Backchannels: Quantity, Type and Timing Matters

Ronald Poppe, Khiet P. Truong, and Dirk Heylen*

Human Media Interaction Group, University of Twente
P.O. Box 217, 7500 AE, Enschede, The Netherlands
{r.w.poppe,k.p.truong,d.k.j.heylen}@utwente.nl

Abstract. In a perception experiment, we systematically varied the quantity, type and timing of backchannels. Participants viewed stimuli of a real speaker side-by-side with an animated listener and rated how human-like they perceived the latter's backchannel behavior. In addition, we obtained measures of appropriateness and optionality for each backchannel from key strokes. This approach allowed us to analyze the influence of each of the factors on entire fragments and on individual backchannels. The originally performed type and timing of a backchannel appeared to be more human-like, compared to a switched type or random timing. In addition, we found that nods are more often appropriate than vocalizations. For quantity, too few or too many backchannels per minute appeared to reduce the quality of the behavior. These findings are important for the design of algorithms for the automatic generation of backchannel behavior for artificial listeners.

1 Introduction

Listening is an important aspect of conversation. In a dialog, the listener actively contributes to the conversation by signalling attention, interest and understanding to the speaker [1]. One particular type of signal is the *backchannel* [2], a short visual (e.g. nod, smile) or vocal (e.g. "uh-huh" or "yeah") signal from the listener that does not interrupt the speaker's speech and is not aimed at taking the turn. There are several types of backchannels [3]. Here, we focus on those with a *continuer* function that convey continued attention but carry no additional affective meaning. From the analysis of human-human conversations, much is known about the *timing* and *type* of such backchannels. We discuss this work in Section 2.

Our goal is to use this knowledge to develop *artificial listeners*, virtual agents that can listen attentively to a human speaker [4]. This requires reliable prediction of backchannel opportunities from observations of the speaker's nonverbal visual and vocal behavior. In addition, appropriate listening behavior needs to

* The research leading to these results has received funding from the European Community's 7^{th} Framework Programme under Grant agreement 211486 (SEMAINE).

H. Högni Vilhjálmsson et al. (Eds.): IVA 2011, LNAI 6895, pp. 228–239, 2011.

be generated, which includes choosing the proper type of backchannel, and making sure that the number and spread of backchannels over a certain period of time is human-like.

Previous work has mainly focused on prediction of backchannel opportunities, initially in telephone-style dialogs [5] and more recently also in face-to-face settings [6]. These works have been evaluated using corpus-based measures such as precision and recall, which are informative of how well the prediction matches the backchannels that are performed in the corpus. However, good approximation of backchannel timings is not a guarantee that the predicted behavior will be *perceived* as human-like. This is partly due to the optionality of backchannels. For example, a predicted backchannel that is not performed by the human listener in the corpus is not necessarily incorrect, and vice versa. Moreover, other factors such as the type of backchannel and the number of backchannels in a period of time are not taken into account in corpus-based research.

To address these issues, Poppe *et al.* [7] have investigated how backchannel behavior, generated using different algorithms, was perceived by human observers. Participants in the perception experiment were shown stimuli of real speakers and animated listeners and were asked to rate how human-like the backchannel behavior appeared to them. Closer analysis revealed some effects of timing, type and quantity of backchannels in short fragments on how they were perceived. However, these factors had not been varied systematically. In addition, they were analyzed at the fragment level, and each fragment typically consisted of multiple backchannels of randomly chosen type.

Therefore, in this research, we conducted a perception experiment where the timing, type and quantity of listener backchannels were varied systematically. Upon viewing a fragment, participants indicated how likely they thought it was that the backchannel behavior had been performed by a human listener. For the three factors under investigation, we briefly explain how we expect each to influence the perception of backchannel behavior. We also present our hypotheses, which we will test in Section 4.

In [7], a significant positive correlation was found between the number of backchannels and the rating of the fragment. In our experiment, we expect to observe the same effect. Our *quantity* hypothesis is therefore formulated as:

Hypothesis 1. *Fragments with higher numbers of backchannels per minute are perceived as more human-like.*

In this study, we consider two types of backchannels: visual (nod) and vocal ("uh-huh"). While both types have the same continuer function, there are differences in timing within the speaker's turn. For example, nods are more often produced during mutual gaze, whereas vocalizations tend to be produced around the end of a segment of speech [8]. We therefore expect that there is no such thing as a general backchannel opportunity, but rather an opportunity for a nod or an opportunity for a vocalization. Although both might partly overlap, in general, we expect that changing the type from that was actually performed would result in lower subjective ratings. The *type* hypothesis is thus:

Hypothesis 2. *Fragments with backchannel types performed by the actual listener are perceived as more human-like compared to fragments in which backchannel types are changed.*

While backchannels are optional, there are many known systematics in the production of a backchannel as a reaction or anticipation of the speaker's verbal and nonverbal behavior. We expect that contingent timings, rather than random timings, will be rated as more human-like. The *timing* hypothesis is therefore:

Hypothesis 3. *Fragments with backchannel timings performed by the actual listener are perceived as more human-like compared to random timings.*

An important addition to the experiment procedure was that participants were not only asked to give a rating per fragment, but also to judge individual backchannels. A common observation in the area of virtual agents is that humans are sensitive to the flaws in animated behavior. With this in mind, we introduced the *yuck* button approach: a button is pressed every time a human observer thinks the behavior displayed is inappropriate. In our experiment, this approach allows us to obtain subjective ratings for both fragments and individual backchannels without additional time requirements. In turn, we can analyze how the rating of individual backchannels influences the perception of an entire fragment.

The paper proceeds with a discussion of related work, followed by a summary of our experiment setup. Results are presented and discussed in Section 4.

2 Related Work

Backchannels, or listener responses, are short visual or vocal signals from the listener to express attention, interest and understanding to the speaker without the aim to take the turn [1,2]. Research into backchannels can be grouped into two directions [9]: the lumping approach and the splitting approach. The former treats backchannels as a single class and is mainly concerned with the timing within a speaker's discourse. The latter approach has investigated specific forms of backchannels and their role in a turn-taking context.

Our goal is to develop artificial listeners, virtual agents that can listen attentively to a human speaker. This requires analysis of the speaker's verbal and nonverbal behavior to identify moments where backchannels might be produced. Research following the lumping approach has investigated the structural properties of backchannels, i.e. the relation between the speaker's behavior and the occurrence of backchannels. Nowadays, the occurrence of backchannels within the speaker's discourse is reasonably well-understood. Dittmann and Llewellyn [10] and Duncan [3] noted that backchannels are often produced after rhythmic units in the speaker's speech, and specifically at the end of grammatical clauses.

Motivated by the goal of automatically identifying backchannel opportunities, recent work has focused on identifying lower-level structural properties

of backchannels. Initially, telephone-style conversations have been addressed. Here, the relation between the speaker's speech and the occurrence of listener backchannels was investigated. A region of low or rising pitch [5,11], a high or decreasing energy pattern [12] and a short pause [13] in the speaker's speech have been found to precede backchannels from the listener. More recent work has shifted towards face-to-face conversations. These differ in backchannel behavior due to the additional visual modality that can be used to signal attention. In particular, the relation between gaze and backchannels has been investigated. For example, Kendon [14] and Bavelas [15] observed that backchannels were more likely to occur during a short period of mutual gaze, usually at the end of the speaker's turn.

These low-level structural properties can be extracted in real-time, and have been used to automatically identify backchannel opportunities using rule-based [5,16] and machine learning [6,17,18] algorithms. Typically, these algorithms work with a short time scale, which does not enforce consistent backchannel behavior over time.

Our aim is to develop artificial listeners, that should display active listening behavior that is human-like. In addition to the identification of backchannel opportunities, this requires the generation of human-like backchannel behavior. Attempts in this direction have been taken by Huang *et al.* [19], who generated head nods at moments that had been identified off-line based on multi-modal features. Maatman *et al.* [20] used the rule-based prediction algorithm of Ward and Tsukahara [5] and displayed nods in an online setting. Even though both works considered a face-to-face setting, none of them have addressed the type of backchannel. Poppe *et al.* [7] used either head nods or vocalizations, but the type was chosen randomly.

In the lumping approach, backchannels have been treated as a single class, without distinguishing between the type (e.g. visual and vocal). However, there are systematic differences in the structural properties of backchannels of different types. For example, Dittmann and Llewellyn [21] observed that, on average, a nod is produced 175ms earlier than a vocalization. In addition, Truong *et al.* [8] found that a visual backchannel was more likely to occur during mutual gaze, whereas vocal backchannels were more often produced during a pause in the speaker's speech.

Given these differences in occurrence, it is likely that there is no such thing as a backchannel opportunity, but rather the opportunity for a specific type of backchannel. We therefore expect that a different type of backchannel, produced in the same structural context, will be perceived differently by human observers. Although some researchers have addressed the perception of different types of backchannels [22,23] in isolation, none of them have investigated their perception in a conversational context.

Therefore, in this paper, we investigate how the quantity, type and timing of backchannels influences how human-like the backchannel behavior is perceived by human observers.

3 Experiment Setup

To investigate our hypotheses, we conducted a user experiment where human observers rated fragments from dialogs. We replaced the human listener by a virtual agent, and systematically varied the backchannel behavior. In this section, we describe the setup of the experiment.

3.1 Stimuli

We used dialogs between a speaker and a listener from the Semaine Solid SAL corpus [24], which contains emotionally-colored dialogs between a human listener and a human speaker. Specifically, we selected 12 fragments from [7] with at least two backchannels. The fragments are between 14 and 31 seconds in length.

We showed the speaker and listener side-by-side but replaced the video of the listener by a virtual agent, animated using BML realizer Elckerlyc [25] (see Figure 1). The backchannel behavior performed by this agent was systematically varied along three dimensions: quantity, type and timing. For each dimension, we took the manually annotated backchannels performed by the actual listener as a basis. For the quantity dimension, we defined three conditions. All backchannels were used in the *original* condition. In the *odd* and *even* condition, we selected every second backchannel, starting with the first or second one, respectively. The three conditions contained 46, 26 and 20 backchannels, respectively. As backchannel type, we used a nod, a vocalization ("uh-huh") or a combination of both. We animated either the *original* types, or the *switched* types, with nods replaced by vocalizations, and vocal and bimodal backchannels by nods. For the timing dimension, we used the *original* onsets or *random* onsets. In the latter case, there was at least one second between two onsets. Also, the order of the types of backchannels was left unchanged. The three dimensions were crossed to yield 12 conditions. In addition to the backchannels, we animated the listener's blinks where they occurred in the actual recording.

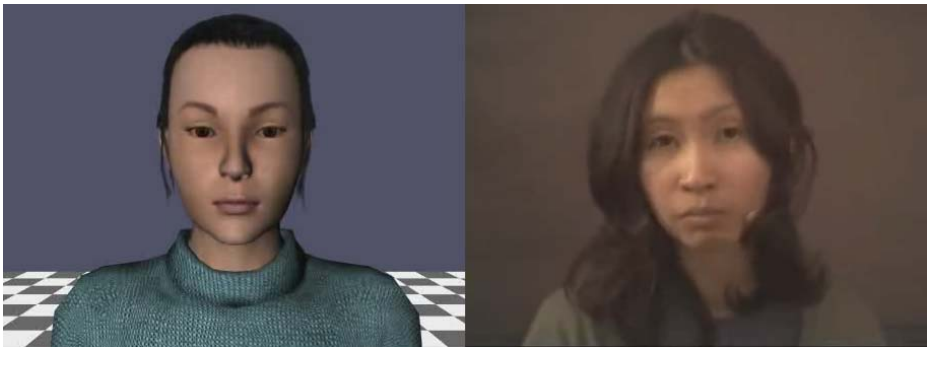


Fig. 1. Example stimulus with artificial listener (left) and actual speaker (right)

3.2 Procedure

The participants were explained they would be participating in an experiment to determine the quality of backchannel behavior. After the briefing, they were shown a set of stimuli. They were instructed to press the yuck button (space bar) every time they thought a listener's backchannel was inappropriate, either in type or timing. Participants could replay the video as often as desired, and adapt their yucks if needed. After watching a video fragment, they were prompted to rate how human-like they perceived the listener's backchannel behavior. They could set a slider that corresponded to a value between 0 and 100.

We divided the 144 condition-fragment combinations into six distinct sets of 24 stimuli. We adopted a Greco-Latin square design to control for order. In addition, this ensured that each participant rated each fragment and condition twice and six participants together rated all possible combinations of both.

3.3 Participants

We recruited 24 colleagues and doctoral students (6 female, 18 male) with a mean age of 32.8 (min 23, max 58). Each of the participants was assigned randomly to a set of stimuli with the lowest number of respondents.

4 Results and Discussion

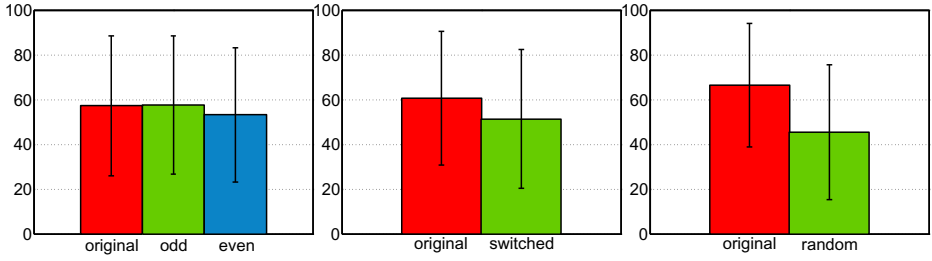
We collected ratings over fragments, and yucks for individual backchannels. Both are discussed separately in the following sections.

4.1 Fragment Ratings

We analyze how quantity, type and timing of backchannels affects how human-like participants perceived a fragment. We ignore the variable fragment as the number of ratings per fragment-condition combination is limited. We performed a repeated measures ANOVA with order as between-subjects variable and quantity, type and timing as within-subjects variables. There are differences in ratings for different participants. While these do not affect the significance of the observed effects, we will use z-scores of the fragment ratings as our dependent variable unless explicitly stated otherwise.

In Figure 2, the results of un-normalized ratings for quantity, type and timing are shown. Overall, these scores are rather low, which is in line with [7]. We partly attribute this observation to the fact that only backchannels and blinks were animated. The high standard deviations are due to the grouping of ratings from different fragments.

For quantity, we did not find a significant effect ($F(2) = 2.062, p = ns$). We observe in Figure 2(a) that the difference between 46 and 26 backchannels in the *original* and *odd* conditions is minimal. However, there is an interaction effect between quantity and timing ($F(2) = 5.891, p < 0.01$). Specifically, the effect is



(a) Quantity (b) Type (c) Timing

Fig. 2. Average un-normalized fragment ratings per dimension

opposite; the *odd* condition was rated the best with the *original* timings, but lowest with *random* timings. The differences between the quantity conditions for the *random* timings are less pronounced.

In this analysis, we did not account for the duration of the fragment. If we correlate the average ratings per fragment-condition combination with the average number of backchannels per minute, we find no significant effect. We therefore have to reject hypothesis 1 that more backchannels per minute are perceived as more human-like. However, when only the conditions with *original* timing and type are taken into account, the correlation is significant ($r(36) = 0.340, p < 0.05$). Closer analysis reveals that the fragment with the highest number of backchannels per minute (20.18) appears to be an outlier. Leaving this fragment out results in a correlation of $r(35) = 0.537, p < 0.001$. This analysis suggests that too few and too many backchannels will reduce the quality of the backchannel behavior. We expect that a reasonable number of backchannels per minute lies between 6 and 12.

Type proved to be significantly different for the *original* and *switched* conditions ($F(1) = 18.233, p < 0.001$). Apparently, different types of backchannels are performed in different contexts. We therefore accept hypothesis 2 that the original type is rated more human-like. We will investigate this more thoroughly in Section 4.2.

We also found a main effect for timing ($F(1) = 94.684, p < 0.001$). Apparently, participants rated *random* timing lower. This confirms hypothesis 3 that original timings are perceived as more human-like. However, the difference between the two conditions is moderate. The same observation was also made in [7,19] and can be partly attributed to inter-personal differences, the optional nature of backchannels and the fact that, apart from backchannels and blinks, no other behaviors were animated.

While these results reveal differences in perception for different quantity, type and timing conditions, each fragment contains multiple backchannels. In the next section, we will analyze the perception of individual backchannels.

4.2 Individual Backchannel Ratings

For each backchannel, we are interested in how often participants rated it as inappropriate. We obtain this information by linking the yucks to the performed backchannels. In addition, we obtain a measure of optionality for each backchannel using parasocial consensus sampling (PCS), which we explain next.

Parasocial Consensus Sampling. Given that backchannels are often optional and that there are inter-personal differences in backchannel behavior, we are interested in the optionality of specific backchannels. We used PCS [19] as a tool to obtain backchannel opportunities from multiple raters. Specifically, we had nine participants watch the video of the speaker from the fragments that were used in the perception experiment. We asked them to press a button whenever they would perform a backchannel. In total, we obtained 240 responses, which is approximately half the number of actually performed backchannels. Still, we expect that the ratings give a more general idea at which moments backchannels are common, and when they are more optional. Next, we discuss how we linked the PCS and yuck responses to the backchannels generated in the stimuli.

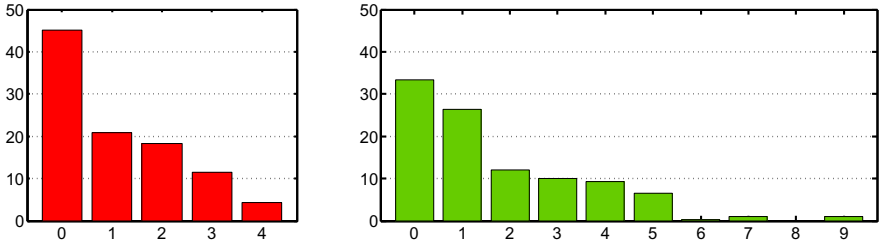
Data Processing. As our aim is to report on the appropriateness of individual backchannels in the stimuli, we need to associate the yucks and the PCS responses to these backchannels. For the yucks, there is a time delay between the stimulus onset and the participants' key press. We analyzed this delay and associated a yuck response with the closest preceding backchannel, provided that the time between them was between 300 and 2500ms.

One would expect that the timings of PCS responses are similar to the actual backchannel onsets. On closer analysis, a PCS response appears to be approximately 200ms later. We use a matching window of 500ms and therefore, we associate a PCS with a backchannel if it is between 300ms before and 800ms after a backchannel onset.

The total number of generated backchannels in all fragments and conditions is 368. Figure 3 shows the frequency of yucks and PCS responses per backchannel. Each fragment-condition combination has been judged by four participants, so the maximum number of yucks per backchannel is four. As the numbers of yuck and PCS responses are a measure of a backchannel's unsuitability and suitability, respectively, it is not surprising that the numbers of these responses are negatively correlated ($r(368) = -0.400, p < 0.001$).

Quantity. For now, we only consider the quantity dimension, and use only the data of the *original* type and timing conditions. As the *odd* and *even* quantity conditions contain a subset of the backchannels in the *original* quantity condition, we expect similar numbers of PCS responses in all conditions. These numbers are 2.35, 2.58 and 2.06, respectively. They are reasonably equal and correlate with the fragment ratings in the previous section.

If quantity would not be an important factor in backchannel behavior, we would expect similar numbers of yucks as well. However, we found the average



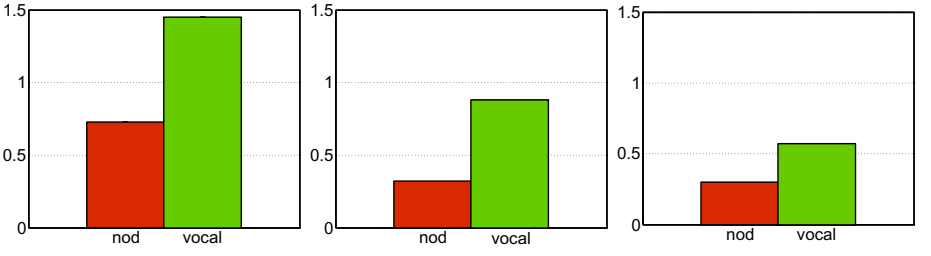
(a) Yucks (b) PCS responses

Fig. 3. Relative frequency of yucks and PCS responses per backchannel (%)

numbers of yucks per backchannel to be 0.54, 0.19 and 0.25 for the *original*, *odd* and *even* conditions, respectively. Apparently, more backchannels is not always better. This is somewhat at variance with findings in [7]. Closer analysis reveals that eight out of 25 yucks in the *original* setting originate from the fragment with the highest number of backchannels per minute (20.18). Again, it appears that too few or too many backchannels reduces the perceived quality of the backchannel behavior.

Type. In Section 4.2, we investigated differences between the *original* and *switched* condition as an indication that changing the type of backchannel affects how it is perceived. Given the yucks, we can also analyze whether the type of an individual backchannel matters, disregarding the specific condition. As we expect that the vocal aspect of bimodal backchannels is most salient, we treated these backchannels as vocalizations. The average numbers of yucks for nods and vocalizations are, respectively, 0.32 and 0.88 with *original* timing, and 1.15 and 2.01 with *random* timing (see Figure 4). Over both conditions, the percentage of backchannels that did not receive a yuck was 57.6% and 32.6% for the nods and vocalizations, respectively. We can further narrow down the class with *original* timings and distinguish between the backchannels performed in the *original* and *switched* type conditions. Changing vocalizations and bimodal backchannels to nods caused a slight increase in number of yucks per backchannel, from 0.30 to 0.36. However, changing nods to vocalizations led to an increase from 0.57 to 1.02.

These numbers indicate that a nod is less often perceived as inappropriate. We expect this can at least be partly explained by the fact that nods are communicated over the visual channel, without directly interfering with the main channel of communication. Therefore, it might be that vocalizations are more precisely timed, whereas nods can be performed throughout the speaker's turn. If this would be the case, one would expect higher numbers of PCS responses for a vocalization compared to a nod for the actually performed backchannels. This is indeed the case, with on average 3.20 responses for a vocalization and 1.97 for a nod. These findings are important for the design of backchannel generation algorithms for artificial listeners. High confidence in the backchannel



(a) All (b) Original timing (c) Original timing and type

Fig. 4. Average number of yucks per nod or vocalization in different conditions

prediction could result in the production of a vocalization, whereas a nod might be produced otherwise.

Timing. The effect of timing on the perception of backchannel behavior can also be observed from the PCS and yuck responses. The average number of PCS responses for a backchannel with *random* timing is 1.03, compared to 2.35 with the *original* timing. Randomly timed backchannels are thus twice less likely to occur. Not surprisingly, the number of yucks in the random condition is much higher than in the original condition, 1.58 versus 0.60. This again shows that timing matters.

5 Conclusion and Future Work

We have conducted a perception experiment where the factors quantity, type and timing of backchannels was varied systematically. Participants in the experiment were shown stimuli of real speakers and animated listeners and were asked to rate how human-like the generated backchannel behavior appeared to them. In addition, we obtained measures of appropriateness and optionality for each backchannel from yuck responses and parasocial consensus sampling (PCS) responses. This approach allowed us to analyze the influence of each of the factors on entire fragments and on individual backchannels.

From the fragment ratings, the number of the backchannels per minute over all conditions was not a significant factor. However, with original timings and type, there was a trend that more backchannels led to a more human-like perception of the fragment. Closer analysis showed that a very high number of backchannels per minute resulted in much lower subjective ratings. In addition, individual backchannels were more often regarded as inappropriate when the rate of backchannels was higher. This was especially true for randomly timed backchannels. In summary, there appears to be a lower and an upper bound on the number of backchannels per minute, around 6 and 12 respectively.

The type of backchannel (originally performed or switched) was a significant factor in the fragment ratings. Apparently, different types of backchannels are

performed in different contexts. Analysis of individual backchannels revealed that nods are less often rated as inappropriate, disregarding their timing. This knowledge has implications for the design of backchannel generation algorithms. If the prediction confidence is low, it is probably more appropriate to generate a nod.

For the timing of backchannels, both fragment ratings and yucks indicated that random timings are perceived as less human-like. This again stresses the importance of accurate backchannel prediction algorithms.

While corpus-based research is useful to identify contexts where a specific type of backchannel is more likely, we argue that the models derived from this research should be validated using perception studies. Also, we propose to abandon the concept of a general backchannel opportunity, and focus on predicting specific backchannels with their own structural properties instead.

The combination of PCS and yucks proved to be valuable in the analysis of individual backchannels. In future work, we expect they will continue to be useful tools to unravel the factors involved in designing a human-like backchannel generation algorithm. Specifically, we plan to analyze at which moments backchannels are perceived human-like, and how these moments differ from each other. Our aim is to conduct these studies in online settings as well. In addition, we continue to look for other ways to predict, generate and understand human behavior, its optionality and dependence on social context.

References

1. Bavelas, J.B., Coates, L., Johnson, T.: Listeners as co-narrators. Journal of Personality and Social Psychology 79, 941–952 (2000)
2. Yngve, V.H.: On getting a word in edgewise. Papers from the Sixth Regional Meeting of Chicago Linguistic Society, pp. 567–577. Chicago Linguistic Society (1970)
3. Duncan Jr., S.: On the structure of speaker-auditor interaction during speaking turns. Language in Society 3, 161–180 (1974)
4. Heylen, D., Bevacqua, E., Pelachaud, C., Poggi, I., Gratch, J., Schröder, M.: Generating Listening Behaviour. In: Emotion-Oriented Systems Cognitive Technologies - Part 4, pp. 321–347. Springer, Heidelberg (2011)
5. Ward, N., Tsukahara, W.: Prosodic features which cue back-channel responses in English and Japanese. Journal of Pragmatics 32, 1177–1207 (2000)
6. Morency, L.P., de Kok, I., Gratch, J.: A probabilistic multimodal approach for predicting listener backchannels. Autonomous Agents and Multi-Agent Systems 20, 80–84 (2010)
7. Poppe, R., Truong, K.P., Reidsma, D., Heylen, D.: Backchannel strategies for artificial listeners. In: Safonova, A. (ed.) IVA 2010. LNCS, vol. 6356, pp. 146–158. Springer, Heidelberg (2010)
8. Truong, K.P., Poppe, R., Kok, I., Heylen, D.: A multimodal analysis of vocal and visual backchannels in spontaneous dialogs. In: Proceedings of Interspeech, Florence, Italy (to appear, 2011)
9. Xudong, D.: Listener response. In: The Pragmatics of Interaction, pp. 104–124. John Benjamins Publishing, Amsterdam (2009)

10. Dittmann, A.T., Llewellyn, L.G.: The phonemic clause as a unit of speech decoding. Journal of Personality and Social Psychology 6, 341–349 (1967)
11. Gravano, A., Hirschberg, J.: Backchannel-inviting cues in task-oriented dialogue. In: Proceedings of Interspeech, Brighton, UK, pp. 1019–1022 (2009)
12. Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., Den, Y.: An analysis of turn-taking and backchannels based on prosodic and syntactic features in japanese map task dialogs. Language and Speech 41, 295–321 (1998)
13. Cathcart, N., Carletta, J., Klein, E.: A shallow model of backchannel continuers in spoken dialogue. In: Proceedings of the Conference of the European chapter of the Association for Computational Linguistics, Budapest, Hungary, vol. 1, pp. 51–58 (2003)
14. Kendon, A.: Some functions of gaze direction in social interaction. Acta Psychologica 26, 22–63 (1967)
15. Bavelas, J.B., Coates, L., Johnson, T.: Listener responses as a collaborative process: The role of gaze. Journal of Communication 52, 566–580 (2002)
16. Truong, K.P., Poppe, R., Heylen, D.: A rule-based backchannel prediction model using pitch and pause information. In: Proceedings of Interspeech, Makuhari, Japan, pp. 490–493 (2010)
17. Noguchi, H., Den, Y.: Prosody-based detection of the context of backchannel responses. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP), Sydney, Australia, pp. 487–490 (1998)
18. Okato, Y., Kato, K., Yamamoto, M., Itahashi, S.: Insertion of interjectory response based on prosodic information. In: Proceedings of the IEEE Workshop Interactive Voice Technology for Telecommunication Applications, Basking Ridge, NJ, pp. 85–88 (1996)
19. Huang, L., Morency, L.-P., Gratch, J.: Learning backchannel prediction model from parasocial consensus sampling: A subjective evaluation. In: Safonova, A. (ed.) IVA 2010. LNCS, vol. 6356, pp. 159–172. Springer, Heidelberg (2010)
20. Maatman, R.M., Gratch, J., Marsella, S.C.: Natural behavior of a listening agent. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 25–36. Springer, Heidelberg (2005)
21. Dittmann, A.T., Llewellyn, L.G.: Relationship between vocalizations and head nods as listener responses. Journal of Personality and Social Psychology 9, 79–84 (1968)
22. Granström, B., House, D., Swerts, M.: Multimodal feedback cues in human-machine interactions. In: Proceedings of the International Conference on Speech Prosody, pp. 11–14. Aix-en-Provence, France (2002)
23. Bevacqua, E., Pammi, S., Hyniewska, S.J., Schröder, M., Pelachaud, C.: Multimodal backchannels for embodied conversational agents. In: Safonova, A. (ed.) IVA 2010. LNCS, vol. 6356, pp. 194–200. Springer, Heidelberg (2010)
24. Valstar, M.F., McKeown, G., Cowie, R., Pantic, M.: The Semaine corpus of emotionally coloured character interactions. In: Proceedings of the International Conference on Multimedia & Expo., Singapore, pp. 1079–1084 (2010)
25. Van Welbergen, H., Reidsma, D., Ruttkay, Z., Zwiers, J.: Elckerlyc - A BML realizer for continuous, multimodal interaction with a virtual human. Journal of Multimodal User Interfaces 3, 271–284 (2010)