Acume: A Novel Visualization Tool for Understanding Facial Expression and Gesture Data

***Abstract*— Facial and head actions contain significant affective information. To date, these actions have mostly been studied in isolation because the space of naturalistic combinations is vast. Interactive visualization tools could enable new explorations of dynamically changing combinations of actions as people interact with natural stimuli. This paper describes a new open-source tool that enables navigation of and interaction with dynamic face and gesture data across large groups of people, making it easy to see when multiple facial actions co-occur, and how these patterns compare and cluster across groups of participants.**

We share two case studies that demonstrate how the tool allows researchers to quickly view an entire corpus of data for single or multiple participants, stimuli and actions. *Acume* yielded patterns of actions across participants and across stimuli, and helped give insight into how our automated facial analysis methods could be better designed. The results of these case studies are used to demonstrate the efficacy of the tool.

The open-source code is designed to directly address the needs of the face and gesture research community, while also being extensible and flexible for accommodating other kinds of behavioral data. Source code, application and documentation are available at [website address to be added in final version].

I. INTRODUCTION

Facial and head actions serve a communicative function and contain significant affective and cognitive information [1], [2]. These actions are dynamic and may unfold differently over time for different people and in response to different stimuli. To date it has not been easy to interpret the individual responses people exhibit and the data has mostly been looked at in isolation, as the space of naturalistic behaviors of different people over time is vast. To even identify the similarities and differences in responses to a fixed stimuli researchers need to be able to analyze data across large populations, investigating both inter and intra group differences. In addition, researchers need to be able to understand dynamics on micro and macro timescales. Visualization tools are becoming integral to the research process in many fields as they enable new questions about the structure and form of data to be asked and answered. However, few applications exist that allow a user to visualize and interact with human behavioral data in such a way.

As an example, in the study of facial expressions, psychologists seek to understand the significance, dynamics and affective interpretation of different actions [3]. An objective way of evaluating facial expressions is the Facial Action Coding System (FACS) [4], [5]. The latest FACS update describes 54 uniquely identifiable facial and head actions and there are thousands of combinations in which these facial and head movements can occur in natural situations, both instantaneously and dynamically. It is a challenging problem to learn which ones are reliable and relevant in a particular context. As such, understanding facial expression data, featuring many action units over time, for many different people in different contexts can be difficult and time consuming. We present an open-source tool that enables navigation of human behavioral data across large corpora and is, to the best of our knowledge, the first of its kind. While it can be applied generally, *Acume* was developed for and is applied here to facial and head actions. For convenience, in the remainder of this paper we use refer to action units (AUs) as our labels, but these can be seen as generic behavioral annotations and not specifically FACS labels.

With the increasing amounts of automatically labeled facial action and gesture data [6], [7], [8], [9] as well as larger and richer corpora of hand labeled data, there is great benefit in having efficient tools to analyze this information. *Acume* is an open-source application that allows researchers to quickly view an entire corpus of behavioral data over time for a large set of participants, stimuli and actions. The goal of this work is to demonstrate how this tool that can be used with a generic set of labeled behavioral data across different contexts. This toolkit enables a researcher to explore aggregated responses across a large population, to automatically cluster responses into groups, to compare individual responses side-by-side, to interactively select and deselect actions and participants of interest, and to efficiently find specific frames in videos of the stimuli and responses by interacting with the data. In addition, *Acume* allows researchers to interact with the data across different time scales, from single frames to many hours.

Acume was used to answer existing, and new, research questions for two sets of ecologically valid data. We were able to 1) identify high-level patterns of facial responsiveness, 2) identify the low-level AUs involved, 3) analyze specific patterns of co-occurrence and sequential occurrence of these AUs, and 4) reevaluate how labeling systems were performing. It was possible to make qualitative and quantitative judgments about the different clusters within the population of responses.

In the remaining part of this paper we describe the design of the toolkit, its structure and the form of input required, two data sets that are used to demonstrate its capabilities and two case studies that demonstrate the efficacy of the application. Finally, we describe results obtained using the application, conclusions and future work.

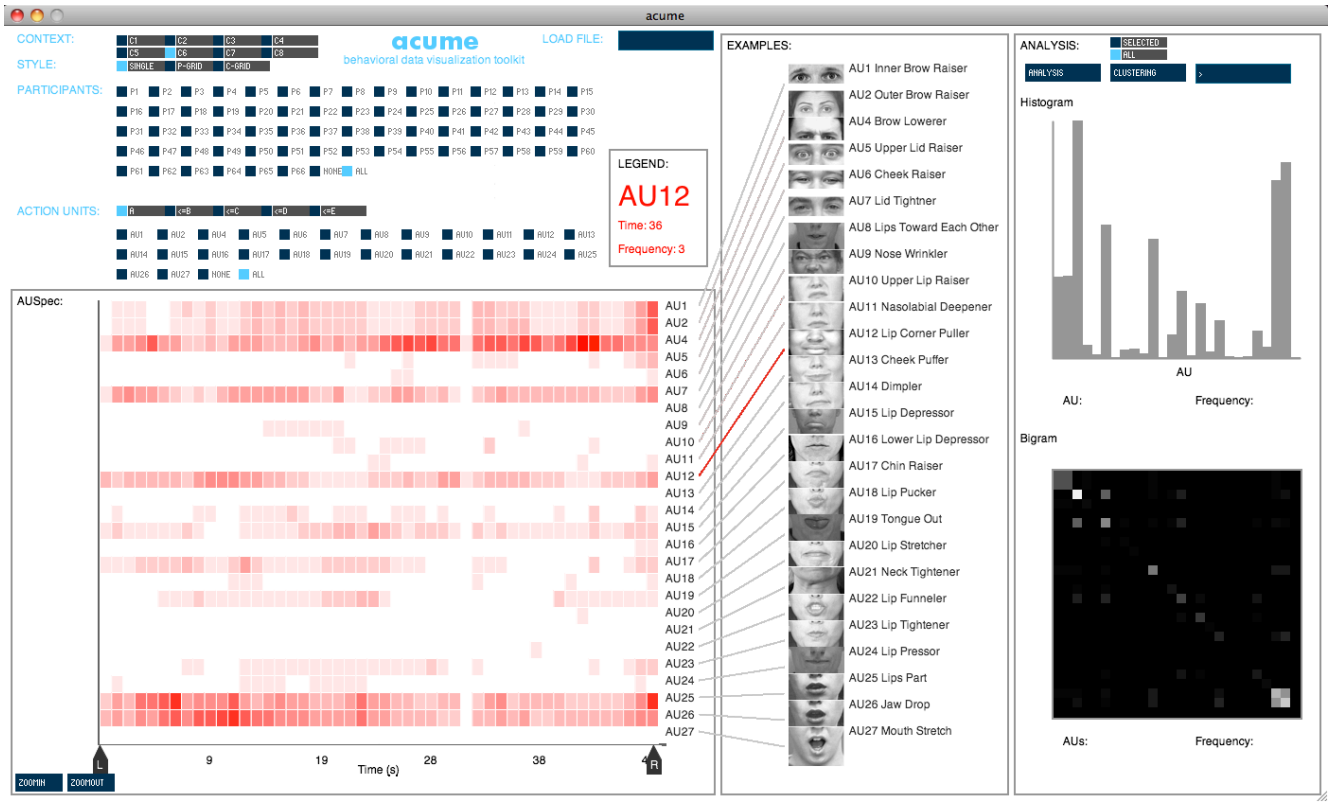

Fig. 1. A screenshot of *Acume*, an open-source toolkit for exploring and visualizing behavioral data on multiple scales.

II. BACKGROUND

A number of automated facial expression and gesture labeling systems exist [6], [7], [8]. These provide the ability to annotate videos with behavioral labels, such as FACS codes. These can be used for efficient labeling of large amounts of data that previously was not possible. In this work our automated labels are generated by FaceSense [10], a computational model for mapping video input to FACS labels and affective-cognitive states. FaceSense uses the displacement and velocity of 21 tracked feature points on the face in order to determine a binary set of action unit labels for each frame.

A useful open-source tool for hand labeling videos with behavioral data is VIDL [11]. This provides an efficient interface for human labeling by allowing distributed labeling over place and time. It can be adapted to add any set of appropriate labels that are specific to a particular task. Amazon's Mechanical Turk is another popular facility for distributed labeling of videos. The Observer XT software [12] is one example of commercially available software for labeling and displaying action unit data. These systems do not provide the ability to compare and aggregate data across groups, looking at individual responses and trends across a population. However, the output from VIDL and other such labeling systems could be easily formatted as input to *Acume*.

Data visualization applications have been utilized to great effect in other fields, such as genomics [13], [14]. These systems allow researchers to explore high dimensional data at different scales and to ask new questions related to this data. Few attempts have been made to provide a structured way to visualize and interact with facial expression and gesture data from large populations, across different dimensions and scales. Inspired by recent developments on the data visualization front, we provide the first open-source application that allows researchers to explore multi-person dynamic face and gesture data across multiple dimensions.

III. ACTION UNIT SPECTROGRAMS (AUSPEC) AND GRID REPRESENTATIONS

One feature of *Acume* is its ability to facilitate the exploration of facial expression and gesture data from large populations. It incorporates different views to summarize the data from multiple participants over time. The first method is through the use of Action Unit Spectrograms (AUSpec), a novel way to aggregate facial action unit data [15], but which we generalize to other behavioral data. Figure 2 shows an AUSpec encoding the frequency of each AU every second across all viewers for two movie clips, an excerpt from 'When Harry Met Sally' and an excerpt from 'The Lion King'. The color intensity of the plot at a particular point is proportional to the number of times a specific intensity action unit occurred at a particular second across the group.

AUSpecs are an informative way of displaying aggregated information across large groups of subjects, clearly visualizing similar responses at a particular point in time. However, an individual's response can become lost in the aggregation. *Acume* addresses this challenge with the GRID

view, which provides the ability to view spectrograms of specific participants side-by-side. The GRID view allows researchers to compare the 'fingerprint' of each response, to visually cluster responses, to identify responses that were significantly more expressive than others and to easily determine those participants whose responses deviated from the aggregate response.

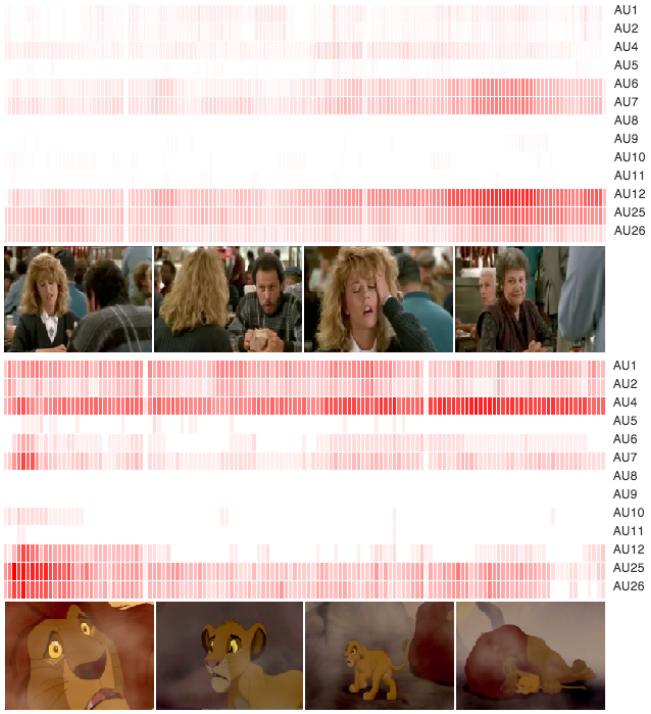

Fig. 2. Action Unit Spectrograms (AUSpec) from the 66 participants watching 'When Harry Met Sally' (top) and 'The Lion King' (bottom). The action unit numbers refer to FACS codes.

IV. APPLICATION

The main contribution of this work is a new open-source application *Acume*, designed to facilitate visualization of dynamic face and head action data across large groups of people, and to compare and cluster specific groups of responses. It is particularly effective when comparing time aligned responses of participants to a stimuli in cases where the presentation and duration of particular stimuli are known. The application allows the user to load in multiple sets of data attaining to different 'contexts' simultaneously. This provides easy comparisons of the responses during different events and ability to search the stimuli and response videos by clicking on the corresponding region of the AUSpec or GRID view.

In the design of the application several key criteria were evaluated. The application was designed as an open-source tool. It was developed in the Processing language for its ease of use and design friendly properties. The adaptability and the extensibility, allow others to use the tool and to add different functions to the front end (e.g. different AU labeling systems) and to the backend (e.g. different methods of clustering or visualizing the data).

There are no inherent limitations on the number of contexts, participants or action units that can be loading in any one instance. However, with a 1440 by 900 pixel display data for up 10 contexts, 88 participants and 34 actions can be loaded without overcrowding the interface. The duration of the data for each context can be unlimited and is restricted only by the visual resolution required.

The input to the application are matrices indicating the presence or absence, and intensity (if available), of each AU across time for each participant in each context. The first three columns indicate the participant's numerical code, the context's numerical code and time sample respectively. The subsequent columns, one column per AU, contain the AU data. The top row indicates the AU's numerical code.

The radio button controls on the interface allow the user to explore all, or a subset of data. Any combination of participants and any combination of AUs, at different intensities, can be selected for analysis. The interface consists of multiple linked panels, the Spectrogram Panel, the Examples Panel and the Analysis Panel.

A. *Spectrogram Panel*

The Spectrogram Panel has two views the AUSpec view, which represents the aggregated data from all participants, and the GRID view that displays side by side plots, one for each of the selected participants. In both cases action units can be turned on and off using the AU radio buttons. In the AUSpec view the user can zoom into a specific time segment and as the cursor is held over a region of the plot the AU, frequency and time instance are indicated in the legend above the plot. In the GRID view the AU and time instance information are displayed in the legend as the mouse hovers over each plot. The user can also select a particular GRID plot and blow it up.

B. *Examples Panel*

The application provides the option to display the definition and/or a visual example, of the selected AUs in the Examples Panel. Figure 1 shows the Examples Panel with relevant FACS AUs loaded. The images are taken from the Cohn-Kanade database [16] and the MMI database [17]. The user can load a definition text file and add a folder of example images specific to their task.

In the spectrogram mode the user has the ability to left-click on a particular time point and an image of the corresponding frame of the stimuli will appear to the right of the Examples Panel. In the GRID view a similar facility is available. This enables the user to click on the GRID plot of a particular subject at a particular time point and images of the corresponding frames of the videos of the subject and the stimuli will appear to the right of the Examples Panel.

Generic file naming systems should be used for the stimuli and participant videos, as for the example images, these are documented at [website address to be added in final version].

C. Analysis Panel

The Analysis Panel is divided into two areas. In both cases the following representations can include all data or just data for the selected participants and AUs.

The frequency tab provides a histogram and bigram representation of the AU data. The color intensity of each square in the bigram represents the relative frequency of AU combinations within the particular dataset. If the cursor is held over a region of the histogram or bigram plot the relevant AU (or AUs) and the frequency associated with them will be displayed below the plot. The bigram allows for immediate identification of frequently occurring AU combinations, information that is not available from the AUSpec.

The clustering tab provides the ability to apply an unsupervised clustering algorithm to the data for each response. A distance matrix is calculated to provide a measure of the similarity between responses. To calculate the similarity between responses a hamming measure is used and the absolute difference in bits calculated between the binary GRID spectrograms. The rows within the distance matrix are then clustered using K-Means. The number of clusters can be set by entering the number into the text box on the Analysis Panel.

The mode button on the Analysis Panel allows the user to view the mode spectrogram for each of these clusters. Indicating which action units at a particular time instance were seen across a majority of the members of that cluster.

D. Implementation

Acume is implemented using the Processing programming language [18] and the ControlP5 GUI library [19]. The application, source code and documentation are freely available at [website address to be added in final version]. Executables are available for MacOS and Windows.

V. CASE STUDIES

The following case studies demonstrate the efficacy of the application. The results and discussion show how the users were able to gain additional insights into the data using the application that were not previously possible.

A. DATA

This section describes two ecologically valid datasets used in the case studies. The first dataset consists of manually coded facial action unit data collected by Karim Kassam [20]. He obtained 528 FACS-coded videos consisting of 66 viewers each watching eight different movie clips. The movie clips were previously identified as strong elicitors of the six basic emotions [21]: Robin Williams Live (amusement, 81s), When Harry Met Sally (amusement, 154s), Cry of Freedom (anger, 155s), Pink Flamingos (disgust, 55s), Silence of the Lambs (fear, 204), Capricorn One (surprise, 47s), Sea of Love (surprise, 12s), The Lion King (sadness, 131s). The viewers were asked to watch the movie clips (11 to 204 seconds) and were simultaneously filmed. These videos were labeled with validated human-coded AUs. The videos were AU coded at one second intervals. These data were formated into a comma-separated values (CSV) file with 55100 samples for our study. We use data for AU1-AU27, excluding AU3.

TABLE I
ACTION UNITS CODED BY FACESENSE. * INDICATES AN ACTION NOT CODED IN FACS.

1 Inner eyebrow raiser	52 Head turn right	63 Eyes up
2 Outer eye brow raiser	53 Head up	64 Eyes down
12 Lip corner puller	54 Head down	Head motion left*
15 Lip corner depressor	55 Head tilt left	Head motion right*
18 Lip puckerer	56 Head tilt right	Head motion up*
25 Lips part	57 Head forward	Head motion down*
26 Jaw drop	58 Head back	Head motion forward*
27 Mouth stretch	61 Eyes turn left	Head motion back*
51 Head turn left	62 Eyes turn right	

The second dataset consists of automatically coded videos. Procter and Gamble obtained 80 videos of 20 participants watching the presentation of four International Affective Picture System (IAPS) database [22] images; a stool, a tiger, a baby and a dirty toilet. The three second period after the subject was initially exposed to the image was analyzed. The videos were automatically coded using FaceSense [23]. This system provided AU labels for the action units shown in Table I, these are based on the FACS system although where indicated the labels are variations on the conventional labels. Every frame of these videos was labeled. The videos, recorded using Quicktime (Apple Inc.), had a frame rate of 30 fps. As such, 90 frames were coded for each participant and image combination. These data were formatted into a CSV file with 7200 samples.

In both sets of data the AUs were labeled on a binary basis, the AU was present (1) or the AU was not present (0). We did not discriminate between the levels of the AUs observed. However, *Acume* provides the ability for the user to discriminate between five levels of AU intensity if this information is available. This was inspired by FACS that provides the ability to code AUs on five levels from the most subtle, A, to the least subtle, E. By changing the radio button the user can chose to display the data for all A level AUs or A+B level AUs etc.

VI. RESULTS AND DISCUSSION

In the first case study we were able to coherently represent AU data from 66 participants each viewing 8 unique stimuli. The data sets were labeled at 11 to 314 instances, depending on the length of stimuli. Figure 2 shows AUSpecs generated using *Acume* from the data for 'When Harry Met Sally' and 'The Lion King.' Figure 3 shows a screenshot of the AU bigram of the data for 'When Harry Met Sally.' By looking at Figure 2 it is immediately identifiable that there were significant occurrences of AU6, AU7, AU12, AU25 and AU26 across the group during 'When Harry Met Sally.' The legend on the interface identified that the highest frequency of occurrence was AU12 at 139s, this was present in 65% of

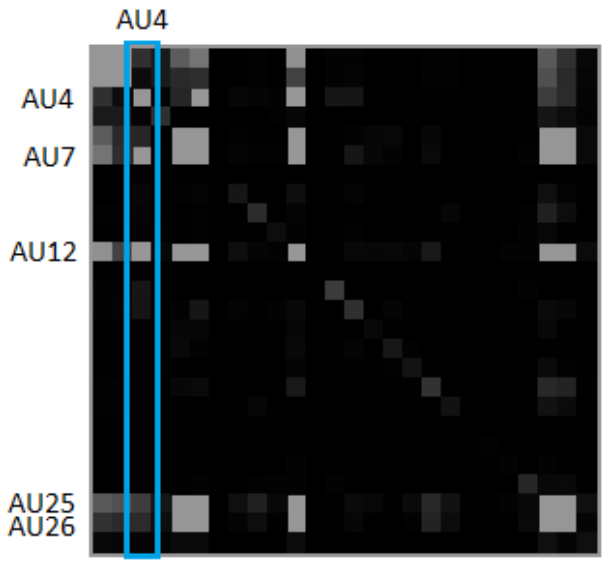

Fig. 3. Action unit bigram (AUBigram) from the 66 participants watching 'When Harry Met Sally'. The greyscale intensity is proportional to the number of occurrences of each combination. The action unit numbers refer to FACS codes. Occurrences of AU4 were significantly greater with AU7 and AU12 than with AU25 and AU26.

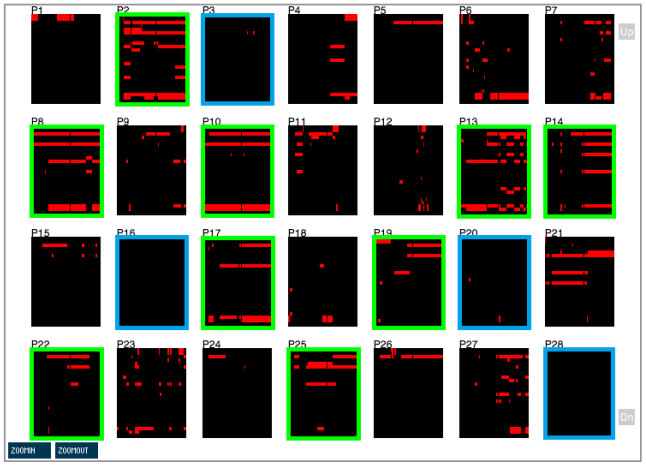

Fig. 4. Example of the responses of 28 of the participants from case study one whilst watching a clip from 'Pink Flamingoes'. Visual clustering can be used to identify those that are inexpressive (blue) and those that have similar components in their facial response (green).

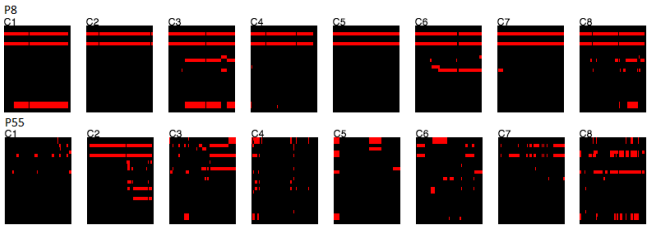

Fig. 5. Comparison of the responses of two participants across eight movie clips. It is clear that the responses of participant 8 (top) are much more uniform than those for participant 55 (bottom). Although both responses in movie two (Cry of Freedom) are similar.

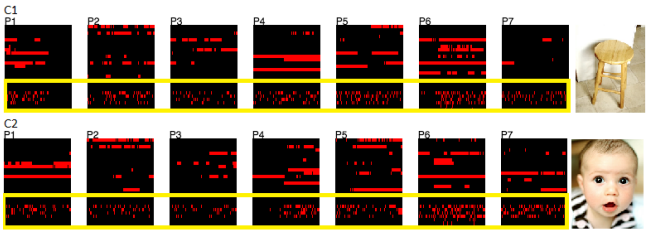

Fig. 6. Screenshots of the GRID view showing the responses of seven participants to two IAPS images. A stool (top) and a baby (bottom). The yellow box highlights the head motion action units which show higher order dynamics relative to the static action units.

participants. A more unexpected observation is the significant number of occurrences of AU4 between 81s and 130s. The bigram, Figure 3, enabled us to easily identify that AU4 occurs a majority of the time with AU7 and frequently with AU12 but occurred infrequently with AU6, AU26 and AU27. These relationships were also reflected in the corpus as a whole. The application allow us to choose to display only the data for the participants and action units selected using the radio buttons. Using this facility we could quickly iterate through combinations of research interest and discover which participants contributed the data of interest.

Figure 4 shows a subset of the GRID view for 28 of the participants whilst watching a clip from 'Pink Flamingoes.' The GRID view enabled the responses to be viewed simultaneously, complementing the AUSpec, helping to identify sub groups of similar responses and to isolate outliers. Figure 4 shows 28 responses to 'Pink Flamingoes.' It was immediately apparent which of the participants showed significant facial activity during the experiment. We also identified those subjects that were inexpressive, highlighted in blue. The GRID view was used to visually cluster the responses of the subjects, despite having data from 66 participants. We noted the similarity in responses for a sub set of participants, highlighted in green.

Figure 5 shows a comparison of the responses of two participants across all eight movies in case study one. Comparing the responses reveals that some responses are more uniform across contexts (participant 8 - top) than others (participant 55 - bottom).

Figure 2 demonstrates that the aggregated responses for 'The Lion King' and 'When Harry Met Sally' have different forms. This was reflected across the other stimuli. The AUSpecs for the amusement clips were similar but different from those for sadness and anger. *Acume* can be used to identify which stimuli induced less expressive responses from the population. The least expressive responses were during 'Silence of the Lambs', the most expressive responses were during the amusement clips.

The AUSpec representation was used to identify the frames of interest within the data. In case study one the users were able to quickly skip through the movie clip by clicking on points on the spectrograms. For 'When Harry Met Sally' the maxima in intensity of the plot corresponded to the climax of the scene.

The processing of labeled behavioral data is strongly dependent on its structure. As such, it is vital to understand the way the labeling systems perform. Previously, without visual representation, evaluating this performance was difficult. In both studies *Acume* was able to help the users understand the labels. Figure 6 shows screenshots of responses from seven participants to two IAPS images (a stool and a baby). The rows of the spectrograms corresponding to head motion action units are highlighted by the yellow box. Figure 6 shows that there are different dynamics in labels for motion action units as for static action units. Understanding this will have implications to the techniques used to learn this data and how the labeling system is tuned.

FACS provides an objective system for measuring facial activity. However, to date the interpretation of these action units has been naive. For instance, interpreting that AU12 indicates a "happy" state. The data from the two case studies described show that responses to media comprise of complex combinations of action unit sequences. *Acume* allows users to better understand how AUs map to states and how gestures manifest in different populations, in part, by simultaneously presenting the data on different scales, aggregate information (AUSpecs, histograms, bigrams) and also specific numerical values (mouse hover and zoom features).

VII. CONCLUSIONS

For two ecologically valid datasets of a total of 86 participants, 62300 samples of facial expressions and 12 stimuli *Acume* yielded new insights into the analysis of facial actions through an efficient, visual interface. The toolkit can be used to effortlessly create intuitive visualizations of multi-dimensional, dynamic behavioral data for large populations.

Acume gave the users new insights into the presence and co-occurrence of facial actions in different contexts and answered questions regarding the stimuli and participant behaviors. *Acume* helped identify those participants that were expressive compared to those that were inexpressive, those that responded in uniform ways to those with greater variance, and helped cluster responders efficiently. The interface provided visual evidence of the characteristics that defined group responses to particular affect inducing stimuli and also the ability to search video stimuli to find the climax. The users were better able to understand how AUs map to states and how gestures manifest in different populations, in part, by presenting the data on different scales.

Acume has an intuitive interface and simple input structure. Source code and documentation are publicly available making the application adaptable and extensible.

VIII. FUTURE WORK

This application is an open-source tool for the community to use for analysis of behavioral data. There is scope to integrate it with existing automatic behavioral labeling software and with manual labeling systems. Additional tools for post processing of the data could be incorporated into the system, such as more complex clustering algorithms and add-ons relating to specific mapping task, such as mapping responses to self-report or behavioral labels.

IX. ACKNOWLEDGMENTS

This work was funded by the Media Lab Consortium. We thank Procter and Gamble for funding a fellowship to support this work and providing data for case study two.

REFERENCES

[1] P. Ekman, W. Friesen, and S. Ancoli, "Facial signs of emotional experience," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1125–1134, 1980.

[2] S. Goldin-Meadow, "The role of gesture in communication and thinking," *Trends in Cognitive Sciences*, vol. 3, no. 11, pp. 419–429, 1999.

[3] K. Schmidt and J. Cohn, "Dynamics of facial expression: Normative characteristics and individual differences," 2001.

[4] P. Ekman and W. Friesen, "Facial Action Coding System (FACS): A technique for the measurement of facial action," *Palo Alto, CA: Consulting*, 1978.

[5] T. Wehrle and S. Kaiser, "Emotion and facial expression," *Affective interactions*, pp. 49–63, 2000.

[6] K. Anderson and P. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Transactions on Systems, Man, and CyberneticsPart B: Cybernetics*, vol. 36, no. 1, 2006.

[7] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *Journal of Multimedia*, vol. 1, no. 6, pp. 22–35, 2006.

[8] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 39–58, 2009.

[9] T. Moeslund, A. Hilton, and V. Kruger, "A survey of advances in vision-based human motion capture and analysis," *Computer vision and image understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[10] R. Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," *Real-time vision for human-computer interaction*, pp. 181–200, 2005.

[11] M. Eckhardt and R. Picard, "A more effective way to label affective expressions," *Proceedings of Affective Computing and Intelligent Interaction ACII-2009. Amsterdam*, 2009.

[12] Noldus, "Observer XT," 2006.

[13] M. Meyer, T. Munzner, and H. Pfister, "MizBee: a multiscale synteny browser," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 897–904, 2009.

[14] M. Meyer, B. Wong, M. Styczynski, T. Munzner, and H. Pfister, "Pathline: A Tool For Comparative Functional Genomics," in *Computer Graphics Forum*, vol. 29, no. 3. Wiley-Blackwell, 2010, pp. 1043–1052.

[15] D. McDuff, R. El Kaliouby, K. Kassam, and R. Picard, "Affect valence inference from facial action unit spectrograms," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 17–24.

[16] T. Kanade, Y. Tian, and J. Cohn, "Comprehensive database for facial expression analysis," *fg*, p. 46, 2000.

[17] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *2005 IEEE International Conference on Multimedia and Expo*. IEEE, 2005, p. 5.

[18] C. Reas and B. Fry, *Processing: a programming handbook for visual designers and artists*. The MIT Press, 2007.

[19] A. Schlegel, "ControlP5: A GUI Library For Processing."

[20] K. S. Kassam, "Assessment of emotional experience through facial expression," *Harvard University*, 2005.

[21] J. Gross and R. Levenson, "Emotion elicitation using films," *Cognition & Emotion*, vol. 9, no. 1, pp. 87–108, 1995.

[22] P. Lang, A. Ohman, and D. Vaitl, "The international affective picture system (photographic slides)," *Gainesville, FL: Center for Research in Psychophysiology, University of Florida*, 1988.

[23] R. El Kaliouby, "Mind-reading machines: automated inference of complex mental states," *Computer Laboratory, University of Cambridge*, 2005.