Towards More Comprehensive Listening Behavior: Beyond the Bobble Head

Zhiyang Wang, Jina Lee and Stacy Marsella

University of Southern California - Institute for Creative Technologies
12015 Waterfront Drive, Playa Vista, CA 90094
zhiyangw@usc.edu, {jlee, marsella}@ict.usc.edu

Abstract. Realizing effective listening behavior in virtual humans has become a key area of research, especially as research has sought to realize more complex social scenarios involving multiple participants and bystanders. A human listener's nonverbal behavior is conditioned by a variety of factors, from current speaker's behavior to the listener's role and desire to participate in the conversation and unfolding comprehension of the speaker. Similarly, we seek to create virtual humans able to provide feedback based on their participatory goals and their partial understanding of, and reaction to, the relevance of what the speaker is saying as the speaker speaks. Based on a survey of existing psychological literature as well as recent technological advances in recognition and partial understanding of natural language, we describe a model of how to integrate these factors into a virtual human that behaves consistently with these goals. We then discuss how the model is implemented into a virtual human architecture and present an evaluation of behaviors used in the model.

Keywords: artificial intelligence, listener feedback, context based feedback, nonverbal behavior

1 Introduction

Two people are having a heated conversation in a cafe. Around the cafe, various bystanders are listening to the interaction. Some avert their gaze, pretend to do something else, hoping not to become participants in the interaction but nevertheless eavesdropping on the exchange. They are hopelessly drawn to the unfolding scene, glancing at the main protagonists to glean information on the interaction from their dialog and nonverbal behavior, but careful to avoid the mutual gaze that might draw them into the interaction. Meanwhile, the owner of the cafe, wanting to calm the situation, is signaling his intention to join the interaction.

Developing virtual humans that can handle such ranges of participation has become an increasingly important area of research, more so as work has sought to realize more complex dramatic scenarios [16]. Work on listening behavior has tackled various aspects of this challenge. For example, there is work on dyadic interactions between human and *rapport* agents that have an implicit, fixed goal of establishing rapport but often have limited understanding of the content of the speaker's utterance [13]. The agents rather rely on low level analysis of the nonverbal and perceptual features of the human speaker's behavior that are correlated with listener feedback, such as pauses in the speaker's utterance.

Although effective in establishing rapport, this approach suffers from several limitations. First, such approaches only provide *generic feedback* [3] signaling such factors that the agent is attending. They cannot provide *specific feedback* [3], feedback tied to a deeper understanding of, and reaction to, the personal relevance of what the speaker is saying as the utterance unfolds. Another limitation is the fixed, implicit goal of establishing rapport. In practice, however, people can have very different kinds of stances towards a conversation, including even their lack of interest in understanding the speaker or a desire to leave the conversation. One approach to addressing this limitation is to have the listener's behavior be conditional on attitudinal factors [4]. Finally, the focus for listening behavior has been largely on dyadic conversations, where the listener agent is main and sole addressee, though there have been notable exceptions [18].

In this work, our interest is to realize this richer form of interaction in a multiparty setting where there may be several virtual humans interacting with one or more humans, playing a variety of roles (e.g. main addressee, side-participants, overhearer, bystander, etc.) with varying degrees of participation in, and commitment to, the conversation. The question that interests us is how these characters respond nonverbally according to their current role in the conversation, their desire to participate, their understanding of the speaker's partial utterance, as well as behavioral signals from the speaker.

This raises technical challenges of how to integrate the various factors that influence a listener, including the perception of the speaker's verbal/nonverbal behavior as well as the listener's reactions to the speaker in light of their goals for participation. In this paper, we review the relevant literature on listener feedback and propose a model that tailors behaviors based on how the various roles of participants influence their nonverbal behaviors and how those behaviors can signal their goals to change roles. To provide both generic and specific feedback, the model integrates information from perceptual and comprehension processes. We then discuss how the model is implemented into a virtual human architecture, relying on prior work to provide perceptual processing of the nonverbal and prosodic features of speaker behavior [26] as well as to provide natural language understanding of a speaker's partial utterance [7] and emotional reaction [25] to it. Finally, we present a preliminary evaluation of behavioral signals used in the model and discuss future directions.

2 Related Work

Listener's feedback [28] has been studied both in social science and humanities research on human behavior as well as in technology work on the design of virtual agents. This section discusses the virtual agent work. Literature on human behavior that has informed this work is discussed and referenced in subsequent sections.

Research on listening behavior for virtual agents has largely focused on dyadic interactions between virtual agent and human, where the virtual agent is the main addressee. The Rapport Agent created by Gratch et al. [13] provides listening feedback based on the nonverbal and prosodic features of the speaker's behavior, such as head movements, body postures, and vocal pitch and intensities. They demonstrated that mimicry of the speaker's behavior, including head movements and gaze aversion, improves the human speaker's sense of rapport and speech fluency. Morency et al. [26] learned probabilistic models that predict listener's nonverbal feedback from the human speaker's multimodal output features (e.g., prosody, spoken words and eye gaze).

Because such designs are driven by the speaker's behavior and more importantly do not incorporate the listener's (virtual human) interpretation and reaction to the utterance, they are arguably more important for generic feedback as opposed to specific feedback [3]. To drive listener's specific backchannel behaviors, the virtual agent needs to interpret the utterances and generate feedback based on personal relevance, as the human speaker's utterance is in progress. Research has sought to address this technological challenge in several ways. Jónsdóttir et al. [19] collected human listeners' feedback data, summarized a set of speaker's key phrases in a limited topic domain, and built a system to generate virtual listener's feedbacks when input utterance match those lexical feedback markers (key phrases). Kopp et al. [21] designed an event-based feedback model for their virtual agent Max. The model generates listener's feedback by using multimodal information including the speaker's pauses and lexical information. DeVault et al. [7] used a classifier to classify partial utterances in terms of semantic frames that the agent understands.

In addition to such work on dyadic conversation, there also has been work in multiparty conversation. Jan and Traum [18] involves movement for modeling agents' participation restriction with group conversation. They developed a social force model to control the distance between the agent and the group center. The force would push two agents apart if they were over close to each other, while the virtual bystander may be dragged towards group if they were outside of the circular participation domain.

In contrast to prior work, the focus of this paper is on a model for generating listener nonverbal feedbacks for multiparty conversations that includes both generic and specific feedback, as well as taking into account that there may be a variety of participants with varying roles and goals for their participation.

3 Conversational Roles and Goals

In this section we discuss the relationships between conversation roles and goals which we later use when mapping listener feedback behaviors to in our model. First we define the various conversation roles by adopting the terminology used by Goffman [11]. In a conversation, the core participants are the speaker and nonspeaking participants (ratified participants), which includes the *addressee* ("addressed recipient") and the *side-participants* ("unaddressed recipients"). In addition, unofficial-participants are called *bystanders*. Goffman identifies two types of bystanders: *eavesdroppers,* who purposefully listen to the conversation, and *overhearers,* who accidentally and unintentionally hear the conversation. However, these conversation roles are not static and can change during social interaction [15, 27].

We can characterize these various roles from the perspective of the goals that the role normatively presumes. Here we define two types of conversation goals: *participation goal* and *comprehension goal*. Since addressees and side-participants are part of the core conversation participants, they hold positive participation goals and to maintain this status they must act appropriately. However, bystanders (overhearers and eavesdroppers) normatively have a negative participation goal (i.e. they are not or do not want to be perceived as participants) and should act in ways that do not increase their level of participation. The conversation roles can also be further distinguish based on the comprehension goals. Eavesdroppers clearly have stronger intentions to understand the conversation, whereas overhearers do not intend to comprehend the conversation. In contrast, addressees and side participants are expected to have positive comprehension goals and to behave consistently.

We can then summarize the relationships between conversation roles and goals as the following. Addressees have positive participation and comprehension goals; side-participants have positive participation goal and either positive or negative comprehension goal; eavesdroppers have negative participation goal but positive comprehension goal; overhearers have both negative participation and comprehension goals.

Several aspects of this classification must be stressed. First, we assume that all of the agents, regardless of their roles, have freedom to change their participation or comprehension goals. For example, although side-participants are part of the conversation group, they may want to leave the conversation at any time. Second, there is a distinction between having a goal and openly appearing (or signaling) that one has a goal. For instance, eavesdroppers may wish to comprehend the conversation (and thus have a positive comprehension goal), but because they do not want to *participate* in the conversation, it is important to appear so since they could be drawn into the conversation and endanger their role as eavesdroppers. Third, these goals are the norm for the roles. For example, side-participants are presumed to be committed to participate and comprehend the conversation and should act consistently, but in reality they may not be concerned with understanding the context of the conversation. For this reason, it is important to consider the individual's goals for participation and comprehension distinct from the role, since the realization of behaviors may depend on both. In this paper we simplify the goals to have a binary value (positive or negative), but one can also imagine the goals having numerical values to specify the degrees of participation or comprehension.

4 Modeling the Impact of Roles and Goals on Behavior

The literature describes various listening behaviors depending on the conversation roles and goals, which we use to inform the knowledge used in our model. Table 1 categorizes the knowledge currently used in the model. The behaviors are categorized according to the agent's goal to participate in the conversation and its desire to comprehend the speech content. In this section, we discuss that knowledge and in the next section we cover how that knowledge is used by the model.

For addressees, we use gaze and mutual gaze to signal goals of participation and comprehension as well as continued attention [1]. This also helps addressees to get clearer visual and vocal information from the speaker [20]. Addressees also glance at other side-participants to seek social comparison [9] or avert gaze as a signal of cognitive overload when comprehending speech [1, 12]. In addition, we use various forms of nodding behaviors to signal that the addressee is attending [26], comprehending [5, 8] or reacting to the speaker [17] and thereby to signal participation and comprehension.

Table 1. Relationship between conversation goals, roles, and listener feedbacks

Conversation Goals		**Conversation Roles**	**Rule and Behavior**
Participating	Not specified comprehending	Addressee or Side-participant	*Attendance*: gaze speaker [1, 14] and head nod [26].
			Mimicry: mimic gaze direction: listener mimics speaker's gaze direction when the speaker has gazed away for a long period. Mimic head gesture: Listener repeats speaker's shaking or nodding behavior. [24]
		Switch from Eavesdropper/Overhearer to Addressee/Side-participant	*Enter group*: decrease distance by moving towards the group [18]
		Addressee or Side-participant	*Respond feedback request*: respond to other participant's communication request by gazing at the speaker [17]; Glance at speaker to indicate continued attention and willingness [1]
			Mirror Emotion: adjust its own emotion to group's emotion status [10]
	Comprehending	Addressee or Side-participant	*Understand*: head nod [5, 8]
			Think: gaze aversion [1]
			Gather Information (addressee/side-participant): glance at speaker to study speaker's facial expressions and direction of gaze [1, 20] or generate social comparison behavior [9].
			Confusion: head tilt and frown [5]
			Emotion reaction: different head movement, gaze behavior and facial expression according to different emotion types.
Not Participating		Eavesdropper	*Gather Information(eavesdropper)*: glance at speaker but with faster speed and less magnitude [1] and avoid mutual gaze [2]. Show less reaction [9] .
	Not Comprehending	Overhearer	*Avoid mutual gaze*: gaze aversion [6, 11]
		Switch from Addressee/Side-participant to overhearer	*Leave group*: increase distance by moving away from group [18]

On the other hand, head tilts and frowns are used to signal confusion [5] and various facial expressions are shown to signal emotional reactions to the content of the speech.

Side-participants are also ratified by the speaker and exhibit similar behaviors as addressees. However, they may be less committed to comprehend the current dialog. If side-participants do not care about understanding the speaker's utterance (i.e. comprehension goal is negative) but the goal is to maintain the participation status, they use glances toward the speaker [1, 14]. The glances here are not to further comprehend but rather to act as a ratified participant. Mimicking or mirroring the speaker's behavior [10, 24] are also exhibited to hold his/her current conversation role.

Eavesdroppers have the goal to understand the conversation but their status as anonymous eavesdroppers may be threatened if they openly signal their comprehension. Thus, they avoid mutual gaze and restrain from showing reactions to the conversation [9]. Furtive glances at the speaker are occasionally used for better comprehension but gaze is quickly averted to prevent providing visual feedback [2] and signs of attention to the speaker [1, 2, 20].

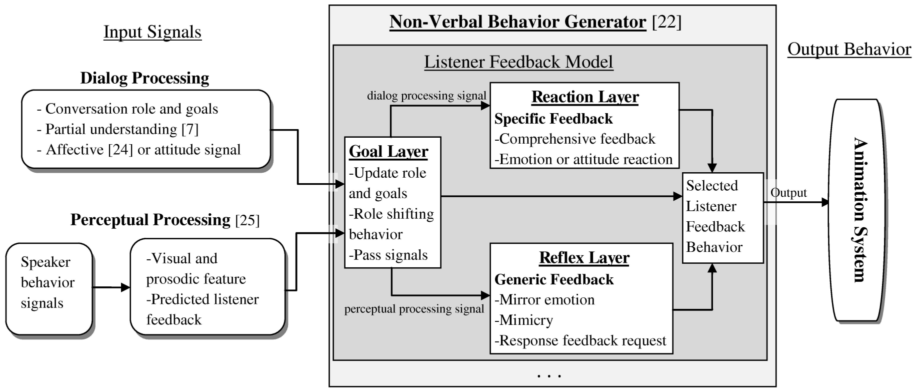

Fig. 1. Architecture of the Listener Feedback Model

Overhearers have neither goals for participation nor comprehension and have fewer concerns about the conversation. Gaze aversion from conversation participants is used to prevent mutual gaze [6, 11] since gaze may be considered as a request signal to be included into the current conversation [1]. However, in a highly dynamic conversation, an overhearer may have difficulty avoiding comprehension of, and reactions, to the conversation.

In addition to the behaviors associated with the conversation roles, there are behaviors associated with role shifts. To signal a change in the conversation role, behaviors associated with the current role are avoided and those associated with the new role are adopted. For example, gazing at the speaker and making mutual gaze signal role shifting from a bystander to a side-participant or an addressee [1, 11]. To shift from an overhearer to an eavesdropper, increased glances at the speaker is adopted to show desires for better comprehension. When the role shift involves changes in the participation goal, interpersonal distance is also adjusted by either moving toward or away from the group to join or leave the conversation [18].

Finally, note that we have not discussed the varieties of turn-taking behaviors associated with the participant seizing the dialog turn or a speaker relinquishing his role as speaker. Such behaviors are more common components of virtual human systems so we have not discussed them here.

5 The Listener Feedback Model

Based on the listener behaviors categorized in the previous section, we constructed the listener feedback model as a set of feedback rules and incorporated them as an extension to the Nonverbal Behavior Generator [23], the behavior planner of our virtual human system. Figure 1 represents the architecture of our model. In this model, we make a distinction between generic feedback and specific feedback as described in section 1. Given the agent's conversation roles and goals, the Goal Layer updates this information and propagates the input signals to the Reaction layer or the Reflex layer that is responsible for generating specific feedbacks or generic feedbacks, respectively. Each layer contains a set of listener feedback rules (see Table1) that gets triggered to generate the behaviors. The following sections discuss the details of the input signals and the different layers in the model.

5.1 Inputs

The listener feedback model receives signals from the virtual human system's cognitive module, broadly classified as dialog processing and perceptual processing signals, as shown in Figure 1. The Dialog Processing signal provides (a) the virtual human's current conversational role as well as participation and comprehension goals, (b) incremental partial understanding information and (c) affective or attitude signal. The conversational role and goals are sent by the virtual human's dialogue module at the start of the conversation and are updated as the interaction between participants unfold. The listener's incremental interpretation of partial utterances is realized by DeVault et al.'s classifier [7], which provides a semantic interpretation as well as a measure how confident the agent is of their partial understanding and a measure of whether the agent believes it will understand better if it continues listening. The affective signal comes from the system's emotion model [25], which predicts the agent's emotional state through an appraisal process of the interpretation of the current interaction and its relationship to the environment.

The Perceptual Processing signal is provided by the virtual human's perceptual model which includes information about the *speaker's* behaviors such as the head movements, gaze direction, pitch accents, and speech pauses. It also includes predictions of the listener's backchannel nods, based on the work of Morency et al. [26].

5.2 Layers of the Model

Upon receiving the input signals, the Goal Layer first updates the agent's role and goals and determine whether to generate a role shifting behavior or to pass the incoming signals to the Reaction and Reflex Layers. The role shifting behavior occurs when the agent's updated participation goal differs from the current participation goal. For example, if the agent's current role is *overhearer* (participation goal=negative) and the updated role is *addressee* (participation goal=positive), he will enter the conversation group and generate attendance behavior by gazing at the speaker and nodding. The role shifting behaviors refer to rule*Enter group* and *Leave group* in Table 1.

If the agent's participation goal is unchanged, the input messages are passed to the Reaction and Reflex Layers and corresponding feedback behaviors are generated depending on the comprehension goal. In particular, the dialog processing signals are passed to the Reaction Layer and the perceptual processing signals are passed to the Reflex Layer. In our model, both layers are active, generating feedbacks concurrently. However, one might instead argue for a more complex interaction. For example, once the partial understanding has achieved high confidence, the reaction layer may dominate the reflex layer.

The Reflex Layer generates generic feedback behaviors when the agent's participation goal is positive but the comprehension goal is negative or unknown. It processes the speaker's perceptual information and generates behaviors such as gazing at the speaker

Table 2. Selection of Comprehension Feedback Rules

<table>
<tr><td colspan="3"></td><th colspan="3">Feedback Rules with Different Roles</th></tr>
<tr><td rowspan="8">Input</td><th>Confidence</th><th>maxf</th><th>addressee/side-participant</th><th>eavesdropper</th><th>overhearer</th></tr>
<tr><td rowspan="2">[0.0, 0.5)</td><td>0</td><td>Confusion</td><td rowspan="2">Idle</td><td rowspan="6">Idle</td></tr>
<tr><td>1</td><td>Attendance</td></tr>
<tr><td rowspan="2">[0.5, 1.0)</td><td>0</td><td>Partial Understand/ Think/ Idle</td><td>Gather Info.(eavesdropper)</td></tr>
<tr><td>1</td><td>Partial Understand/ Think/ Attendance/
Gather-Info.(addressee/side-participant)</td><td rowspan="3">Idle</td></tr>
<tr><td rowspan="2">1.0</td><td>0</td><td rowspan="2">Understand</td></tr>
<tr><td>1</td></tr>
</table>

or mimicking the speaker's gaze direction and facial expressions, in addition to propagating the listener head nods predicted by the visual module. This layer triggers listener feedback rules *Attendance, Respond feedback request*, and *Mirror Emotion* in Table 1.

The Reaction Layer processes affective or attitudinal information as well as the comprehension information. In this model, the agent's emotional reaction is stronger than the reactions related to the partial understanding of the speaker's utterance, therefore any incoming affective or attitudinal signal will have higher priority than the comprehension information. The affective reactions include behaviors such as smiles for joy and furrowed eyebrows for anger (rule *Emotion reaction*).

The comprehension information contains two parameter values: *confidence* (range [0.0, 1.0]) and *maxf* (0 or 1). The confidence value indicates how confident the agent believes it understands the speaker's utterance and maxf indicates whether the agent believes it will understand the utterance more if it keeps listening. We define three categories of understanding based on the confidence value: confusion ([0.0, 0.5)), partial understanding ([0.5, 1.0)), understand (1.0). The *maxf* value further determines which specific feedback is generated. Table 2 shows how the *confidence* and *maxf* values determine the selection of listener feedback rules.

6 Example

We now go through an example scenario to demonstrate how the listener's nonverbal backchannel behaviors are generated. Note that in the current state the mapping between conversation roles, goals, and feedback behaviors have been implemented in our virtual human system whereas the incoming signals from the external modules have not yet been fully integrated. Therefore, here we emulate the incoming dialog processing and perceptual processing signals to demonstrate the model.

In this example, a human user plays the role of a ranger (Ranger) and tries to convince Utah, the local bartender, to take his place as the new sheriff. Utah is in favor of this offer and initially shows positive reactions, but then becomes concerned about Harmony's reactions, who is the saloon owner. On the other hand, Harmony hears the conversation first as an overhearer and shows negative reactions to Ranger's proposal then switches her role to a side-participant to join the conversation.

Below we take an excerpt from the scenario when Ranger offers the job to Utah and describe the input signals and corresponding output behaviors along seven different

Table 3. Feedback Behaviors for the Example Utterance. Here we list the feedback rules for each feedback points. Refer to Table 1 for the specific behaviors. “Idle” indicates idle behavior.

Index	Input Signal				Output Feedback			
	Perceptual	**Dialog**			**Utah**		**Harmony**	
	(Common)	(Common)	(Utah)	(Harmony)	**Generic**	**Specific**	**Generic**	**Specific**
①	Predict: nod prob.(0.6)	PU: maxf(0), confidence(1.0)	Role(P,C): A(1,1)	Role(P,C): O(0,0)	Response feedback request	Role shifting: Enter group, Attendance	-	Idle
②	-	PU: maxf(1), confidence(1.0)	-	-	-	Understand	-	Idle
③	-	PU:maxf(0), confidence([0.5,1.0))	-	Role(P,C): E(0,1)	-	Partial Understand/ Think/ Idle	-	Role shifting: Avoid Mutual Gaze
④	-	PU: maxf(1), confidence(1.0)	-	-	-	Understand	-	Idle
⑤	-	PU: maxf(1) confidence(1.0)	Affective: surprise	Affective: surprise	-	Emotion (surprise)	-	Idle
⑥	-	PU: maxf(1) confidence(1.0),	-	Role(P,C): SP(1,1)	-	Understand	-	Role shifting: Enter Group, Attendance
⑦	-	PU: maxf(1) confidence(1.0)	Attitude: like	Attitude: dislike	-	Attitude: (like)	-	Attitude: (dislike)

P-Participation Goal; C-Comprehension Goal; 1-positive; 0-negative; PU-Partial Understanding A-Addressee; SP-Side Participant; E-Eavesdropper; O-Overhearer

points in the utterance. We represent the agent’s roles and goals as *“Role(participation goal, comprehension goal).”* For example, *“Eavesdropper(0,1)”* denotes that the role is eavesdropper with negative participation goal and positive comprehension goal. Table 3 presents the feedbacks according to different input signals for each agent. The columns are the index for the seven points, input signals and output feedback.

Ranger (Speaker):
“Utah①, it’s time for me to move on② and the③ town will need a strong leader④ like yourself⑤ to⑥ maintain law and order⑦.”

From Table 3 we can see that even with the same input perceptual and partial understanding signals, the agent’s feedbacks are significantly different according to the different conversation roles and goals. This example demonstrates that the feedback model enables the agents with a rich set of reactions that go beyond simply establishing rapport.

7 Behavior Assessments

A key question is whether people can interpret, or decode, the behaviors the model employs, especially the behaviors related to the comprehending goal posited in the model related to the comprehension goal: Gathering Information, Eavesdropping, Thinking, Understand and Confusion. As opposed to the decoding of emotional states which has been extensively studied, there is less evidence that the behaviors we posit for these states can be effectively decoded. If the behaviors are highly ambiguous to observers,

Table 4. Behavior Assessments Results

Think ① and ②: gaze aversion with different direction, magnitude, speed and duration.
Gather Information(ratified participant)①: glance between speaker's head and chest; ②: glance between speaker's head and lumbar.

Rule/Behavior	**Interpretation**					**Recog. Rate**
	Confusion	Think	Gather-Info.	Eavesdrop	Understand	
Confusion/Head Tilt & Frown	**11**	2	1	0	1	73.33%
Think/Gaze Avert ①	2	**11**	0	0	2	73.33%
Think/Gaze Avert ②	1	**11**	2	1	0	73.33%
Gather-Info(ratified)/Scan①	0	1	**13**	1	0	86.67%
Gather-Info(ratified)/ Scan②	0	4	**10**	0	1	66.67%
Gather-Info(eavesdropper)/ Glance at speaker	0	1	3	**10**	1	66.67%
Understand/Nod	1	0	0	0	**14**	93.33%

it undermines the rationale for employing the partial understanding component of the model.

To do a preliminary assessment of this, we created seven video clips of virtual listener nonverbal feedback, based on the rules and behaviors listed in the "Comprehending" signal row of Table 1. In each video, there is a hidden speaker (behind the camera) talking to a virtual human in front of the camera who provides nonverbal feedback (e.g. head nods, facial expressions, etc.) to the speaker. Each subject watched all seven videos. The speech played in the background is the same for each video, while the agent's behaviors were different. The speech is gibberish (nonsense content), so the subject is not influenced by the utterance content itself. After watching each video, the subject was given a forced choice questionnaire that asked them to select the best interpretation from a list of the alternative comprehension goals [1]. We recruited 15 subjects to participate in the experiment. Table 4 shows the results. The rows are the rules and behavior exhibited in the video and the columns are the subject's interpretation of the behavior with each cell listing how many subjects picked that interpretation. The hypothesized interpretation is in bold.

The result shows that for every category, the dominant choice was the hypothesized interpretation. However, some behaviors clearly could be improved if our goal was to reduce decoding ambiguity further. Of course, this is an assessment of just one aspect of the design. We discuss additional evaluation goals in the next section.

8 Conclusion and Future Work

In this paper, we have described the Listener Feedback Model for virtual agents in multi-party conversations. The vision behind this model is that the agent will generate both generic feedback and specific feedback conditioned on a variety of factors, including the speaker's behavior, the listener's role and the desire to participate in the conversation as well as the unfolding comprehension of partial utterances. The model has

[1] The forced choice obviously simplifies this decoding task for the observer but the use of gibberish makes it harder.

been implemented within the nonverbal behavior generation component of our virtual human system and drives the agent to perform feedback automatically and dynamically.

This work will be extended in several ways. A range of extensions to the model are being considered. In particular, we are interested in incorporating other factors which may influence listener's feedback, such as interpersonal relationship, personality, and culture. There are alternative ways in achieving this; the current listener feedback rules could be further added to and modified according to the varying factors or a data-driven approach (e.g., [22]) could be employed to learn models using different sets of data reflecting variations of those factors. Also, as mentioned earlier, there are alternative approaches to how the reactive and reflex layers interact that need to be assessed.

One pressing empirical question concerns how the specific feedback influences the human-virtual human interaction. There have been studies looking at the impact of the generic feedback of rapport agents, but the kind of specific feedback we are discussing here may have a more profound impact. The feedback might facilitate the interaction, providing the human with important information to guide the interaction. On the other hand, the virtual human's reaction to its partial understanding of the utterance, such as a look of anger, could also conceivably cause pauses or disfluency in the human speaker. This in turn may well throw off speech recognition/natural language understanding, thereby impacting the virtual human's ability to recognize and understand the utterance. Regardless, we expect the feedback to to impact the human user's impression of, and expectations about, the virtual human as well as impact potentially a range of relational factors such as trust. Overall, the design of the virtual human may have to fundamentally change to take into account this finer grain interactivity.

Acknowledgments The authors would like to thank Dr. Louis-Philippe Morency and Dr. David DeVault for providing technical assistance to this work. This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

1. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press (1976)
2. Argyle, M., Lalljee, M., Cook, M.: The effects of visibility on interaction in a dyad. Human Relations 21, 3–17 (1968)
3. Bavelas, J.B., Coates, L., Johnson, T.: Listeners as co-narrators. Journal of Personality and Social Psychology 79, 941–952 (2000)
4. Bevacqua, E., Pammi, S., Hyniewska, S.J., Schroder, M., Pelachaud, C.: Multimodal backchannels for embodied conversational agents. In: Proceedings of the 10th Int. Conf. on Intelligent Virtual Agents. pp. 194–200 (2010)
5. Brunner, L.: Smiles can be back channels. Journal of Personality and Social Psychology 37(5), 728–734 (1979)
6. Callan, H., Chance, M., Pitcairn, T.: Attention and advertence in human groups. Social Science Inform. 12, 27–41 (1973)
7. DeVault, D., Sagae, K., Traum, D.: Incremental interpretation and prediction of utterance meaning for interactive dialogue. Dialogue & Discourse 2(1), 143–170 (2011)

8. Dittmann, A., Llewellyn, L.: Relationship between vocalizations and head nods as listener responses. Journal of Personality and Social Psychology 9, 79–84 (1968)
9. Ellsworth, P., Friedman, H., Perlick, D., Hoyt, M.: Some effects of gaze on subjects motivated to seek or to avoid social comparison. Journal of Experimental Social Pscyhology 14, 69–87 (1978)
10. Friedman, H.S., Riggio, R.E.: Effect of individual differences in non-verbal expressiveness on transmission of emotion. Journal of Nonverbal Behavior 6(2), 96–104 (1981)
11. Goffman, E.: Forms of Talk. University of Pennsylvania Press, Philadelphia (1981)
12. Goodwin, C.: Conversational organization: interaction between speakers and hearers. Academic Press, NY (1981)
13. Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R.J., Morency, L.P.: Virtual rapport. In: Proceedings of the 6th Int. Conf. on Intelligent Virtual Agent. pp. 14–27 (2006)
14. Gu, E., Badler, N.: Visual attention and eye gaze during multipartite conversations with distractions. In: Proceedings of the 6th Int. Conf. on Intelligent Virtual Agents (2006)
15. Hanks, W.F.: Language and Communicative Practices. Westview Press (1996)
16. Hartholt, A., Gratch, J., Weiss, L., Team, T.G.: At the virtual frontier: Introducing gunslinger, a multi-character, mixed-reality, story-driven experience. In: Proc. of the 9th Int. Conf. on Intelligent Virtual Agents, pp. 500–501. Springer Berlin / Heidelberg (2009)
17. Ikeda, K.: Triadic exchange pattern in multiparty communication: A case study of conversational narrative among friends. Language and culture 30(2), 53–65 (2009)
18. Jan, D., Traum, D.R.: Dynamic movement and positioning of embodied agents in multiparty conversations. In: Proc. of the 6th Int. conference on Autonomous agents and multiagent systems. pp. 59–66 (2007)
19. Jónsdóttir, G.R., Gratch, J., Fast, E., Thórission, K.R.: Fluid semantic back-channel feedback in dialogue: Challenges and progress. In: Proc. of the 7th Int. Conf. on Intelligent Virtual Agents (2007)
20. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge University Press, Cambridge (1990)
21. Kopp, S., Allwood, J., Grammer, K., Ahlsen, E., Stocksmeier, T.: Modeling embodied feedback with virtual humans. In: Modeling Communication with Robots and Virtual Humans, vol. 4930, pp. 18–37. Springer Berlin / Heidelberg (2008)
22. Lee, J., Marsella, S.: Predicting speaker head nods and the effects of affective information. IEEE Transactions on Multimedia 12(6), 552 –562 (October 2010)
23. Lee, J., Marsella, S.: Nonverbal behavior generator for embodied conversational agents. In: Proc. of the 6th Int. Conf. on Intelligent Virtual Agents. pp. 243–255 (2006)
24. Maatman, R., Gratch, J., Marsella, S.: Natural behavior of a listening agent. In: Proc. of the 5th Int. Conf. on Intelligent Virtual Agents. pp. 25–36 (2005)
25. Marsella, S., Gratch, J.: EMA: A process model of appraisal dynamics. Cognitive Systems Research 10(1), 70 – 90 (2009)
26. Morency, L.P., de Kok, I., Gratch, J.: A probabilistic multimodal approach for predicting listener backchannels. In: Proc. of the 8th Int. Conf. on Intelligent Virtual Agents. pp. 70–84 (2008)
27. Vertegaa, R., der Veer, G.C.V., Vons, H.: Effects of gaze on multiparty mediated communication. In: Proc. of Graphics Interface. pp. 95–102 (2000)
28. Yngve, V.: On getting a word in edgewise. Papers from the 6th regional meeting pp. 567–578 (April 1970)