Processing Negative Emotions Through Social Communication: Multimodal Database Construction and Analysis

Nurul Lubis, Michael Heck, Sakriani Sakti, Koichiro Yoshino, Satoshi Nakamura
Augmented Human Communication Laboratory
Graduate School of Information Science
Nara Institute of Science and Technology
{nurul.lubis.na4, michael-h, ssakti, koichiro, s-nakamura}@is.naist.jp

***Abstract*—Emotion-rich data is pre-requisite in the efforts of transferring emotional aspects of human communication into Human-Computer Interaction (HCI). An important facet of human social-affective interaction is its ability to facilitate social sharing of emotion, a fundamental part of the emotional processes. When conducted properly, such an interaction can give a positive effect to emotion-related problems. However, there is still a lack of resources that are: 1) explicitly designed for studying the emotional problems commonly encountered in everyday life, and 2) involving a professional as an expert in the conversation. In this paper, we present recordings of dyadic social-affective interactions between a professional counselor as an expert and 30 participants, summing up to 23 hours and 41 minutes of material. In each interaction, a negative emotion inducer is shown to the dyad, and the goal of the expert is to aid emotion processing and elicit a positive emotional change through the interaction. Specifically, we aim to observe how an external party can guide and facilitate emotion processing, especially after a negative emotional response in a commonly encountered social situation. The construction, development, and analysis of the database is detailed in this paper.**

1. Introduction

A number of studies have showed a consistent inclination of humans to talk about and socially share their emotional experiences, especially for an intense and/or negative emotion exposure [1]. This is argued to be an essential part of the emotional processes [2]. Social-affective interactions can provide social support, giving positive effect to the treatment of emotion-related problems [3], [4]. On the contrary, the absence of such support in times of need can result in more serious, longer-term consequences, such as withdrawal from social life, difficulties to form meaningful relationships, and sky-rocketing stress levels. As humans impose the emotional aspect of social communications in their interaction with computers and machines [5], an emotionally-competent computer agent could be a valuable assistive technology in addressing this need.

To provide support for healthy users, there exist technologies developed towards a listening oriented system [6] and a companion conversational agent [7]. However, these studies have not yet considered negative emotional experiences and the recovery from them. On the other hand, there exists efforts in addressing a number of clinical emotional disturbances, such as depression and suicide risk evaluation with speech processing techniques [8], and a distress clues assessments through a simulated interviewer agent [9]. Unfortunately, these works are not applicable for the larger, general audiences as they are focusing on clinical circumstances.

To the best of our knowledge, existing studies have not yet examined negative emotional exposures commonly encountered in everyday life, such as those in situations that may elicit a negative emotional response. For example, reading the news, or having a debate on social issues. As discussed previously, prompt recovery of such experience will prevent its accumulation into a more serious emotional problem.

In this paper, we present an emotion-rich corpus containing recordings of dyadic social-affective interactions involving a professional counselor, posing as an expert. We design the corpus to highlight how an external party can guide and facilitate emotion processing through an interaction, especially after a negative emotional experience. To clarify, the corpus is not designed for situations involving clinical emotional problems, such as emotional distress, depression, and anxiety. Instead, we attempt to capture emotional support in everyday situations that may trigger negative emotions. This emotion-rich resource has potential for research efforts pursuing emotionally intelligent Human-Computer Interaction (HCI) to improve user's emotional state. The construction, development, and analysis of the database is detailed in this paper.

2. Related Works

Emotion-rich data is pre-requisite for incorporating emotion in HCI. The majority of existing corpora are constructed for the purpose of recognizing emotion-related phenomena in humans, such as facial expression [10], physiological signals [11], emotion perception [12], and physical motions or gestures [13]. Even though these corpora contain important information for understanding emotion, their scope does not reach the dynamics of emotion at the interpersonal level.

978-1-5386-0563-9/17/$31.00

There exist a handful of social affective corpora, each with its own focus and scope. An example is the Vera am Mittag corpus, containing recordings of spontaneous emotional conversation from television talk shows [14]. Talk show recordings are suitable for affective computing research as they provide natural emotion occurrences in a typical social setting [15]. Despite this quality, an explicit conversation goal that could provide emotional benefits for the participants is still missing from such data.

The SEMAINE Database [16] consists of interactions between a user and a wizard Sensitive Artificial Listeners (SALs) with four distinct personalities, each tends to elicit a certain emotion in the user. Even though this scenario induces fluctuations of emotion in the conversation, it does not reflect human social interaction – it instead reflects user behavior when interacting with a conversational agent.

On the other hand, the Distress Analysis Interview Corpus (DAIC) contains clinical interviews designed for the development of an automated agent for psychological diagnoses [17]. It includes interviews with distressed and non-distressed participants and highlights verbal and non-verbal dialogue actions. However, the corpus does not provide emotion annotation nor employ an expert as the interviewer of the distressed participant.

Even though various conversational scenarios have been considered, there is still a lack of resources that show common emotional problems in a everyday social setting. Furthermore, a great majority of existing corpora does not involve any professional who is an expert in handling emotional reactions in a conversation. Knowledge from such situations is highly potential in constructing assistive technology for emotion-related problems in everyday situations.

To fill these gaps, we design the presented corpus to 1) contain recordings of dyadic spontaneous social-affective interactions before and after a negative emotion exposure, and 2) involve a professional counselor as an expert in the conversation. In each interaction, a negative emotion inducer is shown to the dyad, and the goal of the expert is to aid emotion processing and elicit a positive emotional change through the interaction. This allows the observation of emotion fluctuation in a conversation, and how an external party can guide and facilitate emotion processing through an interaction.

3. Emotion Definition

In this work, we define the emotion scope based on the *circumplex model of affect* [18]. Two dimensions of emotion are defined: *valence* and *arousal*. Valence measures the positivity or negativity of emotion; e.g., the feeling of joy is indicated by positive valence while fear is negative. On the other hand, arousal measures the activity of emotion; e.g., depression is low in arousal (passive), while rage is high (active). Figure 1 illustrates the valence-arousal dimension with respect to a number of common emotion terms. This model is intuitive and easily adaptable to either discrete or other dimensional emotion definitions.

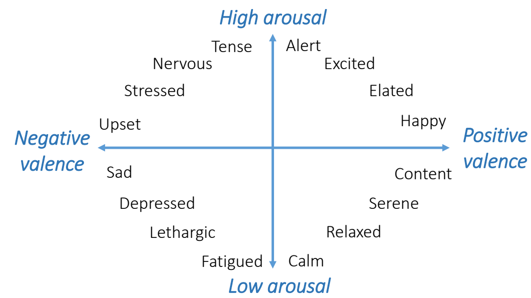


Figure 1. Emotion dimensions and common terms.

4. Database Design

4.1. Recording Scenario

We focus on capturing social support in everyday situations that may trigger a negative emotion, such as reading the news or debating on a social issue. Specifically, we would like to observe how an external party can guide and facilitate emotion processing through an interaction after a negative emotional response. Thus, we arrange for the dyad to consist of an *expert* and a *participant*, each with a distinct role. The *expert* plays the part of the external party who helps facilitate the emotional response of the *participant*.

We design the recording scenario as follows. The session starts with an opening talk as a neutral baseline conversation. Afterwards, we induce negative emotion by showing an emotion inducer to the dyad. The recording then continues with a discussion phase that targets at emotional processing and recovery. In this phase, the expert is given the objective to facilitate the emotional process that follows the emotion induction, and to elicit a positive emotional change through the conversation.

Throughout the process, we ask the participants to assess their emotional states with the use of a questionnaire. In particular, these assessments are collected after briefing, before the emotion inducer, after the emotion inducer, and after the discussion. This allows us to keep track of their emotional states as they occur before and/or after the moments where fluctuations are expected. After the recording, the participants are asked to fill out a post-recording questionnaire to rate their experience of the interaction. Figure 2 illustrates the design of a recording session.

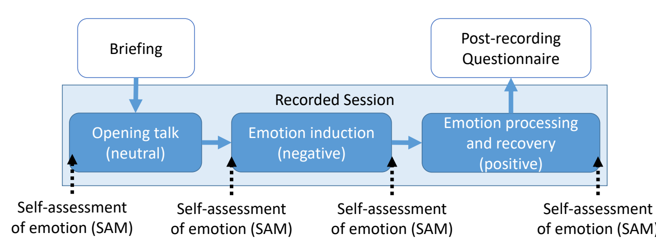


Figure 2. The flow of a recording session.

To approximate an everyday social situation, we focus on various social issues as the conversation topic, such as politics and environmental issues. More specifically, we target issues where opinions with negative sentiment might arise. We intended for the initial emotion to be generated by an external factor to promote openness in the discussion. Furthermore, a more personal topic is likely to take longer to process and recover from, which is undesirable given the limitation of the interaction circumstances.

4.2. Emotion Inducer

We opt for short video clips (a few minutes in length) as emotion inducers in the sessions. The use of video clips as emotion elicitor is well established and has been studied for several decades [19], [20]. One study shows that amongst a number of techniques, the use of video clips is the most effective way to induce both positive and negative emotional states [21]. Furthermore, this technique offers practical replication in constrained environmental settings, such as the recording room. Finally, in terms of ethical concerns, this technique is less personally involved for the inducee compared to others such as autobiographical recollection [22] or the real life method, where inducees are asked to perform various physical tasks [23].

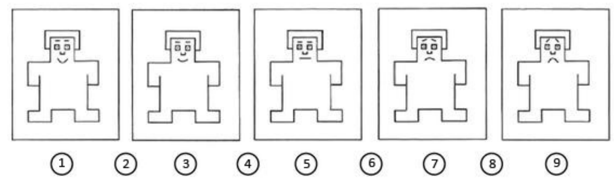
However, unlike the majority of previous studies which uses excerpts of films or movies showing hyper-realistic fictional situations [22], we look for clips that depict real life situations and issues, i.e., non-fiction and non-films. Our concern is the unpredictability from person to person when emotionally responding to clips knowing that it is fictional. Furthermore, the use of a non-fictional inducer reflects real everyday situations better. We intend for the clips to contain enough information and context of a certain subject to serve as conversation topic throughout the recording session.

To fit these requirements, we select short video clips of news reports, interviews, and documentary films as emotion inducers. First, we manually selected 34 of videos with varying relevant topics that are provided freely online. Two human experts are then asked to rate them in terms of intensity and the induced emotion (sadness or anger). Finally, we selected 20 videos, 10 of each emotion with varied intensity level where the two human ratings agree.

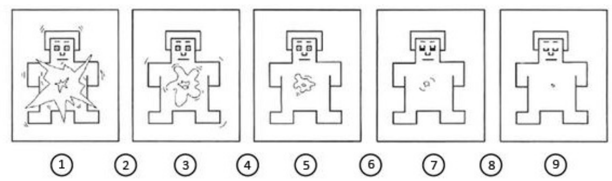
Among others, the anger inducers include reports on an unfair working environment, animal cruelty, and domestic violence. The sadness inducers include but are not limited to stories on environmental changes, a person who went through child abuse, and a child bride.

4.3. Questionnaires

4.3.1. Self-Assessment Manikin (SAM) Questionnaire. Throughout the process, we ask the participants to assess their emotional states through the use of a Self-Assessment Manikin (SAM). The SAM is a pictorial rating scheme designed for easy-to-use non-verbal assessment of emotional state and reaction [24]. Following the emotion definition in Section 3, we exclusively use the valence and arousal SAM. The pictorial scale for valence and arousal is depicted in Figure 3.



(a) Valence SAM.



(b) Arousal SAM.

Figure 3. SAM for self-assessment of emotional state and reaction [24].

For both dimensions, the scale ranges from 1 (most positive) to 9 (most negative) with matching illustrations of the emotional states. During assessment, the participants are simply asked to choose a number for each dimension that matches their current state of emotion, as written below the illustrations.

4.3.2. Post-Recording Questionnaire. We ask the participants to fill out post-recording questionnaires to quantitatively measure 1) the effectiveness of the emotion inducer, and 2) the satisfaction of the session with the counselor. The participants are asked to answer the following questions on a Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree):

- I noticed a negative emotional change in myself after watching the video.
- I noticed a positive emotional change in myself during and after the conversation with the counselor.
- The conversation helped me deal with and process my emotion.
- I felt understood by the counselor.
- I enjoyed the conversation with the counselor.
- I found a kind of emotional connection between myself and the counselor.
- I would like to talk again with the counselor in the future.

5. Data Collection

5.1. Participants

We recruit a professional counselor as the *expert* in the recording. The expert obtained a Diploma in Counseling and has been an accredited member of the British Association for Counseling and Psychotherapy. The expert has more than 8 years of professional experience, and has been practicing with the following areas of expertise: general counseling (e.g., depression, anxiety), relational issues, sexuality, childhood trauma, identity, cultural adjustment, and family problems.

As *participants*, we recruit 30 individuals that speak English fluently as first or second language. The group of speakers covers 13 nationalities in total. All of the participants are residing in the same area and are embedded in an international academic environment during the time of our recordings. The group consists of 20 males and 10 females.

5.2. Set Up

We record the videos of the dyad with two cameras, each facing a single person for a portrait shot. The two cameras are the SONY Handycam HDR-CX670 and the SONY Handycam HDR-PJ675. We record with 29.97 frames per second and a resolution of 1280x720 pixels. The video recordings are stored with the H.264 video compression standard in the yuv420p color space.

The audio signals are captured with two Crown CM-311A cardioid condenser head-worn vocal microphones, both of them wired to a USB audio interface of the type Roland QUAD-CAPTURE UA-55. We record the speech of the dyad as two mono audio signals, one for each speaker, at a sound rate of 44.1 kHz with 16-bit PCM quality, stored in a single .wav file format. After the recording, the data from the camcorders and the microphones are synchronized manually. The layout of the recording room is illustrated in Figure 4.

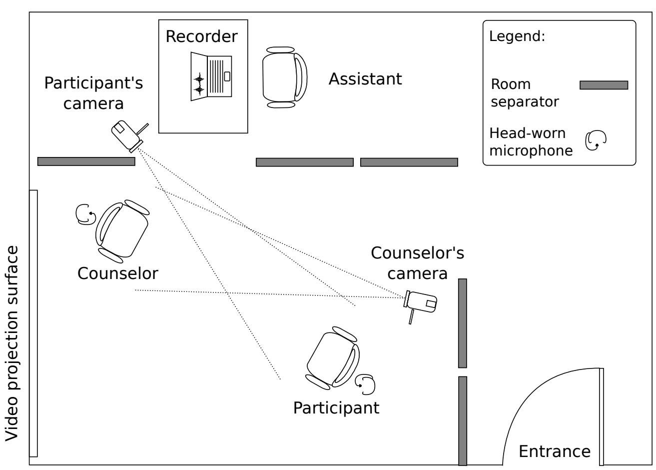


Figure 4. The recording room layout.

5.3. Recording Procedure

We record 60 sessions in the course of 6 weeks, i.e. 10 sessions are recorded in a week. Each of the 30 participants attended 2 sessions with at least one week period between them. For each participant, one session is assigned to an anger-inducing video clip, and another to a sadness-inducing clip. Each video clip is shown to 3 different participants.

Each session is allocated 30 minutes and the procedure is as follows. The camcorders and the expert's microphone are set up prior to the recording. The expert waits for each participant in the recording room while they are briefed in a separate room. Afterwards, the participant enters the recording room, takes a seat, and the recording assistant places his or her microphone. After the equipment is set up, the assistant retreats behind the room separators.

Every session is opened with a brief talk covering topics such as the general well-being, the current research, study or career progress, and weekend plans. Afterwards, the expert hands over the SAM questionnaire for self-assessment. Upon completion, the emotion inducer is shown on the video projection surface. The expert and participant watch the video at the same time. The assistant leaves the room after the playback of the video stops. The participant is then asked to fill out another SAM questionnaire.

Afterwards, the conversation between the dyad begins for the remaining of the allocated 30 minutes. The participant leaves the room when either the allocated time has passed, or when the conversation comes to a natural conclusion. The participant continues to the post-recording procedure, which consists of debriefing and filling out the post-recording questionnaire. In the meantime, the expert does verbal assessment regarding the participant and the conversation. The assistant returns to the recording room at this point to secure the data from the current audio and video recordings and to prepare for the next session.

5.4. Collected Data

During the course of the recordings, a large quantity of audio and video data has been recorded. After removing the overheads of pre- and post-recording periods, the combined duration of all sessions sums up to 23 hours and 41 minutes of material. On average, one session yielded 23.6 minutes of parallel audio and video data that is relevant for annotation. This time includes the opening talk prior to showing the emotion inducer, the video playback period and the discussion. The shortest and longest sessions are 10 and 33 minutes long, respectively.

6. Data Annotation

6.1. Emotion Annotation

The emotion occurrences are annotated using the FEELtrace system [25] to allow the recording of perceived emotion in real time. While an annotator is watching a target person in a recording, he or she is moving a cursor along a linear scale on an adjacent window to indicate the perceived emotional aspect (e.g., valence or arousal) of the target. This results in a sequence of real numbers ranging from -1 to 1, called a *trace*, that shows how a certain emotional aspect falls and rises within an interaction. Statistical analyses of validation experiments have confirmed the reliability and indicated the precision of the FEELtrace system [25].

Following the emotion definition of Section 3, we annotate both the valence and arousal dimensions of each recording. A screen capture of the annotation tool is shown in Figure 5.

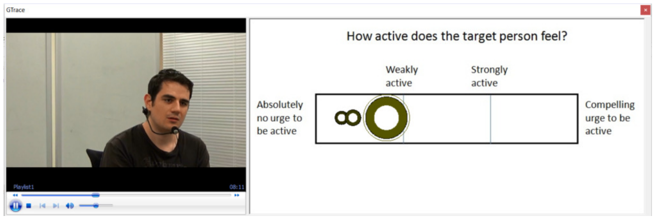


Figure 5. Annotating the arousal dimension.

At this stage of development, we focus on annotating the participant's emotional state. We aim to provide two types of annotations: self-reported and perceived. Self-reported emotion is annotated by the subjects themselves (in this case, the participants), while perceived emotion is annotated by another party according to the communication clues that the subject expresses (in this case, the expert).

For each session, the annotation is performed twice: once for valence, and once for arousal. The self-reported emotion is annotated by the participants directly after the recording is finished. Due to the tight recording schedule with the expert, this arrangement is not possible for the perceived emotion annotation. Instead, the expert performs the task off-site post-recording. All 60 sessions have been annotated with self-reported emotion and perceived emotion traces.

6.2. Transcription

We transcribe the spoken language of each recorded session. The speech of each speaker is split into separate channels, providing clean non-overlapping speech for the transcription task. We employ a paid Automatic Speech Recognition (ASR) service[1] to obtain an automatic transcription of the data. The automatic transcription is then subject to manual revision and inspection of a professional human transcriber.

During manual revision, we maintain non-speech information that potentially gives emotional state clues. The following parts of speech are given special notations: laughter, back-channel utterances, lip, nose and throat noise. All of the sessions have been automatically transcribed manually corrected.

7. Analysis

7.1. SAM and Post-Recording Questionnaires

The details of the SAM and post-recording questionnaires collected throughout the data collection are laid out in Section 4.3. The analysis in this subsection is based on all 60 recorded sessions.

Figure 6 presents the proportion of ratings for all metrics in the post-recording questionnaire. The rating ranges from 1 (strong agreement) to 5 (strong disagreement). Aside from the first metric, low ratings or agreement on the questionnaire indicates a satisfying session. On average, the participants express an agreement to varying degrees for all of the evaluated metrics. Looking at the proportions as well as the average of the ratings for each metric, we found that:

- the emotion inducer videos are effective in eliciting a negative emotional response (video_neg),
- the participants reported an agreement towards the positive emotional effect of the conversation (chat_pos),
- the participants feel that the conversation helps them to process their emotion (helps_emo),
- strongest agreement is observed on the enjoyment of the conversation (enjoyed), followed by the feeling of being understood by the expert (understood),
- emotional connection appears to be the most difficult feeling to achieve through the interaction (emo_connect), possibly due to the limited time and lack of continuity of the interaction, and
- in general, the participants express that they would like to interact with the expert again in the future (chat_again).

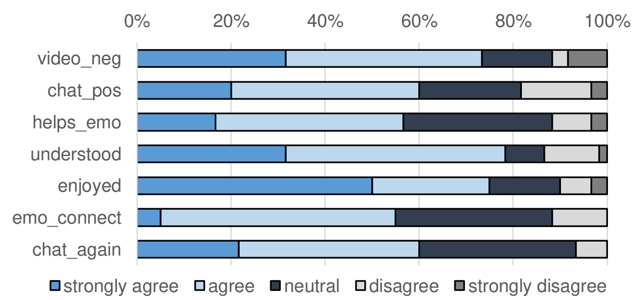
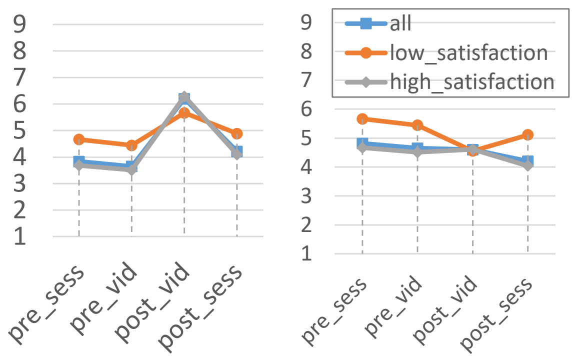


Figure 6. The proportion of ratings of the post-recording questionnaire. The statements of the questionnaire are detailed on Section 4.3.2.

Using the rating of all metrics except the first one (video_neg), we divided the participants into two groups: low and high satisfaction. When the average rating of the metrics is larger than 3, i.e., suggesting disagreement, the participant is put into the low satisfaction group. Otherwise, the participant is put into the high satisfaction group.

For both groups, different trends can be observed in the SAM questionnaire, as shown in Figure 7. The results show significant statistical difference between the two groups' pre-recording valence ($p \leq 0.1$). Furthermore, the impact on valence of both the emotion inducer (negative) and the session (positive) is significantly more intense on the high satisfaction group compared to to the low satisfaction group ($p < 0.1$). On the other hand, the two groups show opposing emotional changes in terms of arousal. We observe a statistically significant difference between the two groups in terms of arousal change after interacting with the counselor ($p < 0.05$). Figure 7(a) also confirms the consistent negative effect of the inducers and the role of the interaction with the counselor in recovering from it.

1. https://www.popuparchive.com



(a) Valence. (b) Arousal.

Figure 7. Average levels of emotion throughout the recording process. The scale ranges from 1 (strongly positive for valence and strongly activated for arousal) to 9 (strongly negative for valence and strongly deactivated for arousal).

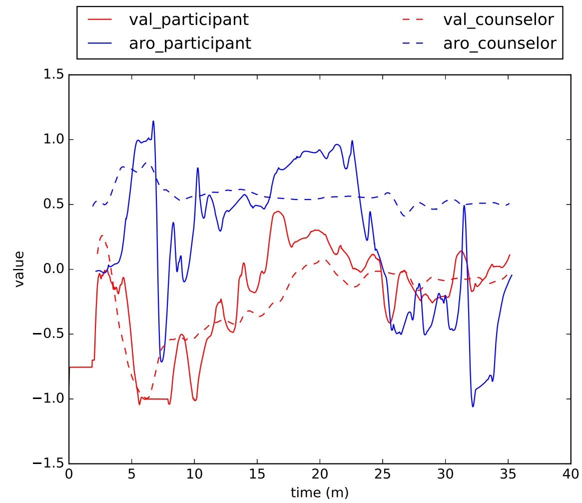
7.2. Self-Reported and Perceived Emotion Annotation

The following analysis is based on all of the sessions in the database. We investigate the correlation between the participant's self-report of their emotion and the corresponding emotion as perceived by the expert (Section 6.1). We suspect that there are differences between emotion as reported by the person who experiences it and by another person who perceives it from the outside. To quantify the agreement between the two annotations, we utilize Pearson's correlation coefficient r. Pearson's r measures the strength and direction of a linear relationship between two variables. Prior to computation, we apply the Savitzky-Golay filter to smooth the annotation as well as increase the signal-to-noise ratio [26].

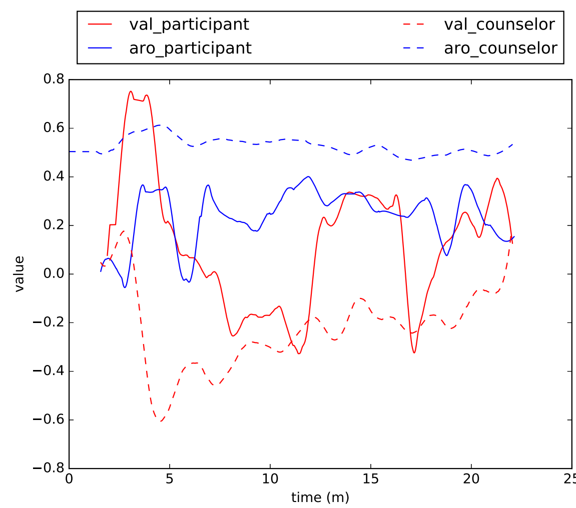
We found that the correlations for valence annotation are consistently stronger than that of activation. Strong correlation ($r \geq 0.5$) for valence are observed in 68.33% of the annotated sessions, while only 8.3% of the sessions have strong correlation for arousal. The average correlation is 0.585 for valence and 0.044 for arousal. Annotations from two sessions with respectively strong and weak correlations are depicted in Figure 8. We notice that in general the self-reported and perceived annotations are correlated more strongly when the emotion of the participant is more intense, i.e., emotion with more drastic changes, and values that reach the extremes of the scale.

8. Conclusion

In this paper, we present recordings of dyadic social-affective interactions between a professional counselor and 30 participants, amounting to 23 hours and 41 minutes of annotated data. The proposed database differs from existing ones in that it is explicitly designed to allow the observation of emotion changes at interpersonal level, involving an external party that guides the emotional process that follows negative emotion induction. We recruited a professional counselor to fill the role of an expert in facilitating this process.



(a) r for valence: 0.78, r for arousal: 0.46.



(b) r for valence: 0.32, r for arousal: 0.08.

Figure 8. Emotion of the participants of two sessions as annotated by the participant (*_*participant*) and as annotated by the expert (*_*counselor*).

We are aiming to provide high quality information of relevance to maximize the potential use of the collected data. We have completed the human annotation of self-reported and perceived emotions, as well as the manually refined transcriptions of the conversation. Collection of multiple perceived emotion annotations may be beneficial in improving the reliability. Annotating the expert's emotion could let us observe sympathetic reactions in the conversation. Additionally, in the future we hope to provide higher level information of the recorded sessions, such as dialogue act, state, and strategy labels.

The presented corpus is designed to support affective computing research that focuses on emotion at interpersonal or social level. Towards this direction, in the future we hope to utilize the corpus for developing a dialogue system with an emotionally intelligent dialogue strategy.

Acknowledgments

This research is supported by the MIC/SCOPE #152307004.

References

[1] O. Luminet IV, P. Bouts, F. Delie, A. S. Manstead, and B. Rimé, "Social sharing of emotion following exposure to a negatively valenced situation," *Cognition & Emotion*, vol. 14, no. 5, pp. 661–688, 2000.

[2] B. Rime, B. Mesquita, S. Boca, and P. Philippot, "Beyond the emotional event: Six studies on the social sharing of emotion," *Cognition & Emotion*, vol. 5, no. 5-6, pp. 435–465, 1991.

[3] A. N. Cohen, C. Hammen, R. M. Henry, and S. E. Daley, "Effects of stress and social support on recurrence in bipolar disorder," *Journal of affective disorders*, vol. 82, no. 1, pp. 143–147, 2004.

[4] E. Zech and B. Rimé, "Is talking about an emotional experience helpful? Effects on emotional recovery and perceived benefits," *Clinical Psychology & Psychotherapy*, vol. 12, no. 4, pp. 270–287, 2005.

[5] B. Reeves and C. Nass, *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge university press, 1996.

[6] T. Meguro, Y. Minami, R. Higashinaka, and K. Dohsaka, "Learning to control listening-oriented dialogue using partially observable markov decision processes," *ACM Transactions on Speech and Language Processing (TSLP)*, vol. 10, no. 4, p. 15, 2013.

[7] M. Cavazza, R. S. De La Camara, and M. Turunen, "How was your day?: a companion eca," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 2010, pp. 1629–1630.

[8] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015.

[9] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet *et al.*, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1061–1068.

[10] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, "The Japanese female facial expression (JAFFE) database," 1998.

[11] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[12] C. Busso, S. Parthasarathy, A. Burmania, M. AbdelWahab, N. Sadoughi, and E. M. Provost, "MSP-IMPROV: An acted corpus of dyadic interactions to study emotion perception," *IEEE Transactions on Affective Computing*, vol. 8, no. 1, pp. 67–80, 2017.

[13] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *Language resources and evaluation*, vol. 42, no. 4, p. 335, 2008.

[14] M. Grimm, K. Kroschel, and S. Narayanan, "The Vera am Mittag German audio-visual emotional speech database," in *Multimedia and Expo, 2008 IEEE International Conference on*. IEEE, 2008, pp. 865–868.

[15] N. Lubis, S. Sakti, G. Neubig, T. Toda, and S. Nakamura, "Construction and analysis of social-affective interaction corpus in english and indonesian," in *O-COCOSDA/CASLRE, 2015 International Conference*. IEEE, 2015, pp. 202–206.

[16] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder, "The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent," *Transactions on Affective Computing*, vol. 3, no. 1, pp. 5–17, 2012.

[17] J. Gratch, R. Artstein, G. M. Lucas, G. Stratou, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella *et al.*, "The distress analysis interview corpus of human and computer interviews." in *LREC*. Citeseer, 2014, pp. 3123–3128.

[18] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

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

[20] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cognition and Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.

[21] R. Westermann, G. Stahl, and F. Hesse, "Relative effectiveness and validity of mood induction procedures: analysis," *European Journal of social psychology*, vol. 26, pp. 557–580, 1996.

[22] A. Schaefer and P. Philippot, "Selective effects of emotion on the phenomenal characteristics of autobiographical memories," *Memory*, vol. 13, no. 2, pp. 148–160, 2005.

[23] C. Landis, "Studies of emotional reactions. II. general behavior and facial expression." *Journal of Comparative Psychology*, vol. 4, no. 5, p. 447, 1924.

[24] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[25] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröder, "'FEELTRACE': An instrument for recording perceived emotion in real time," in *ISCA tutorial and research workshop (ITRW) on speech and emotion*, 2000.

[26] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.