The Distress Analysis Interview Corpus of human and computer interviews

Jonathan Gratch, Ron Artstein, Gale Lucas, Giota Stratou, Stefan Scherer, Angela Nazarian, Rachel Wood, Jill Boberg, David DeVault, Stacy Marsella, David Traum, Skip Rizzo, Louis-Philippe Morency
USC Institute for Creative Technologies, 12015 Waterfront Drive, Playa Vista CA 90094-2536, USA
{gratch, artstein, lucas, stratou, scherer, nazarian, rwood, boberg, devault, marsella, traum, rizzo, morency}@ict.usc.edu

Abstract

The Distress Analysis Interview Corpus (DAIC) contains clinical interviews designed to support the diagnosis of psychological distress conditions such as anxiety, depression, and post traumatic stress disorder. The interviews are conducted by humans, human controlled agents and autonomous agents, and the participants include both distressed and non-distressed individuals. Data collected include audio and video recordings and extensive questionnaire responses; parts of the corpus have been transcribed and annotated for a variety of verbal and non-verbal features. The corpus has been used to support the creation of an automated interviewer agent, and for research on the automatic identification of psychological distress.

Keywords: multimodal corpora, virtual humans, dialogue systems, nonverbal behavior

1. Overview

Untreated mental illness creates enormous social and economic costs, yet many cases go undiagnosed. Up to half of patients with psychiatric disorders are not recognized as having mental illness by their primary care physicians (Higgins, 1994). Within health-care settings, a first step in identifying mental illness is a semi-structured clinical interview, where health-care providers ask a series of questions aimed at identifying clinical symptoms in an open-ended fashion. Recently, there is considerable research interest in developing tools to analyze the verbal and nonverbal content of these interviews as a means for building decision-support tools (Gratch et al., 2013) and computer-assisted self-administered screenings (Bickmore et al., 2005), and for answering fundamental questions about language, nonverbal behavior and mental illness (Scherer et al., 2013b; Yang et al., 2013; Alvarez-Conrad et al., 2001).

The Distress Analysis Interview Corpus (DAIC) is a multimodal collection of semi-structured clinical interviews. Designed to simulate standard protocols for identifying people at risk for post-traumatic stress disorder (PTSD) and major depression, these interviews were collected as part of a larger effort to create a computer agent that interviews people and identifies verbal and nonverbal indicators of mental illness (DeVault et al., 2014). The corpus contains four types of interviews:

Face-to-face interviews between participants and a human interviewer (Figure 1);

Teleconference interviews, conducted by a human interviewer over a teleconferencing system;

Wizard-of-Oz interviews, conducted by an animated virtual interviewer called Ellie (Figure 2), controlled by a human interviewer in another room;

Automated interviews, where participants are interviewed by Ellie operating as an agent in a fully automated mode.

Sample interview excerpts are shown in Figure 3.

Figure 1: Face-to-face interview setup.

Figure 2: Ellie, the virtual interviewer.

Participants are drawn from two distinct populations living in the Greater Los Angeles metropolitan area – veterans of the U.S. armed forces and from the general public – and are coded for depression, PTSD and anxiety based on accepted psychiatric questionnaires.

Besides informing the development of computer-assisted interviews that improve rates of diagnosis, the corpus has been used to examine several fundamental questions about language, nonverbal behavior, psychophysiology and human-computer interaction. This article describes the de-

Face-to-face

Interviewer Let's see, is there someone in your life that's been a really positive influence for you?
Participant Uh yeah, my husband, yeah.
Interviewer Yeah.
Interviewer What kind of values did you take away from him?
Participant Uh he's always uh thinking ahead and looks at the big picture and doesn't uh mull over trivial things so that's something that helped me.
Interviewer Mhm yeah, those are good traits to have.
Participant Yeah, yes.
Interviewer Um how did you guys meet?

Wizard-of-Oz

Ellie Who's someone that's been a positive influence in your life?
Participant Uh my father.
Ellie Can you tell me about that?
Participant Yeah, he is a uh
Participant He's a very he's a man of few words
Participant And uh he's very calm
Participant Slow to anger
Participant And um very warm very loving man
Participant Responsible
Participant And uh he's a gentleman has a great sense of style and he's a great cook.
Ellie Uh huh
Ellie What are you most proud of in your life?

Autonomous

Ellie Who's someone that's been a positive influence in your life?
Participant My mom
Participant Has been a positive influence
Ellie Tell me more about that
Participant Well she's just always really nice and
Participant Considerate and upbeat and
Participant Sh just positive person
Ellie How would your best friend describe you?
Participant Outgoing funny
Participant A little bit
Participant Foolish <laughter>

Figure 3: Sample excerpts from three types of interviews

velopment and content of the corpus and summarizes the key scientific findings it has supported to date.

2. Collection

Collection began with the interpersonal interviews (both face-to-face and via teleconferencing), followed by the Wizard-of-Oz interviews and the automated interviews. Participants were recruited through two channels: on-line ads posted on Craigslist.org, and on-site at a US Vets facility in Southern California (US Vets is a non-profit organization that helps veterans re-integrate into civilian life after deployment). Face-to-face interviews for the Craigslist population took place at the USC Institute for Creative Technologies (ICT) in Los Angeles, California, while the US Vets population were interviewed on-site. Teleconference, Wizard-of-Oz, and automated agent interviews were all conducted at ICT: teleconference and Wizard-of-Oz participants were only recruited through Craigslist, while for the automated agent, US Vets participants were transported to ICT. All participants were fluent English speakers and all interviews were conducted in English. Face-to-face and teleconference interviews typically took 30–60 minutes, while Wizard-of-Oz and automated inteviews were shorter (Wizard-of-Oz 5–20 minutes, automated 15–25 minutes). A summary of collected interviews can be seen in Table 1.

All collection efforts used the same experimental protocol except where changes were required by the nature of the interaction (human vs. computer interviewer) or instrumentation (see below). Participants first completed a consent form (which included optional consent that allowed their data to be shared for research purposes). They then completed set of questionnaires alone on a computer, then went through the interview, followed by additional questionnaires after the interview. Participants were recorded only during the interview. Interviews were semi-structured, starting with neutral questions designed to build rapport and make the participant comfortable; progressing to more specific questions about symptoms and events related to depression and PTSD; and ending with a "cool-down" phase, to ensure that participants would not leave the interview in a distressed state of mind.

Each face-to-face and teleconference interview was conducted by one of two female interviewers. In the face-to-face condition, only the participant and interviewer were in the room during the interview (Figure 1 above). In the teleconference, Wizard-of-Oz and automated interviews, participants were alone in a room in front of a large computer screen, showing the human interviewer in the teleconference interviews, and the animated character Ellie in the Wizard-of-Oz and automated interviews (Figure 2 above).

Ellie's behavior in the Wizard-of-Oz collection was controlled by two wizards, responsible for non-verbal behaviors (e.g., nods and facial expressions) and verbal utterances, respectively (the wizards were the same two interviewers from the face-to-face and teleconference interviews). Two wizards were necessary because controlling both verbal utterances and non-verbal behaviors proved too difficult for a single person to handle in real time. Ellie had a fixed set of utterances (these consisted of pre-recorded audio of the wizard that controlled Ellie's verbal behavior and pre-animated gestures and facial expressions based on those typically employed during the face-to-face interviews). Small changes were made to the interview protocol throughout the data collection effort: the wizards followed a written policy which gradually became stricter and more structured.

The autonomous agent's behavior was guided solely by its implemented policies, without any manual intervention. The policies were refined over time as the agent development progressed.

Starting partway through the Wizard-of-Oz collection and continuing through the automated agent collection, participants were randomly assigned to one of two framing conditions, presenting the character as either an autonomous computer system or a system controlled by a person.

Condition	Total	Framing		Distress		Transcribed	Biopac
		Human	Computer	Yes	No		
Face-to-face	120	—	—	49	71	74	20
Teleconference	45	—	—	16	29	0	20
Wizard-of-Oz	193	53	140	60	133	193	66
Automated agent	263	47	216	99	164	95	71

Table 1: Interviews collected

3. Corpus Composition

3.1. Verbal, Nonverbal and Physiological Instrumentation

The corpus contains audio, video, and depth sensor (Microsoft Kinect) recordings of all the interactions. For the face-to-face and teleconference interactions, the interviewer and participant were recorded by separate cameras, lapel microphones, and Kinects; additionally, face-to-face interviews used an overhead camera to capture the general orientation of the interviewer and participant in the environment. In the Wizard-of-Oz and automated agent interviews, the participant was recorded by a camera, high-quality close-talking microphone, and Kinect, while the agent was recorded through screen-capture software.

A subset of the collections also include physiological data (Biopac).[1] We record galvanic skin response (GSR), electrocardiogram (ECG), and respiration of participants. Sensors were attached following the pre-questionnaires but before the interview. Sensors were connected to the participants' trunk and lower-extremities to avoid interference with natural gestures. Additionally, participants were required to sit at rest for three minutes, then presented a series of standardized emotional pictures (Lang et al., 2008) to help calibrate the instrument and provide baseline measures of physiological responsiveness.

3.2. System logs

The Wizard-of-Oz and automated agent interviews include generated logs of the character's speech and nonverbal behavior events. Additionally, the automated agent logs contain real-time segmentation, recognition and understanding of the participants' speech and language, which drive the agent's actions.

3.3. Questionnaire data

Participants completed a series of questionnaires prior to the interview, including basic demographic questions, established measures of psychological distress, and a measure of current mood. The Positive and Negative Affect Scale (PANAS) was used to assess mood (Watson and Clark, 1994). Measures of psychological distress included the PTSD Checklist – Civilian Version (Blanchard et al., 1996), the Patient Health Questionnaire, depression module (Kroenke and Spitzer, 2002), and the State-Trait Anxiety Inventory (Spielberger et al., 1983), all of which were highly correlated, reflecting the typical comorbidity found between these clinical conditions (Gaudiano and Zimmerman, 2010; Scherer et al., 2013b). After the interaction, participants completed the PANAS again, Kang and Gratch's (2012) Rapport scale, and a measure of social desirability (Li and Bagger, 2007). They also rated their interaction partner on 32 adjectives using a 7-point likert scale with response options ranging from a positive adjective (e.g., polite, kind, warm) to a negative adjective (rude, cruel, cold, respectively).

Before the interview, some participants completed a measure of the five factors of personality (John et al., 1991), and others completed a measure of emotion regulation (Gross and John, 2003). After the interview, participants in the Wizard-of-Oz and automated agent conditions rated their fears of being evaluated negatively during the interview (Leary, 1983) as well as the system's usability (Brooke, 1996); they also rated success at specific design goals (such as "Ellie was sensitive to my body language" and "Ellie was a good listener") using a 5-point scale from "strongly disagree" to "strongly agree".

3.4. Transcription

A portion of the interviews was segmented and transcribed using the ELAN tool from the Max Planck Institute for Psycholinguistics (Brugman and Russel, 2004).[2] Each transcription was reviewed for accuracy by a senior transcriber. Utterances were segmented at boundaries with at least 300 milliseconds of silence. The face-to-face and early Wizard-of-Oz interviews were transcribed from a composite video, combining both participant and interviewer; later Wizard-of-Oz and automated interviews were transcribed from the audio stream of the participant only, while the interviewer utterances were recovered from the system logs.

3.5. Annotation

De-identification All the transcribed interviews were annotated to remove identifying information. Utterances were tagged for mentions of personal names, specific dates, addresses, schools, places of employment, and locations that can be used to narrow down an event. Utterances were not considered to be personally identifying if they only included large locations (e.g. "I live in Santa Monica"), very large institutions ("I served in the Marines"), or nonspecific dates such as age in years. De-identification was

[1] http://www.biopac.com

[2] http://tla.mpi.nl/tools/tla-tools/elan

performed independently by two annotators, and differences were reconciled by a senior annotator. Utterances marked as personally identifying will not be shared in accordance with our institution's ethical guidelines.

Explicit psychological conditions The corpus includes annotations of questions and statements that give an explicit indication of a past or existing condition of psychological distress (e.g. "She diagnosed me with a type of depression"). These are useful when developing systems to detect the more subtle distress signals.

Dialogue annotation Parts of the transcribed corpus have been annotated with dialogue-level information to support the development and training of natural language understanding for the agent. Annotations include: (1) Identification of clarification questions by the participant (e.g. "What was the question?"). (2) Places that are appropriate for the agent to provide a positive or negative empathy response (e.g. "That's great" or "I'm sorry"), used to tune the thresholds on the valence classifier for the agent (DeVault et al., 2014); inter-rater agreement (Krippendorff, 2011) was 0.73 for positive empathy and 0.81 for negative empathy. (3) Domain-specific dialogue acts in participants' responses to specific questions, to support follow-up by the agent.

Non-verbal behavior annotation Several non-verbal behaviors were annotated (Waxer, 1974; Hall et al., 1995): gaze directionality (up, down, left, right, towards interviewer), listening smiles (smiles while not speaking), self-adaptors (self-touches in the hand, body, and head), fidgeting behaviors, and foot-tapping or shaking behaviors. Each behavior was annotated in a separate tier in ELAN. Four student annotators participated in the annotation; each tier was assigned to a pair of annotators, who first went through a training phase until the inter-rater agreement (Krippendorff's alpha) exceeded 0.7. Following training, each video was annotated by a single annotator; to monitor reliability, every 10–15 videos each pair was assigned the same video and inter-rater agreement was re-checked. Annotators were informed that their reliability was measured but did not know which videos were used for cross-checking (Wildman et al., 1975; Harris and Lahey, 1982).

In addition, automatic annotation of non-verbal features was carried out using a multimodal sensor fusion framework called MultiSense, with a multithreading architecture that enables different face- and body-tracking technologies to run in parallel and in realtime. Output from MultiSense was used to estimate the head orientation, the eye-gaze direction, smile level, and smile duration. Further, we automatically analyzed voice characteristics including speakers' prosody (e.g. fundamental frequency or voice intensity) and voice quality characteristics, on a breathy to tense dimension (Scherer et al., 2013a).

4. Usage

The corpus has been used to support the automated agent's interactive capabilities by developing custom acoustic and language models for speech recognition, training classifiers for natural language understanding, and informing the creation of dialogue policies; for details, see DeVault et al. (2014). The corpus has also been used to support the agent's capabilities for distress detection, using multiple types of information including visual signals, voice quality, and dialogue-level features.

Visual signals from the face-to-face data show that several features can serve as indicators of depression, anxiety, and PTSD (Scherer et al., 2013b; Scherer et al., 2014). Specifically, these forms of psychological distress are predicted by a more downward angle of the gaze, less intense smiles and shorter average durations of smile, as well as longer self-touches and fidget on average longer with both hands (e.g. rubbing, stroking) and legs (e.g. tapping, shaking). Moreover, the predictive ability of these indicators is moderated by gender (Stratou et al., 2013). A crossover interaction was observed between gender and distress level on emotional displays such as frowning, contempt, and disgust. For example, men who scored positively for depression tend to display more frowning than men who did not, whereas women who scored positively for depression tend to display less frowning than those who did not. Other features such as variability of facial expressions show a main effect of gender – women tend to be more expressive than men, while still other observations, such as head-rotation variation, were entirely gender independent.

Voice quality from the Wizard-of-Oz data, particularly differences on the breathy to tense dimension, is also a predictor of psychological distress (Scherer et al., 2013a; Scherer et al., 2014). Depression and PTSD are both predicted by more tense voice features, such that those with depression and PTSD exhibit more tense voice characteristics than those without depression or PTSD. Tense voice features were, specifically, able to distinguish interviewees with depression from those without depression with an accuracy of 75%, and distinguish those with PTSD from those without PTSD with an accuracy of 72%.

Dialogue-level features in the Wizard-of-Oz data also serve as indicators of distress: distressed individuals were slower to begin speaking and used fewer filled pauses than non-distressed participants (DeVault et al., 2013). Moreover, the type of distress may moderate which dialogue-level features are most predictive: standard deviation in onset time of first segment in each user turn yielded was the best unique predictor of depression, yet, for PTSD, mean number of filled pauses in user segments was among the most informative. For overall distress – across depression and PTSD, mean maximum valence in user segments was the most valuable. As moderating by gender improved the ability of visual signals to predict distress, moderating by type of question improves the ability of dialogue-level features to detect depression (Yu et al., 2013). For example, time to onset of speech in response to intimate questions predicts distress, whereas length of speech is more predictive of distress from responses to rapport building questions.

Overall, the corpus helps with the identification of subtle indicators of psychological distress along multiple behavioral dimensions; these indicators will be implemented in the agent, to allow it to identify people who should be referred to further evaluation. In addition, the corpus has been used for research that does not directly support development of the agent; for example, some research has considered how interviews might differ depending on the method of data collection (face-to-face, Wizard-of-Oz, automated agent interviews). A comparison of face-to-face and wizard dialogues found that participants use twice as many filled pauses when talking to the animated character than when talking to a live interviewer – precisely the opposite of previous results on people talking to task-oriented dialogue systems, where they were less disfluent when talking to the computer (Faust and Artstein, 2013). An investigation of the effects of framing the character as human-controlled or autonomous showed that participants felt lower fear of negative evaluation and engaged in less impression management when the character was framed as autonomous than when it was framed as human-controlled (Gratch et al., 2014b; Gratch et al., 2014a). In fact, actual method of data collection (Wizard-of-Oz versus automated agent interviews) had no impact on fear of negative evaluation or impression management, but who participants believed they were interacting with (human versus computer) effected both fear of negative evaluation and impression management. Moreover, participants also displayed sad emotional expressions more intensely when they believed they were interacting with a computer compared to a human. This robust dataset has the potential to help various researchers address questions across areas of mental health, human-agent interactions, and verbal and non-verbal behavior.

5. Distribution

Currently, the corpus is being shared on a case-by-case basis by request and for research purposes. Longer-term we intend to make significant portions of the data more broadly available to the research community.

Acknowledgments

This work is supported by DARPA under contract W911NF-04-D-0005 and by the U.S. Army RDECOM. Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

6. References

Alvarez-Conrad, J., Zoellner, L. A., and Foa, E. B. (2001). Linguistic predictors of trauma pathology and physical health. *Applied Cognitive Psychology*, 15(7):S159–S170.

Bickmore, T., Gruber, A., and Picard, R. (2005). Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 59(1):21–30, October.

Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., and Forneris, C. A. (1996). Psychometric properties of the PTSD checklist (PCL). *Behaviour Research and Therapy*, 34(8):669–673, August.

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., and McClelland, I. L., editors, *Usability Evaluation in Industry*, chapter 21, pages 189–194. Taylor and Francis, London.

Brugman, H. and Russel, A. (2004). Annotating multimedia / multi-modal resources with ELAN. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, pages 2065–2068, Lisbon, Portugal, May.

DeVault, D., Georgila, K., Artstein, R., Morbini, F., Traum, D., Scherer, S., Rizzo, A. S., and Morency, L.-P. (2013). Verbal indicators of psychological distress in interactive dialogue with a virtual human. In *Proceedings of the SIGDIAL 2013 Conference*, pages 193–202, Metz, France, August. Association for Computational Linguistics.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A., and Morency, L.-P. (2014). SimSensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'14)*, Paris.

Faust, L. and Artstein, R. (2013). People hesitate more, talk less to virtual interviewers than to human interviewers. In Fernández, R. and Isard, A., editors, *Semdial 2013 DialDam: Proceedings of the 17th Workshop on the Semantics and Pragmatics of Dialogue*, pages 35–43, Amsterdam, December.

Gaudiano, B. A. and Zimmerman, M. (2010). Does comorbid posttraumatic stress disorder affect the severity and course of psychotic major depressive disorder? *Journal of Clinical Psychiatry*, 71(4):442–450.

Gratch, J., Morency, L.-P., Scherer, S., Stratou, G., Boberg, J., Koenig, S., Adamson, T., and Rizzo, A. (2013). User-state sensing for virtual health agents and telehealth applications. In Westwood, J. D., Westwood, S. W., Felländer-Tsai, L., Haluck, R. S., Robb, R. A., Senger, S., and Vosburgh, K. G., editors, *Medicine Meets Virtual Reality 20*, volume 184 of *Studies in Health Technology and Informatics*, pages 151–157. IOS Press.

Gratch, J., Lucas, G. M., and King, A. (2014a). It's only a computer: Virtual humans increase willingness to disclose. Manuscript under review.

Gratch, J., Lucas, G. M., King, A., and Morency, L.-P. (2014b). It's only a computer: The impact of human-agent interaction in clinical interviews. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'14)*, Paris.

Gross, J. J. and John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personal-*

ity and Social Psychology, 85:348–362.

Hall, J. A., Harrigan, J. A., and Rosenthal, R. (1995). Nonverbal behavior in clinician-patient interaction. *Applied and Preventive Psychology*, 4(1):21–37.

Harris, F. and Lahey, B. (1982). Recording system bias in direct observational methodology: A review and critical analysis of factors causing inaccurate coding behavior. *Clinical Psychology Review*, 2(4):539–556.

Higgins, E. S. (1994). A review of unrecognized mental illness in primary care: Prevalence, natural history, and efforts to change the course. *Archives of Family Medicine*, 3(10):908–917, October.

John, O. P., Donahue, E. M., and Kentle, R. L. (1991). *The Big Five Inventory–Versions 4a and 54*. University of California, Berkeley, Institute of Personality and Social Research, Berkeley, CA.

Kang, S.-H. and Gratch, J. (2012). Socially anxious people reveal more personal information with virtual counselors that talk about themselves using intimate human back stories. In Wiederhold, B. and Riva, G., editors, *Annual Review of Cybertherapy and Telemedicine 2012*, volume 181, pages 202–207. IOS Press, Amsterdam.

Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication Methods and Measures*, 5(2):93–112.

Kroenke, K. and Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, 32(9):509–515, September.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8, University of Florida, Gainesville, FL.

Leary, M. R. (1983). A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin*, 9:371–376.

Li, A. and Bagger, J. (2007). The balanced inventory of desirable responding: A reliability and generalization study. *Educational and Psychological Measurement*, 40:131–141.

Scherer, S., Stratou, G., Gratch, J., and Morency, L.-P. (2013a). Investigating voice quality as a speaker-independent indicator of depression and PTSD. In *Proceedings of Interspeech 2013*, Lyon, France, Augist.

Scherer, S., Stratou, G., Mahmoud, M., Boberg, J., Gratch, J., Rizzo, A. S., and Morency, L.-P. (2013b). Automatic behavior descriptors for psychological disorder analysis. In *10th IEEE International Conference on Automatic Face and Gesture Recognition*, Shanghai, China, April.

Scherer, S., Stratou, G., Lucas, G. M., Mahmoud, M., Boberg, J., Gratch, J., Rizzo, A., and Morency, L.-P. (2014). Automatic audiovisual behavior descriptors for psychological disorder analysis. Manuscript under review.

Spielberger, C., Gorssuch, R., Lushene, P., Vagg, P., and Jacobs, G. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press, Inc.

Stratou, G., Scherer, S., Gratch, J., and Morency, L.-P. (2013). Automatic nonverbal behavior indicators of depression and PTSD: Exploring gender differences. In *Proceedings of International Conference on Affective Computing and Intelligent Interaction (ACII)*, Geneva, Switzerland, September.

Watson, D. and Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule – Expanded Form*. University of Iowa, Iowa City.

Waxer, P. (1974). Nonverbal cues for depression. *Journal of Abnormal Psychology*, 83(3):319–322.

Wildman, B., Erickson, M., and Kent, R. (1975). The effect of two training procedures on observer agreement and variability of behavior ratings. *Child Development*, pages 520–524.

Yang, Y., Fairbairn, C., and Cohn, J. (2013). Detecting depression severity from vocal prosody. *Affective Computing, IEEE Transactions on*, 4(2):142–150.

Yu, Z., Scherer, S., DeVault, D., Gratch, J., Stratou, G., Morency, L.-P., and Cassell, J. (2013). Multimodal prediction of psychological disorders: Learning verbal and nonverbal commonalities in adjacency pairs. In Fernández, R. and Isard, A., editors, *Semdial 2013 DialDam: Proceedings of the 17th Workshop on the Semantics and Pragmatics of Dialogue*, pages 160–169, Amsterdam, December.