The USC CreativeIT Database: A Multimodal Database of Theatrical Improvisation

Angeliki Metallinou [†], Chi-Chun Lee [†], Carlos Busso [‡], Sharon Carnicke [£], Shrikanth Narayanan [†]

[†] Electrical Engineering Department, University of Southern California, Los Angeles CA 90089,
[‡] Electrical Engineering Department, University of Texas at Dallas, Dallas TX 75080,
[£] School of Theater, University of Southern California, Los Angeles CA 90089
metallin@usc.edu, chiclee@usc.edu, busso@utdallas.edu, carnicke@usc.edu, shri@sipi.usc.edu

Abstract

Improvised acting is a viable technique to study human communication and to shed light into actors' creativity. The USC CreativeIT database provides a novel bridge between the study of theatrical improvisation and human expressive behavior in dyadic interaction. The theoretical design of the database is based on the well-established improvisation technique of Active Analysis in order to provide naturally induced affective, goal-driven interaction. The carefully engineered data collection and annotation processes provide a gateway to quantify and investigate various aspects of theatrical performance and human communication.

1. Introduction

Human interaction is a complex blend of intents, communicative goals and emotions, which are expressed, among others, through body language, prosodic cues, speech content. The study of human communication and expressive behaviors has attracted interest from multiple domains including psychology, social sciences, engineering, theater, etc. This paper describes the design, collection and annotation process of a novel, multimodal and multidisciplinary interactive database, the USC CreativeIT database. The database is a result of the collaborative work between the USC Viterbi School of Engineering and the USC School of Theater. The database is collected using cameras, microphones and motion capture and contains detailed audiovisual information of the actors' body language and speech cues. It serves two purposes. First, it provides insights into the creative and cognitive processes of actors during theatrical improvisation. Second, the database offers a well-designed and well-controlled opportunity to study expressive behaviors and natural human interaction.

The significance of studying creativity in theater performance is that improvisation is a form of real-time dynamic problem solving (Mendonca and Wallace, 2007). Improvisation is a creative group performance where actors collaborate and coordinate in real time to create a coherent viewing experience (Johnstone, 1981). Improvisation may include diverse methodologies with variable levels of rules, constraints and prior knowledge, concerning the script and the actor's activities. Active Analysis, introduced by Stanislavsky, proposes a goal-driven performance to elicit natural affective behaviors and interaction (Carnicke, 2008), and is the primary acting technique utilitized in the database. It provides a systematic way to investigate the creative processes that underlie improvisation in theater. The role of acting has been considered as a viable research methodology for studying human emotions and communication. Theater has been suggested as a model for believable agents; agents that may display emotions, intents and human behavioral qualities (Perlin and A.Goldberg, 1996). Researchers have advocated the use of improvisation as a tool for eliciting naturalistic affective behavior for studying emotions and argue that improvised performances resemble real-life decision making (Fig. 1, (Busso and Narayanan, 2008)). Furthermore, it has been suggested that experienced actors, engaged in roles during dramatic interaction may provide a more natural representation of emotions, avoiding exaggeration or caricatures (Douglas-Cowie et al., 2003).

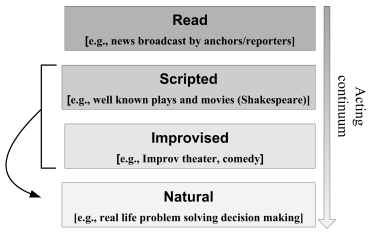


Figure 1: Acting continuum: From fully predetermined to fully undetermined (Busso and Narayanan, 2008)

A variety of acted emotional/behavioral databases exist in the literature. As argued in (Enos and Hirschberg, 2006) valuable emotional databases can be recorded from actors using theatrical techniques. Examples of databases which explore acting techniques include the audiovisual IEMOCAP database (Busso et al., 2008), which contains improvised and scripted acting, and the speech Genova Multimodal Emotion Portrayal (GEMEP) database (Banziger and Scherer, 2007). In (Anolli et al., 2005), authors describe the collection of a multimodal database where contextualized acting is used.

The USC CreativeIT database is a novel, multimodal database that is distinct and complemets most of the existing ones. Its theoretical design is based on the well-established theatrical improvisation technique of Active Analysis and results from a close collaboration of theater experts, actors and engineers. We utilize Motion Capture technology to obtain detailed body language information of the actors, in addition to microphones, video and carefully designed post-performance interviews of the participants. Annotation of the data includes continuous emo-

This work was supported in part by funds from NSF

tional descriptors (valence, activation) as well as theatrical performance ratings (naturalness, creativity) from various perspectives (e.g actor, expert, observer). The database aims to facilitate the study of creative theatrical improvisation qualitatively and provides a valuable source to study human-human communicative interaction.

The rest of this paper is organized as follows. Section 2 describes the theatrical methodology and design hypotheses, section 3 contains the experimental protocol and the technological equipment and section 4 describes the data annotation process. Finally, section 5 contains discussion of future research directions.

2. Theatrical Methodology

2.1. Active Analysis

In Active Analysis, the actors play conflicting forces that jointly interact. The balance of the forces determines the direction of the play. The scripts used in the case play the role of guiding the events (skeleton). The course of the play can be close to or different from the script. This degree of freedom provides an flexibity to work at different levels in the improvisation spectrum. A key element in Active Analysis is that actors are asked to keep a verb in their mind, while they are acting, which drives their actions. As a result, the interaction and behavior of the actors may be more expressive and closer to natural, which is crucial in the context of emotion recognition. For instance, if the play suggests a confrontation between two actors, one of them may choose the verb *inquire* while the other may choose *evade*. If the verbs are changed (e.g.*persuade, confront*) the play will have a different development. By changing the verbs, the intensity of the play can be modified as well (i.e.ask versus interrogate). As a result, different manifestations of communication goals, emotions and non-verbal behaviors can be elicited through the course of the interaction. This flexibility allows us to explore the improvisation spectrum at different levels and makes Active Analysis a suitable technique to elicit emotional manifestations.

2.2. Design of Data Collection

The USC CreativeIT database utilizes two different theatrical techniques, the two-sentence exercise and the paraphrase, both of which orginate from the Active Analysis methodology. We also perform a post-performance survey after the recording.

In the two sentence exercise, each actor is restricted to saying one given sentence with a given verb. For example, one actor may say "Marry Me" with verb *confront*, and another one may say "I'll think about it" with verb *deflect*. Given the lexical contraint, the expressive behaviors and the flow of the play will be primarily based on the prosodic and non-verbal behaviors of the actors. This type of controlled interaction can bring insights into how human/actors use their expressive behaviors, such as body language and prosody, to reach a communication goal. Also, this approach is suitable to study emotion modulation at a semantic level, since the same sentences are repeated different times with different emotional connotation.

In the paraphrase, the actors are asked to act out a given script with their own words and interpretation. Examples of plays that are used are "The Proposal" by Chekhov or "Taming of the Shrew" by Shakespeare. In this set of recordings, actors are not lexically constrained. As a result, the performance is characterized by a more natural and free-flow interaction between the actors which bears more ressemblance to real-life scenarios, compared to the two-sentence exercise. Therefore, behavioral analysis and findings on such sessions could possibly be extrapolated to natural human interaction and communication.

Finally we perform a brief interview of the actors right after each performance. Examples of the questions asked are 'What verbs did you and the other actor use?', 'What was the goal of your character?', 'How would you describe your and the other actor's emotion during the interaction?'. These questions are designed to help understand the cognitive planning process of the actors as they improvise on the scenes.

3. Data Collection

3.1. Session Protocol

An expert on Active Analysis (the 4th author of the paper) directed the actors during the rehearsal and the recording of the sessions. Prior to the scheduled data collection date, the actors had to go through a rehearsal with the director to become familiar with active analysis and the scene. Just before the recording of the paraphrase, there was another 5-minute session to refresh actors' memory and give the director a chance to remind actors of the essence of the script. A snapshot of an actor during the data collection is shown in Figure 2(a)

The data collection protocol consists of the following steps:

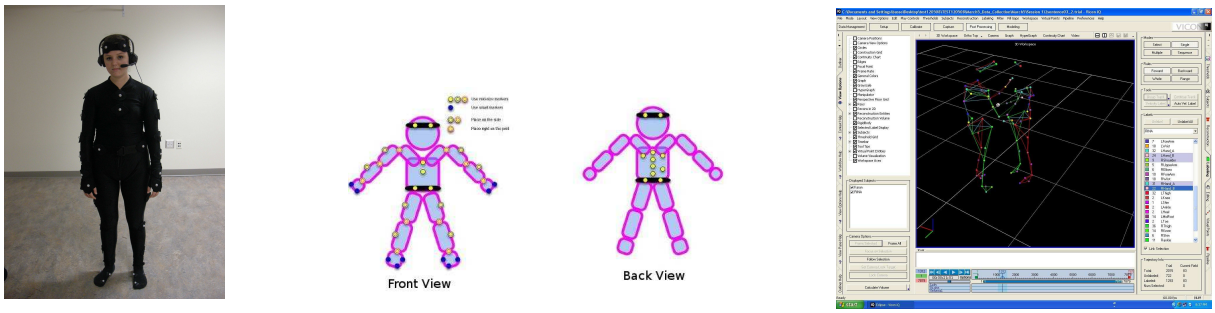
1. Two-Sentence Exercise (unknown verbs)
2. Two-Sentence Exercise, using the same sentences as previously but different verbs (known verbs)
3. Paraphrase of Script (known verbs)
4. Paraphrase of Script, using the same script as previously but different verbs (known verbs)
5. Two-sentence Exercise (unknown verbs)
6. Two-sentence Exercise, using the same sentences as previously but different verbs (known verbs)

Verbs are chosen either by the actors or the director prior to each performance. Some of the commonly chosen verbs are *to shut him out, to seduce, to deflect, to confront, to force the issue* etc, which introduce a large variety of communication goals. *Unknown verbs* indicate that actors are not aware of each other's verb prior to the performance. This setting provides a variety in the interaction dynamic of the two-sentence exercise. During the paraphrases the actor's verbs are always known to each other in advance.

3.2. Equipment and Technical Details

The following is the list of equipment that is utilized in the data collection:

- **Vicon Motion Capture System:** 12 motion capture cameras to record 45 marker's (x, y, z) position for each actor. The markers are placed according to Figure 2(b).
- **HD Sony Video Camcorder:** 2 Full HD cameras are places at each corner of the room to capture the performance of the actors.
- **Microphones:** Each actor has a close-up microphone to record actors' speech at 48KHz with 24 bits.



(a) Actor wearing microphone and markers

(b) Positions of markers

(c) Snapshot of the motion capture post-processing software

Figure 2: Snapshots of an actor during data collection, the marker positions and the post-processing software

3.3. Motion Capture Post-Processing

The first data post processing step is to map each of the markers captured into a subjects defined body model. There are two subjects each with 45 markers, and also there are about 5000 - 10000 frames per interaction session. Since actors are asked to be expressive with body language and gesture, occulusion of markers happens fairly often. Because of this, the computer software is unable to perform all the labeling automatically and accurately. For example, when two subjects are close to one another, one's hand marker may be labeled as the other person's shoulder if we rely on computer labeling. In order to obtain reliable and detailed marker information, the motion capture data was manually corrected frame by frame. The spline function was used to interpolate any missing markers. Such post-processing of one actor in one performance may require approximately 1 - 2 hours, which is a fairly time consuming task. Figure 2(c) shows a snapshot of the post-processing software.

3.4. Data Collection and Annotation Progress

The database contains the recording of nine full sessions, each of which contains approximately one hour of audiovisual data. In total we have recorded 40 two-sentence exercises and 19 paraphrases with 19 actors. The motion capture post-processing step is approximately 95% complete. One full session has been annotated by five different annotators.

4. Data Annotation

4.1. Annotation Attributes and Annotator Groups

The design of the annotation process depends on the collected data as well as the research scope of the database. During improvised dyadic sessions there is a continuous flow of body language and dialog and a diverse expression of emotions and intentions. In order to preserve this flow, we annotate the sessions using continuous labels instead of chopping them into sentences or other arbitrary chunks. Furthermore, since commonly used categorical emotional attributes (angry, happy etc) may not be applicable or sufficient for our data, we choose a more comprehensive set of attributes. These contain dimensional emotional descriptors (valence, activation, dominance) as well as theater performance ratings (interest, naturalness), which may facilitate future theatrical performance analysis.

The attributes that are annotated for each session are described in Table 1. For each attribute, it is mentioned whether the annotation is continuous or a discrete label per session, or both, and if the attribute is annotated per actor or per session as a whole.

All continuous annotations are performed by watching the session videos and using the Feeltrace software. Feeltrace is a publicly available emotional annotation tool, described in (Cowie et al., 2000), which we slightly modified to suit our purposes. The Feeltrace interface enables the user to continuously move the mouse along the computer screen so as to indicate the attribute value, ranging from -1 to 1. For the discrete annotations, annotators are asked to provide a label ranging from 1 to 5.

The annotated attributes are, to a large extent, subjective. In addition to using multiple annotators for the same videos, we are also interested in examining how diverse audience groups may perceive and rate a video, according to their expertise. We categorize the annotators into three groups; theater experts, actors and naive audience. The first group consists of professors of the USC theater school and experts in active analysis while the second consists of students of the theater school, who may or may not have performed in the session. Finally, the naive audience(observer) group consists of USC students who have no technical knowledge of theater.

4.2. Multiple Annotator Correlations

In order to examine the correlations between different annotators for a certain attribute, we performed statistical analysis of the annotations of one two-sentence exercise recording. An example is shown in Figure 3, where we present a segment of the annotation of the activation of an actor, annotated by 5 people; 3 students (naive audience) and 2 actors. Although various annotations differ, the correlations between them are evident. In order to examine linear relationships between the annotations, we computed the Pearson correlation coefficients between all pairs of annotations, which are all found significant at the 0.01 level (2-tailed). We also investigate the prediction success of an annotation using linear regression with the rest of the annotations as predictors. In Table 2, for each of the 5 cases, we present the adjusted R^2 as a measure of the goodness of fit of the linear regression model. The relatively large numbers of R^2 indicate that an annotation can be well-predicted using the rest annotations, suggesting linear relationships

Attribute	Definition	Type	Rating
Continuous Emotional Descriptors			
Valence	Positive vs Negative	continuous	per actor
Activation	Excited vs Calm	continuous	per actor
Dominance	Dominant vs Submissive	continuous	per actor
Theatrical Performance Ratings			
Interest	How interesting do you find the session	continuous and discrete	per session
Naturalness	How natural do you find the performance	continuous and discrete	per actor
Creativity	How creative,in terms of novelty, do you find the performance	discrete	per actor
Actor Verbs	How successful are actors in performing their verbs	discrete	per actor

Table 1: Annotated attributes

between the different annotations.

Furthermore, we computed the Pearson correlation coefficients between all pairs of annotations for all continuous emotional descriptors of the two-sentence exercise (activation, valence, dominance). They were all found statistically significant at the 0.01 level (2-tailed) and positive, which indicates consistency between different annotators despite their possibly different styles. All statistical tests were performed using the SPSS software.

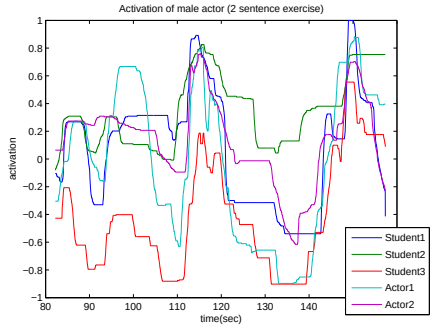


Figure 3: Continuous annotation of the activation of one actor in a two-sentence exercise recording

Linear Regression Fit		
Dependent	Predictors	adjusted R^2
student1	students:2,3, actors:1,2	0.731
student2	students:1,3, actors:1,2	0.561
student3	students:1,2, actors:1,2	0.501
actor1	students:1,2,3 actor:2	0.530
actor2	students:1,2,3 actor:1	0.650

Table 2: Adjusted R^2 values of linear regression for the annotations of the activation of an actor for a two-sentence exercise

5. Future Research Directions

The USC CreativeIT database is a novel, multimodal and multidisciplinary database which represents a unique opportunity to marry engineering methods with the theory and practice of acting. Future research directions that could be pursued using these data include:

- Analysis of prosody and nonverbal behaviors of the actors, such as facial expression and body language. Investigation of how these behaviors are affected by the communication goal, which is specified by the improvisation verb.
- Analysis of the interaction flow and possible synchronization patterns between the actors during the performance, in relation to the pair of improvisation verbs used.
- Analysis of the theatrical performance ratings and possible differences in ratings among the different evaluator groups. Investigation of the body language and expressive choices that may lead to higher overall performance ratings.
- Application of the insights gained from the database analysis to the design of affect-sensitive believable agents.

6. References

L. Anolli, F. Mantovani, M. Mortillaro, A. Vescovo, A. Agliati, L. Confalonieri, O.Realdon, V. Zurloni, and A. Sacchi. 2005. A multimodal database as a background for emotional synthesis, recognition and training in e-learning systems. In *ACII 2005, Beijing*.

T. Banziger and K. R. Scherer. 2007. Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. In *Int'l Conference on Affective Computing and Intelligent Interaction (ACII)*.

C. Busso and S. Narayanan. 2008. Recording audio-visual emotional databases from actors: a closer look. In *In Language Resources and Evaluation (LREC 2008), Marrakech, Morocco*, pages 17–22, May.

C. Busso, M. Bulut, C-C Lee, A.Kazemzadeh, E. Mower, S. Kim, J. Chang, S.Lee, and S.Narayanan. 2008. IEMOCAP: interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42:335–359.

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

R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schroeder. 2000. Feeltrace: An instrument for recording perceived emotion in real time.

E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach. 2003. Emotional speech: Towards a new generation of databases. *Speech Communication*, 40:33–6, April.

F. Enos and J. Hirschberg. 2006. A framework for eliciting emotional speech: Capitalizing on the actor's process. In *LREC Workshop on Corpora for Research on Emotion and Affect, Genova, Italy.*

K. Johnstone. 1981. *Impro: Improvisation and the Theatre.* Routledge / Theatre Arts, New York.

D. Mendonca and W. Wallace. 2007. A cognitive model of improvisation in emergency management. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, 37(4):547–561.

K. Perlin and A.Goldberg. 1996. Improv: A system for scripting interactive actors in virtual worlds. In *Proceedings of the 23rd Annual Conference on Computer Graphics.*