Title

Backchannel Behavior in Child-Caregiver Zoom-Mediated Conversations

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors

Bodur, Kübra
Nikolaus, Mitja
Fourtassi, Abdellah
et al.

Publication Date

2022

Peer reviewed

Backchannel Behavior in Child-Caregiver Video Calls

Kübra Bodur[1] (kubra.bodur@univ-amu.fr)
Mitja Nikolaus[1,2] (mitja.nikolaus@univ-amu.fr)
Laurent Prévot[1] (laurent.prevot@univ-amu.fr)
Abdellah Fourtassi[2] (abdellah.fourtassi@univ-amu.fr)

[1]Aix-Marseille Université, CNRS, LPL, Aix-en-Provence, France
[2]Aix Marseille Université, Université de Toulon, CNRS, LIS, Marseille, France

Abstract

An important step in children's socio-cognitive development is learning how to engage in coordinated conversations. This requires not only becoming competent speakers but also active listeners. This paper studies children's use of backchannel signaling (e.g., "yeah!" or a head nod) when in the listener's role during conversations with their caregivers via video call. While previous work had found backchannel to be still immature in middle childhood (i.e., 6 to 11 years of age), our use of both more natural/spontaneous conversational settings and more adequate controls allowed us to reveal that school-age children are strikingly close to adult-level mastery in many measures of backchanneling. The broader impact of this paper is to highlight the crucial role of social context in evaluating children's conversational abilities.

Keywords: cognitive development; language acquisition; conversation; nonverbal; backchannel

Introduction

Conversation might seem to adults as an effortless social activity. In reality, it relies on complex linguistic and socio-cognitive skills that children have to master starting in early childhood and well into middle childhood (e.g., Matthews, 2014). The healthy development of these skills has a large impact on children's ability to learn from others, maintain relationships in life, and collaborate with peers. If this development is impaired, it can have negative consequences from the risk of developing mental health issues to the quality of academic attainment and employability (e.g., Murphy et al., 2014).

This paper focuses on the development of an important conversational skill — that has received surprisingly little attention in the language development literature — known as Back-Channelling (hereafter, BC).

Backchanneling in Conversation. A backchannel (Yngve, 1970) is a communicative feedback that the *listener* provides to the speaker in a non-intrusive fashion such as short vocalizations like "yeah" and "uh-huh", and/or nonverbal cues such as head nods and smiles. Despite not having a narrative content, BC is a crucial element in successfully coordinated conversations, signaling, e.g., attention, understanding, and agreement (or lack thereof) while allowing the speaker to make the necessary adjustments toward achieving mutual understanding or "communicative grounding" (Clark, 1996).

Types of Backchanneling. Bavelas et al. (2000) have established experimentally the difference between two types of BC transmitted by the listener in a conversation: "generic" and "specific."[1] As its name indicates, *specific BC* is a reaction to the content of the speaker's utterance. It might indicate the listener's agreement/disagreement, surprise, fear, etc. As for *generic BC*, it is performed to show that the listener is paying attention to the speaker and keeping up with the conversation without conveying a narrative content (see examples in Table 2).

It is important to note that, while generic BC does not target the narrative content, it does not mean that it is used randomly. Both generic and specific BCs should be timed precisely and appropriately so as to signal proper attentive listening to the speaker. Otherwise, they can be counter-productive and perceived, rather, as distracting and interrupting (Park et al., 2017). In that sense, both specific and generic BC are *collaborative* mechanisms.

Development of Backchanneling. While some studies have documented early signs of children's ability to both interpret and provide BC feedback in the preschool period (Shatz & Gelman, 1973; Peterson, 1990; Park et al., 2017), a few have pointed out that this skill continues developing well into middle childhood.

For example, Dittmann (1972) analyzed conversations of 6 children between the ages of 7 and 12 in a laboratory setting where children conversed with adults and other children as well as children interacting with each other at school. The results found there to be fewer BC signals produced by young children in this age range compared to the older group (between 14 and 35 years old).

Following Dittmann (1972)'s study, Hess & Johnston (1988) aimed at providing a more detailed developmental account of BC behaviors in middle childhood using a task where children listened to board game instructions from an experimenter. In particular, the authors analyzed children's BC production in various *speaker's cues* such as pauses greater than 400ms, speaker's eye gaze toward the listener, and speaker's clause boundaries. These cues/contexts were not spontaneous, however. They were predesignated by the experimenter to essentially become invariant throughout the task. Hess and Johnston found that younger children produced less BC compared to older ones.

[1]A roughly similar distinction has been proposed in Conversational Analysis literature (Schegloff, 1982; Goodwin, 1986), using the terms "continuers" and "assessment."

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

Both Dittmann (1972) and Hess & Johnston (1988) have results indicating that children continue learning to provide collaborative BC and to better anticipate the speaker's cues during middle childhood, suggesting that their behavior as an age group is still immature compared to that of adults.

The current study and novelty of our work

While Dittmann (1972) and Hess & Johnston (1988) provided important insights about BC development, they both analyzed BC when children were engaged in conversations that may not be the ideal context to elicit and characterize children's full conversational competence. Some conversations were recorded with strangers (or in the presence of a stranger), in a laboratory, and/or with non-spontaneous predesigned scripts.

We argue that children's conversational skills could have been *under-estimated* in these previous studies because of their focus on contexts that are unnatural to the child and are not similar to how they communicate spontaneously in daily life. Indeed, research has shown that context can influence the nature of the conversational behavior more generally (e.g., Dideriksen et al., 2019).

In the current work, we study BC behavior in middle childhood (i.e., ranging, roughly speaking, from 6 to 11 years of age) while increasing, to the extent possible, the ecological validity of the data. The goal is to provide a socially comfortable context where children could show a more natural use of their conversational skills.

More precisely, we collect data where children a) talk one-on-one with their caregivers (as opposed to a stranger/experimenter), b) are at home (as opposed to the lab, or even the school), c) converse to play a fun, easy, and natural game (word-guessing) that many children are already familiar with, as opposed to scripted turns or complex conversational games such as the map task (Anderson et al., 1991) (typically used to prompt conversations in adults).

To achieve these goals, we capitalized on the recent increase in familiarity and use of online video calls by children due to Covid-19 pandemic. Children and caregivers were recorded talking to each other via video call while both at home (but sitting in different rooms and using different personal devices).[2]

To evaluate children's BC behavior compared to adult-level mastery, it is not enough to consider only the caregiver's behavior when talking to children. The reason is that caregivers tend to adapt to children's linguistic and conversational competencies (e.g., Snow, 1972; Misiek et al., 2020; Fusaroli et al., 2021; Leung et al., 2021). Thus, in addition to child-caregiver conversations, we need to examine how adults behave in situations involving other adults. We collect similar data involving the *same* caregiver talking either to another family member or to a non-family member.

[2]See Discussion for some possible limitations of this data acquisition method.

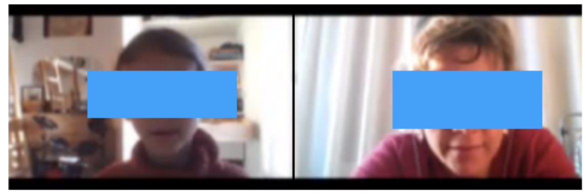

Figure 1: A snapshot of one of the recording sessions involving a child and her caregiver communicating through Zoom.

Using this data, we ask how children's BC behavior in middle childhood compares to adult-level mastery both in terms of the overall rate of production and in terms of its distribution across the speaker's broad contextual cues (e.g., BC overlapping with speech vs. during pauses).

We expect that our more natural setting would allow children to manifest more advanced BC skills compared to previous work. Furthermore, we predict that children's BC behavior would be closer to that found in adult-adult conversation, especially in the adult *family* dyads which provides a similar social context to that of child-caregiver.

Method

Data and participants

We collected a corpus of multimodal conversations where dyads of participants were recorded communicating via video call in two conditions:

Child-Caregiver conversation. We recorded 10 dyads of children talking to their caregivers using different laptops and communicating from different rooms at home. Children were between 6-and 11-years old (mean age was 8.7 years).

Adult-Adult conversation. We recorded the same caregivers talking to adults using the same procedure. Half of the caregivers talked to a family member and the other half talked to a non-family member (e.g., a colleague). In this adult-adult condition, we are not interested in differences between interlocutors (unlike the asymmetrical context of child-caregiver condition). Rather, we are interested in possible differences in behavior across the family vs. non-family dyads, which provide us with two characterizations of adult-level mastery against which we can compare children's maturity in terms of BC behavior.

Procedure

In both conditions, dyads of interlocutors engage in conversations that involve a combination of a word guessing game and spontaneous discussions. Lasting around 15 minutes in total, each recording consists of three stages: first, the Caregiver/Adult 1 explains the task, then the dyad starts playing the game for around 10 minutes and finally they initiate a more spontaneous conversation.

The word guessing game was straightforward: One interlocutor thinks of a word and the other tries to guess it by asking questions. After a word has been guessed, the interlocutors change their roles. In the third, more spontaneous

	Verbal		Head nod		Smile	
	G.	S.	G.	S.	G.	S.
Child	2.5	13.3	3.6	3.3	10	24.7
Caregiver	3.9	9.3	0.4	4.9	14.8	16.5
Family	3.3	12.1	2.0	6.7	21.9	14.9
Non-family	6.0	20.8	19.9	17.7	34.0	22.1

Table 1: Average number of Generic vs. Specific BC produced per participant in each modality.

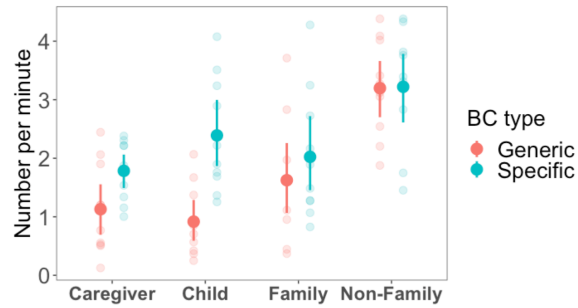

Figure 2: The rate of BC production per minute for both child-caregiver condition and family vs. non-family in the adult-adult condition. Each data point represents a participant and ranges represent 95% confidence intervals.

phase, the caregivers were (minimally) instructed to discuss freely with the interlocutor about how the game went and to comment on each other's performance in the game.

Video Call Software We used Zoom (Zoom Video Communications Inc., 2021) for video calls since participants were more familiar with this software. It is worth mentioning that Zoom has built-in audio-enhancing features that can be problematic for the study of BC. In particular, one feature consists in giving the stage to one speaker while suppressing background noise that may come from other participants' microphones. We made sure, in preliminary testing, that backchannel is not suppressed as "background noise," at least in our case where there are only two active participants in the zoom session.

Annotating Listener's BC.

We manually annotated the entire video recordings for the listener's BC. While a multitude of verbal and non-verbal behaviors can convey some form of *specific* active listening, here, we made this question more manageable by focusing on a subset of multimodal behaviors that can, a priori, be used not only for specific BC but *also* for generic BC.[3] In fact, this subset is a good test for children's ability to encode and decode specific vs. generic BC even when both can be expressed by the same behavior/modality. For example, a smile can be both an accommodating gesture (generic) or an expression of amusement (specific) depending on the context (see Table 2).

In addition to short vocalizations (e.g., "yeah", "m-hm"),[4] head nods and smiles were observed in our data to be used by children and adults both as specific and generic BC. Other non-verbal signals (e.g., head shakes, eyebrow displays, and laughs) were almost always specific, and therefore — as we indicated above — they were not included in the analysis.

The annotation of short vocalizations, head nods, and smiles for BC proceeded in two steps: 1) the target behavior was first segmented in time, and 2) if the behavior was recognized as a BC, it was tagged as specific or generic. For details about the first step, including methods for calculating inter-annotation agreement of segmentation in time, please see our resource conference paper (Bodur et al., 2021). As for the second step, the first author annotated the entire set of (already segmented) behaviors into specific vs. generic BC.

In order to estimate inter-rater reliability for the second step, around 20% of the data were annotated independently by a second annotator. We obtained a Cohen Kappa value of $\kappa = 0.66$, indicating "moderate" agreement (McHugh, 2012).

Table 1 shows some average statistics of Generic vs. Specific BC produced per participant in each modality. It shows that children and adults do produce *both* types of BC in each of the modalities we consider. Given our limited sample size and in order to optimize statistical power, all analyses below were done by collapsing BC across all these modalities.

Annotating the Speaker's cues

We also annotated the recordings for the speaker's cues where the listener's BC may occur. We considered two broad speaking contexts/cues: **a) Overlap with speech**, that is, when the listener produces a verbal or non-verbal BC without interrupting the speaker's ongoing voice activity, and **b) During a pause**, i.e., when the listener produces a BC after the speaker pauses for a minimum of 400ms (Following Hess & Johnston, 1988). In addition, we study the subset of cases where these cues overlap with the speaker's eye gaze, that is, when the listener produces a BC while the speaker is looking at them vs. while the speaker is looking away.

The speaker's continuous segments of speech (or Inter-Pausal Units, IPUs) were annotated automatically using a Voice Activity Detection software (Bigi, 2012; Bigi & Meunier, 2018). The speaker's gaze (looking at listener vs. looking away) was annotated manually (See Bodur et al., 2021).

Results

Rate of BC production

We computed the total number of specific and generic BC produced by each interlocutor, then we divided this number by the length of the conversation in each case to have a mea-

[3]Indeed, preliminary inspection of the data had suggested that behaviors used for generic BC are also used for specific BC (but not vice versa).

[4]If a short vocalization functioned as an answer to a question (e.g., "yeah" as answer to a yes-no question), it was not considered as a BC.

	BC type	Modality
- Parent: So, [ ] the game is a simple word guessing game - Child: Yeah! **Comment:** The child' BC shows that they are engaged in what the caregiver is telling them while still remaining as audience.	Generic	Verbal
- Adult1: It was a good way of training [ ] and also publishing .. - Adult2: Of course! **Comment:** Adult2 becomes involved in the narration process by acting upon Adult1's utterance: They display agreement.	Specific	Verbal
- Adult1:..in addition to having knowledge of EEG because I've never.. - Adult2: [Head nod] **Comment:** Adult2 nods while Adult1 hesitates during the utterance to show that they are anticipating the interlocutor to communicate their engagement.	Generic	Head nod
- Parent: That's right, it's a metronome [ ] Oh too strong! - Child : [Head nod] **Comment:** The child's head nod is internal to the narrative plot of the caregiver, they nod to add information to the narration by showing approval for what was said by the caregiver.	Specific	Head nod
- Adult 1: Ah because in the objects category you did not put living beings - Adult 2: [Smile] **Comment:** Adult2's smile is directed at the narrator while actively listening as the speaker to communicate general understanding and involvement.	Generic	Smile
- Adult 1: No, it's gonna be something so silly too. Um... - Adult 2: [Smile] **Comment:** Adult2's smile is a direct response to what was said by the speaker. It is specific to that point in the narrative and shows amusement.	Specific	Smile

Table 2: Excerpts from conversations in the corpus exemplifying both types of BC in each of the three modalities we consider in this study. BC either occurs during a pause "[ ]" in the speaker's turn or it overlaps with a segment of the speaker's speech (the underlined part). The **comment** explains the decision to classify the BC as generic or specific, following the distinctions made in Bavelas et al., 2000.

sure of the *rate* of production per minute. Results are shown in Figure 2. When comparing children to caregivers, we found what seems to be a developmental effect regarding the rate of production of specific vs. generic BC. More precisely, children produced specific BC at a higher rate than generic BC and this difference was higher for children than for caregivers. Statistical analysis confirmed this observation: A mixed-effect model predicting only children's rate as a function of BC type[5] yielded an effect of $\beta = 1.47$ ($SE = 0.17$, $p < 0.001$), meaning there was a difference between specific and generic BC rate in children. A second model predicting rate of production (by both children and caregiver) as a function of both BC type and interlocutor (child or caregiver)[6] showed there to be an interaction $\beta = 0.81$ ($SE = 0.37$, $p < 0.05$), meaning that the difference between BC types in children is larger than it is in caregivers. As can be seen in Figure 2, and although children produce slightly fewer generic BC, the developmental difference can be largely attributed to children producing more specific BC compared to caregivers.

In the same Figure, we also have the rates of production in the Adult-Adult control conditions. While these controls were meant to be contrasted with the child-caregiver dyads, here we noticed an interesting difference among them: Both BC types were higher in the non-family dyads than in the family dyads. Using a mixed-effects model predicting the rate of BC production as a function of type and family membership,[7] we found a main effect of family membership: $\beta = 0.81$ ($SE = 1.57$, $p = 0.01$) but there were no effect of BC type nor an interaction.

By comparing the child-caregiver condition to family vs. non-family of the adult-adult conditions in Figure 2, we make the following observations. Caregivers, when talking to children, produced BC at a rate roughly similar to the one used among adults in the family dyads. The same thing can be said about children: Although they tend to produce slightly more specific BC and slightly fewer generic BC than adults, these differences were small and not statistically significant. This comparison suggests that, overall, children are not less prolific in terms of BC production than adults, at least when adults converse in a family context.

Distribution of BC over Speaker's cues

Child-Caregiver condition While the results shown in Figure 2 inform us about the overall rate of BC production, they do not show the speaker's cues/contexts in which this production occurs. As explained in the Methods' section, we examined the nature of BC production of the listener during the speaker's speech vs. pause and in the subset of cases

[5] Specified as `Rate ~ BC_type + (1 | dyad)`

[6] Specified as `Rate ~ BC_type*Interlocutor + (1 | dyad)`

[7] Specified as `Rate ~ BC_type*Family + (1 | dyad)`

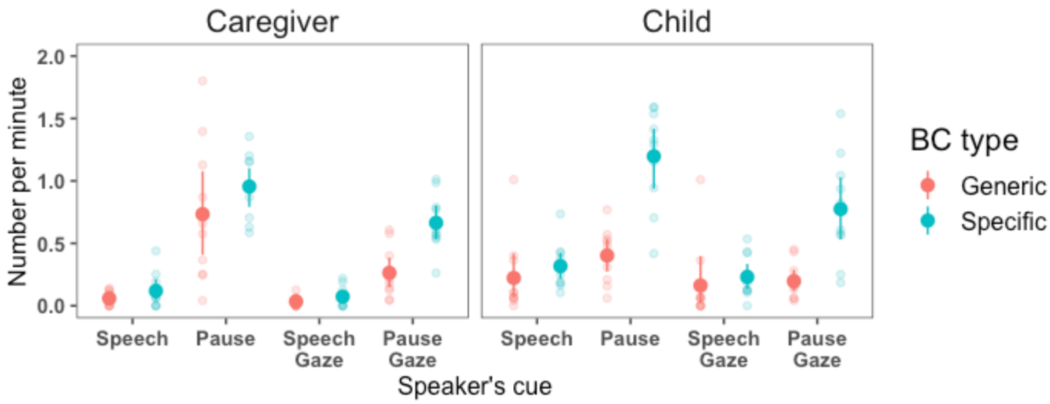

Figure 3: The rate of listener's BC production of children and caregivers across the speaker's speech and pauses (mutually exclusive). We also show the subset of cases where these cues overlap with the speaker's gaze at the listener. Each data point represents a participant and ranges represent 95% confidence intervals.

where these cues overlap with the speaker's gaze (Kendon, 1967; Kjellmer, 2009; Morency et al., 2010).

In Figure 3 we show the rate of BC production of children and caregivers across these cues. The main observation is that children and caregivers produce BC across speaker's cues in roughly similar proportions.

Another observation is that, for both interlocutors, while BC in speech almost always coincided with the speaker's gaze at the listener (though this is likely due to a floor effect), this was not the case for BC produced during pauses where part of this production did not coincide with the speaker's gaze (in other words, listeners also provided BC when the speaker paused and looked away).

One minor difference between children and caregivers is the following. While their rate of BC during speech was generally low, this rate was slightly higher for children (while almost totally absent for caregivers). The origin of this small difference is unclear: it could be due to children providing less BC opportunities for adults to capitalize on, or to adults not providing BC despite such opportunities. Another difference is that caregivers seem to produce slightly more generic BC after pauses than children do (though this difference was not statistically significant).

Adult-Adult (control) conditions Next, we examined how BC production varied across the speaker's cues between family and non-family dyads in the adult-adult control conversations. The results are shown in Figure 4. Unsurprisingly, and in line with findings in Figure 2, we observe a general increase in BC production rate among non-family dyads relative to family dyads.

However, this increase was interestingly not similar across the speaker's cues. In particular, while average BC production rate remained similar during the speaker's *pauses*, we observed a striking increase of generic BC produced by non-family listeners during the speaker's *speech*, going from almost zero to around 1 BC per minute. Indeed, a mixed-effects model predicting rate of generic BC as a function of family membership showed there to be a strong difference: $\beta = 0.84$ ($SE = 0.2$, $p < 0.001$).

Another observation is that, for both family and non-family dyads, only part of the BCs co-occurred with the speaker's eye gaze towards the listener in both speaker's speech and pause, i.e., many BC occurred when the speaker was looking away.

By comparing Figure 3 and Figure 4, we conclude that patterns of BC distribution (across speaker's cues) of children and caregivers are not only largely similar to each other, but also similar to the patterns of BC distribution of adult-adult conversations in the *family* context. In fact, we observed much more differences due to family membership between adults than differences due to developmental age.

Discussion

This paper studied BC in child-caregiver conversations and compared children's behavior to adult-level mastery in family and non-family contexts. While previous work (Dittmann, 1972; Hess & Johnston, 1988) found BC to be still relatively infrequent in middle childhood, here we found that children in the same age range produced BC at a similar rate as in adult-adult conversations where the adults were family members, which is a more pertinent control condition than when adults are not family members. In the latter, the rate of production of BC was much higher.

The findings confirm our prediction that improving the naturalness of the data collection context (i.e., conversation with a caregiver, at home, and using a fun/easy game) allows us to capture more of children's natural use of BC, which we found to be strikingly close to adult-level mastery.

BC opportunities An alternative interpretation is that caregivers provide children with more BC *opportunities* than what they would have received otherwise, "scaffolding" their BC behavior. While it is difficult to quantify exhaustively and

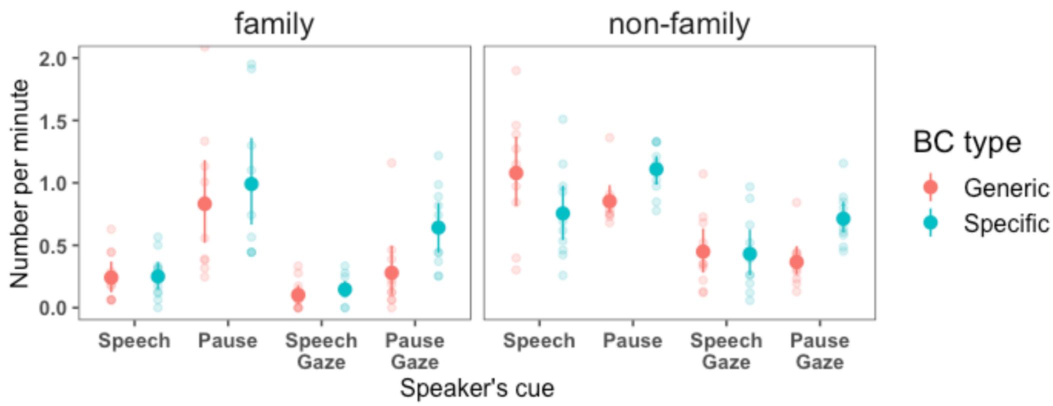

Figure 4: The rate of listener's BC production in family and non-family adult dyads across the speaker's speech and pauses (mutually exclusive). We also show the subset of cases where these cues overlap with the speaker's gaze at the listener. Each data point represents a participant and ranges represent 95% confidence intervals.

precisely all BC opportunities, we can have an approximation by counting all pauses of at least 400ms between two successive sound segments (or IPUs) of the same speaker.[8] Using this rough estimate, Figure 5 shows that, indeed, children are offered slightly more opportunities from the caregiver than the other way around. However, the number of BC opportunities is higher in the adult-adult conversations. Thus, the number of BC opportunities alone does not explain why children still produce BC at a similar rate as adults in the family dyads.

Generic vs. Specific BC We found that children tend to produce fewer generic BC compared to specific BC. However, this result does not necessarily mean children find generic BC harder to learn. Indeed, the rate of generic BC was very low for *both* children and adults in family dyads.

Findings from the non-family control condition suggest a better interpretation for this result. In this condition, adults provided a much higher rate of generic BC. We speculate that the participants used more generic BC (e.g., smiles) to establish social rapport with a stranger. In family dyads (including child-caregiver dyads), however, social rapport is already established, requiring less explicit accommodating signals (Tickle-Degnen & Rosenthal, 1990; Cassell et al., 2007).

Limitations

We used video calls as a way to collect data in a more naturalistic context than previous research did. However, this method involves introducing a medium that has obvious constraints and the participants may be adapting to — or influenced by — these constraints. Indeed, we found some differences with previous work that has studied BC in direct face-to-face conversations in the same culture/country (i.e.,

[8]Note that this measure approximates BC opportunities only in the context of the speaker's pauses, not opportunities for BCs that overlap with the speaker's speech. Estimating the latter requires investigating finer-grained cues within speech such as intonation and clause boundaries.

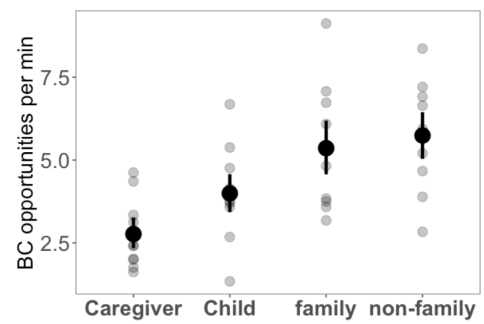

Figure 5: The rate of BC (received) opportunities. Opportunities were defined roughly as pauses of at least 400ms between two successive sound segments (or IPUs) of the speaker.

France) (Boudin et al., 2021; Prévot et al., 2017). For example, our rate of BC production in adults was overall lower. Besides, our number of verbal BC compared to non-verbal BC was also lower (see Table 1).[9] However, our goal in the current work was to *compare* children and adults; the constraints due to introducing an artificial medium of communication applies equally to both populations, thus the comparison remains valid in this specific context.

BC behavior can also be influenced by internet issues such as time lags (Boland et al., 2021), possibly disturbing the appropriate timing and anticipation of BC. That said, our preliminary testing (and, then, the full annotation of the data) have shown that if there were lags, they must have been minimal compared to the time-scale of BC dynamics. The production of BC did not seem to be disrupted (or disrupting) in a noticeable fashion. However, further research is required to precisely quantify the potential (disturbing) effect that video call systems might have on BC as well as other conversational skills.

[9]Part of these differences could also be due to differences in the tasks used to prompt conversations.

Acknowledgments

We would like to thank Fatima Kassim for help with the annotation. We also thank Auriane Boudin, Philippe Blache, Noël Nguyen, Roxane Bertrand, Christine Meunier, and Stephane Rauzy for useful discussion. Finally, we thank all families that have volunteered to participate in data collection. Authors of this work have been supported by funding from ED356, the Institute of Langauge Communication and the Brain (ANR-16-CONV-0002), Archimedes Institute (AMX-19-IET-009), and the MACOMIC project (ANR-21-CE28-0005-01).

References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., ... others (1991). The hcrc map task corpus. *Language and Speech*, *34*(4), 351–366.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941.

Bigi, B. (2012). Sppas: a tool for the phonetic segmentations of speech. In *The eighth international conference on language resources and evaluation* (pp. 1748–1755).

Bigi, B., & Meunier, C. (2018). Automatic segmentation of spontaneous speech. *Revista de Estudos da Linguagem*, *26*(4).

Bodur, K., Nikolaus, M., Kassim, F., Prévot, L., & Fourtassi, A. (2021). Chico: A multimodal corpus for the study of child conversation. In *Companion publication of the 2021 international conference on multimodal interaction* (pp. 158–163).

Boland, J. E., Fonseca, P., Mermelstein, I., & Williamson, M. (2021). Zoom disrupts the rhythm of conversation. *Journal of Experimental Psychology: General*.

Boudin, A., Bertrand, R., Rauzy, S., Ochs, M., & Blache, P. (2021). A multimodal model for predicting conversational feedbacks. In *International conference on text, speech, and dialogue* (pp. 537–549).

Cassell, J., Gill, A., & Tepper, P. (2007, June). Coordination in conversation and rapport. In *Proceedings of the workshop on embodied language processing* (pp. 41–50). Prague, Czech Republic: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/W07-1906`

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Dideriksen, C., Fusaroli, R., Tylén, K., Dingemanse, M., & Christiansen, M. H. (2019). Contextualizing conversational strategies: backchannel, repair and linguistic alignment in spontaneous and task-oriented conversations. In *Cogsci'19* (pp. 261–267).

Dittmann, A. T. (1972). Developmental factors in conversational behavior. *Journal of Communication*, *22*(4), 404–423.

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2021). Caregiver linguistic alignment to autistic and typically developing children.

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

Hess, L. J., & Johnston, J. R. (1988). Acquisition of back channel listener responses to adequate messages. *Discourse Processes*, *11*(3), 319–335.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22–63.

Kjellmer, G. (2009). Where do we backchannel?: On the use of mm, mhm, uh huh and such like. *International Journal of Corpus Linguistics*, *14*(1), 81–112.

Leung, A., Tunkel, A., & Yurovsky, D. (2021). Parents finetune their speech to children's vocabulary knowledge. *Psychological Science*, *32*(7), 975–984.

Matthews, D. (2014). *Pragmatic development in first language acquisition* (Vol. 10). John Benjamins Publishing Company.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Misiek, T., Favre, B., & Fourtassi, A. (2020). Development of multi-level linguistic alignment in child-adult conversations. In *Proceedings of the workshop on cognitive modeling and computational linguistics* (pp. 54–58).

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

Murphy, S. M., Faulkner, D. M., & Farley, L. R. (2014). The behaviour of young children with social communication disorders during dyadic interaction with peers. *Journal of Abnormal Child Psychology*, *42*(2), 277–289.

Park, H. W., Gelsomini, M., Lee, J. J., & Breazeal, C. (2017). Telling stories to robots: The effect of backchanneling on a child's storytelling *. In *2017 12th acm/ieee international conference on human-robot interaction (hri* (p. 100-108).

Peterson, C. (1990). The who, when and where of early narratives. *Journal of Child Language*, *17*(2), 433–455.

Prévot, L., Gorisch, J., Bertrand, R., Gorene, E., & Bigi, B. (2017). A sip of cofee: A sample of interesting productions of conversational feedback. In *Proceedings of the 16th annual meeting of the special interest group on discourse and dialogue (sigdial), prague, czech republic, 2-4 september 2015* (pp. 149–153).

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh'and other things that come between sentences. *Analyzing discourse: Text and talk*, *71*, 93.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development*, 1–38.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 549–565.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry*, *1*(4), 285–293.

Yngve, V. H. (1970). On getting a word in edgewise. In *Chicago linguistics society, 6th meeting, 1970* (pp. 567–578).

Zoom Video Communications Inc. (2021). *Zoom.* Retrieved from `https://zoom.us`