University of Cologne

Faculty of Arts and Humanities

Master's thesis

Navigating Common Ground Using Feedback in Conversation- A Phonetic Analysis

submitted for the degree Master of Arts

by

Alicia Janz

born in Bielefeld, Germany

ajanz1@smail.uni-koeln.de

M.A. Linguistik (specialization: phonetics)

Supervisor: Prof. Dr. Martine Grice

Cologne, August 8th 2022

Table of Contents:

1. Introduction

This thesis deals with backchannel feedback in different conversational contexts. Here, *backchannels* (BCs) are defined following i.a. Ward & Tsukahara (2000) as short vocalizations uttered by the listening interlocutor in a conversation to give positive feedback to the speaking interlocutor, without *claiming the floor.* That is, without taking the role of the primary speaker. In some papers, backchannels have also been described as phatic signals, like head nods or eye blinks (Bangerter & Clark, 2003). I will exclusively focus on German BCs in the vocal modality here, such as *ja* ('yes'), *genau* ('right'), or *mmhm*.

Since backchannels were described as "things that come between sentences" by Schegloff (1982), many researchers have been investigating these feedback signals in human conversation. Today, their importance in the organisation of turn-timing and in guiding conversational dynamics is widely recognized and their role as an essential part of spoken language is mostly undisputed (Gardner, 2001; Savino, 2011; Trouvain, 2014; Truong, Poppe, De Kok, Heylen; 2011; Wehrle & Grice, 2019). In recent years the interest in researching the functions of backchannels in spoken interaction has grown further. Several studies have moved away from investigating BC's purely in task-oriented speech, produced in laboratory settings (Bangerter & Clark, 2003; Dideriksen, Christiansen, Dingemanse, Tylén, Fusaroli, 2021; Fusaroli, Tylén, Garly, Steensig, Christiansen, 2017) in favour of research on feedback-signals in spontaneous interactions. This thesis directly compares feedback signals in these two contexts with the aims of gaining a better understanding of backchannels in real-life conversations.

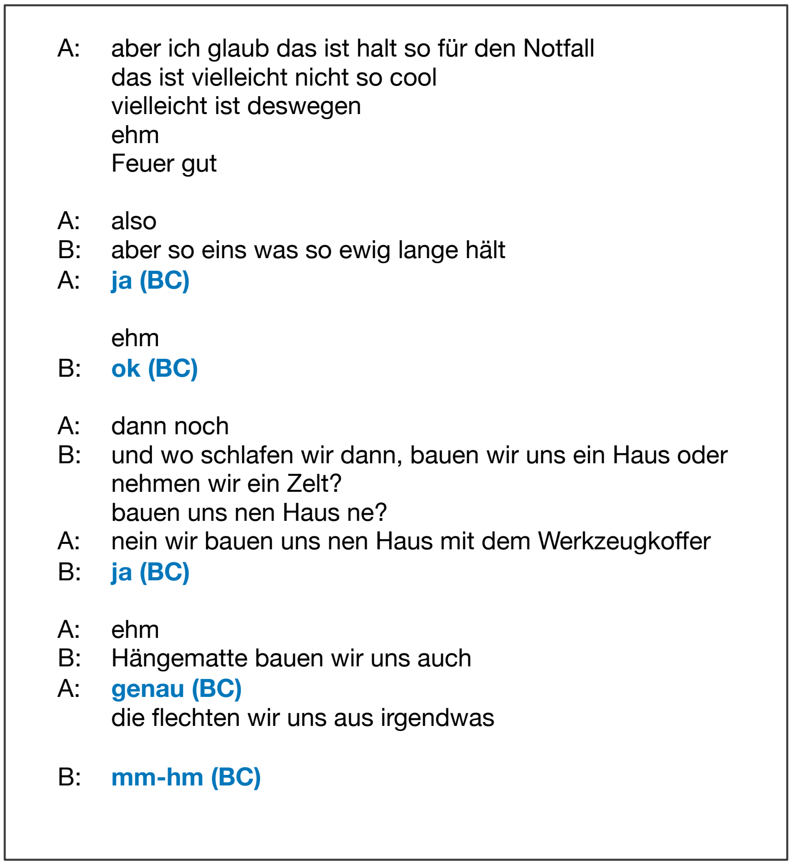
In the following chapters, I first describe backchannels and their general functions (Chapter 2.1.1) in conversation before further elaborating on the impact of lexical choice (Chapter 2.1.2) and intonation (Chapter 2.1.3) in the use of backchannel feedback. The main focus of this thesis is a production experiment investigating potential differences in backchannel use in Task-Oriented as compared to Spontaneous conversation. Chapters three and four are dedicated to first introducing the methods and analyses and second, presenting results on *backchannel rate*, *intonation*, and choice of backchannel token (lexical choice) in the two different conversational contexts. In chapter five the results will be contextualised and interpreted based on the current state of research on the topic. After discussing limitations and future directions, this work is completed by a summary and conclusion in chapter six.

2. Theoretical Background Backchannels

This chapter aims to review and outline existing literature on the topic of vocal feedback signals, specifically backchannels. First, the term backchannel is defined, and a brief outline of the general functions is given (chapter 2.1), followed by the impact of *intonation* (chapter 2.2) and *lexical choice* (chapter 2.3) and their functions in adapting backchannels to varying conversational contexts. Chapter 2.4 is dedicated to presenting the research question, as well as the motivation for the study and predictions about the dataset. Before moving to the function of intonation and lexical choice of feedback-tokens, an overview of backchannels and their functions in general is given in the following section.

2.1. General Functions

How interlocutors coordinate social interaction, especially shared knowledge and behaviour in conversation has always been a topic of great interest in cognitive science and language research (Clark & Schaefer, 1989; Dale, Fusaroli, Duran, & Richardson, 2013). It has often been suggested that backchannels (Yngve, 1979) play an important part in said coordination because they are a key element in the construction and maintenance of *common ground*, some even claim them to be the main mechanism of shared knowledge (Barr, 2004; Clark, 1991; Clark & Carlson, 1981) .The term common ground is used to describe what interlocutors believe to be common knowledge, mutual beliefs and assumptions between speakers in a conversation, but also the mutual awareness about the knowledge they share (Clark, 1991; Fusaroli et al., 2017; Keysar, Barr, Balin, Paek, 1998). In other words, common ground can be described as a kind of meta-knowledge, which is the *knowledge about knowledge* and both interlocutor's awareness of it. In the case of backchannels, the construction of shared knowledge is achieved by the listener giving positive feedback to the interlocutor who is speaking at the time, without interrupting their turn. This concept of backchannels happening *in the background* is based on the idea that there are two channels in a conversation. The main channel, which is occupied by the interlocutor speaking at the time, and the background channel, which can be used to produce feedback (Heldner, Edlund, Hirschberg, 2010; Castello & Gesuato, 2019). Usually, the background channel does not interfere with the main channel. This way, mutual knowledge can be continuously and effectively updated during the conversation without explicit mentioning or interruption (Bangerter & Clark, 2003; Schegloff, 1982). Example 1 illustrates how backchannels can be used in spontaneous conversation to create common ground and communicate to the interlocutor that both communication partners are on the same page (Orestrom, 1983).



A: aber ich glaub das ist halt so für den Notfall
das ist vielleicht nicht so cool
vielleicht ist deswegen
ehm
Feuer gut

A: also
B: aber so eins was so ewig lange hält
A: **ja (BC)**

ehm
B: **ok (BC)**

A: dann noch
B: und wo schlafen wir dann, bauen wir uns ein Haus oder
nehmen wir ein Zelt?
bauen uns nen Haus ne?
A: nein wir bauen uns nen Haus mit dem Werkzeugkoffer
B: **ja (BC)**

A: ehm
B: Hängematte bauen wir uns auch
A: **genau (BC)**
die flechten wir uns aus irgendwas

B: **mm-hm (BC)**

Example 1: *Transcript excerpt of a semi-spontaneous, Task-based conversation between two interlocutors. Backchannels are highlighted in blue.*

This example was drawn from my own data, which is presented more thoroughly in chapters three and four.

According to Clark & Carlson (1981), mutual knowledge plays an important role in any act of comprehension. When a speaker uses e.g. conventional expressions or definite reference, they presume that the expression's metaphorical meaning, or the person they refer to, are known to the listener and are therefore information which is present in common ground. Referents in utterance production are thus chosen according to what the speaker of said utterance believes is common ground. Regarding the comprehension of the same utterance by the listener, it is hypothesized that common ground is crucial in the *search for referents,* a term referring to the act of the listener *searching* their memory for a known referent named as the one which the speaker referred to. Without correctly identifying referents in utterances, gaining an understanding of the same utterance would be impossible.

There is disagreement between theoretical approaches whether the search for potential referents by the listener is restricted by assumptions about mutual knowledge or not (see e.g. Clark & Carlson, 1981; Horton & Keysar, 1996), Therefore, common ground information would either have to be active in a listener's memory during the entirety of the interaction between speakers in order for them to understand any produced utterance, or not at all involved in the search for referents, and as a consequence, not active during conversation (Clark & Marshall, 1981). A slightly newer approach is the one proposed by Keysar et al. (1998), who found that listeners might not be restricted by mutual knowledge in their search for referents. Keysar and colleagues (1998) propose a 'perspective adjustment model' (illustrated by Figure 1) in which the search for referents is not restricted to mutual knowledge, but mutual knowledge is used to adjust the chosen referent in case of common ground violations. Figure 1 illustrates the search for a mentioned referent and subsequent *perspective adjustment* in a case of common ground violation.

Disregarding how common ground is used in comprehension, there is strong evidence for the fact that interlocutors in a conversation establish and update their mutual knowledge using backchannel feedback (Bertrand, Ferré, Blanche, Espesser, Rauzy, 2007; Cutrone, 2014; Fusaroli et al., 2017; Wehrle, 2021). However, the purpose of the conversation has an impact on how this is achieved.

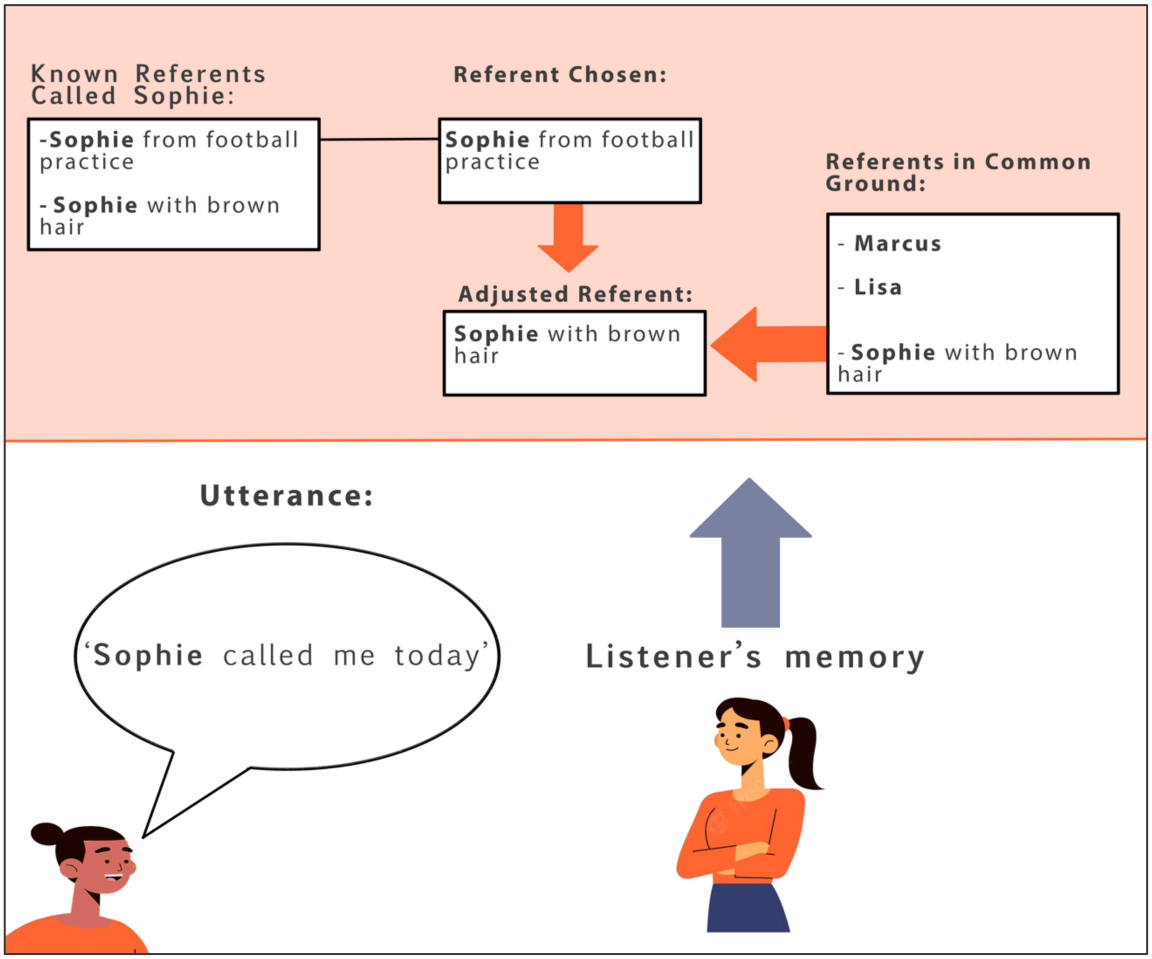


Figure 1: Example 'search for referents' in a listener's memory when hearing the utterance 'Sophie called me today'.

Task-oriented conversations such as Maptask conversations (Anderson, Bader, Bard, Doherty, Garrod, Weinert, 1991) usually contain a large number of feedback signals. In this task, two participants are provided with a map and collaborate to transfer a given route from one participant's map to the other without any visual contact, i.e. using only spoken language. As all task-based conversations, Maptask dialogues serve a clear purpose. That is to reach the goal of the task and consequently, with its completion the conversation ends. Spontaneous, casual conversation on the other hand, often lacks a clear goal and its purpose might be more relevant to forming social bonds than simply exchanging information. In theory it could therefore be continued indefinitely (Gilmartin, Cowan, Vogel, Campbell, 2018). In contrast to task-based interactions in which speakers have clear roles as, e.g. instruction giver or instruction follower in a Maptask, participants in informal conversation have ‘equal speaker rights and can contribute at any time’ (Gilmartin et al., 2018: 298).

Fusaroli et al. (2017) compared the use of *conversational devices* in Maptask and free conversations to gain insights into the coordination of knowledge, behaviour, and social interaction in conversation. In their study they investigated the use of backchannels, interactive alignment, and conversational repair. They found higher levels of repair and reduced levels of syntactic and interactive alignment in Task-Oriented (Maptask) conversation compared to free conversation. More importantly, they observed a *lower* number of backchannels in Task-Oriented dialogue as compared to free conversation.

In a more recent study, Dideriksen et al. (2021) investigated conversational devices (such as backchannels, repair and linguistic entrainment) in different conversational contexts and it was shown that these devices were adjusted to the context. Their results showed that the increased need for precision in Task-Oriented as compared to Spontaneous conversation led to an *increase* of conversational devices such as backchannels in Task-Oriented contexts. These findings are in direct opposition to Fusaroli et al’s (2017) findings. Based on the results of lower backchannel rates in Spontaneous conversation (Dideriksen et al., 2021), one could speculate that in this conversational context, common ground is not updated as regularly through the use of backchannel feedback. Fusaroli and colleagues (2017) agree on the general assumption that different conversational contexts ‘are likely to afford different degrees of explication […] of common ground’ (Fusaroli et al., 2017: 2056). What role BCs play in creating these degrees of explication, remains unclear.

In Example 1, which was already referred to earlier, participants *A* and *B* were asked to discuss items they would take with them on a desert island if they were to go into exile. All backchannels are highlighted in blue. As illustrated in this example, none of the backchannels interrupts the turn of the speaking interlocutor. Backchannels are rather used to support the ongoing turn of the speaking interlocutor. Still, the speaker’s use of backchannels plays an important part in organising who is speaking when and how a turn transition takes place.

Research by Pammi & Schröder (2009) has shown that listener intentions such as the intention to take the floor, are often communicated using backchannel vocalizations. In 1997, Carletta et al. developed a coding scheme for Maptask conversations in which all utterances were organised into conversational moves. According to the coding scheme, backchannels were often categorized as *acknowledge* moves or acknowledge tokens, 'a verbal response that minimally shows that the speaker has heard the move to which it responds, and often demonstrates that the move was understood and accepted' (Carletta et al 1997: 19).

Scholars do not always agree on how to describe and categorise BCs according to their tun-taking functions and the terminology is often used inconsistently. In a concept in which BCs are categorised according to their functions as turn-holding or turn-yielding signal, backchannels are described as reflecting either *passive recipiency* or *incipient speakership*. A token reflecting passive recipiency (Cutrone, 2014) is thereafter acknowledging that the other speaker still has the turn and will continue speaking (oftentimes also called *continuers*). Signals reflecting incipient speakership indicate the intention by the current listener to take the floor, signalling 'preparedness to shift from recipient to speakership' (Jefferson 1983: 4). But listeners are not limited to using backchannels as acknowledgement tokens in turn-taking, they can also be used as tools managing and maintaining social relations between speakers (Dunbar, Mariott & Duncan, 1997). Research has shown that BCs are used to transmit affective states, for example that a speaker is excited, bored, confused, or surprised (Pammi & Schröder, 2009).

Research on backchannel-placement also suggests that backchannel-feedback is highly time-sensitive and can vary vastly in meaning when it is a few milliseconds earlier or later. Too early might signal impatience, too late might possibly imply doubts or a lack of understanding by the listener (Li, 2006). Additionally, a relationship between a feedback tokens exact placement in an utterance and its meaning has been proposed. Duncan (1974), e.g. suggested that backchannels occurring right after the first syllable indicate that the listener is not quite following.

Generally it has been claimed that listeners are not consciously aware of their own use of backchannels, therefore, interlocutors are not fully in control of their backchannelling behaviour (Castello & Gesuato, 2019). Consequently, even though backchannel feedback seems to follow systematic rules, this set of rules is not consciously learned by the speakers of a language (Wehrle, 2021). Still, or even because of their intuitive use, listeners are highly sensitive to the exact realisation of backchannels (Cutrone, 2014; Wehrle, Röttger & Grice, 2018) and deviations from the typical forms are often implicitly judged negatively (Clark & Krych, 2004; Li, 2006). Previous research has also shown that the type of backchannel, and their lexical and intonational realisation can have a profound influence on the communicative success and mutual understanding, as well as the way subjective judgements are perceived (Wehrle, 2021). Therefore, the next two chapters deal with the specific functions of lexical choice, as well as the role of intonation in backchannel feedback in spoken conversation.

2.2. Functions of Lexical Choice

In his paper from 1982, Schegloff describes discourse or conversation as an *interactional achievement*. As a conversation analyst, he understands backchannels primarily as a tool facilitating smooth turn-taking and the production of multi-unit turns by one speaker. In his work, he (like many others e.g. Gibbon et al., 2007;Bertrand et al., 2007; Savino, 2011; Trouvain, 2014) mostly discusses the non-lexical tokens *uh-huh* and the lexical token *yeah* which are the two tokens predominantly used in task-based interaction. However, Schegloff (1982) just briefly touches upon the fact that there are many more possible lexical tokens being used as backchannels. In studies on Task-based conversations it was shown that native speakers of German, for example, not only use the two most dominant backchannels *mmhm* and *ja* (‘yes/yeah’) (Liesenfeld & Dingemanse, 2022), but additionally frequently use the tokens *genau* (‘exactly’) and *okay* (Gibbon et al., 2007; Janz et al., 2022; Sbranna et al., 2022; Wehrle et al., 2018; Wehrle, 2021; Wehrle & Grice, 2019). Token choices have been demonstrated to differ depending on the backchannel’s specific function as turn-holding (incipient speakership) or turn-yielding (passive recipiency) signal. While *ja* and *mmhm* are the preferred tokens for signalling passive recipiency, *genau* and *okay* are used with higher proportions in incipient speakership (Sbranna et al., 2022).

However, while the lexical choice of backchannel token might have pragmatic, as well as turn-management functions, the term *interactional achievement* as proposed by Schegloff (1982), might also include social components of interaction. Besides descriptive information, backchannel tokens in their function as *interactional tools* (Liesenfeld & Dingemanse, 2022) can also serve to express the utterers' stance towards contents that are under discussion e.g. surprise (via exclamations), or reservations regarding whether content should be accepted (via tokens otherwise used in hesitations like 'uh') (Heinz, 2003), as well as communicate social information about interlocutors and their relationship. For instance, they may convey to the speaker that the listener is open to the establishment and maintenance of a social relation. Investigations of the German feedback-token *ja*, for example, have shown it to communicate a variety of emotional states (Gibbon, Stocksmeier, Kopp, 2007). It has also been established that interpersonal synergies and social relations influence lexical choices in conversation (Fusaroli et al., 2014; Krauss, Garlock, Bricker, Mcmahon, 1977). Positive attitudes towards the interlocutor can be signalled by mimicking the interlocutor, for example by making similar lexical choices, i.e. using repetitions of what was said before as backchannels (Dideriksen et al., 2021; Gonzales et al., 2010). Negative attitudes, on the other hand, can cause interlocutors to diverge from each other**.**

Some studies have suggested a relationship between linguistic entrainment (i.a. lexical choice), and a better task-performance in Task-Oriented conversations (Himberg, Hirvenkari, Mandel, Hari, 2015). According to Dideriksen et al. (2021), this kind of lexical entrainment is oftentimes modulated to fit contextual demands. In other words: It is likely that the degree to which interlocutors lexically entrain is affected by the social relation and attitudes of the interlocutor towards each other but also by the setting and purpose of the conversation.

As already mentioned, most existing studies investigated only the most frequent backchannelling tokens. Considering that more context-specific tokens are unlikely to be investigated in a quantitative manner because they are not easily reproducible, this approach is more than understandable. Some studies, such as the one by Wehrle (2021) have, however at least reported tokens that deviate from the *standard* categories (such as *ja, mmhm, genau, okay*) in their lexical form as an *other* category. Studies concerned with the affective functions of feedback signals oftentimes do not report the lexical tokens themselves, but their functions as a category in a quantifiable manner (see e.g.: Nakamura et al., 2021; Bangerter & Clark, 2003; Allwood et al., 1992).

While the choice of backchannel type was shown to play an important role in managing conversation on a social, as well as on a functional level by being involved in turn-management and establishing and maintaining common ground, their meaning and function cannot be construed without considering the intonation they carry and the context in which they are produced. Therefore, the following chapter is dedicated to the function of intonation in backchannel feedback.

2.3. Functions of Intonation

As described above, backchannels have been shown to serve a multitude of functions in conversation. The specific functions, however, might differ depending on the context, and consequently, the purpose and requirements of the conversation (Fusaroli et al., 2017). Because vocalic tokens used as backchannels are often short and some of them contain very little or no lexical content, intonation is an important component to determine a token's exact meaning (Bolinger, 1989). The monosyllabic *ja* ('yes/yeah') or non-lexical *mmhm* for example contain very little or no lexical information. Accordingly, the intonation contour carries most, or all of these tokens' meaning. Several studies to date have investigated the prosodic characteristics of backchannel feedback (Savino, 2010, 2011; Sbranna et al., 2022; Wehrle, 2021). Generally, backchannel prosody has been described to, in very broad terms, somehow differ from the prosodic characteristics of similar short vocalizations (Heldner, Edlund & Hirschberg, 2010). By definition, backchannels usually signal positive feedback as their meanings most often imply understanding and agreement (Savino, 2010; Dideriksen et al., 2021; Sbranna et al., 2022). Previous work on Germanic and Romance languages has shown that positive feedback tokens in general, and those tokens functioning as continuation signals (or *continuers*) in particular, predominantly carry rising intonation contours (Beňuš, Gravano, & Hirschberg, 2007; Caspers, Huang, Yuang, Tang, 2000; Savino, 2010; Wehrle & Grice, 2019 ; Sbranna et al., 2022). Sbranna and colleagues (2022) investigated backchannel intonation in German (and Italian) with regards to their intonation and function as signalling passive recipiency or incipient speakership and demonstrated a relationship between the lexical form of a token, its intonation contour, and its function.

To date, backchannel intonation has been explored from various different angles, are e.g. in intercultural communication or atypical speech, to investigate functional and affective purposes, or to develop backchannel prediction models in the benefit of human-machine interaction. Several studies demonstrated that backchannel conventions are culture- or rather language-specific (White, 1986; Ward & Tsukahara, 2000; Wehrle & Grice, 2019). Therefore, transferring one's native language feedback conventions into a second language (L2), might lead to confusion and misunderstandings. Wehre & Grice (2019) for example, investigated backchannel feedback in Maptask conversations of Vietnamese learners of German. They found the language learners to use high proportions of flat or falling intonation contours on *mmhm* tokens. This behaviour would be appropriate and polite in Vietnamese, whereas the native German speakers produced *mmhm* predominantly with a rising intonation contour which is the appropriate behaviour in German. Wehrle et al. (2018) carried out a perception experiment to test how sensitive native German listeners are to the 'correct' or 'incorrect' prosodic realisation of feedback-signals. Their results show that listeners who produce *inappropriate* backchannels are rated more negatively in terms of character attributions and attentiveness to the conversation as compared to those who produced *appropriate* BCs. Additional studies on backchannelling behaviour of L2 learners and autistic subjects have shown that this is also the case when e.g. a very high backchannel frequency is used in contexts which do not require such a high frequency (Sbranna et al, 2022; Lebra, 1976).

Other studies pursued an approach which attempted to develop a model to reliably predict backchannel responses by looking for backchannel inviting cues in the preceding and following utterances by the interlocutor in a speaking role (Ward & Tsukahara, 2000; Kawahara, Yamaguchi, Inoue, Takanashi, Ward, 2016; Cathcart et al., 2003). Based on a corpus of English and Japanese speech production data Ward and Tsukahara (2000) suggested that prosodic features preceding the backchannel cue its response and prosodic realisation. The most consistent prediction factor were low pitch regions preceding a backchannel. However, they were only able to correctly predict backchannels with roughly 20% accuracy for English and around 35% accuracy in Japanese, suggesting that low pitch regions in Japanese speakers are more reliable as compared to English.

In a newer study investigating BC tokens themselves in comparison to preceding and following utterances, backchannels were found to usually be higher in pitch and more likely to bear a rising pitch accent than other categories of short vocalizations (Heldner, Edlund and Hirschberg, 2010). They were also found to be more similar in pitch to the directly preceding utterance by the other speaker, as compared to the directly following utterance by the same speaker. However, backchannels were usually higher in pitch as compared to utterances by the same speaker, as were the utterances preceding the backchannel which is on contrast to what Ward and Tsukahara (2000) found. Hence, while Ward and Tsukahara found low pitch regions to be backchannel-inviting, Heldner et al. (2010) found high pitch regions to be backchannel-inviting. However, both these studies agreed on the fact that the utterance directly preceding a backchannel is somehow important for its (prosodic) realisation.

Studies on the functional and affective purposes of backchannels and their intonation, tend to focus on the tokens themselves rather than the prosodic features of surrounding utterances. Savino (2010), for example investigated turn-taking and conversation-management functions of backchannel intonation. She showed that participants conveyed their intention to take the floor by using a higher proportion of falling intonation contours while the intention not to take the floor was more often signalled using a rising, as opposed to a flat intonation. In the same study, it was found that Italian subjects use backchannels in Maptask conversations not only to give positive feedback but in some cases also to signal disagreement or uncertainty by using feedback tokens with a falling intonation contour.

The prosodic make-up of backchannel utterances communicates important extra-linguistic information about e.g. mental states of a listener and serve other affective purposes (Gibbon, Stocksmeier & Kopp, 2007; Scott & Sauter, 2006). Gibbon et al. (2007) investigated the connection between mental states and interjections by looking at the prosodic realisation of German *ja* backchannels. Allwood, Nivre and Ahlsen (1992) describe four essential communicative functions of backchannels which are, *contact* (willingness to continue the present interaction)*, perception* (willingness and capability of the listener to receive the message)*, understanding* (willingness and capability to understand the message)*,* and *attitudal reaction* (willingness and capability of the listener to react and respond to the message including approval or disapproval). To my knowledge, none of these studies investigating affective functions of backchannels have investigated differences between specific lexical tokens and their relation to signalling e.g. agreement or excitement.

Studies on the perception of prosodic realisations of BC tokens have shown that listeners are extremely sensitive to a backchannel's exact production. For the German backchannel token *ja,* for example, it was shown that listeners perceive the token as signalling different intentions depending on the exact intonation contour and duration. While a simple rising contour with a duration of 300 milliseconds, for example was perceived as neutral and straightforward, a rise-fall movement was perceived as very strong agreement. As for the perception of other contours, participants reported i.a. annoyance, hesitation, boredom, or anger as possible impressions (Gibbon et al., 2007). Interestingly, the phonetic variation among backchannelling tokens with falling and flat intonation contours (in terms of e.g. slope, amplitude, voice quality) was shown to be much wider than the variation in tokens with a rising intonation contour (Savino, 2010). Considering that listeners have been found to be more sensitive to pitch rises as compared to falls (Hsu, Evans, Lee, 2015), one could draw the conclusion that rising pitch contours are more closely mapped to specific intentions or affective states as compared to pitch falls. Therefore, variations from the intended form in pitch rises would have to be kept as small as possible to avoid misinterpretations by the other interlocutor.

2.4. Research question and predictions

In the previous two chapters the impact of backchannels as predominantly positive feedback signals uttered by the listener to manage conversation and common ground were summarised based on previous literature. Findings on the specific roles played by the two components intonation and lexical choice as found in task-oriented conversations were also described.

Most knowledge we have about feedback production to date in terms of intonation, frequency and lexical choice is not based on data from spontaneous, natural conversations. It rather stems either from task-oriented game-like conversations, such as Maptask conversations (e.g. Heldner et al. 2010; Ha et al., 2016; Wehrle, 2021; Savino, 2011 ; Koiso, Hanae, Horiuchi, Tljtiyad, 1998), often without eye-contact between the two subjects, or corpora of telephone calls (e.g Cathcart et al., 2003; Heinz, 2003). Both these types of communication are rather specific in purpose. In game-like contexts, for example, any interaction has a clear goal, assigned by the experimenter which is missing in spontaneous, casual interactions.

To my knowledge, there are only two studies investigating backchannel in relation to the context of the conversation, both of which are of recent date. In their study on conversational devices, Dideriksen et al. (2021) investigated backchannels in relation to repair, and linguistic entrainment in different conversational settings in native speakers of Danish. Among the contexts they investigated were Maptask conversations as well as spontaneous conversational settings. Dideriksen and colleagues (2021) found task contexts to be more informationally dense as compared to spontaneous conversation contexts, meaning that more information is introduced and transmitted during the dialogue. Conversational devices, such as backchannels, which enable higher referential precision in the construction and maintenance of shared knowledge were found to be more frequent in Task-Oriented contexts, as compared to Spontaneous conversations. Moreover, a rising intonation contour on backchannel tokens was shown to suggest understanding, agreement and attention which is more important in Task-Oriented conversations. Spontaneous conversation, on the other hand, was found to oftentimes be less informationally dense. Its main purpose is rarely to solve a problem with the interlocutor but rather it has social and affective functions (Dideriksen et al., 2021).

However, a slightly older study by Fusaroli et al. (2017) which too, investigated conversational devices in Danish, using i.a. Maptask and Spontaneous conversations, found conflicting results concerning backchannel rate. In this study, backchannels were found to be more frequent in Spontaneous conversations as compared to Maptask dialogue which is in direct opposition to Dideriksen et al.'s (2021) findings.

Evidently, there is a noteworthy gap in the literature when it comes to backchannelling behaviour in task-oriented as compared to spontaneous conversation contexts, not only regarding backchannel rate but also considering backchannel intonation, lexical choice and speakers of different languages. It is an open question whether and how backchannels differ in spontaneous conversation as opposed to Task-based dialogue 1) in their intonation contour as well as 2) in their lexical load or participant's lexical choice and the functional implications these differences might have.

I therefore conducted a production experiment investigating backchannel feedback in native speakers of German in two conversational settings. For reasons of comparability, I chose to use Maptask conversations (Anderson et al., 1991) for the Task-based context and compare them with Spontaneous dialogue of the same participants.

Based on the studies by Dideriksen et al. (2021) and Fusaroli and colleagues (2017), as well as a preceding pilot study regarding backchannel intonation and lexical choice in which we found speakers to use a greater variety of tokens in spontaneous contexts (Janz, Wehrle, Sbranna, 2022), I predicedt:

i) backchannels to be more frequent in Task-Oriented as compared to a Spontaneous conversation context

ii) differences in the intonation contours between the two conversational contexts, specifically more rising backchannels in Task-Oriented conversation as compared to Spontaneous dialogue

iii) differences in the proportions of token types used in the different conversational contexts

3. Method

The present study was conducted within project A02 of the collaborative research centre SFB 1252[1] at the University of Cologne, which is funded by the German Research Foundation (DFG). This chapter will present the methods, including participants (chapter 3.1), recording conditions (chapter 3.2), and experimental procedure (chapter 3.3). Finally, data processing and measurements (chapter 3.4) including the semitone (chapter 3.4.1) and bayesian analyses (chapter 3.4.2) will be reported.

3.1. Participants

In the following sections I will report on recorded data from fourteen participants who took part in the experiment. All participants gave informed consent before the start of the experiment and were paid for their participation after the second of two recordings.

All subjects were native speakers of German and all except one subject grew up monolingually in Germany. One speaker (Speaker 05) reported to have grown up bilingually with German as their dominant language and Italian as their less dominant language. None of the participants reported any diagnosed voice or speaking disorder or hearing impairment. In case of visual impairment, participants wore appropriate glasses or contact lenses. None of the subjects were diagnosed with a reading or writing disorder. Also, none of the participants had in-depth knowledge in the field of phonetics, phonology, or speech analysis. They were all students at the university of Cologne and/or the German Sports University Cologne. The recorded individuals were aged between 23 and 28 at the time of the recording. The mean age was 25.4 years. Nine out of fourteen participants identified as male, the remaining five participants identified as female (none identified as gender-diverse).

[1] Project A02 – Individuum-spezifisches Verhalten in der En- und Dekodierung prosodischer Prominenz

All subjects were recorded in dyads (pairs), resulting in three male-male, three male-female, and one female-female dyad. At the time of the recording, all subjects from the same pair were living in a shared flat. They had been living together for at least six months and at most six years prior to the recording (average: 3.42 years).

Unfortunately, two dyads (dyads 03_04, and 05_06) could only participate in the first out of two recordings and were not available for the second recording due to one participant moving out of the shared flat in the meantime. As a result, Maptask data for these two dyads was not recorded and cannot be presented.

3.2. Recording Conditions and Experimental Procedure

The experimental procedure was constructed around two different conditions. The aim was to create a Task-Oriented conversation environment as well as an informal, Spontaneous conversation environment to analyse and compare backchannels uttered in these two conversational contexts.

Therefore, two separate recordings were carried out with the same set of speakers with a minimum of two weeks apart (maximally five weeks apart**)**. Recordings were always conducted in the same order. The first recording took place in the participant's homes, preferably in the room in which both individuals felt most comfortable (often in the living room or kitchen). In this setting, portable recording equipment, consisting of a pair of clip-on condenser microphones (AKG C417PP), placed in a range within 15cm of the mouth, and a Focusrite Scarlett 2i2 interface (using Adobe audition, sampling rate 48 kHz and bit depth of 30 bit on a MacBook Pro) was used for auditory recordings. Two GoPro cameras visually captured the experiment from opposite positions in the room.

During the first recording, participants were instructed to sit down in a position and at a distance they were both comfortable with before carrying out two tasks with a five-minute break in-between in which the recording continued, and the subjects were instructed to not leave the room. The dialogues that were recorded during the five-minute break will later be referred to as Spontaneous conversations. The instructions were provided in written form in two separate envelopes for each speaker. One of the two participants received text messages to their mobile phone as signals for the start of the experiment, five-minute break, start of the second task, and end of the experiment. During the whole course of the experiment, the experimenter was not inside the room.

The two task descriptions were provided in German and read as follows:

> Task 1:
>
> 'Stellt euch vor, ihr würdet morgen ins Exil auf eine einsame Insel geschickt werden und dürftet nur fünf Gegenstände dorthin mitnehmen.
>
> Bitte diskutiert gemeinsam und einigt euch, welche fünf Gegenstände mitgenommen werden sollten. Wenn ihr euch auf die ersten fünf Gegenstände geeinigt habt, diskutiert welche weiteren Gegenstände mitgenommen werden sollten, obwohl ihr nur fünf Gegenstände mitnehmen könnt'
>
> ('Imagine going into exile on a desert island. Discuss with your partner which five items you would take to the island and why. When you are done, discuss which additional items should be taken even though you can only take the five items you discussed before.')
>
> Task 2:
>
> 'Bitte stellt euch gemeinsam ein fünf-gängiges Menü (plus Getränke) für ein Abendessen zusammen. Das Menü sollte aus folgenden Gängen bestehen: Suppe, Vorspeise, Hauptgang inclusive Beilagen, Käseplatte, Dessert.
>
> Bitte diskutiert zusammen verschiedene Varianten mit dem Ziel letztlich ein Menüs zusammenzustellen, welches ausschließlich aus Gerichten und Getränken besteht, die ihr beide NICHT mögt.'
>
> ('Work together to compose a five-course menu including drinks consisting of the following courses: Soup, starter, main course, cheese, desert. Discuss various versions with the aim to compose a menu consisting only of ingredients that you and your partner both do NOT like to eat.')

The second recording took place in a laboratory setting, at the IfL phonetics of the University of Cologne, an environment none of the subjects was familiar with. Acoustic recordings were carried out using an AKG C544 headset microphone and a Tascam US-4x4 interface at a sampling rate of 48 kHz and a bit depth of 16 bit. The microphone was placed at a distance of approximately 7 cm from the subject's mouth at an angle of 45-90 degrees.

Similarly to the first recording, subjects were given two tasks. All participants were instructed through a video recording of a German speaker explaining each task. The task instructions were presented on a shared screen simultaneously for both participants. The subjects were seated at a table facing each other with an opaque barrier in-between them, making the visual channel unavailable for the time of the recording. The first task consisted of a game-like task. A written task description can be found in the appendix. Following the first task, participants were allowed to take a five-minute break in which the recording, again, continued and they were allowed to move in a range constricted by the length of the microphone cable. Most participants chose to either stay seated, or to seek visual contact to the interlocutor for a brief moment and then sit down again. For all participants who conversed during the five-minute break while remaining seated with the barrier between them and their interlocutor, I later decided to use these conversations as a spontaneous control condition. Afterwards, subjects were provided with video instructions for the second task, i.e. a Maptask (Anderson et al., 1991). In the Maptask the two participants were equipped with a map and worked together to transfer a given route from one map to the other.

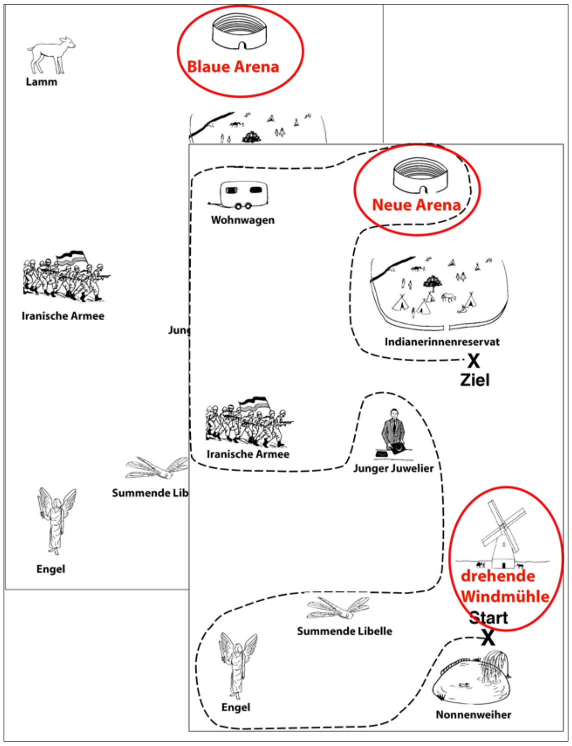


Figure 2*: Example maps for instruction giver on the right and instruction follower on the left. Two example mismatches are highlighted with red circles.*

In the laboratory setting described here, subjects were not allowed any visual contact during the task and therefore had to use only spoken language to solve the task. Figure 2 shows an example set of maps. The instruction giver's map is displayed on the left and the instruction follower's map is shown on the right-hand side. Some of the landmarks differed between the two maps (circled in red), about which the subjects were not aware at the beginning of the task.

All subjects gave their written consent and personal information before the start of the recordings. Altogether, all individual sessions (recording session I and II) took about 45 to 60 minutes for each dyad of which the actual recordings excluding instructions, filling out forms, and levelling in the microphones lasted approximately 30 minutes. Figure 3 depicts an overview of the recording procedures in the two conversation contexts.

During the annotation phase, it was decided to include additional data in the analysis to approximate the influence of visual contact (or lack thereof) on the production of BCs. For this third category of Spontaneous Control conversations speech which was produced during the break in the laboratory setting was used. As only three out of five dyads conversed during the break while having no visual contact, only the data of these three dyads was used for the Spontaneous Control condition.

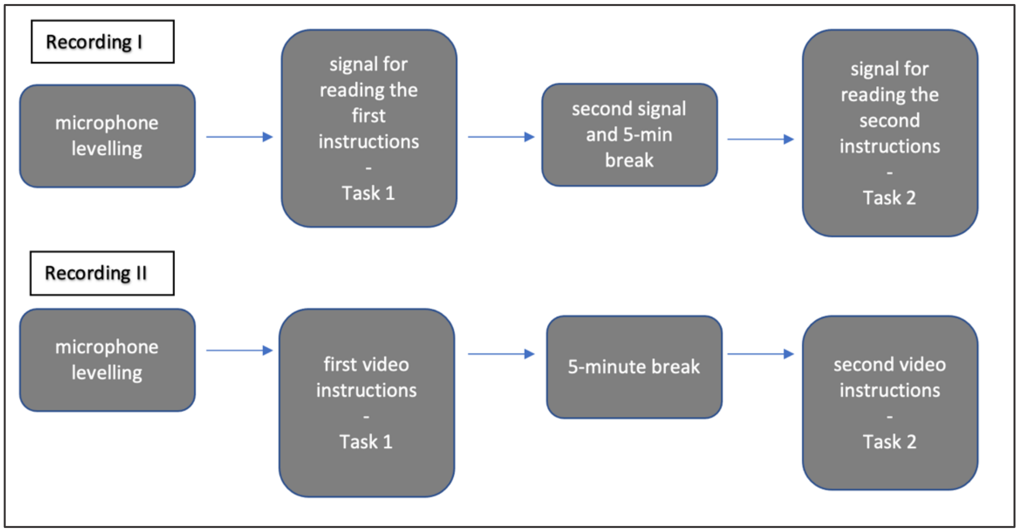

Recording I
microphone levelling
signal for reading the first instructions - Task 1
second signal and 5-min break
signal for reading the second instructions - Task 2
Recording II
microphone levelling
first video instructions - Task 1
5-minute break
second video instructions - Task 2

***Figure 3**: Brief overview of the structure of the recordings I and II. In recording I, tasks and breaks were signalled using text messages to one of the speakers, in recording II, participants were notified via loudspeaker.*

3.3. Annotation Procedure

The acoustic data were processed using *Praat* (Boersma & Weenink, 2021). Two trained annotators (native speakers of German) orthographically transcribed the data. For the backchannel annotations, the Carletta et al. (1996) coding scheme was applied to all speech uttered by a speaker in a listening role. All tokens that met the criteria according to the coding scheme were labelled as backchannels. In the analysis, answers to yes-no or tag questions, as well as repetitions or turn-initiating backchannels were excluded. I also excluded all backchannels consisting of repetitions with more than one word from a previous utterance as they would not have been analysable with the available tools and methods. The resulting set of analysable tokens therefore consisted only of *clear cases* (Ward & Tsukahara, 2000) of acknowledge and positive reply tokens (Savino, 2010), that did not directly initiate a new turn (passive recipiency). However, it should be noted that in the Spontaneous condition, considerably more *unclear cases* were found as compared to Task-Oriented speech.

The backchannel tokens were orthographically annotated (*token*) and divided into five main categories (*type*), i.e. *ja* (‘yes/yeah’), *okay*, *genau* (‘right/exactly’), and the two non-lexical categories transcribed as *hm* (monosyllabic), and *mmhm* (disyllabic). Even though the difference between the latter two tokens is often not clear cut (previous related work subsumed both tokens under the *mmhm* category, (see Wehrle, 2021), I decided to differentiate between the two. During the annotation phase I gained the impression of 1) systematic differences in the proportion of *hm* and *mmhm* tokens in the Spontaneous dataset as compared to the Task-Oriented dataset and 2) structural differences between the intonational realisation of *hm* and *mmhm*. Furthermore, the perceived intonational realisation of the two tokens was perceived to differ as well. All remaining tokens were subsumed under an *other* category with the two subcategories *ah+other,* and *ja+other*. Tokens which were previously reported in studies from Maptask conversations (*ja, okay, genau, mmhm*) will hereafter be referred to as *standard* tokens, all other token-types (*ja+other, ah+other, hm, other*) will be referred to as *non-standard* or *unusual* tokens.

3.4. Data Processing and Measurements

All data was pre-processed and extracted using *Praat*-scripts.(Boersma & Weenink, 2021) (kindly provided by Francesco Cangemi, Simona Sbranna, and Simon Wehrle). All backchannels were extracted as individual sound files. Further data processing, such as calculations of backchannel rates, as well as the descriptive statistics in their entirety, were carried out using *RStudio* (R Core Team, 2019), applying the packages *tidyverse* (Wickham et al., 2019) and *dplyr* (Wickham et al., 2022) for data pre-processing and *ggplot* (Wickham, 2016) for visual presentation.

Applying the procedure described in 3.4.1., I measured the intonation contour (pitch movement in semitones) for each individual token continually, and categorised each value into the category of rising, falling, or level contours. Further, BC rate per minute was calculated individually for each conversation, as well as across contexts. Additionally, choice of backchannel type was analysed, followed by an exemplary investigation of conversation topics.

3.4.1. Semitone Analysis

For the intonation analysis, all F0 values were extracted using *Praat*-scripts. Intonation contours were then manually checked, and hand corrected and smoothed using *mausmooth* (Cangemi, 2015). Thereafter, F0 trajectories were extracted by measuring the absolute pitch at 10% of the utterance and at 90% of the utterance and calculating the difference in semitones between the two points (Ha et al., 2016; Wehrle & Grice, 2019) as illustrated by Figure 3. If due to e.g. an unvoiced segment there was no pitch information available at either one of these points, the extraction point was moved by ten percent. Pitch differences between the two points were measured in semitones. Positive values indicate an F0 rise between the first and second point, negative values indicate an F0 fall between the two points. For the categorical analysis, all values were automatically categorised as rising, level, or falling, using a threshold of one semitone. This threshold was used following i.a. Wehrle (2021).

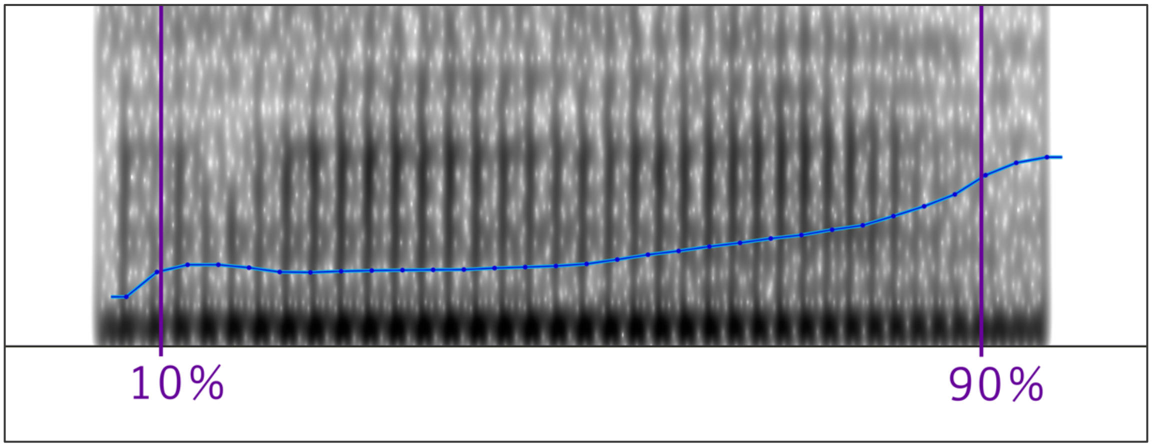


Figure 4: *Example of semitone analysis of rising F0 contour. pitch is measured at 10% and 90% of each token In this case, the contour is rising which would be reflected by a positive value of pitch movement in semitones.*

3.4.2. Bayesian Analysis

To statistically test the descriptive results on differences in the intonational realisation of backchannel tokens between the two main conditions (and the Spontaneous Control condition), Bayesian analysis was used. Two Bayesian mixed effects logistic regression models (Baayen et al., 2008) were generated. For both models, weakly informative priors were used combined with an underlying skew normal distribution. The model structure included speaker as random effect and condition as fixed effect. For model one, the aim was to predict the absolute pitch movement in semitones for each token, to test whether it is greater in the Spontaneous conditions as compared to the Task-Oriented condition. For the first model, four sampling chains ran for 1000 iterations each, including a warm-up period of 500 iterations, resulting in a total of 2000 samples.

The second model aimed at testing whether the distribution of tokens with positive and negative pitch movements differed in the Spontaneous and Task-Oriented conditions. For this model, the same priors, as well as random and fixed effect structure were used combined with a gaussian distribution. Four sampling chains ran for 1500 iterations each, including a warm-up period of 500 iterations, resulting in a total of 4000 samples. All Bayesian analyses were carried out in RStudio (version 4.1.1.; 2021-08-10) (R Core Team, 2019), using the *brms* package (version 2.17.0., Bürkner, 2017).

In the results section, expected values (*ß*) under the posterior distribution and their 95% credible intervals (CI) are reported. I also report the posterior probability that a difference *δ* is greater than zero. The 95% CI represents the range within which an effect is expected to fall with a probability of 95%. Full details on posterior distributions, as well as all other implementations are reported in the appendix (chapter 9.3).

In all analyses, the Task-Oriented conversation context was used as a reference level. This view might seem odd regarding the fact that Spontaneous conversation is more natural than Task-Oriented conversation but considering the amount of literature on backchannels in Maptask and Spontaneous speech it becomes clear that the knowledge about Task-Oriented, or even Maptask conversation, is much greater as compared to the knowledge we have about Spontaneous dialogue.

Given the context of this study and the rather small dataset, the work at hand is necessarily exploratory in nature. Bayesian inference gives outcome based on the data at hand, the chosen models, and the specified prior assumptions. Due to the small dataset and the great variability in tokens, statistical tests are only of limited meaningfulness.

4. Results

Application of the above-mentioned procedures resulted in a small corpus of dyadic Task-Oriented and Spontaneous, German conversations (Task-Oriented: 59.8 minutes, Spontaneous: 47.6 minutes, Spontaneous Control: 17.9) in two different settings (in participants' home and in the laboratory).

In the following chapter, an in-depth descriptive analysis of 568 backchannel tokens in German (Maptask: 329 tokens, Spontaneous: 179, Spontaneous Control: 60), combined with an exemplary statistical analysis applying Bayesian mixed effects logistic regression models (see chapter 3.4.2) will be presented. The aim is to provide a comprehensive understanding of the experimental results emphasizing a transparent and visually rich descriptive analysis.

For a transparent and meaningful report of the experimental results, it is important to emphasise that the surroundings in the two main contexts Spontaneous and Task-Oriented differed. All Task-based conversations were recorded in a laboratory setting without visual contact between the two speakers of a dyad. All Spontaneous conversations were recorded in the speakers' home with visual contact. The recordings made during the task-breaks of the laboratory setting function as an additional control condition (Spontaneous Control). I want to stress that the results presented here, as well as their interpretation in the following chapters is in no way exhaustive.

4.1. Backchannel Rate

For the overall analysis of backchannel rate per minute, all backchannels in all conditions were extracted (as described in the Methods section) counted and divided by the total time of all conversations and respectively by the time of each individual conversation. It must be noted that, while all conversations in the two main contexts (Maptask and Spontaneous) were ongoing conversations without very long stretches of silence, the overall conversational dynamics and the proportions of speech and silence differ between the two contexts, which should be kept in mind when interpreting the following results. In general, the most silence and longest silent intervals were observed in the two Spontaneous conditions. In the Maptask condition subjects uttered proportionally more speech during the conversations as compared to the Spontaneous condition (more information on conversational dynamics and proportions of speech and silence can be found in the Appendix).

Figure 5 shows the general backchannel rate per minute for the whole group of speakers in all three conditions. The purple bar represents backchannel rate for the Maptask condition, BC rate for the Spontaneous condition is depicted by the turquoise bar. The backchannel for the Spontaneous Control condition (Spont_Control) is illustrated in yellow. In the Maptask condition the subjects uttered 5.5 backchannels per minute as opposed to only 3.76 backchannels per minute occurred in the Spontaneous condition (Spont_Control: 3.35 BC/min). Although it should be noted that data for the Spontaneous Control condition was only available from three dyads and should be interpreted with caution, the results for backchannel rate per minute confirm the overall tendency toward a lower BC rate in Spontaneous conversations as compared to Maptask conversations.

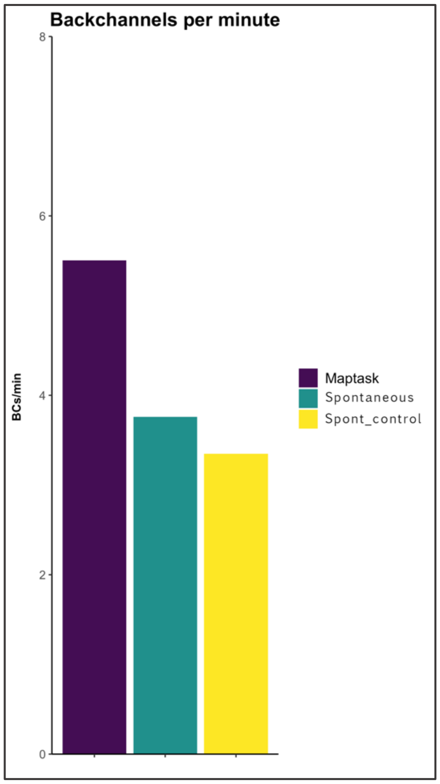


Figure 5: *Backchannels per minute for all speakers. Backchannels per minute on the y-axis, Maptask in purple, Spontaneous in turquoise. Spontaneous Control condition in laboratory setting in yellow.*

Figure 6 (below) displays the backchannel rate per minute on the speaker (top figure) and dyad level (bottom figure). In the top figure, individual speakers are shown as purple squares in the Maptask condition and turquoise squares in the Spontaneous condition. Where available, BC rate per minute for the Spontaneous Control condition is depicted by yellow squares.

In the bottom half of the figure, mean values for all dyads are, again, represented by squares in purple (Maptask), turquoise (Spontaneous), and yellow (Spontaneous Control) squares. As the behaviour of a single speaker in dyadic interaction cannot be fully interpreted without acknowledging the other speakers' influence on an individuals' behaviour, I chose to depict, values for individual speakers as well as mean values per dyad.

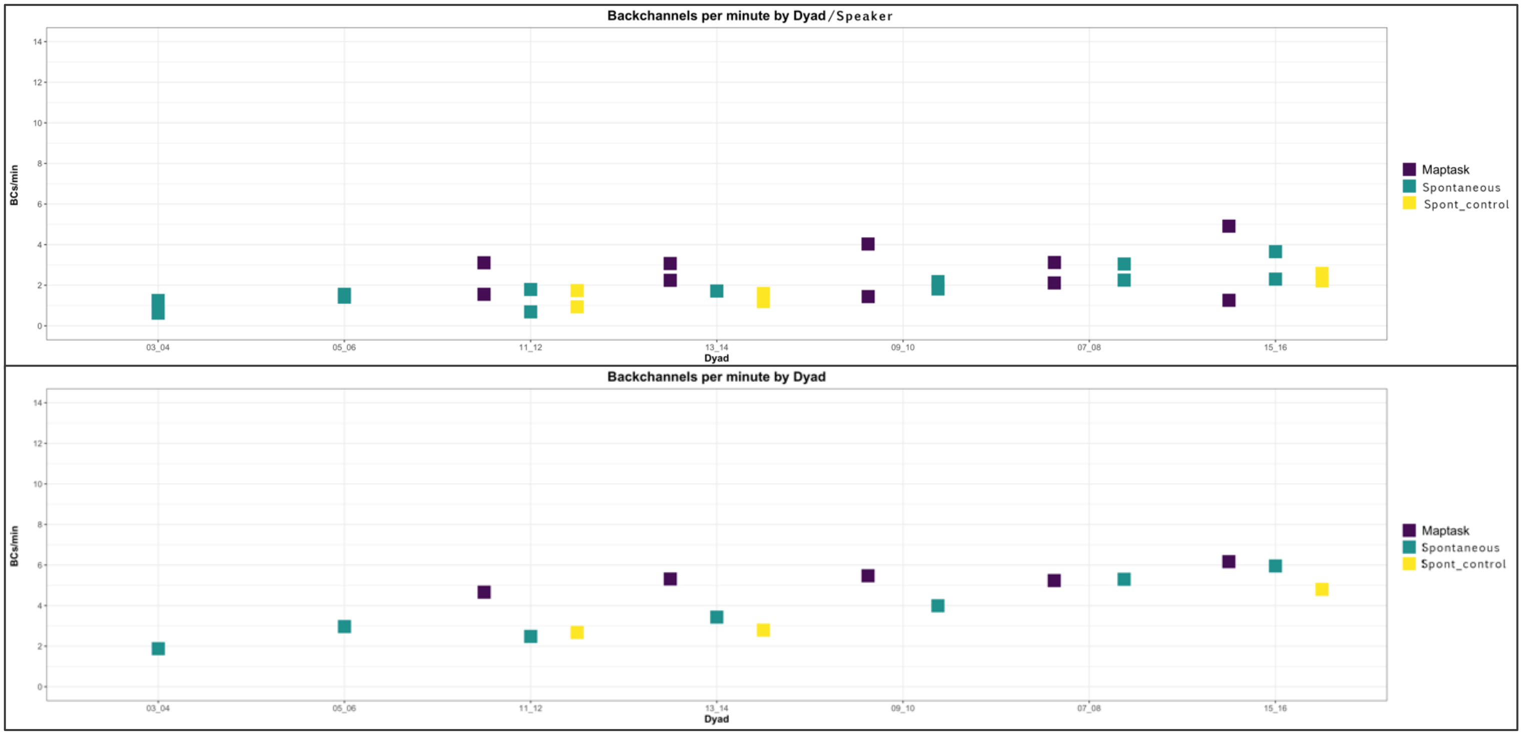


Figure 6: *Backchannels per minute by dyad and speaker with dyads on the x-axis and rate per minute on the y-axis. Note that Maptask values are missing for dyads 03_04 and 05_06. Values for Maptask conversation in purple, Spontaneous in turquoise and Spontaneous Control condition in yellow.*

Analysis on the dyad or speaker level, as shown in Figure 6, confirms the impression that all dyads and about half of all individual speakers used a higher BC rate in the Maptask condition and a lower BC rate in the Spontaneous condition (highest rate per dyad Spontaneous: 5.9BC/min, Maptask: 6.1BC/min, lowest rate per dyad Spontaneous: 1.9BC/min, Maptask: 4.7BC/min).

Additionally, most dyads showed a greater difference in BC rate between the two speakers in the Task-Oriented condition, and a more similar BC rate in the two Spontaneous conditions. This differentiation between individual speakers is lost when investigating only mean values per dyad. Considering mean values only, dyads 11_12, and 13_14 would show the greatest difference in mean BC rate between the Maptask and Spontaneous condition. However, an investigation of individual speaker rates shows that in these two dyads, the two speakers simply behaved more similarly than compared to e.g. dyads 09_10, and 15_16, where BC rates per dyad differed considerably but the two values balanced each other out.

All in all, the data for BC rate per minute suggests different behaviours with generally higher BC rates per minute for the Task-Oriented context and lower BC rates for the Spontaneous contexts. Also, the two speakers of a dyad behaved more similarly in the Spontaneous condition with similar BC rates and showed greater differences in BC rate in the Task-Oriented conversation context. It should be noted that for two dyads (dyads 03_04, and 05_06), only Spontaneous data were available. As already mentioned, in both cases one of the subjects moved out of the shared flat, therefore the dyads did not qualify for the second recording anymore

4.2. Lexical Choice

In this chapter findings from the exploration of lexical choice of backchannel types in Task-Oriented (Maptask) and Spontaneous conversation contexts (Spontaneous and Spontaneous Control) will be reported briefly. The present chapter focuses on reporting descriptive statistics only, to evaluate the amounts of lexical load and differences in lexical choice between the two contexts of Task-Oriented and Spontaneous conversation.

Overall, the proportions of backchannel type by conversation context differed to a certain degree (see Figure 7, and Table 1). In all three conversational contexts, the most frequent backchannel types were the monosyllabic *ja* (Maptask: 32.5%, Spontaneous: 37.7%, Spontaneous Control: 28.6%), and disyllabic non-lexical *mmhm* (Maptask: 37.7%, Spontaneous: 25.4%, Spontaneous Control: 31%). In all other categories, clear differences were observed. While in Task-Oriented speech, tokens from the *okay* category made up 17.5%, in the Spontaneous contexts the amount of these tokens was much lower (Spontaneous: 1.6%, Spontaneous Control: 4.8%). Similarly, the proportion of *genau* tokens in Task-Oriented conversations was much higher (6.4% of all tokens) as compared to the Spontaneous conversation contexts (Spontaneous: 1.6%, Spontaneous Control: 0.0%). Tokens from the categories of *ja+other* , *ah+other,* and *hm* on the other hand, were much more frequent in Spontaneous contexts as compared to the Maptask context (*ja+other*: Maptask: 0%, Spontaneous: 9.8%, Spontaneous Control: 9.5%; *ah+other:*Maptask: 0%, Spontaneous: 3.3%, Spontaneous Control: 7.1%; *hm*: Maptask: 3.1%, Spontaneous: 9.8%, Spontaneous Control: 14.3%).

,

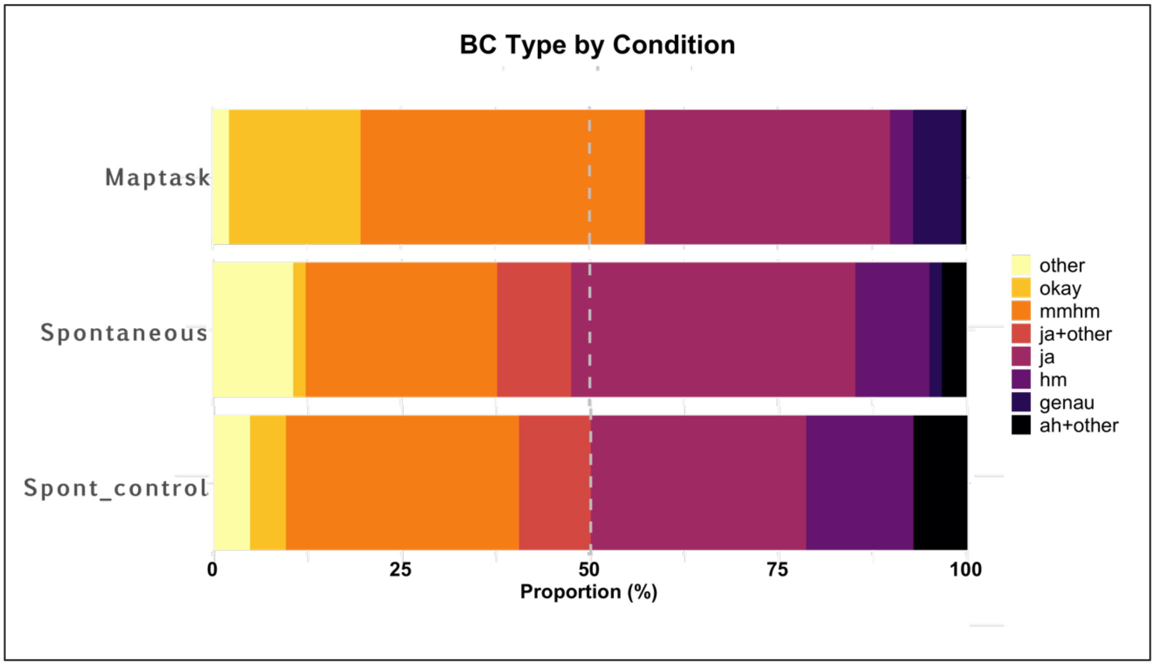


Figure 7: Proportions of backchannel type by condition. Maptask on top, Spontaneous in the middle, and Spontaneous Control at the bottom. Different BC types are indicated by different colours, the grey, dotted line indicates a proportion of 50%.

Analysis on the level of individual speaker (Figure 8) reveals that many speakers used a wider range of standard backchannel types (such as *ja, mmhm, okay, genau*) in the

Maptask condition, but a much higher proportion of *other* tokens in the Spontaneous condition.

Table 1: *Proportions of BC types in the three contexts with Maptask on top, Spontaneous in the middle and Spontaneous Control at the bottom. The two most used BC types per condition are highlighted in grey.*

Condition	Type	Percentage
Maptask	ah+other	0.61
Maptask	genau	6.44
Maptask	hm	3.07
Maptask	ja	32.52
Maptask	mmhm	37.73
Maptask	okay	17.48
Maptask	other	2.15
Spontaneous	ah+other	3.28
Spontaneous	genau	1.64
Spontaneous	hm	9.84
Spontaneous	ja	37.70
Spontaneous	ja+other	9.84
Spontaneous	mmhm	25.41
Spontaneous	okay	1.64
Spontaneous	other	10.66
Spont_Control	ah+other	7.14
Spont_Control	hm	14.29
Spont_Control	ja	28.57
Spont_Control	ja+other	9.52
Spont_Control	mmhm	30.95
Spont_Control	okay	4.76
Spont_Control	other	4.76

A closer look at those lexical backchannel tokens categorized as *other* shows that there were differences in the variety of *other* tokens uttered in Maptask conversations as compared to those uttered in Spontaneous dialogue. *Other* tokens from the Task-Oriented conversation context were exclusively the lexical tokens *ah* and *aha*, while tokens from the Spontaneous conversation contexts were much more variable. In the Spontaneous condition the subjects used tokens like *stimmt* ('true'), *nice,* or *schön* ('nice'), to list only some examples. These tokens differed from the standard lexical tokens not only in their form, but also in their function. While standard tokens might be mostly used to manage the conversation from a functional perspective, the *non-standard* tokens mentioned above presumably mostly signal the listeners attitude towards the content of the interlocutor's utterance (a full list of tokens can be found in the Appendix).

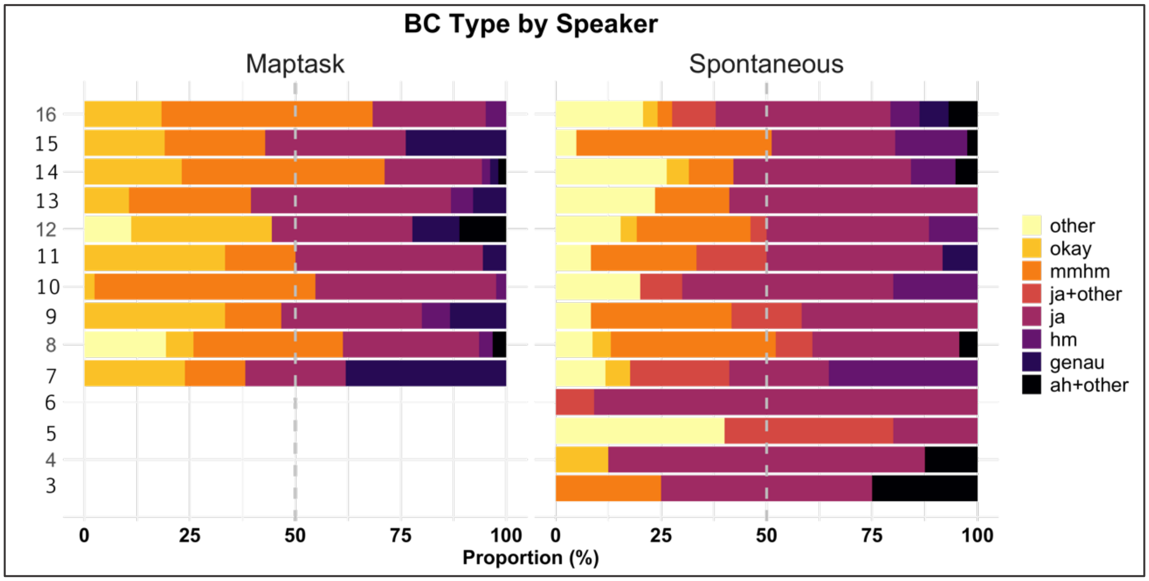


Figure 8: *Proportion of backchannel types per speaker in the two main conditions Maptask (left) and Spontaneous (right)*

These differences in BC function were used to categorise all tokens into the two subcategories affective and functional based on their lexical form. All tokens signalling affective states such as excitement, boredom, astonishment, or agreement were therefore categorised in the affective category, all other tokens remained in the functional category. In the Maptask condition, 99.1 percent of all tokens fell within the category of functional backchannel tokens, while in the Spontaneous condition only 81.1 percent of all tokens fell in this category. The remaining 18.9 percent of BC utterances fell within the category of affective tokens as illustrated by figure 9.

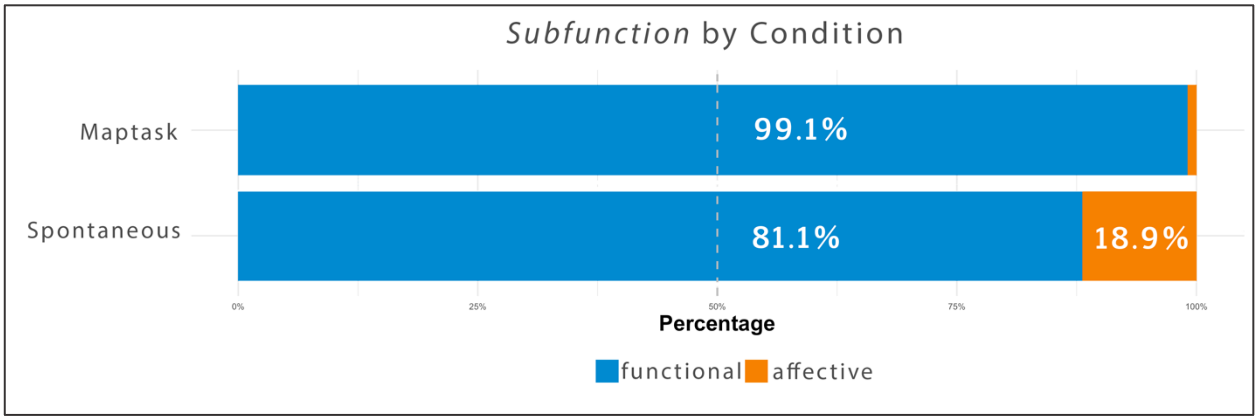


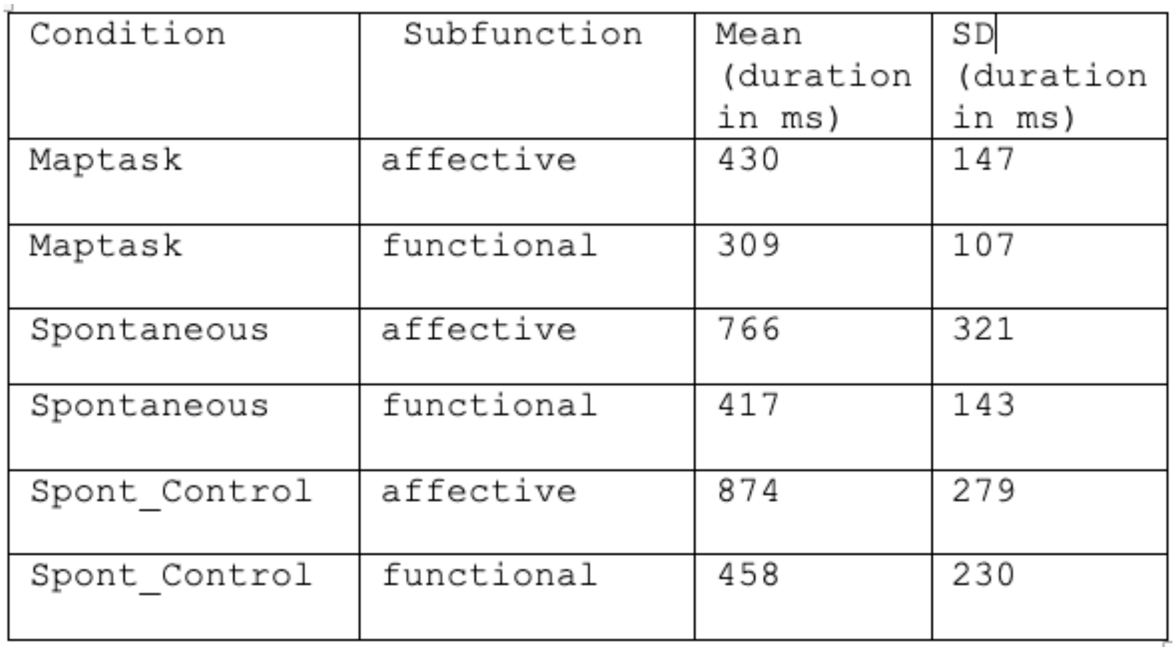
Figure 9: Proportions of subcategory per condition for spontaneous and task-oriented contexts

Backchannel utterances as reported from Task-Oriented conversations usually only lasted a few milliseconds.

Table 2 summarises the mean and standard deviation values in milliseconds for token durations in the two contexts Maptask and Spontaneous, as well as the Spontaneous Control condition in the subcategories affective and functional (as a reminder: Speech from the Spontaneous Control condition consisted of Spontaneous conversations from the laboratory setting and without visual contact). An exploration of the mean durations shows that overall, BC tokens within the subfunction category affective were longer and more variable in duration. Additionally, there were differences found between conversation contexts with the shortest utterance durations and smallest standard deviations in Maptask conversations (mean: 430ms, SD:147ms), followed by tokens from Spontaneous conversations (mean: 766ms, SD: 321ms). The longest mean duration, but smaller variation was found the Spontaneous Control condition without visual contact (mean: 874ms, SD: 279ms).

In the functional category, mean token durations in the two Spontaneous conditions differed only marginally with the higher values occurring the Spontaneous Control condition without visual contact (mean: 458ms, SD: 230ms), and slightly lower values in the Spontaneous condition with visual contact (mean: 417ms, SD:143ms). In the Task-Oriented condition, however, the mean value was clearly lower and the standard deviation smaller (mean: 309ms, SD: 107ms)

Table 2: *Mean and SD of token duration given in milliseconds, for all tokens per condition and subfunction.*



Condition	Subfunction	Mean (duration in ms)	SD (duration in ms)
Maptask	affective	430	147
Maptask	functional	309	107
Spontaneous	affective	766	321
Spontaneous	functional	417	143
Spont_Control	affective	874	279
Spont_Control	functional	458	230

It should be noted, however, that in the Maptask and Spontaneous Control condition only very few tokens fell within the affective category. Nevertheless, one can see a clear tendency towards longer, and more variable token durations in the category of affective backchannels as compared to functional tokens, especially in Spontaneous contexts, as well as a higher proportions of these affective tokens in Spontaneous conversation contexts.

4.3. Intonational Realisation - Continuous

Besides the investigation of backchannel rate and lexical choice, one of the two main foci in the analysis is the intonational realisation of all the extracted tokens from the different lexical categories. For that purpose, the semitone analysis described in chapter 3.4.1 was applied to all backchannel utterances. In this analysis, the absolute pitch was measured in Hz at the beginning (after 10%), and end (after 90%) of each token and the pitch movement in semitones was calculated (see figure 3, chapter 3.4.1).

Tokens which were rated not suitable for intonational analysis due to creaky voice, microprosody, or non-reliable F0 contour were excluded. Of the 508 tokens originally extracted, 448 were considered suitable for analysis. Results on these tokens will be reported in the following sections.

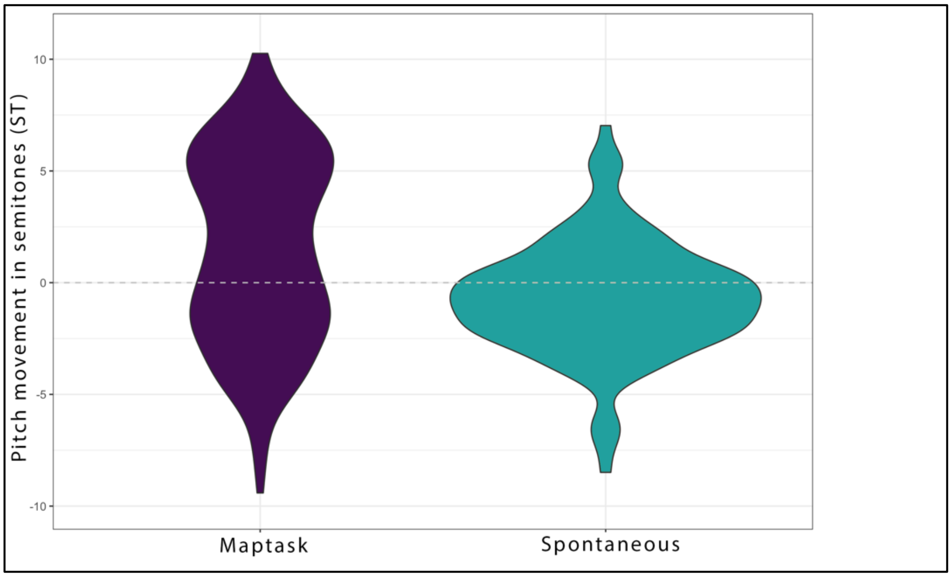


Figure 10: *Violin plots of pitch movement in semitones for all BC tokens from Maptask conversations in purple and for all BC tokens from spontaneous conversations in yellow. Values below zero indicate falling contours, values above zero indicate rising contours.*

First, the overall results from descriptive, continuous analysis will be reported. Then, each token will be categorised into one of the three pitch movement categories 1) rising, 2) falling and 3) level/flat for a subsequent categorical analysis. Results for the categorical analysis will be reported in chapter 4.4.

Figure 10 illustrates the pitch movement in semitones for all tokens uttered in the Maptask (purple) and Spontaneous (turquoise) conversation context. Positive values for pitch movement in semitones (ST) indicate a rising pitch movement, while negative values indicate a falling pitch movement. Based on this figure, a first important observation can be made: In the Maptask condition, values seemed bimodally distributed with two maxima 1) rising pitch movements of around 6 semitones, as well as 2) falling pitch movements of around 1.5 semitones. Values from the Spontaneous condition seemed to be normally distributed (mean value slightly falling around 1 semitone). Figure 11 illustrates the mean values for all tokens per dyad and individual speaker with values from Maptask conversations on the left and values from Spontaneous conversations on the right.

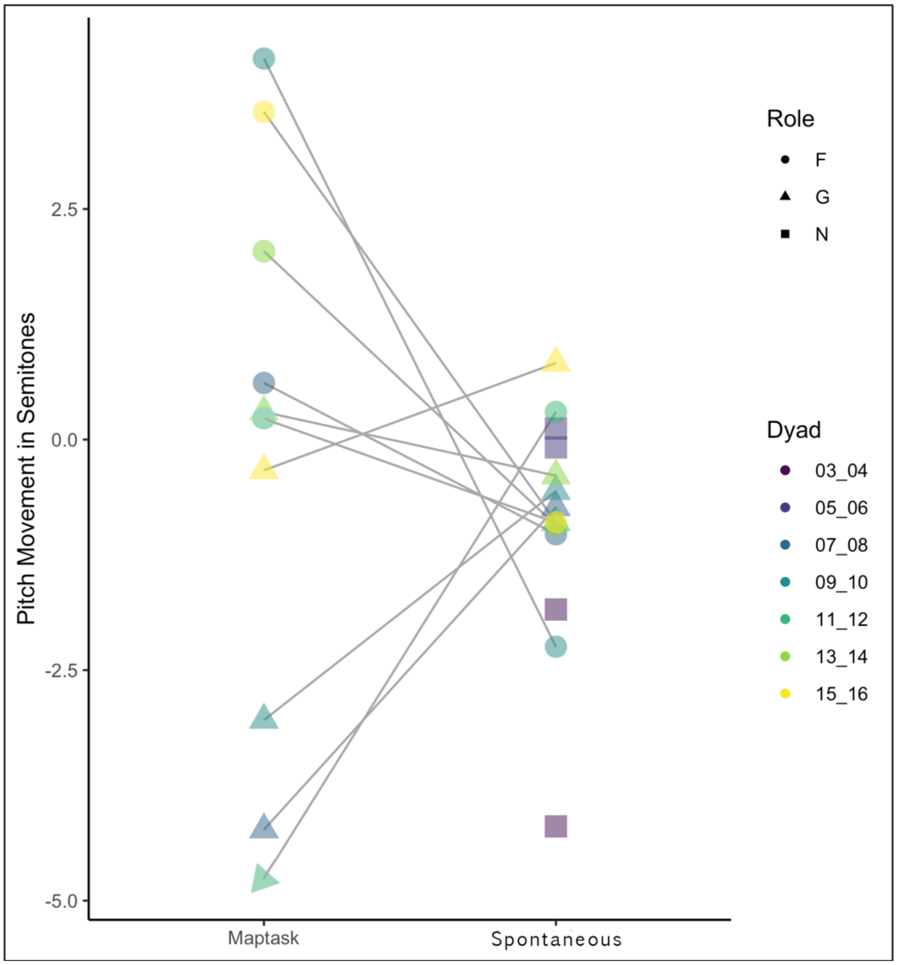


Figure 11: *Mean values for pitch movement in semitones for individual speakers in the Maptask (left) and spontaneous (right) conversation contexts. Speakers of the same dyad are displayed in the same colour, grey lines connect datapoints of one speaker in the two conditions. Instruction followers (Maptask condition) are represented as circles, instruction givers as triangles, speakers who only participated in the spontaneous recording are displayed as squares*

In this figure, participants who had the role of instruction follower in the Maptask are represented by circles, while instruction givers are represented by triangles. Subjects who did not take part in any Maptask conversation are depicted as squares. Speakers of the same dyad are always represented in the same colour.

Investigating Figure 11, it seems that instruction followers in the Maptask to use more feedback signals with clearly rising pitch movements as compared to instruction givers who on average used predominantly falling pitch movements. In the Spontaneous condition on the other hand, all participants used, on average, mostly flat or slightly falling pitch movements. One dyad (03_04) depicts an exception with both speakers using predominantly clearly falling F0 contours (an additional figure containing mean values also for the Spontaneous Control condition can be found in the appendix).

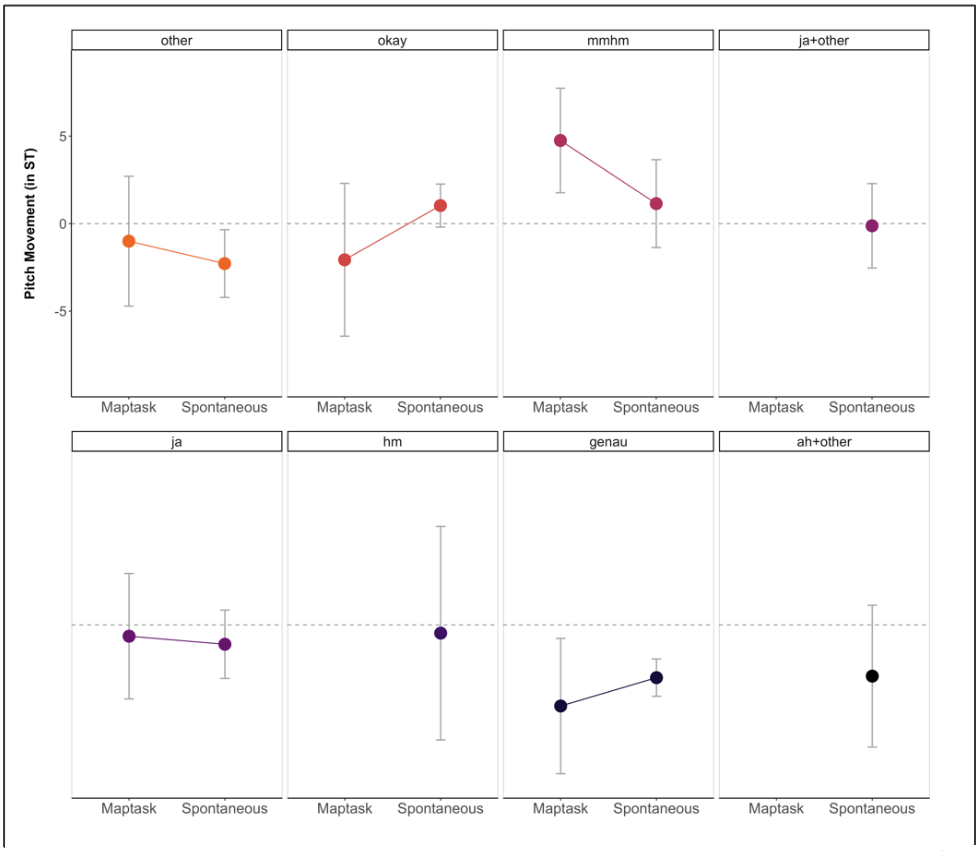
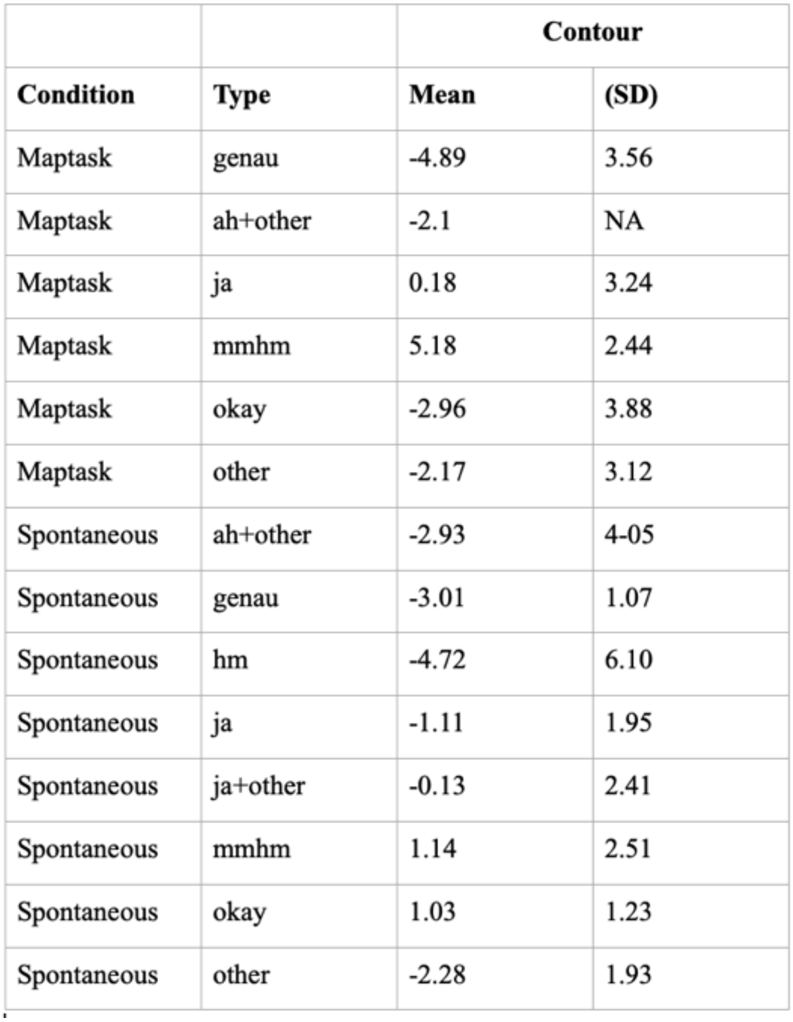


Figure 12: *Pitch movement in Semitones for all lexical BC types in the two main contexts. Tokens from three categories only occurred in the spontaneous condition. Values below zero indicate falling pitch contours, values above zero indicate rising contours. Grey bars represent standard deviations*

Different lexical tokens are likely to carry either rising or falling token-specific intonation contours. For a meaningful analysis of intonation contours it is therefore useful to investigate intonation contours on the level of backchannel type. Continuous ST values for each backchannel type as illustrated by Figure 12, reveal differences between the individual types. In general, two important observations can be made based on this figure.

First, standard deviations of pitch movement in semitones for all token categories were usually higher for the Maptask condition as compared to the Spontaneous condition. Second, for almost all categories with occurrences in Maptask, and Spontaneous conversations, the mean values of pitch movement in semitones for the same tokens in the two different conditions differ to a certain degree. Exact values for all standard, and non-standard token categories are displayed in table 3.

Table 3: *Mean pitch movement in semitones of BCs by type and condition. Negative values indicate falling contours, positive values indicate rising contours.*



		Contour	
Condition	**Type**	**Mean**	**(SD)**
Maptask	genau	-4.89	3.56
Maptask	ah+other	-2.1	NA
Maptask	ja	0.18	3.24
Maptask	mmhm	5.18	2.44
Maptask	okay	-2.96	3.88
Maptask	other	-2.17	3.12
Spontaneous	ah+other	-2.93	4-05
Spontaneous	genau	-3.01	1.07
Spontaneous	hm	-4.72	6.10
Spontaneous	ja	-1.11	1.95
Spontaneous	ja+other	-0.13	2.41
Spontaneous	mmhm	1.14	2.51
Spontaneous	okay	1.03	1.23
Spontaneous	other	-2.28	1.93

4.3.1. Standard Tokens

In the following I will report values on the four *standard* backchannel categories *ja, mmhm, okay,* and *genau* which have been investigated in past studies concerning Maptask conversations. In the *ja* category, I mostly observed tokens with slightly falling pitch movements in both categories. In the Maptask condition an average pitch movement of -0.18 semitones with a standard deviation of 3.24 semitones was measured, while in the Spontaneous condition a slightly more falling mean value of -1.12 semitones with a much smaller standard deviation of 1.95 semitones, was observed. In summary, *ja* tokens carried a slightly stronger falling pitch movement in the Spontaneous condition as compared to the Task-Oriented condition, but values were also much more variable in the Task-Oriented condition as compared to the Spontaneous condition.

For non-lexical backchannel tokens, I differentiated between disyllabic *mmhm*, and monosyllabic *hm* tokens (results for the monosyllabic *hm* will be reported in chapter 4.3.2). For the disyllabic *mmhm* category, mean values for the Maptask condition were exclusively rising with an average pitch movement of 5.1.8 semitones and a standard deviation of 2.44 semitones. In tokens from the Spontaneous conversation context, tokens were predominantly rising as well, with a mean value of 1.11 semitones and a standard deviation of 1.95 semitones. Generally, it can be noted that *mmhm* was observed to be predominantly rising, but with much stronger rises in the Maptask as compared to the Spontaneous condition.

Concerning tokens of the lexical *genau* category, only falling mean values of pitch movements in semitones were observed in both conditions. In the Maptask condition, the mean pitch movement valued -4.89 semitones with a standard deviation of 3.56 semitones. In the Spontaneous condition, average pitch falls were slightly less extreme with a mean value of -3.01 semitones and a relatively small standard deviation of 1.07 semitones.

The last standard token category is the lexical *okay* category. In this category, the greatest differences in mean pitch movements between the Maptask and Spontaneous condition were observed. In the Task-Oriented context, almost all *okay* tokens carried a falling pitch movement with a mean value of -2.96 semitones and a corresponding standard deviation of 3.88 semitones. In opposition to that, most *okay* tokens in the

Spontaneous conversation context carried a slightly rising F0 contour with a mean value of 1.03 semitones and a standard deviation of 1.23 semitones.

4.3.2. Non-Standard Tokens

In the following, I will report results on the non-standard token categories and their intonation contours. The monosyllabic *hm* tokens only occurred in Spontaneous conversation contexts. Tokens from this category carried falling, as well as rising pitch contours, as illustrated in figure 12 and table 3, with a mean value of -4.72 semitones and standard deviation of 6.10 semitones.

All compound tokens consisting of *ja* plus another component, or *ah* plus another component were categorized into the classes of *ja+other*, and *ah+other*. In the Maptask context, no utterance of *ja+other* and only one utterance of *ah+other* occurred, therefore no mean values will be reported for these two token types in the Task-Oriented condition. Mean values for tokens uttered in the Spontaneous condition averaged -0.13 semitones and standard deviation was 2.41 semitones for *ja+other*, and a mean of -2.93 semitones with a standard deviation of 4.05 semitones for *ah+other*.

All tokens for which the lexical content did not qualify for the above-mentioned categories were consolidated in an *other* category. The pitch movements of these tokens were predominantly falling with a mean of -2.17 semitones and standard deviation of 3.12 semitones in the Maptask condition, and an average of -2.28 semitones and corresponding standard deviation of 1.93 semitones in the Spontaneous context.

4.4. Intonational Realisation - Categorical

For the categorical analysis of intonational realisation, all contours with pitch movements of less than one semitone were categorized as level or flat (Wehrle, 2021). All tokens above or below that threshold were counted as rising (positive values above threshold) or falling (negative values above threshold). Overall, the categorical analysis supports the impression of a higher proportion of tokens with rising pitch contours in the Maptask (rise: 52.7 %, level: 12.7 %, fall: 34.6 %) condition as compared to the Spontaneous condition (rise: 20.5 %, level: 34.4 %, fall: 45.1 %). This is true both, considering all tokens from one condition, as shown in figure 13, as well as for some of the individual lexical categories as illustrated in figure 14.

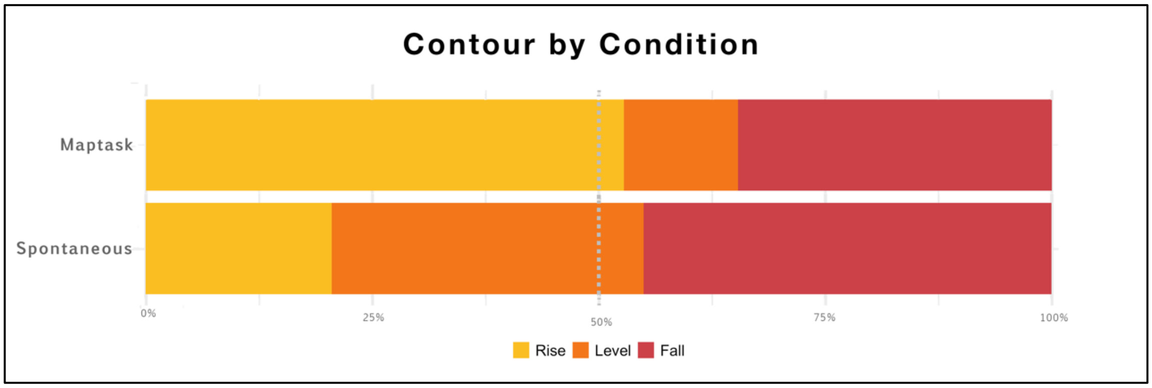


Figure 13: *Proportions of pitch contours per condition with Maptask on top and Spontaneous at the bottom. Dotted grey line indicates fifty percent. Rising contours are depicted in yellow, level contours in orange and falling contours in red.*

On the level of type the highest proportion of rising tokens was observed in *mmhm* tokens from the Task-Oriented context (rise: 94.3 %, level: 4.1 %, fall: 1.6 %) while in the Spontaneous category, only 45.2 % of all tokens carried a rising, 35.5 % a level, and 19.3 % a falling intonation contour. Tokens from the *ja* category uttered in a Task-Oriented context were predominantly rising or level (rise: 36.8 %, level: 25.5 %, fall: 37.7 %), while over half of the *ja* tokens from the Spontaneous conversation context carried a falling F0 contour (rise: 13.0 %, level: 37.0 %, fall: 50.0 %).

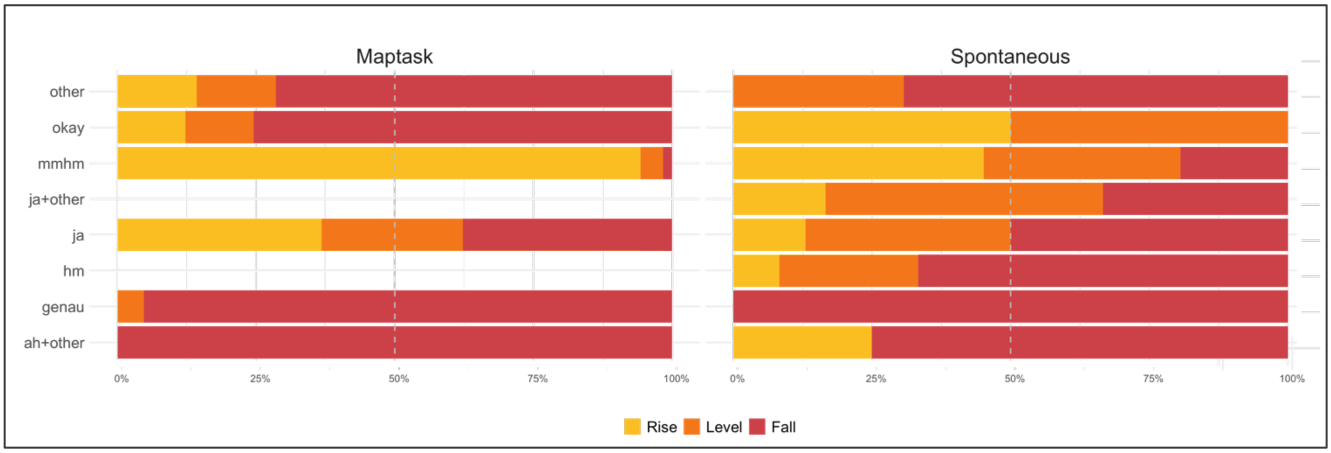


Figure 14: *Proportions of rising(yellow), level(orange), and falling(red) intonation contours on BC tokens from eight different categories in Spontaneous (top) and Maptask (bottom) conversations. Vertical grey, dotted lines indicate a proportion of 50 percent.*

Tokens from the *okay* and *genau* categories on the other hand, were predominantly falling in Task-Oriented speech (*okay:* rise: 12.3 %, level: 12.3 %, fall: 75.4 %, *genau:* rise: 0.0 %, level: 4.8 %, fall: 95.2 %). But while tokens from the *genau* category were exclusively falling in their F0 contour in Spontaneous dialogue, 50 % the *okay* tokens from a Spontaneous context were rising, and 50 % carried a level intonation contour. Those tokens categorized as *ah+other* were exclusively falling in Task-Oriented, and predominantly falling in the Spontaneous conversation context (Spontaneous: rise: 25 %, level: 0.0 %, fall: 75 %). Tokens from the non-lexical *hm* and lexical *ja+other* type only occurred in the Spontaneous context and were mostly level and falling in their intonation contour (*hm:* rise: 8.3 %, level: 25 %, fall: 66.7 %; *ja+other*: rise: 16.7 %, level: 50 %, fall: 33.3 %).

All tokens which deviated in their lexical or non-lexical content from the above-mentioned categories, were categorised as *other* tokens. In both conversational contexts, these tokens were mostly falling in intonation contour (Task-Oriented: rise: 14.3 %, level: 14. 3%, fall: 71.4 %; Spontaneous: rise: 0.0 %, level: 30.8 %, fall: 29.2 %).

On the level of individual speaker (Figure 15), in the Maptask condition, most speakers with the role of instruction follower predominantly used tokens with rising contours, while most instruction givers used predominantly falling or level tokens, also in comparison to the other speaker from the same dyad. Speaker 13 (instruction giver dyad 13_14) was an exception. As instruction giver in this dyad the speaker used 46.0 % rising, 24.3 % level, and 29.7 % falling tokens.

In the Spontaneous condition proportions for individual speakers were very similar to the overall proportions with most speakers using predominantly level or falling contours and lower proportions of rising contours as compared to the Task-Oriented condition. Speakers 3 (of dyad 03_04), 10 (of dyad 09_10), and 11 (of dyad 11_12), were exceptions in that regard. Speakers 03, and 11 used 50 % falling, and 50 % level contours, while speaker 10 used exclusively falling contours in the Spontaneous condition.

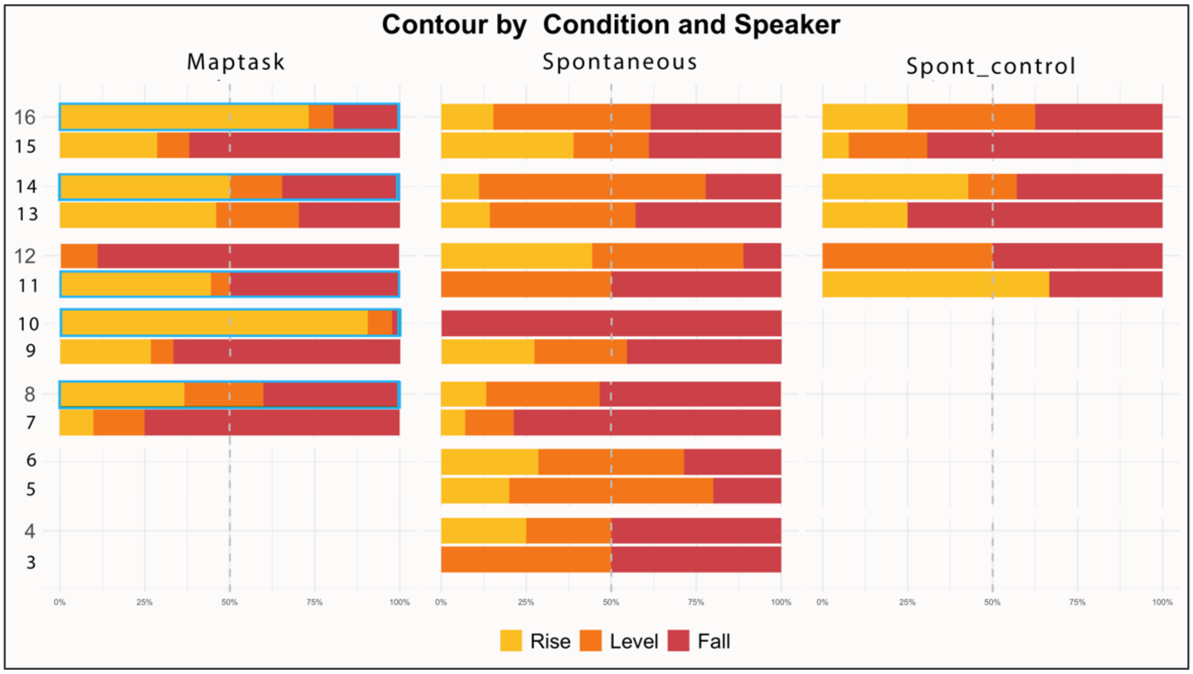


Figure 15: Proportions of rising, level, and falling contours per speaker and condition. Turquoise boxes indicate role as instruction follower in the Maptask condition, speakers of a dyad are paired together.

4.4.1. Bayesian Analysis

For the purpose of testing the overall hypothesis that participants used greater pitch movements in semitones in the Task-Oriented condition as compared to the Spontaneous condition, a Bayesian analysis was carried out using only absolute values of pitch movements in semitones. The model supports the interpretation that subjects used greater pitch movements in Task-Oriented (β_Maptask = 3.82, CI = [3.57, 4.09]), as compared to Spontaneous conversation (β_Spontaneous = 3.26, CI= [2.67, 3.82]), and as compared to the Spontaneous Control condition (β_Spont_Control = 3.65, CI= [2.91, 4.34]). There is compelling evidence for this difference P (β_Spontaneous < 0) = 1.

Hypothesis testing showed no reliable difference between the Spontaneous Control condition and the Task-Oriented condition P (β_Spontaneous < 0; = 0.78). The corresponding model including hypothesis testing can be found in the appendix.

A second model was run using all *original* (positive and negative) values for pitch movement in semitones from the two most frequent token types *ja* and *mmhm*. Although the model did not fit the data perfectly due to the small amount of data points, it did confirm a narrower distribution of pitch movement values around the mean for the Spontaneous condition (β = -0.22, CI= [-2.25, 1.71]), as compared to the Task-Oriented condition (β = 2.12, CI= [1.05, 3.08]). There is compelling evidence for this difference P (β_Spontaneous < 0) = 1.

4.5. Contextual Differences

In this section, some contextual differences between the two conditions Spontaneous and Task-Oriented will be reported. Five out of seven dyads participated in both the Spontaneous and the Maptask conversation which allows for a direct comparison of conversation topics. After listening to all conversations, conversation topics for the Spontaneous conversations were roughly grouped into 1) discussion of the task that preceded the conversation, 2) chatting about every-day-topics, 3) chatting about future plans, 4) scheduling, 5) storytelling, and 6) chatting about the surroundings. For the Maptask conversations, all participants were considered to be *on task* if they did not talk about topics unrelated to the task. Only if participants wandered from the subject of the Maptask, it was noted.

Table 4 shows an overview of all conversational contexts, as well as examples of the lexical tokens used by the participants in Maptask and Spontaneous speech. As shown in the table, most dyads were *on task* during the Maptask conversation, meaning they did not talk about things unrelated to the task during the Task-Oriented conversations. During the Spontaneous conversations on the other hand, participants conversed about a variety of topics. Investigating Table 4, it is apparent that the conversation topics were much more diverse in the Spontaneous context and subjects tended to change topic multiple times during one conversation.

Table 4*: overview table of conversation contexts in Maptask and Spontaneous conversations, examples of lexical backchannels*

Dyad/ Context	Conversation Context		Lexical tokens (examples)	
	Maptask	Spontaneous	Maptask	Spontaneous
03_04	--	discussing previous task	--	‘ja’ ‘okay’
05_06	--	chatting about upcoming semester abroad of	--	‘krank digga‘ ‘ja nice‘ ‘geil‘ ‘ja’ ‘ja safe’
07_08	‘on task’	chatting about every-day topics	‘ja’ ‘okay’ ‘genau’	‘okay‘ ‘cool‘ ‘schön‘ ‘ja voll‘ ‘klar‘
09_10	‘on task’	discussing previous task, chatting about every-day topics	‘ja’ ‘okay’ ‘genau’	‘stimmt’ ‘jaja voll’ ‘ja’
11_12	‘on task’	storytelling by one participant, chatting	‘ja’ ‘okay’ ‘genau’	‘ach crazy‘ ‘okay‘ ‘genau‘ ‘stimmt‘
13_14	‘on task’ high proportions of laughter	discussing previous task, chatting about every-day topics	‘ja’ ‘okay’ ‘genau’	‘ja’ ‘stimmt’ ‘eben”

15_16	‘on task’	chatting about surroundings, every-day topics, scheduling, storytelling	‘ja’ ‘okay’ ‘genau’	‘ah mega‘ ‘oh stimmt‘ ‘nice‘ ‚ja’

5. Discussion

In this thesis, I investigated how backchannel intonation, frequency and lexical choice differs across two conversational contexts. In chapter 2.4, it was predicted that:

i) backchannel use would be more frequent in Task-Oriented as compared to Spontaneous conversation

ii) there would be differences in the intonation contours between the two conversational contexts, specifically more rising backchannels in Task-Oriented conversation as compared to Spontaneous speech

iii) there would be differences in the proportions of tokens used in the different contexts.

In this chapter the results presented above are discussed in relation to the research question and predictions.

5.1. Backchannel Characteristics in Task-Oriented vs. Spontaneous Speech

In chapter four, the results on backchannel rate, intonation, lexical choice, and contextual differences were presented. Table 5 (below) summarises the most important findings in relation to each other. Task-Oriented and Spontaneous conversations showed structural and contextual differences. In Maptask conversations, all subjects were *on task*, meaning they did not digress from the topic of the task. In the Spontaneous conversations on the other hand, many changes in conversation topic and context were observed. This distinction between the two conditions should be kept in mind regarding the discussion and interpretation of the results on BC rate, lexical choice, and intonation.

Table 5*: Overview of all results. '+', '-', and '=' are used to indicate higher, lower or similar rates and values for all contexts in relation to each other.*

Condition	**Context**	**BC rate per minute**	**Lexical choice**	**Intonation**
Maptask	**- No changes** in conversation topics/context	**+ 5.5** BC per minute **+** clear **differences** in rate between speakers	**+ 99%** *functional* backchannels **- Small variations** in BC token duration **-** very **few deviations** from *standard* tokens	**+** over **50% rising** tokens **-** around **30% falling** tokens **+** greater pitch movement in ST as compared to spontaneous
Spontaneous	**+ many changes** in conversation topic/context	**- 3.76** BC per minute **- similar** rates for both speakers	**-** Only **81%** functional backchannels **+ Longer** BC token **durations** compared to Maptask **+** great duration variation in *attitude* tokens **+ many *unusual*** BC tokens	**-** less than **25% rising** tokens **+** Almost **50% falling** tokens **-** less *extreme* rises and falls
Spontaneous control	**/**	**- 3-35** BC per minute **- similar** rates for both speakers	**= similar durations** compared to spontaneous **- fewer unusual** tokens as compared to spontaneous	**/**

5.1.1. Backchannel Rate

In chapter 4.1, I presented results on backchannel rate in Task-Oriented and Spontaneous conversations, as well as the Spontaneous Control condition. My findings regarding a higher backchannel rate in Task-Oriented as compared to Spontaneous conversation are in line with the findings reported by Dideriksen et al. (2021), who also investigated backchannels in Maptask and Spontaneous conversational contexts in native speakers of Danish. As predicted, a higher rate of feedback signals was found in Task-Oriented contexts, which supports the assumption that these conversations are more informationally dense. In Maptask conversations, which were chosen here as the Task-Oriented condition, every new landmark, its position on the other interlocutor's map, and every new direction is new information. It therefore makes sense for the listening interlocutor to give a lot of feedback to signal that they have received and understood the information.

Additionally, results on backchannel rate in the Maptask reflect a clear separation between instruction giver and instruction follower, with the instruction follower using more backchannels, potentially encouraging longer stretches of speech by the instruction giver. In the Spontaneous conditions, these clear differences in rate between speakers were not found. Backchannel rates, in general, were much lower and differences between the two speakers smaller. As the strong separation between speakers, found in the Maptask, is not present in Spontaneous speech, it can be assumed that in this conversational context no clear separation is desired. The goal in Spontaneous dialog rather seems to be an essentially balanced conversation with similar portions of speech by each interlocutor (Gilmartin et al., 2018).

When comparing the BC frequencies found in the data at hand to similar studies investigating BC rate in Maptask conversations (e.g. Wehrle, 2021; Sbarnna et al., 2022) it becomes evident that rates observed in the data set presented here were generally lower. this is true especially in comparison to the data investigated by Wehrle et al. (2021) in which subjects were not familiar with each other. It is therefore likely to assume that the higher familiarity between my subjects influenced the BC frequency in the Maptask.

5.1.2. Lexical Choice

Turning to the subject of lexical load or lexical choice it becomes clear that despite the overall similarities in choice of backchannel type, there are subtle differences between the two conditions. In, Task-Oriented, as well as Spontaneous conversation contexts, the tokens *ja* and *mmhm* are the most used. This finding was expected, at least for Task-Oriented contexts, considering that these token types are the ones most extensively studied (e.g. in Bertrand et al., 2007; Heinz, 2003; Trouvain, 2014; Wehrle, 2021). Additionally, few deviations from standard tokens such as *ja, okay, genau* and *mmhm* were observed in the Maptask context. Disyllabic tokens made up for the highest proportion in the Maptask condition but only few were uttered in the Spontaneous condition. As two-syllable tokens have been suggested to signal a listening role (Ward, 2004), one could reason that the proportion of these tokens is influenced depending on the requirements of the conversational context. That is, in Task-Oriented dialogue, one interlocutor clearly takes the role of a listener and can therefore be expected to produce predominantly disyllabic tokens. However, in the exploratory investigation of backchannels at hand, only those tokens uttered during or between two turns of the other speaker were included into the analysis. This is true for both, the Maptask and the Spontaneous conversation context. Therefore all tokens presented here can be assumed to signal a listening role.

In Maptask conversations standard tokens predominated; while in Spontaneous dialogue *unusual* lexical tokens were used with a much higher proportion as compared to the Task-Oriented context. The lexical variety of these tokens also supports the assumption that feedback signals are adjusted to the context, not only in rate, but also in their lexical form. One might argue that this lexical variation was evoked by frequent topic changes and variations in the conversational context.

In the study at hand, token duration was measured to roughly estimate the segmental content of lexical tokens. Presumed duration is an adequate measure for this estimation, longer token durations as found in Spontaneous conversations would make these tokens more likely to interfere with the main channel and to be consciously noted by the speaking interlocutor. In addition to context-specific differences on a group level, the choice of backchannel token seemed to be highly speaker-specific. Especially when considering affective backchannels (Pammi & Schröder, 2009) many tokens were only used by single speakers, for example *krank digga* ('sick dude'), *geil* ('awesome'), or *ach crazy*. Familiarity or the social relationship between interlocutors appeared to be an influential factor on lexical choice in past research (Giles, Coupland, J. Coupland, 2010). Further, it has been shown that speakers tend to converge in their speech to vast extents (Fusaroli et al., 2014; Pickering & Garrod, 2004). Still, no obvious effect of lexical entrainment stood out in the data presented above. All subjects in this study shared their living environment at the time of the recordings and stated to regularly spend time together. The lack of obvious convergence effects with regards to lexical choice is therefore more than unexpected because usually more entrainment effects can be observed in dialogue of people who like each other (Fusaroli et al., 2014).

5.1.3. Intonation

In chapter 2.4, prior assumptions regarding the results of the investigation of intonation contours and lexical choice of backchannels were presented. In terms of intonation contours, it was assumed that there would be a higher amount of BC tokens with rising pitch movement in the Task-Oriented context and respectively more tokens with falling contours in the Spontaneous conversation contexts.

Indeed, more than half of all backchannel tokens that were produced in the Maptask conversations carried a clearly rising intonation contour, whereas in the Spontaneous conversation context, the amount of rising BC tokens was less than a quarter of all uttered tokens. Still, descriptive statistics in this study showed that overall, the distribution of BC contours was bimodal with many clearly rising backchannels on the one hand, but also many falling backchannels while BCs with a level intonation contour did not play a particularly big role in Task-Oriented conversation.

Considering the fact that in the Maptask participants had a clear role as either instruction giver or instruction follower, it makes sense that many uttered backchannels were continuation signals by the instruction follower, signalling that they understand the instructions. *Mmhm* is quintessential as a continuer and used with much lower proportions in incipient speakership (Sbranna et al., 2022). In the study presented here, *mmhm* was used in almost 40 percent of all backchannel utterances in Task-Oriented speech but only in roughly 25 percent of all Spontaneous backchannel utterances. More importantly, it was shown that in Task-Oriented contexts *mmhm* tokens carried a token-specific rising intonation contour which was independent of the token's specific function. While evidence for that has been provided by multiple studies to date (Sbranna et al., 2022) my data suggests that this preference for rising contours in *mmhm* tokens is less distinct in a Spontaneous conversation context. In Spontaneous contexts, subjects only used rising contours in less than 50 % of all cases (see figure 14, chapter 4.4). Considering the number of studies on Task-based dialogue in which predominantly or exclusively rising contours were found for *mmhm* tokens, further investigations are needed for the use of *mmhm* in Spontaneous conversation.

However, the data at hand confirms the hypothesis of instruction followers using more rising tokens, and therefore, probably more continuation signals as compared to the instruction giver in the Maptask. Furthermore, rising tokens in the Maptask condition were in most cases much steeper than those uttered in Spontaneous dialogue. Considering that listeners are more sensitive to pitch rises as compared to pitch falls and differences between rises are perceived as greater than those between falls (Hsu et al., 2015), one could draw the conclusion that rising *mmhm* tokens in Task-Oriented speech are perceived as more prominent than those produced in a Spontaneous context. In the latter context, the differences between rising and falling contours are more subtle with many flat and falling contours and less extreme pitch movements.

The falling equivalent which was also shown to carry a token-specific intonation contour is *genau* (Sbranna et al., 2022). In the case of *genau* the token's intonation seems to be not only independent of its specific function but also independent of the conversational context it is produced in. The data at hand showed that in both the Task-Oriented and the Spontaneous conversation context, *genau* carried a falling intonation contour. Even though on average, pitch falls in the Maptask condition were slightly steeper than in the Spontaneous condition, it is likely that the difference would not be perceived by a listener.

Past studies have shown that for the tokens *ja* and *okay*, the mapping of intonation contours is less clear. Both tokens can be used with different pragmatic functions and with rising or falling intonation contours (Ha et al., 2016). In my study, *okay* tokens carried mostly falling intonation contours in the Maptask condition, whereas tokens in

Spontaneous conversations were on average mostly rising. With *ja*, on the other hand there was no such distinction. While Wehrle et al. (2019) found *ja* to primarily carry a rising intonation contour, the data at hand only showed a slight tendency towards more falling *ja* tokens in the Spontaneous condition as compared to the Maptask condition where most tokens carried a flat intonation contour. Thus it appears that results for both contexts are contrary to Wehrle et al.'s (2019) findings. Even though only few data points were available for *okay,* the data seems to suggest a pragmatic difference in how this token is used in Task-Oriented as compared to Spontaneous conversation, but no such difference for how *ja* is used.

In the Task-Oriented conversation context, interlocutors have clear roles attributed to them. This has been shown to lead to a high proportion of rising BC tokens in Maptask conversations (Savino, 2010; Sbranna et al., 2022; Wehrle & Grice, 2019). Most studies working with dyadic conversation do not investigate single speakers but consider speakers to be strongly interdependent so their behaviours cannot be investigated independently. Still, depending on the context of the conversation, the degree to which speakers depend on and influence each is likely to differ. Although there was no attribution of speaker roles in Spontaneous conversation, there also seemed to be a tendency towards one speaker uttering a higher proportion of rising tokens than the other. This is in support of the idea that in every conversation, one speaker is slightly more dominant and has a higher influence on the conversational dynamics (Itakura, 2001). These differences in speaker-specific behaviour cannot be captured when investigating dyads as a unit. Overall however, most speakers behaved rather similarly to their interlocutor in the Spontaneous conversation context.

In Maptask conversations, backchannels with rising and falling intonation contours were generally the ones with the highest proportion. Tokens with level intonation contours, however, are more unusual in this conversational context (Savino, 2010, 2011; Sbranna et al., 2022; Wehrle et al., 2018; Wehrle, 2021). While the data at hand confirms these findings for the Task-Oriented condition, results for the Spontaneous context showed a different picture. Here, almost half of all uttered backchannels had a flat intonation contour, a tendency which was stronger in Spontaneous dialogue *with* visual contact between the interlocutors than in those *without* visual contact. These findings are contradictory to e.g. a study on backchannels in German, in which flat tokens were rated as inappropriate, unfriendly or inattentive by listeners (Wherle et al., 2018). However, in Wehrle et al. (2018), subjects were asked to rate backchannels with different intonation contours while listening to only one part of a conversation. One must consider that in a *natural* setting, the interlocutors are most often speaking while perceiving backchannels and therefore not consciously paying attention to the tokens. One possible explanation for negative ratings of level tokens might therefore be that when actively listening to backchannels, they are perceived differently. It has also been shown that oftentimes backchannels are too quiet to possibly be heard by the speaking interlocutor (Ward & Tsukahara, 2000), which is in support of the theory that backchannels in conversation are not always consciously perceived.

Overall, the study presented here has shown backchannel rate to be higher in Task-Oriented as compared to Spontaneous conversation, and differences in the lexical choice of backchannel tokens to be present. Lexical choice, however, seems to be not only context-, but also speaker-specific. Results for backchannel intonation have shown overall pitch movement to be more extreme in Task-Oriented conversation, and the proportion of rising, falling, and flat tokens to differ between contexts. Visual contact (or the lack thereof) between interlocutors seemed to only make a difference when categorically looking at intonation contours.

5.2. Backchannel Functions in Different Contexts

After backchannel characteristics in the two experimental conditions were examined independently in the previous chapter, the emphasis will now be on the functions of modifications in i) the domains of backchannel rate, ii) lexical choice, and iii) intonation. In the results chapter it was already mentioned that the present study only included backchannel tokens uttered during or between turns of the other interlocutor. Tokens that initiated a new turn were excluded. This selection already narrows down the range of possible backchannel functions. In the study at hand, all extracted passive recipiency tokens were further categorised into functional and affective tokens.

This categorisation generally showed that the proportions of backchannel functions differ between the two conditions. In Maptask conversations, backchannels seemed to predominantly manage the conversation from a functional perspective. That is, they guided the structure of the conversation, enabled smooth turn-taking, signalled comprehension, and told the interlocutor to keep on speaking (Heinz, 2003; Sacks et al., 1974; Wehrle, 2021). In Spontaneous dialogue, most tokens were functional tokens as well, but the proportion of affective backchannels was higher, compared to Task-Oriented dialogue. The following sections elaborate on the different features of functional and affective backchannelling tokens in terms of their intonation, rate and lexical choice, and interpret them based on past research.

5.2.1. Backchannel Rate

Investigation of backchannel rate in the two contexts showed that in Task-based conversations, a higher rate of functional backchannels, such as the non-lexical *mmhm,* was used as compared to affective backchannel tokens. Considering the high amount of new information which is introduced into the conversation in tasks like the Maptask, as well as the imbalance of turn duration between the two speakers, it is not surprising that interlocutors in this study uttered many continuation signals (of which *mmhm* is the quintessential one (Sbranna et al., 2022)). To reach the goal of the conversation, both speakers must be clear and precise in the way they communicate with each other (Dideriksen et al., 2021). Additionally, when speakers are on task, the information they share is mainly guided by the task and less by the individuals and their interests or personalities themselves. Reacting by producing affective tokens (apart from surprise about unexpected information) would not benefit the goal of the task.

In opposition to the information shared in Task-based conversation, information or content shared in Spontaneous conversation is usually guided by the interlocutors themselves rather than an attributed task. At the same time, Spontaneous conversations are less informationally dense (Fusaroli et al., 2017), probably even less in conversations of participants that are very familiar with each other. It can be assumed that this leads to the lower rate of backchannels found in Spontaneous dialogue. In a Spontaneous conversation a higher proportion of lexical tokens signalling affective states (such as *nice*, *super*, and *mega*) or affirmative tokens (for instance *voll* ('totally') and *stimmt* ('true')) was found. It is likely that in Spontaneous dialogue, the relationship between backchannel choice and content of the preceding utterance is closer than it is in Task-Oriented conversations.

5.2.2. Intonation Contours and Visual Contact

A mentioned earlier, previous studies on backchannels and their token-specific intonation contour have shown clear tendencies towards prototypical intonation contours for some tokens such as *mmhm* which is most often rising and functioning as a continuer, as well as *genau* which predominantly carries a falling intonation contour. *Okay*, however, can have a rising or falling intonation contour depending on the pragmatic context, and results for *ja* are often inconclusive with some studies showing a preference for rising contours (Wehrle et al. 2019) while others not finding any preference (Ha et al., 2016). Results from the data at hand mostly confirms these findings for Task-Oriented conversations, but not for Spontaneous dialogue in which all tokens carried predominantly level or falling intonation contours. While the data set investigated here is small and the analysis chosen is rather approximate, my findings still challenge the general assumption that backchannels mostly have a rising intonation contour (Beňuš, Gravano, & Hirschberg, 2007; Caspers, Huang, Yuang, Tang, 2000; Savino, 2010; Wehrle & Grice, 2019 ; Sbranna, Möking, Wehrle, Grice, 2022).

One hypothesis on the structural differences in token-intonation could be that the standard lexical tokens might serve different pragmatic functions in Task-Oriented and Spontaneous contexts. These specific contexts should be taken into account in future investigations. Structural differences were also observed considering pitch movement in semitones in the two conditions. In Task-Oriented contexts, the overall pitch movement was generally larger as compared to Spontaneous contexts. One can assume that the steeper rises, and potentially also falls, produced on Task-Oriented tokens, are also perceived as being more salient as compared to those uttered in Spontaneous dialogue (Hsu et al., 2015). However, while the intonation contours of backchannels seemed to be more subtle and less salient in a Spontaneous setting, they often carried more lexical content and were longer in duration. Both factors made them more likely to actively interfere with the speaking interlocutor's turn and therefore influence the conversation on a content level (Ward & Tsukahara, 2000).

Backchannels with level intonation contours, however, might be less likely to be actively perceived by the speaking interlocutor. As mentioned before, studies investigating loudness in BC tokens were able to show that sometimes, backchannels are produced too quiet to possibly be heard by the other interlocutor (Ward & Tsukahara, 2000). As level contours are the least salient, they are less likely to be heard by the other interlocutor as compared to rising or falling contours. Especially when produced with a non-lexical token like *mmhm* or *hm* backchannels with flat intonation contours are possibly not noted by the speaking interlocutor. Therefore, one could speculate that backchannels with level contours are not produced as other-oriented signals, but rather function as a tool for the listening interlocutor to structure the utterance they just heard. In a perception study, Wehrle and colleagues (2018) have found participants to rate backchannels with level contours as impolite or inattentive. Following the train of thought that flat backchannels might be produced by the listening interlocutor for themselves, one could also hypothesise that as speakers, participants have only prototypical expectations of backchannels as rising or falling tokens. When actively listening to them without producing speech in parallel, level contours might be perceived with much more attention and might therefore sound unusual when diverging from the listener's prototypical expectation.

Turning to the subject of visual contact, the data presented in this thesis shows that whether interlocutors can see each other does not seem to make a difference in backchannel rate and choice of backchannel token. This is in line with findings from telephone conversations, in which participants could only see each other (Cathcart et al., 2003). Investigating the choice of intonation contour and pitch movement in semitones, however, visual contact did seem to make a difference. In Spontaneous conversations in which the interlocutors had no visual contact, there was a much stronger tendency towards one speaker leading the conversation and the other one following the lead by using more rising backchannels, which are generally more likely to be continuers.

Depending on the conversational setting, interlocutors seemed to choose a different speech modality. In situations where precision and specificity are required, as it was the case in the Maptask conversations, speakers used a less varied lexical repertoire and a higher rate of backchannels per minute. In Spontaneous conversational settings in which a task or clear conversational goal was absent, precision and specificity were not as important and overall, interlocutors used lower backchannel rates. In these contexts, lexical tokens used as backchannels were also more closely related to the utterance content and might have served not only to organise the conversation from a functional point of view but also to carry social and emotional information e.g. on the interlocutors' relationship, their attitudes towards each other, or attitudes and opinions about the content of an utterance. In terms of intonation, participants also seemed to adapt their behaviour to the requirements of the conversational context. While in Maptask interactions clear speaker roles were assigned to both interlocutors, which was reflected in the proportions of rising continuation signals uttered by the instruction follower, this clearly assigned hierarchy was missing in Spontaneous dialogue. Still, there seemed to be a tendency towards one speaker taking the lead in the conversation. Visual contact did not seem to have a major influence on lexical choice and backchannel rate. For intonation contour, however, the tendency towards conversational hierarchies was more pronounced in conversations without visual contact between interlocutors.

5.3. Notes on Common Ground

When discussing backchannels, one cannot circumvent the concept of common ground. In the introduction chapter, common ground was defined as mutual knowledge that the interlocutors of a conversation share and are aware of sharing (Keysar et al., 1998). The concept is important for all processes of comprehension because listeners can only comprehend utterances if they correctly identify the entities that are referred to (Keysar et al., 2000). Generally, scholars agree that backchannels play an important role in the construction and maintenance of shared knowledge and common ground (Bertrand et al., 2007; Dideriksen et al., 2021; Fusaroli et al., 2017; Wehrle, 2021) and BCs therefore play a key part in the comprehension of utterances. Backchannels per definition provide positive feedback to the speaking interlocutor (Savino, 2010; Dideriksen et al., 2021; Sbranna et al., 2022), continuously acknowledge when new information is introduced into the body of shared knowledge and reassure that information is in common ground when a speaker refers to it.

As demonstrated in this thesis, backchannels were produced with higher frequencies in Task-Oriented conversation contexts in which precision and efficiency were important to reach the goal of the task. The high number of backchannels uttered, as well as the intonation contours and therefore hypothesised backchannel functions make it tempting to speculate that in Task-Oriented conversation, common ground is simply updated and confirmed every time a backchannel is uttered. However, human behaviour is always guided and influenced by social conventions (Barr, 2004). These conventions vary depending i.a. on the community and cultural setting a person has been socialised in. Research has found that overall, people are good at estimating what other people in their community know, although these estimates are systematically biased towards the estimator's own knowledge (Keysar et al., 2003). In Task-based dialogue it is therefore likely that backchannels are not necessarily uttered only in places where they are beneficial for the speaking interlocutor. Listeners might rather follow a context-specific set of communicative conventions, guiding them to produce backchannels in all places in which they feel they might *potentially* be beneficial for the speaking interlocutor. This way not every single backchannel would be used to actually update common ground, but the listener would simply make sure to give as many updates as possible in conversations where precision and efficiency are important. This train of thought is along the lines of what Schegloff described his 1982. He stated that while backchannels are often described as *signalling attention* they are at most *claiming* to signal attention and/or understanding and are not necessarily correct claims of this behaviour.

Spontaneous conversations by speakers who are very familiar with each other, are often less informationally dense (Fusaroli et al., 2017b), meaning that less new information is introduced into mutual knowledge. Also, the information already present in common ground does not have to be updated constantly because interlocutors might already be more certain about what is common ground when information has been referred to in the past. When considering the high amount of backchannel tokens uttered with level contours, which are not very prominent, in the Spontaneous condition along with the fact that sometimes backchannels are too quiet to possibly be heard by the other speaker (Ward & Tsukahara, 2000), it is more likely for those backchannels to not be uttered as other-oriented signals. A more likely conclusion would be that listeners produce those feedback tokens for themselves to split up utterances for better comprehension and transfer into memory. Consequently, these tokens would not serve the function of updating nor maintaining common ground.

Yet, affective backchannels, as produced in higher amounts in Spontaneous as compared to Task-based conversations (see fig. 9, Chapter 4.3), were longer in duration (table 2, Chapter 4.3). Additionally, they usually carried more lexical load as compared to most functional tokens and were therefore more likely to be actively perceived by the speaking interlocutor. Affective tokens have also been shown to contain more context-specific lexical content, which argues that they not only *claim* to signal attention, but actually signal attention. Following the idea that affective tokens manage the conversation on a content level, and influence social relations between subjects, one could speculate about different levels of common ground. The first level for which using functional backchannels would be beneficial, would be the level of content information, while on a second level social information might be updated using i.a. affective backchannels.

In summary, absolute certainty about the current status of common ground seems to be more important in Task-based conversation as compared to Spontaneous dialogue. Subjects therefore seem to follow different communicative rules for different contexts which are only in part adapted to the other interlocutor. However, the higher number of backchannels uttered in Task-Oriented conversation might not necessarily be analogue to a higher necessity of content-wise common ground updates, but listeners might exaggerate their backchannel use to guarantee precision in their communication. In Spontaneous speech on the other hand, backchannels might serve different functions. In this communicative context, backchannels might mainly serve social and affective functions and might be used less to manage common ground on a content level.

5.4. Limitations and Outlook

At the beginning of the results section, it was already stated that this thesis does not aim at drawing a complete picture of how feedback signals work in Task-Oriented and Spontaneous conversation. Rather, it was my intention to record conversations as natural as possible and to gain insight into possible features of backchannels in real-life conversations that might be missed when investigating backchannels purely in Task-Oriented conversation contexts. This course of action allowed deep and detailed exploration of the data, but also implicates some limitations.

Overall, one must admit that the sample size of seven (Spontaneous), or rather five (Maptask) dyads in itself allows for only limited generalisability of the results. In the following I will elaborate on three limitations of the study namely 1) the familiarity aspect, 2) the task, and 3) the methods chosen to analyse the data, before providing a brief outlook on future directions.

Due to the ongoing pandemic and the concomitant regulations in Germany, all subjects had to be living in the same households at the time of the recordings. I chose to only record subjects who had lived in a household together for at least six months prior to the recording and were not in a romantic relationship. However, the time that participants had lived together ranged from six months up to six years. The fact that in all conversations, the interlocutors 1) knew each other, and 2) some were very close friends who had spent years of their lives together as opposed to some who had only met some months before the recording might have had an influence on their conversational behaviour, especially regarding common ground management. Controlling for familiarity and somehow quantifying it e.g. using questionnaires or recording a more homogeneous group of subjects would potentially allow for more fruitful, and more generalisable conclusions about the data.

Second, the conversation contexts themselves allow for some improvements. The Maptask paradigm produces a very specific kind of Task-Oriented speech and was designed i.a. to elicit many continuation signals (backchannels like *mmhm*). Many studies investigating backchannel behaviour based on Task-Oriented conversation use the Maptask paradigm. To compare my results to findings from these studies, choosing the same task was therefore necessary. However, in my study, the subjects' use of backchannel tokens and intonation contours differed profoundly between the Maptask and the Spontaneous conversations. Considering the design of the Maptask and the lack of visual contact in the Task-Oriented setting of the study, it is possible to speculate that the differences found between the two conditions are also very specific to the task and cannot be generalised to all kinds of Task-Oriented dialogue. It is therefore my intention to further investigate feedback signals in Task-Oriented as compared to Spontaneous speech using a less specific task in the future while controlling for visual contact in all conditions.

Third, the chosen methods to analyse the data have their own flaws. It was already mentioned in the results section that the proportions of speech and silence should be factored in when interpreting backchannel rate. For the analysis of intonation contours I used the semitone analysis described in the methods section. While the method is robust when investigating short, monosyllabic tokens, and can give indications of rising, falling, or flat intonation contours, it fails to grasp the complexity of contours carried by multi-syllabic tokens. It was shown in this thesis that backchannel utterances in Spontaneous conversation were more variable in duration and therefore more likely to contain more than one syllable. For a meaningful analysis of the tokens, an analysis method which takes into account the whole contour instead of simple time points, like the measures provided by the *ProPer* toolbox (Albert et al., 2022), should be considered. Furthermore, considering that F0 rises are perceived as more prominent and listeners are more sensitive to them as compared to falls, different thresholds for rises and falls should be considered for the categorisation of contours.

For an in-depth analysis of backchannel functions in the two conversational contexts, coding these functions is vital. This coding poses considerable difficulties, especially in working with Spontaneous, natural where speakers behave even more unpredictably than in a laboratory setting. To simplify the analysis, I chose to only use passive recipiency tokens for analysis. Therefore, coding passive recipiency and incipient speakership would have made no sense. It does make sense, however, to ascribe more fine-grained functions like affirmation and excitement to the tokens to gain an insight into what backchannels are required to bring across in different conversational settings. In my analysis I chose to categorise tokens into affective and functional tokens to account for the variety of lexical tokens with different pragmatic functions as found in Spontaneous conversations. However, for a more robust analysis this categorisation should be even more fine-grained and supported by strict guidelines specifically adapted to Task-Oriented and Spontaneous conversation contexts.

5.4.1. Outlook

Of course, there are many more possible courses of action to be considered in the research of backchannel behaviour in different conversational contexts. As early as 1982, Schegloff, for example, addressed a potential relationship between the placement of backchannel utterances and gaze-behaviour of interlocutors in a conversation (investigated later by i.a. Truong et al., 2011). It was shown in this thesis, that backchannels differ in their rate depending on the context of a conversation but not necessarily depending on the presence or absence of visual contact between interlocutors. Concerning other features of backchannels, such as intonation and lexical choice, the relationship between modulations of these features and the gaze-behaviour of conversational partners is less clear. Many definitions of backchannels include not only vocalic signals, but also gestural signals like head-nods or even smiles (e.g. Bertrand et al., 2007; Trouvain, 2014). Although it is disputed whether these signals follow the same conversational rules as vocalic backchannels, it is clear that those gestural feedback-signals too, contribute to the organisation of conversation. Especially when comparing conversational contexts with and without visual contact between interlocutors, the relationship between vocalic and non-vocalic feedback could be of interest.

To sum up, this thesis represents a first insight into backchannel behaviour in Maptask, and Spontaneous speech, but is limited by i.a. the small group of subjects, variability in subject familiarity, as well as the specificity of the task and some aspects of the analyses themselves. However, it also shows the great potential for future studies with bigger and more homogeneous subject groups, studies of backchannel behaviour in different conversational contexts and their connection to gaze-behaviour and head-movements

6. Conclusion

In this thesis, backchannel feedback was investigated in Task-based as compared to Spontaneous conversation. I demonstrated structural differences in backchannel frequency, the use of lexical tokens, as well as intonation contour in continuous and categorical terms, such as proportions of rising, falling and level intonation contours. It was found that participants in the study at hand used different sets of behaviours or communicative conventions in the two conversational contexts. In Task-Oriented conversations a higher rate of backchannels per minute was used as compared to Spontaneous conversations. The lack of visual contact between the two interlocutors in a dialogue did not seem to be an eminent factor regarding backchannel rate.

It was also found that interlocutors used more backchannels with a clearly rising or clearly falling intonation contour, as well as steeper pitch movements in semitones in Task-Oriented conversations. Investigations of differences between the two speakers of a dyad revealed a clear distinction between the feedback of instruction followers and instruction givers in the Maptask. In Spontaneous dialogue on the other hand, the majority of backchannels had a flat or falling intonation contour, making backchannels in Task-based conversations overall more salient. It is unclear whether the effect of more salient intonation contours was only elicited by the context and how much the lack of visual contact between interlocutors influenced their intonational realisations. Regarding differences between the individual speakers of a dyad, no clear distinction in their behaviour was found, but rather a slight tendency towards one speaker taking the lead and the other one following. Still, in Spontaneous conversation overall, the lexical content of the token might have been more important as it was shown by higher proportions of unusual lexical tokens.

Concerning lexical choice of backchannel tokens, it was shown that participants used different sets of backchannels depending on the context, the goal, and therefore the requirements as well as the contents of the conversation. While listeners used predominantly functional backchannels in Task-based conversations, they used much higher proportions of affective tokens in Spontaneous conversations. Therefore, one can speculate that feedback in Task-based dialogue has a stronger functional focus while in Spontaneous conversations there is a slight tendency towards more social and affective functions.

The study presented here is limited by multiple factors, such as the small number of participants and the differences in visual contact between the two main conditions. Still, it does give a variety of thought-provoking insights into differences between BCs in different conversational settings and questions the generalisability of findings on Task-based conversations for all conversational contexts.

7. References

Albert, A., Cangemi, F., Ellison, M., & Grice, M. (2022). *ProPer: PROsodic analysis with PERiodic energy.* https://osf. io/28ea5/

Allwood, J., Nivre, J., & Ahlsen, E. (1992). On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics*, *9*, 1–26. https://academic.oup.com/jos/article/9/1/1/1659875

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

Baayen, R.H., Davidson, D.J. & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bangerter, A. & Clark, H.H (2003). Navigating joint projects with dialogue. In *Cognitive Science* (Vol. 27), pp. 195-225

Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, *28*(6), pp.937–962. https://doi.org/10.1016/j.cogsci.2004.07.002

Benus, S., Gravano, A. & Hirschberg, J.B. (2007). *The Prosody of Backchannels in American English.* Proceedings of the 16th International Congress of phonetic Sciences, pp. 1065-1068

Bertrand, R., Ferré, G., Blache, P., Espesser, R. & Rauzy, S.. (2007). Backchannels revisited from a multimodal perspective. *Auditory-Visual Speech Processing*, pp. 1–5.

Boersma, P. & Weenink, D. (2021). *Praat: doing phonetics by computer* [computer program] (2011). Version, 5(3), 74.

Bolinger, D. (1989). Intonation and its uses: Melody in grammar and discourse. *Stanford University Press.*

Bürkner, P. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), pp. 1-28.

Caspers, J. (2000). Melodic characteristics of backchannels in Dutch Map Task dialogues. *Proceedings of the 6th International Conference on Spoken Language Processing* , pp. 611–614.

Castello, E. & Gesuato, S. (2019). Holding up one's end of the conversation in spoken English: Lexical backchannels in L2 examination discourse. *International Journal of Learner Corpus Research*, *5*(2), pp. 231–252.

Cathcart, N., Carletta, J. & Klein, E. (2003). A Shallow Model of Backchannel Continuers in Spoken Dialogue. *European ACL* , pp.51–58.

Clark, H.H. (1991). *Grounding in communication*. In J. M. L. & S. D. T. L. B. Resnick, (Eds.). American psychological association, pp. 127-149.

Clark, H.H. & Carlson, T.B. (1981) Context for Comprehension, *Attention and Performance IX*, *313*(30).

Clark, H.H. & Krych, M.A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), pp. 62–81. https://doi.org/10.1016/j.jml.2003.08.004

Clark, H.H. & Marshall, C.R. (1981). *Definite reference and mutual knowledge*. Psycholinguistics: critical concepts in psychology, pp. 414

Clark, H.H. & Schaefer, E.F.. (1989). Contributing to Discourse. *Cognitive Science*, *13*, pp. 259–294.

Cutrone, P. (2014). A cross-cultural examination of the backchannel behavior of Japanese and Americans: Considerations for Japanese EFL learners. *Intercultural Pragmatics*, *11*(1), pp. 83–120. https://doi.org/10.1515/ip-2014-0004

Dale, R., Fusaroli, R., Duran N.D., & Richardson, D.C. (2013). The Self-Organization of human interaction. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 59, pp. 43–95). Academic Press Inc. https://doi.org/10.1016/B978-0-12-407187-2.00002-2

Dideriksen, C., Christiansen, M.H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2022). Quantifying The Interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General*. Doi: 10.1037/xge001301

Dunbar, R.M., Marriott, A. & Duncan, N.D.C. (1997). *Human Conversational Behavior*. Human Nature, 8, pp. 231-146.

Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K.. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*(1), pp. 147–157. https://doi.org/10.1016/j.newideapsych.2013.03.005

Fusaroli R., Tylén, K., Garly, K., Steensig, J. & Christiansen, M.H. (2017). Measures and mechanisms of common ground: backchannels, conversational repair, and interactive alignment in free and task-oriented social interactions. In G. Gunzelmann, A. Howes, & T. Tenbrink (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, pp. 2055–2060. http://hdl.handle.net/2066/197816

Gardner. (2001). *When listeners talk.* Amsterdam [u.a.]: Benjamins.

Gibbon, D., Stocksmeier, T., & Kopp, S. (2007). Synthesis of prosodic attitudinal variants in German backchannel ja. In *Interspeech 2007*, pp. 1290–1293.

Giles, H., Coupland, N., & Coupland, J. (2010). Accommodation theory: Communication, context, and consequence. In *Contexts of Accommodation*, pp. 1–68. Cambridge University Press. https://doi.org/10.1017/cbo9780511663673.001

Gilmartin, E., Cowan, B. R., Vogel, C., & Campbell, N. (2018). Explorations in multiparty casual social talk and its relevance for social human machine dialogue. *Journal on Multimodal User Interfaces*, *12*(4), pp. 297–308. https://doi.org/10.1007/s12193-018-0274-2

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, *37*(1), pp. 3–19. https://doi.org/10.1177/0093650209351468

Ha, K. P., Ebner, S., & Grice, M. (2016). Speech prosody and possible misunderstandings in intercultural talk – a study of listener behaviour in standard Vietnamese and German dialogues. *Proceedings of the International Conference on Speech Prosody*, *2016-January*, pp. 801–805. https://doi.org/10.21437/speechprosody.2016-164

Heinz, B. (2003). Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics*, *35*(7), pp. 1113–1142. https://doi.org/10.1016/S0378-2166(02)00190-X

Heldner, M., Edlund, J., & Hirschberg, J. (2010). Pitch similarity in the vicinity of backchannels. In *Interspeech 2010*. https://doi.org/https://doi.org/10.7916/D8WS92R4

Himberg, T., Hirvenkari, L., Mandel, A., & Hari, R. (2015). Word-by-word entrainment of speech rhythm during joint story building. *Frontiers in Psychology*, *6*(JUN), 797. https://doi.org/10.3389/fpsyg.2015.00797

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? In *Cognition* (Vol. 59), pp. 91-117.

Hsu, C. H., Evans, J. P., & Lee, C. Y. (2015). Brain responses to spoken F0 changes: Is H special? *Journal of Phonetics*, *51*, pp. 82–92. https://doi.org/10.1016/j.wocn.2015.02.003

Itakura, H. (2001). Describing conversational dominance. In *Journal of Pragmatics*, *33*, pp. 1859–1880.

Janz, A., Wehrle, S., & Sbranna, S. (2022). *Making conversation work: Prominence in the intonation of feedback signals* [Conference Talk].

Kawahara, T., Yamaguchi, T., Inoue, K., Takanashi, K., & Ward, N. (2016). Prediction and generation of backchannel form for attentive listening systems. *Proceedings of the Annual Conference of the International Speech Communication Association,* In *Interspeech 2016*, pp. 2890–2894. https://doi.org/10.21437/Interspeech.2016-118

Keysar, B., Barr, D., Balin, J. A., & Paek, T. S. (1998). Definite Reference and Mutual Knowledge: Process Models of Common Ground in Comprehension. *Journal of Memory and Language*, *39*(1), pp. 1–20.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking Pespektive in Conversation: The Role of Mutual Knowledge in Comprehension. *Psychological Science*, *11*(1), pp. 32–38.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), pp. 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A. & Den, Y. (1998). An Analysis of Turn-Taking and Backchannels Based on Prosodic and Syntactic Features in Japanese Map Task Dialogs. In *Language and speech,* Pp 4-13.

Krauss, R. M., Garlock, C. M., Bricker, P. D., & Mcmahon, L. E. (1977). The Role of Audible and Visible Back-Channel Responses in Interpersonal Communication. In *Journal of Personality and Social Psychology* (Vol. 35, Issue 7), pp. 523-529.

Lebra T. S. (1976). *Japanese patterns of behavior*. University of Hawaii Press.

Li, H. Z. (2006). Backchannel Responses as Misleading Feedback in Intercultural Discourse. *Journal of Intercultural Communication Research*, *35*(2), pp. 99–116. https://doi.org/10.1080/17475750600909253

Liesenfeld, A., & Dingemanse, M. (2022). *Bottom-up discovery of structure and variation in response tokens ('backchannels') across diverse languages*. In *Interspeech 2022,* pp. 1126-1130. https://doi.org/10.31234/osf.io/w8hpy

Nakamura, S., Phung, L., & Reinders, H. (2021). The effect of learner choice on L2 task engagement. *Studies in Second Language Acquisition*, *43*(2), pp. 428–431. https://doi.org/10.1017/S027226312000042X

Orestrom, B. (1983). *Turn-taking in English conversation,* lund studies in English (66), Malmo, Sweden.

Pammi, S., & Schröder, M. (2009). *A corpus based analysis of backchannel vocalizations,* In *Proceedings of the Interdisciplinary Workshop on Laughter and other Interactional Vocalisations in Speech*. Berlin, Germany.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), pp. 169–190. https://doi.org/10.1017/s0140525x04000056

R Core Team (2022) *R: A language and environment for statistical computing*, Vienna, Austria.

R Core Team. (2019). *RStudio*.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), pp. 696–735. https://doi.org/10.1353/lan.1974.0010

Savino, M. (2010). Intonational strategies for backchanneling in Italian Map Task dialogues, In *Third ISCA Workshop on experimental linguistics,* pp. 25-27.

Savino, M. (2011). The intonation of backchannel tokens in Italian collaborative dialogues, In *Proc. Language and Technology Conference*, pp. 28-39.

Sbranna, S., Möking, E., Wehrle, S., & Grice, M. (2022). Backchannelling across Languages: Rate, Lexical Choice and Intonation in L1 Italian, L1 German and L2 German. *Proc. Speech Prosody*, pp. 734–738.

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*, pp. 71–93.

Scott, S., & Sauter, D. (2006). Non-verbal expressions of emotion-acoustics, valence and cross cultural factors. *Third International Conference on Speech Prosody* .

Trouvain, J. (2014). Laughing, Breathing, Clicking-The Prosody of Nonverbal Vocalisations. *Proceedings of Speech Prosody (Vol. 7)*, pp. 598–602.

Truong, K. P., Poppe, R., de Kok, I., & Heylen, D. (2011). A multimodal analysis of vocal and visual backchannels in spontaneous dialogs. *Interspeech 2011*, pp. 2973–2976.

Ward, N. (2004). Pragmatic Functions of Prosodic Features in Non-Lexical Utterances. *Speech Prosody 2004*.

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

Wehrle, S. (2021). *A Multi-Dimensional Analysis of Conversation and Intonation in Autism Spectrum Disorder* [Doctoral Dissertation]. University of Cologne.

Wehrle, S., Roettger, T. B., & Grice, M. (2018). Exploring the Dynamics of Backchannel Interpretation: The Meandering Mouse Paradigm. In *ProsLang: Workshop on the processing of prosody across languages and varieties*.

Wehrle, S., & Grice, M. (2019). Function and Prosodic Form of Backchannels in L1 and L2 German [Poster Presentation], *Hanyang International Symposium on Phonetics and Cognitive Science of Language.*

White, S. (1986). *Functions of backchannels in English: A cross-cultural analysis of Americans and Japanese* [Doctoral Dissertation].

Wickham et al. (2019). welcome to the tidyverse, In *Journal of open source software,* 4(43), pp. 1686.

Wickham, H. (2016). Programming with ggplot2. *ggplot2: Elegant Graphics for Data Analysis*, 241-253.

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: A Grammar of Data Manipulation*.

Yngve, V. H. (1979). On getting a word in edgewise. *Chicago Linguistics Society*, *6th Meeting*, pp. 567–578.

8. List of Examples, Figures and Tables

List of figures:

List of examples:

List of tables:

9. Appendix

Data Tables

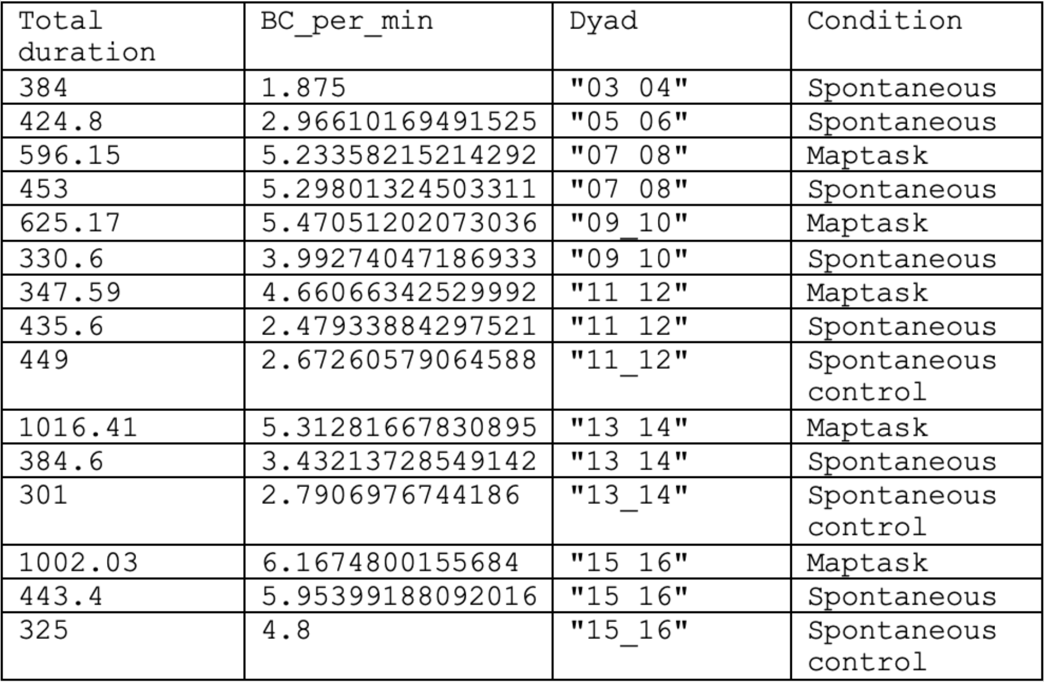
Table A1: *Mean values and standard deviations of pitch excursion in semitones (not rounded) per Condition (Maptask, Spontaneous, and control condition in laboratory setting) for each individual BC type.*

	Condition	Type	Mean	SD
1	Maptask	ah+other	-2.1	NA
2	Maptask	genau	-4.88952380952381	3.56
3	Maptask	hm	5.984	7.04
4	Maptask	ja	0.178490566037736	3.24
5	Maptask	mmhm	5.17910569105691	2.44
6	Maptask	okay	-2.95543859649123	3.88
7	Maptask	other	-2.17285714285714	3.21
8	Spont_lab	ah+other	-4.08666666666667	4.19
9	Spont_lab	hm	-4.14666666666667	1.95
10	Spont_lab	ja	-2.42833333333333	2.61
11	Spont_lab	ja+other	5.095	7.77
12	Spont_lab	mmhm	1.38769230769231	2.55
13	Spont_lab	okay	-0.83	18.37
14	Spont_lab	other	-1.54	4.44
15	Spontaneous	ah+other	-2.9275	4.05
16	Spontaneous	genau	-3.015	1.07
17	Spontaneous	hm	-0.471666666666667	6.10
18	Spontaneous	ja	-1.10717391304348	1.95
19	Spontaneous	ja+other	-0.125833333333333	2.41
20	Spontaneous	mmhm	1.14232258064516	2.51
21	Spontaneous	okay	1.03	1.23
22	Spontaneous	other	-2.28384615384615	1.93

Table A2: *Proportions of tokens with rising, falling, and level contours in the two main conditions Maptask and spontaneous, as well as the spontaneous control condition (Spont_control)*

	Condition	Contour	Percentage
1	Maptask	Fall	34.66
2	Maptask	Level	12.58
3	Maptask	Rise	52.76
4	Spont_control	Fall	52.38
5	Spont_control	Level	21.43
6	Spont_control	Rise	26.19
7	Spontaneous	Fall	45.08
8	Spontaneous	Level	34.43
9	Spontaneous	Rise	20.49

Table A3: *Total conversation duration in seconds and backchannel rates by dyad for all three conditions*



Total duration	BC_per_min	Dyad	Condition
384	1.875	"03 04"	Spontaneous
424.8	2.96610169491525	"05 06"	Spontaneous
596.15	5.23358215214292	"07 08"	Maptask
453	5.29801324503311	"07 08"	Spontaneous
625.17	5.47051202073036	"09_10"	Maptask
330.6	3.99274047186933	"09 10"	Spontaneous
347.59	4.66066342529992	"11 12"	Maptask
435.6	2.47933884297521	"11 12"	Spontaneous
449	2.67260579064588	"11_12"	Spontaneous control
1016.41	5.31281667830895	"13 14"	Maptask
384.6	3.43213728549142	"13 14"	Spontaneous
301	2.7906976744186	"13_14"	Spontaneous control
1002.03	6.1674800155684	"15 16"	Maptask
443.4	5.95399188092016	"15 16"	Spontaneous
325	4.8	"15_16"	Spontaneous control

Table A4: *All lexical and non-lexical tokens used in either of the three conversational contexts.*

Lexical tokens			
ja	ja ja	oh ja	ja ja safe so
ja voll	ja stimmt	ja krass	ja chillig
ja ja schon klar	ja genau	ja wie bei mir	ja ja voll
ja voll gut	ja ja okay	ja nice	ja easy digga
ja das stimmt auch	ja safe	ja voll krass	achso
ah ja stimmt	ah ja okay	genau	okay
schön	klar	cool	nice
das stimmt	stimmt	ah krass	krank
junge	geil	genau ja	okay okay
ah mega	mega	havana pur	sehr gut
das wär geil	eben	ach crazy	ah ja
stimmt	ah okay		
Non-lexical tokens			
ah	aha	oh	mmhm
hm			

Supplementary Figures

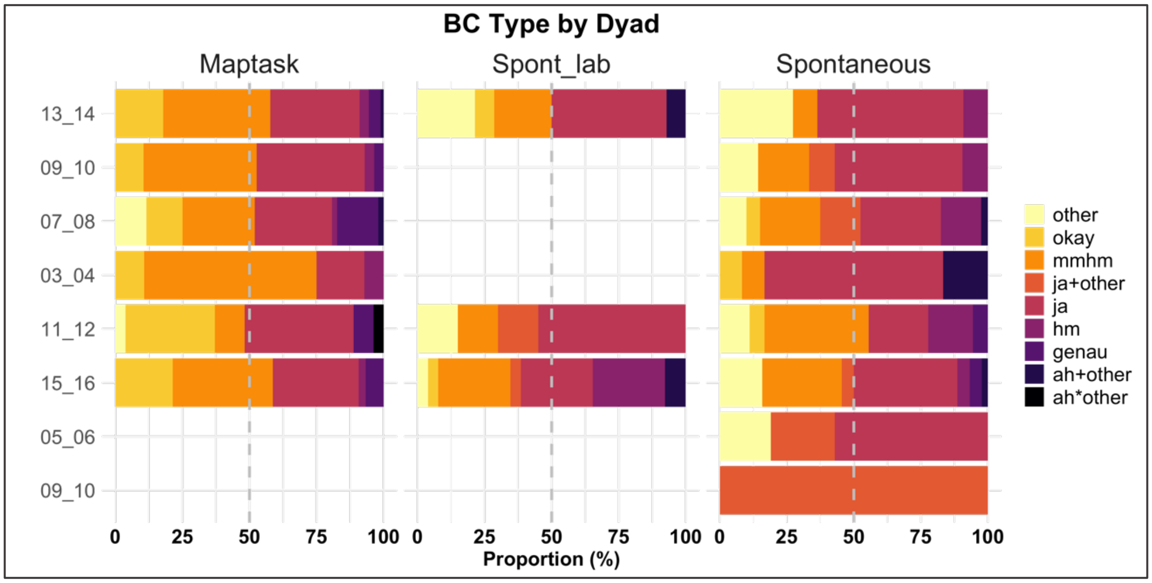


Figure A1: *Proportions of choice of backchannel type by dyad for all three conditions*

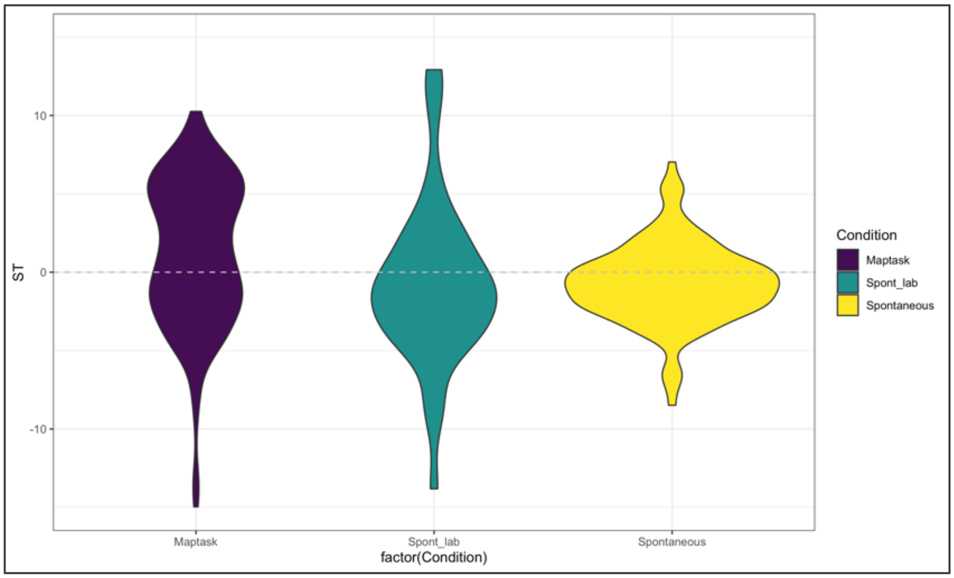


Figure A2: *Distribution of pitch movement in semitones for all three condition, positive values indicate rising contours, negative values indicate falling contours*

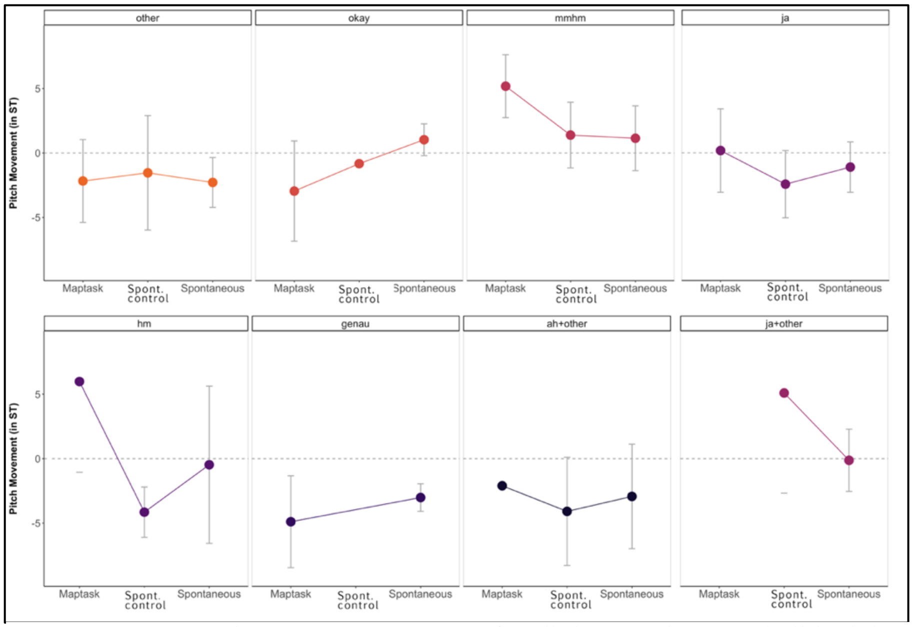


Figure A3: *Mean pitch movement in semitones for all three conditions and all backchannel types. rising contours are indicated by positive values, negative values indicate falling contours, standard deviations are represented by grey bars*

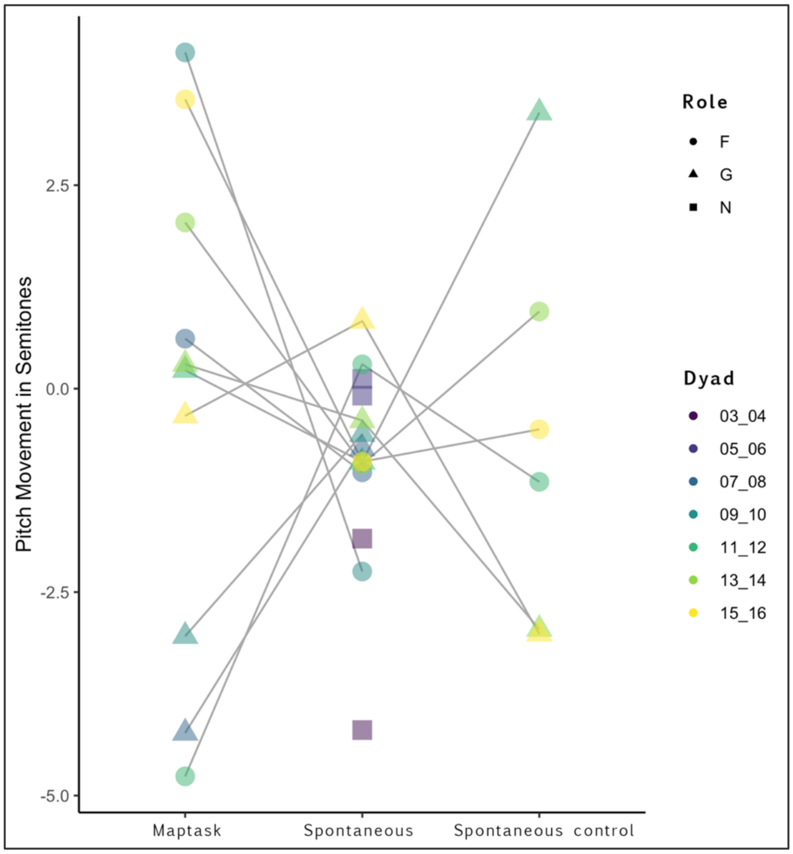


Figure A4: *Mean values for pitch movement in semitones for all speakers in all conditions. Speakers of the same dyad are displayed in the same colour, circles represent instruction followers (in the Maptask), triangles represent instruction givers. Speakers who only participated in the spontaneous recording are represented by squares.*

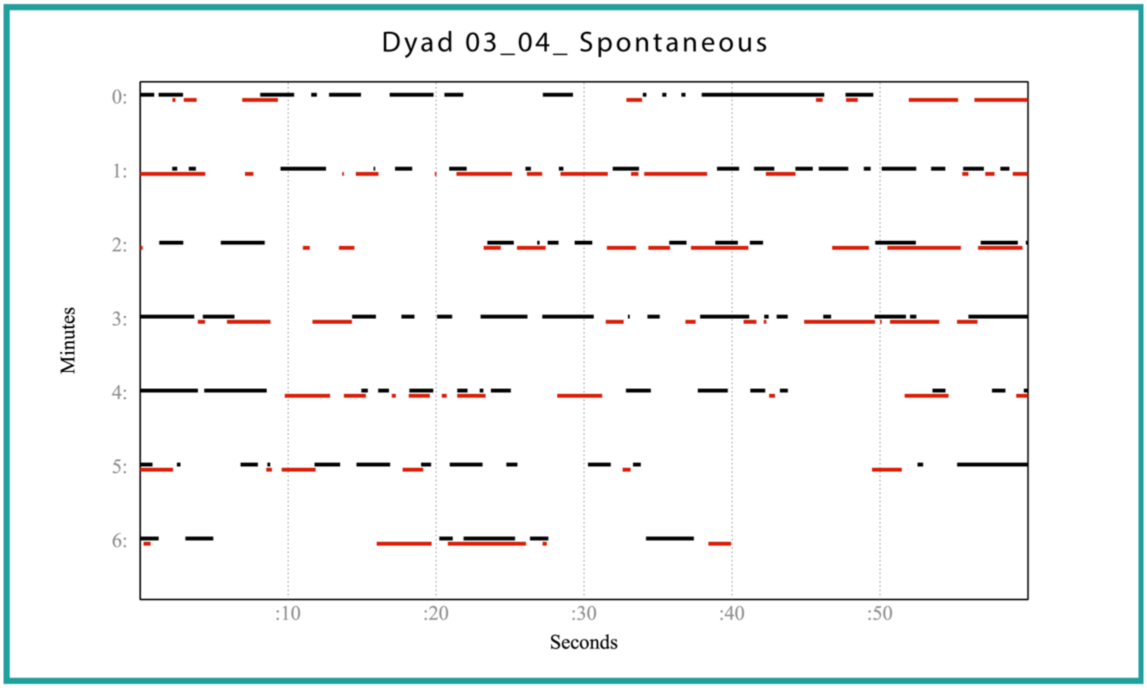


Figure A5: Overview plot for Dyad 03_04 in the Spontaneous condition

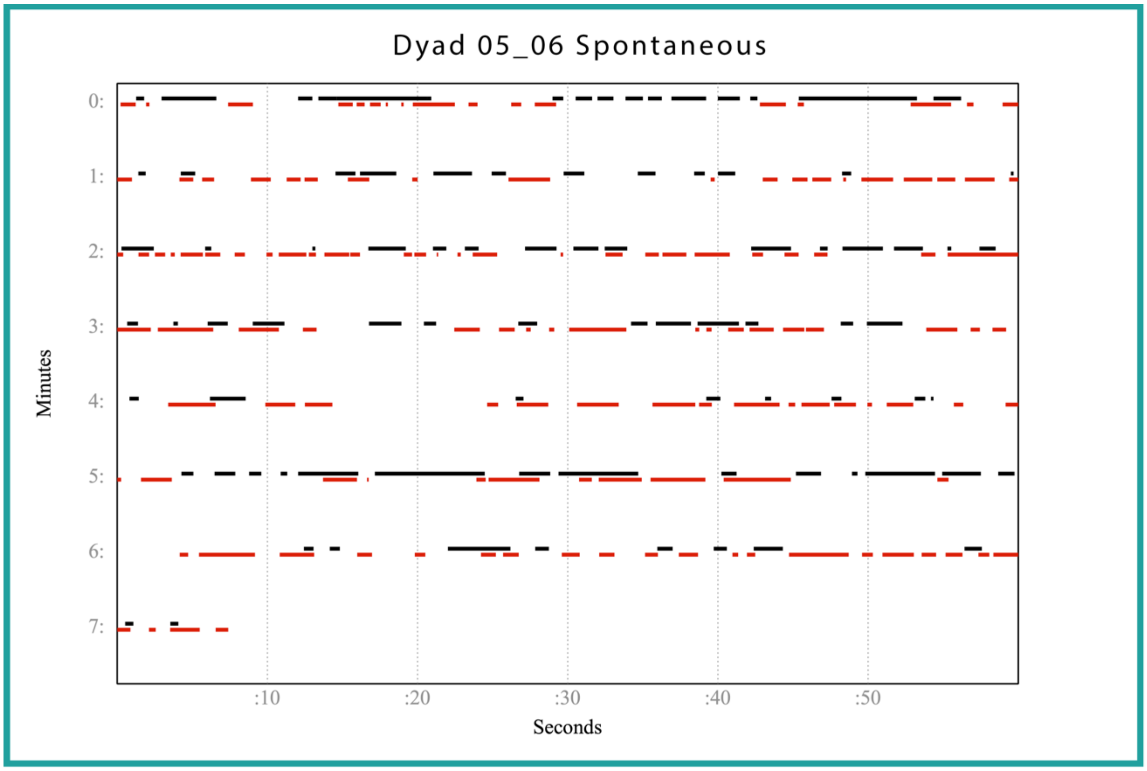


Figure A6: Overview plot for Dyad 05_06 in the Spontaneous condition

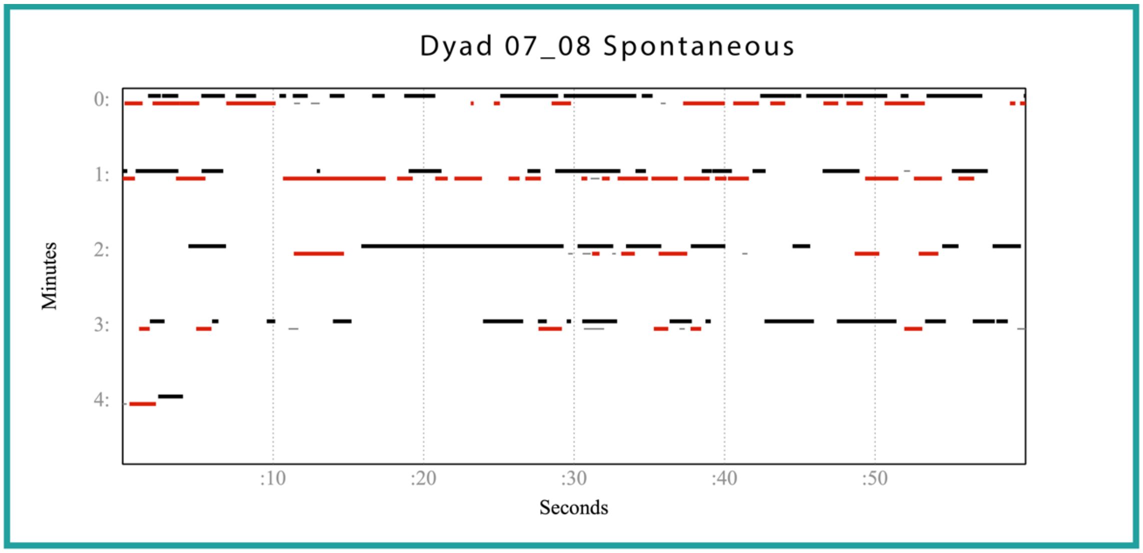


Figure A7: Overview plot for Dyad 07_08 in the Spontaneous condition

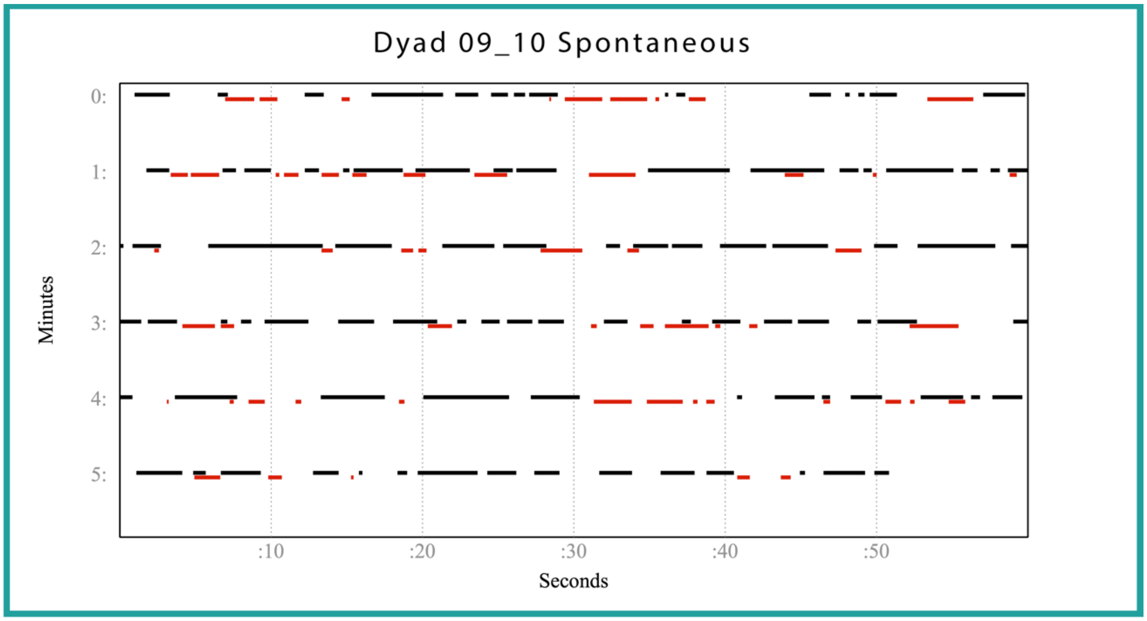


Figure A8: Overview plot for Dyad 09_10 in the Spontaneous condition

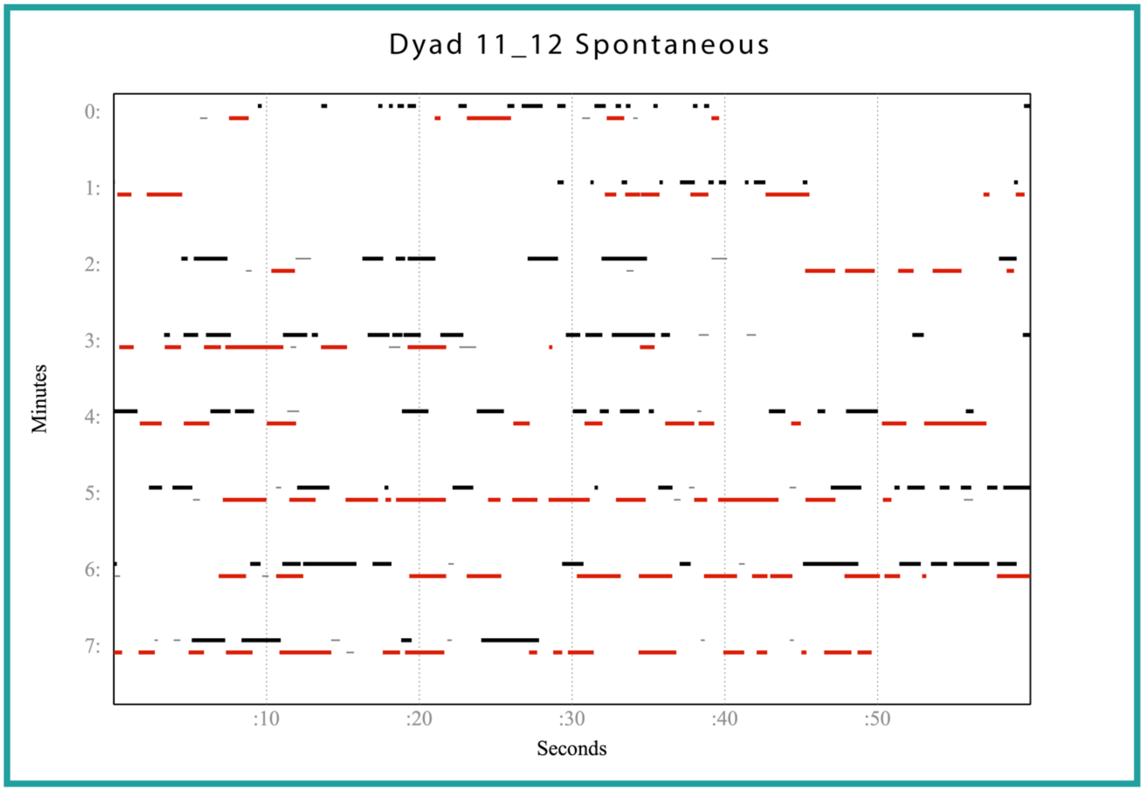


Figure A9: Overview plot for Dyad 11_12 in the Spontaneous condition

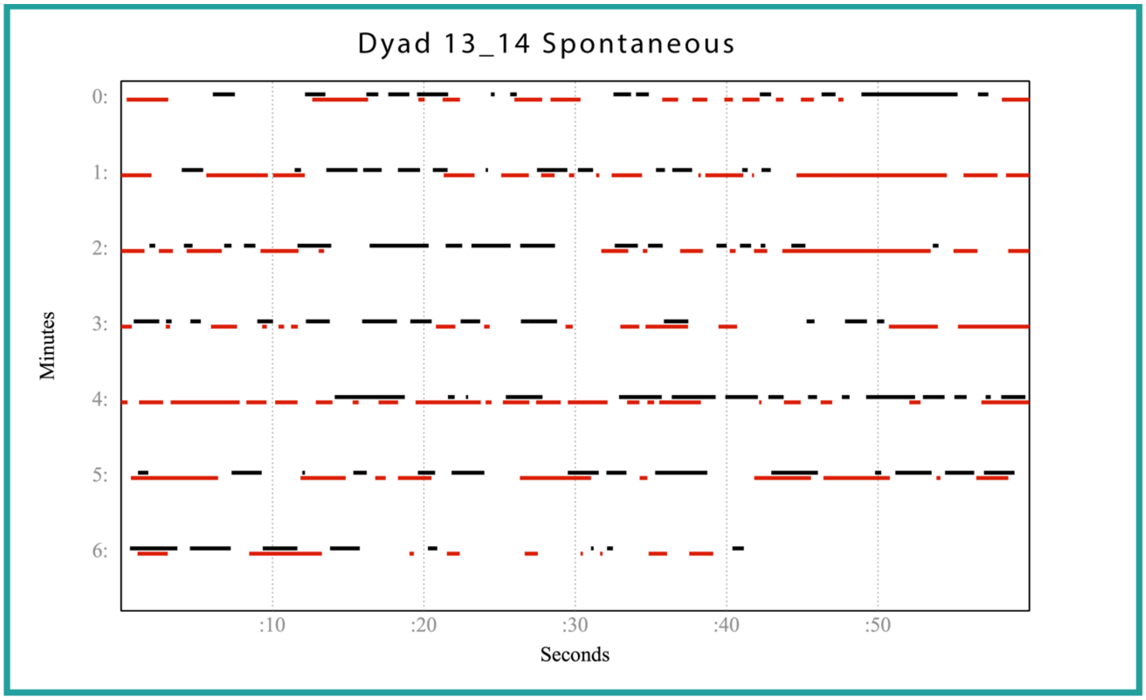


Figure A10: Overview plot for Dyad 13_14 in the Spontaneous condition

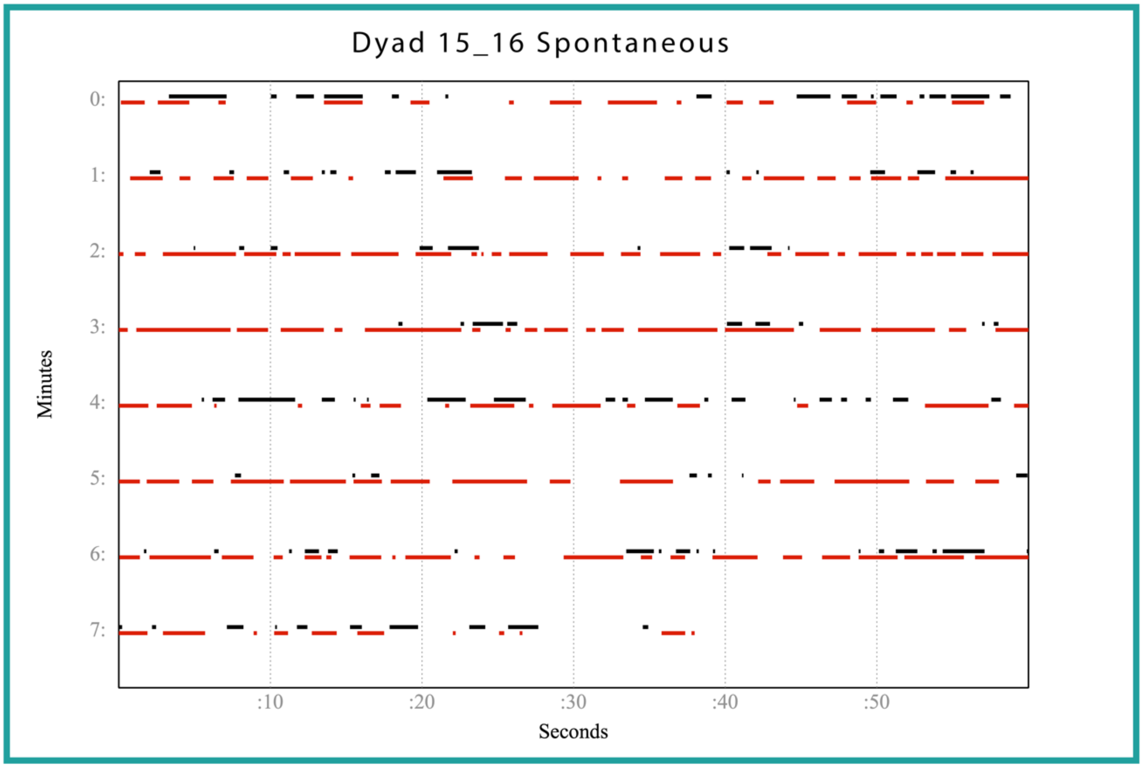


Figure A11: Overview plot for Dyad 15_16 in the Spontaneous condition

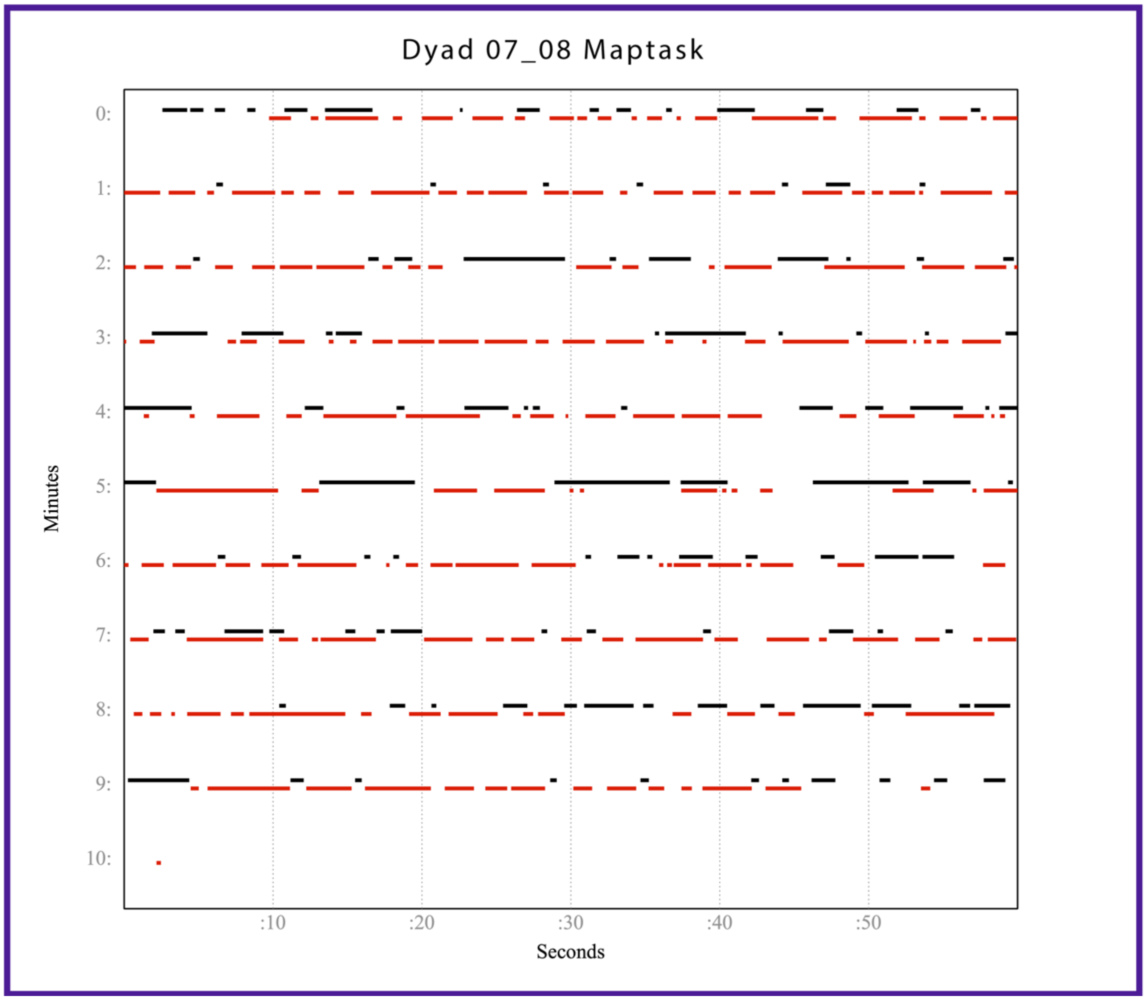


Figure A12: Overview plot for Dyad 07_08 in the Maptask condition

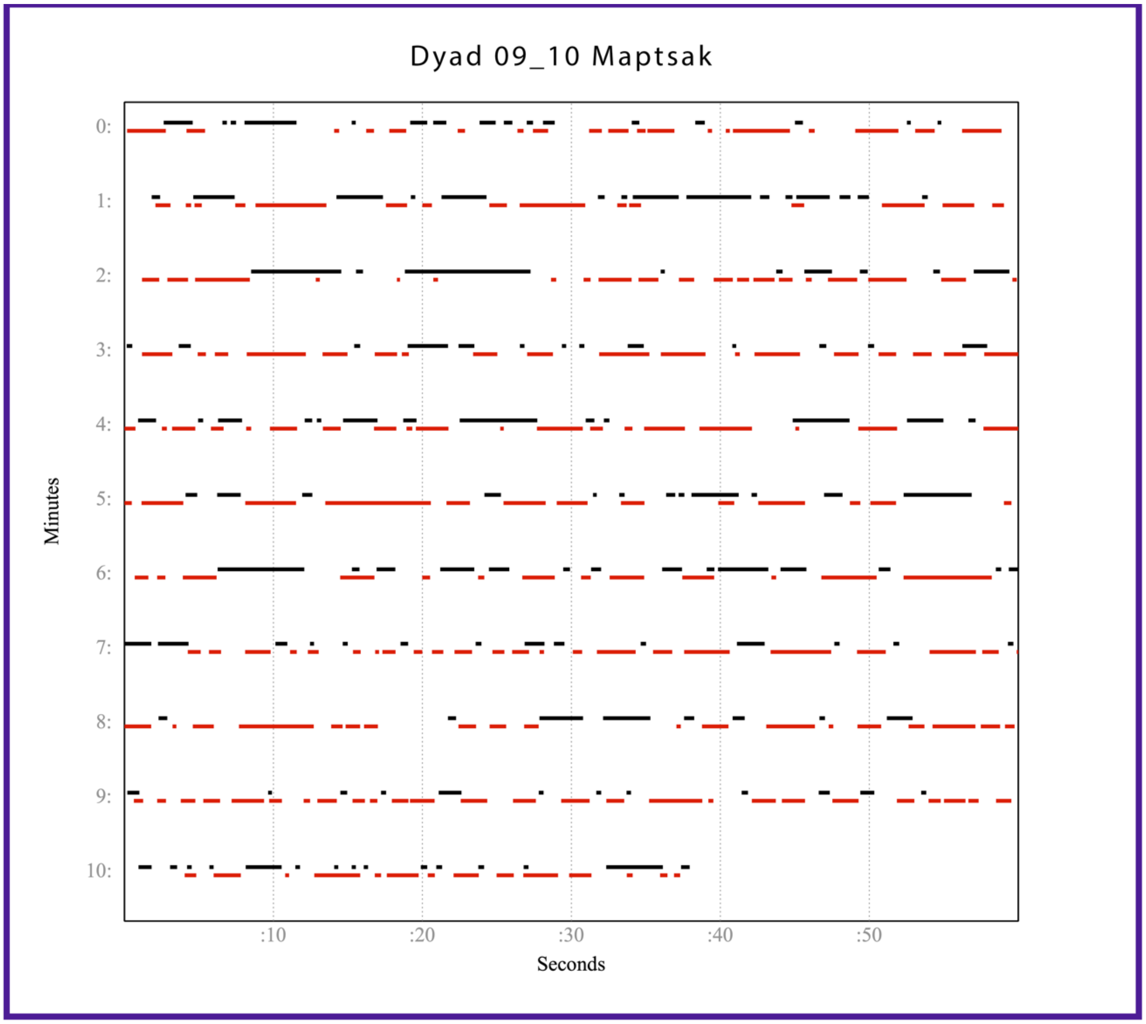


Figure A13: Overview plot for Dyad 09_10 in the Maptask condition

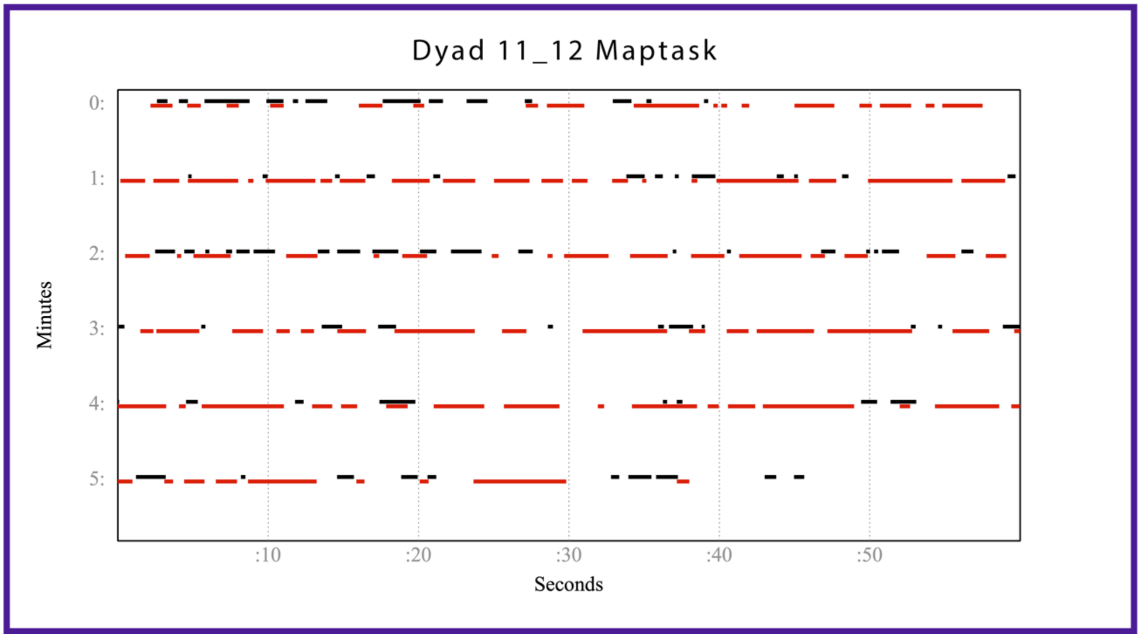


Figure A14: Overview plot for Dyad 11_12 in the Maptask condition

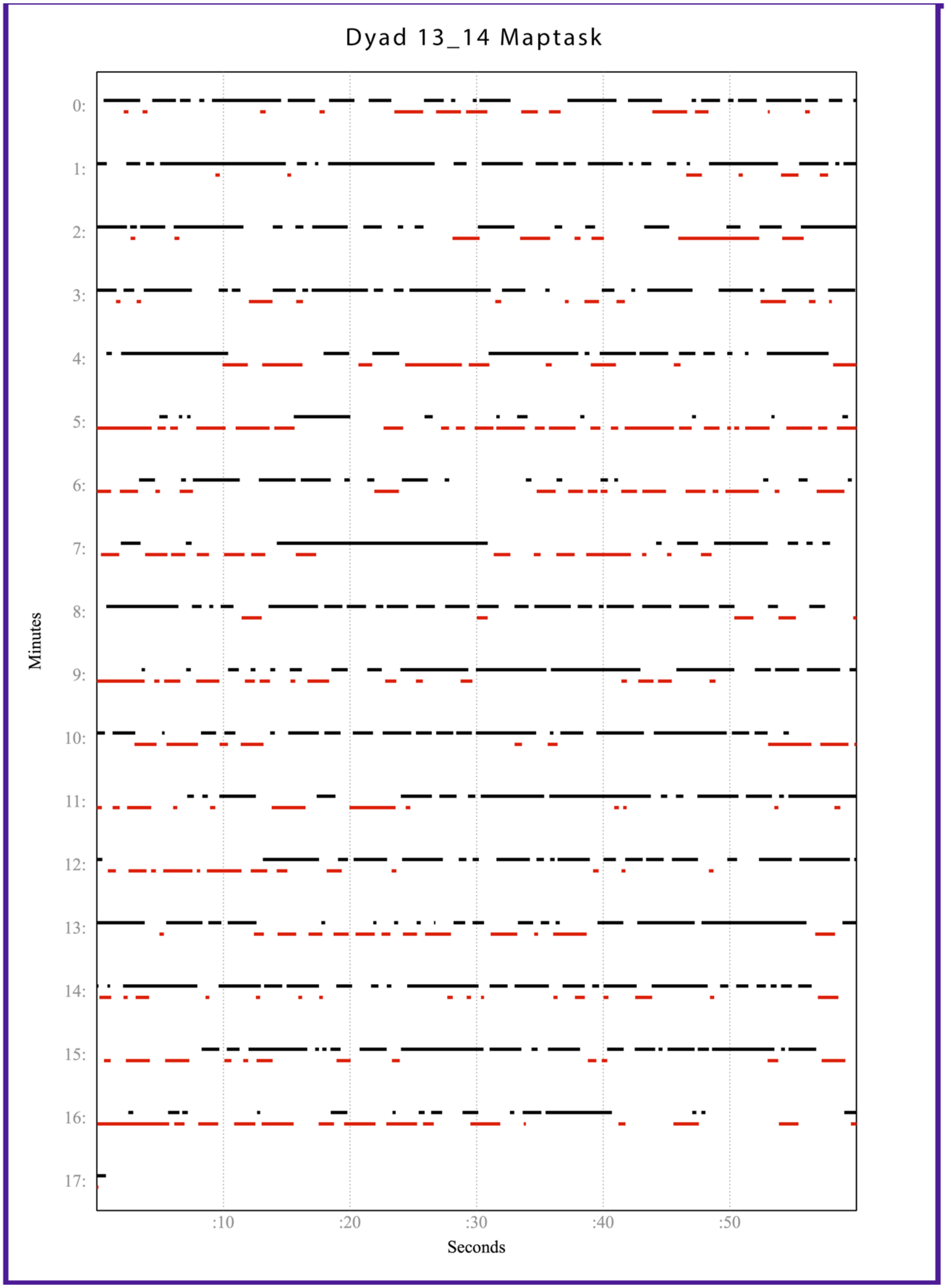


Figure A15: Overview plot for Dyad 13_14 in the Maptask condition

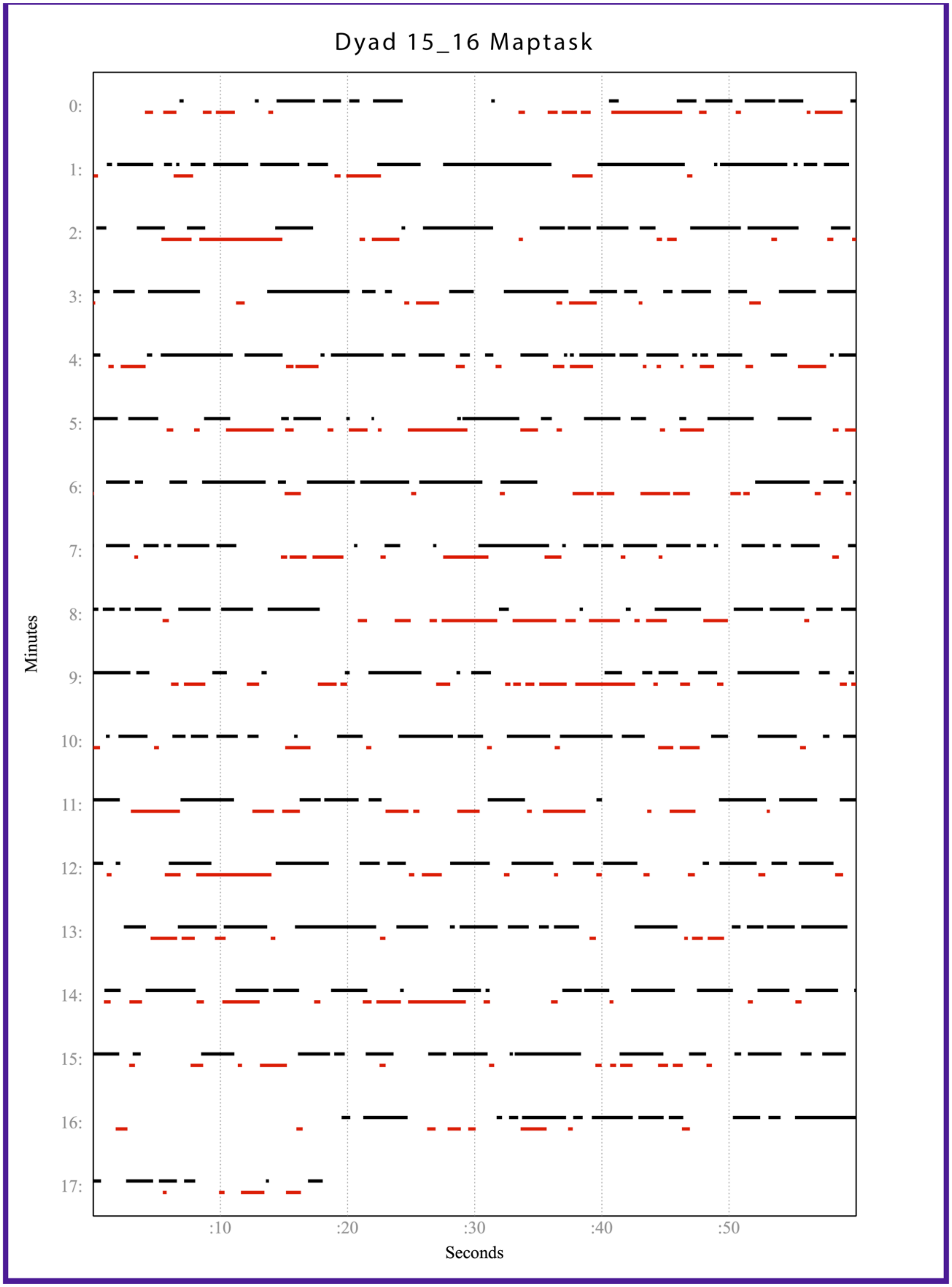
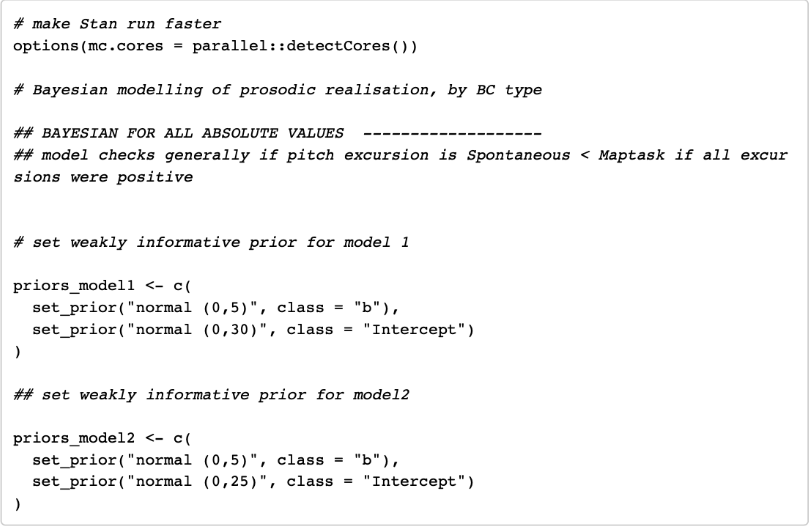
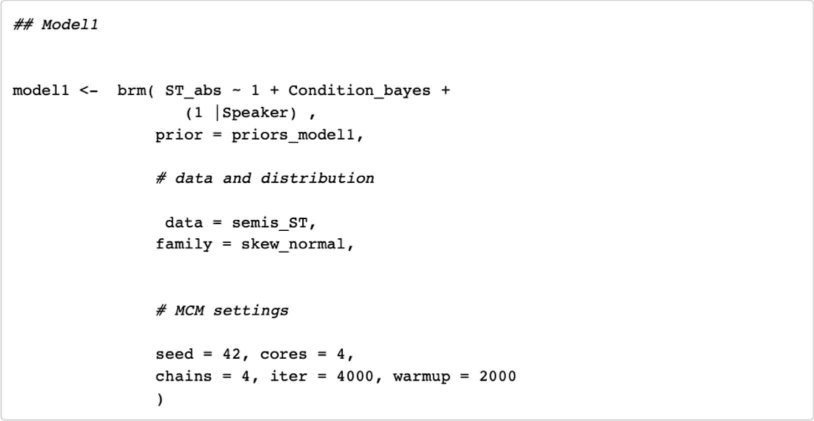


Figure A16: Overview plot for Dyad 15_16 in the Maptask condition

Bayesian Models



```
# make Stan run faster
options(mc.cores = parallel::detectCores())

# Bayesian modelling of prosodic realisation, by BC type

## BAYESIAN FOR ALL ABSOLUTE VALUES  -------------------
## model checks generally if pitch excursion is Spontaneous < Maptask if all excur
sions were positive


# set weakly informative prior for model 1

priors_model1 <- c(
  set_prior("normal (0,5)", class = "b"),
  set_prior("normal (0,30)", class = "Intercept")
)

## set weakly informative prior for model2

priors_model2 <- c(
  set_prior("normal (0,5)", class = "b"),
  set_prior("normal (0,25)", class = "Intercept")
)
```



```
## Model1


model1 <-  brm( ST_abs ~ 1 + Condition_bayes +
                   (1 |Speaker) ,
               prior = priors_model1,

               # data and distribution

                data = semis_ST,
               family = skew_normal,


               # MCM settings

               seed = 42, cores = 4,
               chains = 4, iter = 4000, warmup = 2000
               )
```

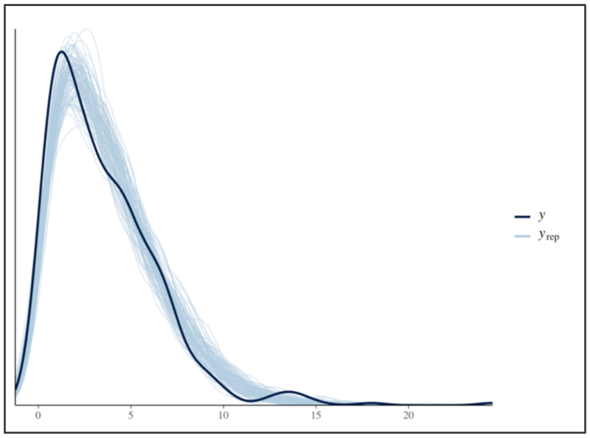
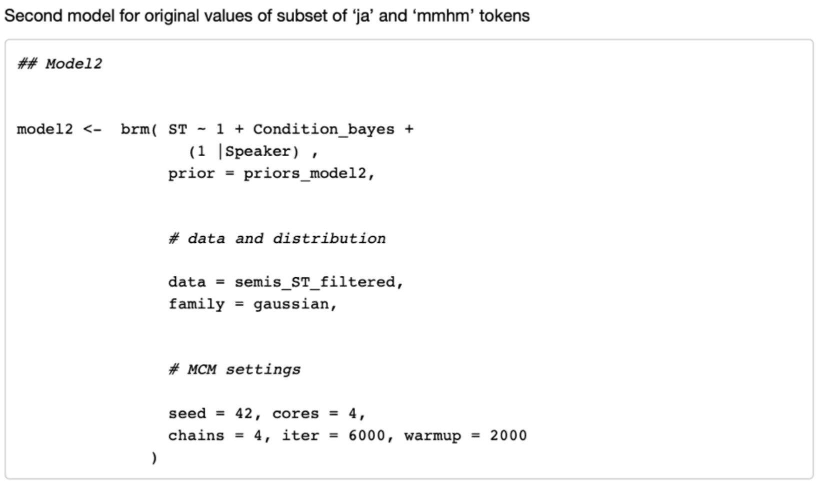


Figure A5: pp_check model 1 with 100 draws



Second model for original values of subset of ‘ja’ and ‘mmhm’ tokens

```
## Model2


model2 <-  brm( ST ~ 1 + Condition_bayes +
                  (1 |Speaker) ,
                prior = priors_model2,


                # data and distribution

                data = semis_ST_filtered,
                family = gaussian,


                # MCM settings

                seed = 42, cores = 4,
                chains = 4, iter = 6000, warmup = 2000
              )
```

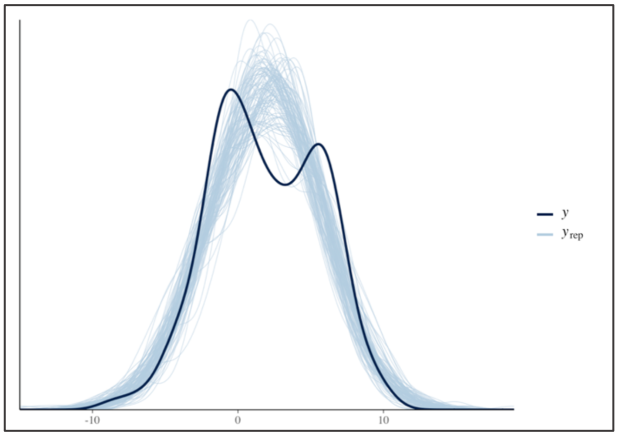


Figure A6: pp_check for model 2 with 110 draws

Task Description Golden Apple Game

Jeder Spieler bekommt ein Spielbrett mit 62 nummerierten Fenstern. Jedes Fenster kann geöffnet werden und zeigt ein Bild. Die Reihenfolge der Bilder ist anders für jeden Spieler und die Bilder unterscheiden sich in Form und Farbe.

Zwei Arten von Bildern sind besonders wichtig, um zu gewinnen: goldene Äpfel und Bomben. Der Spieler, der mehr goldene Äpfel findet, gewinnt – unter einer Voraussetzung: er hat die Reihenfolge der Bilder des Mitspielers korrekt aufgeschrieben. Aber Achtung vor den Bomben! Mit einer Bombe kann man einen goldenen Apfel des Mitspielers zerstören und damit ungültig machen.

Um zu wissen, wie viele goldene Äpfel du hast, nimm dir ein Bild mit einem goldenen Apfel jedes Mal, wenn du einen findest. Wenn dein Mitspieler aber eine Bombe hat, leg das Bild zurück. So kannst du einfach die Bilder am Ende des Spieles durchzählen, um zu wissen, wer gewonnen hat.

Um das Ziel zu erreichen, muss man dem Mitspieler die genaue Reihenfolge der Bilder mitteilen, wie im folgenden Beispiel:

> Du öffnest Fenster Nr. 1, siehst einen schwarzen Mond und fragst:
> Spieler A: *„Hast du einen schwarzen Mond?“*
> Der andere Spieler antwortet mit ja oder mit nein und sagt was er/sie in Fenster Nr. 1 hat:
> Spieler B: *„Ja, ich habe einen schwarzen Mond“* oder *„Nein, ich habe eine grüne Flagge“*

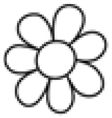
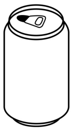
Am Ende dieses Zuges wisst ihr beide, was in der Tabelle des jeweils anderen steht und tragt es in eure leere Tabelle ein.

Im nächsten Zug stellt derjenige, der zuletzt geantwortet hat eine neue Frage, wie im Beispiel oben, d.h. Spieler A fragt immer bei ungeraden Nummern, Spieler B immer bei geraden Nummern.

Am Ende könnt ihr die Äpfel zählen. Wer am meisten davon hat, kontrolliert, ob er die Reihenfolge der Bilder in der Tabelle des anderes korrekt aufgeschrieben hat. Nur in diesem Fall, gewinnt er, sonst hat der andere automatisch gewonnen!

Viel Spaß und danke fürs Mitmachen ☺

LISTE DER MÖGLICHEN OBJEKTEN UND FARBEN:

⬜ Grau	🟪 Lila	🟦 Blau	🟫 Braun	🟩 Grün	🟨 Gelb
 Welle	 Blume	 Vase	 Nonne	 Birne	 Dose