

Multimodal Database of Negative Emotion Recovery in Dyadic Interactions: Construction and Analysis *

☆ Nurul Lubis¹, Michael Heck², Sakriani Sakti^{1,2}, Koichiro Yoshino^{1,2,3},
Satoshi Nakamura^{1,2} (¹NAIST, ²RIKEN AIP, ³JST PRESTO)

1 Introduction

Social sharing of emotion is an essential part in processing a negative emotional experience. As the technology develops, the potential of dialogue systems to address this need has been growing as well. However, there is still a lack of resources that: 1) highlight the emotional problems commonly encountered in everyday life, and 2) involve an expert in the conversation.

To fill these gaps, in this paper we present a multimodal database containing dyadic interactions between a professional counselor as an expert and 30 participants. Each session starts with negative emotion elicitation through short video clips, and in the interaction that follows, the goal of the expert is to aid emotion processing and elicit a positive emotional change. This database allows us to observe how an external party can guide and facilitate emotion processing after a negative emotional response in a common social situation. The construction, development, and analysis of the database are reported in this paper.

2 Database Design

We would like to observe how an external party can guide and facilitate emotion processing through an interaction after a negative emotional response. 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*.

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. Throughout the process, we ask the participants to assess their emotional states with the use of a questionnaire. At the end of the session, we use a questionnaire to measure participant's satis-

faction of the interaction

We opt for short video clips (a few minutes in length) as emotion inducers in the sessions. 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 [1]. We look for non-fiction clips that depict real life issues. To fit these requirements, we select short video clips of news reports, interviews, and documentary films as emotion inducers.

3 Recording and Annotation

We recruit a professional counselor as the *expert* in the recording. The expert is an accredited member of the British Association for Counseling and Psychotherapy with more than 8 years of professional experience. As *participants*, we recruit 30 individuals (20 male, 10 female) who speak English fluently as first or second language. We record 2 sessions for each participant, one induced to “anger,” the other to “sadness.” In total, we record 60 sessions, amounting to 23 hours and 41 minutes of data.

We use the circumplex model of affect, and annotate two emotion dimensions accordingly: valence and arousal [2]. Valence measures the positivity of emotion, and arousal measures the activity. Each dimension is annotated separately using the FEELtrace system [3]. 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. Validation experiments have confirmed the reliability and indicated the precision of the FEELtrace system [3]. We provide two types of emotion annotation: self-reported emotion, annotated by the participants; and perceived emotion, annotated by the expert.

We transcribe the spoken language of each recorded session. We employ a paid Automatic Speech Recognition (ASR) service to obtain an automatic transcription of the data. The automatic transcription is then subject to manual revision and

*負の感情からの回復を目的とする二者対話のマルチモーダルコーパス: 構築と分析

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.

4 Analysis

Figure 1 shows the post-recording questionnaire result. Firstly, the result shows that the emotion inducer videos are effective in eliciting a negative emotional response (video_neg). The participants also reported an agreement towards the positive emotional effect of the conversation (chat_pos) as well as the feeling 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 counselor (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. In general, the participants express that they would like to interact with the counselor again in the future (chat_again).

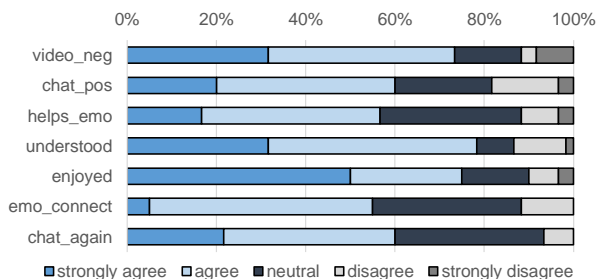


Fig. 1 Post-recording questionnaire result.

We investigate the correlation between the self-report and perceived emotion annotations with Pearson’s correlation coefficient r . 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 but only 8.3% for arousal. The average correlation is 0.585 for valence and 0.044 for arousal. The annotations from a session is depicted in Figure 2. We notice that in general the self-reported and perceived annotations are correlated more strongly when the emotion of the par-

ticipant is more intense.

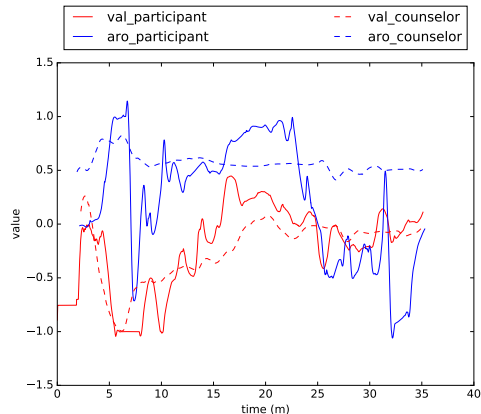


Fig. 2 Emotion annotations of one of the sessions (r for valence: 0.78, arousal: 0.46).

5 Conclusion

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. 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.

Acknowledgement

This research and development work was supported by the MIC/SCOPE #152307004, as well as JSPS KAKENHI Grant Numbers JP17H06101 and JP17K00237.

参考文献

- [1] Rainer Westermann, Gunter Stahl, and F Hesse. Relative effectiveness and validity of mood induction procedures: analysis. *European Journal of social psychology*, 26:557–580, 1996.
- [2] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.
- [3] Roddy Cowie, Ellen Douglas-Cowie, Susie Savvidou, Edelle McMahon, Martin Sawey, and Marc Schröder. ‘FEELTRACE’: An instrument for recording perceived emotion in real time. In *ISCA tutorial and research workshop (ITRW) on speech and emotion*, 2000.