provides only a one dimensional output. To compute the symmetric score, an image is divided into two parts (left and right). Then we simply compute the cosine distance between the two halves.

5. DIC features (DIC)

Digital image correlation and tracking is a method that measures the deformation of an image based upon a reference image. We use the Newton-Raphson method of partial differential correction (Bruck et al., 1989). This classic method supports subpixel accuracy but is computationally expensive (Sutton et al., 2009; Pan and Li, 2011). To track a point in the reference image, this method assumes that the brightness of physical points are constant and the deformation of a small image patch is linear:

$$\begin{aligned} x' &= x + u + u_x \Delta x + u_y \Delta y \\ y' &= y + v + v_x \Delta x + u_y \Delta y, \end{aligned} \tag{1}$$

as illustrated in Figure 8. The parameters $\theta = (u, v, u_x, v_x, u_y, v_y)$ of the deformation are optimized such that the following energy function, which is the difference between the deformed patch and the reference patch, is minimized:

$$\mathrm{E}(\theta) = \sum_{x,y} \left(\frac{\mathrm{I}(x,y) - \mathrm{I_m}}{\mathrm{I}_\sigma} - \frac{\mathrm{I}'(x',y') - \mathrm{I}'_\mathrm{m}}{\mathrm{I}'_\sigma} \right), \tag{2}$$

where I and I′ are the reference and deformed patches; $\mathrm{I_m}$ and I_σ are the mean and standard deviation of all pixels in the patch. With the assumption that the energy function is small compared to the gradient of parameters, the method approximates the second order partial derivative (Hessian matrix) using the gradient only. A bicubic interpolation scheme is applied in order to obtain subpixel values. Also with this interpolation, the energy function become differentiable:

$$\mathrm{I}'(\mathrm{x}', \mathrm{y}') = \sum_{m=0}^{3} \sum_{n=0}^{3} \alpha_{m,n} x^m y^n. \tag{3}$$

A local area of $p \geq 16$ points are used to solve the 16 coefficients $\alpha_{m,n}$. There are several hyperparameters required to be selected for this method. We use cross-validation using this feature only and output the values of hyperparameters that yield the best accuracy. However, in order to compromise the long running time, only four key points are tracked. Figure 9 illustrates the tracking result.

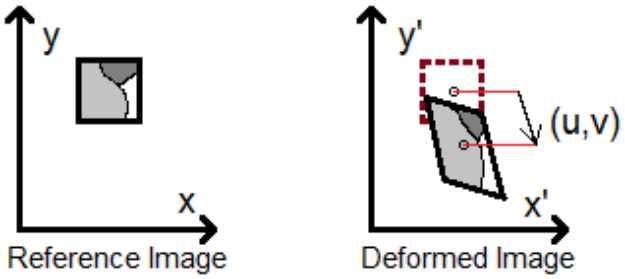

Figure 8: To track the deformation, the Newton-Raphson method of partial differential correction assumes that a local image patch is deformed by a linear transformation.

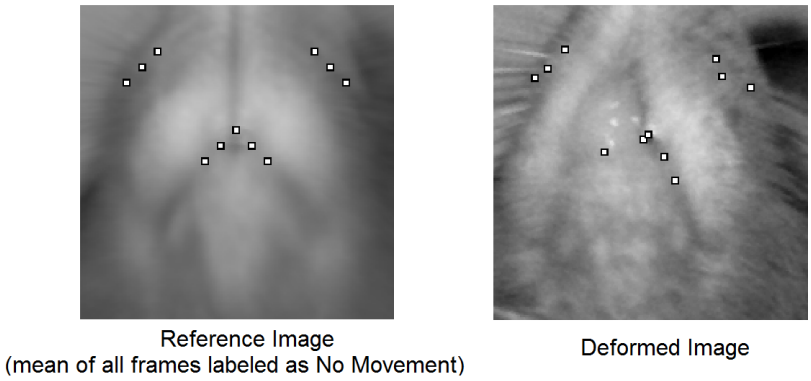

Figure 9: Tracking result by DIC. Every point in the reference image is tracked in the deformed image.

6. Optical-Flow (Opt-Flow)

   Optical-flow is another image registration method. It assumes each pixel in a local patch has the same motion, the brightness of every physical point is constant, and the points are moved within a range in which the brightness of pixels are linear (Lucas and Kanade, 1981). The motion vector is obtained by solving a linear equation. We also include a weight regularizing term in the energy function (least square) to make sure the linear equation is solvable. In order to make the motion constrained in a local area with respect to the scale of objects in the image, the image is convoluted with a Gaussian kernel before the optical-flow is applied. Hyperparameters are selected by cross-

validation. Figure 10 shows tracking results by optical-flow.

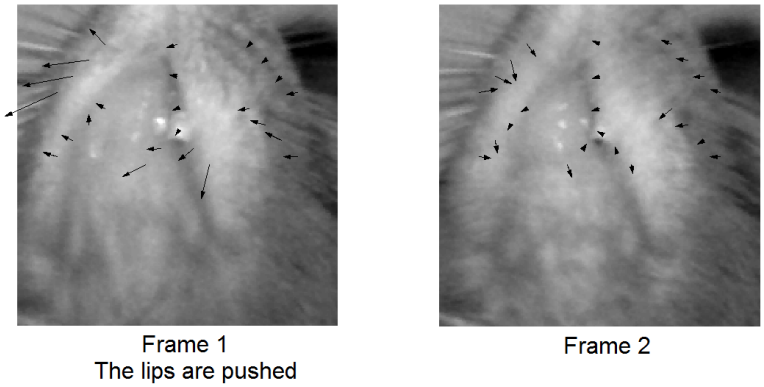

Figure 10: Tracking result by optical-flow. An arrow indicates the motion of the point with respect to previous image.

7. Local Binary Patterns (LBP)

   LBP was first invented as a texture descriptor (Ojala et al., 2002) and later used in facial expression recognition (Shan et al., 2009). It is basically a histogram of binary patterns that describe the neighborhood around a pixel. Figure 11 gives an example of a possible pattern. We follow the hyperparameters of the LBP operator used in (Shan et al., 2009) which generates 59 vectors for an input image window. Furthermore, we can divide the image into several non-overlapping subimages and apply the LBP operator individually to each of them. We tested different dividing methods and found that LBP on 2 subimages yields a good result (0.02 lower than LBP from 9 images in terms of accuracy) with much fewer features. The official implementation of LBP is adopted from Heikkila and Ahonen (Heikkila and Ahonen, 2013).

EMG Features

We have adopted the EMG-based method used in upper limb motion classification (Phinyomark et al., 2012a, 2013). A sliding window of size s seconds moves along the EMG with an increment of $s/2$. The features

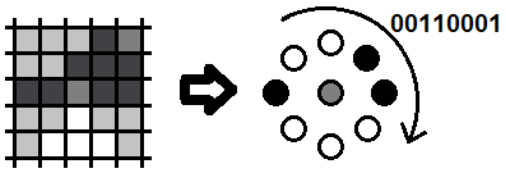

Figure 11: A local binary pattern that defines the type of the pixel or neighborhood. In this case, it shows an intersection of a line with a corner.

are extracted in each window. The hyperparameter s is selected by cross-validation from the set $\{0.05, 0.1, 0.2, 0.4\}$. The increment is fixed based on the size of the window due to the observation that the accuracy is not sensitive to the increment, according to previous results (Phinyomark et al., 2013). The 11 features with best performance (Phinyomark et al., 2013) plus 2 other features are tested. A brief introduction to these features are listed below:

1. Approximate Entropy (ApEn) (Zhao et al., 2006a) measures the unpredictability of a signal. A small ApEn means the signal repeats the same patterns over time. Denote the length n EMG series as $\mathbf{x}_i$; $i = 1, 2, 3, \ldots, n$. The ApEn with hyperparameters m, r is defined below:

$$\begin{aligned} \text{ApEn}_{m,r}(\mathbf{x}) &= -\sum_{i=1}^{n-m} \left(\frac{\exp(A_i) - \exp(B_i)}{n-m} \right) \\ B_i &= \sum_{j=1}^{n-m} \mathbb{1}\left(\left(\max_{k=0,1,\ldots,m-1} |\mathbf{x}_{i+k} - \mathbf{x}_{j+k}| \right) > r \right) \\ A_i &= \sum_{j=1}^{n-m} \mathbb{1}\left(\left(\max_{k=0,1,\ldots,m} |\mathbf{x}_{i+k} - \mathbf{x}_{j+k}| \right) > r \right), \end{aligned} \tag{4}$$

   where $\mathbb{1}()$ is the indicator function. Following previous work (Phinyomark et al., 2013; Zhao et al., 2006a), we select $m = 2$ and $r = 0.2\sigma$, where σ is the standard deviation of all samples of $\mathbf{x}$.

2. Auto-Regressive coefficients (AR) (Tkach et al., 2010; Zardoshti-Kermani et al., 1995) is obtained from fitting each value of the EMG by a linear combination of previous values. The coefficients vector $\mathbf{a}$ of the p-th

order regression model is used as a feature vector and can be computed recursively (Zardoshti-Kermani et al., 1995):

$$\begin{aligned} \mathbf{a}_i &= \mathbf{a}_{i-1} + \mathbf{P}_i \mathbf{X}_i \left(\mathbf{x}_i - \mathbf{X}_i^{\mathrm{T}} \mathbf{a}_{i-1} \right) \\ \mathbf{P}_i &= \mathbf{P}_{i-1} - \frac{\mathbf{P}_{i-1} \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} \mathbf{P}_{i-1}}{1 + \mathbf{X}_i^{\mathrm{T}} \mathbf{P}_{i-1} \mathbf{X}_i}, \end{aligned} \tag{5}$$

where $\mathbf{X}_i = (\mathbf{x}_{i-1}, \mathbf{x}_{i-2}, \ldots, \mathbf{x}_{i-p})^{\mathrm{T}}$. $\mathbf{P}_i$ is the i-th iteration of $\mathbf{P}$ which is initially set as a p by p identity matrix. $\mathbf{a}_i$ is the i-th iteration of the coefficients vector and initialized as a zero vector of length p. The order $p = 4$ was suggested by previous work (Phinyomark et al., 2012a).

3. Cepstral Coefficients (CC) (Phinyomark et al., 2012a; Zecca et al., 2002) "is defined as the inverse Fourier transform of the logarithm of power spectrum magnitude of the signal data". It is similar with the AR coefficients and can be derived from it (Phinyomark et al., 2012a):

$$\begin{aligned} \mathbf{c}_1 &= -\mathbf{a}_1 \\ \mathbf{c}_p &= -\mathbf{a}_p - \sum_{l=1}^{p-1} \left(1 - \frac{l}{p} \right) \mathbf{a}_p \mathbf{c}_{p-1}. \end{aligned} \tag{6}$$

   As the order of CC, p is also set to 4 in our experiments.

4. Detrended Fluctuation Analysis (DFA) (Phinyomark et al., 2012b; Peng et al., 1995) "is a modified root mean square analysis of a random walk". It contains both information in the time-magnitude domain and time-frequency domain. To compute the DFA of a EMG segment $\mathbf{x}$, the signal is integrated first:

$$\mathbf{y}_i = \sum_{t=1}^{i} \left(\mathbf{x}_t - \mathrm{mean}(\mathbf{x}) \right). \tag{7}$$

   Then $\mathbf{y}$ is divided into sub-segments of size s. In each sub-segment $\mathbf{y}_s$, a quadratic function (vector) $\mathbf{y'}_s$ is fitted by minimizing the squared error. The fluctuation with respect to s is defined by the squared difference between $\mathbf{y}_s$ and $\mathbf{y'}_s$ in all segments:

$$\mathrm{f}(s) = \sqrt{\frac{1}{n} \sum_{s} \| \mathbf{y}_s - \mathbf{y'}_s \|_2^2}, \tag{8}$$

where the operator $\|\cdot\|_2$ gives the euclidean norm of a vector. Finally, the DFA is the slop of function $\mathrm{f}(s)$ on log-log space, which can be derived by applying linear regression to samples of $\{\log_{10} s, \log_{10} \mathrm{f}(s)\}$.

5. Integral Absolute Value (IAV) (Kim et al., 2011) is simply the mean absolute value of the EMG. It can be interpreted as a convolution response between the signal and a boxcar (rectangle) kernel.

6. Modified Mean Absolute Value type 1 (MAV1) (Phinyomark et al., 2012a) is a weighted mean absolute value of the EMG with less weight away from the center of the window.

7. Modified mean Absolute Value type 2 (MAV2) (Phinyomark et al., 2012a) is a weighted mean absolute value of the EMG with smoother kernel compared to MAV1.

8. Maximum Fractal Length (MFL) (Phinyomark et al., 2012b) is basically the sum of squared differences between all adjacent EMG samples. It somehow reflects the maximum fluctuation in the EMG.

9. Root Mean Square (RMS) (Phinyomark et al., 2012a) is the estimated standard error of the EMG assuming the expectation is zero.

10. Sample Entropy (SampEn) (Zhang and Zhou, 2012; Zhao et al., 2006b; Richman and Moorman, 2000) is similar to ApEn and it is an enhanced version. Practically, it can be computed in the following way.

$$\mathrm{SampEn}_{m,r} = -\exp\left(\frac{\sum_{i=1}^{n-m} A_i}{\sum_{i=1}^{n-m} B_i}\right), \tag{9}$$

where A_i and B_i are defined by Equation 4. The adopted value of m and r are the same as those in ApEn.

11. Waveform Length (WL) (Tkach et al., 2010; Hudgins et al., 1993) is the sum of absolute differences between adjacent all adjacent EMG samples.

12. Gardner's Measurement (GM) is proposed by Matthew P.H. Gardner, a domain expert who provided the whole dataset discussed in this paper. It counts the number, duration, and intensity of bouts (an EMG segment that possibly indicates muscle contraction) using a set of predefined thresholds.

13. DownSample (DSamp) is simply the downsampled EMG. Extracted from a short EMG segment, the dimensionality of this feature is acceptable. The downsample rate in our experiment is 0.1.

Support Vector Machine

We use the official implementation of Support Vector Machine (SVM) (Chang and Lin, 2011). A SVM model is usually sparse due to its hinge loss function. The predicted class given the input $\mathbf{x}$ in a two-class problem is defined below:

$$y = \text{sign}\left(\sum_{i=1}^{s} \alpha_i y_i \text{K}(\mathbf{x}, \mathbf{x}_i) + b\right), \tag{10}$$

where $\text{K}(\mathbf{x}, \mathbf{x}_i)$ is the kernel function. $\mathbf{x}_i$ and y_i are support vectors that need to be learned together with α_i. Only Radial Basis Function (RBF) kernel is tested because it has fewer hyperparameters and achieves better performance compared to other kernels (Shan et al., 2009).

Multi-class SVM is also implemented in LIBSVM by the "one-against-one" approach. Denote k as the number of classes. $k(k-1)/2$ classifiers are trained based on all combinations of two classes. Then a voting scheme is applied and the label is assigned as the class which received the most votes.

Linear Discriminant Analysis

Suggested by Phinyomark, LDA performs better than SVM on EMG features (Phinyomark et al., 2013). Thus we also test two-class LDA for classification using the "one-against-all" approach (Bishop et al., 2006). This linear classifier is considered robust because the linear projection of instances minimize the within-class covariance while maximize the between-class covariance. Provided instances in two classes, the predicted class is defined

below:

$$
\begin{aligned}
y &= \mathbf{w}^{\mathrm{T}}\mathbf{x} \\
\mathbf{w} &\propto \mathbf{S}^{-1}(\mathbf{m}_2 - \mathbf{m}_1) \\
\mathbf{S} &= \sum_{n \in C_1} (\mathbf{x}_n - \mathbf{m}_1)(\mathbf{x}_n - \mathbf{m}_1)^{\mathrm{T}} + \sum_{n \in C_2} (\mathbf{x}_n - \mathbf{m}_2)(\mathbf{x}_n - \mathbf{m}_2)^{\mathrm{T}},
\end{aligned}
\tag{11}
$$

where $\mathbf{m}_1$ and $\mathbf{m}_2$ are mean of instances in class C_1 and C_2. Notice that the bias of the linear form is learned by putting a constant feature 1 to $\mathbf{x}$. Implementation of LDA is found online (Dwinnell, 2010).

Multi-class Logistic Regression

Although LDA has no hyperparameter, it would overfit the data when the dimensionality of instances is considerably large. Multi-class Logistic Regression provides a fast and relatively sparse solution. The probability of class y is given by the following formula:

$$
\mathrm{P}(y|\mathbf{x}) = \frac{\exp(\mathbf{w}_y^{\mathrm{T}}\mathbf{x})}{\sum_{i=1}^{k} \exp(\mathbf{w}_i^{\mathrm{T}}\mathbf{x})}, \tag{12}
$$

where $\mathbf{w}_i$ is the vector of parameters for class i. In the training stage a second-norm regularization term is involved. All parameters and can be trained using Newton's method efficiently (Bishop et al., 2006). The regularizer is selected by cross validation.

Chapter 4: Orofacial Activity Recognition

Feature Selection

There are 13 EMG features and 7 video features implemented. Each instance contains 3 frames and 0.4 seconds of EMG. We apply the following feature selection method: First, EMG and video features are selected separately. Combinations of video features or EMG features are tested using a linear classifier. Then, the top 5 best performing combinations of EMG and video features are combined and tested again using SVM with RBF kernel, as illustrated in Figure 12. Linear classifiers are used to select features because they have less hyperparameters to be tuned, which means the time spent on cross-validation is much less compared to SVM with RBF kernel. For the EMG features, combinations of 4 out of 13 features are tested with LDA because there are no hyperparameters in this model and LDA is proven to be a robust classifier for EMG datasets Phinyomark et al. (2013). For video features, combinations of 3 out of 7 features are tested with multi-class logistic regression because the dimensionality of features in video features are much higher and sparsity control is needed. In order to make an equivalent comparison of the performance between the video and EMG features, the best accuracies of the EMG using SVM RBF kernel are also reported.

Results

We apply 12-fold cross-validation on instances labeled in 12 sessions (each session contains a video and EMG series). Table 2 shows the feature selection results. Tables 3,4,5 show the prediction-groundtruth matrix of our model. Notice that the EMG features do not contribute to the final accuracy significantly. Table 6 shows that the instances misclassified by video features are also very likely to be misclassified by the EMG features.

Our OAR dataset consists of three rats: Rat#89 (3 sessions), Rat#90 (5 sessions), and Rat#91 (4 sessions). To apply cross-validation, instead of splitting data according to different sessions, there are other data splitting methods. Cross-validation with each data splitting method reflects different levels of generalization ability. Three data splitting methods are discussed below.

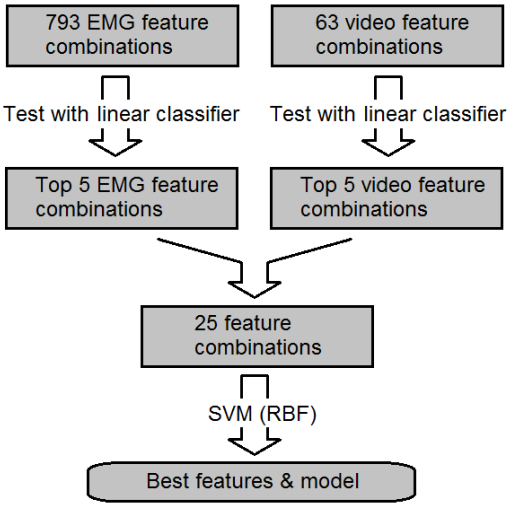

Figure 12: The procedure of feature selection. Linear classifiers are used in the early stage because they need less rounds of cross-validation due to less hyperparameters.

1. Splitting rats. The result of this shows the generalization ability among the animals. For cross-validation, the dataset is splitted by rats. Therefore, the number of sessions in the training set is 7 to 9.

2. Splitting sessions. Because the cropping frame of videos is set manually, the scale, rotation, and illumination condition of all videos are not strictly the same. Thus, cross-validation on 12 sessions shows the ability of learning these invariant video attributes. All sessions are split into three group of sessions each consists of 4 sessions.

3. Splitting instances. The result with this data splitting method is provided for comparing purpose. All instances are randomly split into three groups with the same size.

Table 7 shows the results with the three data splitting methods mentioned above.

Learning Curve of Our Model

Because there are only 3 rats and 12 sessions (each session contains a

Data Utilized	Features	Accuracy
Video+EMG	Eigenfaces, DSP, LBP DFA, MAV1, RMS, WL	0.8200(0.0051)
	Eigenfaces, DIC, LBP DFA, MAV1, RMS, WL	0.8191(0.0051)
	Fisherfaces, DIC, LBP DFA, MAV1, RMS, WL	0.8189(0.0051)
Video	Eigenfaces, DSP, LBP	0.8041(0.0053)
	Eigenfaces, Fisherfaces, LBP	0.8032(0.0053)
	Eigenfaces, DIC, LBP	0.8032(0.0053)
EMG	IAV, MAV1, MFL, RMS	0.6078(0.0065)
	IAV, MAV1, RMS	0.6078(0.0065)
	ApEn, IAV, MAV1, RMS	0.6073(0.0065)

Table 2: The accuracy of our model with selected feature combinations. The numbers in parentheses are the standard error of the accuracy. The SVM (RBF) model is applied to all feature combinations in the table. Also notice that a model with only LBP achieves an accuracy of 0.7963(0.0053). While the best model without LBP achieves 0.7697(0.0056). LBP is a dominating feature for OAR.

Predict	Other	Lick	Gape	LTP	TP	MM	MO	Paw	NM
Other	0.72	0.02	0.07	0.04	0.00	0.04	0.11	0.04	0.00
Lick	0.02	15.96	0.02	0.60	0.04	2.49	0.92	0.00	0.21
Gape	0.72	0.00	3.69	0.02	0.00	0.05	1.02	0.04	0.00
LTP	0.02	0.79	0.00	15.69	0.00	0.78	0.04	0.00	1.06
TP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MM	0.19	2.43	0.12	0.48	0.04	16.05	0.71	0.02	1.22
MO	0.05	0.28	0.74	0.04	0.07	0.37	3.05	0.00	0.04
Paw	0.02	0.05	0.00	0.05	0.00	0.04	0.00	9.56	0.02
NM	0.07	0.11	0.04	0.12	0.00	1.36	0.19	0.02	17.28

Table 3: The prediction-groundtruth matrix of our model with the best EMG and video feature combination. The i, j-th entry of the matrix is the percentage of instances predicted as i-th class but labeled as j-th class.

video and EMG series) in total, we examine the learning curve of our model

Predict	Other	Lick	Gape	LTP	TP	MM	MO	Paw	NM
Other	0.67	0.05	0.07	0.04	0.00	0.07	0.00	0.07	0.00
Lick	0.11	13.88	0.32	1.97	0.07	3.97	2.38	1.97	0.53
Gape	0.18	0.02	3.02	0.51	0.00	0.18	0.62	0.00	0.00
LTP	0.23	1.66	0.48	10.26	0.05	2.17	0.55	1.78	0.12
TP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MM	0.39	3.17	0.16	3.51	0.02	13.52	0.74	4.74	1.38
MO	0.16	0.67	0.60	0.34	0.00	0.42	1.68	0.25	0.04
Paw	0.05	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00
NM	0.04	0.18	0.04	0.41	0.00	0.81	0.07	0.85	17.76

Table 4: The prediction-groundtruth matrix of our model with the best EMG feature combination only. The i, j-th entry of the matrix is the percentage of instances predicted as i-th class but labeled as j-th class.

Predict	Other	Lick	Gape	LTP	TP	MM	MO	Paw	NM
Other	0.35	0.02	0.18	0.02	0.00	0.02	0.12	0.00	0.04
Lick	0.02	15.91	0.04	0.58	0.04	2.54	0.95	0.00	0.25
Gape	0.99	0.00	3.70	0.02	0.00	0.05	1.11	0.05	0.00
LTP	0.02	0.71	0.02	15.62	0.00	0.85	0.05	0.00	1.09
TP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MM	0.16	2.56	0.05	0.53	0.04	15.61	0.74	0.02	1.32
MO	0.09	0.25	0.65	0.05	0.07	0.39	2.80	0.00	0.26
Paw	0.02	0.05	0.00	0.04	0.00	0.04	0.00	9.58	0.02
NM	0.18	0.16	0.04	0.18	0.00	1.68	0.25	0.02	16.84

Table 5: The prediction-groundtruth matrix of our model with the best video feature combination only. The i, j-th entry of the matrix is the percentage of instances predicted as i-th class but labeled as j-th class.

	classified by EMG	misclassified by EMG
classified by video	52.21%	28.20%
misclassified by video	8.57%	11.02%

Table 6: Comparing video and EMG features in OAR. The instances misclassified by video features are misclassified by EMG features with a probability of 56.25%.

Generalization Level	Rats	Sessions	Instances
Accuracy	78.38%	79.76%	89.36%

Table 7: The generalization ability of our model. We apply cross-validation 4 times and the averaged accuracy are shown. Notice that the generalization accuracy between sessions (79.76%) is lower than the accuracy shown in Table 2 (82.00%) because less sessions are used as training data. From this table we conclude that when the training data contains at least two rats, our model can generalize the orofacial activities to other rats well. However, the differences between sessions, e.g. scale, rotation, illumination conditions of the video and possible EMG differences reduced the accuracy of our model significantly.

in terms of the number of rats and sessions. For the learning curve in terms of the number of sessions, we select n sessions randomly as training data and the remaining $12 - n$ sessions as test data. In terms of the number of rats, the number of sessions remains the same while the number of rats changes (some sessions may not be used in the training nor test set). The features used for evaluation are the best feature combination selected before and the hyperparameters of SVM are retuned. This procedure is repeated several times then the curve between n and the averaged classification accuracy is drawn and shown in Figure 13.

The probability of increasing accuracy in our figures is computed in the following way: Assume the true distribution of data (accuracy) is $at + b + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma)$. Then given the observation $\mathbf{x}$ over t and with no prior knowledge of $\mathrm{P}(a, b)$, the following relation is derived:

$$\mathrm{P}(\mathbf{x}|a, b) \propto \mathrm{P}(a, b|\mathbf{x}) \tag{13}$$

Then the probability of the true distribution has a positive trend can be computed below:

$$\mathrm{P}(a > 0|\mathbf{x}) = \frac{\displaystyle\int_0^\infty \int_{-\infty}^\infty \mathrm{P}(\mathbf{x}|a, b)\,\mathrm{d}b\,\mathrm{d}a}{\displaystyle\int_{-\infty}^\infty \int_{-\infty}^\infty \mathrm{P}(\mathbf{x}|a, b)\,\mathrm{d}b\,\mathrm{d}a}, \tag{14}$$

where $\mathrm{P}(\mathbf{x}|a, b)$ is the likelihood of data. Notice that this estimation is invariant to linear transformation to either $\mathbf{x}$ or t.

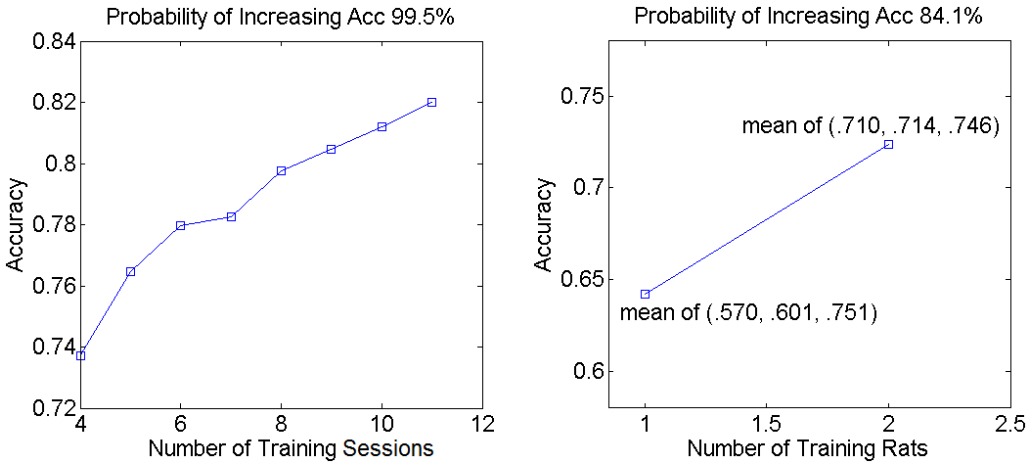

Figure 13: The learning curve of OAR. The accuracy is rising with more labeled sessions and possibly more rats (the number of unique rats of the sessions). The accuracy in the left figure is the averaged accuracy of 24 random split validations. In the right figure, the accuracy is the averaged accuracy of 3-fold cross-validation.

Chapter 5: Cue Prediction

Feature Selection

We also have tested our model and features to retro-classify the cue using the video and EMG of the rat during the time period following the cue. Interesting results about the learning behavior of rats are obtained. Each instance in this problem contains EMG and video frames in 3 seconds. Due to the large amount of frames in each instance, video features cannot be extracted in the previous way because the dimensionality of instances would be unacceptable. Four different video features are tested for this problem:

1. Predicted Orofacial Activities (POA). Each frame in the video of the instance is represented as a 9 dimensional vector and the dimensionality of this feature is $9n$ where n is the number of frames. The entry of the vector is defined as $\mathbf{v}_i = \mathrm{P}(y = i|\mathbf{x})$, where x is the input frame. The probability $\mathrm{P}(y = i|\mathbf{x})$ is generated by the OAR with SVM (RBF) kernel. To output an estimated probability, the LIBSVM basically use SVR and normalize the output to a $[0, 1]$ value (Chang and Lin, 2011).

2. Histogram of Orofacial Activities (HOA). Instead of representing each frame as a 9 dimensional vector, we aggregate all the orofacial activities to a 9-bin histogram.

3. PCA on LBP. Because LBP is a dominating feature for OAR, we simply apply dimensionality reduction method on LBP of all frames. The number of principal components adopted is 50 based on the observation of the eigenvalues.

4. LDA on LBP. Similar to the previous feature, we also apply LDA to LBP of all frames. Notice that the dimensionality of LBP are reduced to 1000 by PCA first in order to run LDA correctly. 50 linear projections are derived such that the external variance between classes is maximized and the internal variance within classes is minimized. We use source codes online to perform multi-class LDA (Alzahrani, 2014).

After all features are computed, the same feature selection method discussed

in the previous Chapter is then applied.

Results

We apply 33-fold cross-validation on instances from 33 sessions. Table 8 shows the feature selection results. Table 9 shows the prediction-groundtruth matrix of our model. Notice that the video features do not significantly contribute to the final accuracy. Table 10 shows that the instances misclassified by EMG features are also very likely misclassified by video features.

Data Utilized	Features	Accuracy(StdErr)
Video+EMG	PCA on LBP, MAV1, RMS, GM	0.4139(0.0042)
	PCA on LBP, MFL, RMS, WL, GM	0.4136(0.0042)
	PCA on LBP, RMS, SampEn, WL, GM	0.4136(0.0042)
Video	PCA on LBP, HOA	0.3908(0.0103)
	PCA on LBP	0.3868(0.0102)
	PCA on LBP, LDA on LBP	0.3837(0.0102)
EMG	MAV1, RMS, WL, GM	0.4118(0.0033)
	ApEn, RMS, WL, GM	0.4097(0.0033)
	MFL, RMS, WL, GM	0.4097(0.0033)

Table 8: The accuracy of cue prediction with selected feature combinations. The numbers in parentheses are the standard error of the accuracy. The SVM (RBF) model is applied to all feature combinations in the table.

Using the Data After Taste Delivery

Since the rats' reaction to the taste should be stronger than the reaction to the cue, we can classify the cue more accurately utilizing the data (video and EMG) after the taste delivery. In other words, we are retro-classifying the taste instead of the cue. The classification accuracy of this problem provides a lower bound of how accurate we can achieve on retro-classifying the cue with the existing features, if the rats are 100 percent sensitive to the cues. Results are shown in Table 11 and 12.

Predicted	CAcid	NaCl	Sucrose	Quinine	Null
CAcid	5.75	2.98	3.38	3.53	1.46
NaCl	3.57	7.35	4.26	4.73	1.54
Sucrose	4.22	3.35	6.91	3.89	0.89
Quinine	4.18	3.54	3.77	5.83	0.81
Null	2.22	2.80	1.66	2.13	15.23

Table 9: The prediction-groundtruth matrix of cue prediction with the model and features of the highest accuracy. The i, j-th entry of the matrix is the percentage of instances predicted as i-th class but labeled as j-th class. Notice that the null cue can be classified most accurately. The best accuracy evaluated by cross-validation of classifying four cues (without the null cue) is 36.00% (chances are 25.14%).

	classified by EMG	misclassified by EMG
classified by video	24.97%	14.04%
misclassified by video	16.23%	44.76%

Table 10: Comparison of the video and EMG features in Cue Prediction. The instances misclassified by EMG features are misclassified by video features with a probability of 76.12%.

Data Utilized	Features	Accuracy(StdErr)
Video+EMG	PCA on LBP, MAV1, RMS, GM	0.6564(0.0100)

Table 11: The accuracy of taste prediction with the best feature combination selected by cue prediction. The numbers in parentheses are the standard error of the accuracy. SVM (RBF) model is applied and its hyperparameters are retuned.

Orofacial Activities to Cues

One way to examine the reaction of rats to different cues, we add up the predicted orofacial activities of the rat. Table 13 shows the result. Also, Table 14 shows the reaction of the rat after the taste delivery.

Predicted	CAcid	NaCl	Sucrose	Quinine	Null
CAcid	18.04	0.57	1.24	1.41	0.09
NaCl	0.22	11.49	6.90	4.82	0.40
Sucrose	0.84	4.24	7.38	4.11	0.13
Quinine	0.57	3.49	4.24	9.59	0.18
Null	0.27	0.22	0.22	0.18	19.14

Table 12: The prediction-groundtruth matrix of taste prediction with the best feature combination selected by cue prediction. The i, j-th entry of the matrix is the percentage of instances predicted as i-th class but labeled as j-th class.

OAR%	CAcid	NaCl	Sucrose	Quinine	Null
Other	0.12(0.8)	0.07(0.6)	0.04(0.4)	0.05(0.5)	0.02(0.3)
Lick	11.8(10.5)	11.3(10.5)	12.1(10.1)	12.6(10.7)	3.2(6.4)
Gape	0.54(3.3)	0.03(0.3)	0.10(1.3)	0.25(2.2)	0.06(1.3)
LTP	0.69(1.8)	0.78(2.3)	0.93(2.8)	0.86(2.0)	0.64(2.1)
TP	0.00(0.0)	0.00(0.0)	0.00(0.0)	0.00(0.0)	0.00(0.0)
MM	59.5(23.0)	60.8(23.5)	62.9(22.4)	62.8(20.8)	27.3(28.8)
MO	1.81(3.7)	1.73(3.6)	1.62(3.4)	1.82(3.6)	0.50(4.2)
Paw	0.44(6.7)	1.18(10.2)	1.05(10.0)	1.54(12.3)	0.48(6.7)
NM	25.1(25.5)	24.1(25.2)	21.3(22.7)	20.1(20.4)	67.8(34.0)

Table 13: Percentage of orofacial activities to cues. The numbers in parentheses are the standard error of the percentage. For example, after a cue of CAcid, $11.8\% \pm 10.5\%$ of the orofacial activities are classified as Lick.

Learning Curve of Rats

The accuracy of cue prediction is lower than the accuracy of taste prediction (using the data after taste delivery). One possible reason is that the rats are still learning the correspondences between the cues and taste. In this case, a rising prediction accuracy with respect to training time (days) should be observed. We evaluate the accuracy of our model on sessions of different dates in two ways:

OAR%	CAcid	NaCl	Sucrose	Quinine	Null
Other	0.05(0.3)	0.21(1.1)	0.12(0.6)	0.18(1.1)	0.04(0.6)
Lick	30.4(10.1)	20.6(11.1)	21.8(10.3)	20.9(11.0)	2.34(4.9)
Gape	0.21(0.8)	0.08(0.4)	0.05(0.3)	0.26(1.0)	0.04(0.7)
LTP	1.07(2.0)	1.31(2.8)	2.37(4.0)	3.34(4.4)	0.46(2.1)
TP	0.00(0.0)	0.00(0.0)	0.00(0.0)	0.00(0.0)	0.00(0.0)
MM	63.7(10.4)	67.2(14.4)	66.6(12.8)	65.7(14.3)	23.3(25.4)
MO	2.13(3.9)	2.91(5.7)	2.63(5.1)	2.61(4.7)	0.48(3.2)
Paw	0.39(6.0)	0.77(8.3)	0.93(9.1)	1.34(11.2)	0.43(5.9)
NM	2.12(5.1)	6.86(12.3)	5.49(9.5)	5.70(11.4)	72.9(29.6)

Table 14: Percentage of orofacial activities in response to different taste deliveries. The numbers in parentheses are the standard error of the percentage. For example, after a delivery of CAcid, $30.4\% \pm 10.1\%$ of the orofacial activities are classified as Lick.

1. Cross-validation. Only the session to be tested is isolated from the training set.

2. Batch-validation. The training set consists of half of the sessions. To illustrate, assume the dates of the sessions are $\{1, 2, 3, 4, 5, 6\}$. The sessions in date $\{1, 3, 5\}$ are merged to the training set and the sessions left are the test set. The advantage of this method is that the training set is consistent, therefore the comparison the more reasonable.

Figure 14 shows the learning curve of rats with above-stated approaches. Notice that in order to show the learning curve of the rats, several training sessions which were not included before are also evaluated. Unfortunately, most training sessions do not have EMG recorded. Thus, we use only the video features for cue prediction without a significant drop in accuracy (see Table 8).

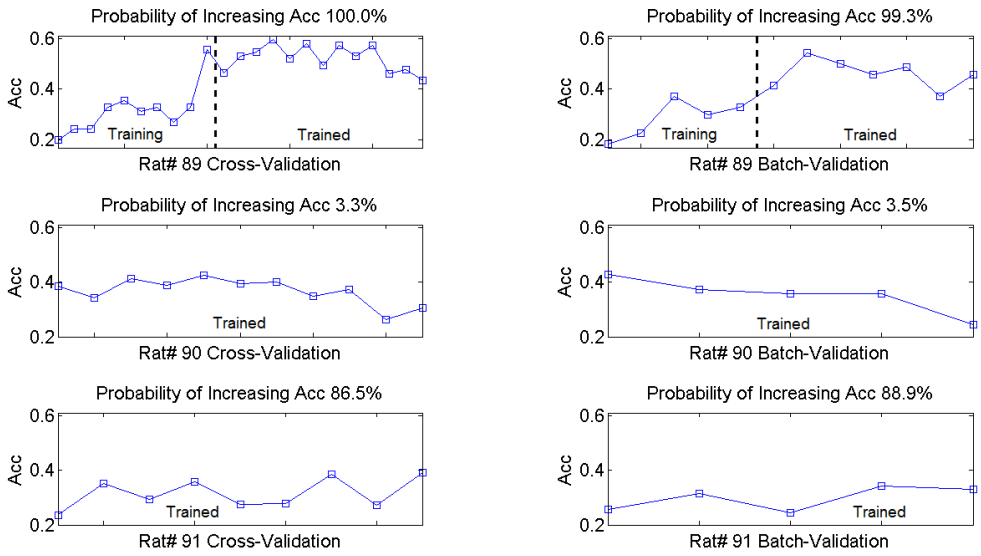

Figure 14: Learning curve of rats evaluated by cross-validation and batch-validation. Notice that in the figure of cross-validation, the first 10 days of Rat#89 are training days. In the figure of batch-validation, the numbers of training days are 5.

Chapter 6: Conclusions

We propose a systematic, automatic, objective, and accurate tool for orofacial activity recognition. This recognition tool is based on SVM and novel features extracted from videos and EMG. The accuracy of our model is 82.0% for recognizing orofacial activities of rats consuming palatable and aversive tastes. We show that in an application of classical conditioning (one of 5 different cues is given to the rat before the corresponding taste delivery), our features can distinguish the reactions of rats to 5 different taste with an accuracy of 65.6%. Furthermore, the reactions to 5 cues before the corresponding taste delivery can also be classified with an accuracy of 41.4%. With this tool, the learning curve of a rat in the training of classical conditioning can be clearly seen and the resulting reactions to all cues can be quantified in terms of orofacial activities. In conclusion, our proposal provides a powerful tool for analyzing orofacial activities and can be applied across various neuroscience research.

References

Md. R. Ahsan, Muhammad I. Ibrahimy, and Othman O. Khalifa. EMG signal classification for human computer interaction: A review. *European Journal of Scientific Research*, 33(3):480–501, 2009.

Sultan Alzahrani. FDA LDA multiclass. `http://www.mathworks.com/matlabcentral/fileexchange/45006-fda-lda-multiclass`, 2014.

Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, Jul. 1997.

K.C. Berridge. Measuring hedonic impact in animals and infants: microstructure of affective taste reactivity patterns. *Neuroscience and Biobehavioral Reviews*, 24(2):173–198, Mar. 2000.

Christopher M Bishop et al. *Pattern recognition and machine learning*, volume 1. springer New York, 2006.

Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152, Jul. 1992.

H.A. Bruck, S.R. McNeill, M.A. Sutton, and III Peters, W.H. Digital image correlation using newton-raphson method of partial differential correction. *Experimental Mechanics*, 29(3):261–267, 1989.

Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

Ira Cohen, Ashutosh Garg, and Thomas S. Huang. Emotion recognition from facial expressions using multilevel hmm. In *Neural Information Processing Systems*, 2000.

Ira Cohen, Nicu Sebe, Fabio G. Cozman, Marcelo C. Cirelo, and Thomas S. Huang. Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. In *IEEE Computer Society*

Conference on Computer Vision and Pattern Recognition, volume 1, pages 595–601, Jun. 2003.

Hong-Bo Deng, Lian-Wen Jin, Li-Xin Zhen, and Jian-Cheng Huang. A new facial expression recognition method based on local gabor filter bank and pca plus lda. *International Journal of Information Technology.*, 11(11): 86–96, 2005.

Will Dwinnell. Lda: Linear discriminant analysis. `http://www.mathworks.com/matlabcentral/fileexchange/29673-lda--linear-discriminant-analysis`, 2010.

Harvey J. Grill and Ralph Norgren. The taste reactivity test. i. mimetic responses to gustatory stimuli in neurologically normal rats. *Brain Research*, 143(2):263–279, Jun. 1978.

Marko Heikkila and Timo Ahonen. A general local binary pattern implementation for Matlab. `http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab`, 2013.

Chao-Kuei Hsieh, Shang-Hong Lai, and Yung-Chang Chen. An optical flow-based approach to robust face recognition under expression variations. *IEEE Transactions on Image Processing*, 19:233–240, Jan. 2010.

Bernard Hudgins, Philip Parker, and Robert N Scott. A new strategy for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 40(1):82–94, 1993.

Kang Soo Kim, Heung Ho Choi, Chang Soo Moon, and Chi Woong Mun. Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions. *Current applied physics*, 11(3):740–745, 2011.

Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, volume 2, pages 674–679, 1981.

Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.

Bing Pan and Kai Li. A fast digital image correlation method for deformation measurement. *Optics and lasers in engineering*, 49(7):841–847, 2011.

Ivan P. Pavlov and G. V. Anrep. Classics in the history of psychology. `http://psychclassics.yorku.ca/Pavlov/`, 1927.

C-K Peng, Shlomo Havlin, H Eugene Stanley, and Ary L Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):82–87, 1995.

Angkoon Phinyomark, Pornchai Phukpattaranont, and Chusak Limsakul. Feature reduction and selection for EMG signal classification. *Expert Systems with Applications*, 39(8):7420 – 7431, 2012a.

Angkoon Phinyomark, Pornchai Phukpattaranont, and Chusak Limsakul. Fractal analysis features for weak and single-channel upper-limb emg signals. *Expert Systems with Applications*, 39(12):11156–11163, 2012b.

Angkoon Phinyomark, Franck Quaine, Sylvie Charbonnier, Christine Serviere, Franck Tarpin-Bernard, and Yann Laurillau. EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert Systems with Applications*, 40(12):4832 – 4840, 2013.

M.B.I. Reaz, M.S. Hussain, and F. Mohd-Yasin. Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological Procedures Online*, 8(1):11–35, 2006.

Joshua S Richman and J Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6):H2039–H2049, 2000.

Chad L. Samuelsen, Matthew P.H. Gardner, and Alfredo Fontanini. Effects of cue-triggered expectation on cortical processing of taste. *Neuron*, 74: 410–422, Apr. 2012.

Caifeng Shan, Shaogang Gong, and Peter W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27:803–816, 2009.

Michael A. Sutton, Jean-Jos Orteu, and Hubert Schreier. *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts,Theory and Applications*. 2009.

Dennis Tkach, He Huang, and ToddA Kuiken. Study of stability of time-domain features for electromyographic pattern recognition. *Journal of NeuroEngineering and Rehabilitation*, 7(1):1–13, 2010.

Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, Jan. 1991.

Max Welling. Fisher linear discriminant analysis. `http://www.ics.uci.edu/~welling/classnotes/papers_class/Fisher-LDA.pdf`, unknown.

Mahyar Zardoshti-Kermani, Bruce C Wheeler, Kambiz Badie, and Reza M Hashemi. Emg feature evaluation for movement control of upper extremity prostheses. *IEEE Transactions on Rehabilitation Engineering*, 3(4):324–333, 1995.

Micera Zecca, Silvestro Micera, MC Carrozza, and P Dario. Control of multifunctional prosthetic hands by processing the electromyographic signal. *Critical Reviews in Biomedical Engineering*, 30(4–6):459–485, 2002.

Ligang Zhang and Dian Tjondronegoro. Facial expression recognition using facial movement features. *IEEE Transactions on Affective Computing*, 2 (4):219–229, Oct. 2011.

Xu Zhang and Ping Zhou. Sample entropy analysis of surface emg for improved muscle activity onset detection against spurious background spikes. *Journal of Electromyography and Kinesiology*, 22(6):901–907, 2012.

Jingdong Zhao, Li Jiang, Hegao Cai, Hong Liu, and Gerd Hirzinger. A novel emg motion pattern classifier based on wavelet transform and nonlinearity analysis method. In *Proceedings of IEEE international conference onrobotics and biomimetics*, pages 1494–1499, 2006a.

Jingdong Zhao, Zongwu Xie, Li Jiang, Hegao Cai, Hong Liu, and Gerd Hirzinger. Emg control for a five-fingered underactuated prosthetic hand based on wavelet transform and sample entropy. In *Proceedings of IEEE/RSJ international conference on intelligent robots and systems*, pages 3215–3220, 2006b.