Differences in Listener Responses between Procedural and Narrative Tasks

Iwan de Kok
Human Media Interaction
University of Twente
Enschede, The Netherlands
koki@ewi.utwente.nl

Dirk Heylen
Human Media Interaction
University of Twente
Enschede, The Netherlands
heylen@ewi.utwente.nl

ABSTRACT

In the long tradition of corpus based research on listener behavior, whether it entails linguistic analysis or social signal processing, many different tasks have been used during the recording of the corpus. So far in no study the task which has been given to the participants has been an independent variable and no studies have looked into the effect of this variable on listener responses. In this paper we present the results of our comparison between listening behavior elicited by procedural and narrative tasks which were used during the recording of our MultiLis corpus. We will show that listeners in the procedural tasks show more agreement in their responses than listeners in the narrative tasks. Furthermore we will show that the long procedural task elicits more responses per minute than the short procedural task. We will reflect on these results in light of cognitive load and grounding theory.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Discourse*

General Terms

Human Factors, Theory

1. INTRODUCTION

In face-to-face conversations and also on the telephone, listeners respond to what the speaker is saying through backchannnel behaviours such as head nods and shakes or vocalisations such as "hmm", "a-ha", and "yeah", through facial expressions that show understanding, surprise or interest in what is being said. They may also respond by repeating a word the speaker just said, or by filling in a word the speaker is looking for. All these kinds of listener response may fulfill various functions. They show that the listener is attending and perceiving the message, has problems or no problems in understanding what is said and they may provide additional information about what the listener's thoughts are about what is being said: agreement or disagreement, liking or disliking, etcetera. On the one hand, the listener responses serve a "structural role", ensuring that the conversation goes on smoothly, by affirming the proper reception of the incoming messages. On the other hand, the responses have a "ritual role", indicating rapport and affiliation or the lack of these. Listener responses are thus not just functioning on the micro-level of conversational organisation, but play an important role with respect to the more social dimensions involved in interaction.

We have been studying listener responses for some years now, both in the context of social signal analysis of multimodal corpora and in the context of the synthesis of backchannels for Embodied Conversational Agents (e.g. [8, 9, 17, 19]). Recently, we completed the collection of a unique corpus specifically designed to study listener responses. The MultiLis corpus, which is described in more detail in Section 3, was recorded to gain more insight into a number of aspects of listener responses.

One of the aspects concerns the role of the individual. People vary in the amount and type of listener responses they provide in a particular context. This may be dependent on their personality, the mood they are in, or their engagement in the conversation. When one wants to study, the timing and placement of backchannels it is important to take this into account. Corpora of one-on-one conversations traditionally include recordings of the responses of one listener in a certain context. Since providing a listener response at a certain moment is optional (according to the definition by Ward and Tsukahara [21]), the moments on which the recorded listener responded to the speaker are not the only moments a listener can respond, nor will all listeners respond at the same moments in the same context. The unique design of the MultiLis corpus provides more insight in this variation that occurs between listeners.

A second aspect that our MultiLis corpus allows one to investigate is the role of some contextual variables. We systematically varied the task that was given to the participants involved in our data collection study. Listeners may tend to respond differently when they are getting instructions as opposed to being told a story. The complexity of the task may also be another factor. The focus of this paper will be this second aspect. We will look at the influence of a procedural or a narrative task on the nonverbal listener behavior and will analyze the speakers actions to find an explanation.

An overview of the different tasks used in other corpora on which research on nonverbal listening behavior has been performed is given in Section 2. In Section 3 the MultiLis Corpus is explained in more detail. The results of our comparison between listening behavior elicited by the narrative and procedural tasks in our corpus are presented and discussed in Section 4.

2. RELATED WORK

Research on nonverbal listening behavior has a long tradition, starting in the sixties with works from e.g. Dittmann et al. [4] and Yngve [24]. Over the years the studies have been performed on various corpora, each with their own characteristics. The corpora varied in gender and culture of the partipicants, the medium through which the conversation took place (e.g. face-to-face, over the telephone, through videoconferencing) and the task the participants received. Several studies have been performed on the influence of gender (e.g. [2, 5, 10, 14]), culture (e.g. [7, 15, 21]) and medium (e.g. [18]), but we are not aware of any study which compares the influence of the task variable, even though different choices were made during the recording of the corpus.

The tasks for the participants in the corpora used for research to listening behavior vary from spontaneous dialogs [2, 5, 7, 15, 16, 21], discussions [5, 14], negotiations [22] and meetings [9, 18] to narrative [13, 17] or procedural tasks [1, 6, 12]. One can not compare these different corpora to each other on, for instance, listener responses per minute based on the numbers reported. There are too many variables that would confuse such a study, like culture, media and probably most importantly: the definition of the authors of what is regarded a listener response and what not. To make a comparison on the influence of different tasks on nonverbal listening behavior, one needs a corpus on which all these variables are controlled.

We could find one such corpus, the JSAI corpus [20]. This corpus contains audio recordings of 29 dialogs with three different tasks: chat, travel navigation and telephone shopping. Even though such a comparison is possible, no comparision between the different tasks has been performed on this corpus at the time of writing. To fill this hiatus in our understanding of listening behavior we will in this paper compare listening behavior elicited by the narrative and procedural tasks in the MultiLis Corpus, in which, as in the JSAI corpus, the other variables are controlled.

3. THE MULTILIS CORPUS

The MultiLis corpus [3] is a Dutch spoken multimodal corpus of 32 mediated face-to-face interactions totalling 131 minutes. Participants (29 male, 3 female, mean age 25) were assigned the role of either speaker or listener during an interaction. In each session four participants were invited to record four interactions. Each participant was once speaker and three times listener.

What is unique about this corpus is the fact that it contains parallel recordings of three individual listeners in interaction with the same speaker, while each of the listeners believed to be the sole listener (see Figure 1). The speakers saw only one of the listeners, believing that they had a one-on-one conversation. We will refer to this listener, which can be seen by the speaker, as *displayed listener*. The other two listeners, which can not be seen by the speaker, will be refered to as *concealed listeners*. All listeners were placed in a cubicle and saw the speaker on the screen in front of them.

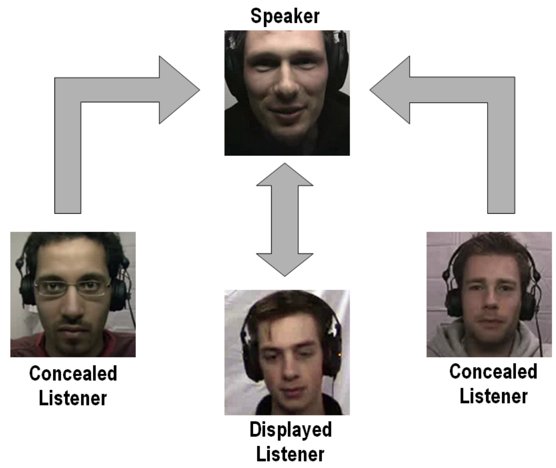

Figure 1: Schematic overview of the setup for this corpus. Each interaction is between a speaker and three listeners. The *speaker* sees only one of the listeners (the *displayed listener*). The other two listeners, the *concealed listeners*, can see the speaker, but the speaker is unaware of their participation to the conversation. All listeners believe to be the only listener in the interaction.

The camera was placed behind an interrogation mirror, positioned directly behind the position on which the interlocutor was projected. This made it possible to create the illusion of eye contact. To ensure the illusion of a one-on-one conversation was not broken, interaction between participants was limited. Speakers and listeners were instructed not to ask for clarifications or to elicit explicit feedback from each other.

3.1 Tasks

During the interactions the speaker needed to perform either a *narrative* task or a *procedural* task. For each type of task two different stimuli were used. Thus for each stimuli we have 8 interactions in our corpus.

In the narrative task the speaker saw a short video clip before the interaction. It was the task of the speaker to retell the events of the story in such a way that the listener could answer questions about the video clip afterwards. To give the speaker an idea what kind of questions were going to be asked, the speaker was given 8 open sample questions, which were taken away again just before the interaction began. We used two different clips as stimuli. The first is the 1950 Warner Bros. Tweety and Sylvester cartoon "Canary Row"[1]. In this clip we see Sylvester undertake a series of failed attempts to infiltrate the appartment where Tweety lives to catch him. The second clip is the 1998 animated short "More" by Mark Osborne[2]. The clip tells the moralistic tale of a factory worker in a distopian future, who is depressed about the routine of his everyday life. He invents a new product, which makes people happy. Initially this makes him the most praised and rich inventor of the world, but after a while he still feels dissatisfied and empty inside, longing for his carefree childhood. These clips where chosen

[1] Canary Row: http://www.imdb.com/title/tt0042304/
[2] More: http://www.imdb.com/title/tt0188913/

to be able to study the effects of affect on listening behavior. Since the Tweety clip is cheerful and the More clip is gloomy, we will refer to them as the *Cheerful* and *Gloomy* stimuli in the remainder of the paper.

In the procedural task the speaker was given 10 minutes to study a recipe. During the interaction the speaker needed to instruct the listener how to prepare the dish described in the recipe. Afterwards the listener needed to write down the recipe. We used two recipes as stimuli: Tea smoked salmon and risotto with mushrooms. Both recipes were chosen, because they are not very well known to limit the chance that participants knew the recipe. The main difference between the two recipes is the length. The salmon recipe is a little shorter and less complex than the risotto recipe. Both recipes have 11 ingredients, but the salmon recipe has 11 lines of instructions, whereas the risotto recipe has 19 lines. Therefore we will refer to them as the *Short* and *Long* stimuli during the remainder of the paper.

3.2 Responses Annotation

For the speaker we are interested in the gaze behavior and speech patterns. The gaze behavior was manually annotated. The annotators indicated whether the speaker was looking at the listener (directly into the camera) or not, and whether the speaker is blinking with his eyes. Gazes at the listener were occasionally interrupted by blinks of the speaker. Even though the gaze was interrupted for a moment, the listener would still have the perception that the speaker is addressing him/her. Therefore we included the blinks between and after a gaze annotation into the "gazing at the listener" interval.

The speech patterns were extracted using the Dutch automatic speech recognition software SHoUT [11]. From the results of the ASR we extracted the utterances. Utterances in this paper are defined as interpausal units, where the minimum length of the (automatically detected) silence between two utterances is 100 ms.

For the listener the corpus includes annotations of head, eyebrow and mouth movements, and speech transcriptions. What we refer to as a listener response can be any combination of these various behaviors, for instance, a head nod accompanied by a smile, raised eyebrows accompanied by a smile or the vocalization of "uh-huh", occurring at about the same time. For each of these responses we have marked the so-called onset (start time). The onset of a listener response is either the stroke of a head movement, the start of a vocalization, the start of eyebrow movement or the start of a mouth movement. When different behaviors combine into one listener response, priority was in this order, which means that if a response is a combination of a head nod, smile and the vocalization of "okay", the onset of the response is the stroke of the head nod.

3.3 Response Opportunity Identification

One of the assumptions underlying a lot of studies on listening behavior is that there are certain times at which it is more appropriate for a listener to produce a listener response than at other times. This assumption is clearly present in the works that attempt to generate listener responses in the case of virtual humans, where rules are defined or learned that determine at which time the agent can produce a listener response. The rules often rely on cues derived from the speaker's actions.

One way to look at this is to say that during the interaction the speakers provide a series of *response opportunities*. These are moments on which a listener can or wants to provide a response in reaction to the actions of the speaker. These opportunities can either be given deliberately by the speaker, by cueing the moment on which the speaker expects a response (resulting in a response elicited by the speaker), or accidently, by simply ending a sentence, looking at the listener for another reason or any other action which made the listner respond (resulting in a response initiated by the listener). The listener either responds to such an opportunity or not.

Since we have recorded three listeners, we have a more (but still not) complete view of all the response opportunities which are present in the interaction. Some opportunities passed up by the displayed listener (the only listener which would have been recorded in any other corpus) are still identified by one or two of the concealed listeners. Furthermore we can see to which response opportunities one, two or all three listeners responded. To identify the response opportunities with at least one response in our interactions we created an algorithm which links listener responses to response opportunities. Presumably there are also response opportunities in the interaction to which no listener responded, but with the information available to us, we can not know when these moments are. Thus the algorithm will not identify these moments.

To create such an algorithm, you need to determine how big the window of such a response opportunity is. This may vary for each opportunity, but since we create an algorithm, we need a fixed value for this window. We do not want our algorithm to link two responses from the same listener to the same opportunity, since the annotator made a deliberate decision that these two responses should be separated. The minimum gap between two responses from the same listener in our corpus is 714ms. To ensure that no two responses of the same listeners are linked to the same response opportunity we set the *opportunity window* to 700 ms. The algorithm we created works as follows:

A forward looking search is performed. When an hitherto unlinked response is encountered, the algorithm checks whether there are more responses which start within the opportunity window of 700 ms from the start time of this response. If there are, all of these are linked to the same response opportunity. After a response opportunity is identified we continue our forward looking search for the next unlinked response.

The algorithm identified 1733 response opportunities with at least one response. 1140 opportunities elicited only one response from the listener, 465 opportunities elicited two responses and 128 opportunities elicited responses from all three listeners. We performed a qualitative evaluation of the algorithm by randomly selecting 100 out of the 593 response opportunities with at least two responses linked to it and manually checking in the corpus whether the linking of the responses to the same response opportunity was correct. This was the case for 84 instances, confirming the validity of our approach.

4. RESULTS

In the next sections we will present and discuss the results on differences between the narrative and procedural tasks, followed by the differences between the four individual tasks.

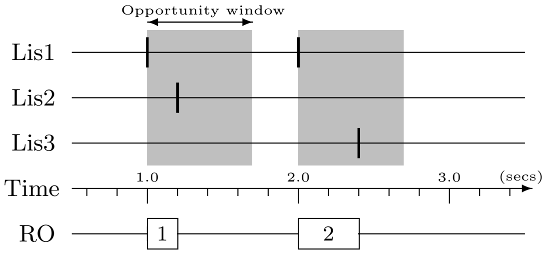

Figure 2: Example of the clustering algorithm. At time 1.0s the algorithm has encountered a listener response from listener 1. It checks whether there are more responses from other listeners within the *opportunity window* of 700ms. There is a response from listener 2 at time 1.2s, thus these are linked to response opportunity (RO) 1. The algorithm continues with the next unlined response and repeats the process. In encounters another responses of listener 1 at time 2.0s. This time there is a response from listener 3 at time 2.4s. Thus response opportunity 2 is created.

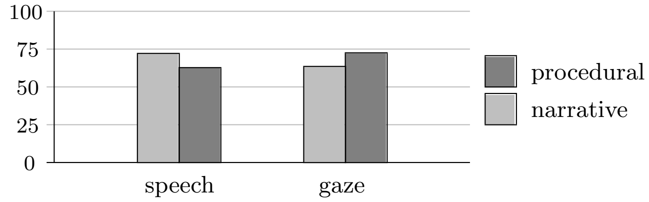

Figure 3: Percentage of the time the speaker spends speaking (left) and looking at the listener (right) during each type of task.

	Narrative	Procedural
Number of responses	1437	1359
Responses / minute	19.4	23.8
Percentage of responses including...		
head movement	86.6	88.7
a smile	16.7	11.0
eyebrow movement	6.7	7.8
vocalization	6.3	7.0

Table 1: Compares the narrative and procedural tasks with respect to the responses of all three listeners on amount and appearance.

Within each subsection we will first present an analysis of the differences of the speakers actions before moving on to the listener responses, since it are these actions which cause differences in listening behavior.

4.1 Narrative versus Procedural

The mean duration of the 16 narrative interactions is 4 minutes and 28 seconds, and the mean duration of the 16 procedural interactions is 3 minutes and 45 seconds. Of this time the speaker speaks 72.0% of the time in the narrative case and 62.6% of the time in the procedural case (see left graph in Figure 3). A paired-sample t-test[3] shows this difference is significant, $p = 0.01$. The mean length of the utterances is also significantly smaller in the case of the procedural tasks (1.32 seconds versus 1.73 seconds in the narrative tasks, $p < 0.05$). This is caused by summation of the ingredients list in this task, which typically consists of short utterances.

The time the speaker looks at the listener is 63.4% of the time for narrative and 72.4% for procedural tasks (see right graph in Figure 3)), but this difference is not significant ($p = 0.21$), due to the big variance in gaze behavior between speakers. Though there is a significant difference between the amount of gaze shifts[4] between the two types of tasks. During the narrative task the speaker shifts his gaze on average 16.4 times per minute and in the procedural task 10.4 times per minute ($p < 0.01$). This means that the duration of the periods the speaker looks at the listener are significantly longer during procedural tasks (4.2 seconds on average) than during narrative tasks (2.3 seconds), $p = 0.06$.

So, now we know the differences in speaker actions, but what kind of effect do these differences have on the listening behavior? We first look at the responses given by the three listeners combined, presented in Table 1. The number of responses given by the three listeners combined is higher for the narrative task than for the procedural task, but normalizing them to the average length of the interactions, shows that in fact the procedural task elicited more responses per minute (23.8) than the narrative task (19.4), but this difference is not significant ($p = 0.37$). If we take a look at what these responses look like, we can see that for both tasks the majority of the responses (86.6% of the responses in the narrative and 88.7% in the procedural tasks) include head movements. Furthermore we can see a trend that the narrative task elicited more smiles (16.7% of the responses) than the procedural task (11.0%), $p = 0.12$.

So far, we simply looked at the responses of the three listeners individually, but part of the responses of the three listeners will be reactions to the same response opportunity. Therefore we identified the response opportunities as discussed in Section 3.3 and compared the two different kind of tasks on these aspects. Table 2 shows the results of this comparison. We can see that the difference in the amount of responses per minute has decreased. Where the procedural tasks elicited 22% more responses per minute than the narrative tasks ($23.8/19.4 = 1.22$), the procedural tasks only include 15% more responses opportunities per minute ($14.3/12.4 = 1.15$). This means that more listeners responded to the response opportunities in the procedural tasks than in the narrative tasks. This shows in the amount of responses opportunities to which three listeners reacted. All three listeners responded in 10.3% of the response opportunities in the procedural task, opposed to 4.8% for narrative tasks, a marginally significant difference ($p = 0.06$). Thus there is more agreement between the listeners during the procedural task. We suspect that this is due to the fact that the procedural tasks are more structured. Participants follow the structure of the recipe (a summation of the ingre-

[3]For all further significance tests in this research the paired-sample t-test was used as well.

[4]A gaze shift is counted in this analysis every time the speaker directs his gaze towards the listener. So, for instance, a gaze shift from left to right is not counted.

	Narrative	Procedural
Number of response opportunities	916	819
Response opportunities / minute	12.4	14.3
Percentage of responses opportunities with...		
one response	68.3	63.0
two responses	26.9	26.7
three responses	4.8	10.3

Table 2: Compares the narrative and procedural tasks with respect to the response opportunities.

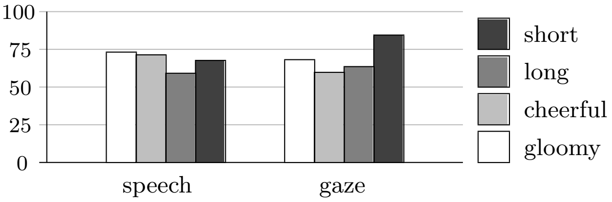

Figure 4: Percentage of the time the speaker spends speaking (left) and looking at the listener (right) during each task.

dients followed by the instructions) when reproducing the recipe. Most speakers also declare this at the start of the interaction and prepare the listener for this. A structure of the upcoming interaction is presented to the listener, so during the interaction the listener does not only acknowledge the content of the interaction through responses, but also the transitions to different phases in the interaction. Especially these transitions are very clear response opportunities to which listeners are more inclined to respond

4.2 Individual Tasks

Now we take a look at the four individual tasks. Of each task 8 interactions were recorded. The mean durations of each task were: Gloomy (4:11 minutes), Cheerful (5:04), Long (4:06), Short (3:03). So the difference between the Long and the Short recipe does not only show in the length of the recipe, but also in the length of the interactions. In the left graph of Figure 4 the percentage of speech time for each task is presented. We can see a significant difference ($p = 0.05$) in speech time between the Long (59.0%) and the Short (67.5%) recipe task. The difference in gaze time (63.4% versus 84.3%) is also marginally significant, $p = 0.06$ (see right graph of Figure 4). Both these results suggest that the cognitive load for the long recipe task was higher than for the short task. The speaker needed less time to think and formulate his sentences in the case of the Short recipe (according to the percentage of speech time result). Furthermore the speaker averted his gaze less. Gaze aversion is regarded as an action which is used to lower the cognitive load [23].

If we look at the mean utterance length we see an insignificant ($p = 0.12$) difference between the Gloomy clip (2.1 seconds) and the Cheerful (1.5 seconds). There is no difference in utterance length between the two recipes (both 1.3 seconds).

Table 3 presents the amount and type of listener responses the speaker elicited from the three listeners during each task. The Long recipe elicits significantly more responses per minute (26.2) than the Short recipe (20.4), $p = 0.02$. The Long recipe was more complex and had more instructions. Because of this complexity, the speaker needed to check if the listener was still paying attention and understood the instructions more often.

The slight difference between the Gloomy (19.7 responses per minute) and the Cheerful clip (19.2 responses per minute) is insignifiant ($p = 0.12$). Also the differences in type of listener response are insignificant for all tasks, including the percentage of responses including a smile the Gloomy (15.9%) and the Cheerful (17.4%) elicited.

In the previous section we saw that the three listeners in the procedural tasks had more agreement in their response behavior to response opportunities than the listeners in the narrative task. In Table 4 we can see that this is mainly caused by the Long recipe task. In this task there are significantly less response opportunities where only one listener responded to (54.1% of all responses), than in the Short recipe task (75.6% of all responses), $p < 0.01$.

5. CONCLUSIONS

In this paper we have presented the results of our comparison between listening behavior elicited by procedural and narrative tasks. We have seen that the three parallel recorded listeners in the procedural tasks showed more consensus in their responses than the three parallel recorded listeners in the narrative tasks. We have attributed this difference to the more structured way in which the interactions with the procedural tasks went. Furthermore we have seen that the *long* procedural task elicited more responses than the *short* task. This difference was attributed to the increase in cognitive load for the speaker. The speaker needs more time to recall the details of the recipe. Once he has recalled these details the listeners are more inclined to acknowledge this information, to encourage the speaker to continue. This explains the high consensus between listeners in the interactions with the long recipe task as well.

6. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 211486 (SEMAINE).

7. REFERENCES

[1] N. Cathcart, J. Carletta, and E. Klein. A shallow model of backchannel continuers in spoken dialogue. *European ACL*, pages 51–58, 2003.

[2] P. M. Clancy, S. A. Thompson, R. Suzuki, and H. Tao. The conversational use of reactive tokens English, Japanese, and Mandarin. *Journal of Pragmatics*, 26(3):355–387, Sept. 1996.

[3] I. de Kok and D. Heylen. The MultiLis Corpus - Dealing with Individual Differences of Nonverbal Listening Behavior. *in press*, 2010.

[4] A. T. Dittmann and L. G. Llewellyn. Relationship between vocalizations and head nods as listener responses. *Journal of personality and social psychology*, 9(1):79–84, May 1968.

	Gloomy	Cheerful	Long	Short
Number of responses	659	778	860	499
Responses / minute	19.7	19.2	26.2	20.4
Percentage of responses opportunities with...				
head movement	86.8	86.5	88.6	88.8
a smile	15.9	17.4	11.3	10.6
eyebrow movement	9.3	4.5	7.2	8.8
vocalization	4.6	7.7	8.0	5.2

Table 3: Compares the individual tasks with respect to the responses of all three listeners on amount and appearance.

	Gloomy	Cheerful	Long	Short
Number of response opportunities	438	478	479	340
Response opportunities / minute	13.0	11.8	14.6	13.9
Percentage of responses opportunities with...				
one response	73.0	64.0	54.1	75.6
two responses	22.8	30.5	32.6	18.5
three responses	4.1	5.4	13.4	5.9

Table 4: Compares the four individual tasks with respect to the response opportunities.

[5] J. Dixon and D. Foster. Gender, social context, and backchannel responses. *The Journal of social psychology*, 138(1):134, 1998.

[6] A. Gravano and J. Hirschberg. Backchannel-Inviting Cues in Task-Oriented Dialogue. In *Interspeech 2009*, pages 1019–1022, Brighton, 2009.

[7] B. Heinz. Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics*, 35(7):1113–1142, July 2003.

[8] D. Heylen, A. Nijholt, and M. Poel. Generating nonverbal signals for a sensitive artificial listener. In *Proceedings of Intelligent Virtual Agents*, volume 4775, page 264, Paris, France, 2007. Springer.

[9] D. Heylen and R. op Den Akker. Computing backchannel distributions in multi-party conversations. In *Proceedings of the Workshop on Embodied Language Processing*, pages 17–24, Prague, Czech Republic, 2007.

[10] L. Hirschman. Female-male differences in Conversational Interaction. *Language in Society*, 23:427–442, 1994.

[11] M. Huijbregts. *Segmentation , Diarization and Speech Transcription : Surprise Data Unraveled*. Phd thesis, University of Twente, 2008.

[12] R. M. Krauss, C. M. Garlock, P. D. Bricker, and L. E. McMahon. The role of audible and visible back-channel responses in interpersonal communication. *Journal of Personality and Social Psychology*, 35(7):523–529, 1977.

[13] R. E. Kraut, S. H. Lewis, and L. W. Swezey. Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology*, 43(4):718–731, 1982.

[14] T. A. Marche and C. Peterson. On the gender differential use of listener responsiveness. *Sex Roles*, 29(11-12):795–816, Dec. 1993.

[15] S. K. Maynard. Conversation management in contrast: Listener response in Japanese and American English. *Journal of Pragmatics*, 14(3):397–412, June 1990.

[16] M. McCarthy. Talking Back: "Small" Interactional Response Tokens in Everyday Conversation. *Research on Language & Social Interaction*, 36(1):33–63, Jan. 2003.

[17] L.-P. Morency, I. de Kok, and J. Gratch. A probabilistic multimodal approach for predicting listener backchannels. *Autonomous Agents and Multi-Agent Systems*, 20(1):70–84, May 2010.

[18] B. O'Conaill, S. Whittaker, and S. Wilbur. Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication. *Human-Computer Interaction*, 8(4):389–428, 1993.

[19] R. Poppe, K. P. Truong, D. Reidsma, and D. Heylen. Backchannel Strategies for Artificial Listeners. In *Proceedings of Intelligent Virtual Agents*, 2010.

[20] SIG Of Corpus-Based Research For Discourse And Dialogue. Constructing a spoken dialogue corpus as sharable research resource. *Japanese Society for Artificial Intelligence*, SIG-SLUD-9, 1999.

[21] N. Ward and W. Tsukahara. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, 32(8):1177–1207, July 2000.

[22] R. White. Back channelling, repair, pausing, and private speech. *Applied Linguistics*, 18(3):314, 1997.

[23] C. Wickens and J. McCarley. *Applied Attention Theory*. CRC Press, Bocan Raton, FL, USA, Dec. 2007.

[24] V. H. Yngve. On getting a word in edgewise. In *sixth regional meeting of the Chicago Linguistic Society*, volume 6, pages 657–677, 1970.