ANALYSIS OF INTERACTION ATTITUDES USING DATA-DRIVEN HAND GESTURE PHRASES

Zhaojun Yang[1], *Angeliki Metallinou*[2], *Engin Erzin*[3] *and Shrikanth Narayanan*[1]

[1]Signal Analysis and Interpretation Lab (SAIL), University of Southern California, Los Angeles, CA
[2]Pearson Knowledge Technologies, Menlo Park, CA
[3]Multimedia, Vision and Graphics Laboratory, College of Engineering, Koç University, Istanbul, Turkey

zhaojuny@usc.edu, angeliki.metallinou@pearson.com, eerzin@ku.edu.tr, shri@sipi.usc.edu

ABSTRACT

Hand gesture is one of the most expressive, natural and common types of body language for conveying attitudes and emotions in human interactions. In this paper, we study the role of hand gesture in expressing attitudes of friendliness or conflict towards the interlocutors during interactions. We first employ an unsupervised clustering method using a parallel HMM structure to extract recurring patterns of hand gesture (hand gesture phrases or primitives). We further investigate the validity of the derived hand gesture phrases by examining the correlation of dyad's hand gesture for different interaction types defined by the attitudes of interlocutors. Finally, we model the interaction attitudes with SVM using the dynamics of the derived hand gesture phrases over an interaction. The classification results are promising, suggesting the expressiveness of the derived hand gesture phrases for conveying attitudes and emotions.

***Index Terms*—** Interaction attitudes, hand gesture, motion capture, segmentation, clustering

1. INTRODUCTION

In human communication, body language is an essential element of nonverbal behavior for a person to express the attitudes, feelings or emotions towards his/her interlocutors [1]. Among the various types of body language, such as body posture, facial expressions and eye movements, hand gesture is one of the most expressive, common and natural forms in human interactions [2]. Understanding the role of hand gesture in conveying interaction attitudes or emotions is important for the applications of automatic emotion recognition, as well as the design of human-machine interface and virtual environment.

This work focuses on studying the role of hand gesture in expressing attitudes of friendliness and conflict in dyadic interactions within and across interlocutors. Analogously to visemes in lip motion, there are also elementary patterns for hand gesture, i.e., hand gesture phrases [3]. In the gesture model proposed by Kendon [3], a gesture phrase defines the basic gesture element and a complex gesture can be decomposed into multiple gesture phrases. Hence, we can use the gesture phrases to describe and model various gesture dynamics across persons and over different time scales. Our work is on the basis of the hand gesture phrases. Although some efforts have been devoted to exploring affective information in specific low-level movement features such as velocity and acceleration [4] [5], studies on the link of affect and gesture dynamics at the phrase (primitive) level are still limited so far. Our goal is three-fold: 1) Identifying elementary patterns of hand gesture (hand gesture phrases) recurring in interactions through unsupervised temporal segmentation and clustering; 2) Performing analysis to validate the usefulness of the automatically derived hand gesture phrases; and 3) Modeling and classifying the interaction attitudes using the dynamics of the hand gesture phrases over an interaction.

We use the multimodal USC CreativeIT database consisting of goal-driven improvised interactions [6]. It contains detailed full body Motion Capture (MoCap) data, providing a rich resource for studying hand gesture during expressive interactions, e.g., for attitudes of friendliness and conflict assumed by interlocutors. We employ an unsupervised clustering method that uses a parallel HMM structure to extract the recurring patterns of the joint gesture of both right and left hands for a participant in an interaction. We then apply a bigram language model to capture the transition structure of the two-handed gesture phrases over an interaction with respect to different cluster numbers. We use normalized perplexity to evaluate the resulting language model for various numbers of hand gesture clusters, and then select a suitable cluster number. We further investigate the use of hand gesture clustering for expressing interaction attitudes by examining the correlation of dyad's hand gesture for interaction types of friendliness and conflict. Our analysis results show that the correlation patterns generally differ across interaction types, which is consistent with our previous finding in [7], establishing the validity and usefulness of the extracted hand gesture phrases. To further validate the derived hand gesture phrases, we build a prediction model employing Support Vector Machine (SVM) to classify an individual's interaction attitude as well as the interaction type as friendly or conflictive, using only the dynamics of the hand gesture phrases over an interaction. We report promising experimental results, which are demonstrating the expressiveness of hand gesture for conveying interaction attitudes and emotions. This direction could be further exploited for attitude-driven hand gesture synthesis for virtual agents.

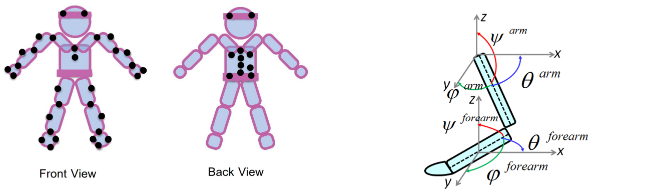
2. RELATED WORK

Attitude or emotion recognition from multimodal cues has been widely studied in recent years. Somasundaran *et al.* exploited attitude information with respect to questions and answers in online discussions and news [8]. Relationships between members in social networks, such as friendship and antagonism, have been studied and modeled in [9]. In addition to text, speech, facial expressions as well as body language are important indicators of attitudes or emotions. Researchers investigated the use of speech prosody features for detecting negative emotions at the utterance level in human-computer interactions [10] [11]. Emotion classification from speech and facial cues was exploited in [12]. More recently, Metallinou *et al.* have applied speech and body language information to track the changes in continuous emotions over an interaction [5]. Bernhardt and Robinson have attempted to detect affective information in the knocking motion from derived motion primitives [13]. However, their work is only specific to the simple scenario (knocking). In this work, we aim at exploring the attitude content in the hand gesture dynamics at the phrase (primitive) level over interpersonal interactions. To this end, we identify hand gesture phrases in a data-driven manner and apply the evolving gesture dynamics during dyadic communication to recognize attitudes at the interaction level.

There is also an extensive literature concentrating on learning primitives of human actions from motion capture data. Levine *et al.* derived gesture subunits from motion data by detecting zero points of the angular velocity [14]. A probabilistic PCA based algorithm has been proposed in [15] to segment motion data into distinct actions assuming that a motion transition occurs when the distribution of motion data changes. Zhou *et al.* proposed an unsupervised hierarchical framework combining kernel k-means and generalized dynamic time alignment kernel, for temporally segmenting and clustering multidimensional time series [16]. Despite the good performance, the computational complexity of this framework limits its applicability to relatively short time series. In this work, we employ a more efficient and flexible clustering model using the parallel HMM structure described in [17] to identify recurring patterns of hand gesture. This model automatically partitions gesture streams into segments each of which is assigned to one of M clusters by maximizing the likelihood through Viterbi decoding.

3. DATABASE DESCRIPTION

We use the USC CreativeIT database in this work, which is a multimodal database of dyadic theatrical improvisations [6]. It contains detailed full body Motion Capture (MoCap) data of participants during dyadic interactions, as shown in Fig. 1(a). The motion capture process retrieves the $3D$ coordinates of the markers in Fig. 1(a) at 60 fps. We manually mapped the $3D$ locations to joint angles of different body parts using MotionBuilder [18]. Fig. 1(b) illustrates the Euler angles of the arm and forearm in the x, y and z directions. The joint angles of left arm, left forearm, right arm and right forearm will be used as gesture features for extracting hand gesture phrases. The joint angles are preferred instead of MoCap $3D$ coordinates, because they are more suitable for animation purposes [14] [17].



(a) Motion Capture Markers. (b) Joint angles for the hand.

Fig. 1. (a) The positions of the Motion Capture markers; (b) The illustration of joint angles for the hand.

The interactions performed by the pairs of actors are either improvisations of scenes from theatrical plays or theatrical exercises where actors repeat sentences to express interaction goals featuring specific interaction stances or attitudes. The interactions were guided by a theater expert (professor/director), and were performed following the Active Analysis improvisation technique pioneered by Stanislavsky [19]. According to this technique, the interactions are goal-driven; actors have predefined goals, e.g., to comfort or to avoid, that they try to achieve through the appropriate use of body language and speech prosody. The goal pair of each dyad defines the attitudes of the interlocutors towards each other and the content of the interaction. As defined by the goals, the attitudes of interacting participants can be naturally grouped into classes of friendliness and conflict. Accordingly, different combinations of attitudes of interacting partners lead to three interaction types: friendliness, medium conflict and high conflict. In friendly interactions both participants have friendly attitudes; in incongruent "medium" conflict interactions one participant is friendly while the other is creating conflict; and in high conflict interactions both participants have conflictive attitudes. The interaction grouping is described in Table 1, along with examples of characteristic goal pairs. Friendliness, medium and high conflict interaction groups contain 12, 26 and 8 interactions respectively, performed by 16 distinct actors (9 female). Thereby, there are 50 friendly and 42 conflictive individuals. The average interaction length is 3.5 minutes.

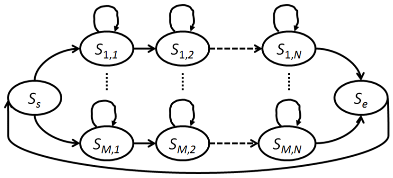


Fig. 2. The parallel HMMs for capturing gesture phrases.

Table 1. Friendly, medium and high conflict interaction types

Interaction Types	Pairs of actors' attitudes	Example goal pairs
Friendly	friendly - friendly	to make peace - to comfort
Medium Conflict	friendly - conflictive	to convince - to reject
High Conflict	conflictive - conflictive	to accuse - to fight back

4. HAND GESTURE CLUSTERING

The elementary gesture patterns are not well established quantitatively, due to the nature of the gesture structure, i.e., the variability across persons and variations in temporal scales. In this work, we identify the recurring patterns of hand gesture in an unsupervised manner. We employ the parallel HMM model in [17] to extract elementary phrases of the joint gesture of both right and left hands, i.e., two-handed gesture phrases, for a participant in an interaction. This model provides flexibility in modeling the variations in the structure and durations of hand gesture phrases. As described in Section 3, the gesture features include the joint angles of the four hand joints, along with their $1st$ order derivatives. The gesture feature vector $\mathbf{f}_k^n$ of the joint n at frame k is:

$$\mathbf{f}_k^n = [\theta_k^n, \phi_k^n, \psi_k^n, \Delta\theta_k^n, \Delta\phi_k^n, \Delta\psi_k^n], n = 1 \cdots 4, \tag{1}$$

where θ_k^n, ϕ_k^n and ψ_k^n are Euler angles of the joint n respectively in the x, y and z directions (see Fig. 1(b)), and $\Delta\theta_k^n$, $\Delta\phi_k^n$ and $\Delta\psi_k^n$ are their corresponding $1st$ order derivatives. Then the gesture feature vector for the four hand joints is: $\mathbf{f}_k = [\mathbf{f}_k^1, \mathbf{f}_k^2, \mathbf{f}_k^3, \mathbf{f}_k^4]$.

The parallel HMM model $\mathbf{\Lambda}$ is composed of M parallel left-to-right HMMs $\{\boldsymbol{\lambda}_i\}_{i=1}^M$, where each branch $\boldsymbol{\lambda_i}$ has N states, as shown in Fig. 2. Here M corresponds to the number of clusters. The feature stream of hand gesture $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_T\}$ is used to train the HMM model $\mathbf{\Lambda}$, where T is the length of the feature sequence. The unsupervised process performs segmentation and clustering by maximizing the likelihood using Viterbi decoding:

$$\{\epsilon_l, m_l\}_{l=1}^L = \arg \max_{\{\epsilon_l, m_l\}} \mathbf{\Pi}_{l=1}^L P(\epsilon_l | \lambda_{m_l}), \tag{2}$$

where $\{\epsilon_1, \epsilon_2, \cdots, \epsilon_L\}$ are the L number of phrase segments of hand gesture produced by the model $\mathbf{\Lambda}$, and each phrase segment ϵ_l is assigned to one of the M clusters with label m_l. As a result, the original feature sequence of hand gesture has been transformed into a sequence of cluster labels. These sequences of labels will be used for the analysis and experiments that follow.

5. CLUSTERING ANALYSIS

In this section, we first apply bigram language models to capture the evolution of two-handed gesture of a participant over an interaction with respect to different cluster numbers. A bigram model is a first-order Markov model, popular in modeling word sequences (here sequences of hand gesture phrases) in language processing. We use normalized perplexity to evaluate each language model, and determine an appropriate cluster number. Next, we investigate the validity of the derived hand gesture phrases by examining the correlation of dyad's hand gesture for different interaction types.

5.1. Bigram Hand Gesture Modeling

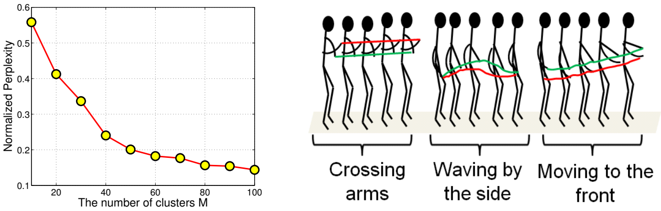
We use the sequences of hand gesture labels to calculate the transition (bigram) probabilities of hand gesture phrases over an interaction. Our objective here is to identify a suitable number of clusters which corresponds to a high-quality bigram language model. Perplexity is a popular way to evaluate language models [20]. This measure quantifies the confusion of the current gesture phrase, i.e., the average number of possible successors, in an information theoretic way. A lower perplexity indicates a better language model. The perplexity ppl is defined as:

$$ppl = P(S)^{-\frac{1}{|S|}}, \tag{3}$$

where S is a sequence with $|S|$ hand gesture phrases. The probability $P(S)$ is computed using the bigram model as: $P(S) = P(g_1)\mathbf{\Pi}_{i=2}^{|S|}P(g_i|g_{i-1})$, where $P(g_i|g_{i-1})$ is the bigram probability that the gesture g_i occurs if the previous gesture g_{i-1} has been observed. However, this measure depends on the vocabulary size (the number of clusters M), i.e., a larger M leads to a higher perplexity. To alleviate this dependency, the normalized perplexity $\overline{ppl}$ is applied in [21] by taking the ratio of ppl and M: $\overline{ppl} = \frac{ppl}{M}$.

Our clustering method is performed with the number of clusters (M) ranging from 10 to 100, and a bigram language model is learned with respect to each cluster number. The normalized perplexity is adopted to evaluate each language model. Fig. 3(a) shows the normalized perplexity of a bigram model as a function of the number of clusters. Overall, we can observe that the $\overline{ppl}$ decreases as the number of clusters increases. Specifically, the $\overline{ppl}$ drops rapidly with increasing number of clusters initially, then the decrease of the normalized perplexity slows down for cluster numbers of around 50 or above. This result suggests that the transition dynamics of hand gesture phrases can be adequately captured by the bigram language model computed using 50 clusters. Higher cluster numbers bring only minor variations to the computed structure, while greatly increasing the computational cost. In the analysis and experiments that follow, we fix the number of clusters at 50 accordingly.

To better understand the semantic meaning of the clusters, we visualize three examples of hand gesture phrases each respectively from one of three distinct clusters, as shown in Fig. 3(b). The red and green lines are the moving trajectories of right and left hands respectively. In the first gesture phrase which shows a person crossing arms, both hand trajectories are constant, indicating a static hand gesture. In contrast, the other two gesture phrases have both hand trajectories with temporal dynamic changes, representing dynamic hand movements. However, the right and left hand trajectories in the waving gesture phrase vary distinctly, whereas both hands move symmetrically in the third gesture phrase. In addition, the two dynamic hand gesture phrases are distinct in the aspect of hand locations, i.e., moving by the side and moving in the front.



(a) Normalized perplexity. (b) Example hand gesture phrases.

Fig. 3. (a) Normalized perplexity of a bigram model varying with the number of clusters; (b) Example hand gesture phrases from three distinct clusters.

5.2. Correlation of Dyad's Hand Gesture

The coordination between human behavior in an interaction has been studied in diverse areas [22]. In our previous work [7], we empirically verified that the coordination patterns between dyad's body language differ depending on the attitudes of the two interlocutors. For example, people with friendly attitudes may tend to adapt more to the behavior of their interlocutors, resulting in a higher level of behavior correlation along an interaction [23]. Herein, we examine the dyad's correlation using the derived hand gesture phrases for different interaction types, to establish the validity of the gesture phrases.

Given an interaction, the hand gesture sequences of two interaction participants are transformed into two parallel sequences of cluster labels. To compute the dyad's correlation at frame t, we set a window (5 sec) centered at frame t respectively for each of the two sequences, count the histogram of cluster labels over the window, and compute the Pearson's correlation between the histograms of the two participants. When shifting the window along the interaction, we get a correlation curve for the entire interaction. This process is illustrated in Fig. 4(a). Similar to [24] showing that interlocutors tend to mimic the gesture of each other, the histograms of the two interlocutors as computed in our work could measure their similar hand gesture patterns over the window. Hence the correlation between gesture histograms could be used to define the similarity of the dyad's behavior, i.e., their behavior coordination.

To represent the average correlation pattern of dyad's hand gesture for each interaction type, we apply dynamic time wrapping (DTW) among the correlation curves of interactions within the interaction group of friendly, medium conflict or high conflict, and average the wrapped curves. Fig. 4(b) presents sample mean correlation curves between dyad's hand gesture for different interaction types. Overall, we can observe that the temporal dynamics of the correlation curve in the friendly interaction are quite distinct from those in the medium or high conflict ones. In particular, there is a generally higher correlation (above 0.5) along the friendly interaction whereas a lower correlation (below 0.2) is observed in the medium and high conflict ones. In addition, we take the mean value of a correlation curve to represent the correlation level of dyad's hand gesture in an interaction. Through Student's t-test, we find that the mean correlation of friendly interactions is significantly higher than that of the medium and high conflict ones ($p < 0.05$). Thus, the coordination dynamics between dyad's hand gesture differ depending on whether the interaction is friendly or (medium or high) conflictive, which is consistent with our previous finding in [7]. This empirically validates the usefulness of the extracted hand gesture phrases. The correlation observed in the high conflict case where both participants also have the same attitudes is lower than the one observed in

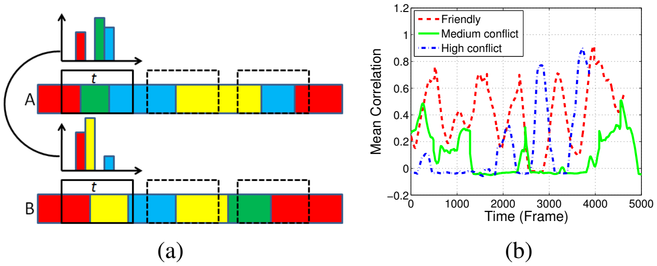


Fig. 4. (a) Illustration of dyad's correlation computation. The cluster labels of gesture phrase segments are represented by different colors; (b) Sample mean correlation curves for friendly, medium conflict and high conflict interactions.

the friendly case. This may be because people with conflictive attitudes are inherently more self-initiated, possibly reducing the dyad's behavior synchrony along interactions.

6. RECOGNIZING ATTITUDES FROM HAND GESTURE

Section 5 focused on the analysis of the automatically derived hand gesture phrases in order to establish their usefulness and validity. Herein we aim at applying the features representing the dynamics of hand gesture phrases for classifying an individual's interaction attitude, as well as the type of an interaction, i.e., friendly, medium conflict and high conflict. Both classification experiments are performed using SVM with an RBF kernel. The leave-one-sample-out scheme is adopted, where "sample" denotes an individual (Section 6.1) or an interaction (Section 6.2).

6.1. Classification of Individual's Interaction Attitudes

In this experiment, we classify an individual's interaction attitude, viz., friendly (50 samples) or conflictive (42 samples). To capture the dynamics of hand gesture flow, we compute the unigrams and bigrams from the sequence of cluster labels (see Section 4) for each individual. Both unigram and bigram counts are utilized as features for attitude classification. Based on the perplexity analysis in Section 5.1, the cluster number is set to 50, which results in a 2500 dimensional bigram feature vector. We reduce the dimensionality using Principal Component Analysis (PCA) by preserving 90% of the total variance.

6.2. Classification of Interaction Types

In this experiment, our goal is to predict the type of an interaction (see Table 1): friendly (12 samples), medium conflict (26 samples) or high conflict (8 samples). Similarly to Section 6.1, for each interaction, we compute the unigram and bigram features from both interlocutors. In addition, we consider the cross-interlocutors bigrams to describe the interaction dynamics between participants, i.e., transitions from gesture g_i^A of participant A to gesture g_{i+1}^B of participant B and vice versa, which we denote as C-bigram. The dimensionalities of bigram (5000-d) and C-bigram (5000-d) features are both reduced by PCA preserving 90% of the total variance.

6.3. Experimental Results

Table 2 presents the classification results for both experiments. The second column of Table 2 shows the results of classifying an individual's attitude using unigram and bigram features as well as their combination. We obtain an accuracy of 87.8%, an improvement of 33.5% over the chance rate, when using only the unigram features. The performance further increases to 91.3% with the inclusion of the bigram features. The third column of Table 2 shows the experimental results of classifying interaction types. We compare the classification performance using each type of features as well as combinations of different feature types. We can observe that the best accuracy of 89.4% is achieved when utilizing all the three types of dynamic features of hand gesture.

Table 2. Summary of classification results for both experiments.

	Accuracy (%)	
	Attitude classification	Interaction classfication
Chance	54.3	56.5
Unigram	87.8	83.6
Bigram	78.4	73.6
C-bigram	/	75.4
Unigram + Bigram	**91.3**	88.1
Unigram + Bigram + C-bigram	/	**89.4**

For both experiments, the unigram features alone generally have better performance than other types of features, indicating that the holistic context of hand gesture is quite distinct depending on the interaction attitudes. In addition, the improvement from bigram features suggests that the evolving dynamics of hand gesture phrases over an interaction add further discriminative power for classifying interaction attitudes. For the experiment of classifying interaction types, it is interesting to observe the benefits from the C-bigram features that characterize interaction dynamics between dyad's hand gesture, reinforcing our observation that the coordination patterns of dyad's behavior differ depending on interaction types.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we studied the role of hand gesture in conveying attitudes of friendliness or conflict towards the interlocutors, during dyadic interactions. To this end, we first employed a parallel HMM model to extract recurring patterns of hand gesture in a data-driven way. Then, we computed the bigram language model to describe the transition structure of hand gesture phrases with respect to different cluster numbers, and evaluated each bigram to identify an appropriate cluster number. We further investigated the validity of the derived hand gesture phrases by examining the correlation of dyad's hand gesture for different interaction types. The analysis results showed that the correlation patterns differ depending on the interaction types, and that friendly interactions are characterized by greater hand gesture coordination between interlocutors. Finally, we employed an SVM to model and classify interaction attitudes as well as interaction types using the dynamics of hand gesture phrases over an interaction. Experimental results showed that an individual's interaction attitude can be classified with an accuracy of 91.3%, and the interaction type can be classified with an accuracy of 89.4%, suggesting the usefulness of the derived hand gesture phrases for discriminating interaction attitudes.

In the future, our goal is to work towards interaction-driven and attitude-driven hand gesture synthesis based on the extracted hand gesture phrases. Specifically, we would like to synthesize expressive hand gesture, and other body gesture, for virtual agents that can display various attitudes and can respond appropriately to the multimodal cues of the human user.

8. ACKNOWLEDGEMENT

This material is based upon work supported by DARPA and Space and Naval Warfare Systems Center Pacific under contract number N66001-11-C-4006.

9. REFERENCES

[1] J.A. Harrigan, R. Rosenthal, and K.R. Scherer, *The new handbook of Methods in Nonverbal Behavior Research*, Oxford Univ. Press, 2005.

[2] J.P. Wachs, M. Kölsch, H. Stern, and Y. Edan, "Vision-based hand-gesture applications," *Communications of the ACM*, vol. 54, no. 2, pp. 60–71, 2011.

[3] A. Kendon, "Gesticulation and speech: Two aspects of the process of utterance," *The relationship of verbal and nonverbal communication*, vol. 25, pp. 207–227, 1980.

[4] A. Kapur, A. Kapur, V-B. Naznin, G. Tzanetakis, and P. F. Driessen, "Gesture-based affective computing on motion capture data," in *Affective Computing and Intelligent Interaction*, pp. 1–7. Springer, 2005.

[5] A. Metallinou, A. Katsamanis, and S. Narayanan, "Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information," *Image and Vision Computing, Special Issue on Continuous Affect Analysis*, 2012.

[6] A. Metallinou, C.-C. Lee, C. Busso, S. Carnicke, and S. Narayanan, "The USC CreativeIT database: A multimodal database of theatrical improvisation," in *Proc. of Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality (MMC)*, 2010.

[7] Z. Yang, A. Metallinou, and S. Narayanan, "Toward body language generation in dyadic interaction settings from interlocutor multimodal cues," in *Proc. of ICASSP*, 2013.

[8] S. Somasundaran, T. Wilson, J. Wiebe, and V. Stoyanov, "Qa with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news.," in *Proc. of ICWSM*, 2007.

[9] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proc. of international conference on World wide web*, 2010, pp. 641–650.

[10] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke, "Prosody-based automatic detection of annoyance and frustration in human-computer dialog.," in *Proc. of INTERSPEECH*, 2002.

[11] C-M. Lee, S. Narayanan, and R. Pieraccini, "Recognition of negative emotions from the speech signal," in *Proc. of ASRU*, 2001, pp. 240–243.

[12] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C-M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *Proc. of ICMI*, 2004, pp. 205–211.

[13] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions," in *Affective Computing and Intelligent Interaction*, pp. 59–70. Springer, 2007.

[14] S. Levine, C. Theobalt, and V. Koltun, "Real-time prosody-driven synthesis of body language," in *ACM Transactions on Graphics*. ACM, 2009, vol. 28, p. 172.

[15] J. Barbic, A. Safonova, J-Y. Pan, C. Faloutsos, J.K. Hodgins, and N.S. Pollard, "Segmenting motion capture data into distinct behaviors," in *Proc. of Graphics Interface*, 2004, pp. 185–194.

[16] F. Zhou, F. De la Torre, and J. Hodgins, "Hierarchical aligned cluster analysis for temporal clustering of human motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 582–596, 2013.

[17] M.E. Sargin, Y. Yemez, E. Erzin, and A. Tekalp, "Analysis of head gesture and prosody patterns for prosody-driven head-gesture animation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1330–1345, 2008.

[18] Installation Guide, "Autodesk®," 2008.

[19] S. M. Carnicke, *Stanislavsky in Focus: An Acting Master for the Twenty-First Century*, Routledge, UK, 2008.

[20] S. Young, G. Evermann, D. Kershaw, G. Moore, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, "The htk book," *Cambridge University Engineering Department*, vol. 3, pp. 175, 2002.

[21] U-V. Marti and H. Bunke, "On the influence of vocabulary size and language models in unconstrained handwritten text recognition," in *Proc. of ICDAR*, 2001, pp. 260–265.

[22] E. Delaherche, M. Chetouani, A. Mahdhaoui, C Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: A survey of evaluation methods across disciplines," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 349–365, 2012.

[23] P. Ekman, "Body position, facial expression, and verbal behavior during interviews.," *The Journal of Abnormal and Social Psychology*, vol. 68, no. 3, pp. 295, 1964.

[24] I. Kimbara, *Interpersonal influences on gesture production: evidence for gesture form convergence across speakers in dyadic interaction*, Ph.D. thesis, University of Chicago, Department of Linguistics, 2006.