An exploration of the neural correlates of turn-taking in spontaneous conversation

Ambika Kirkland

Department of Linguistics
Magister Thesis 15 HE credits
Phonetics
Spring Term 2020
Supervisor: Dr. Mattias Heldner

Stockholms universitet

An exploration of the neural correlates of turn-taking in spontaneous conversation

Ambika Kirkland

Abstract

This project added to the sparse body of research on the neural underpinnings of turn-taking with an electroencephalography (EEG) investigation of spontaneous conversation. Eighteen participants (3 male, 15 female, mean age 29.79), recruited and participating in pairs, underwent EEG hyperscanning as they conversed on a freely chosen topic for 45 minutes. In line with previous research, it was predicted that a time-frequency analysis of the EEG might reveal either increased power at around 10 Hz (the location of one of two components of the mu rhythm, an oscillation possibly involved in motor preparation for speech), or reduced alpha (8-12 Hz) power (reflecting non-motor aspects of turn preparation) prior to taking one's turn. Increased power between 8-12 Hz was observed around 1.5 and 1 second preceding turn-taking, but similar power increases also occurred prior to turn-yielding and the conversation partner continuing after a pause, and a reduction in alpha power was found in turn-taking relative to listening to the other speaker continue after a pause. It is unclear whether this activity reflected motor or non-motor aspects of turn preparation, but the spontaneous conversation paradigm proved feasible for investigating brain activity coupled to turn-taking despite the methodological obstacles.

Keywords

Alpha suppression, conversation, electroencephalography (EEG), hyperscanning, mu rhythm, turn-taking

En utforskning av neurala aspekter av turtagning i spontant samtal

Ambika Kirkland

Sammanfattning

Detta forskningsprojekt bedrar till ett ämne där relativt få studier har genomförts med en elektroencefalografi- (EEG-) undersökning av hjärnaktivitet som är kopplad till turtagning i spontant samtal. Arton deltagare (3 män, 15 kvinnor, medelålder 29,79) som rekryterades och deltog i par, genomgick EEG-hyperscanning medan de pratade om ett fritt valt ämne i 45 minuter. Det förutsades att en tidsfrekvensanalys av EEG kan avslöja antingen ökad effekt vid cirka 10 Hz (vilket motsvarar en av två komponenter i mu-rytmen, en oscillation som eventuellt är involverad i motoriska förberedelser för tal) eller reducerad alfaeffekt (8 -12 Hz) (vilket möjligen återspeglar icke-motoriska aspekter av turtagningsförberedelser) innan man tar sin tur. Ökad effekt mellan 8-12 Hz observerades ungefär 1,5 och 1 sekund före turtagning, men liknande ökningar inträffade också innan samtalspartnern tog sin tur eller fortsatte efter en paus, och en minskning av alfaeffekt observerades när turtagning jämfördes till kontexter där försökspersonerna lyssnade när den andra talaren fortsatte efter en paus. Det är oklart om denna aktivitet återspeglade motoriska eller icke-motoriska aspekter av turtagningsförberedelser, men det visar sig vara möjligt att undersöka hjärnaktivitet kopplad till spontant samtal på ett rimligt sätt trots paradigmens metodologiska svårigheter.

Keywords

Alpha-suppression, elektroencefalografi (EEG), hyperscanning, mu-rytmen, samtal, turtagning

Contents

1. Introduction

Turn-taking is a fundamental element in the organization of dialogue. Who speaks when, for how long, and how parties in a conversation manage the task of allocating turns, has been a frequent object of research in sociology, cognitive psychology and linguistics since Sacks and colleagues (1974) broke ground on this topic within the area of conversational analysis.

The messy and unpredictable nature of spontaneous conversation made many researchers prior to Sacks et al. (1974) hesitant to consider it a topic of serious academic investigation. The organization of turns, however, provided schemes for delineating the events that take place within a dialogue in a concrete manner. We can consider points in a conversation where a change from one speaker to another takes place (e.g., Sacks et al., 1974), talk about who has the *floor* (the role of speaker) and how the floor is managed (see, for example, Cappella et al., 1985), and discuss the structure of conversation on the basis of where segments of speech and silence from each conversation party overlap versus occur in isolation (e.g., Heldner & Edlund, 2010).

Nonetheless, spontaneous dialogue has confounded attempts to study the cognitive processes that underlie these turn-taking events, particularly as regards brain activity. The methodologies most applied to directly investigate online cognitive processing, such as the event-related potential (ERP) method for analyzing electroencephalography (EEG) data, require tight experimental control and relatively clean data. Hence, typical approaches to investigating the neural correlates of speech cautiously approach something resembling natural conversation, stopping at a very safe distance (for example, by asking participants to overhear conversations without actively taking part in them) or just within reach (e.g., by asking participants to give relatively brief responses to a mix of scripted and spontaneous questions delivered by an experimenter), sacrificing as little experimental control as possible.

This exploratory project will add to the small handful of studies on the neural correlates of turn-taking in spontaneous conversation and work towards developing a paradigm for EEG studies of spontaneous dialogue, identifying the most troublesome obstacles and perhaps providing some insight into how to overcome them in the future.

2. Background

2.1. Turn-taking in conversation

"The orderly distribution of opportunities to participate in social interaction is one of the most fundamental preconditions for viable social organization."
Emanuel Schegloff (2000)

Turn-taking is one of the most fundamental aspects of human conversation, a structural groundwork for communication which infants seem to grasp before they can even speak. Babies as young as two months old engage in "proto-conversations", in which they actively participate in and initiate turn-taking (Gratier et al., 2015). Even non-human primates such as marmosets engage in turn-taking while vocalizing with one another (Takahashi et al., 2013) which further emphasizes the centrality and universal nature of turn-taking in social communication.
Unlike in structured interactions (such as debates or scripted dialogues) the length and order of turns in spontaneous conversation are not fixed, but must be determined on the fly (Sacks et al., 1974). Speakers use various means of allocating turns, indicating when they wish to speak and when others may speak. Sometimes speaker switches are indicated explicitly: one speaker can select the next speaker by directly addressing them, e.g., asking someone a question or calling on them by name (Sacks et al., 1974). Often, however, the conversational floor is managed with the use of less explicit cues. For example, speakers can project a desire to yield their turn with prosodic information, such as a rise in fundamental frequency corresponding to the upcoming end of the current speaker's turn (e.g., Schaffer, 1983). Gaze also plays a role in next-speaker selection. Auer (2018), for example, notes that in three-party conversations, the current speaker indicates privileged access to the next turn by gazing longer at the party to whom they intend to yield the floor. Gestures can also play a role in managing the floor. For example, shifts to a more upright posture (Harrigan, 1985) or rapid raising and lowering of the eyebrows (Guaïtella et al., 2009) can signal a desire to begin speaking. Speakers who have the floor may also indicate to other speakers that they are not finished speaking, for example with hand gestures (Zellers et al., 2016) or through a combination of syntactic information and the modulation of pause length (Wennerstrom & Siegel, 2003). Breathing behavior can also provide turn-taking cues; for example, several very short utterances coupled with a pronounced inhalation can indicate a desire to take the floor (Włodarczak & Heldner, 2016b). Listeners may also indicate their agreement with or understanding of the current speaker's statements while encouraging the speaker to continue with gestures such as nodding, or short utterances (such as "mhm", "yes", or "right" in English) that overlap the other speaker's utterance without interrupting it. This behavior is referred to as backchanneling (see, for example, Heldner et al., 2010).
There is a degree of cultural specificity to the rules and patterns of turn-taking (e.g., Hayashi, 1991), and different types of conversational contexts may result in different turn-taking behaviors. For example, competitive conversations such as arguments are characterized by shorter pause lengths than cooperative conversations, as noted by Trimboli and Walker (1984). However, some basic principles remain constant across languages and cultures; namely, speakers generally attempt to minimize both gaps and overlaps between speaker turns (Stivers et al., 2009). This could be thought of as the ultimate goal of the various cues discussed above, which help to maintain the smooth flow of conversation by indicating who should speak when and hence keeping transitions as short as possible.

2.2. Minimizing gaps and overlaps in conversation

Speakers seem to accomplish the challenge of maintaining small gaps and overlaps remarkably well, despite the complexity and multimodal nature of conversation. Heldner and Edlund (2010), for example, have found that the modal duration of between-speaker gaps falls around 200 milliseconds. An implication of this extremely small gap is that speakers must begin to prepare for their upcoming turn well in advance, since this latency is significantly smaller than the 600 ms needed to begin articulating the name of even a single object in picture-naming studies (Indefrey and Levelt, 2004), what to speak of an entire sentence (Bögels & Levinson, 2017).

The capacity of speakers to manage such a minimal degree of interruption in the flow of conversation has been a frequent topic of investigation, with Sacks et al.'s (1974) study as an early example. However, there have been few studies to date that have examined the online cognitive/neural processes that underlie turn-taking as they unfold. Conversational analysis, which has its roots in sociology, has traditionally used offline, qualitative analysis of speaker behavior from transcripts of spontaneous conversation to examine the organizational features of conversation. Sacks et al. (1974) did acknowledge some of the potential cognitive demands of the turn-taking system, noting the need for listeners to analyze a conversation partner's speech over the course of an utterance, and other conversational analysts have carried out more fine-grained investigations of how speakers process incoming information over the duration of an utterance (e.g., Schegloff, 2000), but these analyses are still based on offline data and largely focus on comprehension, whereas the greatest processing bottleneck likely involves speech production processes (Bögels & Levinson, 2017).

2.3. EEG and neuroimaging studies of turn-taking

It is perhaps unsurprising that research on the online processes underlying turn-taking is somewhat sparse. The inherently messy and unpredictable nature of spontaneous conversation makes it difficult to apply typical methods of measuring online cognitive processing, such as electroencephalography (EEG) (Bögels & Levinson, 2017). This obstacle is twofold. First, the most common technique for analyzing brain activity connected to specific events or processes with EEG requires tight experimental control. The event-related potential (ERP) method generally involves presenting many similar stimuli over and over so that small changes in electrical activity measured at the scalp can be summed across participants and events in order to identify systematic fluctuations in a relatively weak signal. Spontaneous conversation does not lend itself well to this kind of control. Secondly, speech generates muscle activity, which can result in artifacts in the EEG signal. Unlike the highly stereotypical artifacts generated by eyeblinks, electromyographic (EMG) artifacts are not so easily identified and removed. They can pervade long stretches of the EEG, and are hard to distinguish from brain activity in some cases (Vos et al, 2010).

Previous studies have attempted to address the aforementioned problems in various ways. One means of tackling the issue with muscle artifacts is to avoid them entirely. The overhearer paradigm operates on the assumption that at least some of the same processes that we engage in during conversation take place while acting as passive participants, listening in without actively taking part in a dialogue. Studies using this paradigm (e.g., Schober & Clark, 1989; Tolins & Fox Tree, 2016) ask participants to listen in on a pre-recorded conversation, and in some cases think about what they would say at various points in the discussion were they active participants, but no actual production is involved. The obvious shortcoming of this method is that while there is evidence that overhearers engage in similar comprehension processes as addressees (Tolins & Fox Tree, 2016) the paradigm provides no means of disentangling processes that are strictly related to speech production from comprehension processes. Furthermore, there may be cases in which overhearers differ from addressees even in terms of comprehension (Schober & Clark, 1989).

Other paradigms attempt to elicit production, but with limitations. The interactive quiz paradigm, for example, uses a mixture of live dialogue and pre-recorded quiz questions (which the participants are led to believe is live) to attain a high degree of experimental control in a setting that still feels natural and involves production. Bögels et al. (2015) used this setup to investigate the time course of production planning with EEG and found support for the early planning account of turn-preparation. A positive large positive event-related potential (ERP) and decreased power in the alpha frequency band following the availability of information crucial to answering the questions indicated that non-motor aspects of turn preparation were set into motion as soon as it was possible to begin planning a response.

A more recent study (Bögels, 2019) has investigated the neural correlates of turn-taking processes with EEG in a more natural setting. An interviewer posed a series of pre-scripted yes/no questions to the participants, which were based on their response to a questionnaire that they had filled out prior to taking part in the experiment. After each question, the interviewer asked a spontaneous follow-up question. In this way, the interviewer engaged in a mix of scripted and unscripted dialogue with the participants. Native speakers of Dutch were asked to indicate the point in each question at which they believed a listener would have enough information available to begin planning their response. This annotation of the spontaneous questions allowed the researchers enough control to hone in on their research questions while also coming closer to natural conversation. This study provided further support to the early planning account of turn preparation, replicating the findings of Bögels et al. in a context that was more similar to spontaneous dialogue.

At least one neuroimaging study, a magnetoencephalography (MEG) study by Mandel et al. (2016) which will be discussed in more detail later, did involve entirely spontaneous dialogue. However, this study used relatively short (7 minute) stretches of conversation and contrasted periods of listening vs. speaking, not considering a range of different types of turn-taking events (for example, turn-holding, where the speaker pauses but then continues speaking, turn transitions and situations where the conversation partner continues after a pause). While this was one of a handful of studies on conversation which recorded brain activity from two participants simultaneously, the recording was carried out while participants communicated via audio only from a distance of 5 km, whereas the present study will utilize hyperscanning of two participants situated in the same room.

Many of the studies discussed so far have asked questions about the time course of production planning, and have looked at events whose starting points are reliant on the content of a conversation partner's utterances and hence highly variable in terms of when they begin relative to the start of a speaker's turn. Answering such questions also requires some consideration of the content of utterances, either by using scripted dialogue (e.g., Bögels et al., 2015) or by asking raters to evaluate (a relatively limited amount of) spontaneous dialogue (e.g., Bögels, 2019) after the fact. Neither of these approaches would be practical when it comes to looking for neural correlates of production and production planning in a lengthy, entirely spontaneous dialogue where both parties are experiment participants. So what can we examine in such a naturalistic context?

Using turn switches as the events of interest provides a large set of discrete events for analysis, but in order for such an analysis to be meaningful, there must be some sort of brain response related to turn-taking which occurs at a relatively consistent interval of time preceding a turn switch. Response planning, therefore, can't be meaningfully investigated with such an approach, as it is too reliant on the content of an utterance and occurs in too large of a range of possible latencies.

The choice of analysis techniques is also an important consideration. The ERP technique is a popular and effective method of identifying changes in neural activity measured by EEG that occur in response to an event (such as the onset of a speaker's turn, in this case). This technique provides excellent temporal resolution and is useful in determining precisely when a brain response occurred following the onset of an event, but is less suited to capturing the "multiple neural processes co-occurring and interacting in the service of integrative and dynamically adaptive information processing" (Roach & Mathalon, 2008). Essentially, the ERP technique treats many of the ongoing fluctuations in phase and amplitude present in the EEG signal as background noise. Decomposing the EEG signal into its various frequency components with spectral analysis and observing fluctuations in the magnitude of these components over time may provide a more dynamic and nuanced view of the many parallel processes that underlie turn-taking.

One frequency oscillation which has actually been studied in spontaneous conversation is the mu rhythm, which consists of two frequency components at roughly 10 Hz (which overlaps with the 8-12 Hz alpha frequency band) and 20 Hz (which lies within the 12-30 Hz beta frequency band), both originating from sensorimotor areas of the cortex. Mandel et al. (2016) used magnetoencephalography (MEG) to investigate the modulation of this rhythm relative to the speaker vs. listener roles in spontaneous conversation. They found that over the course of a 7 minute conversation, both components of the mu rhythm were dampened in sensorimotor regions of the left hemisphere for speaking, relative to listening. However, the 10 Hz component increased in power at two points (1 second and 2.3 seconds) before the end of the conversation partner's turn, which the researchers speculated might point to its involvement in preparation for one's own turn, specifically respiratory preparation, when the partner seems about to end their turn.

Alpha suppression is another phenomenon which seems to be related to turn preparation. Ahn et al. (2018), found reduced power in the alpha frequency band (originating in left temporal and centro-parietal regions) during a task where participants took turns counting numbers back and forth to one another. Bögels et al. (2015) also found evidence that alpha suppression is involved in what they suggest are non-motor aspects of response planning, at variable latencies related to the availability of information needed to formulate a response. While the mu rhythm originates within sensorimotor regions (Mandel et al., 2016), the alpha suppression observed by Bögels et al. (2015) was found to originate in occipital and parietal regions, mostly in the left hemisphere. This lent support to their conclusion that the activity is unrelated to motor processes. They further proposed that this activity might be related to a switch in attention from comprehending the other speaker's utterance to preparing their own speech. It might be speculated that the phenomenon of corollary discharge is involved in this switch of attention, as this process underlies one's ability to distinguish between one's own, self-generated speech and speech produced by others (Ford et al., 2001). However, corollary discharge is typically associated with the N100 component (Ford et al., 2001) and with gamma band synchronization between Broca's area and the auditory cortex (Chen et al., 2001), neither of which were found by Bögels et al. (2015).

Finding alpha suppression related to non-motor response planning in this data is unlikely given that it is not reliably associated with any specific position in time before the onset of a speaker's turn, but rather varies depending on the availability of information. However, should a reduction in alpha power be observed prior to the onset of a speaker's turn, this would need to be distinguished from the 10 Hz component of the mu rhythm, since this component occurs within the alpha frequency band (8-12 Hz). The topographical distribution of such components would provide a means to distinguish between processes related to motor readiness that the mu rhythm seems to correlate with (which arise in the sensorimotor cortex) and alpha activity related to non-motor turn preparation (which as noted, has a more parietal/occipital distribution).

3. Aims and research questions

The relative shortage of EEG studies of turn-taking in spontaneous conversation gave this project something of an exploratory character. Many researchers have avoided using spontaneous conversation in neuroimaging studies, choosing instead to approximate conversation in a very constrained fashion or to avoid directly studying production in the first place, because of the difficulty of obtaining meaningful data or being able to apply common analysis techniques to the data. Given these issues with collecting and analyzing brain data from spontaneous conversation, the pool of studies using this type of paradigm is extremely small. Hence, this project will hopefully make contributions both to our knowledge of the neural processes involved in turn-taking in conversation, and to the development of a spontaneous conversation paradigm for EEG studies. Another aspect that lends this project novelty is the use of EEG hyperscanning, whereby brain activity from each pair of participants is measured simultaneously.

The first aim of this project is to investigate the neural correlates of turn-taking in spontaneous conversation with time-frequency analysis and to find frequency components that correspond to different types of turn-taking events. The second aim is to evaluate the limitations of the spontaneous conversation paradigm for neuroimaging studies of turn-taking and discuss possible future uses of the dataset produced in the course of the study.

The following questions will be addressed in achieving the aims of the project:

1. Can evidence of power increases in the 10 Hz component of the mu rhythm and/or suppression in the alpha band be found preceding the start of a speaker's turn during spontaneous conversation?
2. Can other spectral EEG components be found in connection to the onset of a speaker's turn?
3. Which issues with the spontaneous conversation paradigm will prove to be most troublesome in practice, and to which degree can these issues be mitigated without sacrificing spontaneity altogether?
4. How might the hours of synchronized audio, EEG and breathing data collected for this study be mined for additional insights into turn-taking in conversation?

4. Method and data

4.1 Participants

Eighteen participants (3 male, 15 female) took part in the study in pairs. The participants were between 19 and 53 years old, with a mean age of 29.79. Each pair of participants was recruited together and had an existing relationship with one another (friends, classmates, spouses, etc.). All participants were right-handed native speakers of Swedish (one participant had both Swedish and Persian as native languages) recruited via information fliers on the campus of Stockholm University. In addition to Swedish, all participants spoke English, and several spoke additional languages to varying degrees of proficiency (French, 7; German, 3; Spanish, 3; Italian, 2; Turkish, 1; Greek, 1; Swedish Sign Language, 1; Dutch, 1; Arabic, 1; Norwegian, 1). Information about handedness and language proficiency was collected via questionnaires administered when the participants arrived to take part in the study. Each participant was compensated for their participation with three movie theater tickets. Participants were informed of the purpose of the study and of their right to stop participating at any point, and signed a consent form prior to their participation. Participants' personal information was collected and handled in compliance with the General Data Protection Regulation 2016/679 (GDPR).

4.2 EEG recording

EEG recording was carried out with two daisy-chained BioSemi ActiveTwo systems (one 32-channel and one 64-channel system, each with 6 external electrodes) with a sampling frequency of 2 kHz, recorded with BioSemi's acquisition software (ActiView). This setup allowed for hyperscanning of both participants simultaneously.

4.3 Audio and breathing recordings

Audio was captured both via Sennheiser MKE 2 microphones clipped to each participant's clothing, and via miniature accelerometers (Knowles BU-27135) attached to the skin on the tracheal wall below the cricoid cartilage with cosmetic glue. The throat microphones were used to obtain a clear audio signal for each participant, uncontaminated by cross-talk, in order to simplify automatic segmentation into speech and silence. This type of recording can also capture aspects of vocal quality (Heldner et al., 2018). Breathing was recorded with the RespTrack system (Heldner et al., 2019) using Respiratory Inductance Plethysmography (RIP, Włodarczak & Heldner, 2016a) which uses two elastic bands around participants' chest at the armpit and navel to measure changes in the cross-sectional area of the ribcage and abdomen. Audio, as well as the breathing DC-signals, were recorded with an audio interface with DC-coupled inputs (Expert Sleepers ES-9) at 48 kHz, 24 bit, and the DAW software Logic Pro X

In order to sync the various recordings (breathing bands, audio and EEG) two beeps generated in EPrime (Psychology Software Tools, Pittsburgh, PA) were recorded to an additional track in the audio, one at the start of the recording and one at the end. A signal was sent simultaneously to the EEG recording software as an event from EPrime to allow for syncing of the various files.

4.4 Procedure

The study took place at the Stockholm University Brain Imaging Center (SUBIC) which provided a location for carrying out the study, as well as the additional EEG amplifier required for collecting EEG data from two participants simultaneously.
Participants were seated in a sound-proof room and asked to hold a spontaneous conversation in Swedish on a topic of their choice for 45 minutes. It was suggested that participants avoid sensitive personal topics, as the conversations would be recorded and potentially listened to by the researchers, but they were otherwise free to discuss whatever they wished. During the conversation they were seated back to back and asked to look at a fixation cross, placed at each participant's eye level. This was done to minimize eye movements, which occurred at a high rate during early piloting of the experiment setup when participants sat facing each other, and can lead to artifacts in the EEG signal which are challenging to identify and remove. Additionally, participants were asked to avoid excessive blinking and exaggerated movements.

4.5 Generating turn-taking events

The audio recording from the throat microphone was automatically segmented into periods of silence and speech for each speaker using the MPI-PL Silence Recognizer for ELAN (version 5.6). The minimum silence duration was 200 ms and the minimum sound duration was 80 ms. The periods of silence and speech annotated for each speaker were used to generate a list of turn-taking events and their latencies with the help of a Python script. The script used the TextGridTools toolkit (Buschmeier & Włodarczak 2013) to identify the same types of silences and overlaps described in Heldner and Edlund (2010) and illustrated in Figure 1 below: gaps (between-speaker silences, abbreviated BSS), which indicate silences bounded by speech from two different speakers, pauses, which are periods of silence bounded by speech from a single speaker (within-speaker silences, or WSS), within-speaker overlaps (WSO) where one speaker starts and ends a segment of speech within the other speaker's utterance, and between-speaker overlaps (BSO) where a segment of simultaneous speech is bounded by the silence of two different speakers.

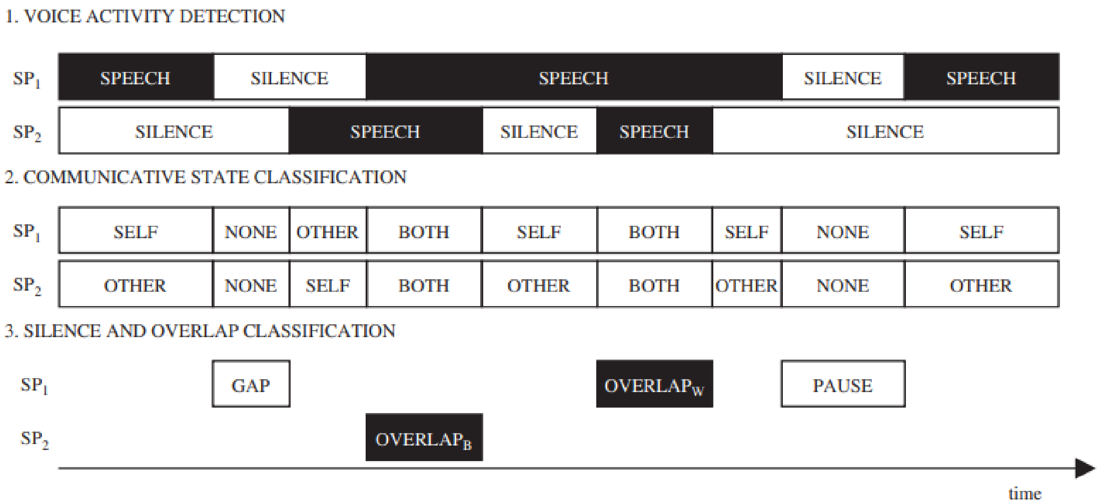

Figure 1. From Heldner and Edlund (2010), used with permission. Illustration of how gaps, overlaps ($OVERLAP_B$), pauses, and within-speaker overlaps ($OVERLAP_W$) are defined and classified in the interaction model. The illustration shows all three steps from the perspectives of both speaker 1 (SP_1) and speaker 2 (SP_2).

Based on these labels, four types of turn-taking events were defined for EEG analysis and used to create an event list that could be read by the EEGLab plugin for Matlab (Delorme & Makeig, 2004) in order to epoch the continuous EEG files for later analysis. These events are shown below in Figure 2. Speech from speaker 1 following a *between-speaker interval* (both BSS and BSO was classified as a speaker 2 to speaker 1 switch (taking the turn from the perspective of speaker 1). Speech from speaker 2 following a between-speaker interval (BSS and BSO) was classified as a speaker 1 to speaker 2 switch (yielding the turn from the perspective of speaker 1). Speech following a pause (WSS) was classified as a continuation by the respective speaker. The latency of the event was the onset of the speech that followed the within-speaker or between-speaker interval.

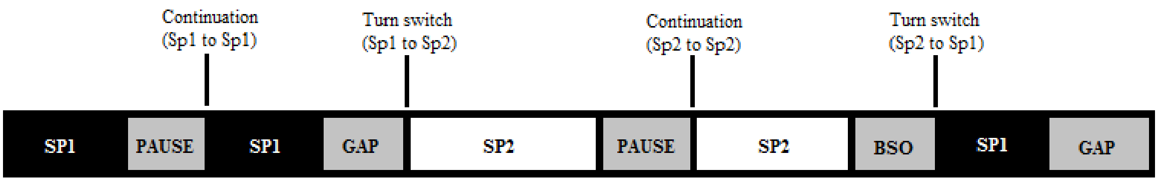

Figure 2. Classification of turn-taking events. The onset of the turn-taking event is the onset of speech following a between-speaker or within-speaker interval.

Note that since EEG activity was recorded for both speakers simultaneously, there are two EEG recordings associated with each point in time. Hence, every turn-taking event is included from the perspective of each participant, and the turn-taking events are defined in terms of "speaker 1" as the speaker whose brain activity is analyzed for that time-locked epoch. If Participant A stops speaking and Participant B begins speaking, this same event will be included as a "speaker 1 to speaker 2 switch" with Participant A's brain activity time-locked to the epoch, and as a "speaker 2 to speaker 1 switch" with participant B's brain activity time-locked to the same point in Participant B's EEG data.

4.6 EEG import and preprocessing

EEG data were imported with the EEGLab plugin for Matlab (Delorme & Makeig, 2004). The data were rereferenced to the mastoid electrodes and 4000 ms epochs (2000 ms before and 2000 ms after speech onset) were created using the turn-taking events described previously. In preparation for independent component analysis (ICA, Delorme & Makeig, 2004), the EEG data were high-pass filtered with a cutoff of 1Hz, low-pass filtered with a 40 Hz threshold, and trimmed with the EEGlab plugin TrimOutlier (Lee and Miyakoshi, 2019), which removes the most extreme parts of the EEG data. The filters had consequences for the time-frequency analysis, as no frequencies below 1Hz or above 40 Hz were included in the data, meaning that parts of the delta and gamma frequency bands were excluded or attenuated. However, given the large amount of noise in the data, especially low-frequency noise, this tradeoff was deemed necessary to obtain clean data. Individual thresholds between 500 and 2500 µV were set to exclude extreme noise, but not blinks (which were removed in a later step). This resulted in the exclusion of up to one minute of data per participant. TrimOutlier was also used to investigate bad channels. Five recordings had one or two bad channels which were removed before carrying out ICA.

4.7 Artifact rejection

Artifacts were rejected in several steps. First, ICA was carried out on the ICA-optimized data set. The resulting components were manually inspected, and based on their topography and spectral profile, components typical of eye blinks were identified and removed. Next, the ICLabel plugin (Pion-Tonachini et al., 2019) was used to classify the source of ICA components. Those components classified with a 90% or higher likelihood of having a non-brain source (e.g., eye or muscle movements, line noise, etc.) were also removed. In addition, components were identified in 13 participants which contributed to a potential artifact just around speech onset. These components were removed. After ICA-rejection, removed channels were extrapolated, and the 64-channel data was reduced to the same subset of 32 head channels present in the data recorded with the 32 channel system. Finally, epochs were rejected automatically by testing various criteria by visual inspection of grand averages and subject averages. A balance between keeping as much data as possible while also removing particularly noisy epochs was found by rejecting head electrodes at +-500 µV, VEOG at +-300 µV and mastoids at +-100 µV. Based on these criteria, 7% of all epochs where rejected. Using these criteria 7% of all epochs where rejected leaving an average of 404 epochs per participant (with a range of between 151 and 972 epochs per participant).

4.8 Computation of time-frequency representations

Time-frequency analysis was carried out with the MNE-Python package, version 0.19.2 (Gramfort et al., 2013) using the Anaconda distribution of Python (Anaconda Software Distribution, 2016). Before calculating group power averages for epochs time-locked to each of four types of turn-taking events, epoched files for each participant were concatenated into a single file per event type. Baseline correction was carried out by subtracting, then dividing by the mean of the entire epoch. This percent change from baseline correction method has the advantage of reducing overall scale differences between frequencies, making them more directly comparable (Roach & Mathalon, 2008). Using the mean of the entire epoch for baseline correction (as opposed to choosing an earlier time window which excludes the event of interest as the baseline) means that changes in power reflect differences in average power over the epoch, as opposed to a direct comparison of differences between pre-event and post-event power (Alday, 2019). While this may be undesirable when looking for causal effects of experimentally manipulated events in a typical EEG experiment, it was appropriate for this study design, because activity both before and after speech onset were of interest, and because the "events" in this case were not experimental manipulations, but spontaneously occurring speech. The use of a whole-epoch baseline allows for a picture of how power fluctuates relative to average power both before and after speech onset.

A complex Morlet wavelet decomposition with 4 cycles was carried out to compute a time-frequency representation (TFR) of the data, using data from all 32 head channels. The frequencies included in the analysis were 2-40 Hz, covering part of the delta frequency band (1-4 Hz), the entire theta (4-8Hz), alpha (8-12 Hz) and beta (12-30 Hz) bands, and the low end (30-40Hz) of the 30-100Hz gamma band.

5. Results

5.1 Single-sample permutation cluster tests

For each condition, non-parametric single sample cluster-based permutation tests were carried out with MNE-Python (Gramfort et al., 2013) to find significant clusters in the time-frequency power estimates. The method has the advantage of addressing the problem of multiple comparison with the use of permutations and cluster-level significance testing. Single-sample t-tests were computed across each frequency/time point in the TFR, and samples which passed the chosen threshold ($p < 0.01$) were clustered based on their proximity to one another in time and frequency and then compared to a null distribution of clusters generated on the basis of 1000 permutations of random sign flips of sample values, comparing the observed clusters with clusters that would occur by chance. Significant clusters for each type of turn-taking event are shown below. Power for each turn-taking event is plotted on the left in Figures 3 through Figure 6, showing the change in power relative to the baseline averaged over all electrodes. The clusters found to be significant with the permutation cluster tests are shown on the right. Clusters where average power is significantly higher are plotted in red, while clusters with lower average power are plotted in blue.

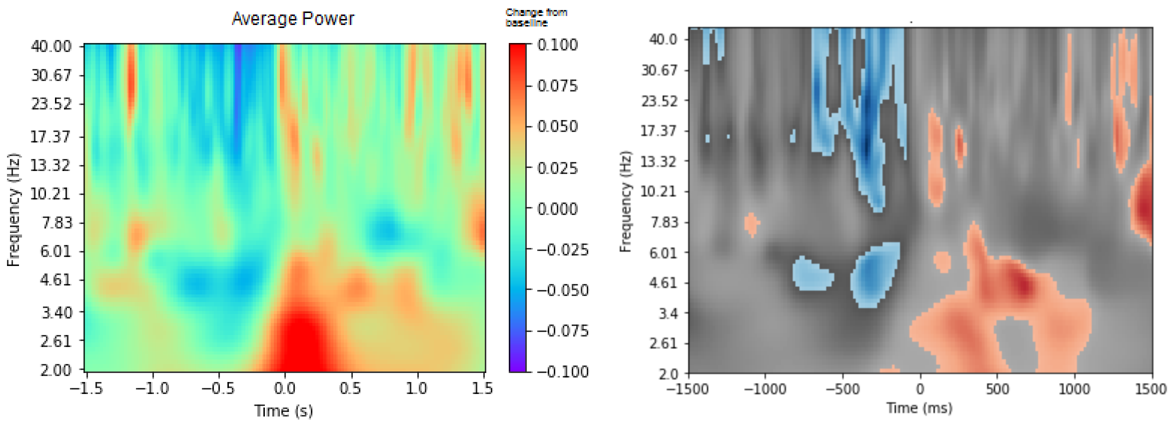

Figure 3. Average power and significant clusters for all 32 channels for continuations after a pause (Speaker 1 to Speaker 1).

Figure 3 shows power across epochs where the speaker whose brain activity was time-locked to the epoch (Speaker 1) continued speaking after a pause, with zero marking speech onset after a silence which was preceded by speech from the same speaker. Significant clusters were found across a range of time points and frequency bands. A summary of the time and frequency intervals for each of these clusters is presented in Table 1, below.

Table 1. Summary of significant time-frequency components for continuations after a pause (Speaker 1 to Speaker 1).

Frequency range	Power decreases: *Before speech onset*	Power decreases: *After speech onset*	Power increases: *Before speech onset*	Power increases: *After speech onset*
delta (<4 Hz)	75-0 ms	0-1050 ms		
theta (4-8 Hz)	787-525 ms 1087-987 ms			87-200 ms 125-437 ms 312-837 ms 1275-1500 ms
alpha (8-12 Hz)	375-212 ms			62-162 ms
beta (12-30 Hz)	750-175 ms			50-150 ms 212-300 ms 962-1000 ms 1175-1287 ms
gamma (>30 Hz)	1412-1275 ms 750-162 ms	912-962 ms 1187-1337 ms		

Figure 4, below, shows power for epochs where the participant whose brain activity was time-locked to the epoch (Speaker 1) began speaking (the zero point in the figure) following a between-speaker interval; in other words, when they took over the turn. Figure 4 also shows the topographical distribution of alpha power across the scalp at 1000-762 ms before the onset of speech, and beta power from 87 ms before speech onset to 162 ms after speech onset.

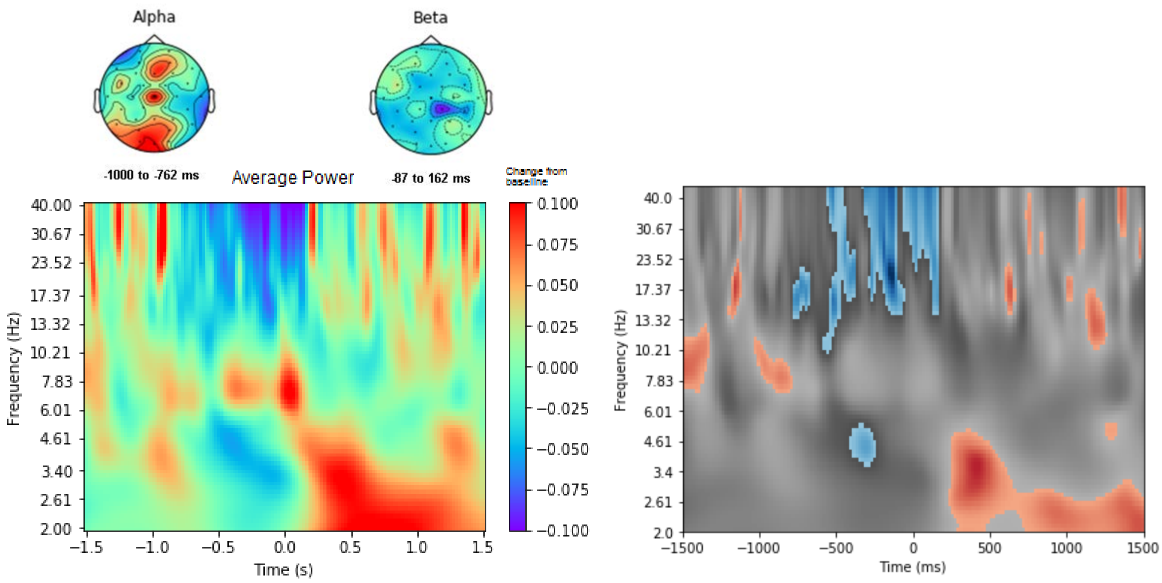

Figure 4. Average power and significant clusters for all 32 channels for epochs where the participant takes over the turn from the other speaker (Speaker 2 to Speaker 1). The distributions of alpha and beta power across the scalp are shown for two different time periods.

In Table 2, below, time and frequency ranges for the significant clusters shown in Figure 4 are summarized.

Table 2. Summary of significant time-frequency components for epochs where the participant takes over the turn from the other speaker (Speaker 2 to Speaker 1).

Frequency range	Power decreases		Power increases	
	Before speech onset	*After speech onset*	*Before speech onset*	*After speech onset*
delta (<4 Hz)				200-1500 ms
theta (4-8 Hz)	425-250 ms			225-500 ms
alpha (8-12 Hz)			1500-1287 ms 1000-762 ms	587-525 ms
beta (12-30 Hz)	775-650 ms 550-475 ms 87-0 ms	0-162 ms	1175-1087 ms	562-560 ms 950-987 ms 1062-1212 ms
gamma (>30 Hz)	562-437 ms 325-112 ms 87-0 ms	0-162 ms	1162-1137 ms	675-725 ms 1025-1062 ms 1275-1337 ms

Figure 5 shows power for epochs where the participant whose brain activity was time-locked to the epoch (Speaker 1) yielded the turn and the other party in the conversation (Speaker 2) began speaking, with zero denoting the onset of Speaker 2's speech following a between-speaker interval.

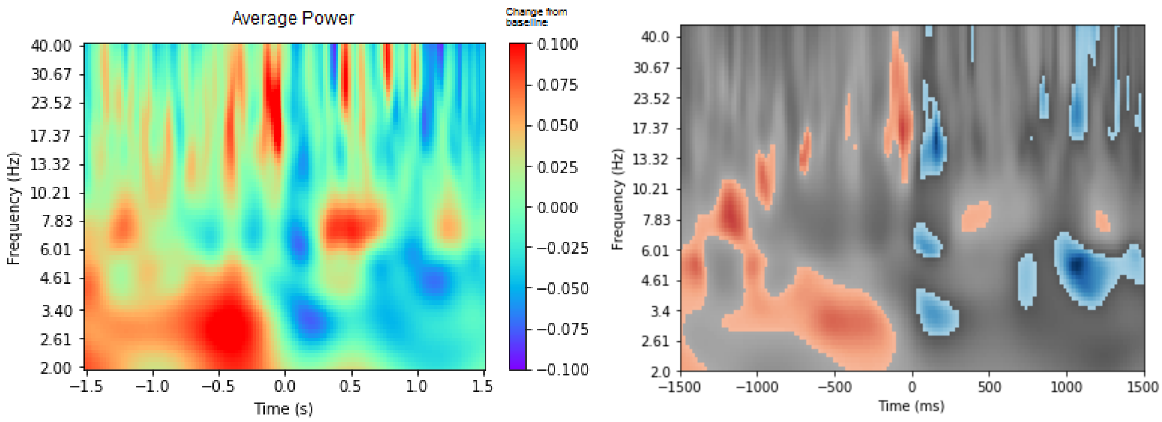

Figure 5. Average power and significant clusters for all 32 channels for epochs where the participant yields the turn to the other speaker (Speaker 1 to Speaker 2)

The time and frequency ranges of the significant clusters in Figure 5 are summarized below in Table 3.

Table 3. Summary of significant time-frequency components for epochs where the participant yields the turn to the other speaker (Speaker 1 to Speaker 2)

	Power decreases		**Power increases**	
Frequency range	***Before speech onset***	***After speech onset***	***Before speech onset***	***After speech onset***
delta (<4 Hz)		25-300 ms	1500-100 ms	
theta (4-8 Hz)		0-187 ms 650-775 ms 875-1500 ms	1500-887 ms	300-475 ms 1200-1312 ms
alpha (8-12 Hz)			1262-1012 ms and 975-850 ms	
beta (12-30 Hz)		50-225 ms 775-850 ms 975-1087 ms 1212-1287 ms 1312-1500 ms	975-850 ms 712-637 ms 437-387 ms 187-0 ms	0-12 ms
gamma (>30 Hz)		75-125 ms 362-387 ms 1012-1162 ms 1250-1275 ms	212-0 ms	0-25 ms

Figure 6 shows power for epochs where the participant whose brain activity was not time-locked to the epoch (Speaker 2) continued speaking after a pause, with zero marking speech onset after a silence which was preceded by speech from the same speaker. Hence, these epochs involve situations where the participant whose brain activity is shown is a listener while the other party pauses and then continues speaking.

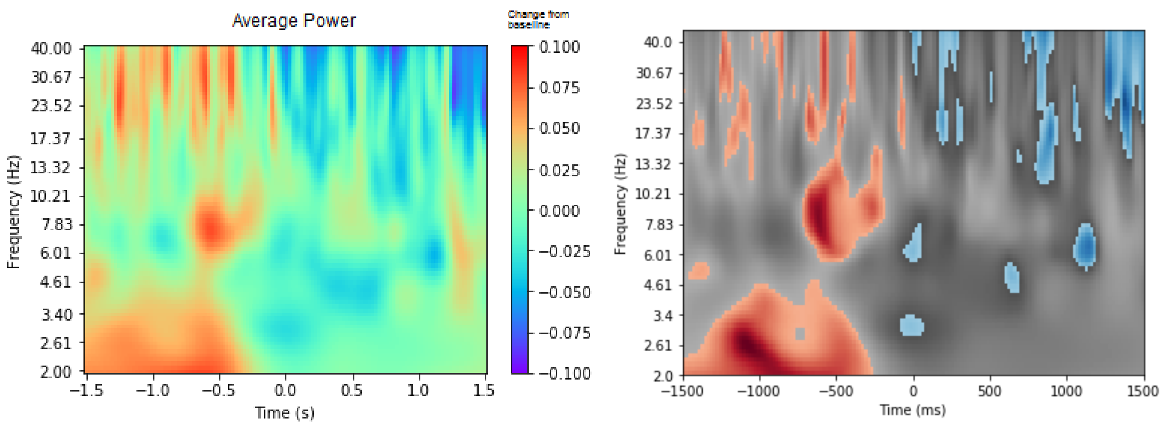

Figure 6. Average power and significant clusters for all 32 channels for epochs where the other speaker continues after a pause (Speaker 2 to Speaker 2)

A summary of the time and frequency ranges for the significant clusters in Figure 6 are presented below in Table 4 (though due to the very large amount of clusters in some frequency ranges, not every cluster is presented in detail).

Table 4. Summary of significant time-frequency components for epochs where the other speaker continues after a pause (Speaker 2 to Speaker 2)

		Power decreases		Power increases	
		Before speech onset	***After speech onset***	***Before speech onset***	***After speech onset***
Frequency range	delta (<4 Hz)	87-0 ms	0-87 ms	1500-262 ms	
	theta (4-8 Hz)	75-0 ms	0-62 ms 575-675 ms 1012-1150 ms	712-187 ms 1450-1287 ms	
	alpha (8-12 Hz)			712-187 ms	
	beta (12-30 Hz)		Numerous	Numerous	
	gamma (>30 Hz)		Numerous	Numerous	

5.2 Statistical analysis of power TFRs for listening vs. turn-taking

In order to compare time-frequency power estimates for epochs where participants listened while the other speaker continued after a pause (Speaker 2 to Speaker 2) with those where they took over the turn from the other party (Speaker 2 to Speaker 1), an additional permutation cluster test with 1000 permutations was carried out to find clusters of power estimates that significantly differed between conditions. This comparison was intended to provide a contrast between turn-taking and a situation as close as possible to passive listening by computing power values for the difference between these two event types.

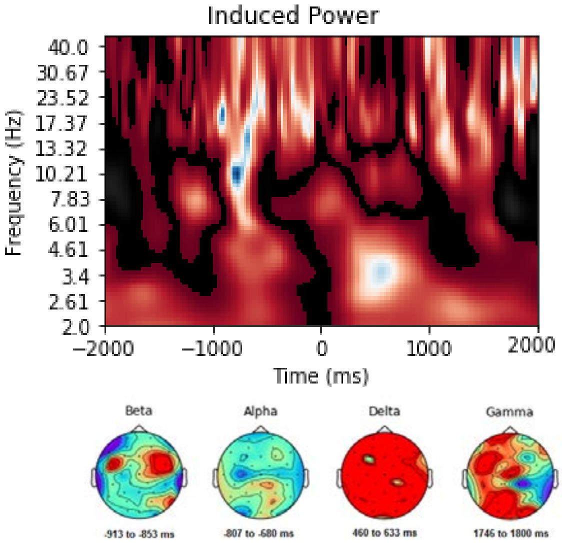

Figure 7. Clusters of power values which significantly differed between epochs where participants continued their own turn after a pause, versus those where participants took over the turn from the other speaker, with topological maps showing the distribution of beta (12-30 Hz), alpha (8-12 Hz), delta (2-4 Hz) and gamma (30+ Hz) power over the scalp for four clusters of significantly decreased power.

One-way ANOVAs were carried out for each pair of frequency/time points in the TFRs, and samples with an F score which passed the chosen threshold ($p < 0.01$) were clustered based on their proximity to one another in time and frequency. The procedure corrects for multiple comparisons with permutations and cluster-level corrections, comparing the resulting clusters to a null distribution which assumes no difference between conditions. Regions where power differs significantly between the two event types are shown in Figure 7. The red regions are regions with significantly higher power in the turn-taking epochs, while the blue regions show areas of significantly decreased power in the turn-taking epochs. Topological maps show the power distribution of four frequency bands in four time ranges where power in those frequency bands was significantly lower in the turn-taking epochs (Speaker 2 to Speaker 1) compared to the listening epochs (Speaker 2 to Speaker 2). The significant clusters shown in Figure 7 are summarized in Table 5, below.

Table 5. Summary of significant clusters of power values for epochs where the other speaker continues after a pause (Speaker 2 to Speaker 2) which significantly differed between epochs where participants continued their own turn after a pause, versus those where participants took over the turn from the other speaker.

	Power decreases	
Frequency range	***Before speech onset***	***After speech onset***
delta (<4 Hz)		460-633 ms
theta (4-8 Hz)		
alpha (8-12 Hz)	807-680 ms	
beta (12-30 Hz)	913-853 ms 680-560 ms 253-80 ms	
gamma (>30 Hz)	746-706 ms	1746-1800 ms

Power was significantly higher in the turn-taking epochs compared to the listening epochs over a large swath of all frequency ranges. Several smaller clusters of significantly decreased activity were also observed: from 746-706 ms before speech onset and 1746-1800 ms after speech onset in the small portion of the gamma band included in the analysis (30-40 Hz), from 913-853 ms, 680-560 ms and 253-80 ms before speech onset in the beta (12-30 Hz) band, from 807-680 ms before speech onset in the alpha (8-12 Hz) band, and from 460-633 ms after speech onset in the part of the delta band included in the analysis (2-4 Hz).

6. Discussion

6.1 Method discussion

As one aim of this project was to assess the feasibility of the method, namely the use of EEG methodologies in spontaneous conversation, a reflection on the strengths and weaknesses of this method is particularly relevant.

The greatest strength of the method was also a significant challenge, namely, the fact that participants engaged in entirely spontaneous dialogue. The two main concerns with this method, factors which have stood in the way of such a method being implemented for this sort of research in the past, were 1) the possible introduction of EMG artifacts from the muscle activity involved in speech production, and 2), the infeasibility of maintaining the sort of tight experimental control which is usually required for EEG studies.

Independent components analysis (ICA) was able to remove muscular artifacts from the data, though it is difficult to be sure of the extent to which muscle activity was successfully separated from brain activity. In the future, more sophisticated methods of removing muscle artifacts could be implemented, such as Canonical Correlation Analysis (Vos et al., 2010). A somewhat unexpected source of artifacts which cropped up during the piloting phase of the study was eye movement artifacts. During piloting, participants sat facing one another. However, this resulted in a large number of horizontal eye movements which were troublesome to separate from the brain activity, hence the decision to sit participants back-to-back looking at a fixation cross. Of course, another consequence of this setup is that participants were not able to see one another, and hence could not react to visual cues that might signal the start of a turn.

As for the uncontrolled nature of the “stimuli”, the only option was to accept that no meaningful events based on the actual content of the dialogue could be used without substantial annotation work and to instead look at turn-taking events, using structural aspects of the conversation rather than analyzing its content. One issue with preparation of the data for analysis that should be considered when working with this (or similar data) in the future is that the turn-taking events occurred in a wide range of different contexts. The length of pauses, gaps or overlaps before the onset of speech varied, and some turn changes occurred after a gap while others occurred after an overlap. Narrowing down some of these parameters might result in cleaner, less noisy data. The segmentation into periods of silence and speech also proved to be somewhat troublesome. The automatic segmentation tool provided by ELAN sometimes failed to correctly identify segments of silence or speech within the specified parameters, perhaps because it was not designed to segment audio from the throat microphones used in this study and expected a noisier signal. Based on a visual inspection of the segmented audio, the end result seemed acceptable. However, no detailed evaluation of the quality of the segmentation was carried out. This may have resulted in some miscategorized turn-taking events, introducing more noise.

However, this study has resulted in the creation of a rich corpus of over 7 hours of spontaneous spoken dialogue, with synced electrophysiological data; both breathing activity and brain activity. The breathing activity could prove helpful in answering questions about the time course of turn-taking events, and a possible connection between breathing behavior and brain activity could be examined in the future. For example, Mandel et al. (2016) speculated that the transient power increases in the 10 Hz component of the mu oscillation that they observed preceding the end of a conversation partner’s turn might be related to respiratory motor preparation. They were not able to confirm this speculation, but it could be addressed by looking at brain activity in conjunction with breathing. This could also potentially clarify whether the changes in alpha activity that was observed preceding turn-taking in this study was related to motor or non-motor aspects of turn preparation (i.e., whether they corresponded to the ~10 Hz oscillation of the mu rhythm or to a non-motor alpha component).

The data collected from the accelerometers on participants' throats was primarily used to obtain clean audio data for segmentation into silence and speech. However, as noted by Heldner et al. (2018) measures of vocal quality could also potentially be extracted from such data. It is possible that this might provide an additional source of cues related to turn-taking.

Annotation of the content of the dialogues would have proven incredibly time consuming, and was outside of the scope of this thesis project. However, selective annotation of parts of the dialogue (for example, isolating questions and asking native speakers of Swedish to pinpoint where in an utterance a listener could have sufficient information to form a response) could allow for the same sort of questions that Bögels and Levinson (2017) posed about the time course of production planning to be answered in the context of fully spontaneous dialogue.

Finally, the fact that a full set of data has been collected for both parties in the conversation means that various forms of coordination between participants in the conversation could be investigated in the future. Coordination of breathing during conversation, for example, has been found to relate to whether or not an attempt to take over the turn from another speaker succeeds (Rochet-Capellan and Fuchs, 2014). Perhaps there is also some insight to be gleaned into the brain activity that relates to this coordination of breathing cycles. Ahn et al. (2018) also found phase synchronization in EEG and MEG data between participants in their study who engaged in interactive counting. Data collected for this study could be used to address whether such phase synchronization takes place in more complex, spontaneous conversation.

6.2 Results discussion

The results of the time-frequency analysis are somewhat hard to interpret in light of the sheer number of significant time-frequency clusters. Addressing the potential significance of every single change in power in every frequency band around each type of turn-taking event would be laborious and probably not provide much clarity. Instead, components relevant to the research questions will be addressed and a few other components which may correspond to turn-taking, auditory processes or speech production will be briefly discussed.

The most specific question about the neural correlates of turn taking was whether increased power in the 10 Hz component of the mu rhythm, and/or alpha suppression, would be observed prior to the onset of a speaker's turn. There is no evidence of alpha suppression before the onset of a participant's turn in the time-frequency results for individual event types. Rather, there seem to be increases in alpha power in a variety of situations—before yielding the turn, before turn onset when taking the turn, and before the other speaker continues after a pause.

The power increase in the alpha band before taking a turn, however, may correspond to the increase in the 10 Hz component of the mu rhythm observed by Mandel et al. (2016). As in their study, this increase occurred at around 1 second before the onset of a turn. While it covered a broader frequency band in the results presented here, this band did include 10 Hz. It was also accompanied by an increase in beta activity around 20 Hz, which could correspond to the 20 Hz component of the mu rhythm. The distribution of alpha activity on the scalp at this time interval does not help much with disambiguating between the kind of non-motor alpha component observed by Bögels et al. (2015) and the mu rhythm. The distribution of activity across the scalp does not necessarily correspond to sources of activity in the brain (Jatoi and Kamel, 2018), so interpreting such a relationship should be done with caution anyhow, but the topography of the pre-turn alpha activity observed in this study shows activity in electrodes over the motor cortex as well as at occipital electrodes. This could mean that this activity reflects a mix of motor and non-motor turn preparation processes, but given the timing, the fact that an increase rather than a decrease in power was observed, and the likelihood that non-motor alpha would be too distributed in time to observe, it seems more likely that the activity is related to motor preparation of some sort.

Relative to the baseline of average power, there was also a period of decreased beta activity right around speech onset when the participant whose brain activity was being measured began their turn following a between-speaker interval, which aligns with another of Mandel et al.'s (2016) observations about the mu rhythm, namely that decreased activity in the 20 Hz component occurs during speech. It has already been noted that topological maps of EEG activity are not an accurate reflection of brain sources (Mandel and colleagues used MEG rather than EEG). However, the beta suppression is distributed over electrodes on the center and left side of the head, a similar pattern to that observed by Mandel et al. (2016). This is far from conclusive but is at least consistent with Mandel and colleagues' findings.

Another pattern of activity in the alpha band was observed, however, when comparing epochs where the participant takes over the turn following a between-speaker interval (Speaker 2 to Speaker 1) to those where the other speaker continues speaking after a pause (Speaker 2 to Speaker 2). Alpha power was significantly lower between 807 and 680 ms before Speaker 1 took their turn, compared to when Speaker 2 continued after a pause (which could be considered something of a "control" since Speaker 1 is a listener in the latter scenario). Once again, however, the suppression is distributed both occipitally/parietally on the scalp, and over centrally-located electrodes above the motor cortex.

While it is a bit difficult in all of the cases discussed above to be sure of the source of the activity, it is nonetheless encouraging that this spontaneous conversation paradigm was able to find activity in the frequency and time ranges where activity related to turn preparation processes has been found previously with more constrained paradigms.

The various changes in power in the other frequency bands are somewhat harder to interpret. Theta and delta activity seem to have differential roles in dealing with speech that is difficult for the listener to understand, with the theta band seeming to relate to speech clarity while delta relates to speech comprehension (Etard and Reichenbach, 2019). Delta activity, which increased in power over particularly large swaths while the participant whose brain activity was measured was actively speaking, also seems to be involved in attention and working memory, possibly increasing during the suppression of afferent signals (perception of one's own motor activity) which interfere with internal concentration (Harmony, 2013). Perhaps this relates to the need to continue attending to the other speaker while speaking (possibly made more difficult by the lack of visual cues due to the back-to-back position, or even by the various pieces of equipment such as breathing bands and microphones the participants had to wear during recording).

Finally, there is evidence that gamma and beta activity are involved in timing cues and the perception of rhythm (Fujioka, 2009), both of which are relevant in conversation. It is important to note that only the lower 10 Hz of the gamma frequency band (which extends to 100 Hz) was included in the analysis, making it difficult to draw conclusions about gamma activity. However, one might speculate that the lower beta power around 1 second before taking a turn, relative to the same point before the other speaker continued after a pause, might be related to predicting the end of the other speaker's turn or timing the beginning of one's own turn.

The inclusion of relatively large epochs and a wide range of frequency bands in the analysis made for a fairly wide net, and consequently, a complex array of results that are not easy to interpret. However, it has also provided plenty of fodder for future investigations of turn-taking in spontaneous conversation. The inclusion of breathing data in the analysis, together with more specific hypotheses about some of the time/frequency regions outside of the alpha band, should allow for a clearer interpretation of these data in the future.

6.3 Ethics discussion

This study was approved by the Swedish Ethical Review Authority and carried out in compliance with the General Data Protection Regulation 2016/679 (GDPR). No ethical issues arose during the course of the study.

7. Conclusions

One of the aims of this thesis was to find neural correlates of different types of turn-taking events. Based on previous findings, the following questions were posed:

1. Can evidence of power increases in the 10 Hz component of the mu rhythm and/or suppression in the alpha band be found preceding the start of a speaker's turn during spontaneous conversation?
2. Can other spectral EEG components be found in connection to the onset of a speaker's turn?

While it was a bit difficult to determine the source of the activity observed in the alpha band (and hence to distinguish between non-motor alpha and motor-related mu components) activity was observed in the expected frequency ranges at expected times preceding the start of a turn, making the answer to the first question a tentative "yes". In addition, a wide variety of other frequency components were observed, the interpretation of which was too complex to address fully but which provided potential avenues for more targeted investigations of other time/frequency ranges.

The second aim was to evaluate the limitations of the spontaneous conversation paradigm for neuroimaging studies of turn-taking and seek to surmount some of those limitations. The questions that were posed in relation to this aim were as follows:

1. Which issues with the spontaneous conversation paradigm will prove to be most troublesome in practice, and to which degree can these issues be mitigated without sacrificing spontaneity altogether?
2. How might the hours of synchronized audio, EEG and breathing data collected for this study be mined for additional insights into turn-taking in conversation?

The paradigm did present significant challenges: noisy data, muscle artifacts, a lack of experimentally controlled events for EEG analysis. However, it ultimately proved possible to carry out with a few small adaptations to the setup (such as sitting participants back-to-back and instructing them to gaze at a fixation cross to avoid more complex eye movements) and the use of turn-taking events generated by automatic segmentation of the audio data in lieu of traditional, experimentally controlled events.

Finally, possible future uses of the dataset produced in the course of the study were discussed. This rich set of electrophysiological data paired with hours of spontaneous conversation could be used to answer a range of questions regarding the neural correlates of breathing activity and its relation to turn-taking, phase synchronization between participants in a conversation, etc.

Acknowledgement

I would like to acknowledge the Stockholm University Brain Imaging center (SUBIC) for providing a space to conduct the study, as well as EEG equipment necessary for conducting hyperscanning. Furthermore, Petter Kallioinen, Iris-Corinna Schwarz, Mattias Heldner, Hatice Zora and Marcin Włodarczak were all involved in carrying out data collection, and Petter Kallioinen and Iris-Corinna Schwarz acted as participants during piloting of the method. Petter also conducted preprocessing of the EEG data and provided helpful insights for data analysis. The study was part of a project funded by the Marcus and Amalia Wallenberg Foundation (Marcus och Amalia Wallenbergs Minnesfond Dnr MAW 2017.0034), on which Mattias Heldner was the principle investigator and Marcin Włodarczak was a co-investigator.

References

Ahn, S., Cho, H., Kwon, M., Kim, K., Kwon, H., Kim, B. S., ... & Jun, S. C. (2018). Interbrain phase synchronization during turn-taking verbal interaction—a hyperscanning study using simultaneous EEG/MEG. *Human brain mapping*, *39*(1), 171-188.

Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology, 56*(12), e13451.

Anaconda Software Distribution. (Version 2-2.4.0.) [Computer software]. (2016) Anaconda, Retrieved from https://anaconda.com

Auer, P. (2018). Gaze, addressee selection and turn-taking in three-party interaction. *Eye-tracking in interaction: Studies on the role of eye gaze in dialogue*, 197-232.

Buschmeier, H., & Wlodarczak, M. (2013). TextGridTools: A TextGrid processing and analysis toolkit for Python. In *Tagungsband der 24. Konferenz zur elektronischen sprachsignalverarbeitung (ESSV 2013).*

Bögels, S. (2019). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *PsyArXiv, 20 May 2019*. https://doi.org/10.31234/osf.io/duq4t

Bögels, S., & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction*, *50*(1), 71-89.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific reports*, *5*, 12881.

Cappella, J. N., Siegman, A. W., & Feldstein, S. (1985). Controlling the floor in conversation. *Multichannel integrations of nonverbal behavior*, 69-103.

Chen, C. M. A., Mathalon, D. H., Roach, B. J., Cavus, I., Spencer, D. D., & Ford, J. M. (2011). The corollary discharge in humans is related to synchronous neural oscillations. *Journal of cognitive neuroscience*, *23*(10), 2892-2904.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods, 134*(1), 9-21.

ELAN (Version 5.6) [Computer software]. (2019). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, *39*(29), 5750-5759.

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., & Roth, W. T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry, 158*(12), 2069-2071.

Fujioka, T., Trainor, L., Large, E., & Ross, B. (2009). Beta and gamma rhythms in human auditory cortex during musical beat processing. *Annals of the New York Academy of Sciences, 1169*(1), 89-92.

Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M., (2013). MEG and EEG data analysis with MNE-Python, *Frontiers in Neuroscience, Volume 7*, ISSN 1662-453X, [DOI]

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in psychology,* 6(1167), 1-10.

Guaïtella, I., Santi, S., Lagrue, B., & Cavé, C. (2009). Are eyebrow movements linked to voice variations and turn-taking in dialogue? An experimental investigation. *Language and speech*, *52*(2-3), 207-222.

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in integrative neuroscience, 7*, 83.

Harrigan, J. A. (1985). Listener’s body movements and speaking turns. *Communication Research*, *12*(2), 233-250.

Hayashi, R. (1991). Floor structure of English and Japanese conversation. *Journal of Pragmatics, 16*(1), 1-30.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555-568.

Heldner, M., Edlund, J., & Hirschberg, J. B. (2010). Pitch similarity in the vicinity of backchannels. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*. Makuhari, Japan.

Heldner, M., Wagner, P., & Włodarczak, M. (2018). Deep throat as a source of information. In *The XXXth Swedish Phonetics Conference, Gothenburg, Sweden, June 7-8, 2018* (pp. 33-38). University of Gothenburg.

Heldner, M., Włodarczak, M., Branderud, P., & Stark, J. (2019). The RespTrack system In *Proceedings SEFOS 2019* (pp. 16–18). University of Southern Denmark.

Jatoi, M. A., & Kamel, N. (2018). Brain source localization using reduced EEG sensors. *Signal, Image and Video Processing, 12*(8), 1447-1454.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1-2), 101-144.

Lee, C., & Miyakoshi, M. (2019). Trim Outlier. Retrieved from https://sccn.ucsd.edu/wiki/TrimOutlier

Mandel, A., Bourguignon, M., Parkkonen, L., & Hari, R. (2016). Sensorimotor activation related to speaker vs. listener role during natural conversation. *Neuroscience letters, 614*, 99-104.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage,* 198, 181-197.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://www.pstnet.com.

Roach, B. J., & Mathalon, D. H. (2008). Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia bulletin, 34*(5), 907-926.

Rochet-Capellan, A., & Fuchs, S. (2014). Take a breath and take the turn: how breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1658), 20130399.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* (pp. 696–735).

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in society, 29*(1), 1-63.

Schaffer, D. (1983). The role of intonation as a cue to turn taking in conversation. *Journal of Phonetics*, *11*(3), 243-257.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. Cognitive psychology, 21(2), 211-232.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences,* 106(26), 10587-10592.

Takahashi, D. Y., Narayanan, D. Z., & Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. Current Biology, 23(21), 2162-2168.

Tolins, J., & Fox Tree, J. E. (2016). Overhearers use addressee backchannels in dialog comprehension. *Cognitive science, 40*(6), 1412-1434.

Trimboli, C., & Walker, M. B. (1984). Switching pauses in cooperative and competitive conversations. *Journal of Experimental Social Psychology, 20*(4), 297-311.

Vos, D. M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F. X., Huffel, V. S., & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics*, *8*(2), 135-150.

Wennerstrom, A., & Siegel, A. F. (2003). Keeping the floor in multiparty conversations: Intonation, syntax, and pause. *Discourse Processes, 36*(2), 77-107.

Włodarczak, M., & Heldner, M. (2016a). Respiratory belts and whistles: A preliminary study of breathing acoustics for turn-taking. In *INTERSPEECH 2016, San Francisco, USA, September 8–12, 2016 (*pp. 510-514). International Speech Communication Association.

Włodarczak, M., & Heldner, M. (2016b). Respiratory turn-taking cues. In *Interspeech 2016, San Francisco, USA, September 8–12, 2016* (pp. 1275-1279). The International Speech Communication Association (ISCA).

Zellers, M., House, D., & Alexanderson, S. (2016). Prosody and hand gesture at turn boundaries in Swedish. *8th Speech Prosody 2016, 31 May 2016 through 3 June 2016*, 831-835.

Stockholms universitet/Stockholm University
SE-106 91 Stockholm
Telefon/Phone: 08 – 16 20 00
www.su.se

Stockholms universitet