

PRIMER: An Emotionally Aware Virtual Agent

Carla Gordon
USC Institute for Creative
Technologies
Los Angeles, USA
cgordon@ict.usc.edu

Anton Leuski
USC Institute for Creative
Technologies
Los Angeles, USA
leuski@ict.usc.edu

Grace Benn
USC Institute for Creative
Technologies
Los Angeles, USA
benn@ict.usc.edu

Eric Klassen
CableLabs
Boulder, Colorado, USA
e.klassen@cablelabs.com

Edward Fast
USC Institute for Creative
Technologies
Los Angeles, USA
fast@ict.usc.edu

Matt Liewer
USC Institute for Creative
Technologies
Los Angeles, USA
Liewer@ict.usc.edu

Arno Hartholt
USC Institute for Creative
Technologies
Los Angeles, USA
hartholt@ict.usc.edu

David Traum
USC Institute for Creative
Technologies
Los Angeles, USA
traum@ict.usc.edu

ABSTRACT

PRIMER is a proof-of-concept system designed to show the potential of immersive dialogue agents and virtual environments that adapt and respond to both direct verbal input and indirect emotional input. The system has two novel interfaces: (1) for the user, an immersive VR environment and an animated virtual agent both of which adapt and react to the user's direct input as well as the user's perceived emotional state, and (2) for an observer, an interface that helps track the perceived emotional state of the user, with visualizations to provide insight into the system's decision making process. While the basic system architecture can be adapted for many potential real world applications, the initial version of this system was designed to assist clinical social workers in helping children cope with bullying. The virtual agent produces verbal and non-verbal behaviors guided by a plan for the counseling session, based on in-depth discussions with experienced counselors, but is also reactive to both initiatives that the user takes, e.g. asking their own questions, and the user's perceived emotional state.

CCS CONCEPTS

• **Information Interfaces and Presentation (e.g. HCI)** → *Miscellaneous*; • **Multimedia Information Systems** → *Artificial, augmented, and virtual realities*; • **Natural Language Processing** → *Discourse*.

KEYWORDS

Virtual Reality, Virtual Agents, Spoken Dialogue Systems, Mixed-Initiative Dialogue

ACM Reference Format:

Carla Gordon, Anton Leuski, Grace Benn, Eric Klassen, Edward Fast, Matt Liewer, Arno Hartholt, and David Traum. 2019. PRIMER: An Emotionally Aware Virtual Agent. In *Joint Proceedings of the ACM IUI 2019 Workshops, Los Angeles, USA, March 20, 2019*, 10 pages.

1 INTRODUCTION

The issue of bullying has been the focus of national media attention in recent years. Although this is not a new issue, the problem has been compounded by the rise of social media technology. Today the problem is at an all-time high, with one out of five public school children reporting being bullied, according to the National Center for Education Statistics [6]. However, only 43% of bullying victims say they have reported the incident to a school official [10], indicating that victims of bullying may be uncomfortable disclosing to adults who might be able to provide the support they need.

Recent research has shown that many people feel more comfortable talking about embarrassing or personal issues with virtual agents than with strangers or even family or friends [8], [9]. Virtual agents could be a cost-effective means to provide entry-level support and allow a counselor to reach more troubled students. A virtual agent, in a VR environment, with monitoring by a (possibly remote) counselor, might be able to provide such support for children who may not know how to communicate their struggles with the adults in their life, or know how best to respond to an ongoing bullying situation at their school.

Over the last decade, there has been a considerable amount of success in creating interactive, conversational, virtual agents, including Ada and Grace, a pair of virtual Museum guides at the Boston Museum of Science [14], the INOTS and ELITE training systems at the Naval Station in Newport and Fort Benning [2], and the SimSensei system designed for healthcare support [4]. There is also some precedence for the use of virtual agents in facilitating bullying education, such as the FearNot! application developed by Aylett et al. [1].

In this paper, we introduce a system called PRIMER, that attempts to explore the user interface issues in creating a tool that a counselor could use to deploy virtual agents in helping bullied students. The main student interface involves the student interacting with a virtual character, who takes on the counselor role in guiding the session, but listening and reacting to the student's semantic and emotional expressions. There is also a counselor interface that provides information that the system has gathered from the interaction, aggregated on a "dashboard" and allows intervention.

In addition to the successes of previous virtual agents, immersive VR environments have also been shown to be helpful in the context of mental health services and resources. One such application is the use of Virtual Reality Exposure Therapy (VRET) as a treatment for soldiers suffering from PTSD. VR systems such as the Bravemind system [11] have proven effective VRET tools, creating VR environments that allow for the replication of traumatic events without exposing the patient to any real physical danger. Based on the efficacy of these VR tools, it was decided that the PRIMER student interface should be deployed in an immersive VR environment.



Figure 1: Ellie, PRIMER's virtual agent

2 USER EXPERIENCE

The user experience (UX) for PRIMER was designed to be engaging by creating a User Perceived State model (see section on Emotion Tracking) that detects and tracks the user's emotional state, and drives the system's reactive responses. Inside the HMD, users become immersed in the application's virtual environment, in this case it is a counselor's office (see Figure 1). Analyzed user input affects Ellie's behavior and dialogue responses, as well as the virtual environment itself. Ellie is capable of displaying a range of different reactive emotions, gestures, body language, and linguistic behavior, depending on the user's current perceived emotional state, and the emotional connotations of their utterances.

2.1 Emotionally Adaptive Behaviors

The system behavior was developed focusing on three core adaptive behaviors for Ellie: facial expressions, posture, and gestures. Three base posture poses (leaning back, sitting upright, leaning forward) were created with transition animations to blend between them. Gesture animations were created with three intensities (low, neutral, high). Additionally, the facial expressions adapt to the perceived emotional state of the user.

Ellie was designed to express the primary seven emotions as outlined by noted psychologist Paul Ekman: happiness, disgust, fear, surprise, anger, sadness, and contentedness; as well as blend between expressions. For example, if the system perceives the user to be in a depressed mood, this would cause her to lean forward, display a gentle encouraging facial expression, and move with small, slow gestures. Ellie takes the initiative and drives the interaction based on a pre-determined plan that can take a variety of paths depending on the user's responses and their overall emotional state.

2.2 Emotionally Adaptive Dialogue

Ellie is also capable of being adaptive to the user's perceived emotional state by changing the tone of her dialogue responses. This means that she will provide appropriate feedback to user utterances, as well as appropriate prompting questions. For example, in response to a very positive user utterance, she may say something like "That's great!", while a very negative user utterance may prompt her to say something like "That's rough." She is also capable of providing very nuanced responses if a user's perceived emotional state shifts suddenly in the conversation. For example, if the system senses the user is happy overall, but registers that a user's last utterance was indicative of sadness, Ellie can respond with a highly nuanced response such as "I understand things are worse for you this week, but I'm also sensing that you are in a good mood. Tell me what's going on."

PRIMER takes a mixed-initiative approach to dialogue interaction, meaning not only does Ellie respond to the user's utterances, and the perceived emotional connotations of those utterances, but she will also take initiative to ask the user questions as well. In this way, PRIMER can provide the semi-structured dialogue necessary to emulate a counselling session with a bullied child, creating a unique user experience.

2.3 Emotionally Reactive VR Environment

Not only does Ellie herself adapt and respond to the user's emotional state, the virtual environment itself is also capable of changing in a number of ways. Based on the user's perceived emotional state, changes to the environment work in tandem with Ellie. Users may experience changes to lighting, background ambience, room color, or even a total change of environment from an office to a park or woods or any other desired environment. Pre-selected multimedia options can be played and/or shown to help illustrate points or teach lessons or for an infinite number of possibilities depending on the application. This adaptive virtual environment is one of the features that sets the PRIMER system apart from some of the previously mentioned similar systems, such as SimSensei.

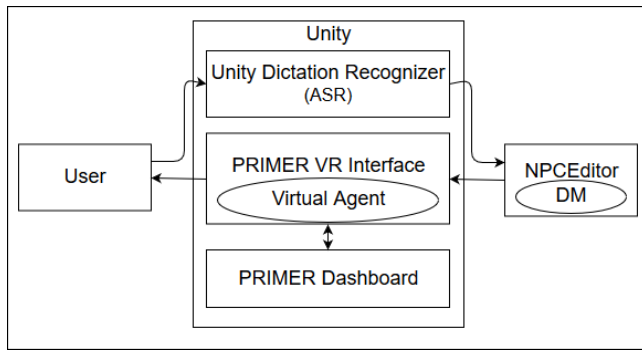


Figure 2: PRIMER system architecture

3 SYSTEM ARCHITECTURE

The PRIMER system is comprised of a number of components, all of which are integrated and/or rendered in a Unity environment. The following components were developed using the VH Toolkit:

- **Virtual Agent:** The virtual agent (Ellie) for the PRIMER system was created using the USC ICT Virtual Human (VH) Toolkit [5].
- **NPCEditor:** Utterance classification was carried out using the NPCEditor, a component of the Virtual Human Toolkit. NPCEditor is a text classification and dialogue management system that serves as the core response selection component of the system. The NPCEditor itself provides the text classification, and integrates a Dialogue Manager script which uses these classification results in the response selection process (more about the DM below). NPCEditor is also a data editor that allows us to collect, organize, and annotate the linguistic data. The NPCEditor text classification algorithm is based on cross-language relevance models and have been used in a number of successfully deployed VH systems [7]. The NPCEditor is the only component of PRIMER which is an external process, not fully encapsulated by the Unity environment.

In addition, the following new components were developed specifically for PRIMER:

- **ASR:** Speech Recognition is handled by the Unity Dictation Recognizer library, which is a light wrapper around the Windows.Speech API, the same engine used by Cortana. This makes the app Windows 10 specific.
- **The VR Interface:** The user interface was developed in Unity (see section on User Experience.) This is the interface the user interacts with, which displays Ellie (the Virtual Agent) in the virtual environment.
- **The Graphical User Interface, or "Dashboard":** In addition to the virtual environment in which the user interacts with Ellie, a second interface, called the Dashboard, was also developed in Unity. This is a separate interface designed to be viewed by a third party observer, such as a school counselor. This interface contains a host of information about the user's interaction and system decision making processes (see section on Graphical User Interface).

- **Dialogue Manager:** The Dialogue Manager (DM) for PRIMER makes use of a persistent User affect model and a conversational model, as well as sensing of current affect and linguistic input. In addition to implementing a rule-based dialogue policy and selecting a next utterance for Ellie, the dialogue manager's information state can be updated by an observer using the Dashboard and sends explanations of its decisions to the Dashboard.

4 SYSTEM FUNCTIONS

In this section, we outline several interface developments that go beyond previous systems using the virtual human toolkit: User perceived emotion tracking, the "dashboard" observer interface, and mixed initiative, emotion-sensitive dialogue processing.

4.1 Emotion Tracking: the User Perceived State

The *User Perceived State (UPS)* is the metric by which PRIMER tracks the user's emotional state. The UPS is represented in the system as an ordered pair of 2 values: *Valence* and *Arousal*. These values are detected through the use of lexical sentiment analysis, in a process that will be further elaborated below in the section on the Dialogue System.

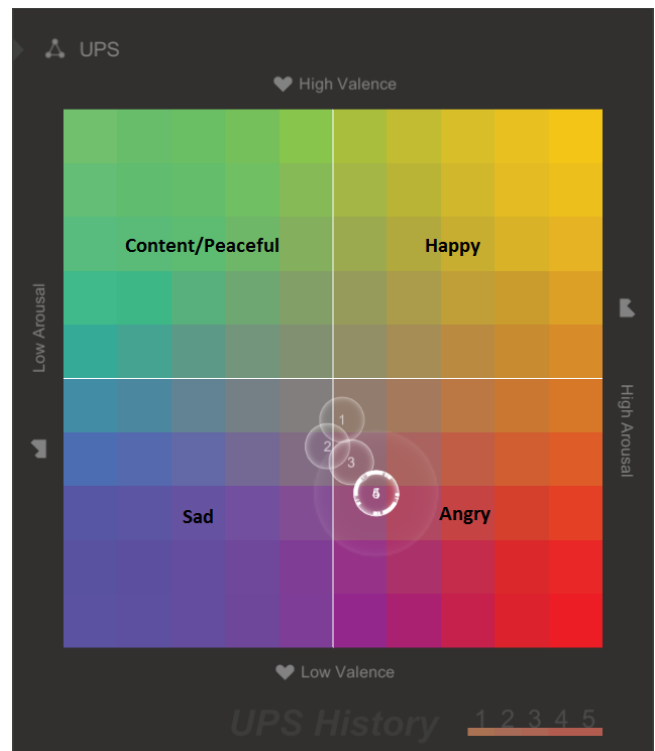


Figure 3: User Perceived State

Valence refers to the overall emotional polarity of a given utterance, which can be positive, negative or neutral. Primer represents this emotional valence as a value between -0.9 (very negative) and 0.9 (very positive), with a valence value of 0.0 being neutral. An example of a positively valenced utterance would be "I'm feeling

great today", a negatively valenced utterance would be "I'm feeling terrible", and a neutrally valenced utterance would be "I'm ok".

Arousal refers to the level of physiological arousal represented by a given utterance. In other words, *Arousal* is a measure of the energy or intensity with which something is said. As with *Valence*, PRIMER represents *Arousal* as a value between -0.9 (low energy) and 0.9 (high energy), with a valence of 0.0 being neutral. While physiological arousal or affect can be difficult to predict from purely text-based input, recent work suggests that lexical markers of arousal do exist, based on psychological word norms [3].

UPS values fall in 1 of the 4 quadrants of the UPS grid, representing a simplified range of human emotions: Happy, Sad, Angry and Content (see Figure 3). In this figure, the circles represent a history of the UPS values for all utterances so far, allowing a third party observer to track the user's emotional state as the dialogue progresses.

PRIMER tracks the UPS at 2 levels: the utterance level, which represents the UPS of a given utterance, and the global level, which is representative of the user's overall emotional state. Global UPS is calculated based on the functions in Equations 1 and 2. It is this tracking of both the Global UPS and utterance level UPS that allows PRIMER to exhibit the kind of emotionally nuanced responses previously mentioned.

$$Arousal_{new} = \alpha * Arousal_{old} + (1 - \alpha) * Arousal_{utt} \quad (1)$$

$$Valence_{new} = \alpha * Valence_{old} + (1 - \alpha) * Valence_{utt} \quad (2)$$

Like [12], we started with a lexical approach to sentiment analysis in order to detect a user's perceived emotional state (UPS). However, the system is designed so that other more sophisticated means of emotional detection could be introduced, and the range of emotions the system is designed to detect could be expanded, allowing for more nuanced responses to the user's perceived emotional state.

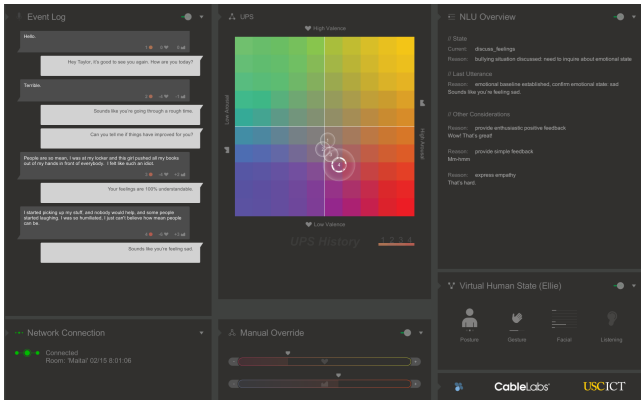


Figure 4: Dashboard

4.2 Graphical User Interface: The Dashboard

In addition to the VR environment described above, PRIMER includes a separate Graphical User Interface, referred to as the "Dashboard" (see Figure 4). The goal in designing the Dashboard visualization was to aid a third party, in this case a counselor or teacher, in understanding the emotional state of the user as well as inform them as to how the system derived that information. It is a visual, easy-to-understand, look behind the curtain of the system. The Dashboard was designed to be monitored in real-time, and allow for real-time adjustments to be made on the fly.



Figure 5: The Event Log, displaying the each dialogue turn and corresponding UPS values

The Dashboard has a modular design that allows for the easy incorporation of components, depending on the desired system. The current dashboard is comprised of five components:

- (1) An Event Log which details the emotional scoring of the user's input (see Figure 5)
- (2) An NLU Overview which provides the reasoning behind the system's state assessment (see Figure 6)
- (3) A display of the UPS; a graphical representation of the user's emotional state over the course of the interaction (see Figure 3)
- (4) A visual representation of the Virtual Human State, such as posture, facial expression blend, gesture intensity, etc. (see Figure 7)

- (5) Manual Override sliders to affect the user's valence and arousal scores (see Figure 8)

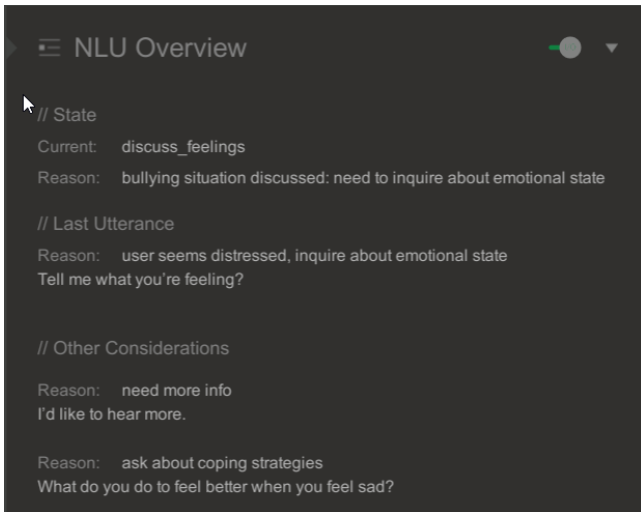


Figure 6: The NLU overview; part of the Dashboard GUI

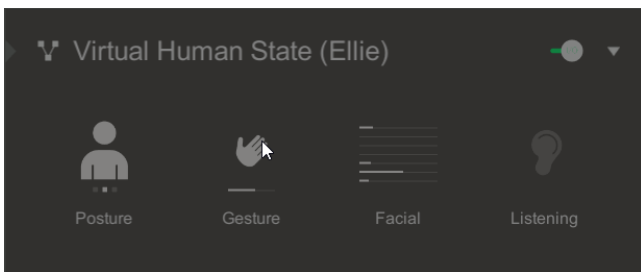


Figure 7: Virtual Human State

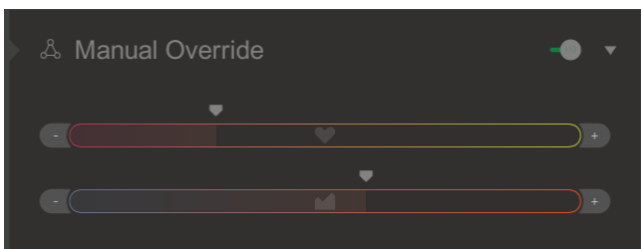


Figure 8: Manual Override

The UPS was designed to be calculated from both the user's input, in this case a bullied child, as well as by manual input, in this case a counselor or teacher. For this initial prototype, it was of particular importance for a trained, experienced clinician to have the ability to include observations of the user's emotional state in the calculation of the UPS.

5 DIALOGUE PROCESSING

The dialogue system for PRIMER was developed using the NPCEditor. As mentioned above, the NPCEditor is a text classification and dialogue management system. It receives the text of the user's utterance transcribed by the ASR, and produces a ranked list of potential responses for each classification domain specified in the training data. NPCEditor is capable of classifying any given response in a number of different classification domains, and for PRIMER these classification domains included two sentiment classifiers (valence and arousal), and a response classifier. Each classifier is trained on a corpus of richly annotated training data. Training data consists of a set of potential user utterances (inputs), which are linked to appropriate system outputs, and may also be annotated with additional information (more on the process of crafting the training data corpus in a later section).

For PRIMER, the three separate text classifiers were trained for each phase of the dialogue using NPCEditor (see section on Dialogue Structure). The first two (sentiment classifiers) use the text content of the user's utterances to establish the utterance's emotional content, i.e., Valence and Arousal. The third classifier (response classifier) ranks all system responses based on their appropriateness to the user's utterance. In this way, for any given user utterance, the NPCEditor produces a ranked list of system outputs for each classifier.

The NPCEditor also incorporates a customizable rule-based dialogue manager (DM). For PRIMER, the DM combines the outputs from all three classifiers to decide which response is the most appropriate to present to the user. The DM models a conversation as a finite state chart with the individual chart states corresponding to dialogue progress. It tracks the user's progress by switching between the states as it detects or initiates shifts in the dialogue. It also tracks the Global UPS by incrementally updating the valence and arousal values using the sentiment classifiers' outputs. The response selection is based on 1) the response appropriateness score from the response classifier; 2) the emotional state of the conversation; 3) the current dialogue state and the set of state-specific rules associated with it.

6 SYSTEM DEVELOPMENT

In addition to the development of the Virtual agent, environment, and interfaces detailed above, the development of the PRIMER system involved content development with subject matter experts, the development of the dialogue structure, and the development of the training data for the classifier and DM.

6.1 Content Development

In order to inform the development of a virtual agent who could be adaptive to the needs of children who have been the victims of bullying, the content development process for PRIMER began with subject matter expert (SME) consultations with a psychologist and school counselor who were familiar with the issue of school bullying. The SMEs provided invaluable insight into the specific issues surrounding providing outreach to children who have been the victims of bullying, and helped inform the training data and dialogue structure of the system.

Speaker	Utterance
Ellie	(1) Sounds like you're feeling angry.
Taylor	(2) I'm not angry, I just don't get it, I just wish things would go back to how they were last year.
Ellie	(3) That's understandable.
Ellie	(4) What do you do to feel better when you feel sad?
Taylor	(5) What do you mean?
Ellie	(6) People do all different things to help themselves feel better. Is there anything you do to make yourself feel better?

Table 1: Excerpt from demo script

The system content for PRIMER is divided into 3 sections: *user utterances*, *system responses*, and *valence and arousal scores*. The core content for this system consisted of a demo script developed to be representative of a typical counseling session between a high school student and a school counselor. This script represents the *user utterances* and the *system responses*. For the purposes of this demo, the script was written to represent the 2nd session between Ellie and Taylor. An excerpt of this script can be found in Table 1.

The *user utterances* consist of a small set of 99 utterances, most of which were hand authored during the script development process to be representative of how a young person might talk to the system. Additionally, some utterances were authored during the system testing process, and tended to be slight variations or rephrasings of the user utterances taken from the demo script. These variations were added to the system content in order to add robustness to the UPS detection. Table 1 shows some examples of the user utterances in the training data.

The *system responses* consist of a set of 97 utterances which were authored during the script development process as responses to the user utterances (refer again to Table 1 for examples). The set of possible system responses was designed to enable Ellie not only to respond appropriately to the user, given their current UPS, but also to allow her to lead the user through the planned dialogue phases, achieving the mixed-initiative dialogue interaction that is so vital to the user experience PRIMER aims to provide. Once the demo script was finalized the *system responses* were recorded, so that Ellie would have a real emotionally nuanced human voice.

The *valence and arousal scores* represent the set of all possible values of valence and arousal which could occur in each dialogue phase, explained in further detail in discussion of Dialogue Structure. This set of valence and arousal values was represented in the dialogue system in the same manner as Ellie's verbal responses, as part of the training data corpus.

6.2 Mixed-Initiative Dialogue Structure

Part of the motivation behind making PRIMER a mixed-initiative dialogue system was to ensure that Ellie could guide the user through a pre-determined set of 7 conversational phases, which would mimic a typical counselling session provided by a school counselor. The goal was to create a semi-structured conversation in which the

user was free to say whatever they would like at any given point, while still enabling the system to exert a measure of control about which topics were being discussed, and the direction in which the interaction progressed. The conversational phases are shown in Figure 9. In each phase, the system tracks the user's UPS, as well as the number of user utterances and the length of those utterances, and there is a certain threshold of emotional and linguistic information that must be reached before it will advance to the next phase. These thresholds differed from phase to phase and were defined in the dialogue manager.

It should be reiterated that for the purposes of this proof-of-concept system, the dialogue structure was designed with the assumption that the system was interacting with a user with whom it had previous interactions. This is apparent in the description of the "Review and Probe" phase below, in which the system will inquire about the user's current emotional state, as compared to previous sessions. A future developmental goal of PRIMER is to create user profiles for each individual user including UPS information from past sessions, in order to build on that information in each following session. This was, however, not implemented in the proof-of-concept.

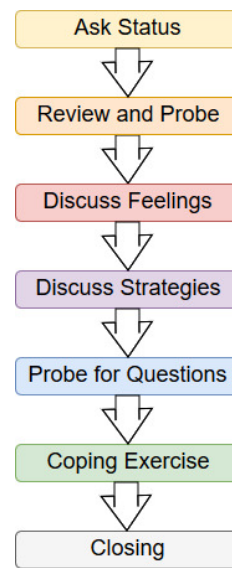


Figure 9: Conversational Phases

Phase 1: Ask Status

In the first phase of the conversation, Ellie attempts to establish a baseline UPS for the user by asking them about their current mood. In this phase, if the user gives only very short, neutrally valenced responses, Ellie will prompt the user for more information. Once the system is satisfied that an emotional baseline has been established, it will prompt an initiative and Ellie will ask the user how they feel compared to last week. This initiative signals the beginning of the Review and Probe phase.

Phase 2: Review and Probe

During the Review and Probe phase, Ellie probes the user to provide more detailed information about their current mood, as compared to the last session. In this phase of the dialogue, once the

user's global UPS has reached a certain threshold in any quadrant, an initiative will be prompted in which Ellie will attempt to confirm with the user what their current emotional state is (see Table 1), line (1). This is done with the intent to encourage the user to discuss their feelings in more detail, and signals the beginning of the Discuss Feelings phase.

Phase 3: Discuss Feelings

In the Discuss Feelings phase, Ellie will continue to prompt the user to talk about their feelings for a few dialogue turns. In this phase the system pays attention to the length of the user's utterances and Ellie will prompt the user for more info if they are only providing very short responses to her inquiries. Once the user has provided a few utterances of sufficient length, a system initiative will be triggered in which Ellie asks the user how they deal with their feelings, signaling the beginning of the Discuss Strategies phase.

Phase 4: Discuss Strategies

In the Discuss Strategies phase, the system attempts to get the user to talk about the strategies they use to cope with their negative feelings. In this phase the system is once again paying attention to the length of the user's utterances, and will continue to prompt the user for more information until it feels it has received sufficient information from the user. At that time, a system initiative will be triggered in which the system asks the user if they have any questions, signaling the beginning of the Probe for Questions phase.

Phase 5: Probe for Questions

The Probe for Questions phase is a very brief phase of the conversation designed to give the user the option to ask Ellie specific questions about coping with negative emotions or with bullying behavior. If the user responses in this phase indicate they have no questions for the system, this will trigger an initiative in which Ellie suggests the user partake in a coping exercise, signaling the beginning of the Coping Exercise phase.

Phase 6: Coping Exercise

In this proof-of-concept version of the system, this particular dialogue phase was not fully implemented, and the system operates under the assumption that a user will respond in the negative when Ellie proposes a coping exercise. The current system could be easily modified to support this interaction during this phase, however at present, the system will be expecting to move through it without such an interaction. In this phase, all user input will prompt simple feedback from Ellie and then trigger an initiative designed to bring the conversation to a close (see Table 2). This signals the beginning of the Closing phase.

Phase 7: Closing

In the closing phase, the system will end the conversation, but encourage the user to come back for further support if needed.

In addition to the conversational initiatives mentioned above, the system was also capable of taking initiative during long periods of silence in which no utterance was received from the user. During periods of silence, Ellie may prompt the user for more information about their previous statement, or make a general inquiry such as "Is there anything you want to tell me?" which is designed to get the user talking again. This feature was designed to keep the interaction going during times when the user may not know what

Speaker	Utterance
Ellie	(1) I recommend we do a roleplay exercise to practice strategies for the next time you're in a similar situation. Okay?
Taylor	(2) Not now. I've gotta go to class.
Ellie	(3) Your parents, teachers, counselors, and I are all here to help you.
Taylor	(4) Ok, bye.
Ellie	(5) Bye for now, I'm here when you need me.

Table 2: Transition between Coping Exercise and Closing phases

to say, and is another way in which the system itself drives the conversation.

6.3 Training Data

Once the initial content development process was done, the demo script was used to create an annotated corpus of training data. Training data for PRIMER consisted of the set of user utterances annotated with valence and arousal scores, and in some cases with appropriate system responses. The valence and arousal scores for each utterance were chosen based on the authors' intuition. In a small number of specific cases, user utterances were linked directly to a system response, but the vast majority of user utterances were annotated solely with information about valence and arousal scores. In this way, the main classifier output for any given user utterance is the ranked lists of arousal and valence scores, as well as a ranked list of potential system responses. The top arousal and valence scores are used to update the UPS according to Equations 1 and 2, respectively. The information provided by these three classifiers is then used by the DM in order to chose appropriate response at any given phase of the dialogue.

In order to facilitate this process, as well as facilitating the mixed-initiative dialogue style, the system utterances were annotated with information about their *type*, *domain*, and whether they signify the beginning of a new conversational phase (*toss*). Examples of system responses and their type and domain annotations can be found in Table 3.

Response Text	Type	Domain
You wanna tell me anything?	probe_for_info	ask_status
That's good.	positive_feedback	discuss_feelings

Table 3: Examples of Type and Domain annotations

Type

The type refers to the general intention of the system response. Table 2 shows the type "probe_for_info", in which the system asks the user for more information, and "positive_feedback" which is related to a response of positive feedback, such as "That's good". By annotating utterances with type information, the DM is able to choose a random utterance from within the set of utterances of a

given type. This provides a more realistic interaction, by ensuring that each time a user interacts with the system, it provides slightly different responses, even to the exact same input.

Domain

The domain is a constraint set on each response as to which phase of the conversation it can appear in. In the examples in Table 2, the first response "You wanna tell me anything?" is constrained to the initial phase of the dialogue "Ask Status" (see Figure 9). These domain annotations enabled the system to provide nuanced responses to similar user utterances, during different phases of the conversation. It should be noted, however, that the same system response could appear in two different domains. In order to achieve this, there would need to be two instances of this response in the training data, each with it's own unique domain annotation. In this way, the system was not limited to a unique set of responses in each conversational phase.

Additionally, a subset of a certain type of utterance could be specified for each given conversational phase. For example, responses of the type `probe_for_info` can be found in both the `ask_status` domain, and the `review_and_probe` domain. The `probe_for_info` response in the `ask_status` domain is very general, such as "You wanna tell me anything?". The `probe_for_info` responses in the `review_and_probe` domain ask the user to elaborate on something they have already said, such as "Can you tell me more about that?". This strategy allowed for the training data to define broader "types" of utterances that can appear in more than one conversational phase, and decisions about which utterances can appear in each phase can be further restrained by using domain annotations.

Toss

Certain system responses were annotated with a "toss", which would indicate to the system that it should enter the next conversational phase. When the dialogue manager detects a toss on a certain response, it will update the internal state tracking, which specifies the particular set of rules for choosing responses within a given conversational phase. This is the mechanism that allows the system to provide different and nuanced responses to the user input in different conversational phases, even if that input is identical to input received in a previous phase.

7 INITIAL TESTING

In initial testing of this demo system PRIMER showed promise at detecting and tracking the user's emotional state, and providing appropriate verbal and non-verbal feedback. Testing of PRIMER using the script developed specifically for the initial proof-of-concept demonstration revealed the system to be capable of appropriate emotional detection and adaptation to the user's emotional state. Further, a broader, albeit informal, test of the system suggests that PRIMER also shows promise in correctly classifying the emotional context of novel utterances, and providing appropriately adaptive feedback as well.

Table 4 shows an excerpt from an "off-script" interaction with the PRIMER system. In this interaction, the majority of the user utterances were not a word-for-word match to any of the training data, yet the system was still capable of having a reasonably coherent conversation, and showing appropriate emotional awareness. The system was also successful at guiding the user through the

Speaker	Utterance
User	(1) Hi.
Ellie	(2) Hello, how are you today?
User	(3) Good.
Ellie	(4) Please continue.
User	(5) There's not that much more to say, it's just a normal day.
Ellie	(6) Can you tell me if things have improved for you?
User	(7) I think things are generally better, yes.
Ellie	(8) Nice!
Ellie	(9) I'd like to hear more.
User	(10) Well, all my exhausting travel is finally behind me.
Ellie	(11) Can you tell me more about that?
User	(12) : Yeah, I've been travelling a lot for work lately.
Ellie	(13) Wow! That's great!
User	(14) : Yeah, it has been a lot of fun, but it is also pretty exhausting to travel so much
Ellie	(15) That's totally normal.
User	(16) Yeah, I guess.
Ellie	(17) Sounds like you're feeling angry.
User	(18) No, I'm not really angry at all.
Ellie	(19) Okay!
Ellie	(20) Can you elaborate on that?
User	(21) Well you said you thought I was feeling angry, I'm definitely not angry.
Ellie	(22) Okay! That's good!.
Ellie	(23) How do you make yourself feel better when you're down?
User	(24) I guess I like to read.
Ellie	(25) Please tell me more.
User	(26) Reading is just a really good way to forget about all the pressures of real life
Ellie	(25) Awesome!

Table 4: An excerpt from an off-script interaction with PRIMER

conversational phases, although it should be noted that the interaction does not represent a novel user with no prior knowledge of the system. This interaction displays the moderate robustness of PRIMER, which can produce a coherent interaction outside of the general domain of bullying, but still within the the general dialogue framework for which it was designed, namely that of a counseling session.

During this interaction, there were not any instances of completely incoherent utterances from the system, however, it does incorrectly gauge the user's global UPS, as evidenced by lines (17) and (18). Here, we can see that the system incorrectly gauged the user's UPS as being in the "Angry" quadrant. However, most of the system's responses are coherent, and display an appropriate emotional valence in response to the user's utterances.

This example also shows how the system leads the user through the conversation by switching between the different conversational

phases. This can be seen in line (6) when the system moves from the *Ask Status* phase into the *Review and Probe* phase by asking the user "Can you tell me if things have improved for you?" The system again leads the conversation in line (17) when it coaxes the user into the *Discuss Feelings* phase by saying "Sounds like you're feeling angry." Although this was not a correct assessment of the user's mood, it quickly recovers from this mistake, continuing to provide appropriately valenced feedback, and soon takes initiative to guide the user into the *Discuss Strategies* phase of the conversation by asking about the user's coping strategies in line (23).

The initial version of PRIMER was given only limited training data, and so was tested mainly on a scripted interaction and some variations rather than a full test with the target population. This limited informal testing did reveal that the purely lexically-based sentiment analysis would probably not be a sufficient means of emotion detection, at least not given the relatively small amount of training data used to train the classifiers for this task. However, for demonstration purposes, lexical sentiment analysis and the small training data corpus was sufficient to provide the necessary emotional awareness, and was also robust enough to handle some off-script interactions as well.

8 FUTURE WORK

The PRIMER system proved to be a successful proof-of-concept system, capable of correctly executing the planned demo, and showing promise as having broader applicability. The next phase of development for the PRIMER system would include:

- **Expanded Training Data and Dialogue Manager:** As the training data was modeled from a script representing the second interaction of user with the agent, we would need transition diagrams and training data to support additional sessions with slightly different interaction plans.
- **Enabling User Profiles:** For a multi-session interaction, we would want to use information gathered from the user in previous sessions (e.g. name, main complaint, prior emotional states and coping strategies), rather than assume it, as was the case from our initial script for session 2. This is evidenced by line (6) in Table 4. Here the system asks if things have "improved", capitalizing on information in the user profile about their emotional state during the last session.
- **Enhanced Emotion Tracking:** In order to make the emotional detection and tracking more robust, PRIMER could be adapted to accommodate additional means of emotional detection and tracking, including audio and visual input from the user. This would allow PRIMER to implement a number of more sophisticated methods such as eye, gaze, and head tracking, voice analysis, and non-verbal behavior [13].
- **Increased Emotional Granularity:** Implementing more advanced methods of emotion tracking would enable the refinement of the currently limited range of emotions PRIMER is capable of detecting. Using the methods mentioned above, PRIMER could be modified to detect and adapt to a far broader range of more subtle human emotional states.

9 CONCLUSION

We have presented PRIMER, a novel user interface for virtual human interactions in virtual reality that are sensitive to the user's emotional state and dialogue behavior, and can be monitored and guided by an external human observer. A proof of concept scenario for counseling victims of bullying was presented in which the agent used mixed initiative dialogue to guide and respond to the user, while a counselor could observe diagnostic information about the system's behavior and the interaction. Results seem promising, though building out a full system will require additional development, training data, and possibly improved techniques.

ACKNOWLEDGMENTS

This research was sponsored by CableLabs. Some of the authors were supported in part by the U.S. Army; 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. Special thanks to our subject matter experts (Todd Adamson PsyD, California School of Professional Psychology and Bill Lemieux, Milwaukee inter-city teacher/counselor), the team responsible for the creation of the virtual agent and VR environment (Jamison Moore, Adam Reilly, Dimitar Tzvetanov, Robert Weaver, Peter Walters, Wendy Whitcup, Joe Yip), and Angela Nazarian (UCSC), the vocal talent who brought life to our virtual human, Ellie.

REFERENCES

- [1] Ruth Aylett, Marco Vala, Pedro Sequeira, and Ana Paiva. 2007. FearNot! – An Emergent Narrative Approach to Virtual Dramas for Anti-bullying Education. *Lecture Notes in Computer Science* 4871 (2007), 202–205. https://doi.org/10.1007/978-3-540-77039-8_19
- [2] Julia C. Campbell, Matthew Jensen Hays, Mark Core, Mike Birth, Matt Bosack, and Richard E. Clark. 2011. Interpersonal and Leadership Skills: Using Virtual Humans to Teach New Officers. In *Proceedings of Interservice/Industry Training, Simulation, and Education Conference (IITSEC) 2011*. IITSEC.
- [3] Houwei Cao, Arman Savran, Ragini Verma, and Ani Nenkova. 2015. Acoustic and lexical representations for affect prediction in spontaneous conversations. *Computer Speech & Language* 29 (January 2015), 203–217. Issue 1. <https://doi.org/10.1016/j.csl.2014.04.002>
- [4] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. IEEE, 1061–1068.
- [5] Arno Hartholt, David Traum, Stacy C. Marsella, Ari Shapiro, Giota Stratou, Anton Leuski, Louis-Philippe Morency, and Jonathan Gratch. 2013. All Together Now: Introducing the Virtual Human Toolkit. In *International Conference on Intelligent Virtual Humans*. Edinburgh, UK. <http://ict.usc.edu/pubs/All%20Together%20Now.pdf>
- [6] Deborah Lessne and Christine Yanez. 2016. Student Reports of Bullying: Results From the 2015 School Crime Supplement to the National Crime Victimization Survey. National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC. (December 2016). Retrieved from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017015>.
- [7] Anton Leuski and David Traum. 2011. NPCEditor: Creating Virtual Human Dialogue Using Information Retrieval Techniques. *AI Magazine* 32, 2 (2011), 42–56.
- [8] Gale Lucas, Jonathan Gratch, Aisha King, and Louise-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100. <https://doi.org/10.1016/j.chb.2014.04.043>
- [9] Gale Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louise-Philippe Morency. 2017. Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. *Frontiers in Robotics and AI* 4 (2017), Issue 51. <https://doi.org/10.3389/frobt.2017.00051>

- [10] Lauren Musu-Gillette, Anlan Zhang, Ke Wang, Jizhi Zhang, and Barbara A. Oudekerk. 2017. Indicators of School Crime and Safety: 2016. National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC. (May 2017). Retrieved from <https://nces.ed.gov/pubs2017/2017064.pdf>.
- [11] Albert 'Skip' Rizzo and Russell Shilling. 2017. Clinical Virtual Reality tools to advance the prevention, assessment, and treatment of PTSD. *European Journal of Psychotraumatology* 8 (2017), 1–20. <https://doi.org/10.1080/20008198.2017.1414560>
- [12] Antonio Roque and David Traum. 2007. A model of compliance and emotion for potentially adversarial dialogue agents. In *Proceedings of the 8th annual SIGDIAL Conference*.
- [13] Marc Schroder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark ter Maat, Gary McKeown, Sathish Pammi, Maja Pantic, Catherine Pelachaud, Bjorn Schuller, Etienne de Sevin, Michel Valstar, and Martin Wollmer. 2011. Building Autonomous Sensitive Artificial Listeners. *IEEE Transactions on Affective Computing* 3 (October 2011), 165–183. Issue 2. <https://doi.org/10.1109/T-AFFC.2011.34>
- [14] William Swartout, David Traum, Ron Artstein, Dan Noren, Paul Debevec, Kerry Bronnenkant, Josh Williams, Anton Leuski, Shrikanth Narayanan, Diane Piepol, H. Chad Lane, Jacquelyn Morie, Priti Aggarwal, Matt Liewer, Jen-Yuan Chiang, Jillian Gerten, Selina Chu, and Kyle White. 2010. Virtual Museum Guides demonstration. In *Proceedings of the 2010 IEEE Spoken Language Technology Workshop*. IEEE. <https://doi.org/10.1109/JPROC.2012.2236291>