SEMDIAL 2017 (SaarDial) Workshop on the Semantics and Pragmatics of Dialogue
15-17 August 2017, Saarbrücken, Germany

Unimodal and Bimodal Backchannels in Conversational English

Gaëlle Ferré
University of Nantes
LLING UMR 6310
Chemin de la Censive du Tertre
BP 81227 Nantes cedex 3
FRANCE

Suzanne Renaudier
University of Nantes
LLING UMR 6310
Chemin de la Censive du Tertre
BP 81227 Nantes cedex 3
FRANCE

Gaelle.Ferre@univ-nantes.fr; suzanne.renaudier@gmail.com

Abstract

This paper presents differences in use of verbal (*(oh) yeah*, *(mh)mh*, *okay* ...) and visual backchannels (*head nods*, *shakes*, *tilts*), e.g. unimodal backchannels, as well as bimodal backchannels that combine a verbal token and a head movement in conversational English. We analyze the participants' gaze-pattern before the production of a BC but also during and immediately after its delivery. We also analyze their placement regarding the main speaker's turn and within the discourse topic. Lastly, we discuss their functions. Our findings reveal that each BC type shows a different picture from the other two both in terms of where they occur within the main speaker's turn and what their functions are. We however do not confirm previous observations regarding the constraints on their occurrence within a discourse topic.

1 Introduction

A conversation needs at least a speaker and a listener. While the two roles can be unbalanced during the telling of a story, when the speaker takes the floor for a long time, the listener still participates in the building of the exchange. Backchannels (BCs), i.e. short responses produced by listeners to signal attention, interest and understanding (Bertrand et al., 2007; Truong et al., 2011, among other studies) play a major role in the process. Indeed, according to many researchers (Terrell and Mutlu, 2012; Yamaguchi et al., 2015, to cite but a few), they regulate speech turns by letting speakers know whether co-participants understand what is being said or not and if they can keep the floor in order to continue their story. More simply, BCs serve to display a continued interest and co-participation in topic development (Gardner, 2001; Lambertz, 2011). Thus, BCs show alignment and they can also be signs of affiliation when the listener takes a stance (Stivers, 2008). We can then say that BCs cannot be disregarded.

However, BCs can show varied forms and functions. But although many studies have focused on unimodal BCs and their occurrences and functions, we can wonder about the specific characteristics of bimodal ones, i.e. BCs that combine a gesture and a verbal token. Do they play a different function than simple visual or verbal BCs? Since it has been shown in previous studies that verbal BCs such as *mhmh* may be produced soon after the beginning of a turn whereas visual BCs such as nods can only be produced later (Dittmann and Llewellyn, 1968; Stivers, 2008; Poppe et al., 2011), what can be said about bimodal BCs that combine both tokens? Is mutual gaze an important cue to the occurrence of a BC and is it different in the case of a bimodal BC?

After presenting the theoretical background in section 2, the paper presents the corpus and the data we worked on for this study in section 3. To answer our research questions, we examined the gaze-pattern throughout whole sequences that contain BCs in section 4. We also considered the placement of BCs with regards to speech turns and (sub)topics as well as their function in conversational English, specifically focusing on the difference between unimodal and bimodal BCs. Section 5 summarizes and discusses our results before we reach a conclusion in section 6.

2 Theoretical background

2.1 Backchannel placement

Many studies have shown that BCs do not appear randomly in a conversation (Bavelas et al., 2000;

10.21437/SemDial.2017-3

McCarthy, 2003; Poppe et al., 2011). First of all, the listener needs to have some information before being able to respond to the speaker: Truong et al. (2011) showed that attention is higher toward the end of speech turns so there is a growing probability of BC production as speech progresses. Furthermore, they often appear at the end of rhythmic units, specifically at the end of grammatical clauses (Dittmann and Llewellyn, 1968; Ike, 2010; Poppe et al., 2011). This way, the listener has the information needed to process what has been said before showing any sign of alignment or affiliation. Nevertheless, as reported by Heldner et al. (2013), there are more backchannel relevance places than actual BCs and a BC would not be appropriate at the end of a speech turn (Bertrand et al., 2007) so their positions are precisely chosen by listeners to help speakers in the building of their story. Whereas *yeah* is preferred to acknowledge the end of a topic, *mhmh* is not appropriate in this position (Jefferson, 1983) and Stivers (2008) further noted that nods occur in mid-telling positions and are considered by speakers as inappropriate when they are produced at the end of narratives.

Actually, BCs seem to be triggered by different cues, prosodic, syntactic or embodied (Tolins and Fox Tree, 2014). Many studies enhance the role of prosody in their occurrence. Among others, Terrell and Mutlu (2012) showed that pauses are very important and that the more pauses there are in the speaker's speech, the more the listener has opportunities to provide BCs and thus facilitate the continuation of the story. Moreover, pitch around BCs has been analysed and researchers agree on saying it has a major influence on the occurrence of BCs (Gravano and Hirschberg, 2009; Poppe et al., 2010; Poppe et al., 2011; Hjalmarsson and Oertel, 2011). However, Benus et al. (2007) noted that BCs seem to follow intonational phrases with rising pitch while Yamaguchi et al. (2015) report that a major prosodic cue preceding a backchannel would be a low pitch region. Hence, the influence of pitch on the listener's BC production may depend on the context and be more important in certain types of interaction as in telephone conversations, for example (Truong et al., 2011). BCs also depend on the language of the speakers (Clancy et al., 1996; Ike, 2010).

It was found as well that participants' gaze plays a major role in triggering a BC on the listener's part (Bertrand et al., 2007; Poppe et al., 2010; Poppe et al., 2011; Truong et al., 2011; Hjalmarsson and Oertel, 2011; Terrell and Mutlu, 2012). Indeed, mutual gaze enables speakers to see if listeners align with them and listeners show this alignment with visual BCs thus avoiding interruption of speech.

Finally, BCs do not appear in the same positions depending on their types. Indeed, visual nods do not interrupt speech whereas verbal ones do. Thus, it was found that verbal and bimodal backchannels were preferably used during pauses whereas visual backchannels such as nods can appear at any moment (Dittmann and Llewellyn, 1968; Lambertz, 2011; Poppe et al., 2011; Truong et al., 2011).

2.2 Backchannel Function

Depending on the listener's intention when providing a BC, these response tokens do not assume the same functions. Even though they all provide some information in the course the talk is taking, they can express different things such as understanding, agreement or simply attention (Gardner, 2001). Researchers agree on saying that BCs can be divided into generic and specific ones (Goodwin, 1986; Bavelas et al., 2000; Tolins and Fox Tree, 2014): generic BCs signal the listener's participation in the conversation while specific ones show one's stance toward what one is being told.

Furthermore, BCs enable speakers to know if listeners align with them, that is if listeners understand what is being said and do not plan to take the floor. However, the tokens can also have a more profound function and show affiliation (Stivers, 2008; Lee and Tanaka, 2016). In this case, the listener shows that s/he agrees with the speaker's stance. Moreover, BCs can be divided into three main functions, according to Gardner (2001): continuers, which give the floor back to the speaker straight away; acknowledgments, which claim agreement or comprehension; news markers, also called assessments in other studies, which mark the prior turn as newsworthy. Lambertz (2011) also distinguishes change-of-activity tokens which mark a movement towards a new topic or action in the conversation.

BCs can take many forms and belong to different types: verbal, visual, or bimodal, but these forms do not correspond to a single function. Intonation can change a generic backchannel into a specific one, for example (Tolins and Fox Tree,

2014). Hence, they all exhibit a great flexibility and multi-functionality of use (Gardner, 2001). However, some BCs seem to be more appropriate as acknowledgments and others as continuers, etc. For example, Lambertz (2011) explained that while both *yeah* and *mhmh* can function as continuers, alignment tokens and agreement tokens, *mhmh* seems to be weaker as an agreement token and appears more neutral whereas *yeah* somewhat expresses an opinion about an utterance. Terrell and Mutlu (2012) reported that nodding is a common non-verbal BC that plays many roles from indicating agreement to conveying sympathy and understanding with the speaker's perspective. It then seems that BCs can assume many functions depending on their position and their utterer's intonation (Gardner, 2001). *Mhmh* can be an encouragement to resume the tale when it occurs during a pause (Morel and Danon-Boileau, 2001) but can also be a follow up or a continuer (Drummond and Hopper, 1993; McCarthy, 2003).

Finally, bimodal backchannels have to be studied as a whole. The verbal part and the visual one cannot be studied separately. Their combination creates a whole new meaning (Bevacqua et al., 2010; Wlodarczak et al., 2012). Some studies have reported that bimodal BCs show a stronger agreement than a nod or a *yeah* on its own (Bevacqua et al., 2010; Terrell and Mutlu, 2012). Their functions are as flexible as the functions of unimodal BCs: Dittmann and Llewellyn (1968) explain for example that *yeah* combined with a nod can signal that the listener wants the floor to ask a question.

3 Data

Considering the research presented in the previous section, that often described unimodal BCs, we want to know what the gaze pattern is in a sequence that contains a unimodal or bimodal BC, where BCs occur in relation to the main speaker's turn and within a discourse unit, and if different types of BCs have different functions. In order to answer these questions, a subsection of the ENVID Corpus (Lelandais and Ferré, 2016) was used that consisted of two 30-minute dyadic interactions. This collaborative corpus was video-recorded between 2000 and 2012 in France and the UK. All participants were native speakers of British English who knew each other well and were video-recorded in soundproof studios to guarantee the sound and image quality of the recordings. They were free to discuss any topic they chose. For the two dialogues in this study, they were seated opposite each other and were filmed by two cameras. Each participant was also wearing a lavalier microphone, providing two separate audio tracks to enable the treatment of overlapping speech.

The corpus had already been edited in FinalCut for previous research to align the images from the two cameras and the soundtracks. It had then been transcribed using PRAAT (Boersma and Weenink, 2009) and gaze direction as well as head movements had also been coded at large previously using ELAN (Sloetjes and Wittenburg, 2008).

Interrater coding reliability between the second author and the initial coder across the three types of head movements described below on one of the two dialogues (364 head movements playing a BC role or not) was .72, as measured by Cohen's K.

3.1 Backchannel identification

None of the previous research on this corpus focused on backchannels, so these had to be coded as such. We coded three types of BCs: verbal, visual and bimodal. The verbal BCs we considered (189 occ) were single occurrences of *(oh) yeah*, *(mh)mh*, *(oh/all) right*, *oh*, *ah*, *really* and *okay*, which were the most common BCs in our corpus. They were not counted if accompanied by further speech or when delivered as an answer to a question. Other single BCs like *wow* or *good* were not numerous enough in our recordings (less than 20 occurrences in total) to be included in this study.

Head movements coded as visual BCs in this study were the same as the ones taken into account in Boholm and Allwood (2010): *nods* (vertical head movements, including what some distinguish as jerks), *shakes* (horizontal head movements) and *tilts* (head leaning towards shoulder). To be considered as BCs, head movements had to be communicative and had to be made by the listener. Head movements coming immediately after questions were not treated as BCs since they could be answers to these questions. The visual category (178 occ) includes head movements that appeared as single BCs and did not accompany any speech.

BCs were coded as bimodal (100 occ) when one of the head movements just described accompanied one of the verbal BCs under study. Head movements accompanying any other stretch of speech (like short responses or the beginning of

Speech/Gesture	nods	shakes	tilts	none	TOTAL
(oh) yeah	49	0	2	69	120
(mh)mh	36	0	0	57	93
(oh/all) right	4	0	1	6	11
oh	1	0	2	38	41
(oh) okay	4	0	0	3	7
ah	1	0	0	7	8
really	0	0	0	9	9
none	142	11	25	0	178
TOTAL	237	11	30	189	467

Table 1: Number of BC occurrences showing the combinations of head movements and speech tokens in two 30-minute dialogues

a full speech turn) were not taken into consideration, nor were any head movements that would come immediately after a question by the speaker. To count as a bimodal BC, the verbal utterance and the head movement had to be in overlap, which may have been partial. The most frequent configuration is that the verbal utterance, being shorter, is fully inserted in the gesture unit as in Example (1) below but we also found examples in which the verbal utterance started quite late in the gesture unit and continued after the head movement was completed or vice-versa as in Example (2).

(1) Hairdresser (ENVID, J:E)

```
1. E: I think she was nicer
2.    than the girl I had
3. J:                    yeah
                       <-nod->
```

(2) Best friends (ENVID, J:E)

```
1. E: I think her best friends
2.    are probably us (.)
3. J:                     yeah
                            <-nod->
```

Table 1 provides the number of occurrences of every possible verbal and visual combination met in the corpus.

3.2 Backchannel placement

Once all the BCs were coded as verbal, visual and bimodal, we noted gaze direction before, during and after BCs in three different ELAN tracks. This included gaze direction of speaker and listener (the participant who backchannels). Gaze direction before and after BCs was considered in the couple of video frames that immediately preceded and followed the BC. Gaze direction during BC production was noted as well but any change of gaze direction occurring during the BC was not considered since it could not have triggered or prevented it.

We also noted when BCs occurred with respect to the main speaker's channel: BCs could occur while the other participant was still speaking, or during a pause. Since BCs can be quite long other possible configurations for their occurrence were: speech + pause (beginning during the other participant's speech and ending during a following pause), pause + speech (beginning during a pause and ending during the start of a new turn by the other participant) or even speech + pause + speech for the longer ones (beginning during speech and being sustained till after speech is resumed by the other participant after a pause).

Still in terms of placement, we defined the conversational topics and subtopics in each dialogue, adopting the methodology of Grosz and Sidner (1986). The corpus counted 109 (sub)topics. Their mean duration was 1 min 63 sec. The shortest one lasted 9 sec and the longest 6 min 10 sec. We divided each (sub)topic into three equal parts to determine the position of each BC as occurring at the beginning (first section), in the middle (second section) or at the end (last section) of the (sub)topic.

3.3 Backchannel functions

In a last step, we coded the perceived function of BCs which could be one of the following: the continuer function (220 occ) was noted when the BC did not reveal any particular stance by the listener and it could be interpreted as "I see what you mean" or "I understand your viewpoint". BCs

were coded as agreements (66 occ) when they expressed a stance and could be interpreted as "I agree with what you say". They were coded as assessments (111 occ) when they conveyed any form (positive or negative) of judgment or evaluation by the listener. And they were coded as follow up (70 occ) when they directly followed another BC in accordance with McCarthy (2003).

Interrater reliability between the authors across the four backchannel functions for the BCs in one of the two dialogues (155 BCs) was .56, as measured by Cohen's κ. Discrepancies in the ratings were resolved by discussion.

4 Results

To answer our research questions, we used a series of Generalized Linear Mixed Models (GLMMs) fit by maximum likelihood estimation using the R 3.4.0 statistical programming language (R Core Team, 2012) and the lme4 package (Bates et al., 2014). Because there was quite a large variation between speakers and dialogues in the production of BCs as shown in Table 2, we systematically included Speaker and Dialogue as random factors in the models.

4.1 Gaze-pattern in sequences that contain a BC

4.1.1 Speaker gaze

We first explored possible interactions among the three BC types (fixed factor = Type; values = bimodal, verbal and visual) and gaze direction of the main speaker (fixed factor = Gaze towards co-participant; values = yes; no) before the co-participant produces a BC. The main effect of gaze direction was significant for bimodal BCs (β = 1.82, SE = .38, p = .001), as well as for verbal BCs (β = -1.06, SE = .32, p = .001) and more marginally for visual BCs (β = -.07, SE = .35, p = .02). The left hand graph in Figure 1 shows that the proportion of bimodal BCs produced as the speaker is gazing at the listener (the one who produces the BC) is very high (86 %). It is a little lower for visual BCs (76 %) and lower still for verbal BCs (68 %). Yet we can say that all BC types are generally triggered by speaker gaze towards listener, their total proportion being well over 50 %.

Considering possible interactions among the three BC types (fixed factor = Type; values = bimodal, verbal and visual) and gaze direction of the main speaker (fixed factor = Gaze towards co-participant; values = yes; no) while the co-participant produces a BC, we found that the main effect of gaze direction was significant for bimodal BCs (β = 1.62, SE = .33, p = .001), as well as for verbal BCs (β = -1.10, SE = .31, $p < .001$) and more marginally for visual BCs (β = -.08, SE = .33, p = .01). As shown in Figure 1, the proportion of bimodal BCs produced while speaker is gazing at co-participant remains quite high (84 %), while it is lower for visual BCs (71 %) and verbal BCs (63 %).

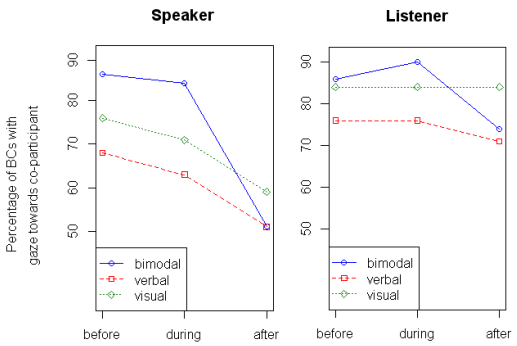

Figure 1: Percentage of speaker/listener gaze towards the other participant before, during and after the production of bimodal, verbal and visual BCs

Lastly, we examined possible interactions among the three BC types (fixed factor = Type; values = bimodal, verbal and visual) and gaze direction of the main speaker (fixed factor = Gaze towards co-participant; values = yes; no) after the co-participant has produced a BC. We found no effect of gaze direction for bimodal BCs (β = .12, SE = .34, p = .7), for visual BCs (β = .09, SE = .26, p = .7) or verbal BCs (β = .12, SE = .34, p = .7). The graph in Figure 1 shows that speaker gaze direction towards co-participant drops to 51 % after the latter has produced a bimodal BCs. The proportion of visual and verbal BCs after which speaker still gazes at co-participant is of the same order as for bimodal BCs (59 and 51 % respectively).

4.1.2 Listener gaze

We applied a similar GLMM model to listeners before, during and after they produced BCs (fixed factor = Gaze towards co-participant; values = yes; no) to see if there was an interaction with BC type (fixed factor = Type; values = bimodal, verbal and visual). We found that the main effect of gaze di-

Speaker	bimodal	verbal	visual	TOTAL
Dial.A: Elena	42	93	78	213
Dial.A: Joey	32	53	6	91
Dial.B: Michelle	21	29	58	108
Dial.B: Zoe	5	14	36	55
TOTAL	100	189	178	467

Table 2: Number of BC types produced by the 4 participants in the two 30-minute dialogues

rection was significant for bimodal BCs (β = 2.16, SE = .58, $p < .001$) before listeners backchannel. There also was an effect of gaze direction before the production of verbal BCs (β = -.8, SE = .34, p = .01). There was however no effect of gaze direction before the production of visual BCs (β = -.07, SE = .36, p = .8). The right hand side of the graph in Figure 1 shows that listeners gaze at speakers 86 % of times before the production of bimodal BCs. Visual BCs are not very different since listeners gaze at speakers 84 % of times before their production. Verbal BCs have a lower percentage than the other two types with 76 %.

Considering possible interactions among the three BC types (fixed factor = Type; values = bimodal, verbal and visual) and gaze direction of the listener (fixed factor = Gaze towards co-participant; values = yes; no) during the production of BCs, we found that the main effect of gaze direction was significant for bimodal BCs (β = 2.56, SE = .63, p = .001), as well as for verbal BCs (β = -1.31, SE = .39, p = .001) but not for visual BCs (β = -.30, SE = .41, p = .4). Here again, Figure 1 shows that listeners gaze at speakers 90 % of times during the production of bimodal BCs. Visual BCs are not very different since listeners gaze at speakers 84 % of times during their production. Verbal BCs have a lower percentage than the other two types with 76 %.

Lastly, we explored possible interactions among the three BC types (fixed factor = Type; values = bimodal, verbal and visual) and gaze direction of the listener (fixed factor = Gaze towards co-participant; values = yes; no) immediately after the production of BCs. We found no effect of gaze direction for bimodal BCs (β = .12, SE = .34, p = .7), for visual BCs (β = .09, SE = .26, p = .7) or verbal BCs (β = .04, SE = .25, p = .8). As shown in the graph in Figure 1, gaze direction of the listener towards co-participant drops to 74 % of times before the production of a bimodal BC, and almost reaches the proportion of gaze direction towards co-participant before the production of verbal BCs (71 %). Interestingly, gaze towards co-participant after the production of visual BCs is sustained at 84 %.

4.1.3 Mutual gaze

The models built so far told us about gaze direction of speaker and listener before, during and after BC production but we also wanted to know if there is an interaction among the three BC types (fixed factor = Type; values = bimodal, verbal and visual) and mutual gaze of both participants throughout the whole sequence (fixed factor = Mutual gaze; values = yes; no) so as to know if gaze is more often sustained by speakers and listeners in some BC types as compared with others. There was a significant main effect of mutual gaze on BC type for visual BCs (β = .57, SE = .26, p = .03) for which gaze towards the other participant is generally sustained throughout the whole sequence. There was also a significant main effect of mutual gaze on BC type for bimodal BCs (β = -.84, SE = .21, $p < .001$) for which mutual gaze is less sustained throughout the whole sequence. There was no effect of mutual gaze on BC type for verbal BCs (β = .08, SE = .26, p = .7).

4.2 BC occurrence within a main speaker's turn (overlap)

We then explored possible interactions among the three BC types (fixed factor = Type; values = bimodal, verbal and visual) and their occurrence within the main speaker's turn (fixed factor = Overlap; values = pause; pause-speech; speech; speech-pause; speech-pause-speech). The main effect of overlap was significant for bimodal BCs (β = 1.62, SE = .36, p = .001), as well as for verbal BCs (β = -1.88, SE = .31, p = .001) and visual BCs (β = 1.06, SE = .40, p = .008).

Figure 2 shows where verbal, bimodal and visual BCs occur with respect to the main speaker's turn. Whereas verbal BCs occur for a large

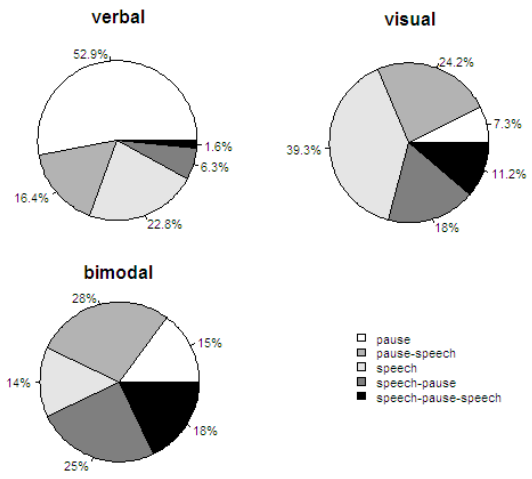

Figure 2: Distribution of verbal, visual and bimodal BCs with respect to the main speaker's turn

majority during a pause of the main speaker, a higher percentage of visual BCs overlap the main speaker's speech and bimodal BCs show a very evenly distributed proportion of each type of overlap, which means they may occur equally during speech or during pauses.

The difference in distribution of the three BCs may be explained by a difference in duration of verbal, visual and bimodal BCs, as represented in Figure 3. The main effect of duration was significant for bimodal BCs (β = 6.95, SE = .11, p = .001, mean duration = 1174.7 ms), as well as for verbal BCs (β = -.93, SE = .07, p = .001, mean duration = 479.5 ms) and more marginally for visual BCs (β = -.13, SE = .05, p = .01, mean duration = 929.6 ms).

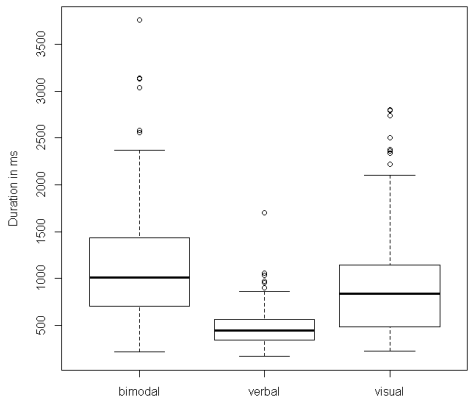

Figure 3: Duration (in ms) of bimodal, verbal and visual BCs

4.3 BC occurrence within discourse units

We first explored possible interactions among the three BC types (fixed factor = Type; values = bimodal, verbal and visual) and their position within discourse units (fixed factor = Position; values = beginning, middle, end). The main effect of type was significant for position (β = 1.19, SE = .33, p < .001) with bimodal BCs occurring more often at the beginning of the discourse topic than other BCs. The middle position showed no significant interaction with BC type (β = .19, SE = .29, p = .50). There wasn't any significant effect of the end position and BC type either (β = 17, SE = .28, p = .52).

In a second step, we also tested a possible interaction between unimodal *(mh)mh*, *(oh)yeah* and *nod* and BC position within discourse units (fixed factor = Position; values = beginning, middle, end). The main effect of position was significant for *(oh) yeah* which occurs slightly less often in the middle of (sub)topics than the other BCs (β = 1.59, SE = .47, p < .001) but there was no effect of position on *(mh)mh* (β = -.23, SE = .43, p = .59) or *nod* (β = -.19, SE = .36, p = .58). Looking at single nods themselves, we found that 33 occurred at the beginning of a (sub)topic, while 52 and 53 occurred in the middle and at the end of discourse (sub)topics respectively so that they are quite evenly distributed among the three positions.

4.4 BC functions

We then tested possible interactions among the three BC types (fixed factor = Type; values = bimodal, verbal and visual) and their functions (fixed factor = Function; values = agreement, assessment, continuer, follow up). The main effect of function was significant for bimodal BCs (β = .98, SE = .26, p < .001), as well as for verbal BCs (β = 1, SE = .31, p = .001) and visual BCs (β = 1.28, SE = .34, p < .001).

Figure 4 shows the distribution of functions for each type of BC and reveals that whereas visual BCs are more often classified as continuers than the others, bimodal BCs have an agreement function more often than the other two types of BCs and verbal BCs are more frequently used to express assessment or follow up than the other two BC types.

Finally, we tested whether there was a possible interaction between the functions of BCs (fixed factor = Function; values = agreement, as-

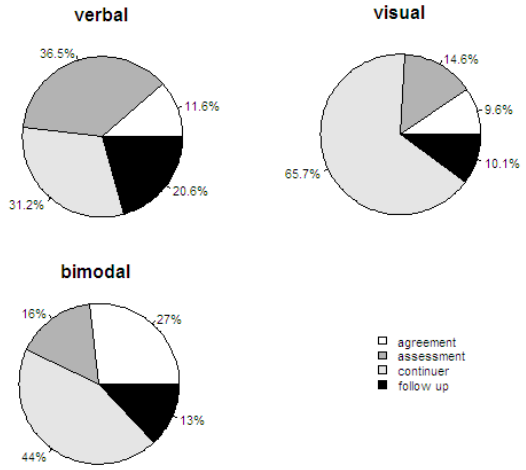

Figure 4: Distribution of functions (agreement, assessment, continuer, follow up) in verbal, bimodal and visual BCs

sessment, continuer, follow up) and their position within the discourse unit (fixed factor = Position; values = beginning, middle, end). The main effect of position was significant for BCs used to mark agreement (β = 1.26, SE = .30, p = .001) as they appear preferentially at the end of the discourse topic. We cannot say however that continuers are distinguished in a significant way from other BCs in terms of placement in the discourse unit and they do not occur significantly earlier than other BCs (β = -.25, SE = .33, p = .43). Lastly, follow ups do not occur later in the discourse topic than other BCs as we might also have expected (β = -.02, SE = .40, p = .95).

5 Discussion

In terms of placement, we have shown that mutual gaze is a strong condition for the production of a BC whatever its type which confirms previous results (Hjalmarsson and Oertel, 2011; Poppe et al., 2011), it also confirms previous findings showing that the condition is stronger for visual and bimodal BCs than for verbal BCs (Bertrand et al., 2007). We add to this that there is a difference not only in the context immediately preceding the BC, but also in the fact that for visual BCs, mutual gaze is more often sustained throughout the whole sequence that contains a BC than for verbal BCs, whereas for bimodal BCs, mutual is less sustained throughout the sequence with a drop of gaze towards co-participant immediately after the production of a BC.

With respect to the main speaker's turn, we refine previous results. We concur with Truong et al. (2011) that verbal BCs are preferably used during pauses whereas BCs with a visual component, being less disruptive, are more likely to occur during speech. However, graphs show that whereas bimodal BCs may occur anywhere within the main speaker's turn, unimodal visual BCs are preferentially produced during speech. That bimodal BCs may occur in overlap with speech although they contain a verbal element can be explained by the fact that they are not just simply a superposition of a verbal and a visual BC, but they are also longer than unimodal visual BCs. Their greater length can be explained by the fact that the head movement itself is longer in a bimodal BC than in a visual BC: the listener initiates a head movement, most of the time during speech, and while sustaining that head movement, adds a verbal token when there is a pause in the main speaker's turn.

Our study did not confirm that nods are preferentially placed in mid-telling position (Stivers, 2008) or that *(oh) yeah* would be preferred over *mhmh* at the end of a topic (Jefferson, 1983). In our corpus, *(mh) mh* and nods were evenly distributed at the beginning, in the middle and at the end of (sub)topics, whereas *(oh) yeah* occurs less in the middle section of the (sub)topic. The only constraint we found concerning the placement within a discourse (sub)topic concerns bimodal BCs which tend to occur more at the beginning than in the middle or at the end. A possible explanation for these differences is that both Stivers and Jefferson examined the occurrence of BCs within narrative parts of speech whereas our study did not distinguish between different discourse types. If there is a constraint in BC placement in narrative, this may not hold for non-narrative parts of speech. Bertrand and Espesser (2017) have also shown that listeners tend to produce more complex BCs as narrative delivery is unfolding in time. The simple BCs considered in the present study may therefore not be constrained by placement.

We did however find differences among the three types of BCs concerning their function as hypothesized by Wlodarczak et al. (2012), although perhaps not the differences one would have expected. Our intuitive idea was that a bimodal BC would have more communicative weight than a unimodal one and would therefore be more likely

to express agreement and assessment, i.e. be a marker of affiliation (Stivers, 2008), than a unimodal BC. Our results show that this is only partly true and that BCs are more specialized than this. Visual BCs are in a large majority used as continuers, which is in perfect agreement with our findings that, being less disruptive, they are also more often produced in overlap with the main speaker's turn. Bimodal BCs are more often used than unimodal BCs to express agreement. Yet, assessment is more often expressed with the use of verbal BCs. This is explained by the fact that verbal BCs are more varied than visual ones and tokens like *all right* or *really* for instance are more likely to express assessment in their semantic content than nods. Another reason for this is that verbal BCs are modulated by intonation contours, which is not the case of visual BCs. One should enquire further however why bimodal BCs, which contain a verbal component (and therefore a possibility of intonation modulation), are not used more often to express assessment than verbal BCs.

Finally, we found that although BC types are not constrained in placement within discourse (sub)topics as they are quite multifunctional as shown in Figure 4, we did find a link between the functions of BCs and their placement within a discourse unit. Contrary to what we expected, the least affiliative BCs (continuers) do not occur earlier in a discourse unit than more affiliative BCs like assessments. However, BCs marking agreement occur later, namely when the listener has sufficient information to be able to express a stance. Follow up BCs do not occur later in the (sub)topic than agreements and assessments which means they are not used as end-of-topic markers, probably because their domain is the speech turn rather than a larger discourse unit, as suggested by McCarthy (2003) who also calls them "third-turn receipts".

6 Conclusion

In this paper, we presented a study of BCs in conversational English, based on a corpus of two 30-minute dialogues. Most studies so far have described verbal and visual BCs, and very little research has been conducted on bimodal BCs in a comparative perspective. Our aim was to establish if there are differences between verbal (*(oh) yeah*, *mhmh*, etc.), visual (*nod*, *tilt*, *shake*) and bimodal BCs both in terms of placement in the main speaker's turn or within the discourse (sub)topics and in terms of their function.

Our main findings were that whereas mutual gaze between participants strongly favors the production of a BC, mutual gaze is more often sustained during and after visual BCs than during and after verbal and bimodal ones. There is also a clear distinction between verbal and visual BCs concerning their placement within the main speaker's turn. Whereas verbal BCs occur preferentially during pauses, visual BCs occur mostly during speech. Bimodal BCs show no such restriction and occur both during speech and pauses. The explanation for this is that they are much longer than the other two types of BC. The only difference we found concerning placement within a discourse topic is that bimodal BCs occur earlier in the (sub)topic than the other two types. Considering their functions, we found that visual BCs are more often used as continuers. Bimodal BCs are more often used as agreement tokens than the other two types and verbal BCs are more often used as assessments than the other two. Finally, we found that there is a correlation between one function played by BCs and BC position within a discourse (sub)topic. More affiliative BCs marking agreement occur later in the discourse (sub)topic, namely when the listener has sufficient information to be able to express a stance, but contrary to expectation, the least affiliative BCs (continuers) do not occur earlier in a discourse unit than more affiliative BCs marking agreement or assessment.

These results are very encouraging but the corpus is still limited in length with only 467 BCs considered. Future research could not only enlarge the corpus, but also vary the type of interaction to give a fuller picture of BCs. If these preliminary results were to be confirmed, this could be a tremendous asset for research on human-machine communication and the development of virtual agents.

Acknowledgments

We would like to thank the participants in the ENVID corpus and are particularly indebted to three anonymous reviewers for their highly constructive comments on a previous version of this paper.

References

Douglas Bates, Martin Maechler, Ben Bolker, and Steven Walker. 2014. Linear mixed-effects models

using eigen and s4 [online: http://cran.r-project.org].

Janet B. Bavelas, Linda Coates, and Trudy Johnson. 2000. Listeners as co-narrators. *Journal of Personality and Social Psychology*, 79(6):941–952.

Stefan Benus, Augustin Gravano, and Julia Hirschberg. 2007. The Prosody of Backchannels in American English. In *ICPhS XVI*, pages 1065–1068, Saarbrücken, Germany.

Roxane Bertrand and Robert Espesser. 2017. Co-narration in French conversation storytelling: A quantitative insight. *Journal of Pragmatics*, 111:33–53.

Roxane Bertrand, Gaëlle Ferré, Robert Espesser, Stéphane Rauzy, and Philippe Blache. 2007. Backchannels revisited from a multimodal perspective. In *AVSP*, pages 1–5, Hilvenbareek, The Netherlands.

Elisabetta Bevacqua, Sathish Pammi, Sylwia Julia Hyniewska, Marc Schrï¿½der, and Catherine Pelachaud. 2010. Multimodal Backchannels for Embodied Conversational Agents. In Jan Allbeck, Norman Badler, and Timothy Bickmore, editors, *IVA 2010, LNAI 6356*, pages 194–200. Springer-Verlag, Berlin, Heidelberg.

Paul Boersma and David Weenink. 2009. Praat: doing phonetics by computer (Version 5.1.05) [Computer program].

Max Boholm and Jens Allwood. 2010. Repeated head movements, their function and relation to speech. In *LREC*, pages 1–5, Valleta, Malta.

Patricia M. Clancy, Sandra A. Thompson, Ryoko Suzuki, and Hongyin Tao. 1996. The conversational use of reactive tokens in English, Japanese, and Mandarin. *Journal of Pragmatics*, 26:355–387.

Allen T. Dittmann and Lynn G. Llewellyn. 1968. Relationships Between Vocalizations and Head Nods as Listener Responses. *Journal of Personality and Social Psychology*, 9(1):79–84.

Kent Drummond and Robert Hopper. 1993. Back Channels Revisited: Acknowledgment Tokens and Speakership Incipiency. *Research on Language and Social Interaction*, 26(2):157–177.

Rod Gardner. 2001. *When Listeners Talk*. John Benjamins, Amsterdam.

Charles Goodwin. 1986. Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies*, 9(2):205–217.

Agustin Gravano and Julia Hirschberg. 2009. Backchannel-Inviting Cues in Task-Oriented Dialogue. In *Interspeech*, pages 1019–1022, Brighton, UK.

Barbara J. Grosz and Candace L. Sidner. 1986. Attention, Intention, and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204.

Mattias Heldner, Anna Hjalmarsson, and Jens Edlund. 2013. Backchannel relevance spaces. In E. L. Asu and P. Lippus, editors, *Nordic Prosody: Proceedings of the XIth Conference, Tartu 2012*, pages 137–146. Peter Lang, Frankfurt am Main.

Anna Hjalmarsson and Catharine Oertel. 2011. Gaze direction as a backchannel inviting cue in dialogue. In *the IVA 2011 workshop on Realtime Conversational Virtual*, pages 1–8, Reykjavik, Iceland.

Saya Ike. 2010. Backchannel: A feature of Japanese English. In *JALT2009*, pages 1–11, Tokyo, Japan.

Gail Jefferson. 1983. Notes on a systematic deployment of the acknowledgement tokens "yeah" and "mm hm". *Tilburg Papers in Language and Literature*, 30:1–18.

Kathrin Lambertz. 2011. Back channelling: the use of yeah and mm to portray engaged listenership. *Griffith Working Papers in Pragmatics and Intercultural Communication*, 4(1-2):11–18.

Seung-Hee Lee and Hiroko Tanaka. 2016. Affiliation and alignment in responding actions. *Journal of Pragmatics*, 100:1–7.

Manon Lelandais and Gaëlle Ferré. 2016. Prosodic boundaries in subordinate syntactic constructions. In *Speech Prosody*, pages 183–187, Boston, USA.

Michael McCarthy. 2003. Talking Back: "Small" Interactional Response Tokens in Everyday Conversation. *Research on Language and Social Interaction*, 36(1):33–63.

Mary-Annick Morel and Laurent Danon-Boileau. 2001. Les productions sonores de l'écouteur du récit : coopération ou subversion. *Revue Québécoise de Linguistique*, 29(1):71–96.

Ronald Poppe, Khiet P. Truong, Dennis Reidsma, and Dirk Heylen. 2010. Backchannel Strategies for Artificial Listeners. In *IVA 2010*, pages 146–158, Berlin, Heidelberg.

Ronald Poppe, Khiet P. Truong, and Dirk Heylen. 2011. Backchannels: Quantity, Type and Timing Matters. In *IVA 2011*, pages 228–239, Reykjavik, Iceland.

Han Sloetjes and Peter Wittenburg. 2008. Annotation by category - ELAN and ISO DCR. In *6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco.

Tanya Stivers. 2008. Stance, Alignment, and Affiliation During Storytelling: When Nodding Is a Token of Affiliation. *Research on Language and Social Interaction*, 41(1):31–57.

R Core Team. 2012. A language and environment for statistical computing. r foundation for statistical computing. [online: http://www.r-project.org].

Allison Terrell and Bilge Mutlu. 2012. A Regression-based Approach to Modeling Addressee Backchannels. In *13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 280–289, Seoul, South Korea.

Jackson Tolins and Jean E. Fox Tree. 2014. Addressee backchannels steer narrative development. *Journal of Pragmatics*, 70:152–164.

Khiet P. Truong, Ronald Poppe, Iwan de Kok, and Dirk Heylen. 2011. A multimodal analysis of vocal and visual backchannels in spontaneous dialogs. In *Interspeech*, pages 2973–2976, Florence, Italy.

Marcin Wlodarczak, Hendrik Buschmeier, Zofia Malisz, Stefan Kopp, and Petra Wagner. 2012. Listener head gestures and verbal feedback expressions in a distraction task. In *Interdisciplinary Workshop on Feedback Behaviors in Dialog, INTERSPEECH2012 Satellite Workshop*, pages 93–96, Stevenson, WA.

Takashi Yamaguchi, Koji Inoue, Koichiro Yoshino, Katsuya Takanashi, Nigel G. Ward, and Tatsuya Kawahara. 2015. Analysis and Prediction of Morphological Patterns of Backchannels for Attentive Listening Agents. In *7th International Workshop on Spoken Dialog Systems (IWSDS)*, pages 1–12, Riekonlinna, Finland.