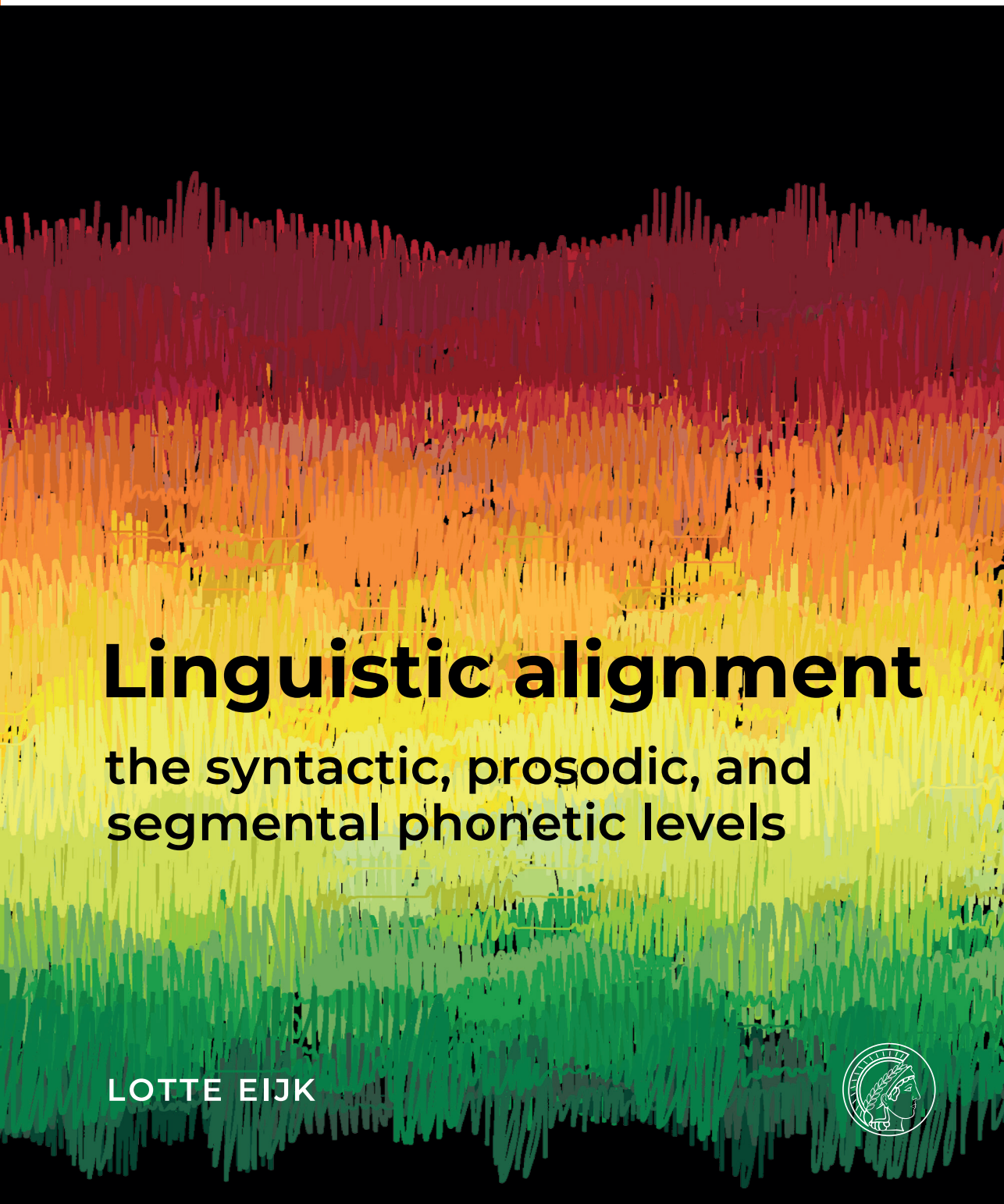
MAX PLANCK INSTITUTE
FOR PSYCHOLINGUISTICS



Linguistic alignment

the syntactic, prosodic, and segmental phonetic levels

LOTTE EIJK

Linguistic alignment:

the syntactic, prosodic, and segmental phonetic levels

Lotte Eijk

Funding Body
This research was funded by the Netherlands Organisation for Scientific Research (NWO) Gravitation Grant 024.001.006 to the Language in Interaction Consortium.

International Max Planck Research School (IMPRS) for Language Sciences
The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs

The MPI series in Psycholinguistics
Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series

ISBN: 978-94-92910-47-9
Cover design by Sofie Rosalien Deen
Layout by Douwe Oppewal
Printed and bound by Ipskamp Drukkers, Enschede

Linguistic alignment:

the syntactic, prosodic, and segmental phonetic levels

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 14 april 2023
om 10.30 uur precies

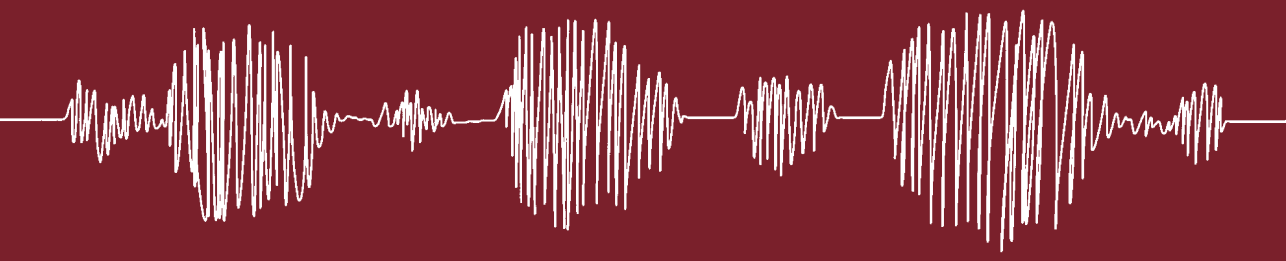
door

Lotte Dorien Eijk
geboren op 12 januari 1995
te Amersfoort

Promotoren:
Prof. dr. M.T.C. Ernestus
Prof. dr. H.J. Schriefers

Manuscriptcommissie:
Prof. dr. M. van Oostendorp
Prof. dr. R.J. Hartsuiker (UGent, België)
Prof. dr. K. Spalek (Heinrich Heine Universität Düsseldorf, Duitsland)
Dr. E. Janse
Dr. R. van Son (Antoni van Leeuwenhoek)

Table of contents



CHAPTER 1

Introduction

Conversation is one of the primary situations in which people use language. During conversation, people tend to synchronise their behaviours. For example, when a speaker calls an object a *couch*, their interlocutor may use this word as well instead of *sofa*. Or, when one speaker speaks faster, their interlocutor may also begin to speak faster. This process of synchronisation occurs both in speech and in non-linguistic behaviour.

Broadly, this synchronisation is referred to, amongst others, as *alignment* (e.g., Pickering & Garrod, 2004), *entrainment* (e.g., Wynn & Borrie, 2022), *accommodation* (e.g., Giles, Coupland & Coupland, 1991), *convergence* (e.g., Pardo, 2006), *mimicry* (e.g., Holler & Wilkin, 2011), *adaptation* (e.g., Brennan & Hanna, 2009), and *coordination* (e.g., Fusaroli et al. 2012). These terms usually denominate slightly different concepts, which differ in the timing, the underlying concept and/or the underlying theory. These different terms are also used in different domains of the language sciences. For instance, whereas the terms *accommodation* and *entrainment* are more commonly used in phonetics, *alignment* is more common in the syntactic literature, and *mimicry* is often used in research on non-linguistic behaviour. Researchers investigating the general phenomenon thus do not align on the terminology they use.

In this dissertation, I adopt the term *alignment*. Alignment can be established over different time spans. If alignment is established over a short period of time, speakers align to the immediately preceding speech, and I will refer to this type of alignment as "local alignment". In contrast, speakers may be assumed to align to a larger number of an interlocutor's utterances, and I will refer to this as "global alignment". Speakers may also continue the alignment when the interlocutor is no longer present (either when there is no interlocutor or a different interlocutor). I refer to this long-lasting alignment as long-term global alignment.

Investigating the timing of alignment can give us insight in underlying cognitive mechanisms. It can furthermore shed light on whether alignment on different linguistic levels behaves similarly over time, or whether alignment on these levels are independent processes.[1] The timing of alignment has been investigated in a number of studies, and for different linguistic levels (i.e., phonetic, syntactic, lexical; e.g., Levitan & Hirschberg, 2011; Ostrand & Chodroff, 2021). Results from these studies show varying patterns, possibly as operationalisations and interpretations of alignment differ for different linguistic levels. The topic of alignment at different linguistic levels is central to this dissertation, focusing on local and global alignment on the syntactic, prosodic, and segmental phonetic levels.

This introduction will first discuss local, global, and long-term global alignment. Next, studies on the three linguistic levels will be described, as well as studies on the

1 In this dissertation, I assume phonetics is a linguistic level.

combination of the different linguistic levels, after which a short description of theories of alignment will be discussed. Lastly, the research question and a description of the contents of this dissertation will follow.

1.1 Timeline of alignment: local, global and long-term global alignment

As mentioned above, most researchers define local alignment as speakers adapting to the behaviour of the interlocutor in the previous turn (e.g., Branigan, Pickering & Cleland, 2000; Gijssels, Casasanto, Jasmin, Hagoort & Casasanto, 2016). They may differ in how they measure local alignment, adopting slightly different definitions of the concept *turn* (e.g., Turn Constructional Units – TCU; Emina & Jan, 2018 versus Inter-Pausal Units – IPU; Levitan & Hirschberg, 2011), or allowing more time and/or speech material between the speaker's and the interlocutor's relevant stretches of speech (e.g., Howes and colleagues (2010) who score alignment of the nearest pertinent syntactic structure). Other researchers measure local alignment over set time intervals (e.g., Time-aligned moving average; e.g., Bonin et al., 2013; Kousidis, Dorran, Mcdonnell & Coyle, 2009).

Similar to local alignment, there are different ways to test for global alignment. One conceptualisation of global alignment is alignment over a whole conversation. Global alignment can be studied by comparing speakers' way of speaking at the beginning and at the end of a conversation (e.g., by dividing the conversation into two halves; Levitan & Hirschberg, 2011), sometimes also comparing this between real versus pseudo-pairs (i.e., pairs of participants who did not interact with each other; Reitter & Moore, 2007), or by comparing utterances separated by large amounts of intervening speech materials (e.g., Bock & Griffin, 2000). In addition, global alignment can be tested by comparing a speaker's way of speaking before and after the conversation (when the speaker no longer receives input from the interlocutor; Gijssels et al., 2016; Troncoso-Ruiz, Ernestus & Broersma, 2019), establishing whether the global alignment is long-lasting.

1.2 Alignment on different levels

Investigations of both local and global alignment have been conducted on different linguistic levels. This dissertation will focus on the syntactic level, the prosodic level, and the segmental phonetic level. For each level, I will discuss several studies, and different options for measurements of alignment – single measures or more holistic measures. This will be followed by a description of some studies focusing on multiple linguistic levels.

1.2.1 Syntactic alignment

There is an abundance of research on syntactic alignment for which the groundwork was done by early studies by Levelt and Kelter (1982) and Bock (1986, 1989). Levelt & Kelter (1982) presented the first study on local syntactic alignment in dialogue. They investigate a so-called correspondence effect, using question (Q)-answer (A) pairs, like Q: "(At) what time do you close?" A: "(At) five o'clock.". They found that participants tended to match their answers to the syntactic structure of the question. A participant would answer "At five o'clock" when being asked *at* what time the store closes and "Five o'clock" when being asked what time they close.

A large share of the syntactic alignment literature is based on Bock's (e.g., 1986, 1989; Bock & Griffin, 2000) early studies. They measured alignment to a particular syntactic structure by comparing the proportions of this syntactic structure in an experimental condition, where participants were presented with speech containing the structure, and in a condition not presenting this structure. This approach is also referred to as structural priming. Branigan and colleagues (2000) extended this approach to dialogue settings, in which participants interacted with confederates, who were instructed to use particular syntactic structures. They found clear evidence for syntactic alignment in such dialogue settings as well.

Since then, some studies have adapted the procedures of the structural priming literature (e.g., Bock, 1986, 1989) to investigate both local and global alignment. Syntactic alternations that have often been documented to show alignment include active versus passive (e.g., Allen, Haywood, Rajendran & Branigan, 2011; Schoot, Hagoort & Segaert, 2019; Schoot, Heyselaar, Hagoort & Segaert, 2016) and the dative alternation (e.g., Bernolet & Hartsuiker, 2010; Branigan et al., 2000, Branigan, Pickering, McLean & Cleland, 2007; Branigan, Pickering, Pearson, McLean & Nass, 2003; Jaeger & Snider, 2013). These different syntactic structures are usually investigated in experimental tasks which allow for control over the occurrence of the syntactic structures of interest such as a picture description task in which participant and interlocutor alternate describing pictures and in which the confederate produces scripted speech (e.g., Branigan et al., 2000).

Other researchers have moved away from studying syntactic alignment for specific syntactic structures and have shown alignment for more holistic measures. An example of such measures is the co-occurrence of combinations of syntactic units (i.e., *n*-grams of Part of Speech tags). Studies applying this measure showed that a specific combination of syntactic units becomes more prevalent over the conversation in the production of both speakers involved in the conversation. Such holistic measures are often applied to spontaneous conversations (e.g., Dale & Spivey, 2006).

A recent study by Ostrand and Ferreira (2019) investigated whether syntactic alignment may be interlocutor-specific, since interlocutor-specificity has been shown

for lexical alignment (e.g., Brennan & Clark, 1996; Wilkes-Gibbs & Clark, 1992). In syntactic alignment, interlocutor-specificity would mean that a speaker retains some knowledge about the syntactic structures used by a specific interlocutor and aligns to these syntactic structures in interaction with this particular interlocutor, but not necessarily in interaction with a different interlocutor. Ostrand and Ferreira's (2019) experiments suggest that, when interlocutor-specificity does not add to communicative utility, syntactic alignment is not specific to the interlocutor, but rather reflects the statistical distribution of the syntactic structure across the entire input received, with this input possibly coming from more than one interlocutor.

1.2.2 Prosodic alignment

Alignment has been studied for various prosodic features, including pitch, articulation rate and intensity (e.g., Schweitzer & Lewandowski, 2013; Levitan & Hirschberg, 2011; Gijssels et al., 2016). Alignment in these suprasegmental features has been found rather consistently. Researchers have investigated alignment by studying one measure for each utterance (e.g., mean F0 per utterance in Gijssels et al., 2016) versus measures that capture the dynamics of (part of an) utterance (e.g., pitch contour in Gorisch, Wells & Brown, 2012).

Similar to syntactic alignment, researchers have also studied more holistic measures by combining a range of features. Apart from prosodic features, these large-scale measure extractions usually included other acoustic measures as well, thereby reflecting more holistic measures. Ostrand and Chodroff (2021), for instance, investigated 323 continuous measures of acoustic-phonetic alignment (including sonorant-specific, obstruent-specific and general temporal specific measures). Some of these measures depended on specific phonemes, others were taken over each utterance, and yet others over the whole recording. The authors used machine learning on these 323 individual measures to compare participants' productions to the confederates' productions, and found that this holistic measure of alignment, as well as other individual features such as speech rate, showed alignment where syntactic or spectral-phonetic measures did not.

1.2.3 Segmental phonetic alignment

Phonetic alignment is often used to refer to alignment to the (articulatory) realisation of certain segments. The phonemes most commonly investigated in alignment research are vowels (e.g., Babel, 2010; Troncoso-Ruiz et al., 2019). Contexts in which these phonemes are investigated are quite often interactions between speakers of different regional variants (e.g., Babel, 2010). These contexts provide a wide range of variation in the realisation of phonemes and thus allow for enough space for speakers to show alignment to their interlocutor.

Studies on regional alignment do not present consistent results. Many studies show large individual differences (e.g., Babel, 2010), which are not always clearly attributable to social factors. Social factors that have shown to have an influence on alignment are, for example, a social bias towards speakers from a certain area (Babel, 2010) and the subjective attitude towards a speaker (Yu, Abrego-Collier & Sonderegger, 2013). In contrast, Gessinger and colleagues (2019b), for instance, found large individual differences in a study on the articulatory realisation of a German suffix, but these differences could not be accounted for by the participant's perceived likeability and competence of the interlocutor.

Next to using acoustic measures for phonetic alignment (e.g., F1-F2 of vowels), phonetic alignment can also be investigated by the use of perceptual measures. An example is the AXB task, where participants indicate whether a speech sample X sounds more like A or B (e.g., Pardo, 2006). This task can be considered as testing a more holistic measure of alignment, since participants have multiple cues to base their choices on.

1.2.4 Studies investigating alignment on multiple linguistic levels

Research comparing alignment at different linguistic levels is rather rare (e.g., Oben, 2015; Ostrand & Chodroff, 2021; Rahimi, Kumar, Litman, Paletz & Yu, 2017; Reitter & Moore, 2007). Oben (2015) studied prosodic, lexical, syntactic and gestural alignment on the basis of a corpus of speech consisting of two parts: a task-based conversation and a free conversation. He found that, in general, alignment was variable, meaning that speakers sometimes aligned locally and sometimes did not. He furthermore found that features belonging to the same linguistic level may behave differently from each other, while features belonging to different levels may behave similarly. More specifically, he found global alignment for some features (gesture, pitch, loudness, function words and syntax), where alignment increased over time, but for other features (speech rate, content words) he did not find this global pattern.

As discussed before, another study investigating multiple linguistic levels of alignment was conducted by Ostrand and Chodroff (2021). The authors studied several features at the phonetic level and syntactic level, combined with a more holistic measure of acoustic alignment consisting of 323 measures. Participants first received all input from a picture description task from two different confederates, one native and one non-native speaker. After receiving these picture descriptions from both confederates, participants described pictures to the two confederates separately. They found that speakers aligned in some individual acoustic features such as speech rate, and in the holistic acoustic measure, but not in other spectral-phonetic measures or in syntactic measures. This study confirms that features of the same linguistic level may show different alignment patterns.

1.3 Theories on alignment

Several theories have been proposed to explain linguistic alignment effects, some aiming at an explanation of multiple linguistic levels, and others mostly theorising about specific linguistic levels. This introduction will not give an exhaustive overview of all available theories; rather, I will highlight those theories and aspects that are central to this dissertation. More exhaustive overviews of theories and studies on alignment can be found in, for example, Wynn & Borrie (2022) and Rasenberg, Özyürek & Dingemanse (2020), also including suggestions for more comprehensive definitions of alignment than I gave above.

One of the most well-known theories in the alignment literature is the Interactive Alignment Model (IAM; Pickering and Garrod, 2004). This theory tries to explain alignment on all linguistic levels and suggests that alignment is a fully automatic process that is resource-free. The underlying mechanism of alignment is assumed to be automatic priming. When a speaker hears an utterance, certain cognitive representations become more active, which consequently makes these representations easier to be used by this speaker in a next utterance. Resulting from automatic priming, alignment is considered a local effect. Local syntactic alignment effects, for example, are often explained, at least partly, to be underlain by priming (see Pickering & Branigan, 1999). Some studies mention that priming could nevertheless also have long-term effects leading to adaptation (e.g., Reitter, Keller & Moore, 2011) or implicit learning (e.g., Chang et al., 2000; In this dissertation, I use the term implicit learning to refer to long-term effects), which could explain global alignment effects.

Another important aspect of the Interactive Alignment Model is that different linguistic levels of alignment can influence each other. One of the examples mentioned by the authors is the *lexical boost effect* in syntactic alignment: Syntactic alignment tends to be stronger when the verb is shared between the prime sentence and the target sentence (Branigan et al., 2000). In such a manner, alignment at one level can percolate to another level, ultimately leading to the alignment of the interlocutors' situation models – their representations of the situation at the moment of the conversation. Alignment of situation models is the basis for successful communication and is also referred to as implicit building of common ground. The Interactive Alignment Model proposes that modelling an interlocutor's situation model, next to a speaker's own, should not be necessary when these two models are aligned, which would be more efficient. Further, the assumption of alignment by an automatic mechanism implies that alignment is not interlocutor-specific, because the priming occurs automatically with each new interlocutor.

The assumption that alignment results from automatic priming contrasts clearly with the theory proposed by Clark (1996). According to this theory, which is mostly based on lexical alignment, alignment is the result of joint action. Speakers create common ground by forming conceptual pacts in order to understand each other. Clark states that by communicating and creating common ground, people align (see also Clark & Brennan, 1991; Brennan & Clark, 1996). Because this common ground is specific to a pair of interlocutors, alignment is assumed to be interlocutor-specific: that is, when two interlocutors have created a shared common ground, they do not stick to this common ground when talking to a new interlocutor.

Ostrand and Ferreira (2019) extend theories (such as the one by Clark, 1996) on alignment, by hypothesising that interlocutor-specificity may differ for different levels of alignment. They propose that alignment is only interlocutor-specific if it has communicative utility. That is, according to the authors, speakers do not align to a specific interlocutor when it does not help the goal of the conversation. In those cases, speakers rather align to recent interlocutor-independent averages. For example, using the same labels to refer to objects could help communication, whereas using the same syntactic structures will most likely not (as hypothesised by Ostrand & Ferreira, 2019). As a consequence, speakers would show interlocutor-specific alignment at the lexical level but not at the syntactic level.

Next to these theories, alignment has been proposed to be influenced by social factors. The most prominent theory on alignment and social factors is the Communication Accommodation Theory, proposed by Giles and colleagues (1991). This theory is mostly based on phonetic alignment studies. The authors argue that speakers modify their speech according to the social situation. When they wish to belong to a certain group of speakers, or wish to be liked by their conversational partner, they align.

Local and global prosodic and segmental phonetic alignment studies are often explained as depending on social relations or communicative factors (e.g., Babel, 2010; Levitan et al., 2012; Ostrand & Chodroff, 2021; Schweitzer & Lewandowski, 2013). Schweitzer and Lewandowski (2013) for example, found an overall effect of divergence in articulation rate in spontaneous conversations. However, when participants liked each other more, the effect of divergence was either absent or changed to alignment.

While the priming and grounding theories are often presented as being two opposing theories, I would like to stress that I do not assume that these theories are mutually exclusive. As, for example, already mentioned by Krauss and Pardo (2004) in their response to Pickering and Garrod (2004), it could well be that alignment is due to a combination of multiple, if not all, underlying mechanisms mentioned above. For example, where one theory could possibly better account for local alignment, another might be better suited to account for global alignment. Also, alignment at different linguistic levels or of different features might be better explained by different theories.

1.4 Research questions and outline

The goal of this dissertation is to add to the knowledge on linguistic alignment and to find out more about the underlying mechanisms by studying different linguistic levels. Firstly, the theories proposed to explain alignment will be investigated by focusing both on very local (turn-by-turn) and more global (over a larger time span) alignment measures. Secondly, possible interlocutor-specificity will be explored. Chapters 2 to 5 present experimental studies investigating local and global alignment in different measures on the syntactic, prosodic and segmental phonetic level. The data for these different levels come from the same participants and the same experimental sessions such that the presence versus absence of alignment at the different levels can be compared. Although the three levels of alignment will be investigated in the same dataset, the combination of the three levels in a single analysis appeared out of the scope of this dissertation. Chapter 6 presents another dataset designed to enable researchers to study alignment in a task-based conversation. The results on the separate linguistic levels and the discussion of Chapter 6 will be brought together in Chapter 7.

The dataset on which Chapters 2 to 5 are based, contains production data from participants in two different sentence completion task experiments: a main experiment and a control experiment. Participants in the main experiment took turns completing sentences with pre-recorded speech from two different interlocutors. They started the experiment by completing sentences by themselves in a pre-test. This was followed by Round 1, where participants interacted with Confederate 1, and then Round 2, where they interacted with Confederate 2. An inter-test and Round 3 followed, during which participants interacted with Confederate 1 again. After Round 3, participants completed a post-test which was very similar to the pre-test. In the control experiment, participants did not receive any auditory input, and some of the confederates' sentences were replaced with filler sentences to control for syntactic alignment effects. This control experiment allows us to distinguish actual alignment effects in the main experiment from fluctuations over the course of the experiment that may be due to other influences, like fatigue.

Chapter 2 focuses on syntactic alignment at a local and a global level. The two interlocutors in the main experiment differed in their use of the different word orders of the auxiliary and the participle in Dutch subordinate clauses. This chapter addresses three questions. Firstly, do speakers align locally to the word order used by the confederates? Secondly, how does long-term global syntactic alignment affect possible local syntactic alignment? And lastly, what are the limits of long-term global syntactic alignment?

Chapter 3 investigates alignment on the prosodic level, focusing on alignment in Pitch (measured as median F0 over the complete utterance) and Articulation Rate (measured as syllables per second phonation time). This study made use of a subset

of the data from the main experiment. Chapter 3 addresses the question whether alignment in Pitch and Articulation Rate can (predominantly) be explained by local alignment, or by global alignment, by asking three questions. Firstly, is there local alignment? Secondly, do speakers align more rapidly to an interlocutor when they have already talked to this interlocutor before? And lastly, how long does alignment persist? Chapter 4 expands on Chapter 3 by adding a control experiment, in which no alignment should occur as the participants did not hear the interlocutors.

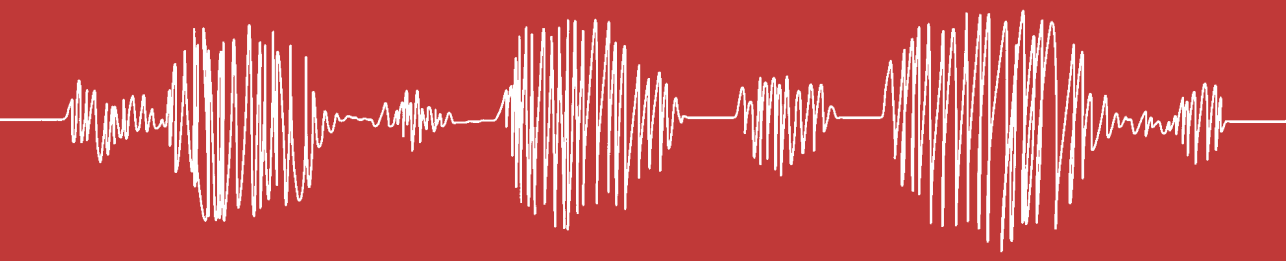
Chapter 5 studies the segmental phonetic level. This chapter focuses on regional variants of a Dutch phoneme, the so-called "hard g" versus the "soft g". Confederate 1 used the "hard g" and Confederate 2 used the "soft g". Phonetic alignment to regional variants was investigated by asking two questions. Firstly, does alignment depend on the prestigiousness of a variant and is this better reflected in local or global measures? And secondly, do speakers need to be exposed to a certain allophone to align, or do they also change their productions when only hearing an interlocutor they have interacted with before, without hearing the allophone?

Chapter 6 describes a dataset created within the CABB team (Communicative Alignment in Brain and Behaviour team) which was designed to investigate alignment at different linguistic and non-linguistic levels, and additionally in pre- and post-brain and behavioural measures. It consists of pre- and post-measures in different behavioural tasks and neural correlates of speakers' representation of certain objects, and interactional data from speakers' interactions about these objects. This dataset was created to combine different levels of alignment, but also offers opportunities to relate the alignment within an interaction to several behavioural and neural pre- and post-measures.

Lastly, Chapter 7 concludes this dissertation. In this chapter, I summarise the experimental Chapters 2 to 6 and bring together the findings and conclusions of the studies described in Chapters 2 to 5 in an attempt to provide more insight into the question what mechanisms underlie linguistic alignment. This chapter also discusses the potential of the two different datasets (the one used for Chapter 2 to 5, and the one presented in Chapter 6) and future directions for the study of alignment.

1.5 Reading guide

Tables and figures are numbered separately per chapter, starting with the chapter number, followed by the table or figure number. The tables and figures in the Appendices, after each chapter, start with the letter of the Appendix, followed by a number per Appendix. Footnotes are numbered throughout the dissertation. All references can be found at the end of this dissertation.



CHAPTER 2

The flexibility of syntactic structures in interaction: alignment to different interlocutors

Abstract

Alignment is the process of adapting speech to another interlocutor's speech. We investigated alignment of a Dutch syntactic structure, where an auxiliary verb is placed either before or after the participle in subordinate clauses, focusing on successive alignment to two different interlocutors differing in their use of the syntactic structure under investigation. Participants first completed sentences in a pre-test by themselves, then in interaction with Confederate 1, with Confederate 2, with Confederate 1 again, and finally by themselves again in a post-test. Our results suggest that participants aligned to both orders of the syntactic structure used by the confederates. We found no evidence of participants re-aligning to a confederate without hearing the pertinent syntactic structure. Furthermore, short-term alignment effects seemed to overrule potential long-term effects. The results show that speakers can align to different interlocutors within a short time span provided that the respective interlocutors produce the relevant syntactic structure.

2.1 Introduction

Speakers tend to adapt their speech to an interlocutor. This adaptation is called alignment, also referred to as convergence or entrainment. In other cases, alignment is denominated as "priming". However, this latter term is sometimes also used to refer to an assumed mechanism underlying alignment or it is used to refer to a process connected to alignment. In this study, we will use the term "alignment" to refer to the empirical observation of adaptation in interactions, without implying any assumption about underlying mechanisms. Alignment occurs on many linguistic levels, for example the syntactic level (e.g., Branigan, Pickering & Cleland, 2000), the prosodic level (e.g., Levitan & Hirschberg, 2011), the phonetic level (e.g., Pardo, 2006) and the lexical level (e.g., Brennan & Clark, 1996). In this study, we contribute to the study of the phenomenon at the syntactic level.

2.1.1 Syntactic priming within and between speakers

A first step in the study of syntactic alignment was taken by Bock (1986). In her seminal study, she investigated the influence of a syntactic structure in a so-called prime sentence on the syntactic structure of a sentence that was produced later (the so-called target sentence) by the same participant. She found that speakers were more likely to use a certain syntactic structure (double object versus prepositional dative and active versus passive) after having used this structure in the prime sentence preceding the target sentence. Other experiments in the same study confirmed that his effect was driven primarily by the abstract syntactic structure of the prime sentence. While Bock's study was carried out with single participants speaking by themselves (i.e., both prime and target sentences were produced by the same participant), Bock and other researchers (e.g., Bock & Griffin, 2000; Branigan et al., 2000; Branigan, Pickering, McLean & Cleland, 2007; Branigan, Pickering, Pearson, McLean & Nass, 2003; Schoot, Hagoort & Segaert, 2019; Reitter & Moore, 2007) have also investigated the phenomenon in situations with primes being produced by an interlocutor, for instance in dialogues. In these situations, the phenomenon can be referred to as "alignment".

Since Levelt and Kelter's (1982) first study on alignment in dialogue, syntactic alignment has been investigated in a variety of studies. Most of these studies provide support for alignment in dialogue (e.g., Branigan et al., 2000, 2003, 2007; Schoot et al., 2019; Reitter & Moore, 2007). Even though syntactic alignment seems to be a rather robust phenomenon, there are a few things to note. Mahowald and colleagues (2016) showed in a meta-analysis of 73 studies that the observed effect sizes of syntactic priming studies, although robust, are often small. In line with this finding, Chia and colleagues replicated and expanded Levelt and Kelter's study (Chia, Axelrod et al., 2019; Chia, Hetzel-Ebben et al., 2020), however with somewhat smaller effects than those observed in the original study.

Furthermore, the most researched syntactic structure is the dative alternation (Double Object vs. Prepositional Object; e.g., Bernolet & Hartsuiker, 2010; Branigan et al., 2000, 2003, 2007; Jaeger & Snider, 2013; Weatherholtz, Campbell-Kibler & Jaeger, 2014), often investigated in a picture description task (e.g., Branigan et al., 2000, 2003, 2007). Even though this syntactic alternation lends itself well to study the phenomenon, it would be important to further consolidate the findings with other syntactic alternations. Similarly, the largest share of syntactic alignment research is done on English (e.g., Branigan et al., 2000, 2003, 2007; Reitter & Moore, 2007), while studies in other languages are less prevalent. Taken together, these points indicate that syntactic alignment should also be investigated in other languages than English, with syntactic structures that have been less well studied, and in other contexts to see how robust the phenomenon is in other languages and in other syntactic alternations.

2.1.2 Short-term versus long-term effects

Having established the basic phenomenon of syntactic alignment, the question arises whether these alignment effects are only short-lived or whether they can also occur over long time intervals. Most of the previously mentioned studies have provided evidence for short-term alignment, and less so for long-term alignment. Short-term alignment refers to the situation in which prime and target are directly adjacent (or very close). The evidence for long-term alignment effects appears to be less clear. It should be noted here that the term long-term alignment is used to refer to two different variants of long-term effects. Both of these versions build on the same underlying principle, namely the continued activation of a syntactic structure. The version mostly investigated in the literature refers to the non-adjacency of prime and target, i.e., to the question whether alignment effects can also be induced when prime and target are not (directly) adjacent to each other, but rather separated by other language material (referred to as long-term adjacency alignment hereafter). We will discuss some studies examining the distance between prime and target, i.e., long-term adjacency alignment in the following paragraphs. The other version, the one that is hardly investigated in the literature but which plays a role in this paper, refers to the question of how long an alignment effect, once being induced, will remain active, i.e., it concerns the (temporal) decay of alignment effects after they have been established (referred to as long-term persistency alignment hereafter). The main difference between the two long-term alignment effects is thus that in long-term adjacency alignment, there is no established alignment yet, whereas in the case of long-term persistency alignment the question is how long alignment persists after it has been established.

A study looking into long-term adjacency alignment versus short-term alignment is by Szmrecsanyi (2005). Szmrecsanyi showed that, in several corpora, the further apart two elements in different utterances (produced by either the same or a different

interlocutor) are situated, the lower the chance that they will be aligned. This study thus suggests that activation from a syntactic structure decays quickly with larger distances between prime and target.

However, Bock and Griffin (2000) showed in two picture description experiments that these effects do not seem to decay with larger distances between prime and target. Prime and target do not necessarily need to be adjacent and priming effects can sustain over rather long distances between prime and target. More specifically, in their first experiment, they showed that structural priming occurs when there are two sentences intervening between prime and target, and in their second experiment, they showed that syntactic priming can even span over ten intervening sentences between prime and target. Thus, this study suggests that activation of a syntactic structure in a priming sentence is long-lived.

Combining both findings, Reitter and Moore (2007) propose that both short-term priming and long-term adaptation effects can manifest in dialogues. The authors set out to investigate Pickering and Garrod's (2004) theory that alignment is caused by priming and leads to communicative success. In their first analyses of the HCRC Map task corpus, they found that there was indeed short-term alignment. Furthermore, in a re-analysis of the data specifically looking at the decay of priming effects, they found that long-term adjacency alignment also occurred. They therefore propose that there are two different mechanisms underlying priming in dialogues, one autonomous mechanism, possibly like the classical priming effect, for short-term priming and one for long-term adaptation that could be closer to Chang and colleagues' (2006) implicit learning theory and that helps interlocutors in aligning their situation models.

In summary, some studies suggest that alignment effects only span short distances (e.g., Szmrecsanyi, 2005), while other studies suggest that alignment effects can span as far as over ten sentences (e.g., Bock & Griffin, 2000). One study, the Reitter and Moore (2007) study, even suggests both effects could manifest in dialogues. All of these studies look at what we have called long-term adjacency alignment, i.e., the distance between prime and target sentence. We do not know of any study focusing on long-term persistency alignment, i.e., investigating how long established alignment effects can last. As mentioned before, both types of long-term alignment assume that some activation of a syntactic structure does not decay quickly, but rather can persist for some time. However, investigating long-term persistency alignment will help us gain insight into how long already established alignment effects can last and can maybe overrule other processes such as short-term alignment. In the experimental design used in the present study in which a participant interacts with two different interlocutors, both short-term alignment and long-term persistency alignment will play a role, which will allow us to look into the interplay between these two types of alignment.

2.1.3 Long-term alignment to different interlocutors

The majority of studies on syntactic alignment have looked into situations in which a participant interacted with one other participant (the other participant being either a confederate or a computer). To our knowledge, only two studies have investigated syntactic alignment in situations with more than two interlocutors: Branigan et al. (2007) and Ostrand and Ferreira (2019). Branigan and colleagues studied short-term alignment (turn-by-turn) effects while varying the participant's role in the experiment: the participant was either the addressee or a side-listener (i.e., simply listening to an exchange between two other speakers). The results showed that participants aligned to the syntactic structure of the input, irrespective of their role as addressee or side-listener. Anticipating on our study (for details see below), two points should be noted. First, in the Branigan study, the participant interacts with two confederates who are both in the same space, thus receiving all input. By contrast, in our study, the participant is successively confronted with only one of the two different interlocutors with these two interlocutors providing different syntactic structures as priming input. Second, the Branigan study looks at short-term priming while our study will look at both short-term priming and long-term (persistency) alignment.

Ostrand and Ferreira (2019) conducted a study that is closer to our present study. First, a confederate A described a set of pictures to the participant who was instructed to sort pictures according to the descriptions provided by the confederate. Then, this procedure was repeated with a new confederate B who again described a set of pictures to the participant. The two confederates consistently used different syntactic structures in their descriptions. Finally, one of the confederates returned and the participant now described a set of pictures to this confederate, after which the same happened with the other confederate. The results showed that the participant's descriptions were not following the syntactic format that had been used previously by the respective confederate, but rather followed an overall partner-independent statistical distribution. Alignment caused solely by the presence of a certain interlocutor, without this interlocutor producing a target syntactic structure, will be referred to as interlocutor-induced alignment hereafter. The Ostrand and Ferreira study resembles our study in that it makes use of two different confederates consistently using different syntactic structures. It differs, however, in that the participants in the Ostrand and Ferreira study do not interact with the confederates on a turn-by-turn basis while they do so in the present study.

2.1.4 The present study

The main goal of the present study is to investigate the effects of a change of interlocutor on syntactic alignment. The studies mentioned above on alignment in situations with more than one confederate either looked at short-term alignment to the same prime-

input under varying roles of the participant (Branigan et al., 2007; addressee or side-listener) or at long-term alignment to two different confederates without turn-by-turn interaction between participant and confederate. In contrast to these studies, in our study participants actually interact (in a turn-by-turn way, i.e., utterances alternatingly produced by participant and confederate) with two different confederates who consistently use different syntactic structures. We expect that interaction with the first confederate will result in (short-term turn-by-turn) alignment. We can then, when the participant switches to the other confederate, observe whether alignment to this new confederate is affected by the preceding alignment to the first confederate (i.e., potential long-term persistency effects on short-term alignment to the second confederate). When participants then switch back to the interlocutor they have interacted with before, we will be able to investigate potential interlocutor-induced alignment. As we will show in more detail below, this will thus allow us to look into the interplay between short-term and long-term persistency alignment, into the reinstatement of a previously established alignment, and into possible interlocutor-induced alignment. More specifically, we will address three questions.

RQ1: Short-term alignment in Dutch: Do speakers align to a syntactic structure produced by a confederate in a relatively short time span?
RQ2: Interplay of short-term and long-term persistency alignment effects: How does long-term persistency alignment affect the short-term alignment to an alternative syntactic structure produced by a different/new confederate?
RQ3: What are the limits of long-term alignment?

a) Interlocutor-induced alignment - After alignment to a given syntactic structure produced by a given confederate has taken place: Can this alignment be re-instantiated by just hearing this interlocutor, even if this interlocutor does not produce the syntactic structure of interest?
b) Is there a general persistence of alignment? That is, does alignment persist when speakers are no longer interacting with the confederate to whom they previously aligned (i.e., long-term persistency alignment)?

As syntactic target structure, we selected a Dutch structure – the combination of the auxiliary and the past participle - that allows for two word order variants in subordinate clauses, which are both perfectly grammatical (e.g., *Het rapport van het jongetje toonde aan dat hij zijn best had gedaan/gedaan had.*, "The little boy's report card showed that he had done his best."). Dutch subordinate clauses have the verb in clause final position, after the clause's subject and (potential) object. In case of a participle in the subordinate clause, the finite verb is an auxiliary verb, forms of *zijn,* "to be", *hebben,* "to have", or less frequently *worden,* "to be/become". The two possible word orders

(finite verb – participle or participle – finite verb) are both valid and correct options in any subordinate clause in Dutch and only know some regional preferences. In Dutch linguistics, the order where the auxiliary verb is placed before the participle is called the "red order" and the other order where the auxiliary verb is placed after the participle is called the "green order" (e.g., Haeseryn, 1990). Participants with a preference for the red order (established in a pre-test) were selected for this study.

The main experiment consisted of a sentence completion task. As illustrated in Figure 2.1, in all parts of the experiment, except for the pre- and post-test where participants completed sentences by themselves, participants interacted with confederates' pre-recorded speech. In these interaction parts, participants were presented with the beginning part of a sentence on a screen in each trial and heard that sentence beginning together with its completion to a full sentence, spoken by the confederate. They then saw a beginning part of a new sentence on the screen, read out loud this beginning part, and completed it with the first completion that came to their minds. On critical trials, sentence beginnings were constructed such that they were likely to elicit the syntactic target structure of interest (called target structure hereafter). Participant and confederate always strictly alternated in the task to complete sentence beginnings (except in the pre- and post-test, where there was no confederate).

There were two confederates, Confederate 1 and 2. Confederate 1 consistently used the green order, i.e., the order that did not have the participant's preference. Confederate 2 consistently used the participant's preferred syntactic order, the red order. The parts of the experiment where participants were presented with the target syntactic structure, are called Rounds (i.e., Round 1, Round 2 and Round 3). Parts of the experiment where

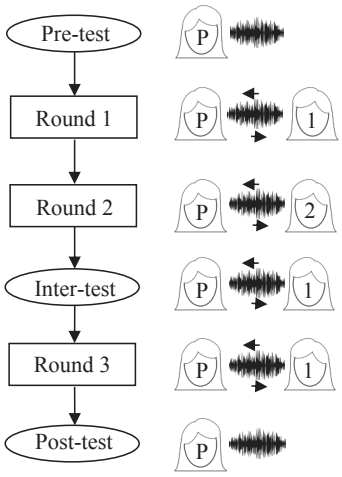


Figure 2.1. Procedure, P = Participant, 1 = Confederate 1, 2 = Confederate 2
For the differences between the rounds, as well as the pre-, inter- and the post-test, see text.

participants are not presented with the target syntactic structure, are called tests (i.e., pre-test, inter-test and post-test).

The study began with a pre-test, that was implemented as a baseline, to determine participants' preferred syntactic order. On the basis of this pre-test baseline, we selected participants with a preference for the red order. In Round 1, the participant interacted with Confederate 1, who consistently produced the participant's unpreferred syntactic order, the green order. If short-term alignment occurs, we expect participants to diverge from their preferred red order and use more of the green order in this round as compared to the pre-test (see RQ1). Round 2 was similar to Round 1, except that participants were now presented with a different confederate, Confederate 2, who consistently used the participant's preferred syntactic order, the red order. In the case of a dominant effect of short-term alignment, we would expect participants to use more of the red order in this round than in Round 1. If long-term persistency alignment can overwrite short-term effects, we would expect to not find a difference between Round 1 and 2 and a difference between the pre-test and Round 2 (RQ2). A so-called inter-test was implemented after Round 2 to see whether participants would switch back to their non-preferred syntactic order by just speaking to Confederate 1, without being presented with this order (i.e., interlocutor-induced alignment; RQ3a). To allow for this to be tested, Confederate 1 only produced filler prime sentences in the inter-test that did not contain the target structure, while the participant completed sentence beginnings of which some should elicit the target structure. In this inter-test, we test for both interlocutor-induced alignment and the longevity of the alignment effects. If alignment effects are interlocutor-induced, we expect participants to switch back to the green order, as this was the order used by Confederate 1 in Round 1. Furthermore, if the effects are only short-term, the inter-test should not differ from the pre-test.

Round 3 was similar to Round 1, but with new materials. Thus, participants were again presented with the opposite syntactic order of their preference, the green order, by Confederate 1. This round allows us to investigate short-term versus long-term alignment effects again (RQ2), but now in a situation where participants interact with an interlocutor they have interacted with before. After Round 3, participants completed a post-test which will reveal what the resulting use of the syntactic structure of interest is at the end of all preceding manipulations (RQ3b). If there is a kind of recency effect such that the most recent alignment persists, we expect this post-test to be similar to Round 3 and different from the pre-test. This post-test, like the pre-test, required participants to read out loud and complete sentence beginnings by themselves, of which a number of sentence beginnings were likely to induce the syntactic structure of interest.

In addition to this main experiment, we conducted a control experiment on an independent sample of participants to test for possible changes in syntactic preference over the course of the experiment if the confederates do not use the target syntactic structure at all. Ideally, such a control experiment should not show any differences in

the use of the critical syntactic structure over the whole experiment. However, such differences might nevertheless occur as a result of, for example, speakers' (conscious or unconscious) tendency to introduce (or to avoid) variability in their utterances. The main procedural differences between the main experiment and the control experiment were that, in the control experiment, participants were not presented with the syntactic target structure in the prime sentences and that participants did not *hear* the speech of the confederates, but instead *read* all the sentence beginnings and their completions from the screen. This latter difference was implemented to allow for the use of this control experiment in an independent study testing for phonetic/phonological alignment.

2.2 Methods

2.2.1 Participants

Seventy-two female native speakers of Dutch participated in the study at the Radboud University Nijmegen in the Netherlands.[2] Participants were included who showed a preference for the red order (auxiliary verb followed by participle) or did not have a preference (one participant in the main experiment and three in the control experiment) in the pre-test, who completed more than half of the target sentences in each part of the experiment with the target structure, and of whom good audio quality was recorded.

Half of the speakers participated in the main experiment. These speakers were aged 18 to 26 years (M = 22.4, SD = 2.0). The other half participated in the control experiment. They were aged 18 to 30 years (M = 21.4, SD = 2.8). All participants were compensated for their participation. None of the participants reported any serious speech, language or hearing impairments that could be relevant for the task. Ethical approval for this study was obtained from the Ethics Assessment Committee Humanities of the Radboud University in the Netherlands (number 6237).

2.2.2 Materials

2.2.2.1 Stimuli

Stimuli consisted of four different types: experimental prime sentences, experimental target sentences, filler prime sentences and filler target sentences. The two confederates recorded the complete prime sentences. Experimental prime sentences consisted of a main clause and a subordinate clause with the syntactic structure of interest in the red or the green order. The participants' stimuli consisted of target sentence beginnings which

2 We only selected female participants to partake in this study to exclude possible influences of gender (e.g., Reichel, Beňuš & Mády, 2018).

had to be completed by the participant. Each experimental target sentence consisted of only a main clause that was likely to evoke a completion with a subordinate clause containing the target syntactic structure. The latter aspect was tested and confirmed in a pilot study. In three separate pilot studies, a total of 14 participants filled in the missing parts of a total of 160 sentence beginnings. Sentence beginnings with the highest number of completions that included the target structure were selected for this study. The main clauses of the experimental prime and target sentences were similar in length and grammatical structure. Examples of experimental prime, experimental target, filler prime and filler target sentences are shown in Table 2.1, where the parts in italics indicate what participants saw on the screen. The parts between brackets indicate what participants heard, but did not see on the screen during the main experiment. In the control experiment, participants did see the parts between brackets (i.e., the sentence completion from the confederate) appear on the screen after seeing the parts in italics, but without hearing anything.

The total stimulus set, for both participants and confederates together in the main and control experiment, contained 502 sentences. Of these sentences, 466 were used in the main experiment. Of the 466 stimuli, 36 were experimental prime sentences, 57 experimental target sentences, and a total of 373 filler prime and target sentences. In the control experiment, the 36 experimental prime sentences were replaced with 36 new filler prime sentences. Participants completed a total of 268 sentences (60 per round, 35 in both the pre- and post-test and 18 in the inter-test), of which 57 were experimental target sentences[3] (12 in each round and 7 in each test). The confederates' stimuli consisted of 198 complete Dutch sentences, of which 36 were experimental prime sentences (12 in each round). Part of the stimuli, that is, 205 stimuli (experimental prime, target and filler prime and target sentences), were adapted from Hartsuiker and Westenberg (2000).

For the main experiment, where participants could hear the confederate's speech, we recorded all the confederates' stimuli in a sound attenuated booth with a Sennheiser K6/ME 64 microphone connected to a pre-amplifier and a Roland R-05 recorder. The confederates' speech was digitised at a sampling rate of 44.1 kHz and a 16-bit quantisation. Intensity of the sentences was normalised to 57 dB before implementing them into the Presentation software (Version 20.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). The confederates were two Dutch female speakers of similar age as the participants (23 and 24 years at the time of recording).

3 There were two sets of target sentences, one set that was randomised over the pre-, inter- and post-test and one set that was randomised over the three rounds.

Table 2.1. Examples of an experimental prime and target sentence and a filler prime and target sentence. The part in italics indicates what participants saw on the screen. The parts between brackets show the sentence completion by the confederate.

Experimental prime	*Het rapport van het jongetje toonde aan dat…* [hij zijn best had gedaan] *"The little boy's report card showed that…* [he had done his best]"
Experimental target	*De voetballer liep juichend naar zijn supporters nadat…* *"The football player walked to his supporters while cheering after..."*
Filler prime	*De boer ontkende dat…* [hij zijn koeien niet goed verzorgde] *"The farmer denied that...* [he did not take good care of his cows]"
Filler target	*Als het dit weekend weer zulk mooi weer is gaan we...* *"If the weather is this nice again this weekend we are going to…"*

2.2.2.2 Lists

Six pseudo-randomised lists were created in order to ensure that sentences did not appear in the same order for all participants and that the sentences and sentence beginnings would occur in different parts of the experiment (i.e., Round 1, 2, 3, pre-test, post-test, inter-test) for different participants. The lists were constructed in the following steps. In the first step, experimental prime-target pairs were made to create the first three lists. This step was later repeated to create another three lists. Experimental prime-target pairs were selected such that the stimuli within each pair were likely to have different auxiliary verbs (e.g., *hebben* in the prime and *zijn* in the target). The auxiliary verb to be expected in the participant's completion of the experimental target sentence beginning was established in the pilot studies. Experimental prime-target pairs were always separated by either three, four or five filler prime-target pairs. The pre- and post-test exclusively contained experimental and filler target sentences. These experimental target sentences were also separated by either three, four or five filler target sentences. The inter-test only contained experimental target sentences and filler sentences (but no experimental prime sentences). The experimental target sentences in the inter-test were separated by either one filler target sentence and two filler prime sentence or two filler target sentences and three filler prime sentences. The experimental prime-target pairs of a given round in list 1 were then transferred to a different round in list 2 and 3 (e.g., prime-target pairs of Round 1 in list 1 were assigned to Round 3 in list 2 and to Round 2 in list 3). This ensured that, across lists (and thus across participants), each experimental prime-target pair contributed to all three rounds. A parallel method was used for the assignment of experimental target sentences in the pre-, inter- and post-test to the lists. Filler prime and target sentences were randomly inserted into the lists, taking into account the position of the experimental sentences; new randomisations were created per list.

After the creation of list 1 through 3, new experimental prime-target pairs were generated by a reassignment of primes to targets under the same restrictions as for the

prime-target pair formation for list 1 to 3. These new prime-target pairs were then used to create list 4 to 6 in the same way as described for list 1 to 3.

For the control experiment, the 36 filler prime sentences, which replaced the experimental prime sentences, were moved across the lists in the opposite direction from the experimental target sentences, as opposed to being moved as pairs as was done in the main experiment (e.g., primes of Round 1 in list 1 were assigned to Round 2 in list 2 and Round 3 in list 3, instead of Round 3 in list 2 and Round 2 in list 3).

2.2.3 Procedure

2.2.3.1 Main experiment

Participants were tested in a sound attenuated booth. Their speech was recorded with the same equipment as was used to record the confederates' speech. Sentences were presented to participants in Times New Roman, font size 34, centered on the screen. After completing the pre-test by themselves, participants were presented with a prime sentence beginning on the screen, and, after approximately a second, heard this sentence beginning being read and completed by the confederate. This was implemented in order to avoid participants having enough time to complete the sentence in their minds before hearing the confederate produce the sentence and thus to prevent possible self-priming of the target structure during the processing of the experimental prime sentence. Participants were told to rate the confederate's completion of each sentence on a 7-point Likert scale on whether they would complete the sentence similarly. This scale was presented on the screen together with the sentence "I would complete this sentence in the same way" in Dutch. We hoped that this rating encouraged participants to pay attention to the confederates' sentences.

On the next trial, participants saw a new sentence beginning on the screen and had to complete that beginning (indicated by the sentence "Complete this sentence" in Dutch). Participants were told that the confederates would also rate their sentence completions. Participant and confederate strictly alternated completing sentences during the whole experiment (except for the pre- and post-test, where there was no confederate). During the parts of the experiment with the confederates (i.e., Rounds 1, 2 and 3 and the inter-test, see Figure 2.1 above), participants saw a photo of a young woman, one for each confederate, to ensure they noticed the change in confederate.

After having completed the experiment, participants filled in a short questionnaire in Qualtrics about demographics (e.g., age, education etc.), the likeability of the confederates and about their own accent. The questions about the likeability of the confederates consisted of three questions to be rated on a 7-point Likert scale: confederates' appearances, confederates' voices, and confederates' accents. After answering the questions about the confederates, participants were instructed to rate

how proud they are of their own accent.[4] The likeability and accent data will not be used in this study. The overall duration of the experiment was approximately an hour.

2.2.3.2 Control experiment

The procedure of the control experiment was identical to that of the main experiment except that participants read the complete filler prime sentences (rather than only the sentence beginning), and did not receive any auditory input. They thus first saw the sentence beginning on the screen and, after 2000 ms, they saw the second part of that sentence (i.e., the sentence completion). As in the main experiment, participants wore headphones, to ensure minimal differences between the control experiment and the main experiment.

2.2.4 Statistical analysis

We analysed the data with generalised linear mixed effects regression models (GLMER) with the binomial link function. R version 4.0.2 (R Core Team, 2020) was used to test these models including the lme4 package version 1.1-23 (Bates, Maechler, Bolker & Walker, 2015) and the car package version 3.0-8 (Fox & Weisberg, 2019). The ggplot2 package version 3.3.2 (Wickham, 2016) and the ggsignif package version 0.6.0 (Ahlmann-Eltze, 2019) were used for visualisation.

The dependent variable of the models was the participant's syntactic choice for a given experimental target sentence (either the red, coded as 1, or the green order, coded as 0; very rare cases where participants used both orders in one sentence were excluded, this accounted for 0.16% of the total number of target sentences that included the target structure in the main experiment, and 0% in the control experiment). The independent variables were Experiment (factor consisting of two levels, the main experiment and the control experiment) and Experiment part (indicating the different rounds and tests). The interaction between these two independent variables (which tests for the critical differences between main experiment and control experiment) was added to the model. Random intercepts for Participant and Item (indicating the sentence beginning) were fitted and random slopes for Experiment part by Participant were also added. This model did not converge, and the random slopes were taken out, leaving the random intercepts for Participant and Item. This model did not converge either, and the optimiser was changed to Bobyqa with 100,000 iterations to ensure convergence.

4 Rating of the accents was done in order to control for possible differences in phonetic alignment, described in Chapter 5.

2.3 Results

Only experimental targets were analysed, thus excluding rare occurrences of the use of the target structure in filler target sentences. Table 2.2 shows the results of the GLMER-model with the pre-test and the main experiment as reference levels (on the intercept). For comparisons between other experiment parts, we releveled the model. Results from those models can be found in Appendix A. A threshold of 1.96 for Z-scores was used for significance testing. Effects are thus seen as statistically significant for Z scores above 1.96 or below -1.96.

Table 2.2. Estimates, standard errors, z-values and p-values with the pre-test on the intercept.

Parameter	Estimate	SE	Z Value	P
Intercept	2.49	0.35	7.03	<0.001
Control experiment	-0.16	0.44	-0.36	0.723
Round 1	-0.65	0.33	-1.98	<0.05
Round 2	0.24	0.34	0.72	0.473
Inter-test	-0.29	0.29	-1.03	0.305
Round 3	-0.77	0.33	-2.35	<0.05
Post-test	-0.57	0.28	-2.06	<0.05
Control experiment * Round 1	-0.26	0.36	-0.73	0.468
Control experiment * Round 2	-1.24	0.37	-3.37	<0.001
Control experiment * Inter-test	-0.18	0.40	-0.44	0.657
Control experiment * Round 3	-0.01	0.36	-0.03	0.976
Control experiment * Post-test	-0.14	0.39	-0.35	0.724

Descriptive data from the main experiment and the control experiment are presented separately in Figures 2.2 and 2.3, respectively. The proportion of the red order is shown per experiment part (i.e., Round 1, 2, 3, pre-test, post-test, inter-test); the brackets indicate the significant differences between experiment parts.

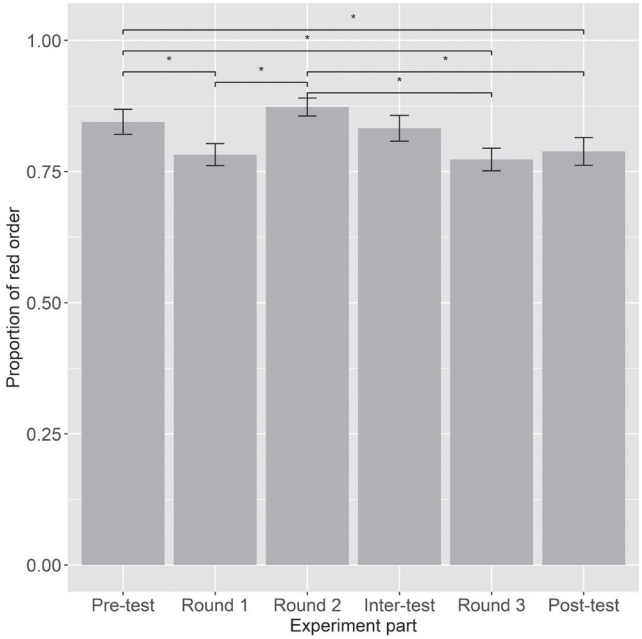


Figure 2.2. Proportion of the red order per experiment part in the main experiment, including error bars.

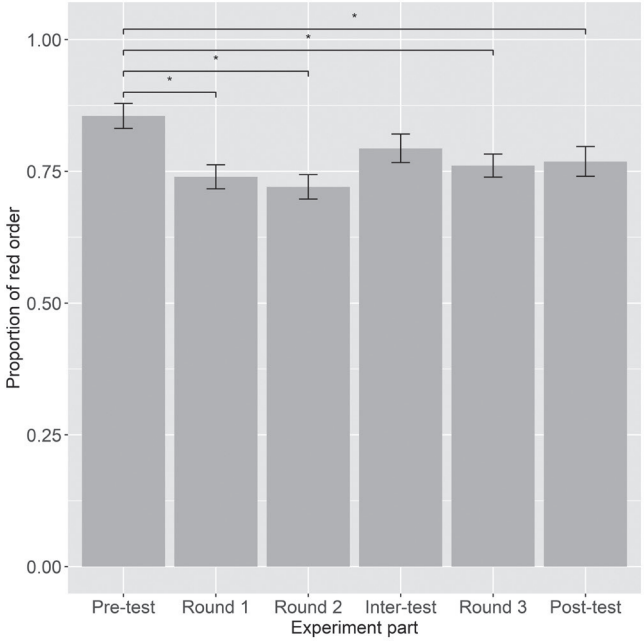


Figure 2.3. Proportion of the red order per experiment part in the control experiment, including error bars.

Figures 2.2 and 2.3 show that participants begin the experiment with a clear preference for the red order. These proportions are similar for the main experiment (84%) and the control experiment (86%) and confirmed by the lack of statistical difference between the main and the control experiment - for the pre-test - as shown in Table 2.2. As we selected participants on the basis of their preference for the red order in the pre-test, this shows that the two groups have indeed similar preferences.

After the pre-test, participants completed Round 1. Figure 2.2 shows that the proportion of the red order declines, with a corresponding increase in the use of the green order, in the main experiment, where participants interacted with a confederate only using the green order. This could be seen as syntactic alignment. However, Figure 2.3 shows that participants also used more of the green order in Round 1 of the control experiment. This means that when participants are not being presented with the green order, as in the control experiment, they still use it more often in Round 1 than in the pre-test. This is confirmed by the significant difference between the pre-test and Round 1 for the main experiment and the lack of a statistically significant interaction between Round 1 and Experiment (see Table 2.2): Round 1 differs from the pre-test in the main and control experiment to a similar extent.

After Round 1, participants completed Round 2. We see a clear increase in the use of the red order in Round 2 compared to Round 1 in the main experiment, where participants now interacted with a confederate only using the red order. This difference between Round 1 and Round 2 is not present in the control experiment, indicating syntactic alignment in Round 2 relative to Round 1 in the main experiment. These descriptive trends in the figures are confirmed by the statistical analyses. There is a significant difference between Round 2 and Round 1 for the main experiment, while a corresponding difference is not present in the control experiment, as indicated by the interaction effect of the experiment and Round 2 as tested with the releveled model with Round 2 on the intercept (see Table A1 in Appendix A). These results - for the main experiment, but not for the control experiment - indicate a return to the originally preferred structure (from the pre-test) after a deviation from this preference in Round 1.

In the next part of the experiment, the inter-test, participants once more interacted with Confederate 1, who now did not produce the target syntactic structure, neither in the control experiment nor in the main experiment. As shown in Figure 2.2, in the main experiment, the proportion of the red order produced by the participants in the inter-test lies between the proportions of Round 2 and Round 1. Thus, it appears that the mere reappearance of Confederate 1 is not enough for participants to fully return to the level of the proportion of the red order of Round 1. However, the proportion is also not the same as that of Round 2, where the other confederate used the red order. This descriptive pattern goes together with a lack of statistically significant differences between the inter-test and both Round 1 and 2 when the intercept represents the inter-test (see Table A2 in Appendix A).

Figure 2.3 suggests that the control experiment shows a slightly different pattern, which is supported by the statistics in Table A2. The participants in the two experiments hardly differ in their use of the red order in this part of the experiment, in accordance with the lack of a statistically significant effect of Experiment in Table A2. They also did not differ in that their use of the red order was similar in the inter-test compared to Round 1. However, we found a statistically significant interaction of Experiment with Round 2 with the inter-test on the intercept, showing that the difference between the inter-test and Round 2 was different between the Control and the Main experiment. While in the main experiment this difference was positive (implying more occurrences of the red order in Round 2 than in the inter-test), this difference was negative in the control experiment (implying more occurrences of the red order in inter-test). However, the difference between the inter-test and Round 2 was statistically significant in neither experiment (see Table A3 in Appendix A for the statistics of the control experiment).

In Round 3, participants continued to interact with Confederate 1, who continues to not use the target structure in the control experiment, but uses the target structure with the green order again in the main experiment (i.e., as in Round 1). We will look at the results in Round 3 from two different perspectives: a comparison between Round 3 and the pre-test to investigate differences between the baseline use and the use when interacting with Confederate 1, and a comparison between Round 3 and Round 2 to see whether participants behave differently when interacting with Confederate 1 as opposed to Confederate 2. First, as Figure 2.2 and Table A4 show, not only the participants in the main experiment but also those in the control experiment showed a different word order pattern in Round 3 than in the pre-test. However, a clear differential pattern is visible between the main and control experiment when comparing Round 3 to Round 2. Figure 2.2 shows that participants in the main experiment used more of the green order as a response to the use of the green order by the confederate, as the proportion of red order is lower again in Round 3 as compared to the preceding parts of the experiment in which they did not hear the green order, the inter-test and Round 2. This proportion is similar to the proportion in Round 1, where the participants interacted with the same confederate using the same structure. In the control experiment, in contrast, participants produced a proportion of green orders in Round 3 that was very similar to those of the other rounds. These patterns in the figures are statistically supported by a significant effect of Round 2 for the main experiment (see Table A4 in Appendix A, with Round 3 on the intercept), which is absent for the control experiment as indicated by the interaction of Round 2 with the control experiment, with a beta that is opposite of the simple effect of Round 2.

Lastly, participants completed a post-test similar to the pre-test (no interaction with any confederate). We see that, in the main experiment, the proportion of the use of the red order remains similar to that in Round 3. This is confirmed by the lack of a

significant difference between Round 3 and the post-test (see Table A5 in Appendix A). Furthermore, we see that the proportion of red order used is statistically lower in the post-test than in the pre-test. The same holds for the control experiment, as we do not see any interaction effects between the main and control experiment and the post-test on the one hand and both Round 3 and the pre-test on the other hand. This means that the main experiment and the control experiment do not differ with respect to these parts.

2.4 Discussion

This study investigated syntactic alignment effects in a situation where participants interact with two different interlocutors. We had three main goals. First, we aimed to expand the available database on short-term syntactic alignment to Dutch, to an understudied syntactic structure. Second, we wanted to find out how long-term persistency alignment to one interlocutor interacts with subsequent short-term alignment to a second interlocutor. Third, we investigated the limits of this long-term persistency alignment.

We ran an experiment in which participants interacted with two different confederates in Dutch. The confederates differed in their word order preference in a Dutch syntactic structure. The first confederate always used the syntactic order (the green order) that was the opposite of the participants' preferred order. After having interacted with the first confederate in Round 1, in Round 2, participants interacted with the second confederate, who used the participant's preferred syntactic order (the red order). After Round 2, participants interacted with the first confederate again in the inter-test and in Round 3. In the inter-test, participants were not presented with the target syntactic structure. After Round 3, in which participants were presented with their unpreferred syntactic order (the green order) again, they completed a post-test in which they did not interact with a confederate. We also conducted a control experiment that was similar to the main experiment, but did not include any experimental prime sentences or speech.

Turning to our first research question, evidence for short-term syntactic alignment in Dutch is not visible in the comparison between the pre-test and Round 1; a difference in the to-be-expected direction was present, but it was present in both the main and the control experiment. In contrast, a differential effect between the main and the control experiment emerged clearly in Round 2 (the round with Confederate 2). There, we see a difference between Round 2 on the one hand and both Round 1 and Round 3 on the other hand (the rounds with Confederate 1) in the main experiment, while there is no

corresponding difference in the control experiment. The difference between Round 2 and 1 reflects the alignment to the red order (the participants' preferred order) used by Confederate 2, after the participant had interacted with Confederate 1 in Round 1, who used the participants' unpreferred structure. Furthermore, the difference between Round 2 and 3 suggests that participants also aligned to the unpreferred order in Round 3. Importantly, while these differences were present in the main experiment, they were absent in the control experiment, supporting the conclusion that the observed effects reflect alignment. We can thus answer the first research question affirmatively. This finding further strengthens the findings in the literature on alignment in other languages, situations, and syntactic structures (e.g., Branigan et al., 2000, 2003, 2007).

The second research question concerned the interplay of short-term alignment and long-term persistency alignment. As discussed above, we did not see any short-term alignment effects in Round 1. In Round 2, however, we did see short-term alignment in the sense of a strengthening of the participants base preference, which differed from the word order presented to them in Round 1. In Round 3, we then saw a difference with Round 2, indicating that the alignment from Round 2 did not persist in Round 3, but rather was overwritten by the short-term alignment within Round 3. This indicates that the short-term alignment to an interlocutor in a round seems to cause a stronger activation of the syntactic structure than any possible long-term activation from a previous round. This finding expands studies on short-term and long-term alignment (e.g., Reitter & Moore, 2007) by indicating that short-term alignment effects seem to be able to overwrite potential long-term effects in situations where both can manifest at the same time.

The third research question concerned the limits of long-term alignment. This question was divided into two questions, one on possible interlocutor-induced alignment and another on the general persistence of alignment. Research question 3a was whether participants would switch back to an unpreferred structure in the main experiment when they were confronted with the confederate who had been using this structure before (in Round 1), without this confederate actually using that structure in the local context (i.e., in the inter-test). This research question was investigated by comparing the inter-test to the different rounds. The inter-test took place after Round 2, where participants aligned to the red order of Confederate 2 in the main experiment. If alignment were interlocutor-induced, we should see a difference between Round 2 and the inter-test in the main experiment, indicating the use of more of the green order (as used by Confederate 1 in Round 1) as opposed to the use of the red order in Round 2. However, we found that the use of the green order in the inter-test lies somewhere in between the levels of Round 2 and Round 1. The level in the inter-test was not significantly different from those of either of the rounds, and also similar to that of the inter-test in the control experiment. We therefore do not find firm evidence

of interlocutor-induced alignment in the inter-test. This finding is in line with the findings in Ostrand and Ferreira (2019), who found that speakers aligned to an overall distribution, that was independent of the confederate. Another explanation could be that participants may have returned to their baseline use of the syntactic orders, since there is no difference between the inter-test and the pre-test either.

Research question 3b was whether the alignment effects persist beyond the direct exposure to the corresponding confederate. If alignment effects would persist, we would expect participants to continue to use more of the green order in the post-test of the main experiment, as they did in Round 3. Furthermore, we would expect to find a difference between the pre-test and the post-test, to confirm a divergence from participants' preference for the red order. Such prolonged activation should have been reflected in a difference between the pre-test and the post-test in the main experiment, but not in the control experiment. Unlike Bock and Griffin (2000) and Ostrand and Ferreira (2019), we do not find clear evidence for such prolonged activation of a syntactic structure. The main experiment and the control experiment both do not show a difference between Round 3 and the post-test nor a difference between the pre-test and the post-test.

Before turning to the general conclusions, we would like to address two potential caveats. First, in the control experiment, we had expected that the proportion of the red versus the green order would remain stable across the different parts of the experiment. However, the results of the control experiment show that this is not the case. We saw that the pre-test was significantly different from almost all of the rest of the control experiment. The proportion of red order in the pre-test is unsurprising as we selected participants based on this preference. However, we cannot explain why participants in the control experiment changed their preference when moving from the pre-test to Round 1. We can only speculate on potential reasons for such a change. This variation may be due to an overestimation of some participants' preference for the red order in the pre-test, which could be due to the relatively small number of items in the pre-test. Another explanation for this difference could be a change in situation: participants interacted with a confederate in Round 1 as opposed to simply completing sentences by themselves in the pre-test. Whatever the eventual explanation of this variation in the control experiment might be, the results of the control experiment show that it is not per se justified to assume that, in a syntactic alignment experiment, variation in the use of different syntactic structures is by definition exclusively induced by the relevant experimental manipulations.

Second, in the present study, we used pre-recorded speech from confederates in order to control the speech input for phonetic analyses, not mentioned in this paper. It is possible that interlocutor-induced effects could be stronger when a real interlocutor would have been present (Schoot et al., 2019). However, studies comparing (beliefs

about) interactions with computer-based interlocutors to interactions with real human interlocutors and studies investigating monologues versus dialogues suggest that differences in alignment effects are minimal (e.g., Branigan et al, 2003; Felker, Broersma & Ernestus, 2021; Ivanova, Horton, Swets, Kleinman & Ferreira, 2020).

2.5 Conclusion

In this study we investigated three questions. The first question was whether speakers align to a syntactic structure that is relatively understudied in a short time span. Our results suggest that speakers align to two different interlocutors in a relatively short time-span in a relatively understudied syntactic structure. The second question was how long-term persistency alignment may affect short-term alignment to an alternative syntactic structure that was produced by a different confederate. Our study allows us to investigate the relative strength of short-term alignment and long-term persistency alignment and the results suggest that short-term alignment effects dominate over long-term persistence of alignment in a situation in which long-term persistency alignment could be established with one interlocutor and is subsequently overridden by short-term alignment with a second interlocutor. Finally, the third question investigated the limits of long-term alignment. Alignment effects only appeared in a situation where the relevant syntactic structure was actually produced by the interlocutor. The mere presence of the respective interlocutor (without using the syntactic structure of interest) does not suffice to trigger alignment with this interlocutor. In conclusion, this study contributes to our knowledge of syntactic alignment by bringing to the table evidence from an understudied structure and showing the relevance of short-term versus long-term persistency alignment, in the presence and absence of the interlocutor, using or not using the relevant syntactic structure.

Appendix A

Round 2 intercept

Table A1. Estimates, standard errors, z-values and p-values with Round 2 on the intercept.

Parameter	Estimate	SE	Z Value	P
Intercept	2.73	0.32	8.63	<0.001
Control experiment	-1.39	0.39	-3.59	<0.001
Pre-test	-0.24	0.34	-0.72	0.473
Round 1	-0.89	0.22	-3.98	<0.001
Inter-test	-0.54	0.33	-1.62	0.105
Round 3	-1.01	0.23	-4.49	<0.001
Post-test	-0.81	0.32	-2.51	<0.05
Control experiment * Pre-test	1.24	0.37	3.37	<0.001
Control experiment * Round 1	0.98	0.29	3.35	<0.001
Control experiment * Inter-test	1.06	0.35	3.07	<0.01
Control experiment * Round 3	1.23	0.29	4.18	<0.001
Control experiment * Post-test	1.10	0.33	3.29	<0.001

Inter-test intercept

Table A2. Estimates, standard errors, z-values and p-values with the inter-test on the intercept.

Parameter	Estimate	SE	Z Value	P
Intercept	2.19	0.34	6.40	<0.001
Control experiment	-0.33	0.42	-0.80	0.425
Pre-test	0.29	0.29	1.03	0.305
Round 1	-0.35	0.32	-1.12	0.265
Round 2	0.54	0.33	1.62	0.105
Round 3	-0.47	0.32	-1.49	0.135
Post-test	-0.28	0.27	-1.03	0.301
Control experiment * Pre-test	0.18	0.40	0.44	0.657
Control experiment * Round 1	-0.08	0.33	-0.24	0.809
Control experiment * Round 2	-1.06	0.35	-3.07	<0.01
Control experiment * Round 3	0.17	0.33	0.50	0.615
Control experiment * Post-test	0.04	0.37	0.11	0.915

Inter-test and control experiment intercept

Table A3. Estimates, standard errors, z-values and p-values with the inter-test and the control experiment on the intercept.

Parameter	Estimate	SE	Z Value	P
Intercept	1.86	0.33	5.56	<0.001
Main experiment	0.33	0.42	0.80	0.425
Pre-test	0.47	0.28	1.69	0.091
Round 1	-0.43	0.31	-1.41	0.159
Round 2	-0.52	0.31	-1.71	0.088
Round 3	-0.31	0.31	-1.00	0.319
Post-test	-0.24	0.26	-0.92	0.358
Main experiment * Pre-test	-0.18	0.40	-0.44	0.657
Main experiment * Round 1	0.08	0.33	0.24	0.809
Main experiment * Round 2	1.06	0.35	3.07	<0.01
Main experiment * Round 3	-0.17	0.33	-0.50	0.615
Main experiment * Post-test	-0.04	0.37	-0.11	0.915

Round 3 intercept

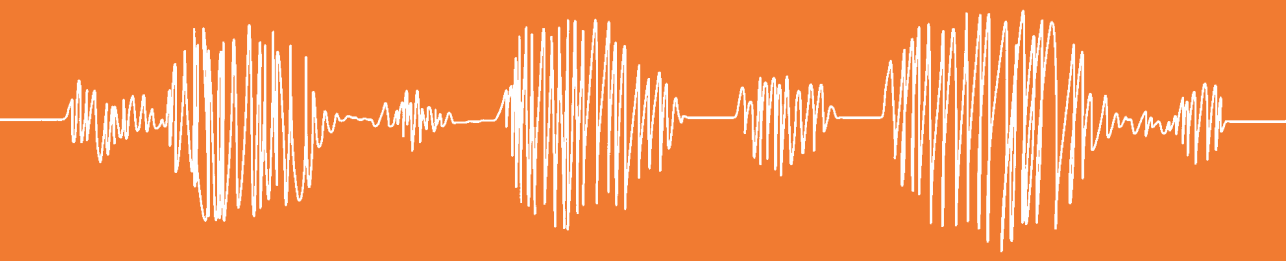
Table A4. Estimates, standard errors, z-values and p-values with Round 3 on the intercept.

Parameter	Estimate	SE	Z Value	P
Intercept	1.72	0.30	5.81	<0.001
Control experiment	-0.17	0.37	-0.44	0.657
Pre-test	0.77	0.33	2.35	<0.05
Round 1	0.12	0.20	0.60	0.548
Round 2	1.01	0.23	4.49	<0.001
Inter-test	0.47	0.32	1.49	0.135
Post-test	0.20	0.31	0.64	0.521
Control experiment * Pre-test	0.01	0.36	0.03	0.976
Control experiment * Round 1	-0.25	0.28	-0.89	0.372
Control experiment * Round 2	-1.23	0.29	-4.18	<0.001
Control experiment * Inter-test	-0.17	0.33	-0.50	0.615
Control experiment * Post-test	-0.13	0.32	-0.40	0.691

Post-test intercept

Table A5. Estimates, standard errors, z-values and p-values with the post-test on the intercept.

Parameter	Estimate	SE	Z Value	P
Intercept	1.92	0.34	5.72	<0.001
Control experiment	-0.29	0.41	-0.72	0.473
Pre-test	0.57	0.28	2.05	<0.05
Round 1	-0.08	0.31	-0.25	0.801
Round 2	0.81	0.32	2.51	<0.05
Inter-test	0.28	0.27	1.03	0.301
Round 3	-0.20	0.31	-0.64	0.521
Control experiment * Pre-test	0.14	0.39	0.35	0.724
Control experiment * Round 1	-0.12	0.32	-0.37	0.709
Control experiment * Round 2	-1.10	0.33	-3.29	<0.001
Control experiment * Inter-test	-0.04	0.37	-0.11	0.915
Control experiment * Round 3	0.13	0.32	0.40	0.691



CHAPTER 3

Alignment of Pitch and Articulation Rate

This chapter is based on:
Eijk, L., Ernestus, M., & Schriefers H. (2019). Alignment of pitch and articulation rate. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019, pages 2690-2694. doi:10.1016/j.wocn.2019.02.004.

Abstract

Previous studies have shown that speakers align their speech to each other at multiple linguistic levels. This study investigates whether alignment is mostly the result of priming from the immediately preceding speech materials, focusing on pitch and articulation rate (AR). Native Dutch speakers completed sentences, first by themselves (pre-test), then in alternation with Confederate 1 (Round 1), with Confederate 2 (Round 2), with Confederate 1 again (Round 3), and lastly by themselves again (post-test). Results indicate that participants aligned to the confederates and that this alignment lasted during the post-test. The confederates' directly preceding sentences were not good predictors for the participants' pitch and AR. Overall, the results indicate that alignment is more of a global effect than a local priming effect.

3.1 Introduction

Alignment (also often referred to as entrainment, convergence or accommodation) refers to the phenomenon that speakers adapt their speech to an interlocutor's speech on multiple levels (e.g., prosodic, phonetic, syntactic). Although alignment has been thoroughly investigated in the (recent) past, e.g., (Bonin et al., 2013; Gijssels, Casasanto, Jasmin, Hagoort & Casasanto, 2016; Levitan & Hirschberg, 2000), many empirical questions are still open.

This study investigates whether alignment is mostly due to priming from the immediately preceding speech materials by addressing three questions. (RQ1) How long does alignment persist when the interlocutor is no longer present? If alignment exclusively results from adaptation to recent input, it should disappear rapidly. (RQ2) Do speakers align more rapidly to a speaker they have been talking to before? If alignment is exclusively driven by the immediately preceding input, this should not be the case. (RQ3) Do the features of the immediately preceding utterance predict how speakers adapt their speech in a given sentence?

We investigated these questions for both pitch and articulation rate, henceforth AR. By investigating two prosodic features, we can see in how far the results are feature specific, that is, whether and to what extent different prosodic features converge or differ in their alignment patterns.

Previous research has shown that both pitch, and AR are susceptive to alignment (Bonin et al., 2013; Gijssels et al., 2016; Levitan & Hirschberg, 2000), although conflicting results have been reported for both features. For instance, research on pitch alignment by Gijssels et al. (2016) has shown that speakers align their pitch to a confederate's pitch on a turn-by-turn basis (see also Levitan & Hirschberg, 2011), that the degree of alignment does not increase over time, and that alignment disappears immediately when the confederate is no longer present. In contrast, Bonin et al. (2013) reported that pitch alignment fluctuates over time and that speakers do not always align in every turn. Research on AR alignment also shows conflicting results. For instance, whereas Levitan and Hirschberg (2011) found alignment, Schweitzer and Lewandowski (2013) found divergence in AR between speaker and interlocutor, though this effect was modulated by how much the participant liked the interlocutor.

We addressed our research questions in a sentence completion task consisting of five parts, which was originally designed to investigate other forms of alignment (phonological and syntactic). Participants first completed sentence beginnings by themselves (pre-test). Then, they alternated between sentence completion and listening to sentences completions from a confederate's pre-recorded speech. They did so, first with Confederate 1 (in Round 1), then with Confederate 2 (Round 2), and then with Confederate 1 again (Round 3). After these parts, they completed sentences by themselves again (post-test).

Our first question can be answered by comparing (the speed of change in) pitch and AR in the post-test with the other parts of the experiment. The second question can be addressed by comparing (the speed of change in) pitch and AR between Rounds 1 and 3 (the rounds with the same confederate). The third question can be addressed by testing whether the pitch or AR of a given sentence is predicted by the confederate's pitch or AR in the directly preceding utterance.

3.2 Methods

3.2.1 Participants

Twenty-five female native Dutch speakers, aged 18 to 26 years (M = 22.4, SD = 2.1) participated in the experiment. Participants received course credits or gift vouchers.

3.2.2 Materials

Two sets of materials were designed. The first set contained 268 Dutch sentence beginnings that had to be completed by the participants. These sentence beginnings were designed to elicit as much speech as possible. An example of a stimulus is shown in (1).

(1) Otto is een stuk vrolijker sinds…
'Otto has been a lot happier since…'

The second set of materials consisted of 198 complete Dutch sentences, which were uttered by the confederates and functioned as auditory primes. During the experiment, participants saw the beginnings of the confederates' full sentences on the computer screen. These beginnings were similar in length and grammatical structures to the sentence beginnings the participants had to complete. The two sets of stimuli included 205 stimuli that were adapted from Hartsuiker and Westenberg (2000).

The complete sentences were recorded by the confederates in a sound-attenuated booth with a table-mounted Sennheiser K6/ME 64 microphone connected to a pre-amplifier and a Roland R-05 recorder. Speech was digitised at a sampling rate of 44.1 kHz, a 16-bit quantisation. Confederate 1 (23-year-old female) had an average median pitch of 224 Hz (ranging from 189 to 256) and an average AR of 5.0 syllables per second (ranging from 3.4 to 6.0), while Confederate 2 (24-year-old female) had averages of 215 Hz (ranging from 193 to 241) and 4.7 syllables per second (ranging from 3.4 to 6.5), see *3.2.4* for the measurement method.

Six pseudo-randomised stimuli lists were generated to make sure that, across participants, a given sentence (beginning) appeared in different parts of the experiment.

3.2.3 Procedure

Participants were tested in a sound-attenuated booth. The participants' speech was recorded using the same equipment as mentioned above. The confederates' speech was presented over Sennheiser HD 215 MKII DJ headphones.

Participants were presented with a sentence beginning via the Presentation software (Version 20.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) in Times New Roman, font size 34, centered on the screen. They were instructed to read aloud the sentence beginning and to complete the beginning with whatever came to mind. In the pre- and post-test (both 35 trials), the participants completed the sentences by themselves. In Rounds 1 (60 trials), 2 (60 trials) and 3 (78 trials), the participants alternated with the pre-recorded speech from Confederate 1, Confederate 2, and Confederate 1, respectively. During these rounds, they saw the picture of the respective confederate on the screen.

Participants were asked to indicate for each sentence produced by the confederates, on a 7-point Likert scale, whether they would finish the sentence in the same way. This way we ensured that they paid attention to the confederates' speech. Instructions ('I would finish the sentence in the same way' plus the scale) were shown on the computer screen during confederates' trials. Participants were told that the confederates would rate their sentences as well. The experiment took less than one hour in total.

3.2.4 Measurements

Median pitch and articulation rate were calculated per sentence in Praat (Boersma & Weenink, 2018). Median pitch was calculated with a script (Marcoux & Ernestus, 2019) which measured F0 values every 10 ms by using the To Pitch... command in Praat with a pitch range of 75 to 500 Hz. The script cleaned the raw values from errors resulting in pitch doubling and halving and from values based on speech produced with creaky voice by removing F0 values that were more than a factor of 1.5 bigger or smaller than the second to last F0 value. Then, the median F0 value per sentence was calculated. We removed all sentences with a minimum F0 lower than 110 Hz or a maximum F0 higher than 400 Hz. After deletion of these outliers, outliers more than 2.5 SD from the mean were deleted, resulting in 6230 data points for analyses (93.22% of the total).

The AR per sentence was calculated with a script (De Jong & Wempe, 2009) using the following parameters: a silence threshold of -25 dB (default), a minimum dip between peaks of 3 dB and a minimum pause duration of 0.3 seconds (default). The script divides the number of syllables (based on a number of syllable-related acoustic properties) of a sentence by the vocalisation time (the total time minus pauses). Outliers more than 2.5 SD from the mean were excluded, which resulted in 6588 data points for analyses (98.58% of the total).

3.2.5 Statistical analysis

Linear Mixed Effects models were performed in R (R Core Team, 2017) using the lme4 package (Bates, Maechler, Bolker & Walker, 2015). Unless otherwise mentioned, our dependent variable was either the participant's median F0 or the AR per sentence. Fixed effects were ExperimentPart (EP) (pre-test, Round 1, Round 2, Round 3 and post-test) and EPtrialnr, which codes the sequential position of sentences within a given part of the experiment. We also tested for a potential quadratic trend of EPtrialnr, but adding the quadratic predictor did not improve the models. We further tested for an interaction of the two fixed effects. Random effects were added for participant and sentence. For the final models, we removed data points deviating more than 2.5 SD from the predicted values. No random slopes were added for participant and sentence, because this caused non-convergence.

3.3 Results

Figures 3.1 and 3.2 show the participants' median pitch and AR as a function of the trial number in the experiment. Different parts of the experiment are indicated by lines in different shades of grey. The figures also show the confederates' average pitch and AR, which were generally higher than the participants' pitch and AR.

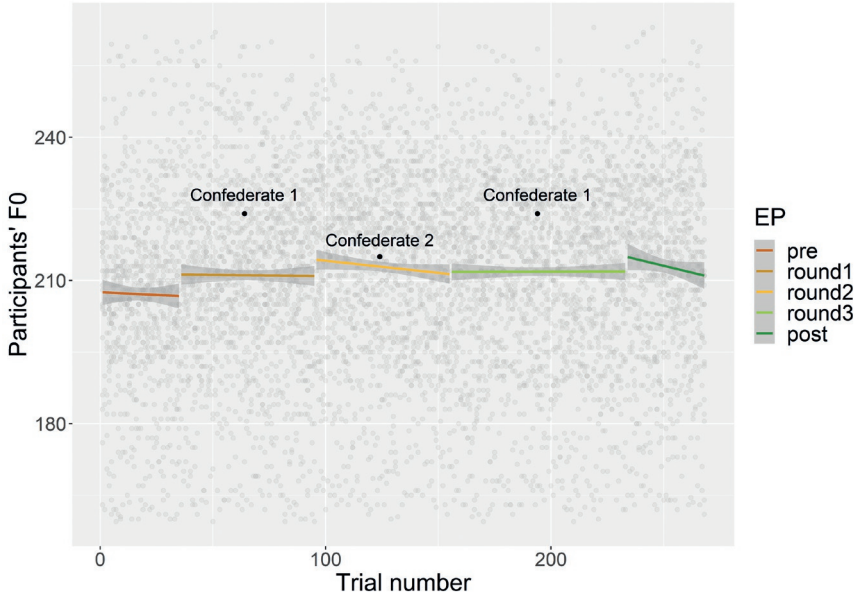


Figure 3.1. Participants' median F0 over pre-test, Rounds 1, 2 and 3 and post-test; lines were fitted using lm. Points represent Confederates' means.

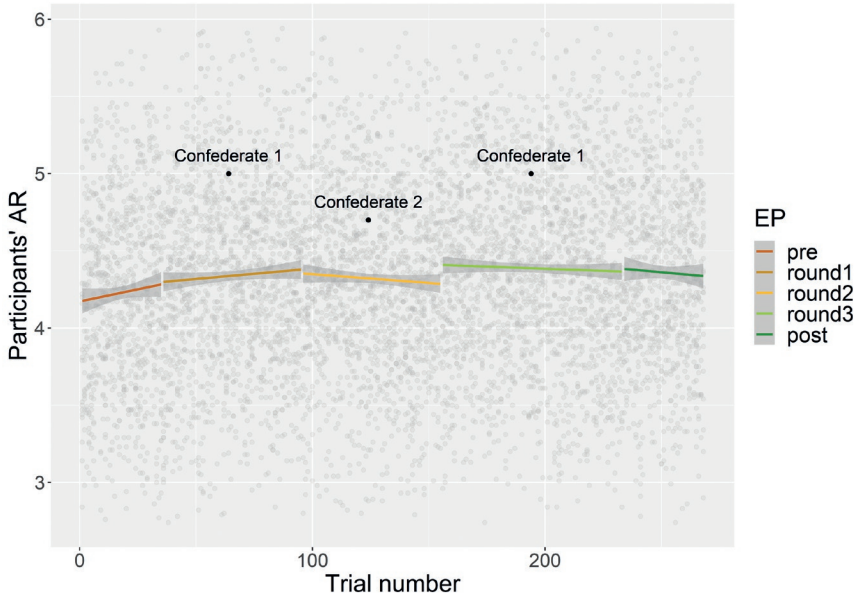


Figure 3.2. Participants' AR over pre-test, Rounds 1, 2 and 3 and post-test; lines were fitted using lm. Points represent Confederates' means.

3.3.1 RQ1: Difference between post-test and other parts

To see whether alignment lasts when the confederate is no longer present, we compared the post-test to the other parts of the experiment. If alignment lasts in the absence of the interlocutor, we would expect a significant difference between the pre-test and the post-test, reflecting that the participant's pitch and AR do not immediately return to the level of the pre-test. We would further expect no difference between Round 3 and the post-test if the alignment of Round 3 lasts in the post-test. Table 3.1 shows the results of the pitch model and Table 3.2 of the AR model, both with the post-test as the reference level.

Table 3.1. Pitch model with post-test as a reference.

Parameter	Estimate	SE	T value
Intercept	212.618	3.704	57.40
EPpre	-5.399[5]	0.869	-6.21
EPround1	-2.784	0.795	-3.50
EPround2	1.261	0.789	1.60
EPround3	-0.650	0.754	-0.86
EPtrialnr	-0.061	0.030	-2.01
EPpre:EPtrialnr	0.031	0.042	0.75
EPround1:EPtrialnr	0.084	0.033	2.52
EPround2:EPtrialnr	0.022	0.033	0.67
EPround3:EPtrialnr	0.062	0.032	1.95

Table 3.2. AR model with post-test as a reference.

Parameter	Estimate	SE	T value
Intercept	4.414	0.063	70.30
EPpre	-0.247	0.052	-4.74
EPround1	-0.128	0.047	-2.72
EPround2	-0.056	0.047	-1.18
EPround3	-0.012	0.045	-0.26
EPtrialnr	-0.003	0.002	-1.77
EPpre:EPtrialnr	0.006	0.003	2.53
EPround1:EPtrialnr	0.005	0.002	2.37
EPround2:EPtrialnr	0.002	0.002	1.16
EPround3:EPtrialnr	0.003	0.002	1.46

Tables 3.1 and 3.2 show that participants did not immediately return to their habitual median pitch and AR in the post-test, as there are statistically significant differences between the pre-test and post-test. This is further supported by the lack of significant differences between the post-test and Round 3. Furthermore, participants gradually returned to their habitual pitch in the post-test as reflected in a significant effect of EPtrialnr within the post-test. This is not the case for AR.

3.3.2 RQ2: Difference between Round 1 and Round 3

To see whether speakers aligned more rapidly to Confederate 1 in Round 3 than in Round 1, we focused on the differences between Round 1 and Round 3. If participants aligned more rapidly, i.e., within the first few trials, in Round 3 than in Round 1, this should result in an overall positive significant difference in median pitch and AR

between Rounds 1 and 3. More rapid alignment could also be reflected in a positive statistically significant difference in the effect of EPtrialnr, i.e., an interaction between EPtrialnr and Round. Tables 3.3 and 3.4 show the models of Tables 3.1 and 3.2, with Round 1 as the reference.

Table 3.3. Pitch model with Round 1 as a reference.

Parameter	Estimate	SE	T value
Intercept	209.834	3.681	57.00
EPpost	2.784	0.795	3.50
EPpre	-2.615	0.778	-3.36
EPround2	4.045	0.661	6.12
EPround3	2.134	0.636	3.35
EPtrialnr	0.023	0.014	1.68
EPpost:EPtrialnr	-0.084	0.033	-2.52
EPpre:EPtrialnr	-0.053	0.033	-1.60
EPround2:EPtrialnr	-0.062	0.019	-3.30
EPround3:EPtrialnr	-0.022	0.016	-1.33

Table 3.4. AR model with Round 1 as a reference.

Parameter	Estimate	SE	T value
Intercept	4.286	0.058	74.18
EPpost	0.128	0.047	2.72
EPpre	-0.119	0.047	-2.55
EPround2	0.072	0.040	1.83
EPround3	0.116	0.038	3.06
EPtrialnr	0.001	0.001	1.82
EPpost:EPtrialnr	-0.005	0.002	-2.37
EPpre:EPtrialnr	0.002[5]	0.002	0.86
EPround2:EPtrialnr	-0.002	0.001	-2.12
EPround3:EPtrialnr	-0.002	0.001	-1.97

Tables 3.3 and 3.4 show statistically significant differences between Rounds 1 and 3 for both pitch and AR. This could mean that speakers aligned very rapidly in Round 3, but see *3.4*. We do not see positive values for the interaction between Round 3 and EPtrialnr. This means that participants did not align more rapidly throughout Round 3 than in Round 1.

5 This is a corrected value, different from the ICPhS paper.

There is one potential caveat to this pattern of results. Because Rounds 1 and 3 do not consist of the same number of trials (see *3.2.3* above), the differences between the rounds could simply be due to this length difference. To control for this possibility, we checked whether the results change when we only analyse the first 60 trials of Round 3 (so it contains the same number of trials as Round 1). This analysis did not show any important changes in the pattern of results.

3.3.3 RQ3: Locality of Pitch and AR alignment

We finally investigated whether participants aligned to the immediately preceding utterance produced by the confederate, i.e., whether they aligned on a turn-by-turn basis. We therefore added the median F0 or AR of the immediately preceding sentence produced by the confederate as a fixed predictor to the models discussed above. Furthermore, we analysed the data from only Rounds 1, 2, and 3, excluding trials with outlier values from the confederates. In these models, turn-by-turn alignment should be reflected as an effect of the pitch or AR of the preceding sentence produced by the confederate on the following participant's sentence. The models showed that the preceding median pitch and AR did not have a significant effect on the participants' pitch ($\beta = 0.012$, $t = 0.91$) and AR ($\beta = 0.005$, $t = 0.37$), indicating that alignment was not a local turn-by-turn effect.

We also studied locality of the alignment effects by analysing the difference between the participant's median F0 and AR and the confederate's median F0 and AR in the directly preceding prime. We tested the same models as in *3.3.1* and *3.3.2*, but replaced the participants' F0 and AR by the absolute values of the difference scores. Results showed that there were no statistically significant effects of EPtrialnr for any of the three rounds. This suggests alignment on a turn-by-turn basis did not increase within any round.

3.4 Discussion

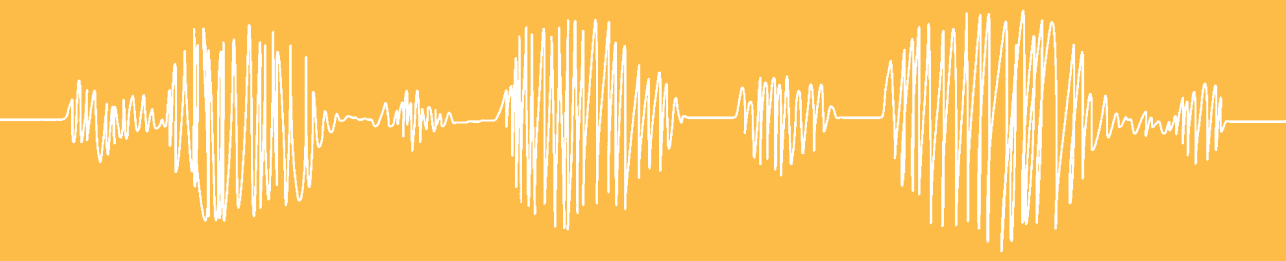
We investigated alignment of two prosodic features. The main results are as follows. First, speakers do not immediately go back to their habitual pitch and AR when they no longer hear the interlocutor. This differs from the findings by Gijssels et al. (2016), who found that participants' pitch immediately returns to a speaker's base value in the interlocutor's absence. Our results thus suggest that alignment has more long-lasting effects than suggested before.

Second, we saw a difference in overall pitch and AR between Rounds 1 and 3, with the same confederate. This could mean that participants aligned very rapidly, within the first few trials of Round 3, when they heard Confederate 1 again. Alternatively, it could be a spill-over effect from Round 2 (with a different confederate). This alternative could be tested, for example, by having participants finish sentences by themselves again in Round 2 instead of alternating with Confederate 2.

Lastly, unlike Gijssels et al. (2016) and Levitan and Hirschberg (2011), we did not find effects from the immediately preceding utterance. Taken together, these results indicate that alignment is not the exclusive result of immediate local priming from an interlocutor's preceding utterance, but rather a more global effect.

Although participants globally aligned to the confederates in both median pitch and AR, our data also show differences between median pitch and AR alignment (e.g., the effect of EPtrialnr in the post-test). Alignment of different prosodic features does thus not behave the same in all aspects in this experiment.

In conclusion, the present study suggests that prosodic alignment of pitch and AR is more than a local reaction to the acoustic characteristics of the immediately preceding utterance.



CHAPTER 4

Investigating prosodic alignment effects: the addition of control data

Abstract

This chapter investigates which effects found in Chapter 3 can be interpreted as actual alignment effects as opposed to other potential effects that are not due to alignment. We will use the data from the control experiment presented in Chapter 2. We take data from 25 participants to match the number of participants in the dataset presented in Chapter 3 (which is a subset of the data of Chapter 2). The effects found in Chapter 3 are absent in this control data, suggesting that participants in the main experiment of Chapter 3 were indeed influenced by the confederates' speech and thus show genuine alignment effects. Data furthermore suggest that Articulation Rate differs between the parts where speakers speak by themselves as opposed to when they are told that they are interacting with an interlocutor, but participants did not modulate their median F0 to a great extent in these situations. These results taken together indicate that the effects found in Chapter 3 are very likely genuine alignment effects.

4.1 Introduction

Chapter 2 presented a main experiment, on syntactic alignment, in which we varied the interlocutors, and a control experiment, in which no spoken input of the interlocutors was present. The data of the control experiment helped us interpreting which effects present in the data from the main experiment were due to the presence of the interlocutors' speech. Because the control data appeared indispensable to correctly interpret the data, we now present control data for Chapter 3, to help interpreting the data from this chapter.

The control data we present here are taken from the same experiment that we used as control in Chapter 2. That is, the control experiment is identical to the experiment presented in Chapter 3, except for two differences. First, participants did not receive any auditory input. Sentence beginnings were presented on the screen and approximately a second later, the completion from the confederate appeared on the screen. Second, 36 stimulus sentences were changed for reasons related to the control for syntactic alignment.

The control experiment consisted of the same Experiment Parts as the main experiment discussed in Chapter 3: a pre-test, Round 1, Round 2, Round 3, and the post-test. In the pre-test and post-test, participants were instructed to complete sentences by themselves. In the three rounds, they alternated completing sentences in interaction with Confederate 1 (in Round 1 and 3) and Confederate 2 (in Round 2). Together, these Rounds and tests are referred to as Experiment Parts.

In Chapter 3, we found that speakers seem to globally align to other interlocutors with respect to median Pitch and Articulation Rate. Speakers furthermore did not immediately return to their habitual F0 and Articulation Rate after interaction with interlocutors, as we found a difference between the pre-test and post-test. Next, we found a difference between a situation where a speaker interacts with an interlocutor for the first time and the second time in Articulation Rate, reflected in a difference in Articulation Rate between Round 1 and Round 3. Lastly, we did not find effects of the immediately preceding utterance on the next utterance.

In this chapter, we will discuss the control data in relation to the findings in Chapter 3. The last finding, however, cannot be tested in the control experiment, since participants did not receive any auditory input from the confederates, and we can thus not investigate the influence of the immediate confederate's Articulation Rate and Pitch on those of the participants.

4.2 Methods

The data were taken from the control experiment described in Chapter 2. To keep the number of data points in the analyses and the procedure of selecting participants comparable to Chapter 3, we selected the first 25 participants of the total 36 mentioned in Chapter 2. They were all female native Dutch speakers and aged between 18 and 30 years (M = 21.9, SD = 3.0). We performed the same cleaning procedure and statistical analyses (Linear Mixed Effects models) as reported in Chapter 3 in R version 3.4.2 (R Core Team, 2017), using the package lme4 version 1.1.13 (Bates, Maechler, Bolker & Walker, 2015), and R version 4.0.2 (R Core team, 2020) with ggplot2 version 3.3.2 (Wickham, 2016) for visualisation.

We measured participants' Pitch and Articulation Rate in the control experiment in the same manner as in the main experiment in Chapter 3.

4.3 Results

Pitch and Articulation rate for participants in the control experiment are visualised in Figure 4.1 and 4.2, respectively. The figures show the course of the two measures over the different Experiment Parts, indicated by the different colours. The light grey dots represent single sentence values.

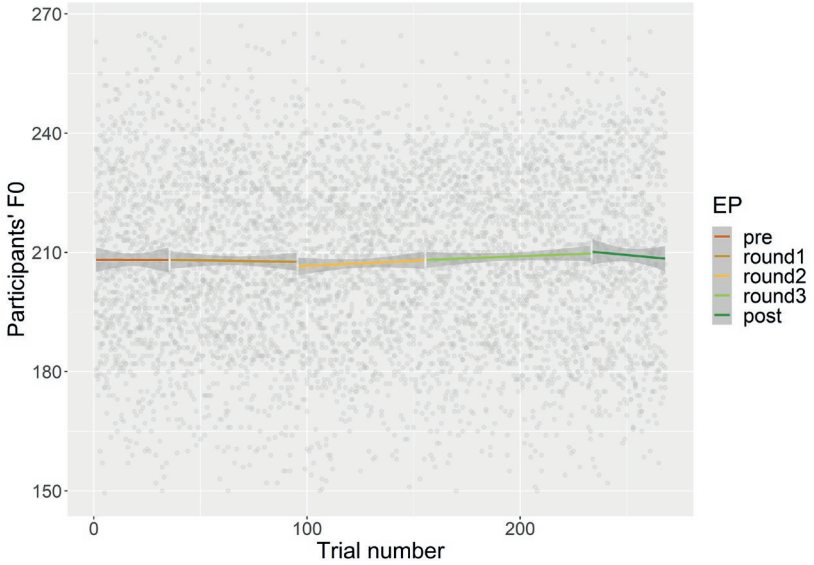


Figure 4.1. Participants' median F0 over pre-test, Rounds 1, 2 and 3 and post-test; lines were fitted using lm. Trial 1 to 35 = pre-test, 36 to 95 = Round 1, 96 to 155 = Round 2, 156 to 233 = Round 3, 234 to 268 = post-test.

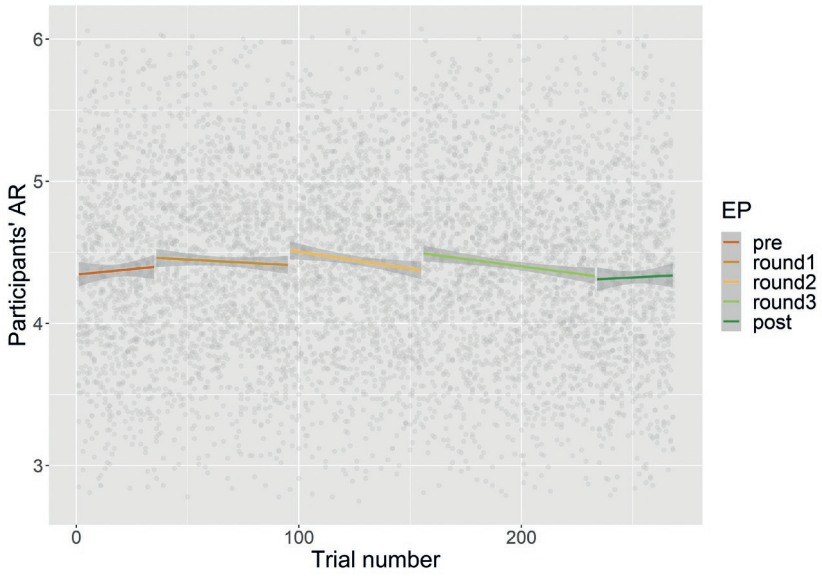


Figure 4.2. Participants' Articulation Rate over pre-test, Rounds 1, 2 and 3 and post-test; lines were fitted using lm. Trial 1 to 35 = pre-test, 36 to 95 = Round 1, 96 to 155 = Round 2, 156 to 233 = Round 3, 234 to 268 = post-test.

To test whether the effects found in Chapter 3 were also present in the control experiment, we computed similar models to the ones in Chapter 3. To match the models in Chapter 3, the models for median Pitch and Articulation Rate include Experiment Part (pre-test, Round 1, Round 2, Round 3, post-test) and Trial number within the Experiment Part as simple fixed effects and in interaction, and random intercepts for Participant and Trial (indicating the sentence to be completed).[6] The results of these models with the post-test on the intercept (as in Chapter 3) are presented in Table 4.1 and 4.2, for Pitch and Articulation Rate, respectively.

4.3.1 Long-term effects

The possible long-term effects of alignment were examined by comparing the post-test to both the pre-test and Round 3. Chapter 3 showed significant differences between the post-test and the pre-test for both Pitch and Articulation Rate and an absence of significant differences between the pre-test and Round 3. This indicated that participants

6 We did not add random slopes for Experiment Part per Participant to keep the models comparable to the ones reported in Chapter 3. We have added these slopes to the models for both Chapter 3 and the current chapter to check. In Chapter 3, this did not change any of the relevant effects. In this chapter, the effects remain similar as well. The only relevant difference is that there no longer is a difference between the pre-test and Round 1 and 3 for Articulation Rate. Furthermore, the Pitch model with Round 1 as the reference level did not converge.

did not immediately switch back to their baseline Pitch and Articulation Rate from the pre-test. As shown in Table 4.1 and 4.2, we do not see a difference between the post-test and the pre-test for either variable in the control experiment. This confirms the interpretation of the findings in Chapter 3, that alignment effects from Round 3 last in the post-test.

Instead, we see that participants change their Articulation Rate when alternating with a confederate as opposed to when they are completing sentences by themselves, as indicated by the difference between the post-test and the three Rounds. We furthermore tested this by putting the pre-test on the intercept of the model, and found that the pre-test also differs from all three Rounds (for the output, see Table B1 in Appendix B).

Table 4.1. Pitch model with the post-test as the reference level. EP stands for Experiment Part, EPtrialnr stands for the trial number within an experiment part.

Parameter	Estimate	SE	T value
Intercept	210.636	1.560	134.97
EPpre	-2.511	2.194	-1.14
EPround1	-2.351	1.943	-1.21
EPround2	-3.980	1.955	-2.04
EPround3	-2.470	1.872	-1.32
EPtrialnr	-0.074	0.075	-0.98
EPpre:EPtrialnr	0.062	0.107	0.58
EPround1:EPtrialnr	0.060	0.082	0.73
EPround2:EPtrialnr	0.098	0.082	1.19
EPround3:EPtrialnr	0.099	0.079	1.25

Table 4.2. Articulation Rate model with the post-test as the reference level. EP stands for Experiment Part, EPtrialnr stands for the trial number within an experiment part.

Parameter	Estimate	SE	T value
Intercept	4.329	0.063	68.92
EPpre	0.021	0.057	0.37
EPround1	0.128	0.051	2.52
EPround2	0.185	0.051	3.66
EPround3	0.140	0.048	2.90
EPtrialnr	<-0.001	0.002	-0.08
EPpre:EPtrialnr	0.001	0.003	0.20
EPround1:EPtrialnr	-0.001	0.002	-0.26
EPround2:EPtrialnr	-0.002	0.002	-1.03
EPround3:EPtrialnr	-0.002	0.002	-0.83

4.3.2 Realignment to an interlocutor

In Chapter 3, we found significant differences between Round 1 and 3 overall, but no effects of Trial number over the Experiment Parts indicating alignment for Pitch and AR. The slopes indicate the development of these measures over time within an Experiment Part. There are different options to interpret these results, as mentioned in Chapter 3. It could be that participants aligned very quickly in the first few trials of Round 3 or that there were spill-over effects from Round 2. As shown in Table 4.3 and 4.4, we do not find any of these differences in the control experiment. This further supports that there are effects due to the input from the confederates in Chapter 3.

Table 4.3. Pitch model with Round 1 as the reference level. EP stands for Experiment Part, EPtrialnr stands for the trial number within an experiment part.

Parameter	Estimate	SE	T value
Intercept	208.285	1.157	180.04
EPpre	-0.160	1.928	-0.08
EPround2	-1.629	1.650	-0.99
EPround3	-0.119	1.551	-0.08
EPpost	2.351	1.943	1.21
EPtrialnr	-0.014	0.033	-0.43
EPpre:EPtrialnr	0.002	0.083	0.03
EPround2:EPtrialnr	0.038	0.047	0.82
EPround3:EPtrialnr	0.039	0.047	0.97
EPpost:EPtrialnr	-0.060	0.082	-0.73

Table 4.4. Articulation Rate model with Round 1 as the reference level. EP stands for Experiment Part, EPtrialnr stands for the trial number within an experiment part.

Parameter	Estimate	SE	T value
Intercept	4.457	0.057	78.20
EPpre	-0.107	0.051	-2.11
EPround2	0.057	0.043	1.34
EPround3	0.012	0.041	0.30
EPpost	-0.128	0.051	-2.52
EPtrialnr	-0.001	0.001	-0.82
EPpre:EPtrialnr	0.001	0.002	0.52
EPround2:EPtrialnr	-0.002	0.001	-1.34
EPround3:EPtrialnr	-0.001	0.001	-1.08
EPpost:EPtrialnr	0.001	0.002	0.26

4.4 Discussion

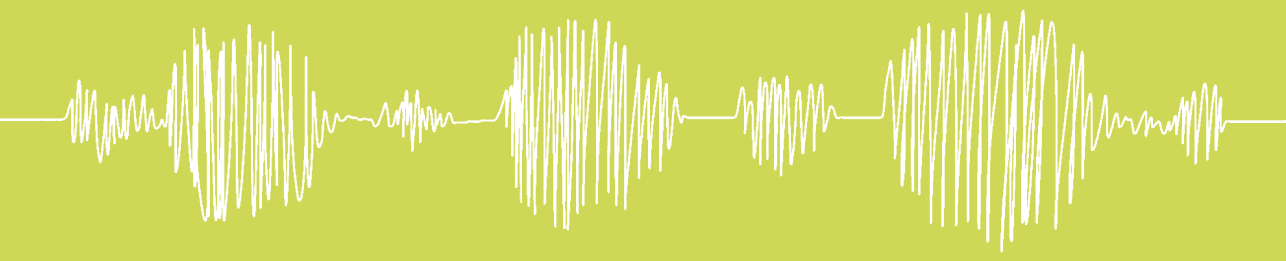
This chapter investigated whether the effects observed in the main experiment discussed in Chapter 3 were also present when participants did not hear any spoken input from the confederates (in the control experiment). This appears not to be the case. Thus, the effects observed in the main experiment (Chapter 3) were not present in the control experiment which suggests that the significant effects of Chapter 3 are due to participants' reactions to the auditory input. We can thus confirm that we found a form of alignment to the median F0 and Articulation Rate. These findings show the importance of a control experiment to be able to confirm findings of alignment.

Furthermore, next to confirming that the effects of Chapter 3 were alignment effects, we also found that when not presented with confederates' speech, participants change their Articulation Rate when they are in different situations: speakers seem to modulate their Articulation Rate differently when they are completing sentences while being presented with (written) input from a confederate as opposed to when they are completing sentences by themselves.

Appendix B

Table B1: Articulation Rate model with the pre-test as the reference level. EP stands for Experiment Part, EPtrialnr stands for the trial number within an experiment part.

Parameter	Estimate	SE	T value
Intercept	4.350	0.063	69.50
EPround1	0.107	0.051	2.11
EPround2	0.164	0.051	3.25
EPround3	0.119	0.048	2.49
EPpost	-0.021	0.057	-0.37
EPtrialnr	<0.001	0.002	0.20
EPround1:EPtrialnr	-0.001	0.002	-0.52
EPround2:EPtrialnr	-0.003	0.002	-1.28
EPround3:EPtrialnr	-0.002	0.002	-1.10
EPpost:EPtrialnr	-0.001	0.003	-0.20



CHAPTER 5

Phonetic alignment to regional variants of allophones

Abstract

Alignment is the process of speakers adapting their speech to the interlocutor. This study investigated phonetic alignment across regional variants, examining whether speakers align to prestigious versus less prestigious variants to the same extent, and whether such potential alignment is better reflected in more local or global acoustic measures. The second goal was to investigate whether speakers only align to a feature of a regional variant when they hear this feature, or whether it is enough to hear other characteristics of the regional variant. We examined whether speakers, in an interaction, aligned to two regional allophone variants in Dutch (the so-called "hard g" versus "soft g", i.e., [χ] versus [x]-[ɣ]) used by two different interlocutors. Participants (who were not selected to have either regional variant) performed a sentence completion task in which they first completed sentences on their own in a pre-test, then in interaction with Confederate 1 in Round 1, with Confederate 2 in Round 2, with Confederate 1 again in an inter-test and Round 3, and lastly by themselves again in the post-test. We measured alignment as adaptation in the participants' duration and Centre of Gravity of their allophone. Results indicated no clear alignment effects to either of the variants. First, we did not find differences between the Rounds with the different confederates or between the pre- and post-test. Second, none of the alignment measures we developed to detect alignment at the local, intermediate or more global level showed any effects: 1) the last produced confederate's hard/soft g, 2) average of the last ten confederate's hard/soft g's, and 3) average of all previous confederates' hard/soft g's. Exploration of the individual differences showed large variation, which could not be explained by the participants' baseline productions, their opinion on the likeability of the confederates' accents or the pride of their own accent. This indicates that phonetic alignment to regional variants does not (only) depend on the prestige, but is susceptive to large idiosyncratic differences among speakers.

5.1 Introduction

The process of alignment, that is, adapting one's speech to that of an interlocutor, has been well-studied. The phenomenon, also known as accommodation, entrainment or convergence, has been investigated at several linguistic levels; e.g., the phonetic level (e.g., Pardo, 2006), the syntactic level (e.g., Branigan, Pickering & Cleland, 2000), and the lexical level (e.g., Brennan & Clark, 1996). Pickering and Garrod (2004) suggest that this alignment is automatic, and makes conversation easier. In contrast, the Communication Accommodation Theory states that different social factors may influence alignment. Dragojevic, Gasiorek & Giles (2016), for example, mention in their book that speakers can behave differently when hearing different accents, depending on how prestigious the accent is. Speakers from a low prestige variant generally align to a more prestigious variant, although some speakers may not align, or even diverge, in order to keep their identity. This study further investigates phonetic alignment in regional variants.

Phonetic alignment seems to be a rather robust phenomenon: speakers tend to adapt to each other's way of speaking at the phonetic level. This has consistently been found for suprasegmental features (such as pitch and articulation rate, e.g., Gijssels, Casasanto, Jasmin, Hagoort & Casasanto, 2016 and in Chapter 3; also referred to as prosodic alignment) and for the second and first formants of vowels (e.g., Babel, 2010), both in first language and in second language speech (e.g., Berry & Ernestus, 2018; Troncoso-Ruiz, Ernestus & Broersma, 2019). With respect to experimental tasks, experimental imitation tasks like the shadowing task (e.g., Pardo, Jordan, Mallari, Scanlon & Lewandowski, 2013) and perceptually focused tasks like the AXB task (e.g., Pardo, 2006) have proven especially sensitive to detecting alignment effects.

5.1.1 Local versus global alignment

A reoccurring topic in theories of alignment (on all linguistic levels) is whether alignment is primarily a local or a global phenomenon. This distinction is relevant because it provides information on the time course of alignment, and subsequently about the underlying mechanisms (for example, local alignment may be interpreted as evidence for a short-term priming mechanism). Many studies on alignment have shown that speakers align to their interlocutors' directly preceding utterance (e.g., Gijssels et al., 2016), thus very locally. Other studies have reported more global alignment effects, with the alignment building up over time (e.g., Chapter 3), or persisting after the interaction (e.g., Ruch, 2015; Troncoso-Ruiz et al., 2019). Studies investigating both the local and the global level find differing results (Berry & Ernestus, 2018; Levitan & Hirschberg, 2011; and in Chapter 3).

In Chapter 3, we investigated both local and global alignment in Pitch and Articulation Rate to two different interlocutors and only found alignment at the global level, where speakers seemed to progressively move their productions towards the confederates' speech during the interaction, which persisted after the end of the interaction. In contrast, Berry and Ernestus (2018) found both local and global alignment in two vowel contrasts in Spanish and Dutch speakers interacting in English. In standard English pronunciation, the two vowel contrasts under study are separated. However, the Dutch speakers used one vowel for the one contrast (/ɛ/–/æ/) and two vowels for the other (/i/–/ɪ/), where the Spanish speakers started out with the opposite. The authors found that the Spanish speakers, conversing with the Dutch speakers in an English conversation, quickly merged one of the vowel contrasts (/ɛ/–/æ/), indicating local alignment to the Dutch speakers in this contrast, and slowly separated the other (/i/–/ɪ/), indicating global alignment to the Dutch speakers in this latter contrast. Levitan and Hirschberg (2011) looked at different alignment measures, which in terms of our terminology, can be interpreted as local alignment and global alignment measures. They investigated intensity mean and intensity maximum, pitch mean and maximum, jitter, shimmer, noise-to-harmonics ratio and speaking rate. For all these features, they found local alignment effects and for some features they also found that speakers become more similar over the conversation, indicating a more global alignment effect.

In summary, these studies show both local and global alignment effects, some of which lasted after the interactions. They furthermore suggest that different measures (i.e., the contrasts in Berry & Ernestus (2018)) and different operationalisations of alignment on the local and the global dimension may lead to different results (e.g., Levitan & Hirschberg, 2011). So far, it is unclear which characteristics primarily trigger local alignment and which ones primarily trigger global alignment.

5.1.2 Previous research on alignment to regional variants

Several studies have investigated whether alignment to an interlocutor can also be found when the interlocutors speak different variants of a language. These studies all show large individual differences that may be driven by social factors. For instance, Babel (2010) investigated alignment of vowels (first and second formants from six different vowels) across two English dialects and whether potential alignment would be influenced by speakers' attitudes towards their interlocutors. Babel found that New Zealand speakers especially align to Australian speakers if they are pro-Australia rather than pro-New Zealand (biases being measured by an Implicit Association Task).

Another example of a study showing individual differences in alignment to regional variants is by Gessinger and colleagues (2019b), who used a Wizard-of-Oz experiment to look at the phonetic alignment of the German suffix <-ig>. This suffix is pronounced [ɪç] in the North of Germany and [ɪk] in the South (Gessinger et al., 2019a). Participants first

performed a baseline task during which they produced several instances of this suffix. Then they interacted with an intelligent computer system called Mirabella. Mirabella always produced the opposite allophone to that preferred by the participant in the baseline task. Results indicate that some participants aligned to Mirabella (and thus diverge from their baseline preference), while others retain their preference or diverge from Mirabella (divergence meaning the opposite of alignment in the rest of the paper). The authors speculated these mixed results may be due to awareness of and attitude towards the contrast. Similarly, Earnshaw (2020), investigated the FACE vowel (i.e., the vowel as it occurs in the word <face>), in West Yorkshire English speakers, and showed that speakers are highly variable in the amount and in the direction of alignment of this vowel.

A study by Salvesen (2016) differs from most studies in that it investigated phonetic alignment of the same speakers to two different regional variants. A small sample of speakers interacted with speakers of two Scottish standard English accents (Anglo Standard Scottish English versus Scots Scottish Standard English) on different days. The study found that the speakers of the Anglo variant aligned more to the speakers of the Scots variant, avoiding characteristics of the Anglo variants overall. Salvesen explained the findings by hypothesising that the Anglo speakers may want to be part of the so-called "in-group" of Scots speakers, while avoidance of the Anglo characteristics may be due to the experimental setting being informal, indicating that the Anglo variant may be more prestigious.

In conclusion, alignment to regional variants may be susceptive to large individual differences. Some speakers align while others do not change their productions or even diverge from the interlocutor. So far, the number of speech characteristics and the number of regional variants studies is too limited to draw any firm and clear conclusions about when alignment may be expected.

5.1.3 The current study

The main goal of our study is to extend the body of evidence on alignment to regional variants. We will do so by studying alignment to the allophones of a phoneme which differ between two regional variants of Dutch, one being more standard and prestigious than the other. We will address two questions. Firstly, do speakers align more to the prestigious variant than to the less prestigious variant, and is this better reflected in a local or global alignment measure? Secondly, if speakers aligned to an interlocutor, and later on they hear this interlocutor again, do they need to be exposed to the specific allophone to show alignment or is hearing the interlocutor enough to also change speakers' production of the phoneme?

We will explore our questions using the Dutch fricative /x/ (in for example <goed>, meaning *good* in English). This phoneme has two main variants: the so-called "hard g"

and "soft g" (e.g., Van de Velde, Van Hout & Gerritsen, 1997; Van der Harst & Van de Velde, 2008). The two variants primarily differ in place of articulation. The "hard g" is uvular and is predominantly used in the north of the Netherlands (above the major rivers). The "soft g" is typically palatal, velar or palato-velar and is commonly used in the south of the Netherlands. The regional variant with the "hard g" is generally regarded as more prestigious and standard than the variant with the "soft g" (e.g., Grondelaers & van Hout, 2010; Pinget, Rotteveel & Van de Velde, 2014). When not referring to the specific allophones, we will use /x/ to refer to the phoneme and its variants in the rest of the paper.

As a result of alignment, speakers may completely change the place of articulation of the fricative. This would be a large, categorical change. However, speakers may also apply more subtle changes, gradually changing some of the acoustic characteristics of the fricative. In order to be able to capture subtle alignment effects, we will analyse the duration of the /x/. The choice for the duration measure is based on van der Harst and colleagues' (2007) observation that the "hard g" tends to be longer than the "soft g". In addition, we will measure the Centre of Gravity (CoG) to validate the duration results. A lower CoG reflects a more frontal fricative, and "soft g" should thus be characterised by a lower CoG than "hard g".

We studied phonetic alignment to the "hard g" and "soft g" in the dataset obtained in Chapter 2 for the study of syntactic alignment. This dataset resulted from an experiment that was designed to investigate both syntactic alignment (see Chapter 2), and phonetic alignment to regional variants (the present study). In the experiment, participants interacted with two different confederates. The first confederate used a "hard g" and the second confederate used a "soft g". Each participant interacted with both confederates, which allows us to investigate whether participants phonetically align to both regional variants.

The experiment consisted of a sentence completion task during which, in most parts, participants interacted with pre-recorded speech from the confederates. Pre-recorded speech was used to allow for a careful and full control over the phonetic input. During these interaction parts of the experiment, confederate and participant strictly alternated completing sentence beginnings to full sentences. Apart from these interaction parts, participants also completed a pre- and a post-test during which they completed sentences by themselves.

Figure 5.1 presents an overview of the experiment, as presented in Chapter 2. Participants were not selected to have a preference for either allophone and all participants therefore first participated in a pre-test, which provided us with a baseline of each participant's use of the /x/ at the start of the experiment. After the pre-test, participants interacted with Confederate 1, who produced uvular [χ], in Round 1. Subsequently, participants completed sentences in Round 2 in interaction with Confederate 2 who

produced a (palato-)velar/palatal [ɣ-x]. Round 2 was followed by an inter-test, where participants interacted with Confederate 1 again, but Confederate 1 only produced one instance of /x/ over the whole inter-test.[7] This inter-test allows us to investigate our second research question - whether participants align to Confederate 1 without hearing the /x/ being produced. After the inter-test, participants continued interacting with Confederate 1 in Round 3 who produced uvular [χ] again, as in Round 1. Lastly, participants completed a post-test, which was comparable to the pre-test – participants completed sentences by themselves. All parts of the experiment will be referred to as Experiment Parts, indicating the pre-test, inter-test, post-test, Round 1, 2, and 3.

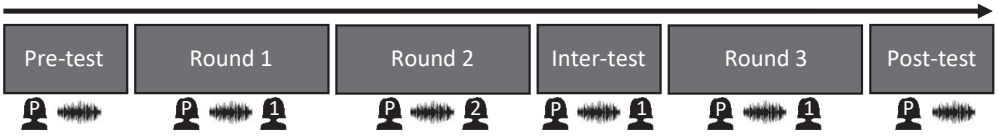


Figure 5.1. Procedure of the experiment

P means participant, 1 and 2 refer to Confederate 1 and 2 respectively.
Pre- and post-test: Participants completed sentences by themselves.
Round 1 and 3: Participants completed sentences in alternation with Confederate 1, who produced [χ].
Round 2: Participants completed sentences in alternation with Confederate 2, who produced [x]-[ɣ].
Inter-test: Participants completed sentences in alternation with Confederate 1, but the total of the confederate's sentences included just one [χ].

We will investigate whether alignment took place in several ways. First, we will compare the duration and Centre of Gravity of the /x/ in different parts of the experiment. The parts where participants interacted with Confederate 1 will be compared to the part where participants interacted with Confederate 2. Furthermore, the pre- and post-test will be compared to investigate possible long-term global effects after the interaction.

Second, we will test for alignment in the duration of the /x/ by means of three continuous predictors reflecting the input that participants received with respect to the /x/. The three predictors differ in the locality of this input. The first predictor only reflects the properties of the immediately preceding /x/. This predictor should detect local alignment and is called the "immediate predictor" hereafter. The value of this predictor changes with every production of the /x/ by the confederate and is thus highly variable. The second predictor is duration of the last ten tokens of the interlocutor's /x/. This predictor reflects more global alignment and is called the "intermediate predictor" hereafter. This predictor will be more stable than the immediate predictor,

7 The inter-test was designed to not include any /x/s, but due to an oversight in the stimulus construction, Confederate 1 did produce one single instance of /x/.

since it averages over ten different productions of confederates' /x/s, but changes more drastically when there are switches from one confederate to the other (i.e., from Round 1 to Round 2). This increased stability, in turn, also means that this predictor is less representative for local effects. Finally, the third predictor is the average duration of the /x/s produced by both confederates so far. This predictor is meant to detect global alignment and is called the "accumulating predictor" hereafter. Together, the use of these three different predictors will maximise the chance that we will find alignment effects and will give us more insight into whether alignment to the /x/ in Dutch is a more local or more global phenomenon.

Next to the main experiment, we conducted a control experiment, as described in Chapter 2. Participants in the control experiment did not hear the confederates (for procedural details, see below). This control experiment was implemented to identify whether any effects we might observe in the main experiment are real alignment effects, or whether they could be due to other potential factors like spontaneous changes in the production of /x/ over repeated production.

5.2 Methods

5.2.1 Participants

All data from the syntactic alignment experiment described in Chapter 2 were analysed in this study. It consists of participants' productions in the main and control experiment, over which 72 participants were divided equally. Participants in the main and control experiments were aged 18 to 26 years (M = 22.4, SD = 2.0) and 18 to 30 years (M = 21.4, SD = 2.8), respectively. Participants were not selected to have a preference for either of the allophones. They did not report any serious speech, language or hearing impairments relevant for the task and were compensated for their participation.

5.2.2 Confederates

The confederates were two Dutch female speakers of similar ages as the participants (23 and 24 years at the time of recording). Confederate 1 was selected to have a so-called "hard g" (/χ/ - she lived above the major rivers of the Netherlands, in the city of Delft, most of her life). Confederate 2 was selected to have a so-called "soft g", (/x/ and /ɣ/ - she lived in the South of the Netherlands, in the city of Venlo, most of her life). Confederate 1's /x/s had an overall mean duration of 72.1 ms (SD = 28.7 ms) and a CoG of 2270.3 Hz (SD = 757.1 Hz), and Confederate 2's /x/s had an overall mean duration of 58.7 ms (SD = 27.6 ms) and a mean CoG of 2028.7 Hz (SD = 1366.3 Hz).

5.2.3 Materials

5.2.3.1 Stimuli

Participants received two types of stimuli: full sentences that were produced by the confederates and sentence beginnings that had to be completed to full sentences by the participants. An example of a sentence produced by the confederate is "De boer ontkende dat [hij zijn koeien niet goed verzorgde.]" (English translation: "*The farmer denied that [he did not take good care of his cows.]*"), where the part between square brackets was not shown on the participants' screen, but only heard over headphones. This example features two tokens of the /x/ (in *goed* and in *verzorgde*). An example of a sentence to be completed by the participant is "Als het dit weekend weer zulk mooi weer is gaan we…" (English translation: "*If the weather is this nice again this weekend we are going to…*").

Each participant received a total of 268 to-be-completed sentences (60 per round, 35 in the pre-test and in the post-test and 18 in the inter-test). The set of sentences produced by the confederates consisted of 198 complete sentences and included 0 to 5 occurrences of the /x/ (there were 33 sentences without a /x/ in total, of which 17 in the inter-test, meaning 9% of the confederates' stimuli over the three rounds did not contain a /x/). When there was no /x/ in the confederate's sentence – the immediate predictor – is slightly less local, since it reflects the Confederate's /x/ in the utterance preceding the immediately preceding utterance. Of the total set of 268 plus 198 sentences, 205 were adapted from Hartsuiker and Westenberg (2000).

Stimuli were pseudo-randomised over six lists. Lists for the control experiment were similar to those of the main experiment, except for some prime sentences that were substituted by other sentences, for reasons related to the parallel study on syntactic alignment (see Chapter 2 for further details), adding 36 confederates' stimuli to the total stimuli set.

The speech of the two confederates was recorded in a sound attenuated booth with a Sennheiser K6/ME 64 microphone connected to a pre-amplifier and a Roland R-05 recorder. The confederates' speech was digitised at a sampling rate of 44.1 kHz, with a 16-bit quantisation. The intensity of the recorded sentences was normalised to 57 dB. The stimuli were presented with the Presentation software (Version 20.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com).

5.2.4 Procedure

5.2.4.1 Main experiment

Participants were individually tested in a sound attenuated booth, using the same equipment as used for the recording of the confederates. During the pre-test and the post-test, participants saw a beginning part of a sentence on the screen, including the instruction to complete the sentence ("Complete this sentence" in Dutch) and to press Enter after completion of the sentence to go to the next sentence.

During the rounds and inter-test, participants and confederates strictly alternated completing sentences. The procedure for completing sentences for the participants was the same as in the pre- and post-test. The confederates were not physically present; they were only present in the form of their pre-recorded speech and their picture was visible as well. Participants heard the full sentences produced by the confederates over headphones, while seeing the beginning parts on the screen. The audio started about one second after the visual presentation of the beginning part of the sentence. The visually presented parts of the stimuli appeared on the screen centred in Times New Roman, font size 34. Participants were instructed to judge whether they would complete the sentence in the same way as the confederate using a 7-point Likert scale ("I would complete this sentence in the same way" in Dutch). This judgement was implemented to encourage participants to listen carefully to the sentences, and mostly pay attention to the content of the sentences. Participants were told that the confederates would also rate their sentence completions.

After completing the experiment, participants filled in a questionnaire in Qualtrics providing demographical data (e.g., their age, education etc.), the likeability of the confederates (i.e., their appearance based on the photos, their voice and their accent) and their own accent. The likeability data is not analysed in this paper, except for the ratings of the confederates' accents and participants' own accents.

5.2.4.2 Control experiment

The control experiment minimally differed from the main experiment. The main difference was that the control experiment did not contain any spoken utterances by the confederates: participants only saw the sentences produced by the confederates on the screen. They first saw the beginning part of the sentence, which was completed after 2000 ms. Despite not receiving auditory input, participants also wore headphones in the control experiment, to keep differences with the main experiment minimal.

5.2.5 Forced alignment

For the measurement of the duration and CoG of the /x/s produced by the participants and the confederates, the occurrences of the /x/ were identified in the speech signal using forced alignment. Participants' and Confederates' speech was forced aligned with

Kaldi (Povey et al., 2011), a speech recognition tool. The forced alignment takes audio files and orthographic transcriptions and produces phonetic transcriptions aligned with the speech signal. This forced alignment was based on 50 acoustic phone models, including models for vowels, consonants, silence, and speaker noise. It was trained on all components with a broadband signal of the Corpus Gesproken Nederlands (CGN; Oostdijk, 2000). For each 10 ms frame, 13 Mel-frequency cepstral coefficients (MFCCs) were computed as well as 13 delta and delta-deltas, providing 39 features per frame. This resulted in nnet3 triphone models (Deep Neural Networks; DNNs). Cepstral mean and variance normalisation were used on utterance basis. The pronunciation dictionary was created by combining the Corpus Gesproken Nederlands (CGN) with the Celex dictionary (Baayen, Piepenbrock & Gulikers, 1996), removing an error (i.e., <zeiden> was changed from [zeɣdə] to [zeidə]), and adding additional pronunciation variations (i.e., we added the pronunciation [ə] for the suffix <-en>; the pronunciation [sr] for the consonant cluster <schr->; [axtban], [axtbanə], [axdban], and [axdbanə] for <achtbaan> and <achtbanen>, respectively; and [t], [s], [f], and [x], respectively, for final /d/, /z/, /v/, /ɣ/, allowing variation in voice). The forced alignment did not make a distinction between the different /x/s as, due to a lack of training material, it was not trained to do so.

This forced alignment was validated by comparing the results for a subset of the data (15 sentences) to the transcriptions of the same data by two trained native Dutch speakers, who were instructed to transcribe the phones as they heard them (i.e., they were asked to faithfully transcribe reduced variants or variations in pronunciation). This process comprised of two steps. The goal of the first step was to investigate how many of the labels matched between Kaldi and the two human transcribers and the second step aimed to find out whether the boundaries of those labels were placed in the same positions by the different transcribers. Levenstein distances were used to compare the transcriptions. The percentage of matched labels between the two human transcribers was 89.6 per cent, while it was 83.1 and 81.2 per cent, respectively between the two human transcriptions and Kaldi's transcription. We then checked how well the time points of the boundaries of the phones matched. We ignored differences up to 20 ms, which is a widely used cut-off point (e.g., Pluymaekers, Ernestus, & Baayen, 2006; Ernestus, Kouwenhoven & Van Mulken, 2017). The two human transcribers had matching timing of their boundaries (with the accepted deviation of 20 ms) in 86.8 per cent of the cases while they agreed with Kaldi in 73.1 and 75.2 per cent, respectively. Given the impossibility of manual transcription of the whole corpus, and the reasonable percentages, we accepted the phonetic transcriptions produced by Kaldi.

5.2.6 Dependent and independent variables in the analyses

Duration and Centre of Gravity serve as dependent variables in this study. Occurrences of /x/ in names, in <gg>, right before or after speaker noise, and right before or after another occurrence of /x/ were not analysed. The dependent variables will be predicted by independent variables of interest and control variables that serve to remove some of the variation in duration and CoG that is not due to alignment. The independent variables of interest in the different models are Experiment Part (only the rounds in some models), Experiment (indicating main or control experiment), and the immediate, intermediate, and accumulating predictors. We furthermore selected control variables on the basis of the many previous studies in which they showed clear effects on duration (e.g., Klatt, 1976; Pluymaekers, Ernestus & Baayen, 2005). Depending on the model, the (subset of) control independent variables are the articulation rate measured over the sentence, the position of the /x/ in the word (onset, middle, offset), the category of the previous and the next phone (back vowel, front vowel, obstruent, sonorant, schwa or silence), the log transformed word frequency of the word the /x/ was in, and the duration of the word minus the duration of the /x/. Random intercepts of Participant, Trial, and Word were added, depending on the model.

A script in Praat (Boersma & Weenink, 2018) by Elvira-Garcia (2014) was adapted to extract some of the variables needed from the phonetic transcription combined with the audio: the duration of the /x/, the Centre of Gravity (referred to as CoG hereafter; using Power: 2 (default)), the phone preceding and following /x/, the word in which the /x/ occurred, and the duration of that word. For the word duration measure, the duration of the /x/ was subtracted from the total word duration, to eliminate an effect of the duration of the /x/ on the word duration. A different Praat script by de Jong and Wempe (2009) was used to extract the articulation rate of each sentence containing one or more occurrence of /x/. The settings used for the script were a silence threshold of -30 dB, a dip between peaks of 2 dB and a minimum pause duration of 0.3 seconds.

Word frequency was extracted from the Subtlex corpus (Keuleers, Brysbaert & New, 2010). Words with a frequency of 0 were taken out of the analyses. Before implementing word frequency into the models, we added 1 and log transformed the values.

5.2.7 Statistical analyses

Multiple Linear Mixed Effects regression models were fitted using R version 4.0.2 (R Core Team, 2020), including the lme4 package version 1.1-23 (Bates, Maechler, Bolker & Walker, 2015) and the car package version 3.0-8 (Fox & Weisberg, 2019). Ggplot2 (Wickham, 2016) was used for visualisation. Models were tested for both duration of the /x/ and CoG.

Since participants were not selected to have a preference for either of the allophones, we investigated whether the duration data showed an effect of the preferred /x/ on alignment. Participants from the main experiment were grouped, based on a perceptual judgement of their productions in the pre-test by the first author, in four different groups characterised by using mostly "hard g" (n = 12), "soft g" (n = 12), "both" (n = 5) or "in between" (n = 7). Group was added as a fixed effect to the analyses of the duration data from the main experiment. These analyses revealed no significant interaction effects with Experiment Part with the three continuous predictors reflecting the confederates' productions, indicating that there were no substantial alignment differences between the groups. We furthermore plotted the data per group, and we did not see any clear differences between the groups in the data for the different Rounds. Since we did not find any differences between the groups, the groups were merged in the analyses of both the duration and the CoG data discussed in this paper. In the following, we will first discuss the models with the duration of the /x/, followed by the models of the CoG.

We investigated potential alignment effects in four different manners: 1) By comparing the different Experiment Parts to see whether participants behave differently over the experiment, 2) By investigating local alignment to the interlocutor in the different rounds, 3) By investigating intermediate alignment to the interlocutor, 4) by investigating global alignment to the interlocutor.

5.2.7.1 Raw duration

To investigate whether duration of the /x/ differed between the experiments as a function of Experiment Part, we fitted a model with the raw duration data excluding 2.5 SD outliers of the data (28733 data points remaining – 96.9% of the original data points). From the final model (reference model) we excluded 2.5 SD outliers (28134 data points remaining in the model), on the main and control experiment data together. This model (referred to as "raw duration model" in the following sections) included fixed effects of Experiment Part (pre-test, inter-test, post-test, Round 1, 2 and 3) and Experiment (main or control), and an interaction between the two, the articulation rate measured over the sentence containing the relevant /x/, the position of the /x/ in the word (onset, middle, offset), the category of the previous and the next phone (back vowel, front vowel, obstruent, sonorant, schwa or silence), the log transformed word frequency of the word the /x/ was in, and the duration of the word minus the duration of the /x/. Random intercepts of Trial (the sentence to be completed by the participant) and Participant were added, as well as random slopes of Experiment Part per Participant. This model did not converge (also not when changing the optimiser to Bobyqa with 100,000 iterations) and the random slopes were taken out.

5.2.7.2 Residuals duration

The confederates' durations of /x/ are influenced by at least articulation rate and the surrounding segments, which we entered as control predictors for the participants' productions in the raw duration model. It may be argued that when aligning, speakers take these factors on the /x/ duration into account. That is, it is not the raw /x/ duration that the speaker may align to, but the /x/ duration given the speech rate, the surrounding segments, etc. The continuous predictors of alignment should therefore not reflect the (average) raw durations of confederate's productions, but durations that are independent of their exact segmental and prosodic context. Our three continuous measures are thus not based on the "raw" durations, but on durations from which the variation resulting from articulation rate and context was reduced.

We computed these "intrinsic" durations as follows. We built a model including only the control variables mentioned before and random effects of Word. In this model (referred to as "covariate model" in the next sections), the dependent variable was the raw duration of all /x/s from both participants and confederates in the main experiment (participants in the control experiment did not receive any acoustic information from the confederates and were thus excluded), including outliers. That is, we had 15085 data points for participants and 935 data points for confederates. We see the residuals of this model as approximations of the intrinsic durations of the participants' and confederates' /x/s and computed the three continuous measures on the basis of these residual durations.

We then performed four analyses, building models that we will refer to as "residual models" in the next sections. In all models, the dependent variable was the participants' intrinsic /x/ durations from the rounds (Round 1, 2 and 3) only, since participants only received input from the confederates in these parts of the experiment. We excluded 2.5 SD outliers of the models. Each of the four models included a different predictor of interest: Round (Round 1, 2 and 3 – 9806 data points – 97.7% of the original data points), replicating and validating the raw duration model, previous confederate's /x/ residual (9807 data points – 97.7% of the original data points) in interaction with Round, average of the previous ten confederate's /x/ residuals (9806 data points – 97.7% of the original data points), and average of all previous confederate's /x/ residuals (9808 data points – 97.7% of the original data points). Only the model testing for the effect of the confederate's immediately preceding /x/ included an interaction of this continuous predictor with Round to see whether participants aligned differently in the different rounds. We could not include interactions with Round in the models of the other two continuous predictors since the intermediate and cumulative measures can span over multiple rounds. All models were fitted with random intercepts of Trial (indicating the sentence beginning to be completed by the participant), random intercepts of Participant and random slopes of Round per Participant. The models did not converge with

random slopes of Round per Participant (also not when changing the optimiser to Bobyqa with 100,000 iterations) and the random slopes were taken out.

5.2.7.3 Raw Centre of Gravity

A similar model as the one used for the raw duration model was fitted for CoG (referred to as "raw CoG model" in the following sections), with CoG from the main and control experiment together as the dependent variable, excluding 2.5 SD outliers (29285 data points remaining – 98.8% of the original data points) and excluding 2.5 SD outliers of the model (28838 data points remaining in the model). This model was fitted to validate the results from the duration data. We only fitted one model with the CoG data to validate the duration analyses, due to smaller differences between the confederates in the CoG data than in the duration data (percentage difference of 20.5% for duration versus 11.2% for CoG – calculated by using (|A-B|/((A+B)/2))*100)) which suggest that the statistical power of these analyses will be smaller than of the duration analyses. The CoG model contained part of the same predictors (excluding articulation rate, word duration and word frequency) and the same random structure as the raw duration model.

5.3 Results

We examined the data descriptively and fitted the different statistical models that were introduced in the previous sections.

5.3.1 Descriptive data duration

Figures 5.2 and 5.3 show box plots of the raw duration data; Figure 5.2 for the main experiment and Figure 5.3 for the control experiment. These plots do not show any clear patterns over the Experiment Parts. More importantly, the plot of the main experiment does not show any clear deviations from that of the control experiment, suggesting that the interaction with the confederates did not have any obvious effects on the raw durations.

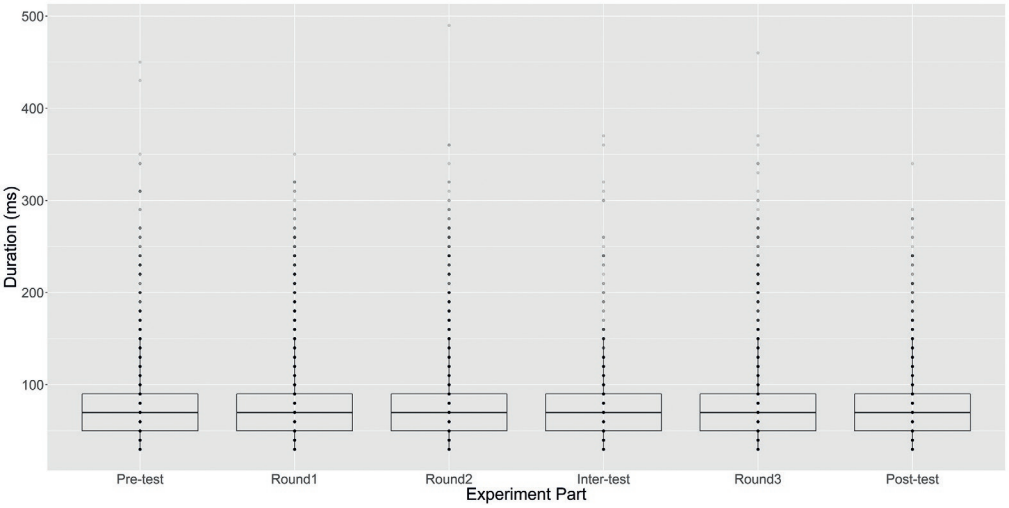


Figure 5.2. Raw duration of participants' /x/ per part of the main experiment.

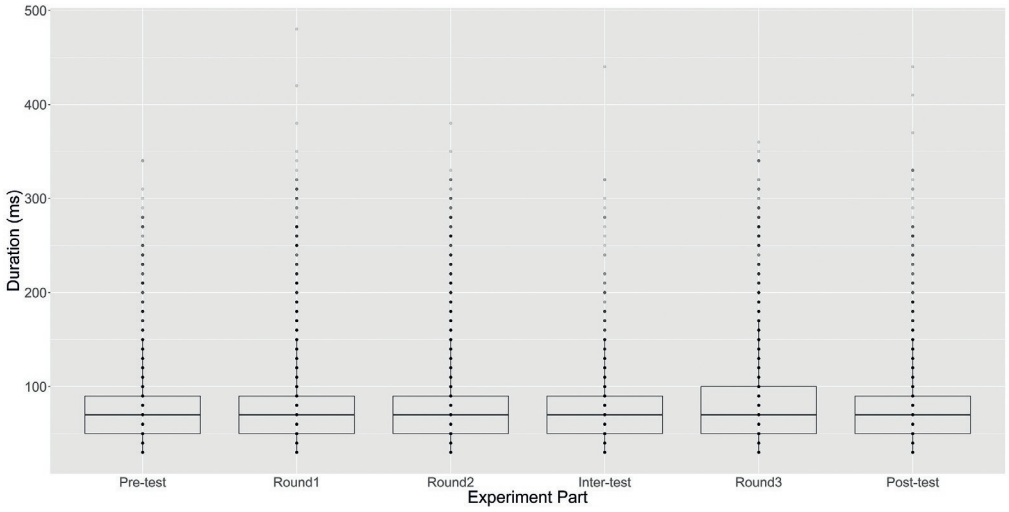


Figure 5.3. Raw duration of participants' /x/ per part of the control experiment.

5.3.2 Raw duration model

We tested a raw duration model on the data of all Experiment Parts of the main experiment and control experiment together. Table 5.1 shows the results of the ANOVA (Fox & Weisberg, 2019) conducted on the final lmer model. The ANOVA shows a significant interaction effect between Experiment Part and Experiment, indicating that

there are differences between the main and the control experiment in the Experiment Parts (in contrast to what the descriptive examination of the data suggested, see Figures 5.2 and 5.3). The other fixed effects, which were included to control for some of the variability in duration, were all statistically significant (we will discuss this in more detail in the covariate model). For the full table of the results of the raw duration model, see Appendix C, Table C1, with the pre-test (for Experiment Part), the control experiment (for Experiment), word final /x/ (for the position of the /x/), and back vowel for the category of the next and previous phones on the intercept.

To further clarify the potential differences between the main and the control experiment, we made subsets of the data of the different Experiment Parts to compare the main versus the control experiment separately for each part of the experiment. The betas and T values for the effect of Experiment in these subset models are reported in Table 5.2 per model. The other simple effects are not included in the table, since we are only interested in the difference between the two experiments (for the output of the full models, see Appendix C, Table C2 to C7). These analyses show no significant difference between the main and control experiment for any of the Experiment Parts, thus supporting the conclusion form the inspection of the descriptive data that there are no systematic differences between the main experiment and the control experiment. The interaction effect in the statistical model probably arose from the direction of the (non-significant) difference between the pre-test of the main experiment versus control experiment which is opposite to the direction of the (non-significant) differences between the other Experiment Parts (i.e., the positive versus the negative estimates in Table 5.2). This interaction is probably no sign of alignment as there are no clear differences between the rounds.

Table 5.1. Results from ANOVA for the raw duration model.

Variable	Chi square	Df	P value
Experiment part	11.97	5	<0.05
Experiment	1.44	1	0.23
Articulation rate	96.37	1	<0.001
Position /x/	908.19	2	<0.001
Category of the next phone	11684.83	5	<0.001
Category of the previous phone	10865.86	5	<0.001
Logged word frequency	49.10	1	<0.001
Word duration without /x/	549.58	1	<0.001
Experiment part*Experiment	19.12	5	<0.01

Table 5.2. The effect of Experiment on duration as the dependent variable in the models ran per Experiment Part with the control experiment on the intercept. Other simple effects are not reported here.

Subset	Estimate	T value
Pre-test	0.71	0.36
Round 1	-2.14	-1.15
Round 2	-2.49	-1.46
Inter-test	-2.48	-1.27
Round 3	-3.03	-1.68
Post-test	-1.86	-0.98

5.3.3 Residual models

5.3.3.1 Covariate model

The covariate model was meant to compute the "intrinsic" durations of the participants' and confederates' /x/s in the main experiment. Results of this model can be found in Table 5.3. We see that the /x/ is shorter when the articulation rate is higher, is longer when it is in word initial rather than final position, and is longer when the word duration is longer, and, unexpectedly, longer when the word is more frequent. Finally, the duration differs for different phonetic contexts. Most of these effects (except for the effect of word frequency) are in the expected directions and thus confirm that our duration measure is sensitive to the context in which /x/ occurs.

Table 5.3. Results from the covariate model; word final /x/ and back vowel for the category of the next and previous phones on the intercept.

Variable	Estimate	Standard error	T value
Intercept	40.64	3.10	13.12
Articulation rate	-2.16	0.28	-7.64
Logged word frequency	3.52	0.27	13.25
Word duration (ms)	0.09	0.002	46.20
Position /x/ – initial	6.11	1.29	4.74
Position /x/ – middle	0.12	1.17	1.10
Category next phone – front vowel	-3.33	1.52	-2.20
Category next phone – obstruent	-10.52	1.32	-7.94
Category next phone – schwa	-31.43	1.24	-25.41
Category next phone – silence	95.68	1.62	59.06
Category next phone – sonorant	-7.42	1.31	-5.67
Category previous phone – front vowel	-6.80	1.31	-5.20
Category previous phone – obstruent	-23.88	1.20	-19.86
Category previous phone – schwa	-18.27	1.16	-15.71
Category previous phone – silence	44.24	1.66	26.62
Category previous phone – sonorant	-23.31	1.27	-18.37

5.3.3.2 Residual models

We tested four different models that had the participants' residuals of the covariate model as the dependent variable. The first model tested for Round to investigate whether participants showed any differences between the three different rounds. This model did not show any effects of Round. Model 2 tested for the effect of just the immediately preceding confederate's /x/ on the participant's /x/ in interaction with Round. Model 3 tested for the effect of the average of ten preceding /x/s of the confederate(s) on the participant's /x/. The fourth model tested for the effect of the average of all /x/s from both confederates that precede a participant's /x/ on this /x/. None of these three models showed any significant effects for the confederates' productions. Results of these models are reported in Tables 5.4 to 5.7. Overall, the results of these residual models confirm the observations on the basis of the descriptive data and the raw duration data: we see no clear indications of alignment effects in these participants.

Table 5.4. Results from the residual model testing for effects of Round with Round 2 on the intercept.

Variable	Estimate	Standard error	T value
Intercept	-0.38	1.20	-0.32
Round 1	0.11	0.41	0.26
Round 3	-0.36	0.41	-0.87

Table 5.5. Results from the residual model testing for effects of the previous /x/ in interaction with Round, with Round 2 on the intercept.

Variable	Estimate	Standard error	T value
Intercept	-0.61	1.22	-0.50
Residual of the last confederate's /x/	-0.02	0.02	-0.92
Round 1	0.38	0.51	0.74
Round 3	0.09	0.51	0.17
Residual of the last confederate's /x/*Round 1	0.01	0.03	0.53
Residual of the last confederate's /x/* Round 3	<-0.01	0.03	-0.05

Table 5.6. Results from the residual model testing for effects of the average of the confederates' last ten /x/s.

Variable	Estimate	Standard error	T value
Intercept	-0.44	1.18	-0.37
Average of the residuals of the last ten confederate's /x/	-0.02	0.02	-1.15

Table 5.7. Results from the residual model testing for effects of the average of all confederates' /x/s preceding a participant's production of a /x/.

Variable	Estimate	Standard error	T value
Intercept	-0.45	1.18	-0.38
Average of the residuals of all heard confederate's /x/	<0.01	0.04	-0.02

5.3.4 Descriptive data CoG

Figures 5.4 and 5.5 show the descriptive raw Centre of Gravity data in the main experiment and in the control experiment, respectively. As for the raw duration data, we do not see any clear patterns over the different Experiment Parts, nor clear differences between the main and control experiment.

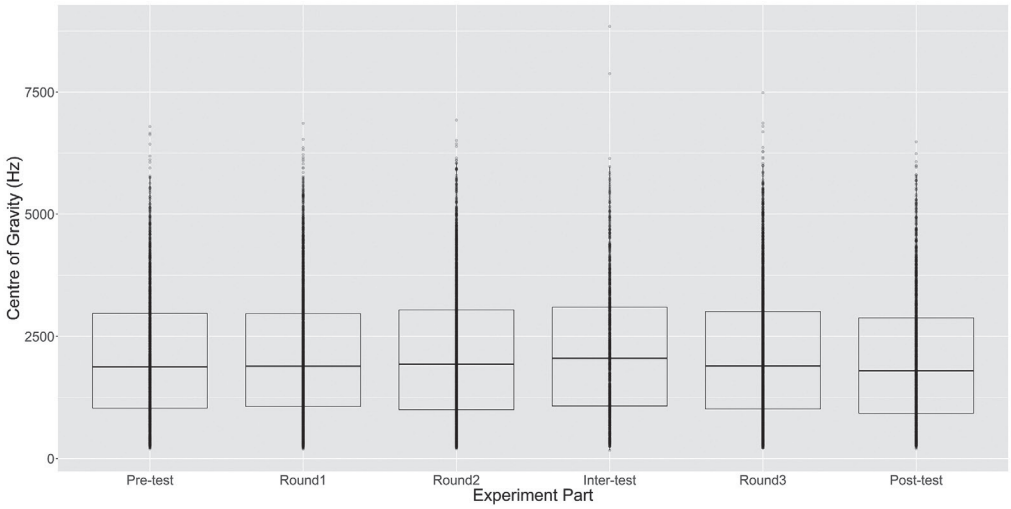


Figure 5.4. Raw Centre of Gravity of participants' /x/ per part of the main experiment.

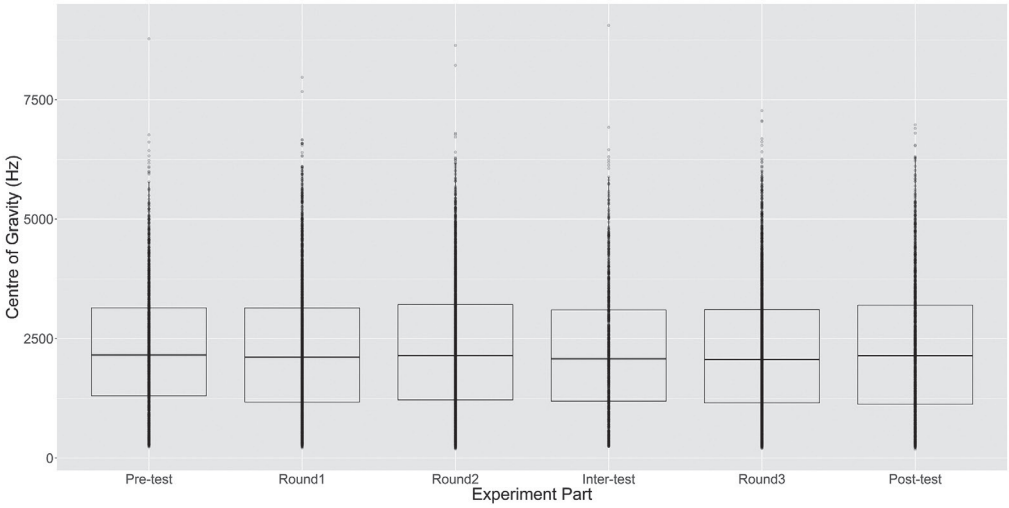


Figure 5.5. Raw Centre of Gravity of participants' /x/ per part of the control experiment.

5.3.5 Raw Centre of Gravity model

To confirm the absence of general alignment effects reflected in CoG as indicated in the descriptive data, we performed statistical analyses. The ANOVA (Fox & Weisberg, 2019; see Table 5.8 for the output) on the final model showed a significant interaction between Experiment and Experiment Part, in the absence of a main effect of Experiment. For the full output of the raw CoG model (on the intercept: pre-test, control experiment, final /x/, and back vowel for the category of the next and previous phones), see Appendix C, Table C8.

Table 5.8. Results from ANOVA for the raw CoG model.

Variable	Chi square	Df	P value
Experiment Part	10.64	5	<0.05
Experiment	2.25	1	0.13
Position /x/	454.15	2	<0.001
Category of the next phone	4566.15	5	<0.001
Category of the previous phone	6382.05	5	<0.001
Experiment part*Experiment	11.20	5	<0.05

In order to investigate the interaction, we analysed every Experiment Part separately. In these models, a main effect of Experiment (see Appendix C, Table C9 to C14) was only present in the post-test (see Table 5.9). This effect for the post-test indicates that, in the post-test, participants in the main experiment had a lower CoG than participants in the control experiment. Participants' /x/ in the post-test of the main experiment thus seems

to be closer to the soft /x/ of Confederate 2 in Round 2 than to the /x/ of participants in the control experiment. This is unexpected, since the participants in the main Experiment heard Confederate 1, not Confederate 2, in Round 3, just before the post-test. We therefore assume that this interaction is not a reflection of CoG alignment.

Table 5.9. The effect of Experiment on CoG in the models ran per Experiment Part with the control experiment on the intercept. Other simple effects are not reported here.

Subset	Estimate	T value
Pre-test	-185.03	-1.77
Round 1	-116.37	-1.28
Round 2	-135.35	-1.44
Inter-test	-138.01	-1.23
Round 3	-144.77	-1.50
Post-test	-216.87	-2.06

Since we did not find any alignment effects in this CoG model, and because of the smaller differences between the confederates in the CoG data than in the duration data, we do not further explore CoG.

5.3.6 Exploration of Individual differences

Since previous literature shows large individual differences in alignment, and since we did not find any clear group alignment effects, we explored the individual participant duration data in three different manners. First, we calculated the averages of each participant's productions in the main experiment in Round 1 and 2, respectively. We then subtracted the average of Round 2 from the average of Round 1 per participant to detect duration differences between the two rounds, in which the participants were presented with different /x/s by different confederates. Figure 5.6 plots these differences. Each dot represents a participant. A positive difference means that a participant had a shorter duration in Round 2 than in Round 1. Because Confederate 1 had a hard /x/ and thus a longer duration, and Confederate 2 had a soft /x/ and thus a shorter duration, a positive difference indicates that participants show a form of alignment. A negative difference means divergence, since participants have a longer duration in Round 2 than in Round 1 in this case.

The data was grouped (see the different colours in Figure 5.6) based on perceptual judgements based on the pre-test, resulting in four groups, a group mainly producing "hard g" in the pre-test, a group mainly producing "soft g" in the pre-test, a group producing both and a group with productions that seem to lie in between the allophones (see the Methods section). Visual inspection of Figure 5.6 does not reveal any systematic pattern - participants from all groups are distributed over the whole scale. This is in line

with the non-significant effect of Group in the statistical models tested for duration in the main experiment, as mentioned in the Methods.

Next, we investigated the differences between the average durations between Round 1 and Round 2, grouping participants based on their ratings of the likeability of the confederates' accents as indicated in the questionnaire. Participants were divided into three groups, one group that had a preference for the accent of Confederate 1 (n = 26), another group with a preference for the accent of Confederate 2 (n = 4), and the last group without a preference (n = 6). In Figure 5.7, these groups are marked with different colours. Again, we see no systematicity within groups; participants from all groups are spread over the whole scale.

Lastly, we investigated the same differences, but now grouping the participants based on their rating of how proud they are of their own accent (Figure 5.8). Participants were divided into three groups, one group that was proud of their accent (positive; n = 13), another group that was neutral about their accent (neutral; n = 9), and a last group that was not proud of their accent (negative; n = 14). We again see that participants in the three groups are spread over the whole scale.

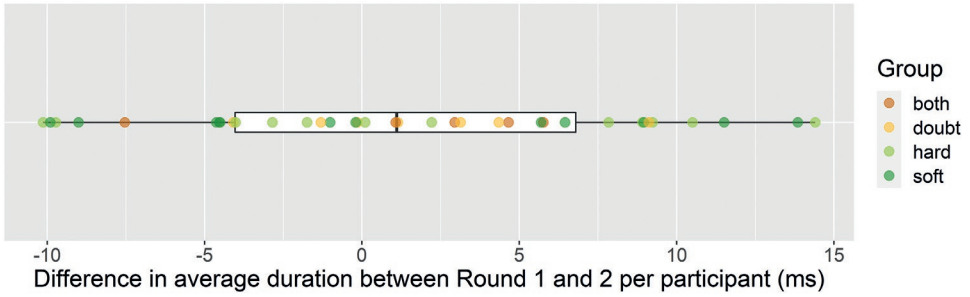


Figure 5.6. Difference in average duration between two rounds with different confederates – Round 1 and Round 2. The dots represent participants, and their colours represent the groups the participants belong to, as based on the pre-test.

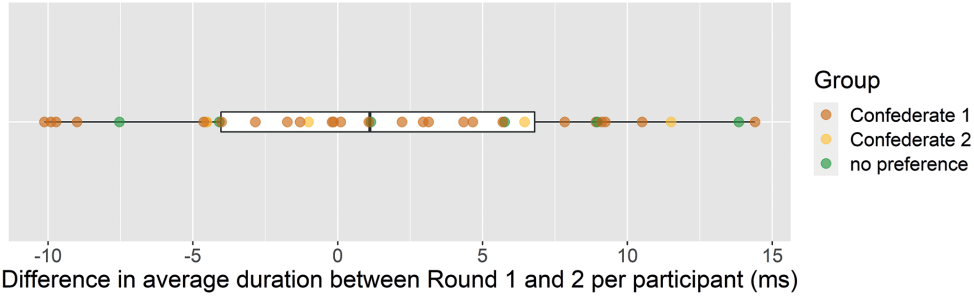


Figure 5.7. Difference in average duration between two rounds with different confederates – Round 1 and Round 2. The dots represent participants, and their colours represent the participants' preference for the confederates' accents.

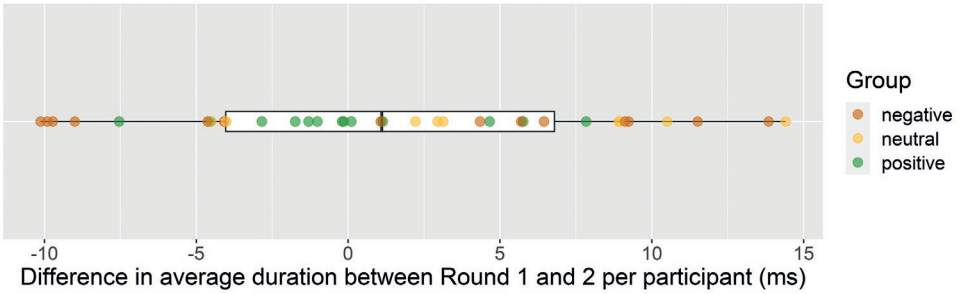


Figure 5.8. Difference in average duration between two rounds with different confederates – Round 1 and Round 2. The dots represent participants, and their colours represent the participants' pride of their own accent.

5.4 Discussion

This study contributes to our knowledge of alignment to regional variants. So far, a restricted number of studies have addressed this topic and the picture emerged so far is not yet clear. We studied speakers' alignment to an allophone in two regional variants of Dutch. This study had two main questions. The first question was whether speakers phonetically align more to a more prestigious variant than to a less prestigious variant and whether this is better reflected in local or global alignment measures. This question was addressed by investigating speakers' productions in different parts of an experiment, in which they took turns in a production task with two different confederates. We compared speakers' productions between the different parts of the experiment and compared any potential differences with those of a control experiment, in which the participants did not hear the confederate. Moreover, we investigated whether their productions are best predicted by the last single production of the interlocutor, by the average of the last ten productions of the interlocutor, or by the average of all productions of the interlocutors so far. Our second research question was whether speakers need to hear the specific feature in order to align to that feature, or whether they also align to that feature when just hearing the returning interlocutor's speech, not containing that feature.

In a sentence completion task, participants interacted with two confederates who differed in their use of the allophonic variant of the fricative /x/. Participants first completed a pre-test by themselves, then interacted with Confederate 1 (who used a so-called "hard g") in Round 1. After Round 1, participants interacted with Confederate 2 (who used a so-called "soft g"), in Round 2. In the inter-test and in Round 3, participants interacted with Confederate 1 again. In the inter-test, this confederate

produced one instance of /x/ (due to a mistake in stimulus construction, the confederate produced one single instance of /x/, where there were meant to be none), which allowed us to test our second question. After these Experiment Parts, participants completed a post-test that was similar to the pre-test, i.e., without interaction with a confederate, which allowed us to investigate possible long-term global effects. This alternation of interaction with different confederates allowed us to test the participants' adjustment of their fricatives to the different regional variants of the interlocutors.

5.4.1 Group effects of alignment

We investigated whether alignment is present by focusing on two acoustic measures: /x/ duration and its CoG. We used these continuous measures instead of subjective perceptual categorisation, because these continuous acoustic measures can also show subtle forms of alignment.

We examined our data in different ways in our search for alignment effects. We first compared the different parts of the Experiments with respect to /x/ duration and CoG. This comparison did not provide any evidence for phonetic alignment to neither of the regional allophones, neither the more prestigious nor the less prestigious variant. Moreover, we did not see any differences between the pre- and the post-tests, which would have indicated possible long-term global effects. The absence of alignment effects between the different Experiment Parts prevents us from drawing conclusions about our second research question, that is, whether speakers need to be presented with a certain feature to align to that feature or whether hearing a returning interlocutor's speech, without this feature, would be enough to change their productions.

Second, we examined our data for local and global effects of alignment by investigating whether the duration of /x/ was co-determined by the most recently produced confederate's /x/, the average of the confederate's last ten /x/s, and the average duration of all instances of /x/s produced by the confederate up to the production of a participant's /x/. These (average) durations are co-determined by the confederate's articulation rate and the segmental and prosodic context of the /x/. We may expect that speakers do not directly copy these (average) durations but rather adapt them to the articulation rate and the segmental and prosodic context in their utterances. Our three continuous measures were therefore not based on the "raw" durations as produced by the confederates, but on durations from which the variation resulting from articulation rate and context was reduced. None of these predictors showed any significant effect. That is, also these more fine-grained analyses did not provide any evidence of phonetic alignment to the regional variants under test.

5.4.2 Individual patterns

Since previous literature has shown that speakers may show considerable interindividual variation in their alignment patterns and since we did not find any evidence at the group level, we proceeded to investigate individual alignment patterns. These analyses indeed showed rather large individual variation. This might explain the absence of a group effect. We investigated whether we could predict whether a given participant showed alignment, no effect of the confederate's speech, or divergence. However, these attempts did not show any clear pattern. That is, we could not identify any clear sub-groups of participants that would allow for a systematic explanation of this large variation. More specifically, we did not see systematic group patterns for: a) groups based on speakers' pre-test productions, or b) groups based on the likeability ratings of the confederates' accents, and c) groups based on speakers' pride of their accent.

Previous studies on alignment to regional variants that are similar to our study, also failed to find overall group effects. Gessinger et al. (2019b), for example, showed that some speakers aligned to the [ɪç] versus [ɪk] contrast in the German suffix <-ig>, some diverged and some simply maintained their use of the suffix. Earnshaw (2020), furthermore, showed that speakers are highly variable in the amount and also in the direction of alignment of an English vowel. Our findings converge with these studies with similar conclusions, but in our case for a Dutch fricative in an interaction with two different interlocutors both using a distinct regional variant.

Finally, one might hypothesise that alignment to regional variants is driven by the relative prestige of these variants. Our findings indicate no group alignment patterns to the regional variant that is most prestigious. Rather, we see large individual variation independent of the prestige. This indicates that a high prestige regional variant does not necessarily elicit alignment. It could be that some speakers do not align because they want to maintain their identity and thus their own accent (see for example Giles and Gasiorek, 2013). We asked participants about their pride of their own accent, but these ratings did not explain the individual differences. This is not unexpected, because only for some speakers pride may overrule the effect of prestige of the other accent while this may not be the case for other speakers. Thus, there are two possibly counteracting forces (i.e., prestige and identity) and the actual result for a given participant will depend on the relative strength of these forces within this given participant.

5.4.3 Phonetic alignment in the literature

In order to see whether there may be other explanations for the absence of clear group alignment effects in our study than social factors, we examined the available studies under which circumstances clear phonetic alignment effects on a group level have been found. There are a few rather clear differences between studies where these effects are found and the present study.

As a first difference, group alignment effects are often reported in studies on suprasegmental features (e.g., Gijssels et al., 2016 and Chapter 3), also referred to as prosodic alignment. Examples of these features are pitch and articulation rate. In contrast to our study, these studies thus do not examine phonemes. Moreover, the suprasegmental features under study did not differ between regional variants and therefore did not carry sociolinguistic information. The differences in presence versus absence of overall alignment effects for these suprasegmental features versus our phonemic properties could be due to the fact that speakers are usually less aware of variation in suprasegmental features than of the variation investigated in our study. This could mean that when speakers are more aware of the variation under study, we find more individual variation in alignment patterns.

Another difference concerns the fact that group alignment effects are often found when alignment is assessed by a perceptual task, most commonly in an AXB assessment (e.g., Pardo, 2006). In these studies, participants are asked to assess potential alignment by deciding whether a lexical item or phrase A or B sounds more like X. Assessing in this manner, as opposed to measuring phonetic features as they are produced by participants (as was done in our study), tends to lead to significant group effects (e.g., Pardo, 2006; Pardo, Jordan, Mallari, Scanlon & Lewandowski, 2013). In the present study, we used automatically extracted measures of phonetic features as produced by the participants in order to more closely investigate alignment effects. On the one hand, it could be that perceptual tasks like the AXB task allow to detect more holistic differences (i.e. there are more cues for a listener to base their judgement on) which are not reflected in single acoustic measures. On the other hand, AXB tasks may be less sensitive in detecting subtle differences among realisations of phones, which we studied here.

A third area within the phonetic alignment literature showing group level phonetic alignment effects, is the shadowing literature (e.g., Pardo et al., 2013). In shadowing tasks, participants are asked to repeat the items they hear. In contrast, our participants were not asked to repeat heard utterances. The task to repeat a heard utterance might induce a stronger tendency to precisely copy the heard input, and this could in turn make any potential alignment effects stronger.

A last area within the phonetic alignment literature where group effects have been found, regards second language speakers (e.g., Troncoso-Ruiz et al., 2019). In most of these studies, second language speakers interact with a native speaker, who implicitly has the role as model speaker, causing the second language speakers to align, in order to improve their second language. In our study, in contrast, speakers might choose to not align to one or the other regional variant because they do not consider that variant to be their model.

5.4.4 Technical challenges

A few other factors could have played a role in the absence of group alignment effects in our study. We used forced alignment to align the phones with the speech signal. Since this alignment works in steps of 10 ms, it cuts off boundaries at 10 ms. It could be that more subtle effects are lost. This is especially relevant since the mean duration difference between the two Confederates was 13.4 ms. Note, however, that the CoG measure confirms the findings from the duration measure.

Another factor that could have led to smaller effects, is the fact that the confederates' and participants' /x/s in our data differed in their prosodic characteristics (e.g., articulation rate and position in the word) and their context (e.g., the preceding phoneme and the following phoneme), which lead to noise in the data. We tried to take care of this noise by predicting participants' /x/ durations on the basis of the residuals from a model reducing the influence of several covariates, rather than on the basis of "raw" durations. However, we most likely have not taken out all irrelevant variation from the productions.

5.5 Conclusion

In conclusion, the present study contributes to the literature on phonetic alignment to regional variants by examining whether speakers align to the regional variants of /x/ in Dutch. We did not find evidence for clear alignment effects at the overall group level. We hypothesise that some speakers did not align because they preferred their own allophonic variant, which reflected their sociolinguistic identity, while others may have aligned to the more prestigious variant. The relative weight of these factors, and possibly others, may explain whether one finds alignment effects.

Appendix C

Duration

Table C1. Output of the raw duration model, with the control experiment, the pre-test, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	92.85	1.84	50.43
Round 1	0.89	0.60	1.48
Round 2	1.68	0.60	2.81
Inter-test	1.85	0.74	2.49
Round 3	1.43	0.60	2.38
Post-test	-0.37	0.63	-0.58
Main experiment	-0.08	1.78	-0.05
Articulation rate	-1.93	0.20	-9.82
Logged word frequency	0.49	0.07	7.01
Word duration (ms)	0.02	<0.01	23.44
Position /x/ – initial	13.41	0.48	27.96
Position /x/ – middle	4.75	0.46	10.24
Category next phone – front vowel	-7.12	0.51	-14.00
Category next phone – obstruent	-16.22	0.55	-29.50
Category next phone – schwa	-32.27	0.44	-72.76
Category next phone – silence	56.25	1.12	50.18
Category next phone – sonorant	-16.37	0.51	-31.80
Category previous phone – front vowel	-5.44	0.43	-12.68
Category previous phone – obstruent	-26.29	0.49	-54.12
Category previous phone – schwa	-15.61	0.48	-32.33
Category previous phone – silence	40.21	0.85	47.58
Category previous phone – sonorant	-23.73	0.45	-52.27
Round 1*Main experiment	-1.79	0.78	-2.30
Round 2*Main experiment	-2.50	0.77	-3.27
Inter-test*Main experiment	-2.81	1.00	-2.82
Round 3*Main experiment	-3.07	0.77	-3.98
Post-test*Main experiment	-1.29	0.86	-1.50

Table C2. Subset duration pre-test output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	101.70	4.34	23.43
Main experiment	0.71	1.96	0.36
Articulation rate	-2.43	0.60	-4.08
Logged word frequency	<0.01	0.22	0.02
Word duration (ms)	0.01	<0.01	5.30
Position /x/ – initial	12.74	1.50	8.52
Position /x/ – middle	0.73	1.42	0.51
Category next phone – front vowel	-3.80	1.53	-2.48
Category next phone – obstruent	-15.44	1.69	-9.14
Category next phone – schwa	-30.77	1.34	-23.02
Category next phone – silence	50.95	3.51	14.52
Category next phone – sonorant	-15.76	1.59	-9.89
Category previous phone – front vowel	-0.40	1.30	-0.31
Category previous phone – obstruent	-26.27	1.48	-17.81
Category previous phone – schwa	-17.22	1.45	-11.88
Category previous phone – silence	43.65	2.49	17.55
Category previous phone – sonorant	-23.17	1.39	-16.63

Table C3. Subset duration Round 1 output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	91.68	3.17	28.95
Main experiment	-2.14	1.86	-1.15
Articulation rate	-1.80	0.42	-4.31
Logged word frequency	0.54	0.15	3.60
Word duration (ms)	0.03	<0.01	12.83
Position /x/ – initial	12.42	1.03	12.02
Position /x/ – middle	4.47	1.04	4.31
Category next phone – front vowel	-8.82	1.07	-8.22
Category next phone – obstruent	-15.72	1.23	-12.75
Category next phone – schwa	-31.80	0.94	-33.86
Category next phone – silence	56.26	2.42	23.20
Category next phone – sonorant	-17.01	1.08	-15.72
Category previous phone – front vowel	-4.59	0.93	-4.92
Category previous phone – obstruent	-27.51	1.05	-26.08
Category previous phone – schwa	-14.56	1.07	-13.56
Category previous phone – silence	37.32	1.74	21.40
Category previous phone – sonorant	-22.50	0.99	-22.78

Table C4. Subset duration Round 2 output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	90.67	3.04	29.87
Main experiment	-2.49	1.70	-1.46
Articulation rate	-1.82	0.43	-4.23
Logged word frequency	0.66	0.14	4.60
Word duration (ms)	0.02	<0.01	11.21
Position /x/ – initial	13.58	1.00	13.57
Position /x/ – middle	6.35	0.96	6.59
Category next phone – front vowel	-9.58	1.02	-9.40
Category next phone – obstruent	-13.78	1.15	-12.01
Category next phone – schwa	-32.13	0.91	-35.26
Category next phone – silence	58.39	2.38	24.54
Category next phone – sonorant	-14.22	1.08	-13.19
Category previous phone – front vowel	-7.74	0.90	-8.63
Category previous phone – obstruent	-25.21	1.00	-25.18
Category previous phone – schwa	-14.49	0.98	-14.84
Category previous phone – silence	40.61	1.68	24.23
Category previous phone – sonorant	-23.16	0.95	-24.35

Table C5. Subset duration inter-test output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	100.06	5.27	18.99
Main experiment	-2.48	1.94	-1.27
Articulation rate	-2.52	0.71	-3.55
Logged word frequency	0.29	0.26	1.10
Word duration (ms)	0.02	<0.01	5.86
Position /x/ – initial	15.31	1.75	8.77
Position /x/ – middle	4.92	1.66	2.97
Category next phone – front vowel	-8.21	1.99	-4.12
Category next phone – obstruent	-18.49	2.03	-9.10
Category next phone – schwa	-35.24	1.65	-21.36
Category next phone – silence	55.62	3.99	13.93
Category next phone – sonorant	-17.91	2.01	-8.91
Category previous phone – front vowel	0.73	1.57	0.47
Category previous phone – obstruent	-24.51	1.69	-14.50
Category previous phone – schwa	-13.60	1.77	-7.71
Category previous phone – silence	38.16	3.52	10.85
Category previous phone – sonorant	-24.26	1.55	-15.63

Table C6. Subset duration Round 3 output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	101.19	3.07	32.97
Main experiment	-3.03	1.80	-1.68
Articulation rate	-2.54	0.43	-5.89
Logged word frequency	0.40	0.14	2.84
Word duration (ms)	0.02	<0.01	10.14
Position /x/ – initial	12.05	0.99	12.21
Position /x/ – middle	3.92	0.96	4.10
Category next phone – front vowel	-6.32	1.05	-6.04
Category next phone – obstruent	-19.43	1.13	-17.21
Category next phone – schwa	-33.28	0.91	-36.54
Category next phone – silence	49.37	2.29	21.60
Category next phone – sonorant	-18.15	1.04	-17.44
Category previous phone – front vowel	-8.24	0.90	-9.14
Category previous phone – obstruent	-27.17	1.06	-25.60
Category previous phone – schwa	-16.28	1.03	-15.87
Category previous phone – silence	42.31	1.92	22.00
Category previous phone – sonorant	-23.22	0.96	-24.14

Table C7. Subset duration post-test output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	79.95	3.93	20.35
Main experiment	-1.86	1.90	-0.98
Articulation rate	-1.09	0.56	-1.93
Logged word frequency	0.73	0.19	3.78
Word duration (ms)	0.02	<0.01	9.56
Position /x/ – initial	19.23	1.36	14.15
Position /x/ – middle	9.31	1.27	7.31
Category next phone – front vowel	-8.91	1.48	-6.03
Category next phone – obstruent	-12.32	1.52	-8.10
Category next phone – schwa	-33.36	1.29	-25.80
Category next phone – silence	63.81	3.33	19.18
Category next phone – sonorant	-14.08	1.43	-9.84
Category previous phone – front vowel	-1.09	1.19	-0.92
Category previous phone – obstruent	-25.45	1.29	-19.66
Category previous phone – schwa	-16.78	1.34	-12.50
Category previous phone – silence	38.44	2.46	15.60
Category previous phone – sonorant	-24.32	1.21	-20.17

Centre of Gravity

Table C8. Output of the raw CoG model with the control experiment, the pre-test, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	249.23	71.57	3.48
Round 1	-77.56	29.00	-2.68
Round 2	-51.41	28.84	-1.78
Inter-test	-83.17	36.09	-2.30
Round 3	-91.29	29.04	-3.14
Post-test	-33.13	30.73	-1.08
Main experiment	-184.65	93.61	-1.97
Position /x/ – initial	479.25	23.30	20.57
Position /x/ – middle	211.91	21.06	10.06
Category next phone – front vowel	1008.78	24.53	41.12
Category next phone – obstruent	1740.47	26.50	65.68
Category next phone – schwa	915.39	20.76	44.10
Category next phone – silence	1606.44	37.47	42.88
Category next phone – sonorant	1183.32	24.73	47.86
Category previous phone – front vowel	208.90	20.33	10.28
Category previous phone – obstruent	1662.99	23.01	72.27
Category previous phone – schwa	637.58	22.73	28.06
Category previous phone – silence	1162.84	38.94	29.87
Category previous phone – sonorant	526.12	21.75	24.19
Round 1*Main experiment	87.19	37.66	2.32
Round 2*Main experiment	66.11	37.10	1.78
Inter-test*Main experiment	87.23	48.63	1.79
Round 3*Main experiment	64.08	37.42	1.71
Post-test*Main experiment	-15.51	41.71	-0.37

Table C9. Subset CoG pre-test output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	362.49	108.44	3.34
Main experiment	-185.03	104.50	-1.77
Position /x/ – initial	348.22	65.09	5.35
Position /x/ – middle	112.40	58.00	1.94
Category next phone – front vowel	881.51	66.01	13.35
Category next phone – obstruent	1588.86	71.66	22.17
Category next phone – schwa	900.99	55.32	16.29
Category next phone – silence	1498.70	101.20	14.81
Category next phone – sonorant	1030.41	67.63	15.24
Category previous phone – front vowel	230.36	54.51	4.23
Category previous phone – obstruent	1691.86	60.94	27.76
Category previous phone – schwa	645.60	59.23	10.90
Category previous phone – silence	1199.55	104.32	11.50
Category previous phone – sonorant	627.79	58.94	10.65

Table C10. Subset CoG Round 1 output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	24.34	87.89	0.28
Main experiment	-116.37	91.06	-1.28
Position /x/ – initial	401.20	49.04	8.18
Position /x/ – middle	224.61	46.15	4.87
Category next phone – front vowel	1111.56	50.64	21.95
Category next phone – obstruent	1910.99	58.06	32.92
Category next phone – schwa	982.62	42.92	22.90
Category next phone – silence	1775.34	79.04	22.46
Category next phone – sonorant	1283.15	50.91	25.20
Category previous phone – front vowel	269.14	43.24	6.22
Category previous phone – obstruent	1824.34	48.81	37.37
Category previous phone – schwa	700.67	49.26	14.22
Category previous phone – silence	1413.01	77.23	18.30
Category previous phone – sonorant	687.72	46.24	14.87

Table C11. Subset CoG Round 2 output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	165.46	87.56	1.89
Main experiment	-135.34	93.85	-1.44
Position /x/ – initial	521.54	47.45	10.99
Position /x/ – middle	264.86	42.95	6.17
Category next phone – front vowel	1190.07	48.15	24.72
Category next phone – obstruent	1843.20	54.38	33.90
Category next phone – schwa	877.32	41.98	20.90
Category next phone – silence	1670.44	77.07	21.68
Category next phone – sonorant	1277.53	51.03	25.03
Category previous phone – front vowel	117.34	41.71	2.81
Category previous phone – obstruent	1640.88	46.69	35.14
Category previous phone – schwa	622.69	45.26	13.76
Category previous phone – silence	1138.90	76.86	14.82
Category previous phone – sonorant	404.35	44.57	9.07

Table C12. Subset CoG inter-test output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

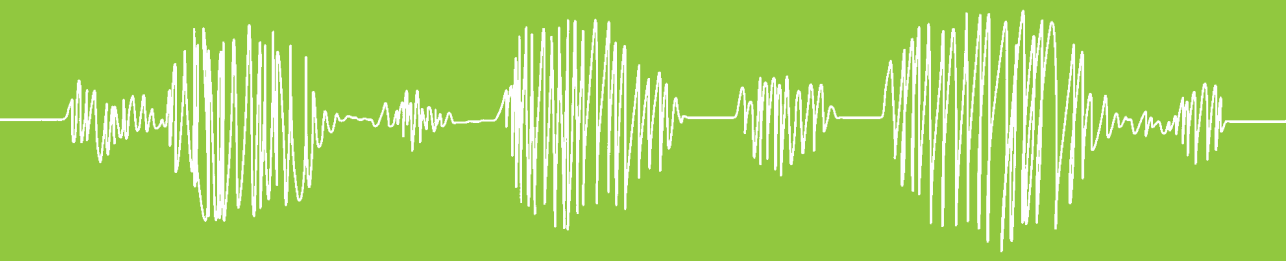
Variable	Estimate	Standard error	T value
Intercept	464.84	128.53	3.62
Main experiment	-138.01	111.97	-1.23
Position /x/ – initial	447.34	82.88	5.40
Position /x/ – middle	91.75	74.40	1.23
Category next phone – front vowel	687.24	93.43	7.36
Category next phone – obstruent	1400.71	94.77	14.78
Category next phone – schwa	753.29	74.21	10.15
Category next phone – silence	1378.50	136.74	10.08
Category next phone – sonorant	956.79	92.38	10.36
Category previous phone – front vowel	216.07	72.09	3.00
Category previous phone – obstruent	1701.90	75.97	22.40
Category previous phone – schwa	674.92	79.50	8.49
Category previous phone – silence	1062.44	162.96	6.52
Category previous phone – sonorant	475.24	73.15	6.50

Table C13. Subset CoG Round 3 output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	177.21	90.08	1.97
Main experiment	-144.77	96.73	-1.50
Position /x/ – initial	597.41	48.47	12.33
Position /x/ – middle	274.19	43.23	6.34
Category next phone – front vowel	997.13	50.85	19.61
Category next phone – obstruent	1771.08	55.14	32.12
Category next phone – schwa	973.47	43.09	22.59
Category next phone – silence	1621.00	78.48	20.65
Category next phone – sonorant	1264.46	50.51	25.03
Category previous phone – front vowel	162.77	43.25	3.76
Category previous phone – obstruent	1475.10	50.85	29.01
Category previous phone – schwa	529.38	48.81	10.85
Category previous phone – silence	893.69	90.48	9.88
Category previous phone – sonorant	430.39	46.39	9.28

Table C14. Subset CoG post-test output - with the control experiment, the /x/ in final position, and the back vowel as category for the next and the previous phone on the intercept.

Variable	Estimate	Standard error	T value
Intercept	288.49	106.09	2.72
Main experiment	-216.87	105.54	-2.06
Position /x/ – initial	540.40	64.52	8.38
Position /x/ – middle	227.62	55.10	4.13
Category next phone – front vowel	901.44	69.18	13.03
Category next phone – obstruent	1605.81	71.20	22.56
Category next phone – schwa	829.22	58.26	14.23
Category next phone – silence	1549.70	108.21	14.32
Category next phone – sonorant	1024.01	66.98	15.29
Category previous phone – front vowel	142.35	55.59	2.56
Category previous phone – obstruent	1651.88	59.69	27.67
Category previous phone – schwa	685.55	62.56	10.96
Category previous phone – silence	1057.89	109.40	9.67
Category previous phone – sonorant	538.24	56.72	9.49



CHAPTER 6

The CABB dataset: A multimodal corpus of communicative interactions for behavioural and neural analyses

This chapter is based on:
Eijk, L.*, Rasenberg, M.*, Arnese, F., Blokpoel, M., Dingemanse, M., Doeller, C. F., Ernestus. M., Holler, J., Milivojevic, B., Özyürek, A., Pouw, W., van Rooij, I., Schriefers, H., Toni, I., Trujillo, J., & Bögels, S. (2022). The CABB dataset: A multimodal corpus of communicative interactions for behavioural and neural analyses. *NeuroImage*, 119734. https://doi.org/10.1016/j.neuroimage.2022.119734

* Shared first author

The Data Use Agreement for this dataset is available in the published version of this chapter.

Abstract

We present a dataset of behavioural and fMRI observations acquired in the context of humans involved in multimodal referential communication. The dataset contains audio/video and motion-tracking recordings of face-to face, task-based communicative interactions in Dutch, as well as behavioural and neural correlates of participants' representations of dialogue referents. Seventy-one pairs of unacquainted participants performed two interleaved interactional tasks in which they described and located 16 novel geometrical objects (i.e., Fribbles) yielding spontaneous interactions of about one hour. We share high-quality video (from three cameras), audio (from head-mounted microphones), and motion-tracking (Kinect) data, as well as speech transcripts of the interactions. Before and after engaging in the face-to-face communicative interactions, participants' individual representations of the 16 Fribbles were estimated. Behaviourally, participants provided a written description (one to three words) for each Fribble and positioned them along 29 independent conceptual dimensions (e.g., rounded, human, audible). Neurally, fMRI signal evoked by each Fribble was measured during a one-back working-memory task. To enable functional hyperalignment across participants, the dataset also includes fMRI measurements obtained during visual presentation of eight animated movies (35 min total). We present analyses for the various types of data demonstrating their quality and consistency with earlier research. Besides high-resolution multimodal interactional data, this dataset includes different correlates of communicative referents, obtained before and after face-to-face dialogue, allowing for novel investigations into the relation between communicative behaviours and the representational space shared by communicators. This unique combination of data can be used for research in neuroscience, psychology, linguistics, and beyond.

6.1 Introduction

Language is a key socio-cognitive human function predominantly used in interaction. Yet, much work in linguistics and cognitive neuroscience has focused on individuals' coding-decoding of signals according to their structural dependencies. Understanding the communicative use of language requires shifting the focus of investigation from individual competencies to the mechanisms used by interlocutors to understand each other during live interactions. Here, we provide a dataset that can be used to study face-to-face, multi-turn referential communication between pairs of interlocutors (through audio/video and motion-tracking recordings), as well as individuals' representations of the dialogue referents (as estimated from behavioural and fMRI data collected before and after the dialogue).

The dataset presented here emerges from CABB (Communicative Alignment of Brain and Behaviour), a research program focused on studying interactive language use. This program builds on the notion that interlocutors can disambiguate referentially flexible signals by building a shared cognitive space (e.g., Clark, 1997; Clark & Brennan, 1991; Hutchins & Hazlehurst, 1995; Stolk et al., 2016, 2022). A shared cognitive space involves not only presumed common ground, the propositions jointly taken for granted or communicated, but also mutual awareness of the circumstances of communication, and thus the likely joint goals, norms, and affordances of the event, embedded in the recent interactional history. Besides the traditional focus on transfer of propositional content, this research initiative considers how language use is organised to achieve interactional goals and to monitor mutual understanding, and how interlocutors create and control a shared cognitive space during live communicative interactions. CABB considers the contribution of multimodal communicative resources (speech, gestures) at different levels of linguistic structure (from phonology to pragmatics) during interactive task-based dialogue.

The interactional part of the dataset consists of audio, video, and body-movement recordings of face-to-face communicative interactions in Dutch between 71 pairs of participants, without restrictions on communicative means (e.g., speech, gestures), timing, turn-taking, or feedback. Participants communicate about 16 novel visual objects which lack conventional labels - called "Fribbles" (Barry et al., 2014). The Fribbles (see Figure 6.1) were designed and pre-tested for their ability to evoke different conceptualisations across individuals. As such, different pairs would need to work together to create their own pair-specific conceptualisations and labels for them, enabling us to see effects of the interaction rather than mere exposure to the stimuli. The participants were instructed to communicate in order to identify (Referential task) and localise (Localisation task) the Fribbles on a screen. These tasks are designed to capture a core element of everyday human communication: each pair needs to create mutually understood

utterances, dependent on the situated context of the ongoing interaction (Clark, 1996; Stolk et al., 2022). Participants were not familiar with the Fribbles at the onset of the study, a task feature designed to amplify this process of negotiating a common referent that arguably occurs in many communicative interactions. More precisely, in the Referential task, participants need to negotiate referential expressions for the Fribbles to be able to identify them amidst the total set, similar to referential tasks with tangrams (see e.g., Clark & Wilkes-Gibbs, 1986; Holler & Wilkin, 2011). In the Localisation task, each pair needs to work out collaboratively whether a particular Fribble is located at the same position on their respective screens. Participants had equal opportunities to speak in the interaction, since they switched roles throughout the task.

The dataset contains high-quality audio recordings using head-mounted microphones, along with time-aligned orthographic transcriptions of the speech for 47 out of the 71 interactions (see *Methods*). These enable different types of linguistic analyses of individual participants' speech (e.g., lexical, semantic, phonetic), as well as investigations of alignment between participants on these levels (e.g., Pickering & Garrod, 2004). Moreover, high-quality video recordings from three different angles, as well as 3D body motion-tracking data from two Microsoft Kinects (V2), allows researchers to analyse participants' movements, postures, and gestures, as well as their alignment between participants. The face-to-face set-up, where participants stood opposite each other and had full vision of each other's torso, facilitated the use of gestures (although this was in no way explicitly encouraged).

The dataset also provides estimates of participants' individual representations of the Fribbles using two behavioural measures and one neuroimaging (fMRI) measure. These measures are taken both before (pre) and after (post) the face-to-face interaction. Behaviourally, participants named each Fribble using one to three words (Naming task), and rated each Fribble on 29 different visual and semantic features (Features task; based on Binder et al., 2016). Neurally, participants' brain responses to the Fribble images were measured using fMRI while they performed a one-back working memory task to monitor their attention to the stimuli, following earlier studies using neural representational approaches (e.g., Bracci et al., 2015; Dobs et al., 2019). By containing both the pre and post measures, this dataset is well suited for measuring changes in estimated individual representations of each referent Fribble, as well as the extent of convergence of such estimated representations within each pair, brought about by the interaction. Comparison of across-voxel activity patterns of fMRI responses to the Fribbles across participants is enhanced by the possibility of implementing so-called "hyperalignment" (Haxby et al., 2011). Namely, the dataset includes fMRI data of participants watching the same eight animated movies (about 35 min in total). This enables the fMRI pre-processing step of aligning individual brains to a common information space across the sample, based on functional (instead of anatomical) similarities between the brains.

That is, voxels from different brains that are similarly activated in response to the same stimuli while watching the movies are aligned to each other. Hyperalignment is especially relevant for the present dataset because it allows for more direct comparisons between activation patterns caused by the same Fribble in different brains.

To date, this dataset is unique in that it combines multimodal interactional data with behavioural and neural characterisation of the representational consequences of a face-to-face communicative interaction. The interactional, behavioural, and neuroimaging data can be used for addressing a wide range of research questions within and across various disciplines such as linguistics, neuroscience, and psychology. Furthermore, the dataset offers the possibility to combine those measures and investigate how face-to-face multimodal naturalistic communication changes the estimated representations of the referents within and across interlocutors.

In recent years, open access brain-imaging datasets have increasingly become available, providing different types of data (e.g., resting state, task-related) from multiple brain-imaging methods (i.e., EEG, MEG, fMRI), such as the Human Connectome Project (Van Essen et al., 2013), the CamCan dataset (Taylor et al., 2017), and the MOUS dataset (Schoffelen et al., 2019). However, none of these quantify the consequences of communicative interactions with (behavioural and) fMRI observations. The unique characteristics of this dataset can also be appreciated by comparing it to existing corpora with recordings of social interaction. Interactional corpora consisting of audio data are rather numerous (for an overview, see Ernestus & Baayen, 2011), containing for example spontaneous face-to-face and telephone conversations in Dutch as in the Corpus Gesproken Nederlands (CGN; Oostdijk, 2000), or task-based interactions in Scottish English as in the HCRC Map Task corpus (Anderson et al., 1991). Examples of multimodal corpora, consisting of both video and audio data, are the InSight Interaction Corpus (Dutch; Brône and Oben, 2015), the IFADV corpus (Dutch; Van Son et al., 2008), the Spontal corpus (Swedish; Edlund et al., 2010), and the Nijmegen Corpus of Casual French (Torreira et al., 2010). These corpora include many aspects of multimodal communication, but do not provide the combination of high-quality audio, video, and motion tracking necessary to implement fine grained integrative analyses of both gestures and speech. At least one other dataset (Rauchbauer et al., 2019) also combines multi-modal interactive data (speech, eye-movements, and face-recordings) with fMRI measurements. Differently from our dataset, the fMRI data were acquired in individual participants while they were interacting with a human or a robot. With 71 interactions (47 fully transcribed), the present corpus provides ample possibilities for rich qualitative and quantitative studies of communicative interactions. This dataset also opens up new research avenues as observations from the interaction can be related to correlates of individuals' representations of the dialogue referents as estimated from the behavioural and neuroimaging measures.

A precursory dataset of the CABB team (with a similar paradigm, but without fMRI data) has been used in earlier reports (Pouw, De Wit, et al., 2021; Rasenberg et al., 2022), and further reports on the present dataset are in preparation. This contribution is intended to describe the dataset with respect to the procedures used in the acquisition as well as some example analyses, and make it available for use by other researchers. From here onwards we refer to this as the Dataset (along with a folder name). See Section 6.2.7.2 for information on how to access the Dataset.

6.2 Methods

6.2.1 Participants

In total, 142 right-handed, native Dutch speakers (71 pairs; 30 all-female, 7 all-male, and 34 mixed gender pairs, according to self-reported data) participated in the study, with an average age of 22.86 years (*SD* = 3.63, *range* = 18–33 with one outlier of 45). All participants reported no neurological or language-related disorders, no metal implants (except for dental) in their body, no history of brain surgery, no hearing impairments, and normal or corrected-to-normal vision. The participants were recruited via the Radboud SONA participant pool system. Data and transcriptions of 37 pairs (74 participants) from all tasks are fully complete and shared (see Section 6.2.7.1 for details on the availability and quality of various parts of the Dataset).

6.2.2 Ethical approval and participant consent

This study met the criteria of the blanket ethical approval for standard studies of the Commission for Human Research Region Arnhem-Nijmegen (DCCN CMO 2014/288). Participants were emailed information about the study in advance and verbally informed on the testing day itself. Written informed consent was obtained before data collection started. Participants agreed to the sharing of the fully anonymised data[8], and could optionally agree to the sharing of potentially identifiable audio/video data with researchers for scientific purposes and/or for educational and/or promotional purposes, through (a) presentations/lectures (not publicly available), (b) newspapers, magazines/journals or other (online) news outlets, (c) social media, and d) television. See the Participants folder in the Dataset for the full overview of data sharing consent.

8 Note that defaced structural MRI data are technically only pseudonymised and not fully anonymised. However, the consent form for standard studies that we used dated from 2018, which was before this issue was recognised in the scientific and legal community.

6.2.3 Materials

The experimental stimuli consisted of 16 pictures of blue 3D objects made up of geometrical figures attached to each other, on a grey background, which we refer to as Fribbles (see Figure 6.1; note that the term "Fribbles" was never mentioned to participants). We adapted these stimuli from objects also called Fribbles (Barry et al., 2014). The adaptation was based on pilot tests, in which participants individually named each Fribble using one to three words (see Naming task explained in Section 6.2.5.1 below) and/or played the Referential communication game in pairs (see Section 6.2.5.5 below). These pilots resulted in a final set of Fribbles (Figure 6.1) which evoked variable conceptualisations (names) across both individuals and pairs. This was important to be able to control for general aspects of the interaction by comparing convergence (e.g., in labels) between real interacting pairs and pseudo-pairs, i.e., pairs who did not interact with each other (see e.g., Section 6.3.5 below).

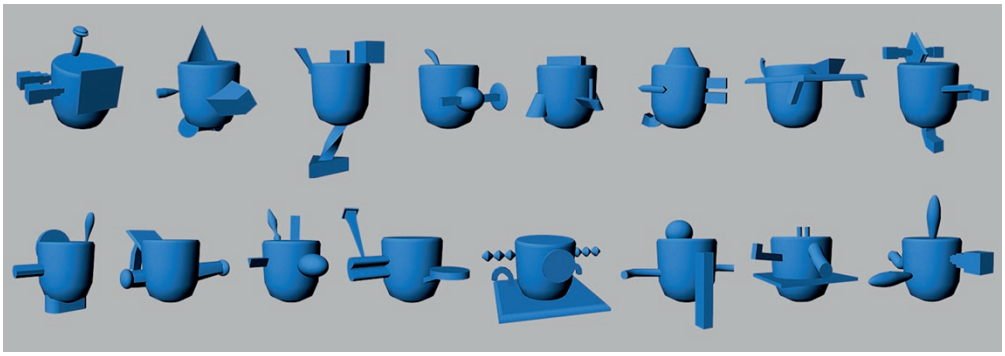


Figure 6.1. The 16 stimuli (Fribbles, based on Barry et al., 2014) used in the different tasks of the study, designed to evoke various conceptualisations.

6.2.4 Set-up and apparatus

6.2.4.1 MRI apparatus & (f)MRI image acquisition

Magnetic resonance images were acquired using two 3T MAGNETROM MR scanners: Prisma and PrismaFit (Siemens AG, Healthcare Sector, Erlangen, Germany). For the functional acquisition a multi-band 2D-EPI sequence released as part of the Human Connectome Project (Uğurbil et al., 2013) was used. Functional images were acquired using a multi-band six sequence. The parameters of the acquisition were: TR/TE = 1,000/34 ms, flip angle = 60°; 2mm^3 isotropic resolution over a FOV = 208 × 208 × 132 mm; Multi-band acceleration of six was used in the slice direction and no parallel imaging was applied in-plane. Phase encoding was applied on the AP direction with a partial Fourier coverage of 7/8, including five volumes with reversed phase encoding (A >> P), which can be used to correct image distortions. Approximately 750 vol were

acquired in each of the four one-back runs (two in the pre session and two in the post session) and 2,074 vol in the movies run (session three).

A T1-weighted scan was acquired at the end of the second session in the sagittal orientation using a 3D MPRAGE sequence with the following parameters: TR/TI/TE = 2,400/1,000/2.22 ms, 8° flip angle. Following, a T2-weighted scan was also acquired in the sagittal orientation using a variable flip angle TSE with the following parameters: TR/TE = 3,200/563ms, echo spacing = 3.52 ms, Turbo Factor = 314. Both the T1 and T2 used a FOV of 256 × 240 × 167 mm, a 0.8 mm^3 isotropic resolution, and parallel imaging (iPAT = 2) to accelerate the acquisition resulting in an acquisition time of 6 min 38 s for T1 and 5 min 57 s for T2. At the start of each of the three sessions, an additional fast T1 weighted scan was obtained using a spoiled gradient echo sequence with the following contrast parameters: TR/TE =6.31/3.2 ms; flip angle = 11°. Acquisition was performed in the sagittal orientation with a FOV of 176 × 256 × 256 mm and 1 mm^3 isotropic. A five-fold controlled aliasing acceleration was used resulting in a total acquisition time of 1 min 17 s.

Stimuli were presented using an EIKI LC-XL100 beamer with a resolution of 1,024 × 768 and a refresh rate of 60 Hz, and were projected onto a screen behind the scanner bore. Participants were able to see the screen via a mirror. Given different characteristics of the two scanners used, the image sizes for the Fribble stimuli were adjusted such that all participants experienced all Fribbles at the same visual angle in both scanners.

During fMRI acquisition, participants' attention levels were monitored by single-eye recording, using an infrared source eye-tracker. Also, respiration and heartbeat were recorded using a respiration belt and a pulse wave sensor, respectively; both required the same MRI-compatible amplifier from BrainAmp ExG MR.

6.2.4.2 Set-up and apparatus of the interaction

The interaction took place in a sound-attenuated booth. Participants of a pair faced each other about two meters apart while standing in front of a table (see Figure 6.3, middle panel). Each participant faced a 24" screen (BenQ XL2430T), slightly tilted for an optimal viewing angle, and positioned at hip height. This ensured that participants could see each other, and prevented interference with the participants' gesture space (McNeill, 1992). All 16 Fribbles were simultaneously presented on each participant's screen, in a random arrangement over a grey background, each Fribble covering 4x4 cm on the screen (see Figure 6.4). The Fribbles were labelled with numbers for one participant and letters for the other. Button boxes (with a red and a yellow button) were positioned below the screen and were used by the participants to provide answers (for the Localisation task, but not the Referential task) and/or to move to the next trial (see Section 6.2.5).

Video recordings were made with a frame rate of 29.97 frames per second (fps) at 1,920 × 1,080 resolution using three HD cameras (JVC GY-HM100/150); cameras 1 and 2 were positioned to the side to yield (semi-)frontal-views of each participant, while camera 3 was positioned in the middle to yield an overview of both participants (Figure 6.2). Two head-mounted microphones (Samson QV) were used to record speech for each participant separately. These microphones were connected to an AudiTon pre-amplifier and then to a Roland R-05 recorder, which were both situated in the control room, where the experimenter could listen to and adjust the volume of the incoming audio. The output of the pre-amplifier (which consisted of two separate audio channels, one for each participant) was transmitted to the recorder (where the audio was digitised at a sampling rate of 44.1 kHz and a 16-bit quantisation), the output of which was transmitted to cameras 1 and 2, respectively (digitised at 48 kHz and 16-bit). Two Microsoft Kinects (V2), positioned next to cameras 1 and 2 (Figure 6.2) were used to collect 3D positional joint tracking data (for 25 joints) at 30 fps. During data collection, the experimenters monitored the Kinect pose skeleton tracking which served as an online quality check of the Kinect tracking.

Since recordings were started manually on the various devices, all audio, video, and motion-tracking data was synchronised off-line (see Section 6.2.6). To facilitate this process, a dedicated "synchronisation signal" device was used: every 60 s the device sent a digital code to the laptops controlling the Kinect (stored in log files), and a beep as audio input to the cameras (recorded on a secondary audio channel, separately from the speech). See Figure D1 in Appendix D for a schematic overview of all materials in the interaction setup and their connections.

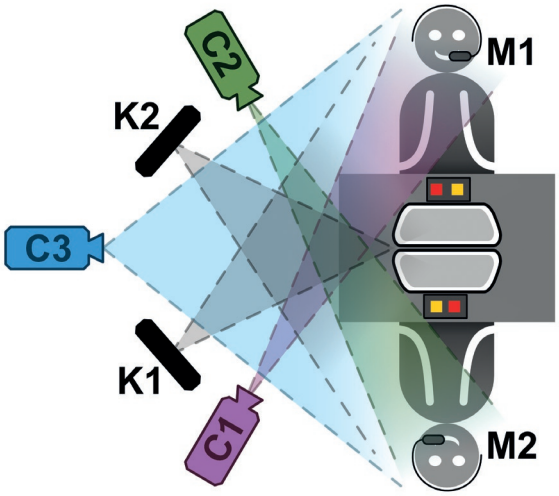


Figure 6.2. Recording set-up for the interaction (C1-C3: cameras, K1-K2: Kinect, M1-M2: microphones).

6.2.5 Procedure

Participants came to the lab in pairs (but did not know each other beforehand) performing several individual tasks (i.e., Naming task, Features task, one-back task in the fMRI) before and after a joint interactional task (i.e., Referential and Localisation tasks) followed by another fMRI session (movie watching) and a questionnaire. All tasks were programmed using the Presentation software (Version 20.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). See Figure 6.3 for an overview of all tasks and the next sections (Section 6.2.5.1 – Section 6.2.5.6) for a detailed description (for full Dutch Instructions for all tasks, see the Presentation NBS scripts in the Presentation_ scripts folder in the Dataset). Before starting, participants provided informed consent (Section 6.2.2) and were asked not to talk to each other before the interaction part of the study and not to talk about the tasks in the break(s) between tasks during the entire session. The session lasted for six to eight hours in total and included a lunch break of at least 30 min immediately after the interaction. Whenever possible, two pairs were tested on the same day in an interleaved fashion, which meant participants had another break of maximally 45 min before the last scanner session and were asked to fill out most of the questionnaire in this break.

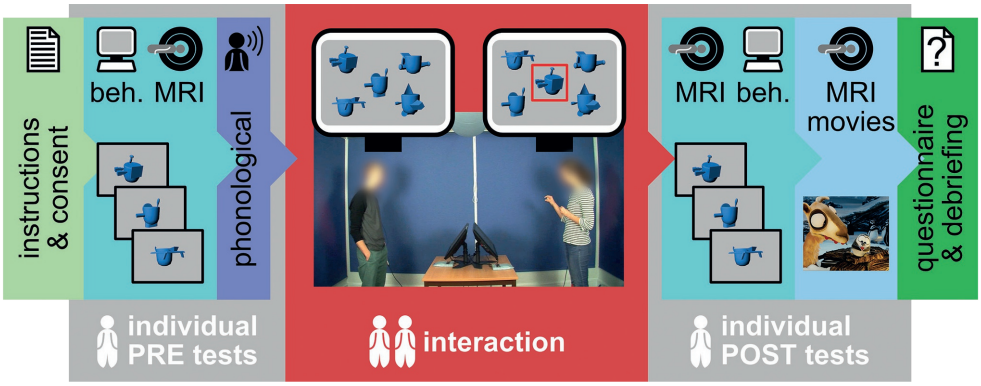


Figure 6.3. Overview of participants' tasks during the testing day. *Beh.* = behavioural tasks (Naming and Features). *MRI* = magnetic resonance imaging (task: one-back task in sessions one and two; movies in session three). *Phonological* = phonological pre-test. *PRE* = before the interaction, *POST* = after the interaction.

6.2.5.1 Naming and Features tasks

Participants performed two behavioural tasks, the Naming and the Features task, individually in sound-proofed cubicles, while sitting in front of a 24 inch, full HD screen and responding using a keyboard and a mouse. The two tasks were presented in an interleaved fashion so that for each Fribble, participants first performed the

Naming task and then the Features task before moving on to the next Fribble. The order of presentation of the Fribbles was randomised per participant. Participants received written instructions for both tasks on the screen, were given the opportunity to ask questions, and then received an oral summary of the instructions from the experimenter.

For the Naming task, all Fribbles were presented on the screen simultaneously. The position of the 16 Fribbles was randomised separately for each participant, but was the same for all trials within participants. On each trial, one Fribble was marked with a red square. Participants were instructed to name or describe that Fribble using one to three words in such a way that the other participant would be able to find it amongst all other Fribbles on the screen (see Figure D2 in Appendix D for an example screenshot of a Naming task trial).

For the Features task, the same Fribble they had just named was presented in the left top corner of the screen with a lead-in sentence next to it ("To what extent do you view this picture as…"). Underneath, 29 different features were shown in the form of linguistic labels (to be read as completing the lead-in sentence). The features were based on a study by Binder and colleagues (2016) and our own selection given the range of results in the pilot Naming task (categories of objects). Some examples of features are "Rounded", "Symmetrical", "Human", and "Positive" (see Table D1 in Appendix D or Dataset: Data for the full list). Participants were instructed to judge to what extent each feature was compatible with their view of the Fribble by moving a slider underneath the feature label (left: "not at all", right: "very strongly"). They were instructed to decide within a few seconds for each feature and to choose the leftmost position on the bar if a feature was not applicable or neutral. Only after participants had moved the slider for each of the 29 features, they could press enter to continue to the next Fribble (see Figure D3 in Appendix D for an example screenshot of a Features task trial).

The tasks were the same when participants performed them for the second time (after the interaction), but with a different random order of Fribble presentations per participant (both on the screen and over trials). Participants were told that they were allowed to give the same name as the first time, but that they did not have to. They were again instructed to describe each Fribble such that their partner would be able to find it amongst the others. Both before and after the interaction 45 min were planned for the Naming and Features task.

6.2.5.2 fMRI one-back task

While still in the cubicles, participants received written instructions for the scanner-based one-back task. They were instructed to press one button when the picture they saw was the same as the previous picture and another button in all other cases (also for the first picture). They were then given an oral summary and performed a short test block with the one-back task using different pictures than the ones used in the actual task

(seven trials). This test block was repeated until all responses were correct. In the MRI scanner, participants read the instructions on the screen again while localiser scans were acquired and then again performed the same practice block (to practice with the response buttons in the scanner) until all responses were correct. When necessary, additional instructions were given over the intercom. In the scanner, participants gave responses on a button-box using the index and middle fingers of their right (dominant) hand. The allocation of the fingers to "same" and "different" responses was counterbalanced over participants, but was the same for all sessions per participant.

After the onset of the fMRI sequence, Fribbles were presented one by one, slightly below the centre of the screen (to avoid vertical head movement at the onset of a Fribble), on a grey background, for two seconds, with a visual angle of about five degrees. This visual angle ensured visibility of the Fribble, while discouraging large saccades. In between Fribble presentations, a fixation cross appeared centred at the same position. In each scanning session, participants saw 12 presentations of each Fribble, as well as 32 catch trials (two per Fribble) in which the Fribble was repeated. These catch trials were used to monitor participants' attention to the stimuli. The Fribbles were presented in a jittered design, with inter-stimulus-intervals (ISI) of three, four, or five seconds. Each Fribble was preceded by all ISIs four times. The order of presentation was different per participant but the same for both one-back fMRI sessions (i.e., pre and post interaction). Each session was divided into two runs, to give participants a break, stopping scanning in between but leaving participants in the scanner for a few minutes. Each run consisted of three blocks, each containing two presentations of all Fribbles in random order plus five to seven pseudo-randomly interleaved catch trials. In between blocks, there was a 20 s break while a summary of the instructions was presented on the screen. The last five seconds of the pause were counted down on the screen. Each session lasted about 30 min in total.

6.2.5.3 Phonological pre-test

After the fMRI session, but before the interaction, we implemented a pre-test to provide a baseline for potential analyses of participants' speech production in the interaction. Participants were tested one-by-one in the same sound-attenuated booth and using the same audio equipment as for the recording of the interaction. When one of the participants was doing the pre-test, the other participant waited in a separate room, wearing Bose Quietcomfort 35 ii noise-cancelling headphones. The pre-test consisted of two parts. The first part provided a baseline for vowel and diphthong productions. Participants were instructed to read aloud 16 Dutch (non-)words that were presented on the screen. These words consisted of Dutch vowels and diphthongs ([ʏ], [ɛ], [ɪ], [ɔ], [ɑ], [a], [e], [ə], [o], [y], [i], [ø], [u], [ɛɪ], [œy], [ɑʊ]) preceded by an <h> and followed by a <t>, which served as a neutral and constant phonetic context across vowels and

diphthongs. The second part of the pre-test served as a baseline for other acoustic characteristics (such as articulation rate, pitch, and /x/ in particular, since this consonant shows clear variation in Dutch) and elicited semi-spontaneous speech. Participants were instructed to read aloud five beginning parts of sentences (ranging from seven to ten words) and to complete them with the first completion that came to their mind. In both parts of the pre-test, participants could click any button on the button box to go to the next (non-)word or sentence. The (non-)words and sentence beginnings were presented centred on the screen. The pre-test lasted about three minutes in total.

6.2.5.4 Interaction

Participants received instructions about the interaction prior to the phonological pre-test. After participants indicated they had finished reading the instructions on their screen, the most important points of the instructions were verbally repeated by the experimenter and participants could ask questions. They then jointly received instructions for the phonological pre-test, which they individually completed in the sound-attenuated booth. When the participants were ready to start with the interaction, they entered the same sound-attenuated booth together and positioned themselves (standing up) behind their respective screens (see Section 6.2.4.2 and Figure 6.3, middle panel). A short summary of the interaction task instructions was presented on the screen, which participants read in silence.

During the interaction, participants saw the 16 Fribbles on their screen in a random arrangement with corresponding numbers or letters (see Figure 6.4). Both participants saw the same 16 Fribbles in the same general spatial layout, but 50% of the Fribbles were not positioned in the same locations within this layout (see Figure 6.4). On each trial, one of the 16 Fribbles was marked by a red square (the target for that trial) for one of the two participants (the "Director").

The participants completed two different Fribble-related tasks in each trial: the Referential task and the Localisation task. In the Referential task, participants were instructed to communicate with one another so that the Matcher would understand which Fribble was the target on any given trial (i.e., the one marked by the red rectangle in the Director's display). They were informed that they could communicate in any way they wanted (without explicitly mentioning speech and gesture). Once the participant without the red square (the "Matcher") was certain what the target Fribble was, the Matcher said the number or letter of that Fribble out loud and clicked on the yellow button of the button box to move to the Localisation task.

In the Localisation task, participants were instructed to communicate the location of the target Fribble on their screens. They then had to decide whether it was located in the same position on the screen for both participants or not. After reaching agreement, the Matcher pressed the yellow button to indicate "same position" or the red button

for "different position". After completing the Referential and Localisation tasks for one Fribble, participants switched roles for the next trial. Participants completed six rounds, each with different spatial layouts for all 16 Fribbles, resulting in 96 trials. The trial order and spatial layout was the same for all pairs. The total interaction phase took about one hour on average (M = 52.24 min, SD = 10.75 min, *range* = 35.20 - 77.48 min).

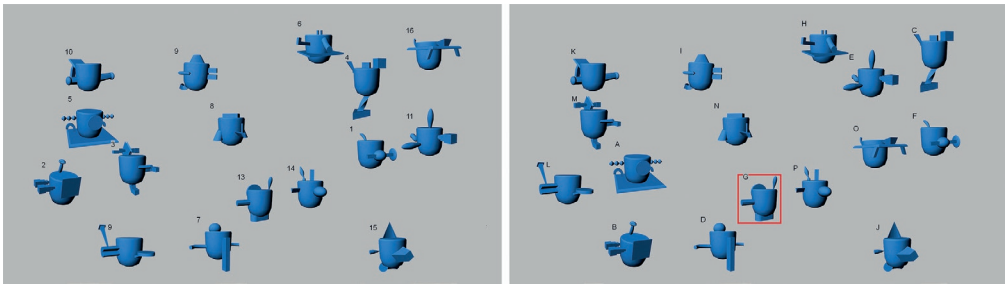


Figure 6.4. Example of a trial in the interactional task as shown to the participants (left: the Matcher in this trial and right: the Director in this trial). The Fribble with the red rectangle was the target Fribble for this trial.

6.2.5.5 fMRI movies

A third fMRI session served to enable later hyperalignment of all participants' brains based on functionally similar responses to complex stimuli (Haxby et al., 2011; see *Introduction*). Participants viewed eight animated movies (see Table D2 in Appendix D for details), presented on a part of the screen slightly below the centre, on a black background with a height of 360 pixels and a width of 640 pixels, and a visual angle of about 9-11° vertically and 16-20° horizontally (see Table D2 for details). The movies were in .avi format and were played at 30 frames per second. Each movie was played in its entirety, except for the start and end of the movie (i.e., titles and credits). These were cut off, so that no text was shown to participants. The duration of the movies was 4.1 min per movie on average (*range* = 2.2 - 6.1 min) and 35 min in total, including breaks. The movies were selected to contain categories of objects that were mentioned often in the pilot Naming tests of the Fribbles (see Section 6.2.3), such as humans, plants, tools, toys, and food. The movies were preceded by a filler video clip that lasted a few seconds. There was a 12 s break between movies. Participants were instructed to simply attend to the movies and to lie still. The movies did not contain (spoken) language, but participants were still provided with the sound of the movies via earphones within the ear, in order to make it easier to stay focused. Before scanning started, the sound of the movies was adjusted to the participants' individual preferences. Note that seven out of the eight movies could not be shared as part of the Dataset, due to copyright issues. The last movie is open source and can be found on the internet (see Table D2 in Appendix D).

6.2.5.6 Questionnaire

The questionnaire consisted of 30 questions in total and was administered via a computer in the cubicles either after the movie fMRI session or before. In the latter case, the last two questions (about the movies) were administered on paper. The questionnaire consisted of 15 questions relating to the different aspects of the study, (e.g., the goal, strategies in the different tasks, difficulty, level of attention, etc.); nine questions about the other participant (e.g., about their personality, voice, whether they were a real participant, etc.): one open question and eight to be indicated on a 7-point Likert scale (1="not at all", 7="very much"); and six questions about the participant: four demographic (age, sex, occupation, and studies) and two judgement questions, to be indicated on a 7-point Likert scale (intro-/extraversion and whether they were proud of their own accent). For the English translation of the questions asked to participants, see Table D3 in Appendix D or the Data folder in the Dataset.

6.2.6 Preprocessing

6.2.6.1 Transforming fMRI data into BIDS structure

All raw MRI data were converted to BIDS (Brain Imaging Data Structure; Gorgolewski et al., 2016) using BIDScoin (version 1.5; Zwiers et al., 2021), including conversion to NiFTi format, and supplemented with standardised metadata. All anatomical MRI scans were defaced to remove identifiable features using a wrapper tool around pydeface (DOI: https://zenodo.org/badge/latestdoi/47563497).

6.2.6.2 Processing and synchronising audio, video, and Kinect data

The cameras saved multiple consecutive .mp4 files for each interaction (of about 14 min / 3.45 GB each), which were first concatenated and saved as a single .mp4 file for each camera per interaction. This was done for all recordings. For the majority of pairs (see Section 6.2.7 and Table 6.1), we also manually synchronised the three videos with Adobe Premiere Pro CC (version 2018) with the help of the auditory synchronisation signals (see Section 6.2.3.2).[9] The videos were trimmed and the audio (from the head-mounted microphones as recorded on the cameras) from participants A and B were set to the left (−100) and right (100) audio channel respectively. We then exported six media files which were used for the transcriptions: three video files (as .mp4 files with H.264 codec) and three audio files (as .wav files). The video files were exported with the recorded audio from both participants as stereo channel (where one participant is audible on the left, and the other on the right channel). The audio files were exported at 16-bit

9 The off-line synchronisation of the audio/video data from the different cameras did not allow for time-alignment with millisecond precision, since the sampling rate was 29.97 fps (i.e., one frame every 34 ms). We checked the lag between the two audio channels from cameras 1 and 2 after synchronisation, and found this to be 11 ms on average (*range*: 0-33 ms).

sample size; the audio of the individual participants was exported as mono channel in two separate files (one for each participant; in which the other may still be slightly audible), and the combined audio of both participants with stereo channel (similarly to the video exports). All video and audio files were exported with a sample rate of 44,100 Hz. Finally, to enable synchronisation of the video and Kinect data, one additional audio file was exported which included the auditory synchronisation signal. To this end we used custom-made Matlab scripts, where the principle for synchronisation was the same as described above for the videos; the script adjusted the timestamps of the Kinect data to match the videos by time-aligning the digital and auditory synchronisation signals (which were transmitted every 60 s; see Dataset: Preprocessing for the script).

6.2.6.3 Orthographic transcription procedure for the interaction data

Orthographic transcription of the speech in the interaction phase was done in ELAN (Wittenburg et al., 2006), see Figure 6.5. The ELAN files included all synchronised media files for each pair (see Section 6.2.5.2): three videos (from cameras 1, 2 and 3) and three audio files (from the head-mounted microphones as recorded on the cameras; one file for each participant and another file containing both channels). This allowed the transcribers to inspect the audio waveforms and to listen to each participant separately or both participants simultaneously (which is particularly useful in case of overlapping speech). Three tiers were used for the transcription: two for the transcribed speech, and one on which the transcriber added comments.

Speech was first segmented into Turn-Constructional Units (TCUs; i.e., potentially complete, meaningful utterances, Clayman, 2013; Couper-Kuhlen & Selting, 2017; Schegloff, 2007). If TCUs exceeded 10 s, they were divided into multiple segments of under 10 s length. This was done to allow for optimal automatic forced alignment of the speech into phones for future phonetic analysis.

Speech was orthographically transcribed based on the standard spelling conventions of Dutch. All words, discourse particles (e.g., "oh", "ah", etc.) and filled pauses (transcribed as "uh" or "um") were transcribed. Unfinished words were also transcribed but marked. When the transcriber was not certain about their transcription, the respective element was placed between parentheses. When the transcriber could not determine what was said at all, this part was transcribed as a question mark enclosed by parentheses. Non-lexical vocalisations and other sounds were transcribed between asterisks (e.g., *laugh*, *cough*, *lip smack*, etc.). In addition to being transcribed between asterisks, long stretches of laughter during speech were also commented on in the comments-tier.

6.2.6.4 Linking transcribed speech to task structure

The task structure of the interaction is indicated on the "trial" tier in ELAN ("1.1_ref" etc.; indicating round number (1 – 6), trial number (1– 16) and task ("ref" or "loc"), see Figure 6.5). The onsets and offsets of these annotations were manually adjusted; by default, a task ended when the Matcher pressed a button to move to the next task. However, sometimes there was a mismatch between the moment at which participants pressed the button and their speech productions relating to either of the two tasks (e.g., participants would start talking about the location of the Fribble before pressing the button to end the Referential task). In these cases, the onset/offset of the task was placed earlier or later such that the speech about the respective tasks would fall under the right trial annotation.

Once these annotations were finalised in ELAN, we derived the answers for the Referential task (i.e., the letters and numbers that participants said out loud) from the transcripts (note that participants indicated answers for the Localisation task with the button box). We then finalised the ELAN files by adding the following tiers about the task structure and performance: target (Fribble number), director, correct_answer, given_answer, and accuracy.

These ELAN files were then exported to text and Praat TextGrid files for further analyses. The text files were transformed into two datafiles for each pair: one containing all speech annotations (which are linked to trials based on the annotation onsets) and one containing all trial information (i.e., trial onset and offsets, and information about task, target, director, answers and accuracy). The annotations in the Praat TextGrid files were readjusted to match the original audio files recorded with the recorder (by moving the boundaries of all annotations using a script that can be found in the Preprocessing folder in the Dataset). This was necessary since the original audio files also included the phonological pre-test, and because the audio files from the camera and the recorder were not exactly aligned due to internal clock drift.

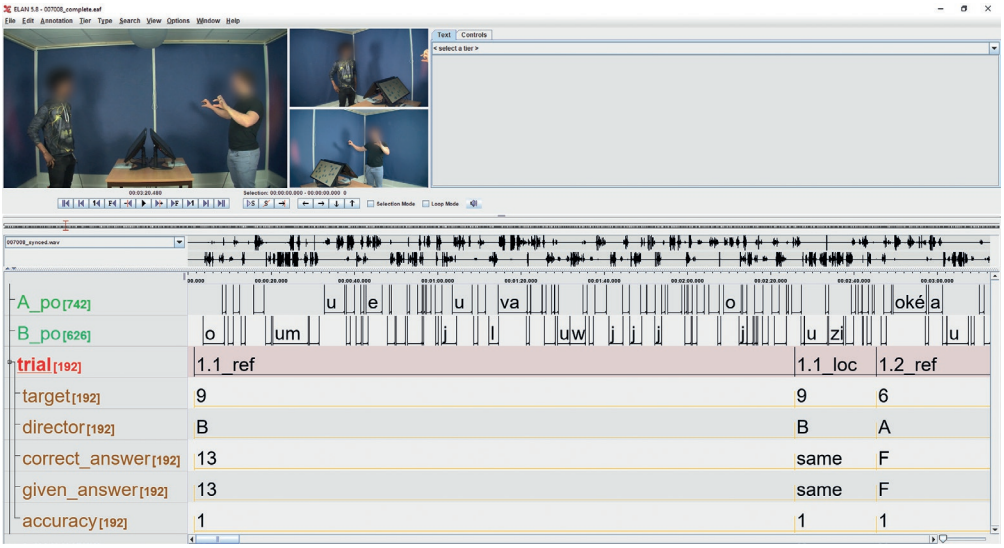


Figure 6.5. Screenshot from ELAN file. It includes the synchronised videos from the three cameras (top left; faces are blurred in this figure but not in the videos in the shared dataset), the synchronised audio from both head-mounted microphones (waveforms in the middle), as well as annotations and transcriptions on various "tiers" (rows at the bottom of the window). The first two tiers include the transcribed speech for the participant on the left ("A_po") and right ("B_po"; where "po" stands for practical orthography). The remaining tiers provide information about the task structure and accuracy (see Section 6.2.6.4).

6.2.7 Dataset

6.2.7.1 Data availability and quality

Incomplete or non-usable data is not shared (see Table 6.1 for an overview of the number of participants and pairs for which data is shared). First, two participants from different pairs exceeded our age criteria and misunderstood instructions, leaving 140 individual participants, and 69 complete pairs.

MRI data from 16 participants were excluded due to MRI scanner/software malfunction (n = 6); missing data due to participant claustrophobia (n = 4); excessive motion within or between sessions (n = 3); bad performance on one-back task (n = 2); and experimenter error (n = 1); see the Participants folder in the Dataset for more details. Note that exclusion of MRI data based on motion or one-back performance was only done for extreme cases. Researchers may still want to deal with motion artifacts and/or errors in the one-back task while preprocessing and analysing the shared data.

There are 124 individual participants and 56 complete pairs with complete and shared (f)MRI data. Furthermore, (f)MRI data of three extra participants is still shared because only session three is incomplete and this may be irrelevant for some users of the data.

Audio/video data from the interactive sessions are shared in the Dataset, except for seven cases where one or both participants did not provide consent. We focused our transcription resources on a selection of 47 interactions, based on their (audio) quality and completeness of the MRI data (see above). However, audio/video data is shared for 42 of these pairs, four pairs contain individuals with non-usable MRI data, and questionnaire data are missing for one pair (these issues were encountered only after transcription had finished). Thus, all data (i.e., fMRI, behavioural, and transcription/interaction data) is usable for 42 pairs (84 participants) in total. For 37 of these 42 pairs, thanks to participant consent, the video and audio data can be shared as well.

Table 6.1. Numbers of participants and pairs for which different types of data are available and/or shared in the Dataset.

Data	Participants (*n*)	Pairs (*n*)
Total collected	142	71
Usable behavioural data (Naming & Features)	140	69
Usable (f)MRI data all sessions	124	56
Usable (f)MRI data sessions one and two	127	59
Usable interactional data *of which audio & video shared (fully preprocessed*)*	138 *126 (98)*	69 *63 (49)*
Transcribed interactional data *of which audio & video shared (fully preprocessed*)*	94 *84 (84)*	47 *42 (42)*
All usable data (interaction/transcription, MRI, behavioural) *of which audio & video shared*	84 *74*	42 *37*

*fully preprocessed = video is synchronised and processed as described in Section 6.2.6.2

6.2.7.2 Data accessibility

See Data and code availability statement for instructions on how to access the data.

6.2.7.3 Data structure and format

The Dataset contains the stimuli and the raw and minimally processed data, as well as data and scripts needed to reproduce the results reported in Section 6.3. It also contains scripts used for running the tasks and scripts or files used for preprocessing the data. Table D4 shows an overview of the different data types and formats in the Dataset.

6.3 Results

In this section, we present analyses implemented on different parts of the Dataset, chosen to offer an intuition of its characteristics and potential for analysis.

First, we present a short overview of participants' answers to the questionnaire (Section 6.3.1). Following, we present several measures obtained within the interaction, in which participants communicated to match and localise novel objects (Fribbles). To give a sense of the interactional data, we present data for time on task and accuracy (Section 6.3.2), number and type of words used (Section 6.3.3), and gesture characteristics based on an automated analysis pipeline for motion tracking (Kinect, Section 6.3.4). In these three sections, we report the results for the participant pairs for which the audio/video data has been processed, trial annotations adjusted, and the interactions transcribed (n = 47).

We also probe correlates of the lexical and conceptual representations of each Fribble in each participant before and after the interaction, using two behavioural measures (Naming and Features) and a brain measure (fMRI). Below, we show how pair members converged on a representation after the interaction (Section 6.3.5) using one of the behavioural measures (Naming). For the fMRI measurements, we show an indication of the data quality based on the fMRI data before the interaction only (Section 6.3.6).

Note that all (pre)processing of the data specific for the analyses reported here is also included in this Results section, so that the Methods section could be reserved for a description of the dataset itself. Scripts used for preprocessing, analysis, and figure generation are shared in the Results folder in the Dataset.

6.3.1 Questionnaire

This section describes participants' experiences and task strategies, as reported in the questionnaire (see Table D3 in Appendix D for all questions asked).

Most participants did not correctly guess the goal of the study. Some participants thought that the study was about object names/perceptions changing (or occasionally: becoming more similar) as a consequence of the interaction.

Main strategies reported for the Naming task were: describing a unique part of the Fribble, describing the objects holistically as existing concepts, or a mix of both strategies. Participants generally did not report using the list of features in their names, only "compact" and "human" were mentioned occasionally. Most participants reported using descriptions from the interaction for all or some of the objects in the post-interaction Naming session.

For the Features task, no clear strategies were mentioned. Participants did report to take into account their name from the Naming task when evaluating the features, especially in the post-interaction session.

During the one-back task in the MRI scanner, most participants used the names they used in the Naming task or in the interaction to remember the relevant Fribble

during the inter-stimulus interval (e.g., through phonological rehearsal), even more so in the post-session.

Strategies for the Referential task in the interaction entailed describing the Fribbles' prominent visual features at first and associating them with known concepts in later rounds. Other mentioned strategies were repeating successful names from earlier rounds and using one's own names from the Naming task. Participants reported using either their own or their interlocutor's initial label or coming up with labels collaboratively. Some participants reported to have retained names that were successful in earlier rounds.

Main strategies for the Localisation task included dividing the screen into four quadrants; dividing the screen into rows and columns, proportions, or percentages; finding patterns or clusters of Fribbles (e.g., triangles, squares); or describing Fribble positions relative to features of the monitor (e.g., centre, logo).

6.3.2 Interaction: Time on task and accuracy

On average, pairs spent almost one hour on the interaction tasks (M = 52.24 min, SD = 10.75 min, *range* = 35.20 - 77.48 min). Pairs spent more time on the Localisation task (M = 28.90 min, SD = 9.00 min, *range* = 15.21 - 54.20 min) than on the Referential task (M = 23.34, SD = 3.44, *range* = 16.54 - 31.85). Figure 6.6, panel A shows that the time spent per trial decreased as the interaction progressed, and that this pattern was more pronounced for the Referential task (in blue) compared to the Localisation task (in orange). As for accuracy, pairs performed well (near ceiling) for both the Referential task (M = 99.24% of trials correct, SD = 1.06%, *range* = 94.79 - 100%) and the Localisation task (M = 95.28% of trials correct, SD = 4.60%, *range* = 79.79 - 100%), see Figure 6.6, panel B.

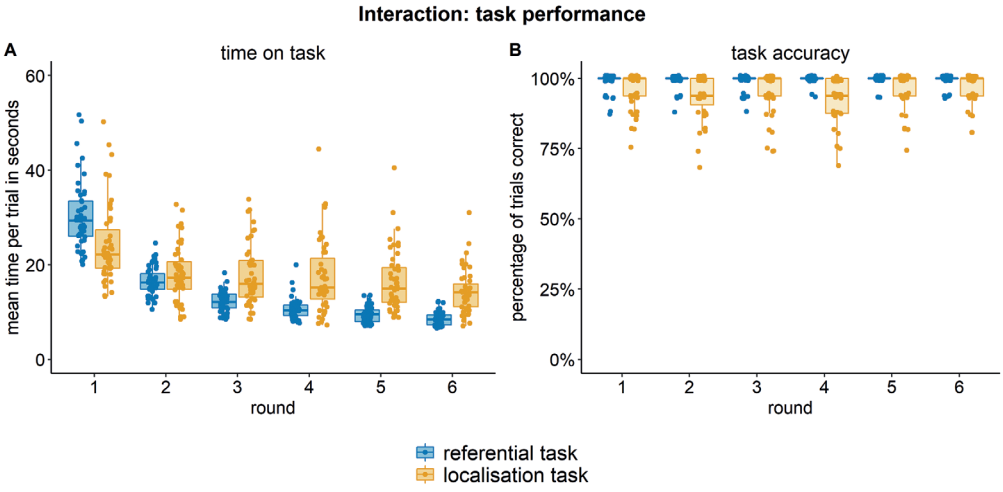


Figure 6.6. Distribution of time spent per trial (panel A) and task accuracy (panel B) in the six rounds of the interaction, for the Referential and Localisation task separately. Dots represent pairs (n = 47).

6.3.3 Interaction: Number of words and word types

Word counts from the interactional tasks (47 pairs) were analysed to give an insight into the content of the corpus. The transcriptions were first cleaned: we removed unfinished words, non-speech noises, punctuation (indicated here between <>: <#>, <()>, <'>, <,>, <.>, <`>, <?>, <->), and converted back- and forward slashes (<\> and </>) into spaces. All words that were not completely clear to the transcribers are included in the analyses.

The overall average word count per pair was 8,552 (*SD* = 2,125, *range* = 5,007 – 15,233, Figure 6.7; see Figure 6.8 for word types). We distinguished between function words, content words, and interactional markers. The function word list was created from the Dutch Molex lexicon (Gigant-Molex, 2019) and can be found in the script in the Results folder in the Dataset. Interactional markers (also known as procedural conventions; Mills, 2011; Knutsen et al., 2019) are linguistic resources used to manage the interaction. To create the interactional marker list, we took all words in the corpus that did not appear in the CELEX lexical database (Baayen, Piepenbrock & Gulikers, 1996). From this list, we manually removed task responses, content words not present in CELEX (such as names), English words, typos, and spelling variations of words present in CELEX. All words that did not fit the function word list nor the interactional marker list were automatically marked as content words (see Dataset: Results).

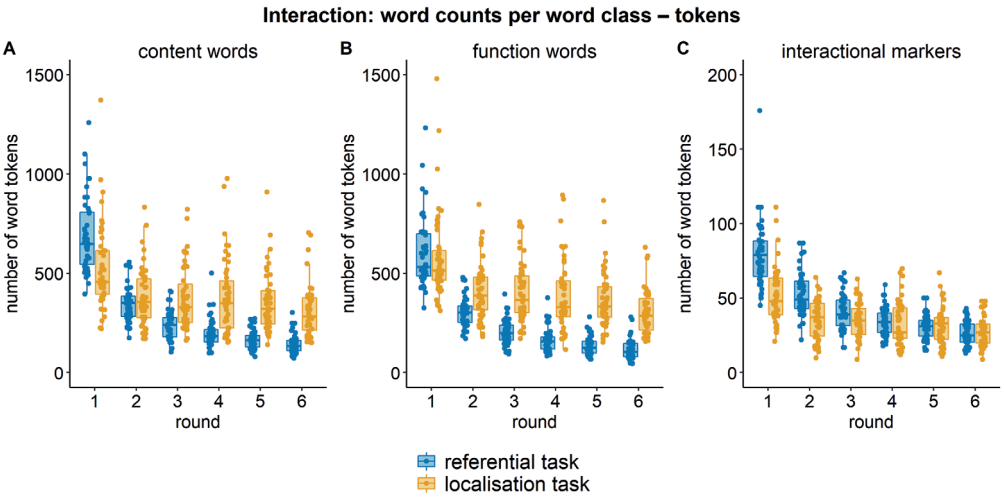


Figure 6.7. Word token counts visualised per round, per task, and per pair. Plots for content words (Panel A), function words (Panel B), and interactional markers (Panel C). Note that the scales of the y-axes differ. Dots represent pairs (*n* = 47).

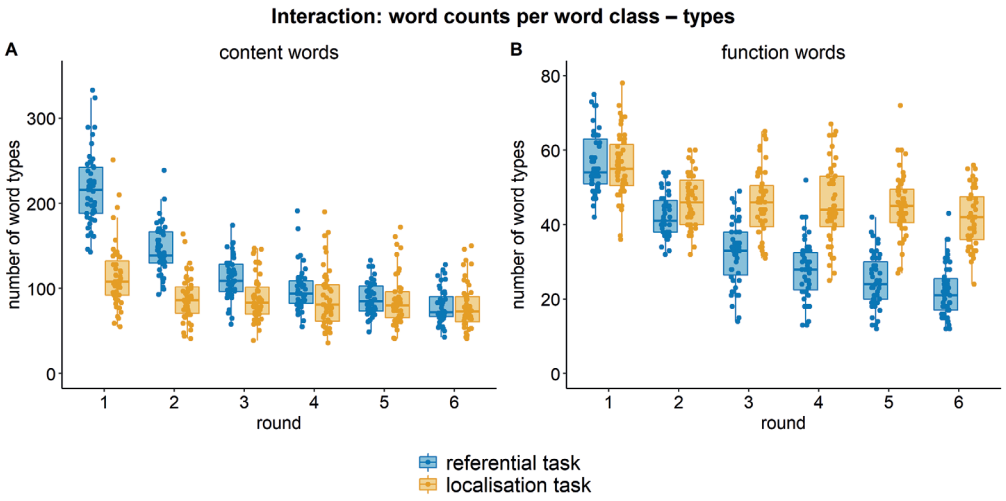


Figure 6.8. Counts for different word types per pair visualised per round and per task. Plots for content word types (Panel A) and function word types (Panel B). Note that the scales of the y-axes differ. Dots represent pairs (*n* = 47).

Visual inspection of the figures shows that both word token counts and word type counts decrease over the rounds for the Referential task (in blue). For the Localisation task (in orange), these counts are more stable and seem to mostly decrease between rounds 1 and 2, but remain rather stable over the rest of the rounds. This is consistent with the observation from Section 6.3.1 above that the time used overall in the Localisation task is more stable over rounds as well.

6.3.4 Motion tracking data and automatic coding of gestures

To inspect the characteristics of manual gestures produced during the interaction, the motion tracking (Kinect) data was analysed for 94 participants (i.e., 47 pairs with complete trial annotations). Kinect sampling rate was regularised at 30Hz (linear interpolation), timeseries of the right hand were high-pass filtered (2nd order Kolmogorov–Zurbenko filter with a span of 3), followed by filtering of the timeseries derivative. The right hand was chosen for this report to illustrate communicative movements of the dominant hand (all participants were right-handed). The hand tip (middle finger) was chosen as this was a point that captures movement of both the upper and lower arm, wrist, and finger. The resulting time series data can be inspected alongside the videos in ELAN (see Dataset: Results for an example).

6.3.4.1 Kinematic measures

Our key descriptive measures for communicative manual movements are (autocoded) gesture counts, submovements, gesture duration, and average vertical height.

6.3.4.1.1 Autocoder of communicative gestures. We employed a rule-based automatic movement coder (Pouw, De Wit, et al., 2021) to approximate communicative gestures that were made during the interactions. The autocoder takes as input the speed and position of the hands to approximate gesture events (see Section D1 in Appendix D for details). Previously we have applied our autocoder on a dataset with a similar design, where we tested the performance of the autocoder relative to human annotations of iconic gestures (Pouw, De Wit, et al., 2021). In that study we found that (p. 11) the human coded iconic gestures were positively related to the auto-coded gestures $r = 0.60$, $p < 0.001$, with a 65.2% accuracy in time overlap between these codings (*true positive = 70%, false positive = 86%, true negative = 93%, false negative = 1%).*

6.3.4.1.2 Submovements. Kinematic submovements are computed on the right hand tip speed by counting the number of positive local peaks that exceed 15 cm/s during an autocoded gesture event. Following earlier work (Trujillo et al., 2018, 2019), we assume that gestures designed to communicate tend to have more submovements. Measures akin to submovements have been found to strongly correlate with the number of information units human annotators perceive in the gesture (Pouw, Dingemanse, et al., 2021). Thus, the number of submovements is a kinematic measure that approximates the number of semantic units of the gesture.

6.3.4.1.3 Gesture duration. Duration in seconds of the autocoded gesture event.

6.3.4.1.4 Average vertical height. Average vertical position within each gesture event. Following earlier work (Trujillo et al., 2018, 2019), we assume that the degree to which a gesture is forefronted in a more prominent gesture space is an informative kinematic quality about the degree of saliency of the gesture.

6.3.4.2 Descriptive results kinematics

Figure 6.9 shows the main results of the kinematic measures as they develop over the rounds. It can be seen that autocoded gesture count drastically decreases over the rounds, and the number of submovements of these gestures also decrease over time. Gesture duration and vertical height also follow this pattern but in a less pronounced way. These quantitative patterns in the kinematic data likely relate to the common ground that is built over the rounds (e.g., Holler & Bavelas, 2017; Holler et al., 2022; but see Hoetjes et al., 2015 for slightly different results in a referential task with stimuli similar to the

Fribbles), and to kinematic optimisation of gestural signalling (e.g., Pouw, Dingemanse, et al., 2021).

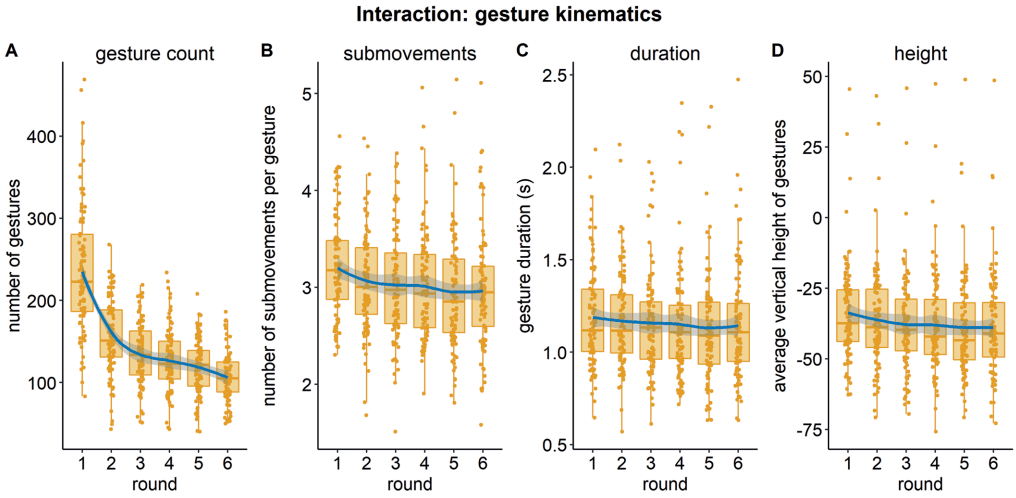


Figure 6.9. Descriptives of the motion tracking data; each jitter point represents kinematic descriptives (of autocoded right-handed gestures) for one round for one participant (*n* = 94 participants). The trend line reflects the smoothed means using a Local Polynomial Regression (loess) Fit.

6.3.5 Lexical similarity in the Naming task

Since the interactional tasks were expected to lead to conceptual alignment between participants of a pair, here we report their degree of lexical alignment, as one of the proxies for conceptual alignment. In order to do so, we calculated the similarity of the names for the same Fribble between two pair members (that formed an actual pair in the interaction), both before and after the interaction, for the 69 pairs that had usable behavioural data (see Table 6.1, second row). These are referred to as "real pairs" from here on. To ensure that such lexical alignment was specific to the individual interactions, we also calculated the increase in lexical alignment from before to after the interaction in all possible "pseudo-pairs": pairs of participants who engaged in the tasks in different roles but who did not interact with each other (n = 4,692). The increase of the real pairs was compared to a permutation distribution (n = 10,000) of lexical alignment difference scores from all possible (real and pseudo-)pairs. Pseudo-pairs provide a rigorous control for systematic but communicatively un-specific effects of task performance and task structure.

The text written by the participants in the Naming task (one to three words per Fribble) was regularised by removing special characters (indicated here between <>: <'>, <">, <()>, <&>, <+>, <.>, <;>) except if they were part of a word, converting

the character </> into a space, <=> into the word <is>, correcting spelling errors for a small number of frequently occurring words/names (i.e., *Pippi Langkous* ("Pippi Longstocking"), *Pinokkio* ("Pinocchio"), *plateau* ("plateau"), and *trofee* ("trophy")), converting uppercase characters to lowercase, and converting numbers into the corresponding words, except if the number was part of a word (e.g., 3d). The words were then checked against the NLPL Dutch CoNLL17 corpus (Zeman et al., 2017). Only words missing from the corpus were changed by correcting their spelling or dividing compounds (or words split with a <-> character) into two (or more) words. Note that this procedure may have led to an underestimation of name similarity if two differently spelled versions of the same word were both present in the corpus. For two unidentified words in the corpus, it was not clear how they should be corrected, so these were left as such.

Naming similarity between names for the same Fribble was operationalised here as lexical similarity, that is, how many words were (exactly) the same between the two names, normalised by the number of words. To compute this, the cosine similarity of the names was taken, resulting in a score between 0 (no words are the same) and 1 (all words are the same; see also Duran et al., 2019; Rasenberg et al., 2022). As an example, the names "trophy triangle plateau" and "trophy with blocks" led to a similarity of 0.33 because one out of three words was the same.

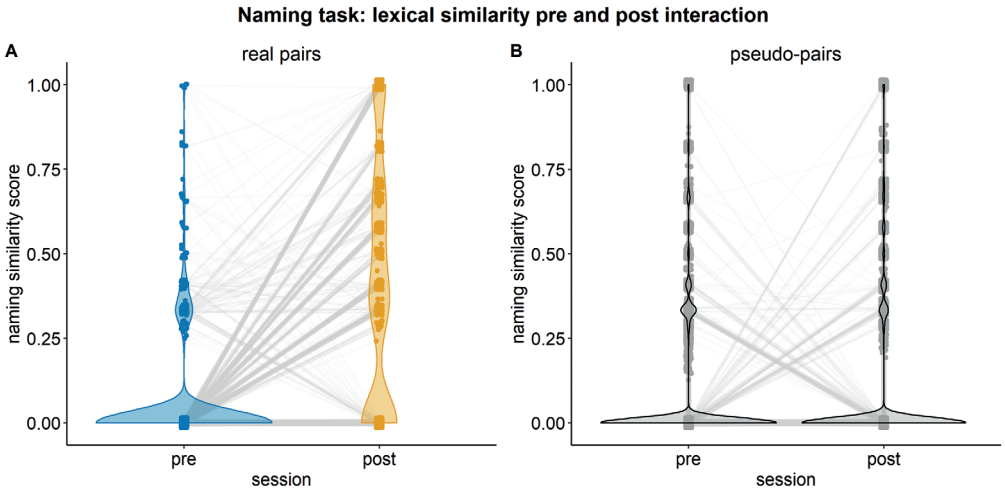


Figure 6.10. Distribution of Naming similarity scores (i.e., cosine similarity between the names provided by two participants of a pair for a particular Fribble), before (pre) and after (post) the interaction. Panel A shows results from real pairs, and panel B from pseudo-pairs. Dots represent individual data points (n = 1,004 (69 pairs by 16 Fribbles) for real pairs and n = 75,072 (4,692 pairs by 16 Fribbles) for pseudo-pairs).

Real pairs numerically showed an increase in lexical similarity from before the interaction (M = 0.06, *Median* = 0) to after the interaction (M = 0.38, *Median* = 0.33), see Figure 6.10, left panel. The average difference score (M = 0.32) was tested against a permutation distribution of 10,000 average difference scores each calculated from 69 pairs that were randomly drawn from a pool of all possible pseudo-pairs (n = 4,692; see Figure 6.10, right panel) and all real pairs (n = 69). The average difference score for real pairs clearly lies above this distribution (p = 0), showing that the increase in lexical alignment for real pairs cannot be due to mere experience with the task.

6.3.6 Relation of fMRI data to visual similarities of the Fribbles

As a general check of the fMRI data quality, we performed a correlation analysis between pairwise visual similarities of the Fribble-images and pairwise brain-pattern similarities related to viewing the Fribbles before the interaction. We entered 112 participants in this analysis, which constitute the 56 pairs for which both participants had usable fMRI data in all sessions (see Table 6.1, third row).

Each fMRI data run was spatially aligned, coregistered to the corresponding anatomical T1 scan, and spatially normalised (MNI space). Then, a general linear model was fitted to the data, considering a regressor for each of the Fribbles per session (as well as 9 nuisance regressors capturing signal variance related to head motion, and signals from the cerebro-spinal fluid, white matter, and out-of-brain compartments). The resulting stimulus specific beta weights were then used to create 16-by-16 dissimilarity matrices containing all pairwise dissimilarities between brain patterns of the 16 Fribbles for searchlights (radius = 9 mm/4.5 voxels, within a grey matter mask) through the brain. The neural dissimilarity matrices were correlated with a 16-by-16 Fribble dissimilarity matrix calculated as one minus the Structural Similarity Index (Wang et al., 2004), a metric of visual similarity between Fribbles. The resulting Fisher-Z transformed correlation values per participant and searchlight were subjected to a second level permutation analysis (n = 1,000 iterations) over participants with TFCE correction for multiple comparisons (Smith and Nichols, 2009). The resulting Z-values are shown in Figure 6.11 with a significance threshold of 1.96 (corresponding to two-sided $p <$ 0.05). As expected, the areas with significant correlations are mainly located around the visual cortex.

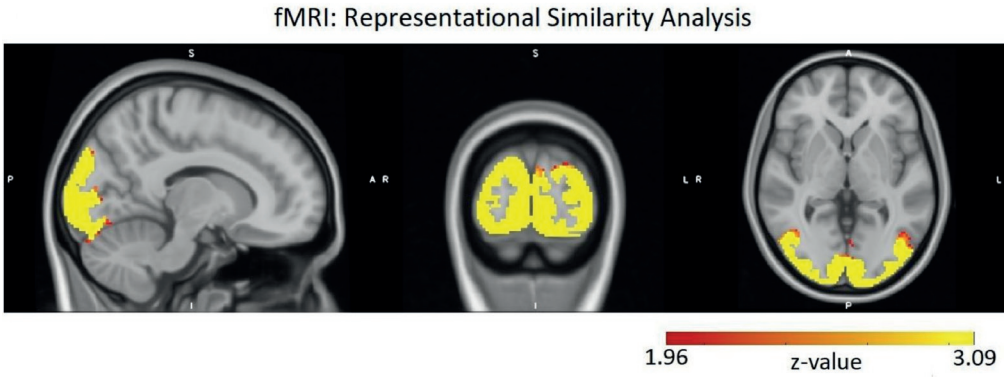


Figure 6.11. Visualisation of significant correlations between visual similarities and similarities of brain patterns between the 16 Fribbles throughout the brain, shown in a sagittal (left), coronal (middle), and axial (right) slice. L = left, R = right, A = anterior, P = posterior. Colours indicate TFCE-corrected Z-values from the second level analysis across participants, above the two-sided significance threshold of 1.96.

6.4 Discussion

This paper describes a large dataset consisting of (transcribed) speech, audio, video, and motion-tracking data during face-to-face task-based interaction about novel objects (Fribbles), as well as pre- and post-interactional behavioural and fMRI measures, estimating representations of the Fribbles from 71 pairs of participants.

We discuss aspects of this dataset on the basis of the reported results to demonstrate its quality and to provide suggestions for potential uses of the data. We deliberately refrain from embedding the dataset in strong theoretical assumptions to avoid biasing potential uses and to allow for different hypotheses to be tested by a wide range of researchers.

On average, each pair spent about an hour performing the two interactional tasks (i.e., the Referential and Localisation tasks), spending slightly more time on the Localisation task. Participants performed near ceiling on both tasks. The time pairs spent per trial, as well as the number of function and content word types and tokens they used, descriptively decreased over the six rounds for the Referential task, whereas these measures appeared more stable for the Localisation task. An exception to this pattern is that participants appeared to use a similar number of interactional markers within the two tasks in later rounds. Over the rounds, decrease in time on task and word counts in the Referential task and, to a reduced extent, in the Localisation task, is likely to be related to the building of common ground. This pattern is in line with earlier work using repeated reference games, which are known to elicit increasingly shorter referential expressions from participants over rounds (e.g., Hawkins et al., 2020; Kraus &

Weinheimer, 1964; Clark & Wilkes-Gibbs, 1986). Given that the Fribbles themselves remained the same in each round, whereas their locations changed, it is to be expected that more common ground can be built up in the Referential than in the Localisation task.

Furthermore, the amount, duration, and average vertical height of gestures, as well as the amount of gesture sub-movements decreased over interactional rounds. This general decrease in gesture count, size, and complexity over the interaction was expected on the basis of previous research showing that such modulations follow from the building of common ground (e.g., Holler & Bavelas, 2017; Holler et al., 2022) and the kinematic optimisation of gestural signalling (e.g., Pouw, Dingemanse, et al., 2021). Taken together with the previously described results regarding speech, it appears that, descriptively, attenuations in speech and gesture over time go hand in hand, which is also in line with earlier work (Holler & Bavelas, 2017).

In the Naming task, participants named all Fribbles using one to three words. Pairs generally showed a larger lexical similarity between their names or descriptions of the Fribbles after the interaction than before, an increase that proved highly reliable when compared to permutations including pseudo-pairs (formed post-hoc by pairing up participants who did not interact with each other). This result confirms that the interaction led to convergence of naming conventions for the Fribbles.

Regarding the fMRI data, correlations between the similarity of brain-activation patterns in response to the Fribbles and objective visual similarity of the Fribbles were highest around the visual cortex. This expected result shows the quality of the (f)MRI data.

These results show that the interactional data (linguistic and kinematic), the computer-based behavioural measures, as well as the brain imaging data are of high quality. This means that this dataset can be used for a large range of analyses on interactions (e.g., phonetic, lexical, syntactic, semantic, pragmatic, and gestural analyses). One key objective of the CABB team for collecting the data was to investigate alignment in the interactions, and therefore the dataset is well suited to quantify the degree to which participants align their linguistic and/or bodily behaviour at different levels (Pickering & Garrod, 2004; see Rasenberg et al., 2020 for an overview of different definitions and measures of alignment). A wide range of analyses is possible, which may be qualitative or quantitative in nature, recruiting manual or (semi-)automatic procedures to analyse the audio-video data (e.g., with OpenPose, Cao et al., 2017) or transcripts (e.g., with the Python package ALIGN, Duran et al., 2019).

For example, one could measure the degree of similarity between participants' realisations of different phonemes over the course of the interaction (e.g., Pardo, 2006). At the prosodic level, one may compare pitch or articulation rate (e.g., Chapter 3). The phonological pre-test (see Section 6.2.5.3) is useful for such analyses, since it

provides a baseline of participants' speech before they start interacting with each other. At the syntactic level one could compare N-grams (e.g., Fusaroli et al., 2017; Reitter & Moore, 2014) or look at specific syntactic constructions (e.g., Hartsuiker et al., 2008). At the lexical level, one could quantify how often participants use the same words (e.g., Bangerter et al. 2020; Brennan & Clark, 1996). At the semantic level, word2vec or similar distributional models (e.g., Mandera et al., 2017) could be used to measure semantic similarity between participants' speech turns (e.g., Dideriksen et al., 2019). In terms of bodily behaviour, one could for example look at how people align their posture (e.g., Shockley et al., 2003) or at the type and form of their gestures (see e.g., Bergman & Kopp, 2012; Chui, 2014; Holler & Wilkin, 2011; Louwerse et al., 2012). The present dataset is especially suited to perform such gestural analyses, given the rich set of (mostly iconic) gestures elicited by the task (see Section 6.3.4.2) and the availability of Kinect measurements for quantitative analyses. It is also possible to test hypotheses about how alignment at different levels and/or modalities is related to each other (e.g., Cappellini et al., 2022; Mahowald et al., 2016; Oben & Brône, 2016; Pickering & Garrod, 2004; Rasenberg et al., 2022).

Furthermore, the task-based interactions – in which the establishment of mutual understanding is a challenge – allow researchers to investigate the interactional mechanisms that people use to solve coordination problems, such as other-initiated repair (Schegloff et al., 1977; Schegloff, 2000). Given the relatively free-form of the interactions (in which people were free to communicate in any way they wanted, without time constraints), the data can be used to analyse various (multimodal) interactional phenomena, such as turn-taking (Sacks et al., 1974) or the use of backchannels or acknowledgements (Allwood et al., 1992; Jefferson, 1984; Yngve, 1970).

In addition, the dataset allows researchers to examine whether the interactions result in changes in the estimated representations of the Fribbles, and whether the representations of pair members tend to converge. Such hypotheses could be tested in several ways using the present dataset, given the availability of both brain data and two types of behavioural data. The results provided here (see Section 6.3.5) show that interacting participants converge in the sense that they more often use the same words to refer to the Fribbles after the interaction than before the interaction. In a similar vein, one could investigate such convergence in terms of semantic similarity of the names, similar scores given to the features, and similar brain activation patterns in fMRI measurements between participants. The latter analysis is further facilitated by the possibility to implement functional hyperalignment of participants (Haxby et al., 2011; see *Introduction*).

Moreover, the unique feature of this dataset is the combination of linguistic, behavioural, and neural data within the same paradigm and for the same stimuli and participants, opening up the possibility for a systematic investigation of the relation

between them. This in turn, may make it possible to find support for or against specific hypotheses regarding the relationship between certain characteristics of the interactions, which may support mutual understanding and convergence between participants in estimated representations. As clearly shown by Figure 6.10, panel A, in Section 6.3.5, pairs display quite some variability with regards to lexical alignment in the Naming task after the interaction. It may be possible to find characteristics of the interaction that can explain such variance to some extent. In conclusion, the present dataset ultimately allows researchers to provide a comprehensive picture of both the behavioural aspects of multimodal interaction and associated changes in representations of the interactional referents, estimated using behavioural as well as neural measures.

Data and code availability statement

The Dataset is stored as a Research Documentation Collection in the Donders Repository (https://data.donders.ru.nl/). Note that the Dataset is not publicly available, since participants specifically consented to their sensitive (audio and video) data being used by researchers for scientific purposes only. To ensure this and to warrant secure data storage and sharing of these sensitive data, a request for access must be submitted to the Dataset managers by signing a Data Use Agreement (provided as a separate pdf file in the published version of this chapter), which specifies the conditions and restrictions under which the data is shared. Specifically, conditions are specified regarding the secure data storage of the data (see the Appendix of the Data Use Agreement for details) and the restriction that the data is used for scientific purposes only. Furthermore, it is specified that users should acknowledge the origin of the data as follows: "Data were provided (in part) by the Radboud University, Nijmegen, The Netherlands" and that they should cite the published version of this chapter in papers or other presentations using the data. Importantly, it is specified that "neither the Radboud University, nor the researchers that provide this data should be included as an author of publications or presentations if this authorship would be based solely on the use of this data."

In short, to be able to access and download the data, two steps are required. First, you need to create a user profile in the Donders Repository by logging in with your SURFconext or ORCID account (https://data.donders.ru.nl/login). For more information about the ORCID option and alternative ways to login, see: https://data.donders.ru.nl/doc/help/helppages/user-manual/login-profile.html?8. Second, the Data Use Agreement needs to be completed and sent to ivan.toni@donders.ru.nl. Upon completion of these steps, users will be granted access to the collection and can view and download files through the Donders Repository website (for more information, see: https://data.donders.ru.nl/doc/help/user-manual/transfer-data.html).

Appendix D

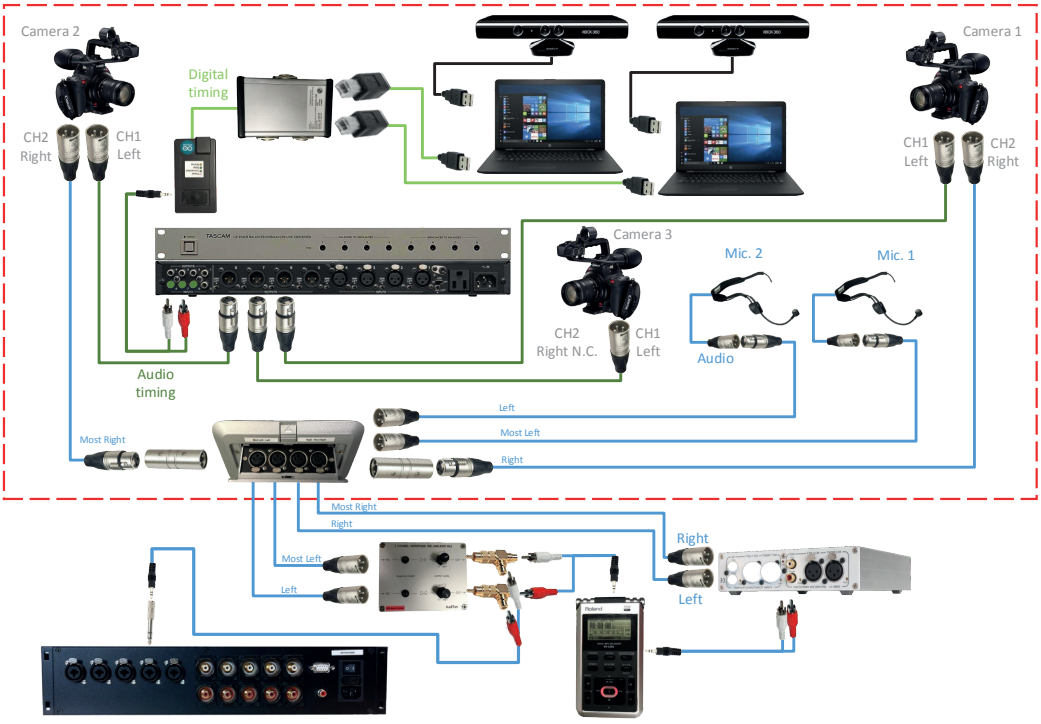


Figure D1. Overview of the equipment used for recording the interaction. Equipment in the area marked in red was situated inside the sound-attenuated booth, the rest of the equipment was situated in the control room. Note that this figure is for illustrative purposes only (it visualises the set-up and the connections between the devices; for accurate brand and product names, see Section 6.2.4.2).

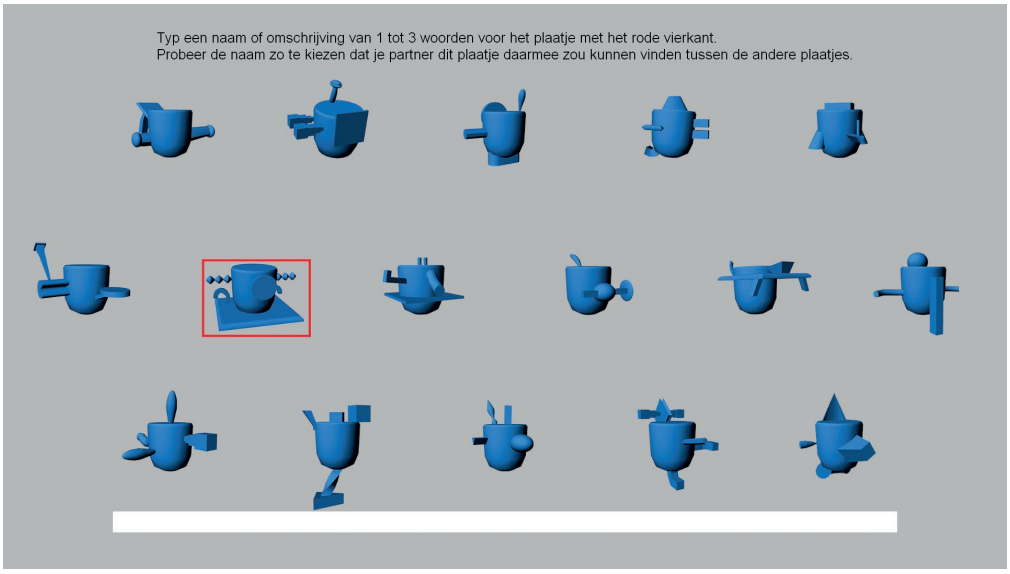


Figure D2. Screenshot of one trial in the Naming task. Instructions are given at the top (in Dutch; English gloss: "Please type in a name or description of 1 to 3 words for the picture with the red square. Try to choose the name in such a way that your partner could use it to find the picture between the other pictures.") and the red square indicates the Fribble that should be named. Participants type in their name/description on the keyboard and it becomes visible in the white bar at the bottom of the screen.

Table D1. The 29 features that participants saw on the screen in the Features task in original Dutch with English translation. On top of the screen a lead-in sentence was displayed (In hoeverre zie je dit plaatje als; "To what extent do you view this picture as...").

Nr	Feature (original Dutch)	Feature (English translation)
1	afgerond	rounded
2	puntig	pointy
3	symmetrisch	symmetrical
4	langwerpig	elongated
5	plat	flat
6	compact	compact
7	visueel complex	visually complex
8	aandacht vragend	demanding attention
9	iets met een kenmerkende kleur	something with a characteristic colour
10	licht of helder om te zien	light or bright on the eyes
11	groot	large
12	klein	small
13	iets met een kenmerkende smaak	something with a characteristic taste
14	iets met een kenmerkende geur	something with a characteristic smell
15	makkelijk hoorbaar	easily audible
16	gerelateerd aan beweging	related to movement
17	menselijk	human
18	iets met een hoofd/gezicht	something with a head/face
19	iets met een lichaam	something with a body
20	gerelateerd aan acties met het gezicht/de mond	related to actions with the face/the mouth
21	gerelateerd aan acties met de hand/arm	related to actions with the hand/arm
22	gerelateerd aan acties met de voet	related to actions with the foot
23	iets met een vaste plaats/locatie	something with a fixed place/location
24	iets waarvoor tijd (tijdstip of duur) relevant is	something for which time (point in time or duration) is relevant
25	iets waar jij directe ervaring mee hebt	something you have direct experience with
26	iets wat jou of anderen helpt	something that helps you or others
27	iets waardoor je verrast wordt	something that you are surprised by
28	positief/plezierig	positive/pleasant
29	negatief/onplezierig	negative/unpleasant

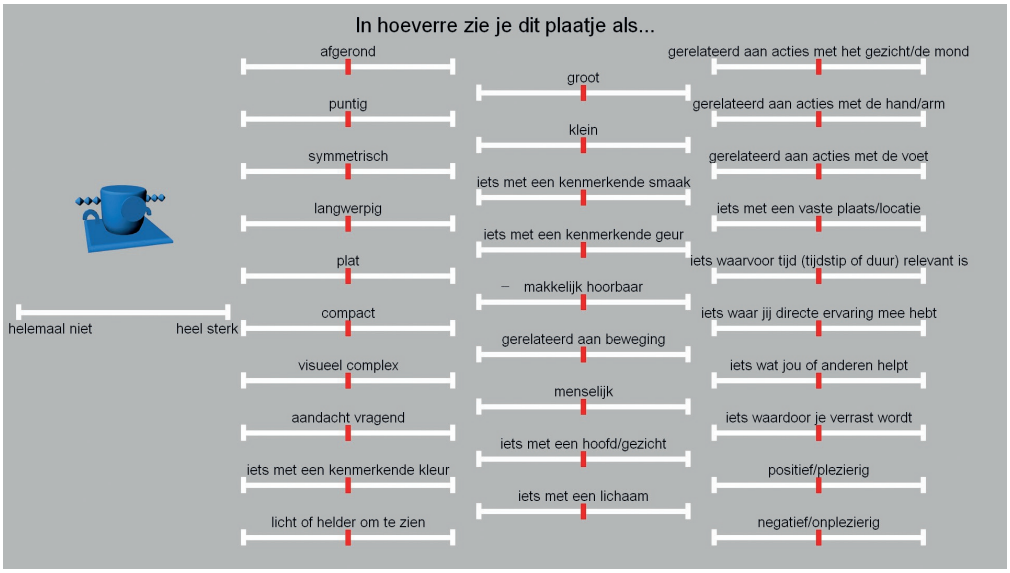


Figure D3. Screenshot of one trial in the Features task. See Table D1 for translations of all features. Participants should rate the Fribble displayed on the left by changing the sliders for all 29 different features.

Table D2. Details on the eight animated movies participants viewed in the movies session.

Movie order	Movie name	Source	Short description	Length shown (s)	Remarks
1	Caminandes Llamigos	Pablo Vazquez, Blender Foundation, http://www.caminandes.com/	A llama chases a penguin in the snow for some berries.	134	
2	Lifted	Pixar Short Films Collection 1 (DVD)	Aliens take away a sleeping man at night.	261	
3	One man band	Pixar Short Films Collection 1 (DVD)	Two men play several instruments at a square for a child.	217	black bar at top/bottom of screen
4	Knick Knack	Pixar Short Films Collection 1 (DVD)	Toys are enjoying music and trying to escape from their shelf.	173	
5	Geri's game	Pixar Short Films Collection 1 (DVD)	An old man plays a game of chess against himself.	245	black bar left/right of screen
6	La Luna	Pixar Short Films Collection 2 (DVD)	A man and child sail to the moon to change its appearance.	365	black bar at top/bottom of screen
7	Presto	Pixar Short Films Collection 2 (DVD)	A magician's show is disturbed by his rabbit.	265	
8	Partly Cloudy	Pixar Short Films Collection 2 (DVD)	A stork brings babies, but appears to always get the difficult ones.	300	

Table D3. Translated questions from the questionnaire (the original questions were in Dutch).

Nr	Translated question
1	Did you notice anything about the experiment or would you like to say something about it? If so, what?
2	What do you think the goal of the experiment is?
3	When you had to name/describe the images for the first time, what strategy did you use? Was it difficult?
4	Did the features influence your naming? If so, how?
5	How did you go about the first features task? Was it difficult?
6	Did the naming influence the features? If so, how?
7	Did you use a different strategy the second time you did the naming task? How? Was it easier/harder?
8	Did you use a different strategy the second time you did the features task? Was it easier/harder?
9	What strategy did you use in the first interactive task with your partner (where you had to describe the images)?
10	What strategy did you use in the second interactive task with your partner (where you had to describe where the image was on the screen)?
11	Did you think you could use gestures during the interactive tasks? If so, did you do so?
12	Did you think the task inside the fMRI scanner with the images was hard? Was it harder/easier/the same the second time?
13	Did you use a certain strategy to do the task in the fMRI scanner? If so, which one?
14	Could you easily keep your attention while watching the videos in the scanner (the last part) or were you distracted sometimes?
15	Have you already seen some of the videos in the fMRI scanner? If so, which ones?
16	You will now get a few questions about the other participant in the experiment. Could you be friends with him/her? Give a score from 1-7 (1 = very unlikely, 7 = very likely)
17	Does the other participant look like you? Give a score from 1-7 (1 = not at all, 7 = very much).
18	How intelligent do you think the other participant is? Give a score from 1-7 (1 = not at all, 7 = very much).
19	How selfish do you think the other participant is? Give a score from 1-7 (1 = not at all, 7 = very much).
20	How shy do you think the other participant is? Give a score from 1-7 (1 = not at all, 7 = very much).
21	How enthusiastic did you think the other participant was during the interaction? Give a score from 1-7 (1 = not at all, 7 = very much).
22	How nice did you think the other participant was? Give a score from 1-7 (1 = not at all, 7 = very much).
23	How pleasant did you think the other participant's voice was to listen to? Give a score from 1-7 (1 = not at all, 7 = very much).
24	Do you think the other participant was a real participant or a collaborator of the researcher?
25	You will now get a few questions about yourself. How old are you?
26	What is your sex?
27	What do you do in your daily life? (studying, working, unemployed, ?)
28	What study programme have you followed/are you following?
29	How introverted/extraverted do you think you are? Give a score from 1-7 (1 = very introverted, 7 = very extraverted).
30	How proud are you of your own accent? Give a score from 1-7 (1 = not at all, 7 = very much).

Table D4. Overview of all data types in the Data folder of the Dataset with associated tasks and formats (extensions).

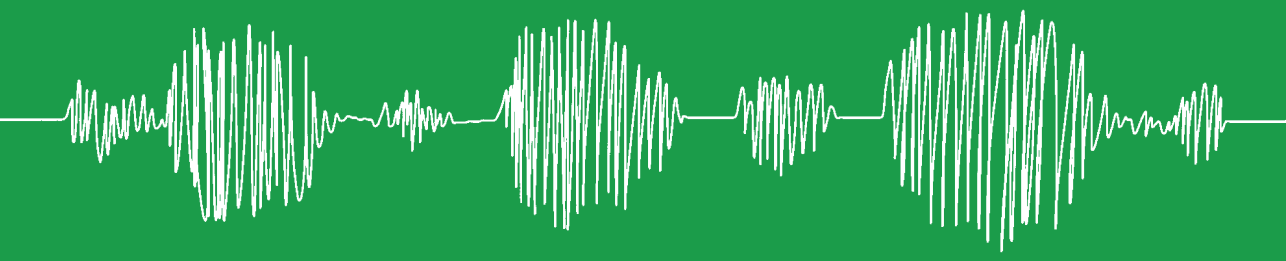
Data type	Task	Format(s)
task log files	Naming and Features, one-back, phonological pre-test, interaction, movies, questionnaire	txt
aggregated data from log files	Naming, Features, questionnaire	csv
MRI	one-back, movies	NIfTI
eye-tracking	one-back, movies	idf
physiological	one-back, movies	eeg; vhdr; vmrk
Audio	phonological pre-test, interaction	wav
Video	interaction	mp4
Kinect	interaction	c3d; csv; txt; log
Transcription	interaction	eaf; pfsx; txt; TextGrid

Section D1. Details on the automatic movement coder

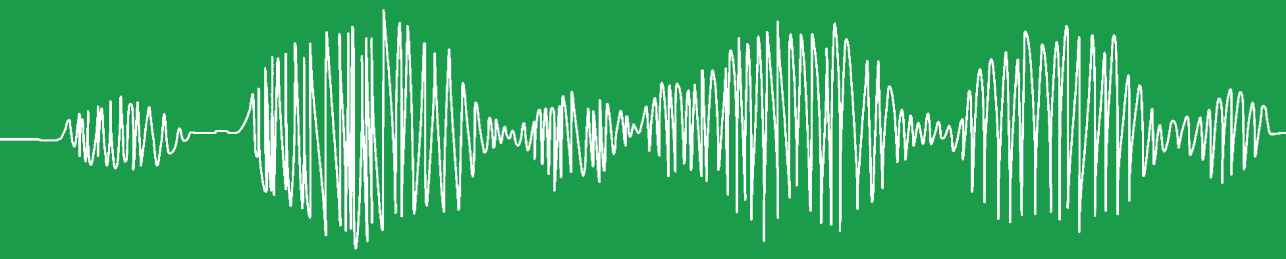
The gesture autocoder was applied to the dominant hand. We determined whether there was a gesture event based on the following rules:

1. A gesture event is considered when the movement speed of the hand tip exceeds 15cm/s.
2. If a candidate gesture event is adjacent to another event within 250 ms, the events are merged, and treated as a single event.
3. If a candidate gesture event is shorter than 200 ms then the event is too short to be considered a gesture event, and such events are excluded. However, by rule 2, if candidate gesture events shorter than 200 ms are adjacent to each other, they will be merged and are therefore not excluded.
4. If the gesture event contains a movement that does not exceed the vertical threshold of 1SD under the average vertical position, the gesture event is excluded. This is because we want to avoid the detection of button presses as communicative gestures.

This autocoder is comparable in nature to other benchmarked approaches (Ripperda, Drijvers, Holler, 2020), with the exception that we did not apply a manual removal for possibly incorrectly detected communicative gestures.



CHAPTER 7



General discussion and conclusions

The main goal of this dissertation was to expand the knowledge on linguistic alignment and gain more insight into the underlying mechanisms by investigating features of three different levels: the syntactic level, the prosodic level and the segmental phonetic level. This dissertation closely investigated local versus global alignment, and possible interlocutor-specificity. Two datasets were developed that are well suited for further future alignment studies.

The current chapter is aimed at presenting the two datasets, briefly summarising the findings in this dissertation, linking them to existing research, bringing studies in Chapter 2 to 5 together, and reflecting on the theories presented in Chapter 1, and throughout the other chapters. Moreover, this chapter will discuss future directions of research into linguistic alignment.

7.1 The two datasets

This dissertation resulted in two different datasets. The first dataset formed the basis for Chapters 2 to 5, and the second dataset is described in Chapter 6. They both contain data from Dutch native speakers.

7.1.1 Dataset 1: The Sentence Completion in Interaction with Two Interlocutors (SCITI) dataset

The first dataset comprised of the data from two experiments: a main and a control experiment. Both experiments are a sentence completion task. A total of 72 participants divided over the two experiments each completed 268 sentences. They all preferred the red syntactic order (auxiliary before the past participle in subclauses), while they differed in their preferred allophone of the /x/.

In both experiments, participants started with a pre-test in which they completed sentences by themselves. They then alternated completing sentences with two different confederates. These confederates differed from each other, first in their order of the auxiliary and the past participle in subordinate clauses and, second, in their allophone of the /x/. Participants alternated with Confederate 1 in Round 1, with Confederate 2 in Round 2, in an inter-test and Round 3 with Confederate 1 again, and lastly in a post-test by themselves. The inter-test and Round 3 differed from each other in that, in the inter-test, the confederate did not produce subclauses with the auxiliary-past participle combination, and hardly any /x/s.

The control experiment differed from the main experiment in two aspects. First, in the control experiment, participants read instead of heard the full sentences produced by the confederates. Second, in the control experiment the confederates did not produce subordinate clauses with the auxiliary-past participle combination. As a consequence, some of the confederates' stimuli were replaced.

7.1.2 Dataset 2: The Communicative Alignment in Brain and Behaviour (CABB) dataset

The second dataset was described in Chapter 6. This large dataset includes data from 71 pairs of participants performing several tasks focused around multimodal referential communication. These tasks were based on, to participants unknown, objects called *Fribbles*. Because these Fribbles were unknown objects, participants would have to negotiate the names of the objects in an interaction.

The dataset consists of audio, video and motion-tracking data of an interaction, pre- and post-measures of fMRI observations and two different behavioural tasks, and data of a questionnaire. Data of three different cameras, motion-tracking, head-mounted microphones, and orthographic transcriptions of a large part of the dataset are available. The interaction consisted of two different tasks, one referential communication task in which participants described the Fribbles to each other, and a task in which participants decided whether a Fribble was located in the same location on their respective screens. The behavioural pre- and post-tasks were a naming task in which participants were to label the Fribbles, and a features task, in which participants were to rate the Fribbles on certain features (e.g., human, rounded). The fMRI data was collected by presenting participants with a one-back working memory task, and furthermore included a post-session in which participants watched different short movies for which the data could be used for functional hyperalignment.

Chapter 6 presented initial analyses showing that the dataset is of high quality and can be used for a range of analyses within and across different disciplines such as linguistics, psychology and neuroscience. Analysis of the behavioural naming task, for instance, comparing pre- to post-interaction, showed that the interaction resulted in more similar labels for the Fribbles.

7.2 Alignment on different linguistic levels

Alignment is a ubiquitous phenomenon, occurring on different linguistic levels. Previous research has investigated all of these levels. In this dissertation, the syntactic, prosodic and segmental phonetic levels were further investigated in the same dataset, as presented in the previous section. The results of the analyses based on this dataset will briefly be discussed per level and how they relate to previous work.

7.2.1 Syntactic level

In Chapter 2, syntactic alignment to a Dutch syntactic alternation was investigated. The alternation under investigation was the so-called "red" versus "green order" – the

order of auxiliary verb-past participle versus past participle-auxiliary verb – in Dutch subordinate clauses (e.g., "Het rapport van het jongetje toonde aan dat hij zijn best had gedaan." versus "Het rapport van het jongetje toonde aan dat hij zijn best gedaan had." English translation: *The little boy's report card showed that he had done his best.*). The presence of syntactic alignment was investigated by assessing the presence of local (also referred to as short-term alignment in Chapter 2) and long-term global alignment (referred to as long-term persistency alignment in Chapter 2). In addition, a goal of this study was to test for interlocutor-specific alignment, which was possible because participants interacted with two different interlocutors, and because of the inclusion of an inter-test in the experiment where participants did not receive any input containing the pertinent syntactic structure.

Local alignment was measured in consecutive turns in the rounds with the interlocutors. Long-term global alignment was measured by comparing participants' use of the syntactic order between the pre- and the post-test. These possible long-term global effects were also studied to find out whether they could be overruled by local alignment effects when participants switched from speaking to one interlocutor to the other (i.e., at the beginning of a new round).

Evidence for local alignment to the interlocutors was found in the rounds where participants interacted with the confederates in the main experiment, and not in the control experiment. Participants in the main experiment used more of the green order in Round 1 and 3, where they interacted with Confederate 1, who used the green order, and more of the red order when interacting with Confederate 2, who used the red order. This is in line with other literature on syntactic alignment, where local alignment effects have often been found (e.g., Branigan, Pickering & Cleland, 2000; Hartsuiker, Bernolet, Schoonbaert, Speybroeck & Vanderelst, 2008). In contrast, there was no evidence of long-term global alignment effects.

Furthermore, no evidence of speakers realigning to the interlocutor without hearing the pertinent syntactic structure was found in the inter-test – indicating no proof of interlocutor-specific syntactic alignment. This is also in line with other studies on syntactic alignment, where no speaker-specific alignment was found (Ostrand & Ferreira, 2019).

Several studies have proposed syntactic alignment could lead to implicit learning (e.g., Bock & Griffin, 2000; Chang, Dell & Bock, 2006; Hartsuiker et al., 2008). Implicit learning should be revealed in our dataset as global alignment effects, for which we found no evidence in our dataset. It could be that participants in our dataset did not receive enough consistent input for the effects to result in clear implicit learning effects. This could be tested by repeating a lengthened version of this experiment.

One can also interpret the results as alignment to only one of the interlocutors – alignment to Confederate 2. Participants in both the main and the control experiment use less of the red order in Round 1 as compared to the pre-test. This indicates that participants selected more of the red order in the pre-test than in Round 1 by chance, and that the difference between the pre-test and Round 1 does not imply alignment in the main experiment. If so, the data still suggest alignment to the second interlocutor, in the sense of a strengthening of the preference of the red order compared to Round 1 and Round 3. Because of the difference in the participants' use of the red order between Round 2 and Round 3 (rounds with different interlocutors), the data still indicate that local alignment is stronger than possible global effects to Confederate 2 in Round 2. Whether participants also align to confederates using the order opposite to their own (i.e., whether alignment can also occur in Round 1) should be investigated in future studies, for example by also including a group of participants with a preference for the green order in the pre-test, and investigating whether they also show alignment in Round 2.

These two different interpretations of the data in the main experiment (i.e., whether there is or whether there is not alignment to the first interlocutor) show the importance of a control experiment which can show whether effects that have been found in an experimental study actually indicate alignment or whether they could be caused by other processes. The importance of the control experiment is also shown in the interpretation of the difference between the pre- and the post-test in the main experiment. Without the control experiment, the data could have been interpreted as showing long-term global alignment effects. However, since this difference was also present in the control experiment, it cannot be attributed to the manipulation in the main experiment. The pre-test baseline was also important for a correct interpretation of the data because it served as a baseline for participants' use of the syntactic structure. Without this baseline, participants' productions when alternating with the confederates could not be compared to participants' baseline use of the syntactic structure.

7.2.2 Prosodic level

In Chapter 3, a subset of the data from the main experiment was analysed to investigate alignment at the prosodic level. More specifically, this chapter focused on local and global alignment of pitch (measured as median F0) and articulation rate (measured as the number of syllables per second). The confederates differed in their pitch and articulation rate: Confederate 1 had an overall higher pitch and articulation rate than Confederate 2. Local alignment was estimated by predicting participants' pitch and articulation rate in a sentence based on those of the confederate in the directly preceding sentence. Global alignment effects were investigated by studying the time course of pitch and articulation rate over the rounds and by comparing the pre-test to

the post-test. Chapter 3 was only based on the data from the main experiment, because control data were not yet available at the time.

Chapter 3 suggests global alignment to pitch and articulation rate. Furthermore, alignment effects lasted in the post-test – when speakers were no longer interacting with an interlocutor – indicating long-term global effects. In contrast, there was no evidence for local alignment. These results are in line with previous studies reporting global effects in pitch alignment (Oben, 2015). They are, however, not in line with a large share of the prosodic alignment literature, which mainly found local alignment effects, and less so global alignment effects (e.g., Gijssels, Casasanto, Jasmin, Hagoort & Casasanto, 2016). Other studies found divergence in articulation rate, modulated by social factors (Schweitzer & Lewandowski, 2013). This shows the importance of conducting more research on alignment at the prosodic level so that a comparison can show the possible sources for these different findings, including the role of methodology.

Chapter 4 added to Chapter 3 by analysing data from the control experiment. None of the effects presented in Chapter 3 indicating alignment were observed in Chapter 4. This indicates that the effects found in Chapter 3 were actually due to alignment rather than to task-related factors.

The confederates in the experiment were not instructed on how to exactly articulate the sentences. As a consequence, their articulation rate and pitch varied from one sentence to the next. This may have obscured local alignment effects. In a future experiment, the pitch and articulation rate of the confederates could be made to vary less. This would lead, however, to less naturally sounding recordings.

7.2.3 Segmental phonetic level

The aim of Chapter 5 was to investigate phonetic alignment to a Dutch allophone – the so-called "hard g" (uvular) versus the "soft g" ((palato-)velar and palatal). The former regional variant is part of a more prestigious accent (e.g., Grondelaers & van Hout, 2010). This study examined whether speakers align more to an allophone belonging to a prestigious accent than to one of a less prestigious accent, whether this can better be explained by local or global measures, and whether hearing the allophone in question is essential for alignment or if hearing an interlocutor again, without this interlocutor using this allophone, can also change speakers' productions. These questions were investigated by comparing the different parts of the experiment and by means of three different alignment measures ranging from more local to global.

Phonetic alignment was investigated in two measures, the duration and Centre of Gravity (CoG) of the /x/, since these continuous measures should be able to show gradual and subtle changes in the participants' productions. Participants' productions were compared between the different parts of the main and control experiments.

Moreover, three different measures of alignment were investigated, ranging from local to more global, to investigate the time course of alignment. These measures were based on the residuals of a statistical model, predicting duration on the basis of control variables like articulation rate for both participants and confederates. The residuals were assumed to contain less variation due to contextual influences such as articulation rate and therefore reflect the "intrinsic" durations of the /x/. The first measure reflected the last single production of the interlocutor, the second measure the average of the last ten productions of the interlocutor, and the third measure the average of all productions of the interlocutors that the speaker had heard up until that point.

None of the analyses showed any significant alignment effects. Comparisons between the rounds with the different confederates or the pre- and the post-test showed no difference for either duration or CoG. There was also no difference between participants' productions between the main experiment and the control experiment, or evidence for alignment in the three measures designed to investigate local versus global alignment in the duration of the /x/.

One of the research questions in this study was whether speakers needed to be presented with a specific feature of an accent to change their productions of that feature, or whether they would also change their production in interaction with an interlocutor they had spoken to before, without this interlocutor using this feature. Since I found no evidence for alignment, I cannot draw any conclusions in regard to this question.

When investigating the duration data, we observed large individual differences in alignment patterns: while some participants showed alignment, others showed divergence and yet others showed no effect at all from the confederates' realisations. These individual differences could not be explained on the basis of participants' baseline productions in the pre-test (whether they produced a *hard g*, *soft g*, both or something in between), participants' preference for the accent of one or the other interlocutor, or participants' pride of their own accent.

A possible other explanation for the individual differences could be an interplay of several factors, where the prestige of an accent may be important for one speaker, while for another speaker the retaining of one's identity as reflected by their accent may be imperative. Moreover, other social factors, such as the competence of the interlocutor or personality traits of the speaker, may also play a role. Future research in alignment to regional variants should further investigate the sources of individual differences for example, by relating alignment effects to the answers to a thorough questionnaire on how the participants see themselves, their interlocutors, and their accents. In this manner, it could be possible to investigate which specific factors, or interplay of multiple factors, could explain the individual differences in alignment to regional variants.

Even though this lack of evidence for phonetic alignment seems to be different from a large share of studies on phonetic alignment (e.g., Pardo, 2006; Berry & Ernestus, 2018), our findings fit well in the phonetic alignment literature on regional variants more specifically. Several studies have found that there are large individual differences (e.g., Earnshaw, 2020; Gessinger, Möbius, Fakhar, Raveh & Steiner, 2019b), and some of these studies explicitly mention that these differences were not directly explicable by the speakers' opinions about their interlocutor (Gessinger et al., 2019b). Chapter 5 expanded these findings in phonetic alignment in response to regional variants to local and global measures of alignment to allophones in Dutch, a language that has until now not been investigated in this respect.

7.2.4 Possible explanations for differences between the present data and the literature

The different studies reported in this dissertation provided different results with respect to alignment. Some of these results deviate from other studies on alignment at the same linguistic level. One reason why our results may seem to differ from other studies is that null results, such as those obtained in Chapter 5, might not always get published. As a consequence, it is unclear how stable alignment effects actually are.

Another possible reason for differences between our results and some of the results reported in the literature may be due to the different ways of testing alignment. Communicative intent has been argued to affect alignment (e.g., Ostrand & Ferreira, 2019). As a consequence, alignment effects may differ between more communicative tasks and a sentence completion task, and between situations in which the interlocutor is present versus absent. Moreover, the interlocutors in the dataset analysed in Chapters 2 to 5 did not change their productions to possibly align to the participants during the experiment since their speech was pre-recorded. The effects of an interlocutor who changes their speech as compared to an interlocutor who does not should be explored in future studies.

7.3 Reflection on theories

There is an abundance of theories on linguistic alignment. One of the most well-known theories hypothesises that the underlying mechanism of alignment is automatic priming (Pickering & Garrod, 2004). Another theory is that by Clark (1996) who states that alignment is a process related to the creation of common ground in interactions. This creation of common ground is needed for successful conversation and is argued to be conscious and speaker-specific. A third theory, proposed by Giles, Coupland & Coupland (1991) states that alignment is socially mediated. These three theories are not

mutually exclusive, but all may be more or less applicable to different contexts, levels or time scales.

In the following, I will discuss how the results presented in this dissertation can contribute to these theories. I will first discuss the different linguistic levels, and I will then speculate on how to combine the different levels relating to underlying mechanisms.

7.3.1 Alignment on the different levels in relation to the theories

In short, I have found evidence for local alignment in a syntactic word order, for global alignment in pitch and articulation rate, an indication of large individual differences for alignment to regional variants on the phonetic level, and no evidence for any interlocutor-specific alignment. These results indicate that alignment on the different linguistic levels and in the different features of a single level is not directly comparable and also does not directly seem to be driven by the same (combination of) mechanisms.

In particular, by comparing several parts of the experiment, the analyses of Chapter 2 helped us understand that automatic priming could be the mechanism underlying syntactic alignment, since we found evidence for local alignment and no evidence for speakers aligning to an interlocutor without this interlocutor using the syntactic structure. In contrast, automatic priming could be less successful in triggering alignment in other linguistic levels, such as the prosodic and the segmental phonetic level, where there was no direct evidence for local priming as the source of alignment (Chapters 3, 4, 5). Thus, the collection of studies included here provide partial evidence for the Interactive Alignment Theory by Pickering and Garrod (2004).

In addition, our study showed that speakers may also adapt their speech to reflect the interlocutors' productions more globally on the prosodic level and that speakers maintain this global alignment after the interaction. This suggests a slow adaptation to pitch and articulation rate of the interlocutors (as also proposed for syntactic alignment by Reitter and Moore, 2007), which could reflect implicit learning, which in turn is reflected in global effects in the studies in this dissertation.

This dissertation did not provide any evidence for interlocutor-specific alignment. Exposure to the syntactic structure seemed to be pertinent for speakers to align to this structure. The prosodic and segmental phonetic level also did not provide any evidence for interlocutor-specific alignment. The results therefore are not in line with Clark's theory, which is mainly based on the lexical level.

Furthermore, social factors potentially influence alignment, at least on alignment to regional variants. Here, large individual differences were found. The underpinnings of the idiosyncratic differences are not yet clear, but can most likely be linked to (an interplay of) social and communicative factors as suggested by Giles and colleagues (1991).

7.3.2 Speculations on the underlying mechanisms

Linguistic alignment is clearly a complex phenomenon that does not behave the same on different linguistic levels or even for different features on the same linguistic level. The broad phenomenon of alignment seems to demand a combination of different mechanisms as described in the various theories. I would like to stress that this paragraph concerns speculations. Speculatively, I would suggest that alignment results from a heightened activation of cognitive representations. It could be that local alignment effects are established when the single use of a feature by an interlocutor results in high activation in a speaker. More global alignment effects could possibly be explained by a more subtle cumulative increase in activation with each occurrence of the feature adding to the activation level. An immediate increase in activation may be expected if the feature is rather salient and categorical (such as the two different orders for the syntactic structure in this dissertation). A more gradual increase may be expected when the input is continuous and may not differ radically between what a speaker usually uses and what they hear. The gradual increase in activation could build up in a cumulative way, and could reflect some implicit learning process (Bock and Griffin, 2000, Chang et al., 2006).

Speculatively, the increase in activation can be influenced by social and communicative factors. Speakers may actively adapt to or actively supress features that are produced by their interlocutors, for instance, if these features are relevant for successful communication or reflect a social group the interlocutor belongs to. In contrast, features that may be less socially or communicatively relevant may be less consciously adapted to by the speaker. This alignment could nevertheless help speakers make conversation less effortful (as suggested by Pickering and Garrod, 2004) as relevant cognitive representations are more activated (such as pitch and syntactic structures in many contexts).

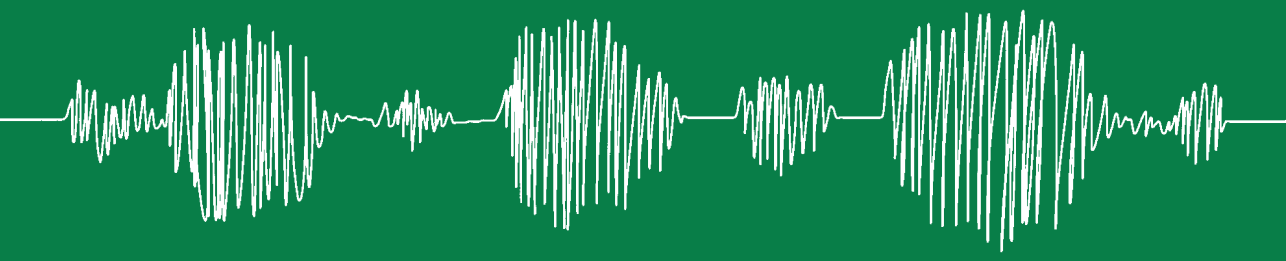
These speculations could be further investigated in both datasets presented in this dissertation. The combination of these two datasets offers a range of possibilities for the investigation of alignment in an experimental setting where several features are manipulated, as well as in an experimental referential communication task. The first dataset, as discussed in Chapters 2 to 5, is possibly better suited for more investigations on the syntactic, prosodic, and segmental phonetic levels as it contains a large amount of speech per participant with a large variety of content words. The dataset presented in Chapter 6 is likely better equipped to be investigated for lexical, semantic, and gestural alignment. In both datasets, different linguistic levels can be investigated for alignment to investigate Pickering and Garrod's (2004) theory suggesting that alignment on different levels of alignment influence each other.

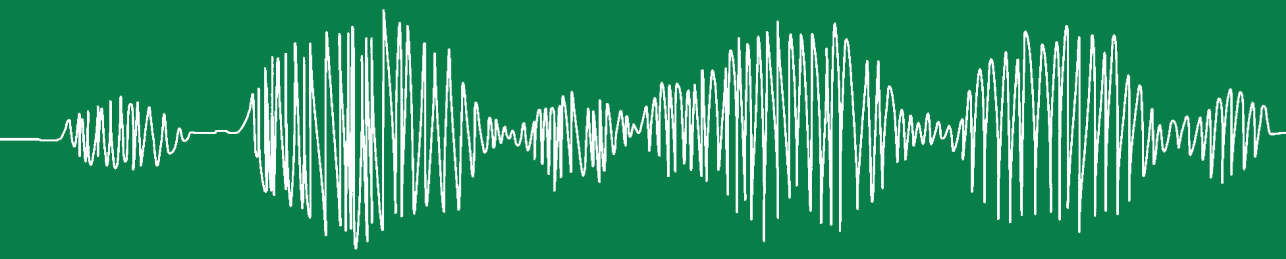
Clark's (1996) theory can be further investigated in the dataset presented in Chapter 6, since this dataset includes interactions in a communicative task. Firstly, alignment on different linguistics levels could be related to task success to assess which levels or features of alignment could be imperative for task success. Secondly, this dataset includes a questionnaire with a range of questions about the speaker and the likeability of their interlocutor, which could shed further light on the social factors influencing different levels of alignment. Thirdly, the efficiency of the task (which could indicate the ease of the conversation) can be assessed to be related to measures of alignment on different linguistic levels. Lastly, alignment in interaction can be related to behavioural and neural pre- and post-measures in the second dataset. Although more spontaneous settings are possible, this dataset is a great step towards more ecologically valid settings for the investigation of alignment.

7.4 Conclusions

This dissertation contributed to the study of linguistic alignment in multiple ways. The first contribution is evidence that linguistic alignment is a complex phenomenon. The underlying mechanisms of alignment were investigated for different features and time lines on different linguistic levels. Alignment is clearly not a straightforward phenomenon and cannot be explained by one single mechanism for all linguistic features. The second contribution is methodological in nature. This dissertation clearly shows the importance of baseline measures and control experiments in alignment research. Without the use of either, our interpretation of the results would have differed, which would have led to false conclusions. The third and last contribution is the creation of two different datasets that allow for careful investigation of alignment. The first dataset was used to study syntactic, prosodic and segmental phonetic alignment, and can further be investigated for other linguistic features, on all linguistic levels, and for a direct comparison of alignment to different features at different levels. The second dataset consists of task-based interactions between participants, including pre- and post-measures of behavioural (names and features) and neural correlates of different objects discussed in the interaction. As in the first dataset, alignment can be investigated in linguistic behaviour, both locally and globally, and both in single measures or more holistically, providing a rich starting point for future research. Moreover, the dataset presented in Chapter 6 also includes video data and pre- and post-measures, which makes it possible to study other non-linguistic behaviour such as gestures.

In conclusion, this dissertation contributed new evidence on linguistic alignment in the syntactic, prosodic, and segmental phonetic domains. This evidence clearly shows that alignment is a complex phenomenon, with variation within and across linguistic levels, and between individuals.





References

Ahlmann-Eltze, C. (2019). ggsignif: Significance Brackets for 'ggplot2'. R package version 0.6.0. https://CRAN.R-project.org/package=ggsignif

Allen, M. L., Haywood, S., Rajendran, G., & Branigan, H. (2011). Evidence for syntactic alignment in children with autism. *Developmental science, 14*(3), 540-548.

Allwood, J., Nivre, J., & Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, *9*(1), 1-26. https://doi.org/10.1093/jos/9.1.1

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC map task corpus. *Language and speech, 34*(4), 351-366. https://doi.org/10.1177/002383099103400404

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). The CELEX lexical database (cd-rom).

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*(4), 437-456.

Bangerter, A., Mayor, E., & Knutsen, D. (2020). Lexical entrainment without conceptual pacts? Revisiting the matching task. *Journal of Memory and Language*, *114*, 104129. https://doi.org/10.1016/j.jml.2020.104129

Barry, T. J., Griffith, J. W., De Rossi, S., & Hermans, D. (2014). Meet the Fribbles: Novel stimuli for use within behavioural research. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00103

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Bergmann, K., & Kopp, S. (2012). Gestural Alignment in Natural Dialogue. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1326–1331). Cognitive Science Society.

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming?. *Cognition, 114*(3), 455-461.

Berry, G. M., & Ernestus, M. (2018). Phonetic alignment in English as a lingua franca: Coming together while splitting apart. *Second Language Research, 34*(3), 343-370.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bock, J. K. (1989). Closed-class immanence in sentence production. *Cognition, 31*, 163-186.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*(2), 177–192. https://doi.org/10.1037/0096-3445.129.2.177

Boersma, P., Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from http://www.praat.org/.

Bonin, F., De Looze, C., Ghosh, S., Gilmartin, E., Vogel, C., Polychroniou, A., Salamin, H., Vinciarelli, A., Campbell, N. (2013). Investigating fine temporal dynamics of prosodic and lexical accommodation. *Proc. INTERSPEECH* Lyon, 539–543.

Bracci, S., Caramazza, A., & Peelen, M. V. (2015). Representational similarity of body parts in human occipitotemporal cortex. *Journal of Neuroscience*, *35*(38), 12977-12985. https://doi.org/10.1523/JNEUROSCI.4698-14.2015

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13–B25. https://doi.org/10.1016/S0010- 0277(99)00081-5

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition, 104*(2), 163–197. https://doi.org/10.1016/j.cognition.2006.05.006

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Nass, C. (2003). Syntactic alignment between computers and people: The role of belief about mental states. In *Proceedings of the 25th annual conference of the cognitive science society* (Vol. 31). Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482– 1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science, 1*(2), 274-291.

Brône, G., & Oben, B. (2015). InSight Interaction: A multimodal and multifocal dialogue corpus. *Language Resources and Evaluation*, *49*(1), 195–214. https://doi.org/10.1007/s10579-014-9283-2

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291-7299).

Cappellini, M., Holt, B., & Hsu, Y.-Y. (2022). Multimodal alignment in telecollaboration: A methodological exploration. *System*, 102931. https://doi.org/10.1016/j.system.2022.102931

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological review, 113*(2), 234.

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of psycholinguistic research, 29*(2), 217-230.

Chia, K., Axelrod, C., Johnson, C., Bressler, M., Cooperman, H., Chu, A., Dash, E., Di Bella, J., Engelhardt, A., Farruggio, V., Folsom, S., Gomariz, H., Greiner, E., Hager, S., Hansen, N., Kenefick, C., King, J., King, K., Lavaud, M., Leone, E., McGuire, G., Montanez, S., Morpeth, J., Neumann, M., Rivera, D., Sotolongo, N., Sparacio, K., Stokes, K., Tarro, D., Treacy, A., Wagler, K., Weitzel, S., Woller, S., & Kaschak, M. P. (2019). Structural repetition and question answering: A replication and extension of Levelt and Kelter (1982). *Discourse Processes, 56*, 2-23.

Chia, K., Hetzel-Ebben, H., Adolph, M., Amaral, M., Arriga, M., Booth, H., Boudreau, V., Carpenter, J., Cerra, C., Clouden, M., Cryderman, J., Darij, R., Dollison, J., Franco, N., Ghougasian, L., Hamilton, L., Karosas, K., Kenoyer, C., Krenz, V., Lancaster, S., Ma, M., Markwell, G., Montoya, F., Nadler, R., Pinto, S., Rojas, M., Sarmiento, D., Stitik, C., St. John, J., Valencia, M., Walker, K., Wells, E., Wolf, J., Wright, D., & Kaschak, M. P. (2020). Examining the factors that affect structural repetition in question answering. *Memory & Cognition, 48*, 1046-1060.

Chui, K. (2014). Mimicked gestures and the joint construction of meaning in conversation. *Journal of Pragmatics*, *70*, 68–85. https://doi.org/10.1016/j.pragma.2014.06.005

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

Clark, H. H. (1997). Dogmas of Understanding. *Discourse Processes, 23(3),* 567–582. https://doi.org/10.1080/01638539709545003

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39. https://doi.org/10.1016/0010-0277(86)90010-7

Clayman, S. E. (2013). Turn-Constructional Units and the Transition-Relevance Place. In *The Handbook of Conversation Analysis* (pp. 151–166). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118325001.ch8

Couper-Kuhlen, E., & Selting, M. (2017). *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge University Press.

Dale, R., & Spivey, M. J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning, 56*(3), 391-430.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods, 41*(2), 385-390.

Dideriksen, C., Fusaroli, R., Tylén, K., Dingemanse, M., & Christiansen, M. H. (2019). Contextualizing Conversational Strategies: Backchannel, Repair and Linguistic Alignment in Spontaneous and Task-Oriented Conversations. In Goel, A.K., Seifert, C.M., & Freksa, C. (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 261–267). Cognitive Science Society.

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature communications*, *10*(1), 1-10. https://doi.org/10.1038/s41467-019-09239-1

Dragojevic, M., Gasiorek, J., & Giles, H. (2016). Accommodative strategies as core of the theory. In H. Giles (Ed.), *Communication accommodation theory: Negotiating personal relationships and social identities across contexts*. Cambridge, UK: Cambridge University Press.

Duran, N. D., Paxton, A., & Fusaroli, R. (2019). ALIGN: Analyzing linguistic interactions with generalizable techNiques - A Python library. *Psychological Methods*, *24*(4), 419–438. https://doi.org/10.1037/met0000206

Earnshaw, K. (2020). *A forensic phonetic investigation of regional variation and accommodation in West Yorkshire* (Doctoral dissertation, University of Huddersfield).

Edlund, J., Beskow, J., Elenius, K., Hellmer, K., Strömbergsson, S., & House, D. (2010). Spontal: A Swedish Spontaneous Dialogue Corpus of Audio, Video and Motion Capture. In *LREC* (pp. 2992-2995).

Elvira-Garcia, W. (2014). Zero-crossings-and-spectral-moments, v.1.3 [Praat script]

Emina, K., & Jan, G. (2018). F0 accommodation and turn competition in overlapping talk. *Journal of phonetics, 71*, 376-394.

Ernestus, M., & Baayen, R. H. (2011). Corpora and exemplars in phonology. In *The handbook of phonological theory (2nd ed.)* (pp. 374-400). Wiley-Blackwell.

Ernestus, M., Kouwenhoven, H., & Van Mulken, M. (2017). The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics, 62*, 50-64.

Felker, E., Broersma, M., & Ernestus, M. (2021). The role of corrective feedback and lexical guidance in perceptual learning of a novel L2 accent in dialogue. *Applied Psycholinguistics*, 1-27.

Fox, J. & Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological science, 23*(8), 931-939.

Fusaroli, R., Tylén, K., Garly, K., Steensig, J., Christiansen, M. H., & Dingemanse, M. (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, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2055–2060). Cognitive Science Society.

Gessinger, I., Möbius, B., Andreeva, B., Raveh, E., & Steiner, I. (2019a). Phonetic Accommodation in a Wizard-of-Oz Experiment: Intonation and Segments. In *INTERSPEECH* (pp. 301-305).

Gessinger, I., Möbius, B., Fakhar, N., Raveh, E., & Steiner, I. (2019b). A Wizard-of-Oz experiment to study phonetic accommodation in human-computer interaction. In *International Congress of Phonetic Sciences (ICPhS)*, Melbourne (pp. 1475-1479).

Gigant-Molex (Version 1.0) (2019) [Data set]. Available at the Dutch Language Institute: http://hdl.handle.net/10032/tm-a2-p9

Gijssels, T., Casasanto, L. S., Jasmin, K., Hagoort, P., & Casasanto, D. (2016). Speech accommodation without priming: The case of pitch. *Discourse Processes, 53*(4), 233- 251.

Giles, H., & Gasiorek, J. (2013). Parameters of non-accommodation: Refining and elaborating communication accommodation theory. In J. Forgas, J. László, & V. Orsolya Vincze (Eds.), *Social cognition and communication* (pp. 155–172). New York, NY: Psychology Press.

Giles, H., Coupland, J., & Coupland, N. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., Poline, J.-B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J.A., Varoquaux, G.& Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, *3*(1), 1-9. https://doi.org/10.1038/sdata.2016.44

Gorisch, J., Wells, B., & Brown, G. J. (2012). Pitch contour matching and interactional alignment across turns: An acoustic investigation. *Language and Speech, 55*(1), 57-76.

Grondelaers, S., & Van Hout, R. (2010). Is Standard Dutch with a regional accent standard or not? Evidence from native speakers' attitudes. *Language Variation and Change, 22*(2), 221-239.

Haeseryn, W. J. M. (1990). Syntactische normen in het Nederlands: een empirisch onderzoek naar volgordevariatie in de werkwoordelijke eindgroep [Syntactic norms in Dutch: an empirical study on order variation in the verbal endgroup]. Unpublished doctoral dissertation. Nijmegen: University of Nijmegen.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition, 75*(2), B27-B39.

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language, 58*(2), 214-238.

Hawkins, R. D., Frank, M. C., & Goodman, N. D. (2020). Characterizing the dynamics of learning in repeated reference games. *Cognitive science*, *44*(6), e12845. https://doi.org/10.1111/cogs.12845

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. *Neuron*, *72*(2), 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language*, *79–80*, 1–17. https://doi.org/10.1016/j.jml.2014.10.004

Holler, J., & Bavelas, J. (2017). Multi-modal communication of common ground: A review of social functions. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 213–240). Benjamins.

Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, *35*(2), 133-153. https://doi.org/10.1007/s10919-011-0105-6

Holler, J., Bavelas, J.B., Woods, J., Geiger, M., & Simons, L. (2022). Given-new effects on the duration of gestures and of words in face-to-face dialogue. *Discourse Processes.*

Howes, C., Healey, P. G., & Purver, M. (2010). Tracking lexical and syntactic alignment in conversation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).

Hutchins, E., & Hazlehurst, B. (1995). How to invent a shared lexicon: The emergence of shared form-meaning mappings in interaction. In E. N. Goody (Ed.), *Social Intelligence and Interaction* (pp. 189–205). Cambridge: Cambridge University Press.

Ivanova, I., Horton, W. S., Swets, B., Kleinman, D., & Ferreira, V. S. (2020). Structural alignment in dialogue and monologue (and what attention may have to do with it). *Journal of Memory and Language, 110*, 104052.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127*(1), 57-83.

Jefferson, G. (1984). Notes on a systematic deployment of the acknowledgement tokens "yeah"; and "mm hm." *Paper in Linguistics, 17*(2), 197-216. https://doi.org/10.1080/08351818409389201

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior research methods, 42*(3), 643-650.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59*(5), 1208-1221.

Knutsen, D., Bangerter, A., Mayor, E., Zwaan, R., & Dingemanse, M. (2019). Procedural coordination in the matching task. *Collabra: Psychology*, *5*(1). https://doi.org/10.1525/collabra.188

Kousidis, S., Dorran, D., Mcdonnell, C., & Coyle, E. (2009). Convergence in human dialogues time series analysis of acoustic feature. In *Proceedings of SPECOM* (p. 2).

Krauss, R. M., & Pardo, J. S. (2004). Is alignment always the result of automatic priming? *Behavioral and Brain Sciences, 27*(2), 203-204.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(1), 113-114. https://doi.org/10.3758/BF03342817

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14*(1), 78–106. https://doi.org/10.1016/0010- 0285(82)90005-6

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of the Annual Conference of the International Speech Communication Association*, (pp. 3081–3084). INTERSPEECH.

Levitan, R., Gravano, A., Willson, L., Beňuš, Š., Hirschberg, J., & Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies* (pp. 11-19).

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior Matching in Multimodal Communication Is Synchronized. *Cognitive Science*, *36*(8), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, *91*, 5–27. https://doi.org/10.1016/j.jml.2016.03.009

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57-78. https://doi.org/10.1016/j.jml.2016.04.001

Marcoux, K., Ernestus, M. (2019). Pitch in native and non-native Lombard speech. *Proc. 19th ICPhS* Melbourne.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

Mills, G. J. (2011). The emergence of procedural conventions in dialogue. In: Bar-el, L. A. & Hölscher, C. & Shipley, T. (Eds.) *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*.

Oben, B. (2015). *Modelling interactive alignment: A multimodal and temporal account*. Unpublished doctoral dissertation, KU Leuven.

Oben, B., & Brône, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics*, *104*, 32–51. https://doi.org/10.1016/j.pragma.2016.07.002

Oostdijk, N. (2000). The Spoken Dutch Corpus. Overview and First Evaluation. In *LREC* (pp. 887-894).

Ostrand, R., & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of Phonetics, 88*. https://doi.org/10.1016/j.wocn.2021.101074 101074.

Ostrand, R., & Ferreira, V. S. (2019). Repeat after us: Syntactic alignment is not partner-specific. *Journal of Memory and Language, 108*, 104037. https://doi.org/10.1016/j.jml.2019.104037

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. The *Journal of the Acoustical Society of America, 119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language, 69*(3), 183-195.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in cognitive sciences, 3*(4), 136-141.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190. https://doi.org/10.1017/S0140525X04000056

Pinget, A.C.H., M. Rotteveel & H. Van de Velde (2014). Standaardnederlands met een accent: herkenning en evaluatie van regionaal gekleurd Standaardnederlands in Nederland. *Nederlandse Taalkunde 19* (1), 3-46.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America, 118*(4), 2561-2569.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2006). Effects of word frequency on the acoustic durations of affixes. In *Ninth International Conference on Spoken Language Processing*.

Pouw, W., De Wit, J., Bögels, S., Rasenberg, M., Milivojevic, B., Ozyurek, A. (2021). Semantically related gestures move alike: Towards a distributional semantics of gesture kinematics. In: Duffy V. G. (ed.) *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body, Motion and Behavior. HCII 2021. Lecture Notes in Computer Science, vol 12777*. Springer, Cham. https://doi.org/10.1007/978-3-030-77817-0_20

Pouw, W., Dingemanse, M., Motamedi, Y., & Özyürek, A. (2021). A systematic investigation of gesture kinematics in evolving manual languages in the lab. *Cognitive science, 45(7)*, e13014. https://doi.org/10.1111/cogs.13014

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Vesel´y, K., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., and Stemmer, G. (2011). The kaldi speech recognition toolkit. In *ASRU*.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.Rproject.org/.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rahimi, Z., Kumar, A., Litman, D. J., Paletz, S., & Yu, M. (2017). Entrainment in Multi-Party Spoken Dialogues at Multiple Linguistic Levels. In *Interspeech* (pp. 1696-1700).

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, *44*(11), e12911. https://doi.org/10.1111/cogs.12911

Rasenberg, M., Özyürek, A., Bögels, S., & Dingemanse, M. (2022). The primacy of multimodal alignment in converging on shared symbols for novel referents. *Discourse Processes*, *59*(3), 209-236. https://doi.org/10.1080/0163853X.2021.1992235

Rauchbauer, B., Nazarian, B., Bourhis, M., Ochs, M., Prévot, L., & Chaminade, T. (2019). Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society B*, *374*(1771), 20180033. https://doi.org/10.1098/rstb.2018.0033

Reichel, U. D., Beňuš, Š., & Mády, K. (2018). Entrainment profiles: Comparison by gender, role, and feature set. *Speech Communication, 100*, 46-57.

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive science, 35*(4), 587-637.

Reitter, D., & Moore, J. (2007). Predicting success in dialogue. in *Annual Meeting - Association for Computational Linguistics*, vol. 45, no. 1, p. 808.

Reitter, D., & Moore, J. D. (2014). Alignment and Task Success in Spoken Dialogue. *Journal of Memory and Language, 76*, 29–46. doi:10.1016/j.jml.2014.05.008

Ripperda, J., Drijvers, L., & Holler, J. (2020). Speeding up the detection of non-iconic and iconic gestures (SPUDNIG): A toolkit for the automatic detection of hand movements and gestures in video data. *Behavior research methods, 52(4)*, 1783-1794. https://doi.org/10.3758/s13428-020-01350-2

Roettger, T. B. (2021). Preregistration in experimental linguistics: Applications, challenges, and limitations. *Linguistics, 59*(5), 1227-1249.

Ruch, H. (2015). Vowel convergence and divergence between two Swiss German dialects. In: *18th International Congress of Phonetic Sciences, Glasgow* UK, 10 August 2015 - 14 August 2015, ICPHS.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735. https://doi.org/10.2307/412243

Salvesen, A. (2016). Investigating Linguistic Prestige in Scotland: An Acoustic Study of Accommodation between Speakers of Two Varieties of Scottish Standard English. *Lifespans and Styles, 2*(1), 35-47.

Schegloff, E. A. (2000). When "others" initiate repair. *Applied Linguistics*, *21*(2), 205–243. https://doi.org/10.1093/applin/21.2.205

Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis* (Vol. 1). Cambridge University Press.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382. https://doi.org/10.1353/lan.1977.0041

Schoffelen, J.-M., Oostenveld, R., Lam, N. H. L., Uddén, J., Hultén, A., & Hagoort, P. (2019). A 204-subject multimodal neuroimaging dataset to study language processing. *Scientific Data*, *6*(1), 17. https://doi.org/10.1038/s41597-019-0020-y

Schoot, L., Hagoort, P., & Segaert, K. (2019). Stronger syntactic alignment in the presence of an interlocutor. *Frontiers in Psychology, 10*, 685.

Schoot, L., Heyselaar, E., Hagoort, P., & Segaert, K. (2016). Does syntactic alignment effectively influence how speakers are perceived by their conversation partner?. *PloS one, 11*(4), e0153521.

Schweitzer, A., Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. In *INTERSPEECH* Lyon. 525-529.

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 326–332. https://doi.org/10.1037/0096-1523.29.2.326

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Stolk A., Verhagen L., & Toni, I. (2016). Conceptual alignment: How brains achieve mutual understanding. *Trends in Cognitive Sciences, 20(3)*, 180–191. https://doi.org/10.1016/j.tics.2015.11.007

Stolk, A., Bašnáková, J., & Toni, I. (2022). Joint epistemic engineering: The neglected process of context construction in human communication. In Ibanez, A., & Saravia, S.S. (eds.). *Routledge Handbook of Neurosemiotics*. https://psyarxiv.com/rwfe6/

Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, *1*(1), 113-150.

Taylor, J. R., Williams, N., Cusack, R., Auer, T., Shafto, M. A., Dixon, M., Tyler, L. K., Cam-CAN, & Henson, R. N. (2017). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, *144*, 262–269. https://doi.org/10.1016/j.neuroimage.2015.09.018

Torreira, F., Adda-Decker, M., & Ernestus, M. (2010). The Nijmegen corpus of casual French. *Speech Communication, 52*(3), 201-212. https://doi.org/10.1016/j.specom.2009.10.004

Troncoso-Ruiz, A., Ernestus, M. & Broersma, M. (2019). Learning to produce difficult L2 vowels: the effects of awareness-rasing, exposure and feedback. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019, 1094-1098. doi:10.5007/2175- 8026.2018v71n3p99.

Trujillo, J. P., Simanova, I., Bekkering, H., & Özyürek, A. (2018). Communicative intent modulates production and comprehension of actions and gestures: A Kinect study. *Cognition, 180*, 38-51. https://doi.org/10.1016/j.cognition.2018.04.003

Trujillo, J. P., Vaitonyte, J., Simanova, I., & Özyürek, A. (2019). Toward the markerless and automatic analysis of kinematic features: A toolkit for gesture and movement research. *Behavior research methods, 51(2)*, 769-777. https://doi.org/10.3758/s13428-018-1086-8

Uğurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P. F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D. A., Smith, S. M., Miller, K.L., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L.R., Behrens, T.E.J., Glasser, M.F., Van Essen, D.C., Yacoub, E. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage*, *80*, 80–104. https://doi.org/10.1016/j.neuroimage.2013.05.012

Van de Velde, H., Van Hout, R., & Gerritsen, M. (1997). Watching Dutch change: A real time study of variation and change in standard Dutch pronunciation. *Journal of Sociolinguistics, 1*(3), 361-391.

Van der Harst, S., & Van de Velde, H. (2008). 17 g's in het Standaardnederlands?. *Taal en tongval, 59*, 172-195.

Van der Harst, S., Van de Velde, H. & Schouten, M. E. H. (2007). Acoustic characteristics of Standard Dutch /ɣ/. In *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken. 1469-1472.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Van Son, R., Wesseling, W., Sanders, E., & van den Heuvel, H. (2008). The IFADV Corpus: a Free Dialog Video Corpus. In *LREC* (pp. 501-508).

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially-mediated syntactic alignment. *Language Variation and Change, 26*(3), 387-420.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of memory and language, 31*(2), 183-194.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *5th international conference on language resources and evaluation (LREC 2006)* (pp. 1556-1559). http://hdl.handle.net/11858/00-001M-0000-0013-1E7E-4

Wynn, C. J., & Borrie, S. A. (2022). Classifying conversational entrainment of speech behavior: An expanded framework and review. *Journal of Phonetics, 94*, 101173.

Yngve, V. H. (1970). On getting a word in edgewise. In M. A. Campbell (Ed.), *Papers from the sixth regional meeting, Chicago Linguistics Society* (pp. 567–578). Department of Linguistics, University of Chicago.

Yu, A. C., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PloS one, 8*(9), e74746.

Zeman, D., Popel, M., Straka, M., Hajič, J., Nivre, J., Ginter, F., Luotolahti, J., Pyysalo, S., Petrov, S., Potthast, M., Tyers, F., Badmaeva, E., Gokirmak, M., Nedoluzhko, A., Cinková, S., Hajič jr., J., Hlaváčová, J., Kettnerová, V., Urešová, Z., Kanerva, J., Ojala, S., Missila, A., Manning, C. D., Schuster, S., Reddy, S., Taji, D., Habash, N., Leung, H., de Marneffe, M.-C., Sanguinetti, M., Simi, M., Kanayama, H., dePaiva, V., Droganova, K., Mart´ınez Alonso, H., C¸ oltekin, c., Sulubacak, U., Uszkoreit, H., Macketanz, V., Burchardt, A., Harris, K., Marheinecke, K., Rehm, G., Kayadelen, T., Attia, M., Elkahky, A., Yu, Z., Pitler, E., Lertpradit, S., Mandl, M., Kirchner, J., Alcalde, H. F., Strnadova, J., Banerjee, E., Manurung, R., Stella, A., Shimada, A., Kwak, S., Mendonca, G., Lando, T., Nitisaroj, R., & Li, J. (2017). CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 1–19. https://doi.org/10.18653/v1/K17-3001

Zwiers, M. P., Moia, S., & Oostenveld, R. (2021). BIDScoin: A user-friendly application to convert source data to the Brain Imaging Data Structure. *Frontiers in Neuroinformatics*, 65. https://doi.org/10.3389/fninf.2021.770608

Research Data Management

Personal data

Will you process personal data? If yes, how will you ensure compliance with legislation on privacy?

Yes, I processed personal data in two different datasets. Dataset 1, the SCITI (Sentence Completion in Interaction with Two Interlocutors) dataset, consists of audio data, related F0 and articulation rate data, and questionnaire data about the experiment and demographic data. Dataset 2, the CABB (Communicative Alignment in Brain and Behaviour) dataset, consists of audio data, video data, kinematics, (f)MRI data, age and gender information, and questionnaire data. Data will be retained for at least 10 years in the Radboud Data Repository and the Donders Repository.

It was necessary to collect this data to address my research questions and to provide datasets for future research. The audio data and questionnaire data were used for analyses in multiple chapters. The age and gender information and some of the questionnaire data were necessary to ensure a homogeneous sample of participants.

In order to protect the privacy of your participants, will you anonymise or pseudonymise the data?

The privacy of participants will be protected because I (pseudo)-anonymised the data. Data has been pseudonymised by giving each participant a participant number upon starting the experiment, for both datasets. There are no documents linking sensitive information to the participant numbers. Forms with participants' names are safely stored either at the Donders Centre for Cognitive Neuroimaging (CABB dataset) or at the Erasmus building (SCITI dataset), both at the Radboud University.

Audio and video data cannot be anonymised, but participants have given approval for the use of their data by other researchers and/or for educational purposes. Only data of participants who have agreed to share their data will be shared.

Do you need approval from an ethics committee for your project?

Yes, I needed approval from the ethics committee for my project. This approval was given to me by the Ethics Committee of the Faculty of Arts on 10-04-2018 with reference number 6237.

Does your research require an informed consent procedure?

Yes, my research required an informed consent procedure. I have followed the standard informed consent procedure as specified by the Centre for Language Studies lab and the Donders Centre for Cognitive Neuroimaging.

Storing and sharing during research

Will you make use of safe storage during your research, including back-up facilities?
Safe storage has been used during my research. For the SCITI dataset, files are stored at the university server in workgroup folders on the university's network drive. These systems are backed-up by the university regularly. When working from home, I made use of the VPN connection to work with sensitive data. Files for the CABB dataset were stored on a P-drive at the Donders, and are now stored in a Research Documentation Collection in the Donders Repository.

With whom will you share your data during research?
During research, the SCITI dataset was shared with my supervisors and student assistants working on the data. The CABB dataset has been shared with the rest of the Communicative Alignment in Brain and Behaviour team. In the near future, other researchers will also be able to use both datasets. Only the personal data of participants who agreed to sharing of their data, will be shared with other researchers.

How will you deal with security issues that arise during your research?
Data for the SCITI dataset are stored in workgroup folders at Radboud University, and were only shared via the workgroup folders. When collecting these data, they were temporarily stored on SD cards, and transferred to the folders after testing. The CABB dataset is stored in the Donders Repository. During testing, part of these data were also stored on SD cards, but transferred to a safe location (the P drive) immediately after testing. Using these two storage locations is conform the policy of the Radboud University. During research, whenever I needed to access personal data, I used the VPN to access the folders. The data are backed up daily.

I organise my project's folder according to the following format:
The structure of both datasets meets the requirements of the Research Data Management of my institute. It includes readme files to further clarify the contents where needed.

Long term archiving and reuse

In the context of scientific integrity, where will you archive your data (including raw data, metadata and documentation) for at least 10 years?
The full SCITI dataset will soon be archived in a data sharing collection in the Radboud Data Repository. F0 and articulation rate data from Chapters 3 and 4 is already available (https://doi.org/10.34973/ka6j-r180). The CABB dataset is stored in a Research Documentation Collection in the Donders Repository (https://data.donders.ru.nl/), and shared.

In the context of data reuse, will you make your research data publicly available?
Yes, I will make my research data publicly available, as far as possible. Data from participants who have consented to sharing their data will be shared with researchers. The SCITI dataset will be published as soon as publication of the relevant papers is finished. The CABB dataset is available for use by other researchers. How to get access to this dataset is explained in Chapter 6: "In short, to be able to access and download the data, two steps are required. First, you need to create a user profile in the Donders Repository by logging in with your SURFconext or ORCID account (https://data.donders.ru.nl/login). For more information about the ORCID option and alternative ways to login, see: https://data.donders.ru.nl/doc/help/helppages/user-manual/login-profile.html?8. Second, the Data Use Agreement needs to be completed and sent to ivan.toni@donders.ru.nl. Upon completion of these steps, users will be granted access to the collection and can view and download files through the Donders Repository website (for more information, see: https://data.donders.ru.nl/doc/help/user-manual/transfer-data.html)."

How will you ensure that your research data will be stored in a FAIR manner?
The SCITI dataset will be stored in the Radboud Data Repository (see above), and the CABB dataset is stored in the Donders Repository (see Chapter 6). Data will be Findable via the data sharing collections in the repositories. Moreover, papers on the datasets will be published to make them better findable. Datasets will be Accessible via the repositories. For the CABB dataset, researchers will need to sign an agreement to ensure the safe use of the data. My data will be Interoperable, because they include metadata, as well as readme files, and standard data formats. The two datasets will be Reusable by other researcher who ask for permission to use the datasets.

English summary

Conversation is one of the primary situations in which people use language. During conversations, people's behaviours tend to become more similar in various ways. For instance, if two speakers use different words to refer to the same object, this may change during conversation. In a conversation, one speaker can use the word *couch*, while the other person uses the word *sofa* to refer to the same object. The speaker using the word *couch* may change their way of referring to the object by calling it a sofa, to match it to the word used by the other speaker. Another example is that speakers tend to become more similar in how fast they speak.

These kinds of changes to each other's speech are called *alignment*. Alignment can occur at different linguistic levels, including the syntactic, prosodic, and segmental phonetic levels. Alignment has been shown to occur at these different levels but researchers rarely investigate more than one level when formulating and testing alignment theories. This makes it challenging to draw conclusions about potential relationships between alignment at different linguistic levels and about possible underlying mechanisms of alignment.

Researchers have proposed different theories to explain the underlying mechanisms of alignment. Pickering and Garrod (2004) propose that the underlying mechanism is automatic priming, where mental representations become more active (e.g., a syntactic structure) when a speaker is exposed to them while they are used by another speaker. It is then more probable that for example this syntactic structure will be used again. An opposing theory by Clark (1996) proposes that alignment is a joint action, where speakers in a conversation actively align. Ostrand and Ferreira (2019) extend these theories by hypothesising about the possibility that a speaker could learn their interlocutor's behaviour and align to it, while then changing their behaviour when conversing with another speaker. The authors say alignment is only specific to an interlocutor when it helps to achieve the goal of the conversation. Next to these theories, Giles and colleagues (1991) have proposed that alignment depends on the social situation, where speakers align more when they want to be liked by their interlocutor or when they want to belong to a certain group.

Alignment can occur on different time scales, both locally and globally. Local alignment means that speakers align their behaviour to what they have heard most recently, usually in the previous speaking turn. Global alignment is alignment that happens over a longer period of time, where speakers' behaviour becomes more similar over time. Studying these different time scales can inform us about the underlying mechanisms of alignment.

The goal of this dissertation is to contribute to the knowledge on linguistic alignment and to better understand its underlying mechanisms. In order to do so, we

studied alignment on different linguistic levels. In addition, we investigated both local and global alignment. Furthermore, this dissertation also investigated to what extent alignment could be interlocutor-specific. These topics were investigated in Chapters 2 to 5. These chapters all concern analyses of different aspects of the same dataset. Chapter 6 concerns the creation of another dataset that is suitable for alignment research.

Chapter 2 focussed on syntactic alignment. We investigated this in a sentence completion task in which speakers interacted with pre-recorded speech of two different interlocutors who differ in their use of a Dutch syntactic alternation (auxiliary-participle versus participle-auxiliary; .e.g *heb gehad* versus *gehad heb*). Next to a main experiment, in which speakers were presented with the syntactic alternations by the interlocutors, we conducted a control experiment. In this control experiment, speakers were not presented with the pertinent syntactic structure, but they did still finish sentences that elicited the use of the syntactic alternation - like in the main experiment. We found that speakers aligned to the interlocutors in the main experiment, while, as predicted, the participants in the control experiment did not. We found evidence of local alignment, and no evidence of alignment specific to an interlocutor.

Chapter 3 investigated alignment on the prosodic level, focussing on alignment in pitch and articulation rate, using a subset of the data from the main experiment presented in Chapter 2. We found indications of global alignment and not of local alignment. Alignment to the prosodic features under study thus cannot solely be due to local priming.

Chapter 4 expanded on Chapter 3 by investigating the same questions in control data. These control data are a subset of the control data presented in Chapter 2. Next to not being presented with the syntactic structure relevant in Chapter 2, participants did not receive any auditory input from the interlocutors. This enabled us to investigate a similar situation as in Chapter 3, but now without any pitch or articulation rate input from the interlocutors. The effects present in Chapter 3 were absent in the control data. This confirms the findings in Chapter 3, that the effects that we found are due to alignment and not to other potential effects.

Chapter 5 studied alignment on the segmental phonetic level. It aimed at better understanding phonetic alignment of regional variants, by investigating alignment to different regional variants that differ in prestige. This was studied in local measures and more global measures. This chapter focussed on regional variants of a Dutch phoneme, the so-called 'hard g' versus the 'soft g' (for example the <g> in the Dutch word <goed> can be pronounced as a hard or soft 'g'). Next to using different syntactic structures described in Chapter 2, the two interlocutors also differed in the variant of the 'g' they used. At the group level, we found no evidence of speakers aligning to the interlocutors, neither to the more prestigious variant (hard 'g') nor to the less prestigious one (soft 'g'), and neither locally nor globally. However, after closer inspection of the data, we found large individual variation.

Chapter 6 presents a new dataset created within the CABB team (Communicative Alignment in Brain and Behaviour team, a research group within the Language in Interaction consortium). This dataset was designed to investigate alignment at different linguistic levels during an interaction, as well as for neural and behavioural analyses of pre- and post-tasks. The experiment was designed around seeing or describing certain objects. This dataset is very well suited to investigate linguistic alignment in an ecologically valid setting, where pairs of speakers interact in a task-based conversation.

The results discussed in this dissertation taken together indicate that alignment is a very complex phenomenon, more so than may be reflected in the literature. Our results indicate that alignment cannot be explained by one single mechanism for all different linguistic levels and features. A combination of different mechanisms should thus underlie the complex phenomenon, which may vary depending on the linguistic level, feature, task context, and individual differences between speakers. Moreover, this dissertation showed that control data and baseline measures are important features of any alignment experiment, and should be standard in future alignment experiments to ensure findings related to theories. Lastly, the two datasets presented in this dissertation are very well suited to further investigate the theories proposed at all linguistic levels, and to be connected to other behavioural and neural data presented in Chapter 6. This dissertation thus contributed new information on linguistic alignment at different linguistic levels – the syntactic, prosodic, and segmental phonetic levels – and has shown that evidence for this phenomenon varies within and across levels, as well as between individuals.

Nederlandse samenvatting

Mensen gebruiken taal meestal in conversaties. Wanneer mensen een gesprek met elkaar voeren, gaat hun spraak steeds meer op elkaar lijken. Dit gebeurt op verschillende manieren, bijvoorbeeld in woordgebruik. Wanneer de ene spreker het woord *friet* gebruikt, terwijl de andere persoon het woord *patat* gebruikt, dan kan dit tijdens het gesprek veranderen. De spreker die het woord *friet* gebruikt, kan de manier van refereren hiernaar veranderen door het ook patat te noemen, om het zo overeen te laten stemmen met het woord dat door de andere spreker wordt gebruikt. Een ander voorbeeld is dat sprekers hun spreeksnelheid aan elkaar aanpassen.

Dit soort aanpassingen aan elkaars manier van spreken wordt *alignment* genoemd. Alignment kan op verschillende taalkundige niveaus optreden, waaronder het syntactische, prosodische, en segmenteel fonetische niveau. Onderzoekers hebben alignmenteffecten gevonden op deze verschillende niveaus, maar bestuderen vaak slechts één niveau om theorieën te formuleren en te testen. Dit maakt het moeilijk om conclusies te trekken over de potentiële relatie tussen alignment op verschillende taalkundige niveaus en over de onderliggende mechanismen van alignment.

Onderzoekers hebben verschillende theorieën voorgesteld om alignment te verklaren. Pickering en Garrod (2004) stellen dat het onderliggende mechanisme automatische priming is, waarbij mentale representaties (bijvoorbeeld van een syntactische woordvolgorde) actiever worden wanneer een spreker ze net gehoord heeft van een andere spreker, wat het waarschijnlijker maakt dat bijvoorbeeld deze syntactische volgorde opnieuw gebruikt wordt. Een andere theorie van Clark (1996) stelt dat alignment een *joint action* is, waarbij sprekers in een gesprek actief alignen. Ostrand en Ferreira (2019) breiden deze theorieën uit door in te gaan op de mogelijkheid dat een spreker kan leren welk gedrag de gesprekspartner gebruikt en zich hieraan zou kunnen aanpassen, en dit gedrag vrijwel meteen kan loslaten in conversatie met een andere gesprekspartner. Zij stellen voor dat alignment alleen specifiek is aan de gesprekspartner wanneer dit helpt om het doel van het gesprek te bereiken. Naast deze theorieën hebben Giles en collega's (1991) voorgesteld dat alignment afhankelijk is van de sociale situatie, waarbij sprekers zich meer aanpassen wanneer ze graag willen dat hun gesprekspartner ze aardig vindt of wanneer ze bij een bepaalde groep willen horen.

Alignment kan over verschillende tijdsspannen plaatsvinden, zowel lokaal als globaal. Lokaal betekent dat sprekers hun gedrag afstemmen op wat ze het meest recent hebben gehoord, meestal in de vorige beurt. Bij globale alignment vindt alignment over een langere periode plaats, waarbij het gedrag van sprekers over de tijd steeds meer op elkaar gaat lijken. Het bestuderen van deze verschillende tijdsspannen kan helpen in het onderzoek naar de onderliggende mechanismen van alignment.

Het doel van dit proefschrift is om bij te dragen aan de kennis over alignment en meer inzicht te krijgen in de onderliggende mechanismen. Hiervoor hebben we alignment op verschillende taalkundige niveaus bestudeerd. Daarnaast hebben we zowel lokale als globale alignment onderzocht. Dit proefschrift onderzocht ook in welke mate alignment afhangt van de gesprekspartner. Deze onderwerpen zijn onderzocht in Hoofdstukken 2 tot 5. Deze hoofdstukken behandelen analyses over verschillende aspecten van dezelfde dataset. Hoofdstuk 6 betreft de creatie van een andere dataset, die erg geschikt is voor alignmentonderzoek.

In Hoofdstuk 2 is gekeken naar syntactische alignment. We hebben een experiment uitgevoerd waarin participanten een zin moesten afmaken na het horen van vooraf opgenomen spraak van twee verschillende gesprekspartners, die verschilden in hun gebruik van een Nederlandse syntactische woordvolgorde (hulpwerkwoord-voltooid deelwoord versus voltooid deelwoord-hulpwerkwoord; bijvoorbeeld 'heb gehad' versus 'gehad heb'). Naast een hoofdexperiment waarin participanten werden blootgesteld aan deze syntactische volgorde van de gesprekspartners, hebben we ook een controle-experiment uitgevoerd. In dit controle-experiment kregen participanten de betreffende syntactische structuur niet te zien of te horen, maar werd deze wel net als in het hoofdexperiment uitgelokt in zinnen die participanten afmaakten. We vonden dat participanten alignden met de gesprekspartners in het hoofdexperiment, terwijl, zoals voorspeld, de participanten in het controle-experiment dat niet deden. We vonden evidentie voor lokale alignment, en geen evidentie voor alignment aan een specifieke gesprekspartner.

In Hoofdstuk 3 is gekeken naar alignment op het prosodische niveau, door de toonhoogte en articulatiesnelheid te onderzoeken. Deze studie onderzocht een subset van de data uit het hoofdexperiment dat in Hoofdstuk 2 werd gepresenteerd. We vonden een indicatie van globale alignment en niet van lokale alignment. Alignment in deze twee prosodische maten kan dus niet alleen verklaard worden door lokale priming.

Hoofdstuk 4 breidde Hoofdstuk 3 uit door dezelfde vragen te onderzoeken, maar nu in de controledata. Deze controledata zijn een subset van de controledata die in Hoofdstuk 2 werden gepresenteerd. Naast het niet blootgesteld worden aan de voor Hoofdstuk 2 relevante syntactische structuur, ontvingen de participanten geen auditieve input van de gesprekspartners. Hierdoor konden we een vergelijkbare situatie onderzoeken als in Hoofdstuk 3, maar nu zonder enige toonhoogte of articulatiesnelheid-input van de gesprekspartners. De effecten die in Hoofdstuk 3 aanwezig waren, waren afwezig in de controledata. Dit bevestigt de bevindingen in Hoofdstuk 3, namelijk dat de effecten die we vonden alignmenteffecten zijn en geen potentiële andere effecten.

Hoofdstuk 5 richtte zich op regionale varianten van een Nederlands foneem, de zogenaamde 'harde g' versus de 'zachte g' (de <g> in <goed> kan bijvoorbeeld worden uitgesproken als harde of zachte 'g'). Naast het gebruik van verschillende syntactische woordvolgordes zoals besproken in Hoofdstuk 2, verschilden de gesprekspartners

in welke variant van de 'g' ze gebruikten. We vonden geen evidentie dat sprekers zich aanpassen aan de gesprekspartners op groepsniveau, noch aan de meer prestigieuze variant (de harde 'g'), noch aan de minder prestigieuze variant (de zachte 'g'), noch lokaal, noch globaal. Nadere inspectie van de data liet grote individuele variatie zien.

Hoofdstuk 6 presenteert een nieuwe dataset die binnen het CABB team (Communicative Alignment in Brain and Behaviour team, een onderzoeksgroep binnen het Language in Interaction consortium) is verzameld. Deze dataset is bedoeld om alignment op verschillende taalkundige niveaus tijdens een interactie te onderzoeken, en te kijken naar neurale- en gedragsmaten in taken voorafgaand en volgend op de interactie. Het experiment was ontworpen rondom het bekijken of beschrijven van bepaalde objecten. Deze dataset is uiterst geschikt om alignment in een ecologisch valide setting te onderzoeken, waar paren sprekers een taakgerichte conversatie voeren.

Alle resultaten samen laten zien dat alignment een zeer complex fenomeen is, meer dan wellicht in de literatuur wordt weergegeven. Onze resultaten suggereren dat alignment niet door één enkel mechanisme voor alle verschillende taalkundige niveaus en maten kan worden verklaard. Een combinatie van verschillende mechanismen is dus nodig. Naast variatie in deze mechanismen afhankelijk van het taalkundige niveau en maten, kunnen ook de taakcontext en interindividuele variatie tussen sprekers invloed hebben. Bovendien heeft dit proefschrift aangetoond dat basismetingen, zoals die in het controle-experiment en in de voormetingen, belangrijke gegevens leveren voor de interpretatie van elk alignmentexperiment. Controle-experimenten en voormetingen zouden dus in de toekomst de standaard moeten vormen in alignmentexperimenten om duidelijkere conclusies te kunnen trekken over theorieën. Ten slotte zijn de twee datasets die in dit proefschrift zijn gepresenteerd zeer geschikt om verder onderzoek te doen naar de voorgestelde theorieën op alle taalkundige niveaus, en om te worden gekoppeld aan andere gedrags- en neurale maten in de dataset gepresenteerd in Hoofdstuk 6. Kortom, dit proefschrift heeft nieuwe informatie geleverd over alignment op verschillende taalkundige niveaus - de syntactische, prosodische en segmentele fonetische niveaus - en heeft aangetoond dat bewijs voor dit fenomeen varieert binnen en tussen niveaus, en tussen individuen.

Acknowledgements

Writing these acknowledgements may be the hardest part of writing this thesis (and not just spelling the word). There are so many people who have helped me throughout these years. Unfortunately, I do not have a fancy metaphor for my PhD journey, so I will just say it has been a crazy but amazing time that would not have been possible without the people in these acknowledgements.

First and foremost, I would like to thank the best supervisory team a PhD candidate could wish for: **Mirjam Ernestus** and **Herbert Schriefers**. I believe we were truly a great team with meetings (always noted down as 'Meeting Mirbert' in my calendar) in which I never failed to have a good laugh. Together we always managed to quickly find answers to any question.

Mirjam, ik weet niet waar ik moet beginnen. Ik heb enorm veel geleerd van al onze meetings; van onderzoek en statistiek tot oud-Hollandse zwemslagen. Als ik me ergens zorgen over maakte, kon ik altijd even op je deur kloppen en voelde ik me al snel beter. Je hebt altijd een antwoord op al mijn vragen of weet precies naar wie je me door kunt verwijzen. We kunnen het overal over hebben. Bedankt voor de steun en begeleiding.

Herbert, ook van jou heb ik ongelooflijk veel geleerd. Wat ik vooral nooit meer zal vergeten is dat descriptieve data mijn beste vriend zijn. Ik ben erg blij dat ik heb mogen profiteren van je oneindige kennis over de psychologie en psycholinguïstiek in het algemeen, en natuurlijk al je kennis over het onderwerp van dit proefschrift. Ik waardeer het enorm dat je altijd enthousiast bent over mijn ideeën, ook al had ik er uiteindelijk vaak geen tijd voor. Zelfs al hebben we nog nooit Duits gepraat, Tausend Dank.

These short paragraphs are not enough to express my thanks to you both. You have made my PhD journey into the amazing experience it was. Unfortunately, I have never gotten that promised selfie of the three of us, so I had to design a different cover…

Besides my supervisors, I would like to thank my manuscript committee for reading this manuscript. **Marc van Oostendorp**, **Esther Janse**, **Rob van Son**, **Robert Hartsuiker**, and **Katharina Spalek**, thank you very much for your time and incredibly useful comments.

I would also like to thank the **Language in Interaction Consortium** for providing me with amazing opportunities. Being part of a large group of researchers who are all interested in language has given me a very broad knowledge in this field. Regularly meeting with PhD candidates from different fields also helped me see the broader picture of my research. I would especially like to thank the **CABB team** for the great team

science, the inspiring meetings and for the fun dinners. **Asli**, **Branka**, **Flavia**, **Herbert**, **Iris**, **Ivan**, **James**, **Judith**, **Laura**, **Marieke**, **Mark B.**, **Mark D.**, **Marlou**, **Mirjam**, **Sara**, **Wim**, and others who helped the team, this thesis would not have existed without you. A special thanks to **Sara** and **Marlou**; setting up the experiment and testing participants was a massive but fun undertaking.

Many thanks to the **Centre for Language Studies** and the **Speech Production and Comprehension group** in particular, who were always there to practice talks and discuss complex statistics – **Amalia**, **Annika**, **Aurora**, **Chen**, **Cong**, **Emily**, **Esther**, **Hanno**, **Joe**, **Katherine, Katie**, **Lieke**, **Lisa**, **Lou**, **Louis**, **Martijn**, **Mirjam**, **Robert**, **Stella**, **Tim**, **Yachan**, and **Xinyu**. A special thanks to **Louis ten Bosch** for always helping me with any statistics or forced alignment questions. I could always stick my head around the corner to bother you with any technical question. And to **Esther Janse**, who inspired me to do a PhD by making sure I had the very best time doing my Master's thesis research in New Zealand. Thank you for always having great advice.

I would also like to thank **Margret van Beuningen** and **Bob Rosbag** for their help in the CLS lab, and **Uriel Plönes** at the DCC, for making sure the quality of my recordings was always up to standard. **Henk van den Heuvel**, thank you for your help with any GDPR questions. Without the help of many research and student assistants it would not have been possible to do the research in this thesis. Thanks to **Anouck**, **Daniëlle**, **Flavia**, **Inez**, **Ivy**, **Joost**, **Lisette**, **Marein**, **Rosa**, **Sebastian**, **Tessa**, **Tessel**, and **Yvonne** for their help. Many participants have taken part in the sometimes long experiments in this dissertation. I would like to thank them for their time and interest.

Being part of both the **Graduate School for Humanities** and the **International Max Planck Research School** has helped me tremendously both socially and academically. I followed many courses in which I learned a great deal. **Peter van der Heiden**, thank you for organising GSH-related events. **Kevin Lam**, thank you for organising so many useful courses and fun activities.

At the end of my PhD, I had the opportunity to visit the **Human Interaction Lab** at Utah State University. **Stephanie Borrie**, thank you for the warm welcome to your lab, even though it was a brief stay. I thoroughly enjoyed our conversations about conversations and your enthusiasm about entrainment research. **Annalise Fletcher**, it was great seeing you again and spending time talking about research and life. **Camille Wynn** and **Samantha Budge**, thank you for taking me to lunches all over Logan and sharing a love for socks. Thank you all for making my stay in Utah a memorable one.

I would also like to express my gratitude to my new colleagues in York. **Sophie Meekings**, your celebratory baking creations helped me get through the very last part of my PhD. Thank you to the **Speech lab** - **Sven**, **Sophie**, **Sarah**, **Alex** and **Emily** - for welcoming me to York and to the lab.

My dear paranymphs, **Aurora** and **Katherine** (in alphabetic order). You are the best.

Just kidding, I have more to say. You both helped me with everything from supporting me from the very start in E8.11 in a writing session to the title of my thesis (thank you for preventing me from calling it "To align or not to align: Depends on the weather"). You were there for me from day one until the very last minute of my PhD. We had so much fun at ICPhS in Melbourne. I could not have done without you these past years, and I am sure I cannot do without you in the future. **Aurora**, we do not always agree on theories, but I am happy we agree on how to have fun. PhD life was never boring with you, from Kinder eggs, to painting dinosaurs, to bubble helicopters, to chair races, to office tosti's, to Nerf guns, to your great taste in hiking shoes, the list goes on and continues to go on. **Katherine**, apart from being my go-to native English speaker, co-Tasmania-driver, and plant and cheese lover, I can also talk to you about everything and anything. I really appreciate that you always listen even though we both also love to talk. I am very happy you both agreed to be my paranymphs, making this the perfect end to a great period.

Next to my paranymphs, I was lucky to have many other colleagues who became good friends. **Chen**, thank you for being the amazing person you are. I hope to spend many more afternoons talking about the UK, research, and life in general while drinking your great tea blends and cuddling Camilia. **Wei**, I am very happy you were in the office next to me. I loved peeking around the corner and seeing your happy face. Even though you live in Germany now, I am happy we still talk a lot. **Elly** en **Thijs**, Fellows en Gamer, bedankt dat jullie altijd maar een appje weg zijn. Elly, je was een Excellerende bijna-kantoorbuur. Kletsen kunnen we altijd erg goed, misschien iets té goed. Bedankt dat ik je stem mocht gebruiken in mijn experimenten. Thijs, je bent altijd bereid te helpen. Bedankt voor je altijd goede grappen en scherpe comments, zelfs als het over fonetiek gaat. **Saskia**, bedankt voor de regendansjes, drankjes en duikjes! Ik ben erg blij dat onze PhDs overlapten en we over de meest random dingen kunnen praten. My two great X-office mates, **Xiaoru** and **Xinyu**, I loved sharing an office with you! Xiaoru, it was always fun being in the office with you. Next to talking about research, you could also tell me everything about any possible hobby. Xinyu, even though we were office mates for a relatively short time, it was a great time. I am happy we found a way to continue our Tuesday chats over text. **Emily**, **Hannah**, **Rehana**, and **Tim** also deserve credit, I always enjoy chatting with you. Days at the office would not have been the same without my colleagues on the 8th and 9th floor of the great tower, and at the MPI. I would like to thank **Aurora**, **Annika**, **Aurélia**, **Candice**, **Chantal**, **Chen**, **Claire**, **Elly**, **Emily**, **Ferdy**, **Figen**, **Gert-Jan**, **Hannah**, **Hanno**, **Imke**, **Katherine**, **Lisa**, **Maria**, **Patricia**, **Saskia**, **Rehana**, **Tashi**, **Theresa**, **Thijs**, **Tim**, **Wei**, **Xiaoru**, **Xinyu**, **Xing**, and **Yu** - spending coffee breaks and lunches with you was amazing.

Naast alle collega's en vrienden die ik heb ontmoet dankzij mijn PhD, zijn er andere mensen die mijn vrije tijd altijd leuker maken. Dit zorgde ervoor dat ik altijd afleiding had van mijn PhD wanneer nodig, en ook soms wanneer niet nodig.

Een deel van deze tijd besteed ik nog altijd in Venlo. **Anke**, **Chantal**, **Claire**, **Evelien**, **Juul**, **Lise**, **Loes**, **Merle**, ik ken jullie al eeuwen en daar ben ik enorm blij om. Bedankt dat jullie jullie zijn en we altijd leuke weekendjes, avondjes en vroege ochtenden hebben, die me soms iets te veel mijn PhD hielpen vergeten. Het is altijd fijn met jullie en **Ad**, **Jaap**, **Marc**, **Max** en **Ricky**! Anke en Evelien, bedankt dat ik uren naar jullie foto's mocht kijken tijdens het testen van participanten. **Shanou**, mijn beste dubbelbuddy, ook zonder tennissen is het altijd top. **Lotte**, naast onze prachtige naam, hebben we genoeg andere dingen om over te kletsen.

Om mezelf aan zo veel mogelijk verschillende /x/s bloot te stellen, reis ik ook graag het land door. **Ineke**, van Italiaans lessen tot treinritjes tot uiteten, we kunnen altijd uren praten over alles en niks (of tot we in Venray waren en je blij was dat je in stilte naar je auto kon). **Elske**, zonder jouw hulp en goede verstopskills in de woonkamer was ik überhaupt nooit aan deze PhD begonnen. Bedankt voor al je hulp en hopelijk kunnen we ooit nog een keer samen in Nieuw-Zeeland zijn. **Berthe**, **Dani**, en **Tessa**, bedankt voor alle keren dat ik keihard alle stress eruit heb kunnen zingen en dansen tijdens concerten.

Na vermoeiende tripjes met veel /x/s, kom ik altijd graag thuis in Nijmegen. **Jorinde**, allerbeste vergane Fagel, wat moet het vermoeiend zijn geweest om continu iemand te hebben die op je deur klopt (gelukkig kon ik je verblijden met mijn snackla). Zonder jou was het begin van mijn PhD zéker niet hetzelfde geweest. **Rosa**, naast een geweldige student-assistent en tennisster, ben je een overall geweldig persoon! Bedankt dat je me iedere week als blije Golden Retriever met buikspierpijn van het lachen over de tennisbaan hebt laten rennen. **Charlotte**, **Juan**, en natuurlijk **Puck**, jullie zijn de beste buren die er bestaan. Bedankt dat ik altijd aan kan bellen om uren op de gang een bal over te gooien, een lekker kopje koffie te drinken of de beste burger op de wereld te eten. **Sofie**, de frequente padelsessies, uitjes, cappuccino's, koude vakantie, jaarlijkse diners, outfits en uitgebreide verkleedsessies hebben mij en mijn proefschrift goed geholpen. Ik ben ook erg blij met het design van de voorkant, dankjewel! Mijn algemene obsessie met sporten, maar vooral mijn recente padelobsessie heeft me enorm geholpen om mijn proefschrift af te krijgen. **Sofie**, **Joost** en **Chris**, bedankt dat jullie altijd zin hebben om me van de baan te slaan.

And of course my newest home in York, without /x/s. Thank you, **Ben**, for being an amazing punny flatmate, making home feel like home, and of course for helping me make all the last decisions about my dissertation.

Arjan, Gerda en **Tessa**, bedankt voor alle weekendjes weg, leuke vakanties en lekkere diners de afgelopen jaren. **Papa**, **mama**, **Bente**, **Rob**, **Ollie**, **Bart**, **Karlijn**, en **Spikey**, jullie verdienen ook veel dank. Papa en mama, zonder jullie had ik niet bestaan. Jullie staan altijd voor me klaar als het nodig is en zelfs als het niet nodig is. Bente, je bent een schatje en mijn geweldige grote kleine zus. Bedankt voor je altijd luisterende oor en mooie stem die ik mocht lenen voor mijn experimenten. Bart, je bent altijd mijn lat geweest en de beste broer die er bestaat. Ooit zal ik misschien ook een Pythontovenaar zijn. En natuurlijk Spikey, mijn allergrootste liefde.

Lieve **Thom**, na alle e-mails die je altijd mocht proeflezen, zal dit het laatste van mijn PhD zijn wat je hoeft te lezen. Uren heb je geluisterd naar mijn verhalen de afgelopen jaren. Ook al zijn fonetiek en syntaxis nog steeds niet de meest bekende woorden in je lexicon, je schreeuwt tegenwoordig wel op random momenten ALIGNMENT! Bedankt dat je er altijd voor me bent.

Curriculum Vitae

Lotte Eijk was born in Amersfoort, the Netherlands, in 1995. In 2015, she finished her Bachelor's in French Language and Culture at the Radboud University, Nijmegen. During her Bachelor's she studied at Paris-Sorbonne IV with the Erasmus programme. Following her Bachelor's, she obtained a Master's in Speech and Language Pathology, and in French Linguistics (cum laude), both at the Radboud University. For her Master's thesis Speech and Language Pathology, she did research at the New Zealand Institute of Language, Brain and Behaviour in Christchurch, New Zealand, where she was also a teaching assistant/guest lecturer. In 2017, she started her PhD within the Language in Interaction consortium. Her research was mainly conducted at the Centre for Language studies of the Radboud University. Here, she was part of the Graduate School of Humanities and the International Max Planck Research School for Language Sciences. During her PhD, she taught as a lecturer at Leiden University. Currently, she is a Postdoctoral Research Associate at the University of York in the United Kingdom.

Publications

Eijk, L., Ernestus, M., Schriefers, H. (2019). Alignment of Pitch and Articulation Rate. *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia (2019), pp. 2690-2694.

Eijk, L., Fletcher, A., McAuliffe, M., & Janse, E. (2020). The effects of word frequency and word probability on speech rhythm in dysarthria. *Journal of Speech, Language, and Hearing Research, 63*(9), 2833-2845. https://doi.org/10.1044/2020_JSLHR-19-00389

Eijk, L.*, Rasenberg, M.*, Arnese, F., Blokpoel, M., Dingemanse, M., Doeller, C. F., Ernestus. M., Holler, J., Milivojevic, B., Özyürek, A., Pouw, W., van Rooij, I., Schriefers, H., Toni, I., Trujillo, J., & Bögels, S. (2022). The CABB dataset: A multimodal corpus of communicative interactions for behavioural and neural analyses. *NeuroImage*, 119734. https://doi.org/10.1016/j.neuroimage.2022.119734

* Shared first author

MAX PLANCK INSTITUTE
FOR PSYCHOLINGUISTICS

VISITING ADDRESS
Wundtlaan 1
6525 XD Nijmegen
The Netherlands

POSTAL ADDRESS
P.O. Box 310
6500 AH Nijmegen
The Netherlands

CONTACT
T +31(0)24 3521 911
F +31(0)24 3521 213
E info@mpi.nl
Twitter @MPI_NL
www.mpi.nl

CLS | Centre for Language Studies
Radboud University