Chips or caviar, lazy or loyal?

Investigating neural correlates of accent-related stereotypes about Frisian speakers of Dutch

Jelle Brouwer s2719118
j.brouwer.29@student.rug.nl
University of Groningen
MA, Language and Cognition

Supervisor: Dr. Hanneke Loerts h.loerts@rug.nl
Second reader: Prof. Dr. Wander Lowie w.m.lowie@rug.nl

August 30, 2019

21,467 words
# Table of Contents

Acknowledgments .................................................................................................................................. 4  
Abstract ................................................................................................................................................... 5  
1. Introduction ........................................................................................................................................ 6  
2. Theoretical background ..................................................................................................................... 8  
   2.1. How are attitudes formed? ......................................................................................................... 8  
   2.2. Speech and expectations ........................................................................................................... 10  
      2.2.1. Accent and identity ............................................................................................................ 10  
      2.2.2. Status and solidarity .......................................................................................................... 13  
      2.2.3. Speaker evaluations of status and solidarity ................................................................... 14  
   2.3. The situation in the Netherlands ............................................................................................. 19  
   2.4. Separating implicit and explicit attitudes ............................................................................... 22  
      2.4.1. The implicit association task: the most common technique ............................................ 22  
      2.4.2. Methodological issues in measuring implicit attitudes ................................................... 23  
      2.4.3. ERPs: the ‘true’ implicit measure? .................................................................................... 25  
      2.4.4. The N400 ............................................................................................................................ 26  
      2.4.4. Speaker-message integration and the N400 .................................................................... 28  
   2.5. Relevance of the study .............................................................................................................. 31  
3. Present study .................................................................................................................................... 33  
   3.1 Research Questions and Hypotheses ........................................................................................ 35  
4. Method .............................................................................................................................................. 37  
   4.1. Participants ............................................................................................................................... 37  
   4.2. Materials .................................................................................................................................... 38  
      4.2.1. Stimuli ................................................................................................................................ 38  
      4.2.2. Stimuli validation .............................................................................................................. 40  
      4.2.3. Stimuli Recordings ............................................................................................................. 42  
      4.2.4 Questionnaires .................................................................................................................... 43  
      4.2.5. EEG lab equipment ............................................................................................................ 44  
      4.2.6. E-prime experiment ........................................................................................................... 45  
   4.3. Design ........................................................................................................................................ 45  
   4.4. Procedure ................................................................................................................................... 46  
   4.5. Analysis ..................................................................................................................................... 47  
      4.5.1. Preprocessing ..................................................................................................................... 47  
      4.5.2. Selecting a Region of Interest ........................................................................................... 48  
      4.5.3. Generalized Additive Mixed Models ................................................................................. 50  
      4.5.4 Explaining our first GAMM model .................................................................................... 52
Acknowledgments

There are a lot of people without whom this thesis would not have existed. Now that I am finished, I would like to express my sincerest gratitude to them.

First of all, Hanneke. Thank you for introducing me to this topic in the first place. While this has by far been the most challenging project I have been a part of during my academic career up until now, both in scale, duration and complexity, you always managed to make it fun even at the most difficult times. Thank you for always making time for short (and sometimes long) meetings even though you were extremely busy yourself. Without your supervision, your expertise in EEG research, and your mental support this project would not have been possible.

A big thank you to Peter Albronda is also in order. Without your constant care for the equipment, we would not have been able to test even half of the participants that signed up. In fact, without your training, I would have never been able to start with EEG research in the first place. I would also like to thank you for the excellent conversations we had while the participants were in the booth.

I would also like to thank Vincent for aiding us with the collection of our data, and for being probably one of the few people on this earth I can spend 10 hours with in a lab that has no sunlight or fresh air. Thank you for putting up with my sometimes terrible sense of humor. I hope my habit of chain-drinking espressos during testing didn’t rub off on you too much.

Thank you to Aziza. You have given me the kick up the ass that I needed when I didn’t know how to continue, you’ve been an inexhaustible source of mental support even at points when we couldn’t see each other because I was in the studio, the lab, or the library. You’ve always reminded me to stay positive.

I am also grateful for my parents and family for the mental support, and of course for helping out a little by filling out the stimuli validation questionnaires. You probably thought some of the stuff we were doing was crazy or plain weird, but you always supported me.

Also a big thank you to all my friends. While I was not able to hang out for long at times, you’ve always been there for a nice bit of distraction. Whether it was coffee, dinner, or simply watching terrible Netflix series together, I could always count on you.

And lastly, of course thank you to the participants and speakers, as well. You not only provided the data that made this study, but you provided us with plenty of funny banter that made every testing session feel like 15 minutes (although after listening to 300+ sentences, I doubt it felt the same for you).
Abstract

A common finding in the literature is that standard accents are associated with status, while regional accents are associated with social attractiveness. An issue with these studies, however, is that a questionnaire-based evaluation method is not informative regarding the neural processes that occur while these association are activated. The present study therefore compared the electrophysiological brain responses in subjects hearing a series of accent-congruent versus accent-incongruent statements; these statements were either made by a Frisian (regional) or Dutch (standard) speaker. It was hypothesized that accent-incongruent items would elicit a larger negativity than accent-congruent stimuli around 400ms (N400) post stimulus-onset, reflecting difficulties integrating speaker and message. Furthermore, in the analyses a comparison was made between Frisian and Dutch subjects. It was expected that participants who heard a speaker from their own group utter negative but accent-congruent stimuli, would show a larger N400 effect compared to subjects from the other group, due to not having the same implicit associations with this accent. Similarly, participants who heard positive but accent-incongruent stimuli from a fellow group-member were expected to display a smaller N400 effect compared to the other group. While some differences between accents or groups were significant, no evidence in favor of these hypotheses was found. A number of possible explanations for this lack of effect is discussed, as well as suggestions to improve further work in this field.

N400; speaker-message integration; implicit associations; accent stereotypes
1. Introduction

Despite our best efforts, humans constantly place others into categories. Think of people who claim to have a ‘gaydar’, for instance (Shelp, 2003). These assumptions regarding one’s underlying personality traits can be based on a large number of variables that define us socially in the eyes of society, such as physical mannerisms (Johnson, Gill, Reichman, & Tassinary, 2007), skin color (Lundman & Kaufman, 2003), but also accent. A myriad of studies have already investigated what explicit associations exist regarding one’s way of speaking (Kinzler & Dejesus, 2013; Grondelaers, Van Hout & Steegs, 2010; Dixon, Mahoney, & Cocks, 2002; Zahn & Hopper, 1985). These studies have shown that assumptions are quickly made about one’s socioeconomic status and social attractiveness once they start speaking. While the Dutch pride themselves on being one of the most open-minded and progressive nations in the world (Buruma, 2007), many instances of stereotyping exist in the Netherlands that are based on accent. The extent of this becomes especially apparent when looking at wage gaps between speakers of standard Dutch and a dialect (Yao & van Ours, 2016; Yao & van Ours, 2018). Previously published work on accent-related associations mainly utilized questionnaire-based methods (e.g. Hilton & Gooskens, 2013). While these studies have been indispensable in showing us the associations people have towards certain accents, they only reflect an explicitly held opinion. These explicit opinions may very well differ from the implicit associations that are held by a participant (e.g. Fazio & Olson, 2014). However, often used measures of implicit associations, such as the Implicit Association Task (Greenwald, McGhee, & Schwartz, 1998), have been heavily criticized
Furthermore, even if these methods were without controversy, they would not tell us anything about the neural processes leading to the activation of the association. A more suitable method for measuring implicit associations, therefore, could be the Event Related Potential (ERP) technique. This method investigates electrophysiological responses to stimuli over time using electroencephalography (EEG). At least one study (Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008) already found that the brain responds strongly when speaker and message cannot be unified by a listener. More specifically, a negative deflection in the signal that is normally associated with semantic errors was found when a speaker made statements that were not in line with expectations regarding them. Similarly, the present study aimed to investigate to what extent implicit associations regarding a non-standard accent could be measured in the brain. Like van Berkum and colleagues, the N400 – a negative deflection in EEG signal in the 300-500ms time-window that is normally associated with semantic errors – was used. More specifically, we investigated ERPs in participants listening to Frisian-accented Dutch. These speakers are often seen as ‘nice, but slightly dim’ (e.g. Hilton & Gooskens, 2013). Participants who grew up in this region, however, could be more prone to rejecting these stereotypes because they consider themselves part of the stereotyped group (Tajfel, 1974; Efferson, Lalive, & Fehr, 2008). We therefore also investigated whether regional identity affected N400 components.
2. Theoretical background

2.1. How are attitudes formed?

Humans constantly judge, categorize, and predict the behavior of others. These assumptions are based on a large number of possible variables: African American men get stopped more often by police due to their skin color (Lundman & Kaufman, 2003), men with a lisp are often rated as gay (Mack & Munson, 2012), and people with glasses are seen as more intelligent (Harris, Harris, & Bochner, 1982). Before we can delve into specific stereotypes pertaining to accent, it is necessary to describe how stereotypes and attitudes form. Stangor (2000) states two different ways in which stereotypes can arise. The first is that stereotypes are generalizations of traits that do in fact occur. He mentioned the stereotype that all women are compassionate and caring because they often take the role as the main parental figure in our society. The second reason is called illusory correlations, which essentially is a type of confirmation bias of something that is not actually true. An example would be that if you think you have seen more bad female drivers, it is likely the case that your perceptions are modulated by your expectations (Stangor, 2010, p. 140).

When we look at accents, then, we can apply similar reasoning from both perspectives. As an example let us look at Texans in the United States. Approximately 15% of Texans statewide are working in agriculture (Gleaton & Anderson, 2005). Texans are known for their distinctive ‘drawl’. For instance, sounds that would be diphthongs in standard English are pronounced as a monophtong. A Texan saying “by” would not say /baI/, but /ba:/ (Thomas, 2003). As a result of generalizations, it is not very strange to assume that someone who
speaks with this highly recognizable Texan accent is a farmer. When it comes to illusory correlations we can also look at U.S. regional accents. A widely held belief is that the so called ‘rednecks’ are generally unintelligent blue-collar workers. However, according to the illusory correlations theory, this is simply a self-perpetuating stereotype: because you are primed to think a certain way, you see everything that fits your narrative as proof, and everything else you simply ignore. These stereotypes can have far-reaching consequences. For example, there is a strong stereotype that women’s mathematical prowess is inferior to that of men. Studies have shown repeatedly that by priming this stereotype (i.e. by telling women that men score better on a certain test), math performance of female participants becomes worse (Spencer, Steele, & Quinn, 1999; Cadinu, Maass, Rosabianca, 2005). Stereotyping, then, is a self-fulfilling prophecy: when a stereotype exists, those targeted start behaving as they are expected. This subsequently strengthens the stereotype’s apparent validity.

One can probably imagine intuitively that it is much easier to stereotype a group that you lack knowledge of than a group that you are a part of. Furthermore, if you belong to a certain group you are more prone to defending others in that group (Efferson, Lalive, & Fehr, 2008). This stems from social identity theory (Tajfel, 1974). Very simply put, an individual considers others who are similar to themselves us (the ingroup). Anyone who differs from said individual, then, is considered them (the outgroup) in this paradigm (Stets & Burke, 2000, p. 225). Many studies have shown that, indeed, intergroup relations are often typified by the assumptions introduced above. A strong ingroup identity, for instance, has been linked to more stereotyping of the outgroup
Additionally, ingroup members tend to accept negative statements about themselves from another ingroup member, while criticism from an outgroup is often responded to defensively (Hornsey, Oppes, & Svensson, 2002). This type of tribalistic thinking is reinforced by a lack of intergroup contact. Indeed, a meta-analysis of over 500 studies found that contact between groups robustly predicted the reduction of outgroup prejudice (Pettigrew & Tropp, 2006). It seems, then, that stereotyping of groups is caused in part by a lack of knowledge of an outgroup. At the same time, stereotypes are ‘accurate’ because those under stereotype threat start behaving as others think they will.

2.2. Speech and expectations

2.2.1. Accent and identity

In the previous section, we have seen how the groups one belongs to, like one’s ethnicity or sex, may influence how one is perceived. Another, perhaps equally salient, variable is the way one talks, or one’s accent. In the present study, accent is defined as an amalgamation of variables that differ between speakers from different regions and social groups. A single utterance can provide a wealth of information regarding one’s social identity. Gay men, for instance, are more likely to have a lisp (Van Borsel et al., 2009), the prosody of non-native speech is distinct from native speech (Kolly & Dellwo, 2014), and pronunciation of certain phonemes can differ between working class and upper class inhabitants of the same town (Stuart-Smith, Timmins, & Tweedie, 2007). This information is not only salient, but it is also extracted quickly. Flege, for instance
found that participants were able to classify a French-English accent after as little as 30ms of exposure (1984).

Differences between accents are not coincidental, either; people seem to be at least subconsciously aware of how identity can be performed through an accent. William Labov's seminal department store study (1972), for instance, showed that employees of several department stores at different price points adjusted their accent based on the perceived socioeconomic status of their interlocutor. This perception was based on the price-range of the store: the stores that stocked the most expensive items, had clerks that displayed the most usage of the standard variants, while the employees of inexpensive stores had more features that were associated with the working classes. Similarly, Eberhardt and Downs (2015) found that employees in one store that sold dresses at multiple price points adjusted their accent based on the budget of the customer. The employees expected that someone who shopped within a relatively expensive price range belonged to a high-status social group. The identity of the addressee (i.e. the client in the store), then, was assumed by the employee, and as a result, they aligned their speech to match that of the perceived socioeconomic status of their interlocutor.

Another example of accent being utilized as a tool to express identity is Moroccan Flavored Dutch (Nortier & Dorleijn, 2008). The authors describe how phonetic features from Moroccan and Berber were employed in the Dutch of youths from ethnic minority groups. For instance, the uvular fricative /χ/, which is present in Dutch, may be emphasized and lengthened, while any words with /sx/ will become /šx/ (e.g. “sleep” will become “shleep”). This could easily be
ascribed to a low level of proficiency. However, the youths displaying these features were often second or third generation immigrants, who were perfectly capable of speaking Dutch with an accent that was more typical of native speakers. Moreover, they only used this accent among friends. Interestingly, Moroccan youths were not the only group using these features. Other groups, such as Turkish and Afghani teens used the same, markedly Moroccan, features in their speech. This non-standard language use, then, could be attributed to these youths’ ingroup identity of being part of minority groups in a country where the majority of people was white and Dutch. The use of a non-standard accent to denote group belonging is called covert prestige. It was first found by Labov (1963), who found that the permanent residents of a popular holiday island had developed markedly different speech patterns from General American. Similar findings were reported by Cutler (2010), who found that a group of white immigrants (e.g. Russian) in the U.S. who identified with hip-hop culture, were more likely to adopt African American Vernacular English (AAVE) accent-features, than General American speech features. Cutler argued, for reasons similar to participants in Labov’s study, that this was an act of identity perhaps aimed to disrupt the expectation their speech would become ‘white’ by default that because they were white: since the immigrant youths did not identify with the dominant white culture, the AAVE accent had higher covert prestige for them.

The sources mentioned above show that an accent is not merely an arbitrary marker showing one’s regional background (Kolly & Dellwo, 2014). Rather, one’s accent could be seen as a performance of for instance ethnic
minority identity (Cutler, 2010; Nortier & Dorleijn, 2008), sexual orientation (Van Borsel et al., 2009), and socioeconomic status (Labov, 1972; Eberhardt & Downs, 2015, Stuart-Smith et al., 2007). These factors, among others, all converge into an individual’s identity, and consequently this combined identity is reflected in one’s accent.

2.2.2. Status and solidarity

As was demonstrated above, accent and identity are strongly linked. It is not strange, then, that people have a myriad of associations with accents, like they have with ethnicity and gender. These associations are very often based on stereotypes. As early as 1960 studies were done that asked English participants to rate speech samples of a familiar (English) and unfamiliar (French) languages on factors such as intelligence, dependability, and likability (Lambert, Hodgson, Gardner, & Fillenbaum). Later on, an influential study was published by Zahn and Hopper (1985), in which participants were asked to rate speech samples from four different English speakers: a young African American speaker from the Southern U.S., a young white Appalachian male, and two Midwestern graduate students (male and female). Their principal axes factor analysis yielded a number of variables that can be grouped under status, such as literacy, education, and income. Additionally, a number of the variables could be categorized as solidarity, like niceness, goodness, and honesty. The status dimension, then, can broadly be put as any trait that expresses socioeconomic status and competence, while the solidarity dimension encompasses traits that are associated with social attractiveness. A third dimension that is sometimes mentioned in the literature is dynamism. This dimension contains traits related to liveliness (e.g. friendly,
hardworking, strong). Many of these traits, however, overlap with both status and solidarity to some extent (for a meta-analysis with dynamism traits explained, see: Fuertes, Gottdiener, Martin, Gilbert, & Giles, 2012, pp. 124-126). Since the present study is explorative in nature, there is a need to use only the more clearly defined status and solidarity because effects of dynamism would be harder to explain.

2.2.3. Speaker evaluations of status and solidarity

A large number of studies have focused on attitudes towards accents with similar approaches to Zahn and Hopper (1985). Accent, for instance can lead us to think of people as more friendly or of higher socioeconomic status. It seems that when it comes to native versus non-native accents, native speakers are almost always rated more positively for both status and solidarity traits. A 1991 study by Tsalikis, Deshields, and Latour, for instance, investigated evaluations of students at Florida university who were presented with a sales pitches from several speakers with either a native (General American) or non-native (Greek) English accent. They not only found that participants were more likely to buy from a native speaker, but also that the American salesperson was deemed significantly more friendly, intelligent, and of higher status. This preference for a native salesperson has been replicated in Guatemalan Spanish versus non-native Spanish (Tsalikis, Ortiz-Buonafina, & Latour, 1992), Australian English versus Indian English (Tombs & Rao Hill, 2014), and American English versus Cuban and Nicaraguan English (Deshields, Kara, Kaynak, 1996). Similarly, other studies have focused on how an accent influences potential employability. Many
studies have reported that those with a stronger non-native accent are often not deemed good candidates (Carlson & McHenry, 2006; Clark & Paran, 2007).

On the one hand, one could argue, that non-native speakers are simply not as able to utilize persuasive techniques as native speakers are. Indeed, DeCarlo reported that an inability to do so would immediately lead to suspicion in the context of a sales pitch (2005). On the other hand, a strong link between a high degree of ethnocentrism and negative foreign accent evaluations has been reported before (Neuliep, Speten-Hansen, 2013; Chakraborty, 2017). Indeed, it appears that even outside the context of sales native accents are preferred (Kinzler, Shutts, Dejesus, and Spelke, 2009). These researchers presented 5-year-old children with recordings from native speakers of their first language, as well as non-native speakers. Additionally, these speakers could be the presented on screen as being the same or a different race as the children. The children were then presented with these stimuli, after which they were asked who they would rather be friends with. It was found that speakers with a native accent were deemed more suitable friends than the non-native speakers. Additionally, when only pictures were shown, children were more likely to pick a same-race friend. The most interesting finding of this study, however, was that accent had more influence on the child’s choice than race. More specifically, when the children were shown a same-race child with a non-native accent, and a different-race child with a native accent, they would be more likely to choose the different-race child as a friend. This study shows not only that attitudes towards accent are strongly ethnocentric (in terms of nationality, not necessarily ethnicity), but also that this favoritism of one’s own accent is already present in early childhood.
This does not mean that native speakers win the popularity contest by default, however. In other instances, an accent from another country is deemed more socially attractive or of higher status. For instance, American undergraduates found their own accent the easiest to understand, but preferred hearing a British accent (outgroup) because it was more euphonious (Scales, Wennerstrom, Richard, & Wu, 2006). Similarly, native New Zealanders rated their own accent lower on the status dimension than a British English or American English accent (Bayard, Weatherall, Galloïd, & Pittam, 2001). Furthermore, Lalwani, Lwin, and Li (2005) found that Singaporean participants rated a British-English spokesperson as significantly more reliable and credible than one speaking Singaporean English. Similarly, non-native accents are sometimes used in commercials to denote a certain image; think of the use of an Italian accent in a wine commercial to portray the brand as authentic (Peters & Hammonds, 1984) or the use of French to seem more alluring or mysterious (Pinet, 1979). It seems then that on the one hand, people are more prone to trusting someone from their own ingroup (Tajfel, 1974), while on the other hand they are aware of the prestige of other accents (a term popularized in: Giles, 1973).

The sources mentioned up until now, however, discussed native versus non-native accents. When looking at regional versus standard accents, on the other hand, the standard accent sometimes loses its social attractiveness. A study conducted in England showed that participants rated a regional West Yorkshire accent much higher than a standard Received Pronunciation accent on sociability, friendliness, and sincerity (Hiraga, 2005). The standard accent,
however, was rated much higher on traits related to status (e.g. intelligence, successfulness). Another study by Kinzler and Dejesus (2013), similarly, found that speakers with a non-standard regional accent (Southern U.S.) were seen as friendlier but less intelligent than those with a standard accent (Northern U.S.). Interestingly, these ratings were the same across respondents from both parts of the United States. This suggests that participants were not only aware of the stereotypes regarding the groups, but they even agreed with them, regardless of whether they concerned the ingroup or outgroup.

Not all non-native accents are created equal, however. Dixon and colleagues (2002) investigated to what extent an accent influenced the attribution of guilt in a matched guise study. This experimental paradigm entails that one speaker records stimuli for both conditions in an experiment. As a result, variation in ratings is more easily ascribed to the accent, instead of inter-speaker differences. Participants listened to an interaction between a policeman and a suspect with a Received Pronunciation (standard) or Birmingham (non-standard) accent. The researchers found that the non-standard guise was deemed guilty significantly more often than the standard guise. Hiraga (2005), similarly, reported that the Birmingham accent was rated lower for solidarity traits than Received Pronunciation. She mentioned that these results seemed to be in line with Wilkinson’s theory (1965), which postulates that rural non-standard dialects will be preferred over urban non-standard dialects. Another study in the U.K. also found that Standard English and the Queen’s English were indeed seen as more socially attractive and prestigious than more urban regional accents, like the ones found in Manchester and Liverpool (Coupland & Bishop, 2007).
Interestingly, however, these researchers found that ratings for solidarity were more positive when asked to rate an accent that was identical to one’s own, while the self-rating scores for status were much lower than that of Standard English.

Conversely, a study by Bishop, Coupland, and Garrett (2005), found that respondents rated accents from their own region (e.g. Wales, Scotland, and North-Ireland) as being more socially attractive and having higher prestige than respondents from other regions did. These results partly go against the previously mentioned studies, as ingroup members apparently did not agree with the stereotype that they were lower in status. This suggests that, while regional accents are not always thought of as having higher status and solidarity, there may be a tendency for ingroup members to prefer their own accent. However, it must be noted that these speakers still rated Standard-English higher on status dimensions. It seems that for status, standardness is more important than group membership. The fact that some non-native ingroup members rated themselves higher than others rated them, however, seems in line with Tajfel’s theory on ingroups and outgroups (1974), and it was even reported before that (Giles, 1970). These different results from Kinzler & Dejesus (2013) could perhaps be caused by the fact that Scotland, Wales, and Northern-Ireland are all countries within the U.K., each with their own language. In the U.S.A., on the other hand, each state speaks the same language. Perhaps the situation in the U.K., then, fosters a stronger sense of identity. It must be noted, however, that this distinction is difficult to substantiate, as the literature comparing the U.K. to the U.S.A. in this aspect is essentially non-existent. These explanations, therefore, should be seen as conjecture.
2.3. The situation in the Netherlands

While the studies mentioned up until now have focused mainly on the Anglosphere, the situation in the Netherlands itself may be completely different. Essentially, at the moment there exists a certain divide between the Randstad (i.e. the larger cities), and the provinces (e.g. the smaller cities and/or more agriculturally based communities). For a visualization of these regions, please refer to Figure 1.

![Figure 1: Map of accent-regions within the Netherlands based on Pinget, Rotteveel, & Van de Velde (2014), who made distinctions of accents based on postcodes (pp. 12-13). North, South, and Mid-Netherlands are all seen as ‘the provinces’. Please note: Friesland, marked with red, is a part of the North. Furthermore, while dialects of Frisian are also spoken on some of the Wadden Islands, these were not taken into account for the sake of ease.](image-url)
The Randstad accent, through the educational system and the media, has been made the standard since the 1920s (Grondelaers, Van Hout, & Speelman, 2011; Pinget, Rotteveel, & Van de Velde, 2014). As a result, there seems to be a strong sense of what ‘correct’ Dutch is supposed to sound like. Only a small number of studies, however, have investigated accent evaluations in the Netherlands. Nevertheless, the studies that have been done, have often found similar results to Kinzler and Dejesus (2013), in that speakers with regional accents are rated as higher in solidarity but lower in status. A study by Heijmer and Vonk (2002), for instance, investigated evaluations of six regional accents. These accents were Standard Dutch (Randstad), Amsterdam (postcode 10 in Figure 1), the Hague (postcode 25), Limburg (postcodes 60-64), Twente (postcode 74), and Flemish (not visible in Figure 1, as it is in Belgium). It was found that standard Dutch was rated much higher than the other accents on status, while it was rated much lower for solidarity. In line with Wilkinson’s theory (1965), the accents that were most associated with ‘the provinces’ (i.e. Twente and Limburg) were rated highest for solidarity.

Conversely, a more recent study found that an accent from Groningen (postcodes 96 to 99 in Figure 1), another accent strongly associated with more rural areas, was rated low for both status and solidarity (Grondelaers, Van Hout, & Steegs, 2010). Another recent study investigated the same status and solidarity dimensions, but included ingroup and outgroup members (Grondelaers, Van Hout, & Van Gent, 2019). Like Heijmer and Vonk, they found that a strong Randstad accent was associated with much lower solidarity ratings, than a strong Southern accent. Additionally, the Southern accent was rated lower on
status than the Randstad accent. Interestingly, in terms of ratings it did not seem to matter whether a participant was from the same or a different region as the speaker. This suggests that the ingroup solidarity, like reported in previous studies (e.g. Bishop et al., 2005; Giles, 1970; Coupland & Bishop, 2007), was not present. Rather, these results reflect more those of Kinzler and Dejesus (2013).

One group from the provinces that experiences certain negative stereotypes are the Frisians (any area marked in red in Figure 1). The Frisians are a linguistic minority who have their own language that is recognized by the government. Despite this formal recognition, however, use of Frisian is on the decline (Rys et al., 2017). Additionally, Frisian seems to be subject to the same attitudes as other regional accents. Hilton and Gooskens (2013) studied attitudes towards Frisian in Friesland versus the rest of the Netherlands using the matched guise technique. They found that informers who were not from the province of Friesland rated the Frisian guise as less clever. On the other hand, results showed that Dutch participants believed that Frisian sounded equally as friendly as Dutch. Contrary to Heijmer and Vonk’s findings, however, Frisian participants rated the Frisian guise much higher on cleverness than the Dutch participants did. This suggests that Frisians indeed display some degree of ingroup bias in favor of Frisian.
2.4. Separating implicit and explicit attitudes

2.4.1. The implicit association task: the most common technique

The studies mentioned above are invaluable because they show us what kind of explicit associations one may have with a certain accent. Explicit attitudes, however, may not reflect implicit ones. Indeed, Quillian (2008) discussed this mismatch between explicit and implicit attitudes in an essay on whether unconscious racism could exist at all. He stated that “rather than replacing explicit attitudes, implicit attitudes form a second level of attitudes that become manifest in certain behaviors and contexts” (p. 7). The question then becomes: how can we measure something that one is perhaps not even aware of? By far the most prevalent method for measuring implicit attitudes is the Implicit Association Test (IAT), which was devised by Greenwald, McGhee, and Schwartz (1998). IATs exist for several purposes, such as measuring implicit racial bias (Saujani, 2002), homophobia (Steffens & Buchner, 2003), and fatphobia (Teachman & Brownell, 2001); the methodological premise, however, is the same for all tests. In one version of the IAT, participants are asked to sort stimuli into career-related words and family-related words by pressing two buttons on the keyboard (Project Implicit, 2011). Simultaneously, they have to press one button for words related to women, and another for words related to men. The hypothesis is that longer RTs for the stereotypically ‘incongruent’ condition (i.e. female-career, male-family) suggest that there are implicit biases against career-oriented women and family-oriented men.
Interestingly, many studies suggest that there is often a mismatch between explicit attitudes and implicit associations. Participants with positive explicit opinions about gay men still displayed negative bias in the IAT (Steffens & Bucher, 2003), health professionals with neutral explicit attitudes towards obese people showed negative implicit associations (Teachman & Brownell, 2001), and participants from the United States who explicitly state that they did not have any bias towards non-native accents still showed an implicit preference to a U.S. accent over a Korean one (Pantos & Perkins, 2013).

2.4.2. Methodological issues in measuring implicit attitudes

It must be mentioned, however, that the IAT is highly controversial. Some authors suggest that the mismatch between implicit and explicit attitudes are indeed caused by motivation to give a socially desirable answers to explicit questions (Fazio & Olson, 2014). The IAT, then, would reveal one’s ‘true’ beliefs one was trying to hide, instead of subconscious beliefs one was not aware of. If this were true, the IAT would be an appropriate measure of implicitly held attitudes. Others, however, state that the IAT may activate cultural stereotypes that the participant may not hold personally (Arkes & Tetlock, 2004). Therefore, if an IAT suggests implicit associations between people of color and negative emotion, it might as well be that the participant in question was reminded of bias from society at large. Yet others believe that IAT results are caused by salience asymmetries (Rothermund & Wentura, 2004), in the sense that a negative word may be highly salient (e.g. Pratto & John, 1991), while a word related to homosexuality, for instance, may be unsalient to a heterosexual participant.
Rothermund and Wentura argue that this mismatch in salience, rather than one's personal attitudes, is the primary cause of the longer response times.

Furthermore, even if IATs could definitively show implicit bias, they would not be the ideal way to measure implicit attitudes towards accent. Pantos and Perkins (2013) found a preference towards the native accent, in the sense that response times were shorter when native speech and 'good' words had to be sorted on the same side than when a foreign and 'good' were to be sorted on the same corner. The mean length of their auditory stimuli, however, was 1.25ms. They, rightfully, stated that “it appears that participants in the present study formed their implicit attitudes based on the non-nativeness of the accent and not on stereotypes related to any particular nationality” (p. 12). If one were to devise an auditory IAT with complete words or utterances, another problem would arise. The online processes that are executed while an implicit association forms, may have completed before the end of a word. As a result, participants’ response times would be meaningless. This, then would merely exacerbate a major issue that the visual IAT also suffers from: it only measures the end-state of processing. While the physical response (i.e. pressing a button) can be seen as a proxy of how long it took for this process to complete, it is completely uninformative about the type of cognitive processes that lead to that response.

Thus far we have seen that current measures of attitudes are problematic for two reasons. Either, they are too explicit, causing responses that are perhaps only given because they are socially desirable. Or they only reflect an end state in processing. If we want to measure the cognitive processes leading up to the activation of an implicit attitude, another measure should be utilized.
2.4.3. ERPs: the ‘true’ implicit measure?

A more suitable, but highly underused, method to measure implicit association is the Event Related Potential (ERP) technique. This method investigates changes in electrophysiological activity in the brain by utilizing electroencephalography (EEG). Steven Luck, one of the founding fathers of the technique as it is known today defined ERPs as “a scalp-recorded neural signal that is generated in a specific neuroanatomical module when a specific computation is performed” (2014, p. 66). Very simply put, an ERP is an electrophysiological response in the brain to a stimulus. ERPs are differentiated from each other by looking at the latency (i.e. how quickly after onset of a stimulus the effect occurs), the polarity (is the deflection positive or negative), and the distribution on the scalp (e.g. does the effect occur in the frontal electrodes). An early posterior negativity (EPN), for instance, is “a negative potential over visual cortex in the N2 latency range” (Luck, 2014, p. 107), which can be found in the posterior (i.e. in the back) electrodes. Different types of ERPs are used to denote several processes in the brain: a P600 (i.e. a more positive signal peaking around 600ms post onset of the stimulus), for instance, reflects difficulties with certain syntactic structures (e.g. Kaan, Harris, Gibson, & Holcomb, 2000), in sentences like “the girl in the white dress eat the mango”. The EPN, an example that was mentioned earlier, on the other hand is often found in highly arousing stimuli (Kissler, Herbert, Winkler, & Junghofer, 2009).
ERPs do not suffer from the same methodological issues as IATs because they are measured over time, instead of after processing. As a result, activity before, during, and after the presentation of a stimulus can be analyzed. Furthermore, no explicit responses are needed from a participant during testing; this prohibits participants from giving socially acceptable answers, which is a risk in explicit measures (Fazio & Olson, 2014).

2.4.4. The N400

A highly suitable ERP component for the present study is the N400. This consists of a negative deflection in the EEG signal that peaks approximately 400ms post stimulus onset, and shows the largest effect in the 300-500ms range (e.g. Balconi & Pozzoli, 2005; Kmiecik & Morrison, 2013). It was first reported by Kutas and Hillyard (1980), who found that the amplitude of the N400 was significantly larger in sentences containing lexico-semantic incongruities, such as “he spread the warm bread with socks” (p. 203). In his book, Luck (2014), cites the two most prevalent theories on what the N400 actually reflects. One of these theories states, in Luck’s words, that “the N400 component reflects neural activity associated with finding and activating meaning” (p. 105). This was posited by Kutas, van Petter, & Kluender (2006). The second theory, posits that an N400 component reflects the difficulty or the ease with which a certain word can be integrated in the preceding context: “the better the semantic fit between a word and its context, the more reduced the amplitude of the N400” (Hagoort, 2007, p. 246).
As the theories mentioned above perhaps already suggest, every word elicits an N400 component (see Figure 2). More typically, when a study refers to an N400 effect, it means a larger negativity elicited in one condition relative to another one. This can be seen in Figure 2, too, where socks (incongruent) elicits a much larger negativity than work (congruent).

In the forty years following its discovery, many studies have reported on this component. For instance, the N400 seems to occur regardless of modality (for an overview see: Kutas & Federmeier, 2011): it has been reported in written language (Kutas & Hillard, 1980), auditory stimuli (Perrin & García-Larrea, 2003), and semantically anomalous pictures (Nigam, Hoffman, & Simons, 1992). Furthermore, semantically correct words with a low probability of occurring, given the sentential context (low Cloze probability words) elicit N400s, as well (Kutas & Hillyard, 1984; Loerts, Stowe, & Schmid, 2013). More specifically, the less expected a word is within a context, the larger the amplitude of the N400.

Figure 2: Example of an N400 component (adapted from Luck, 2014, p. 104). As is clearly visible, every word elicits a negativity of sorts, but the largest negativity is seen when a word is semantically incongruent (i.e. “socks”). Please note: In EEG research it is more common to have the y-axis inverted (i.e. + and – are normally reversed). The graphs in our results section follow this convention.
2.4.4. Speaker-message integration and the N400

If the N400 only reflected difficulty with retrieval or integration of semantic information, it would not be very suitable for the present study. However, many other experimental paradigms have reported this component. Not only the sentential context seems to modulate N400 amplitudes; the wider pragmatic context influences N400 in the same manner. The sentence “the girl comforted the clock” (Nieuwland & Van Berkum, 2006, p. 1098), for example, seems semantically incongruent. Indeed, Nieuwland and Van Berkum found an N400 effect in participants who saw this sentence by itself. However, when it was preceded by a context where an anthropomorphic clock and a little girl had a conversation about the clock’s mental health, the N400 effect disappeared completely. In a similar vein, N400s have also been found in even broader pragmatic contexts, such as real-world knowledge (Hagoort, Hald, Bastiaansen, & Petersson, 2004). Hagoort et al. found that semantically correct sentences that conflicted with real-world knowledge (e.g. “Groningen is located in Melbourne”), elicited N400 components that were similar in amplitude to semantically incorrect sentences.

One could argue, however, that implicit associations are different from factual real-world knowledge or the context provided by an overarching storyline. Since they are based on stereotypes, one may be aware of their possible inaccuracies on a conscious level, while believing their accuracy on a subconscious level. A study from 2008, however, provides evidence against this (Van Berkum, Van den Brink, Tesink, Kos, & Hagoort). Van Berkum and colleagues investigated if a mismatch between the speaker and message would
elicit an N400, as well. They presented participants with phrases that were either congruent or incongruent with stereotypical perceptions of the speaker that said them. For instance, they had an older woman with an upper-class accent say “I have a large tattoo on my back” (p. 581). It was expected that the N400 component would arise after the trigger word “tattoo”. It was indeed found that N400 amplitude was significantly larger after stimuli where speaker and message could not be integrated, such as the example mentioned above. This suggests that the N400 amplitude is in fact modulated by speaker-expectations, implicit attitudes, and stereotypes.

Another interesting study found somewhat similar effects of foreign accent (Hanulikova, van Alphen, Van Goch, & Weber, 2012). Instead, these researchers investigated the P600 effect, which reflects morphosyntactic errors. They specifically investigated participants’ responses to determiner-noun errors (e.g. de* meisje: the* girl, instead of het meisje: the girl). These errors could be made by a native or non-native speaker of Dutch. Native Dutch speakers typically do not make these errors, while non-native speakers, heavily overuse the more prevalent form ‘de’ (e.g. Blom, Polisenská, & Weerman, 2008). Unsurprisingly when participants heard a native speaker make an incorrect determiner-noun combination, a P600 was found. However, when a non-native speaker made the same mistake, no such effect was found. These results are similar to those in van Berkum and colleagues’ (2008) study. They suggest that the accent elicited certain expectations regarding language proficiency (i.e. a non-native speaker will make more mistakes in picking the right determiner), which was reflected in the EEG signal. While the present study will not investigate responses to
morphosyntactic errors, this again strongly suggests that expectations based on accent modulate our brain's response to stimuli.

N400 effects have also been reported in written word pairs that were incongruent with gender stereotypes, such as “Men: Gossipy” (White, Crites, Taylor, & Corral, 2009, p. 193). Again, semantically there was nothing incorrect about this combination, but from a pragmatic viewpoint this reversal of traditional gender roles violated expectations. Osterhout, Bersick, and McLaughlin (1997), also investigated responses to gender stereotypes. For instance, when subjects read items such as “the doctor prepared herself for the operation” (p. 274), an N400 effect was found. Racial profiling also seems to be reflected by this same N400 component (Hehman, Volpert, & Simons, 2013). These researchers showed participants stereotypically white traits (e.g. educated, spoiled), and stereotypically black traits (e.g. athletic, hostile, armed) in combination with faces from either a Caucasian-American or African-American person. Similar to White and colleagues’ study (2009), when the combination of adjective and face did not match the stereotypes an N400 effect was found.

From the studies discussed above it can be concluded that the reach of the N400 goes much further than simply indexing the response to semantic errors. Rather, arguing from Hagoort’s (2007) theory, the N400 could be seen as a reflection of how easily a stimulus can be integrated in the context of what is seen or heard. This indeed encompasses semantic errors (e.g. Kutas & Hillyard, 1980), but evidently it also reflects mismatches between what is expected through stereotypes and what is actually experienced.
2.5. Relevance of the study

The studies mentioned above have comprehensively shown that evaluations of a speaker’s personality are strongly dependent on their accent. Evidence from the Netherlands suggests that those with standard accents are often seen as possessing higher status seen by participants, while those with regional accents receive higher scores for solidarity-related traits (e.g. Heijmer & Vonk, 2002). At first glance, these evaluations seem relatively harmless echoes of stereotypes of the ‘busy work-driven young urban professional’ and the ‘friendly farmer from the small town where time seems to stand still’. We have also seen, however, that some of the negative evaluations regarding non-native accents can lead to accent-based discrimination in experimental courtroom settings (Dixon et al., 2002). Equally worrisome, is that studies utilizing real-life data have found strong evidence for discrimination of people with a regional accent. For instance, a large-scale study utilizing data from 7,000 participants found that Dutch men who speak a dialect (and as a result have a regional accent) earn significantly less than those who speak a standard variant (Yao & Van Ours, 2016; Yao & Van Ours, 2018). Additionally, while somewhat outside the scope of the present study, those with non-native accents are also barred from jobs because they supposedly sound ‘unprofessional’ (Ghorashi & Van Tilburg, 2006). These sources seem to point towards undesirable patterns of economic discrimination of those with non-native or regional accents. If the present study finds strong evidence for a negative bias towards regionally accented speakers, in the sense that they are seen as lower in status, this could for instance be made aware more explicitly in
educational facilities and in the workplace. However, of course, any findings for the present study should be validated before any such drastic steps are taken.

Moreover, this study adds to the existing literature in a number of ways. It expands on findings that link the N400 to difficulties in speaker-message integration (van Berkum et al., 2008) by investigating if and how incongruities with expectations influence the amplitude of the N400. Using this method, then, can provide more insight into the processes behind speaker-message integration. The present study also adds to the existing literature by investigating whether implicit measures (i.e. ERPs) have an advantage over explicit measures (i.e. questionnaires). Lastly, using these implicit measures, we will be able to see the stereotyped group’s reactions to incongruities with stereotypes about themselves. This, then, not only shows how Frisians process their own stereotypes, but it also potentially gives us an insight into the degree that Frisians consider themselves part of their own ingroup.
3. Present study

As we have seen in the background reading, non-standard accented speech often brings with it certain stereotypes. On the one hand these regional accents are seen as pleasant and friendly, but on the other hand they are often seen as less intelligent and lower in socioeconomic status. Additionally, studies that focus on ingroup versus outgroup and attitudes towards accented speech seem to find mixed results: sometimes speakers rate themselves higher on the status and solidarity dimensions, while other times the outgroup is rated higher. We can potentially get a clear view of this issue by investigating the neural correlates (in particular the N400) that underlie the processing of standard and non-native accents, as well as the differences between ingroup and outgroup members.

The aim of the present study, then, as was alluded to before, is to see whether EEG can be used as a method to measure the surprise that stems from hearing an utterance from someone that is not in line with stereotypes about them. In part, this study aims to see whether the results reported by van Berkum et al. (2008) can also be found when one’s accent is the variable making an utterance congruent or incongruent with the message. First and foremost, therefore, our aim is to investigate the difference in EEG signal between accent-congruent and accent-incongruent sentences on the status and solidarity conditions.

Before stating research questions a short introduction is in order to define accent-congruent and accent-incongruent sentences within the present study. These will be revisited in more detail in the method section. The sentence “I became a lawyer after high school”, for instance, is clearly high-status. Standard
Dutch speakers are often rated higher on the status dimensions (e.g. Grondelears et al., 2019) while Frisian accented speakers are deemed less intelligent (e.g. Hilton & gooskens, 2013). Therefore, we see this as a sentence that is accent-congruent for the Dutch speaker, while it is accent-incongruent for the Frisian speaker. The reverse would be true for a sentence such as “I became a plumber after high school”. Since it is low-status, this item is congruent for Frisian-accented speakers and incongruent for Dutch-accented speakers. For solidarity sentences, on the other hand, regional accents are often deemed nicer, while the standard accent is the one that is seen as less socially desirable (Heijmer & Vonk, 2002). Therefore, “I always give money to charity” would be an accent-congruent solidarity sentence for the Frisian speaker, while it would be accent-incongruent for the Dutch speaker. Lastly, to partly validate our results, Cloze sentence pairs were employed (Loerts, Stowe, & Schmid, 2013). These sentences are always grammatically and semantically correct, but one half of the sentence pair is much less predictable (e.g. “He always sleeps on the right side of the bed high probability” versus “He was watching TV on the bed low probability”). A larger N400 component for sentences with low Cloze probability has been reported in a number of studies (Delong, Urbach, Groppe, & Kutas, 2011; Block & Baldwin, 2010). If there is no difference in ERPs between high and low Cloze probability sentences, then, it could suggest that our sample is underpowered. With this in mind, the following research questions were posed:
3.1 Research Questions and Hypotheses

**RQ1:** When looking at the data from the Frisian-accented and Dutch-accented stimuli combined, will accent-incongruent words within a sentential context elicit a larger N400 component relative to accent-congruent items, reflecting a mismatch between speaker and message?

**RQ2:** When comparing data from the Frisian-accented and Dutch-accented stimuli to each other, to what extent will there will be differences in N400 amplitudes when comparing accent-incongruent stimuli from one speaker to accent-congruent stimuli from the other speaker?

**RQ3a:** When ingroup members hear accent-congruent stimuli that reflect negatively on their group, will their N400 amplitudes differ from that of outgroup members exposed to the same stimuli?

**RQ3b:** When members of one group hear accent-incongruent stimuli that reflect positively on their ingroup, will their N400 amplitudes differ from that of outgroup members exposed to the same stimuli?

**RQ4:** How will N400 amplitudes differ between high and low Cloze probability sentences?
**H1:** An accent-incongruent utterance will lead to issues with the integration of speaker and message (Van Berkum et al., 2008), and as a result, the N400 amplitude will be significantly larger for accent-incongruent stimuli than for accent-congruent stimuli.

**H2:** In comparing utterances between accents, we believe that the same issues with speaker-message integration apply (Van Berkum et al., 2008), which will be reflected by significantly larger N400 amplitudes for speaker-incongruent stimuli compared accent-congruent stimuli.

**H3a:** Ingroup members will have issues integrating their own negative stereotypes due to ingroup solidarity (e.g. Bishop et al., 2005; Hilton & Gooskens, 2013), which in turn will be reflected by larger N400 components for accent-congruent stimuli that are negative about their group.

**H3b:** Since ingroup members think more positively about members of their own group (e.g. Tajfel, 1972; Hilton & Gooskens, 2013), we expect that accent-incongruent stimuli that are positive only elicit an N400 in the outgroup.

**H4:** We expect that low-Cloze probability sentences elicit a significantly larger N400 component than high-Cloze probability sentences (e.g. Loerts et al., 2013).
4. Method

4.1. Participants

Twenty-five participants (6 men, 18 women, 1 would rather not say) were recruited through flyers in the university, guest-lectures, and by sending e-mails to study associations. Participants’ ages ranged from 19-26 ($M = 22.1$, $SD = 1.94$), and most were enrolled in university courses ($N = 21$). Of the other four subjects, three were enrolled in vocational level education, and one was enrolled in a university of applied sciences. Participants were either paid 10 euros for their time, or they received course credit.

All participants except three were born in the Netherlands. One of these three was adopted from China at an early age. The other two were born in Luxembourg and Finland, respectively, but had grown up speaking Dutch with their parents. They were therefore deemed native speakers of Dutch. Of the participants born in the Netherlands, the majority came from the North of the Netherlands ($N = 18$), the rest came either from the South ($N = 1$), or from the Randstad ($N = 3$). Of the 25 participants, only four indicated that they spoke Frisian.

All participants in the present sample were right-handed and native speakers of Dutch (and Frisian if applicable). None of the participants indicated having any learning disorders (e.g. dyslexia). None of the tested participants used prescription medication (e.g. Ritalin/Medikinet for ADHD). Lastly, none of the participants indicated having had previous head-injury that could warrant atypical brain activity.
4.2. Materials

4.2.1. Stimuli

Stimuli consisted of spoken sentences and were made by creating a number of bipolar sentence pairs for both status and solidarity (136 pairs for Status, 127 pairs for Solidarity. 526 stimuli in total). Each sentence pair was written in such a way that the sentences were identical up to the trigger word (i.e. the word that caused the sentence to be either congruent or incongruent). The sentences differed in accent-congruency. As was indicated in the theoretical background section, Frisian speakers are generally thought of as lower in status while Standard Dutch speakers are generally seen as high-status (Hilton & Gooskens, 2013). We expect that issues with speaker-message integration will arise when subjects hear an accent-incongruent sentence (like in Van Berkum et al., 2008). An example of a sentence pair for status (with the trigger word underlined) would be:

*Mijn jongere broer is werkzaam als piloot high-status/vuilnisman low-status en zou geen andere baan willen.*

*My younger brother works as a pilot high-status/garbage man low-status and wouldn’t like another job*

For solidarity, sentences were constructed using the same method. An example of a sentence pair is:
Toen mijn nicht twintig kilo overgewicht was kwijtgeraakt in de sportschool zei ik dat ze er geweldig high-solidarity/mannelijk low-solidarity uit zag toen ze op de koffie kwam.

When my cousin lost twenty kilos in the gym I told her that she looked amazing high-solidarity/masculine low-solidarity when she visited for coffee.

Additionally, the design included 124 Cloze sentence pairs from a previous study (Loerts, Stowe, & Schmid, 2013) as fillers. These pairs contained one high Cloze item (highly expected given the sentential context) and one low Cloze item (highly unexpected given the sentential context). These stimuli were neutral, in the sense that they were not related to stereotypes regarding accents. Aside from acting as a filler to obscure the main objective of the study, the stimuli were also chosen because Loerts et al. found an N400 effect using these sentences. As a result, they acted as a control for our data. An example of a Cloze sentence pair would be:

_De herder controleert de kwaliteit van het schaap High Cloze voor de verkoop_  
_The shepherd checks the quality of the sheep High Cloze before the sale_

_De handelaar controleert de kwaliteit van het schaap Low Cloze voor de verkoop_  
_The merchant checks the quality of the sheep Low Cloze before the sale_

A complete overview of the sentences utilized in the study can be seen in Appendix A.
4.2.2. Stimuli validation

Stimuli were presented to a panel of speakers (N = 193) through a Qualtrics (Provo, UT) questionnaire. The stimuli were divided into four lists (two regarding status, two regarding solidarity), such that each half of a sentence pair was rated in two situations: a congruent context, where one where the sentence was expected from the speaker, and an incongruent context, where the sentence was unexpected. All respondents were asked to rate 60 of these sentences that were randomly chosen from two of the four lists of stimuli (30 status-related, 30 solidarity-related). Respondents who were presented with status-related sentences from one list, for instance, would receive the question: “how strange or normal would you find it if a rich/intelligent person said the following?” Raters who received the same sentence in the other list would be prompted with “how strange or normal would you find it if a poor/unintelligent person said the following?” Respondents were then asked to rate the sentences from this list on a scale from 1 (very normal) to 5 (very strange). As mentioned above, the rated stimuli could be congruent (i.e. a high-status stimulus presented after being prompted with the rich/intelligent question stated above), or incongruent (i.e. a low-status stimulus presented after seeing the rich/intelligent question stated above). The same procedure was followed for solidarity sentences.

A number of tactics were devised to ensure each sentence was validated. Like previously mentioned, each stimulus was placed in two lists: one list where the direction of the question was ‘high’ (i.e. making a rating based on the assumption that the person saying it is intelligent or wealthy), and vice versa one where the direction of the question was ‘low’. This ensured that each sentence
was rated in both the congruent and incongruent contexts. Furthermore, the questionnaire was counterbalanced, such that no respondent would rate both the high and low half of a sentence pair. Additionally, stimuli within each list were counterbalanced, such that each stimulus received an approximately equal number of ratings (with a minimum of 5 raters per stimulus). This was not always successful, however. In order to receive an adequate number of participants per stimulus, therefore, the procedure was repeated with sentences that did not have an adequate number of responses. Additionally, sentences that were not given the expected ratings were rewritten and re-entered in these questionnaires. On average each part of a sentence pair in each of the lists was rated 6.69 times ($SD = 1.7$), with the mode being 5 ratings per sentence.

The mean rating for each individual sentence was then determined. Subsequently, this score was compared to the expected score: 1 for congruent sentences, and 5 for incongruent sentences. For congruent sentences a threshold of $<2.5$ was set, and for incongruent sentences a threshold of $>2.5$ was put in place. In the end, the speaker-congruent solidarity sentences had a mean rating of 2.16 ($SD = 0.65$), where the expected score was 1. Speaker-congruent status sentences had a mean rating of 2.25 ($SD = 0.57$). Speaker-incongruent solidarity sentences, with an expected score of 5, were on average given a rating of 3.43 ($SD = 0.93$). Lastly, the speaker-incongruent status sentences received a 3.06 on average ($SD = 0.72$). For an overview of the ratings, please refer to Appendix B.
4.2.3. Stimuli Recordings

Two (female) employees were recruited from the University of Groningen. One of the speakers had a Frisian accent, while the other had a upper-class Randstad accent (specifically Amsterdam). Each speaker recorded the sentences in a sound-proof recording booth on a Shure SM27 microphone with a windscreen fitted over the element to reduce the loudness of plosives like /t/, and /p/.

Recordings were made in Adobe Audition version 12.1 (Adobe Systems, 2019). Each recording was then cropped, such that there were no unwanted pauses before the onset of a sentence. Other undesirable sounds, like breaths, were also removed manually. Every sentence was then normalized for volume in Audacity version 2.3.2. (Audacity Team, 2019), such that peak volume did not exceed -6 decibels relative to full scale (dBFS). This means that the peak volume in any given recording is 6dB softer than the loudest possible sound the microphone can record. The limit of -6dBFS ensured that the audio did not become distorted (this would happen if the signal exceeded the 0dBFS limit).

The sentences were then compressed (threshold -22dB, noise floor -40dB, ratio 2:1) to ensure that each sentence was approximately the same loudness at any given point in the sentence. Essentially, this procedure entailed that any audio signal that was softer than -22dBFS was made twice as loud. However, to limit amplifying background noise, any sound that was softer than -40dBFS did not get amplified. After compression, the recordings were normalized once again at -6dB.

Because the present study was part of a larger project that included two slower speaking non-native speakers, recordings of the Frisian and Dutch
speakers were slowed down by 5% in Adobe Audition using the time stretch function. This tool was chosen because it ensured that speech sounded natural (other methods produced audio clipping). Additionally, the time stretch function did not alter pitch characteristics. As a result, the changes were unnoticeable.

Trigger points were placed at four points: the word preceding the target word, the target word, the word following the target word, and the last word in the sentence. These triggers were first placed in Adobe Audition using the marker function; they were subsequently manually exported to Excel. Triggers were given a hexadecimal code that was then placed in the E-prime file.

4.2.4 Questionnaires

Prior to signing up, each participant was asked to fill out a short questionnaire with questions regarding exclusion criteria. This questionnaire included questions regarding native languages, handedness, medication, and drug use. The full pre-selection questionnaire can be seen in Appendix C.

After participation, participants were given another questionnaire that included more specific questions regarding language use (e.g. frequency of usage, context of usage), and their socioeconomic status. This questionnaire was administered after testing because some of the questions likely would give away the goal of the experiment. For instance participants were asked to guess where the speakers they heard were from. Additionally, participants were asked to rate the speakers on a 5-point Likert scale (completely disagree – completely agree) on traits also utilized in Zahn & Hopper (1985), such as “rich”, and “warm”. The posttest questionnaire can be seen in Appendix D.
4.2.5. EEG lab equipment

All recordings were made in Brain Vision Recorder 2.1. (Brain Products GmbH, 2019) using a 64+8 channel EEG cap. The electrode distribution was according to the 10-20 system (Sharbrough et al., 1991; Jasper, 1958). For an overview of the electrodes please refer to Figure 3. Four separate electrodes were used to filter out ocular disturbances. Horizontal eye-movements were recorded using two electrodes placed parallel to the pupils on the sides of the face near the outer canthi. Blinks were recorded using two more electrodes placed in one line with the pupil above and below the left eye. EEG and EOG signal was amplified with a ReFa amplifier (TMSI 8-64 / 72 channels). Electrical impedance was kept below 20kΩ. The data was recorded with a sampling frequency of 500Hz.

Figure 3: Overview of electrodes utilized in the present study. Originally by Albronda (2019, p. 24).
4.2.6. E-prime experiment

E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was utilized to present stimuli to participants. Every participant would receive one of eight e-prime experiments that were identical in structure and procedure. Each of these experiments differed, however, in the sentences that the participant would receive. There were eight experiments, all corresponded to one of the eight lists that were utilized in the experiment. The lists were counterbalance such that a participant would only hear one part of a sentence pair from one of the four speakers who recorded the sentences (2 parts per sentence pair x 4 speakers in the project = 8 lists).

4.3. Design

A 2 (Frisian/Dutch participant) x 2 (Frisian/Dutch accent) x 2 (accent-congruent/accent-incongruent) design was used. As mentioned before, sentences were presented in such a manner that participants only heard one version of a sentence pair as uttered by one speaker. As a result, while each sentence pair was used twice, once for each accent (i.e. Frisian congr., Dutch congr., Frisian incongr., Dutch incongr.), only one of these four were presented to any one participant. Analyses will were done with generalized additive mixed models, or GAMMs (See section 4.5.3. and 4.5.4. for a justification; for an in-depth explanation, see: Wieling, 2018). The dependent variable was the amplitude of the EEG signal measured in microvolts within the N400 time-window (300-500ms) after the trigger word was uttered (e.g. Van Berkum et al., 2008). Explanatory variables were congruity of the utterance, the accent of the speaker, and the language background of the participant.
4.4. Procedure

After signing a consent form (Appendix E), and reading the letter with information regarding the experiment (Appendix F), participants were briefed on the procedure. They were then seated in a quiet room, approximately 60cm (23.5”) from a 43cm (17”) Samsung Syncmaster 797DF CRT monitor. Once they were connected to the amplifier, and impedance was sufficiently low, they were given both written and verbal instructions regarding the experiment. Participants were told that they would listen to a number of sentences. Additionally, they were instructed to pay attention closely because a true/false question regarding the sentence could be asked. The purpose of these questions was twofold. They ensured that participants paid attention at all time. Additionally, they made participants believe that the study was about how well accents could be understood, which distracted them from the actual subject matter of our study.

Each trial was presented as follows: a fixation cross appeared on the screen. Then, after 500ms, the recording started playing, while the fixation cross remained on the screen. After the sentence the participants saw four asterisks for 2000ms, which indicated that they could blink. The stimuli were presented over one training block (4 stimuli) and five testing blocks (76 or 77 stimuli per block). Between each block, participants were given an opportunity to take a break. After the last block, the conductive gel was washed from the participants’ hair. Lastly, they were asked to fill out the post-test questionnaire. They were then briefed verbally, and given the opportunity to ask questions.
4.5. Analysis

4.5.1. Preprocessing

Firstly, data from five participants was discarded. Four of these were caused by missing markers (i.e. the triggers that tell the EEG recording software to start/stop recording). A fifth participant’s data was discarded due to having overall bad signal. All raw EEG data was first preprocessed in Brain Vision Analyzer 2.1 (Brain Products GmbH, 2019) utilizing a preprocessing pipeline. First, the signal was rereferenced, such that the electrodes at the mastoids were taken as reference electrodes. Data was filtered with a high and low pass filters (0.1 and 50Hz, respectively). Furthermore, the signal was split into segments of 2000ms, each. Subsequently, an ocular correction was applied, which filtered out blinks and other eye-movements. Lastly a baseline correction was done, such that the data became centered (i.e. with the old mean becoming 0).

After the preprocessing, each participant’s data was loaded into R version 3.6.1 (R Core Team, 2019) using pre-processing scripts adapted from Loerts (2012) and Meulman (2014). The signal was then downsampled from 500Hz to 100Hz by averaging the signal per timebin of 10ms. The percentage of artifacts was then calculated ($M = 0.8$, $SD = 0.95$). Subsequently, the E-prime experimental data was added to the data, as to include information regarding the sentence (e.g. accent, sentence direction). Lastly, subject data from the post-test (i.e. ratings for all speakers, and whether the subject was Frisian or Dutch) were added. The data was then split into Regions of Interest (ROIs).
4.5.2. Selecting a Region of Interest

Data per ROI was compared visually by plotting the signal for congruent and incongruent sentences (Figure 4). The regions of interest, starting from the upper-left corner are: central (C1, Cz, C2, CP3, Cpz, CP4, P1, Pz, P2, POz), left-anterior (FP1, AF3, AF7, F3, F5, F7), left-central (FC3, C1, C3, C5, CP3, CP5), left-posterior (P3, P5, P7, PO3, PO7, O1), right-anterior (Fp2, AF4, AF8, F4, F6, F8), right-central (FC4, C2, C4, C6, CP4, CP6), and right-posterior (P4, P6, P8, PO4, PO8). Van Berkum et al. (2008), reported the strongest negativity in the central-posterior electrodes (i.e. p. 586). As can be seen, in Figure 4, the central ROIs for both status and solidarity display the least noisy signal. Therefore, since the cleanest signal was obtained here, and since the electrodes from the central ROI matched those from Van Berkum et al.’s study quite well, only the central ROIs were taken into account for the present analyses.

As can be seen in both plots showing raw signal, a small negativity seems to arise for incongruent items in both status (upper panes) and solidarity (lower panes) conditions over the central electrodes.
Figure 4: Plots per ROI for congruent (green) and incongruent (red) items. Status sentences are plotted in the upper 7 panes, whereas solidarity sentences are plotted on the lower 7 panes.
4.5.3. Generalized Additive Mixed Models

Studies investigating ERPs often suffer from one big issue: inappropriate analyses. While EEG signal is incredibly rich, providing a temporal resolution of up to 2kHz, many studies utilize ANOVAs (e.g. Kuipers & Thierry, 2011; Van Berkum et al., 2008). Firstly, this method is not ideal because it looks at means per group over a specific epoch (e.g. 300-500ms for the N400). As a result, much of the complexity and variation that occurs in the data is reduced to a single number. Additionally, ANOVAs are inappropriate for EEG data because the assumption of normality is rarely met. While this is not very problematic in groups of equal size it can still be problematic if data is very skewed (Skidmore & Thompson, 2013). Another method that is often used for EEG data is linear mixed-effects regression models (e.g. Stites, Payne, & Federmeier, 2017). This approach is also far from ideal, however, as EEG data is highly autocorrelated due to the signal being time-series (Baayen, Van Rij, De Cat, & Wood, 2018).

A better approach, therefore, would be Generalized Additive Mixed Models, or GAMMs (Hastie, 2017; Wood, 2017). GAMMs are ideal for EEG data because they possess “a large array of tools for modeling nonlinear dependencies” (Baayen et al., 2018, p. 1). These nonlinear relations are modelled by combining a number of basis functions together to create a function that closely follows the data. One can imagine, however, that this can lead to overfitting (e.g. Hawkins, 2004). This would be problematic because it becomes impossible to generalize the data to a population if an analysis only matches the dataset. GAMMs, however, include a non-linearity penalty to the data in order to correct for this (Wieling, 2018, p. 90). Additionally, the presence of autocorrelation is not as big an issue
because it can be corrected for (Meulman, Wieling, Sprenger, Stowe, & Schmid, 2015).

All analyses were conducted in R version 3.6.1 (R core Team, 2019), using the \texttt{mgcv} package version 1.8-28 (originally published in: Wood, 2001). More specifically, the \texttt{bam} function was used because it works better on large datasets, such as ours. This is because the \texttt{bam} function allows for the use of multiple processors in parallel. As a result, it is much faster than the \texttt{gam} function. In order to visualize and interpret the results, R package \texttt{itsadug} version 2.3 was utilized (originally published in: Van Rij, Wieling, Baayen, & Van Rijn, 2015). The R-scripts used for building and visualizing the models can be found in \textbf{Appendix G}, the data and other supplemental materials used in this study can be found in \textbf{Appendix H}. 
4.5.4 Explaining our first GAMM model

A number of generalized additive mixed models were built using the bam function. The present study is essentially interested in three variables: the difference between congruent and incongruent sentences, whether the speaker’s accent influences this difference, and the difference between Frisian (ingroup) and Dutch (outgroup) participants hearing a Frisian speaker produce an utterance that is in line with stereotypes about them. These three variables, then, were Congruency over Time, a two-way interaction Congruency x Accent over Time, and a three-way interaction between Congruency x Accent x Dutch/Frisian participant over Time. Initial model building included all three of these terms in the same model. Model inspection, however, showed that a high degree of concurvity (the non-parametric equivalent of multicollinearity) existed between these three terms. This is highly problematic, as it increases chances of making a type I error (Figueiras, Roca-Pardiñas, Cadarso-Suárez, 2005). Figueiras et al. did propose a method of controlling for concurvity by employing a conditional bootstrapping algorithm, but this was deemed too technically challenging to be executed. Therefore, in order to decrease the chance of a type I error, separate models were built per sentence type (status/solidarity) for Congruency, Congruency x Accent and for Congruency x Accent x Dutch/Frisian participant. Additionally, a model was built for Cloze sentences (high vs low Cloze). Separate models had to be built because the packages used do not allow for combining multiple levels of an interaction. The first model specification can be seen in Figure 5.
Since GAMMs are relatively new in EEG research, an explanation of the code in Figure 5 is in order. Our dependent variable is the EEG signal measured in mV. Signal was taken from the electrodes in the central region of interest (C1, Cz, C2, CP3, CPz, CP4, P1, Pz, P2, POz). This signal is then modelled on the explanatory variables (i.e. anything after the tilde symbol). The non-linear patterns in our data are modelled using ‘smooths’. Smooths are made by combining a number of basis-functions until they fit the data (Wieling, 2018). The first term, s(Time, by=Congruency), then, creates a non-linear pattern for both congruent and incongruent sentences over the -500 to 1500ms time-window. In addition to the non-linear functions, Congruency is put in the model by itself as a fixed effect, too. This way, it can be investigated whether congruent and incongruent sentences differ significantly from each other when time is not taken into account. Furthermore, two random effects are added, one for Subject, s(Time, Subject, bs=‘fs’, m=1), and one for Word (s(Time, Target, bs=‘fs’, m=1). These model a non-linear pattern over time per speaker and word. In this case bs=‘fs’ invokes a random effects function. Furthermore, m = 1 refers to which part of the

After the explanatory variables, some other specifications regarding the computation are made. Gc.level specifies how often unused but full memory should be emptied. It was set to 2 here, meaning as often as possible. method=’fREML’ specifies that fast restricted maximum likelihood estimation should be utilized; this method was chosen because it is much faster than other methods

```
mV ~ s(Time,by=Congruency) + Congruency + s(Time, Subject, bs=‘fs’, m=1) + s(Time, Target, bs=‘fs’, m=1), data=dat, gc.level=2, method='fREML', discrete=F,family=scat, rho=rhoval1[2], AR.start=SeqStart, nthreads=16)
```

Figure 5: Example of GAMM used in present study
(Wieling, 2018). Discrete was set to TRUE. This converts the continuous data in our model to discrete (i.e. rounded) values. As a result, computation time becomes much lower. Family was set to scaled-t distribution, because EEG data is often heavy tailed (Meulman et al., 2015). The Rho parameter controls for autocorrelation. It is optional, but necessary, since time-course data is heavily autocorrelated. A value for Rho is typically determined by running the same model without a specified Rho parameter, and then taking the amount of autocorrelation at lag 2 (Wieling, 2018). AR.start specifies where each stimulus starts. Lastly, the nthreads argument specifies how many computer cores can be used. This also speeds up computation, as normally only one core is used.
5. Results

5.1. Status

5.1.1 Status: Speaker ratings

Each participant was asked to rate the speakers they heard in the experiment on a number of traits related to status: intelligence (unintelligent-intelligent), socioeconomic class (working class-upper class), education (low-high), income (poor-rich), and literacy (functionally illiterate-highly literate). An average status rating was then obtained by calculating the mean score for these traits. A paired two-sided Wilcoxon-Signed rank test showed that the Frisian accent was deemed significantly lower in status (Median = 3.33) than the Dutch accent (Median = 4), $Z = -3.68, p < .001$. This difference is visualized in Figure 6. Interestingly, Frisian participants rated the Frisian speaker only marginally higher on status (Median = 3.67, $SD = 0.77$) than Dutch participants did (Median = 3, $SD = 0.6$). This difference was not significant, however, as an unpaired Wilcoxon-Signed rank test pointed out ($Z = -0.13, p = .55$) Furthermore, Frisian participants even rated the Dutch speaker somewhat higher on status (Median = 4.25, $SD = 0.64$) than the Dutch participants (Median = 4, $SD = 0.53$). Again, however, this difference was not significant ($Z = 0.33, p = .63$).

Figure 6: Boxplots with status ratings per accent and subject group (FR = Frisian subject, NL = Dutch subject)
5.1.2. Status: Congruency

The first model (as mentioned above in Figure 5), modelled the effect of Congruency over time for all status sentences. The results of this model can be seen in Table 1. In rows 1 and 2, we can see the estimates for the parametric coefficients (i.e. the effect of Congruency, without taking into account the time-course of the data). We see that, there is a very small but insignificant difference between congruent and incongruent sentences (line 2). In lines 3 through 6 of the table, the estimates for our smooth terms can be seen. Both Congruent and Incongruent items were significant. This is not very informative, however, as the \( p \)-value for smooths only indicate whether or not the smooth differs significantly from 0 (i.e. Wieling, 2018). Therefore, plots were made to visualize the signal amplitude over time for both Congruent and Incongruent items (Figure 8).

Unsurprisingly, significant results were found for our random slopes per subject (Figure 7) and per item (not plotted due to the large number of items). This supported the inclusion of these two random slopes, and it suggested that a large amount of variation was due to individual participants and/or words.

| Table 1: Summary of the results for our GAMM model, modelling the effects of Congruency over Time |
|------------------------------------------|----------------|------------|---------|----------------|
| **Parametric coefficients:**             | **Estimate**  | **Std.Error** | **t-value** | **Pr(>|t|)** |
| 1 (intercept)                            | -1.12         | 0.316      | -3.547   | <.001 ***     |
| 2 Congruency-Incongruent                  | 0.063         | 0.185      | 0.341    | .73           |
| **Smooth terms:**                         | **EDF**       | **Ref.DF** | **F**   | **P-value**   |
| 3 s(Time): Congruency-Congruent          | 5.169         | 6.291      | 4.615    | <.001 ***     |
| 4 s(Time): Congruency-Incongruent        | 5.069         | 6.179      | 5.977    | <.001 ***     |
| **Random effects**                       |               |            |          |               |
| 5 s(Time, Subject)                       | 53.807        | 179        | 2.105    | <.001 ***     |
| 6 s(Time, Target)                        | 396.940       | 2.087      | 2.110    | <.001 ***     |
| * \( p < .05 \)                           | ** p < .01    | *** p < .001 |
| R-Squared (adjusted): 0.122               |               |            |          |               |
| Deviance explained: 9.06%                 |               |            |          |               |
Firstly, the pane on the left in Figure 8 was made to model the signal for both Congruent and Incongruent items. Because we were interested in an N400 effect of incongruent sentences relative to congruent ones, it was necessary to plot the difference between congruent and incongruent sentences. The difference in signal of congruent – incongruent can be seen in the pane on the right. If the two estimates differed significantly, the function plot_diff would have visualized at

Figure 7: Plots signal (in mV) per participants for all status sentences over the central ROI. Green lines indicate congruent sentences, while red lines indicate incongruent items. As it clearly visible, there is a large amount of variation per subject. This warrants the inclusion of a random slope per participant.
what points in time the difference in signal between the two levels became significant by highlighting that portion in red. This was not the case for congruent vs incongruent sentences at any point during the measured time-window (-500 until 1500ms). It is useful to note that for every difference plot in this study, the same procedure was followed: the effect that the largest negativity was expected from (i.e. the one that should be most surprising in that particular situation), was subtracted from the effect the smallest negativity was expected from. This way, if a difference plot shows a significant result on the negative (top) half of the difference plot, it means that indeed the surprising condition elicited a negativity effect.

While there was a negativity for incongruent sentences peaking around 600ms, the same negativity was also present for congruent items. When comparing the two with a difference wave (right plot), a slight negativity effect of incongruent items relative to congruent items seemed to occur around the 1000ms mark. However, this difference was not significant.

**Figure 8:** Estimated effect of Congruency over time for all Status sentences (left panel) with a difference plot (right panel) over the electrodes in the central ROI (measured in mV) during the entire measurement (500ms pre-stimulus to 1500ms post-stimulus. There is no significant difference between Congruent and Incongruent sentences in the present sample.
5.1.3. Status: Congruency x Accent

A second model was specified that replaced Congruency with Congruency x Accent. This model enabled us to see whether any significant differences could be found between items that were congruent for the Frisian-accented speaker and incongruent for the standard Dutch speaker (i.e. low-status sentences), or vice versa (i.e. high-status sentences: congruent for Dutch, incongruent for Frisian).

In Table 2 we can see that for the parametric coefficients (lines 2 through 4) there were no significant differences between the reference level (Congruent-Frisian speaker) and any of the other levels. When looking at the smooth terms, we can see that all levels except Congruent-Frisian items differed significantly from 0 (lines 5 through 8). Again, since this was not very informative, a plot showing each level was created (Figure 9). Lastly, like for our previous model, significant results were found for both random effects (lines 8 and 9). Again, this suggests that a large amount of variation is found on the subject and item level.

Table 2: Summary of the results for our GAMM model, modelling the effects of Congruency x Accent over Time for Status sentences

| Parametric coefficients: | Estimate | Std.Error | t-value | Pr(>|t|) |
|--------------------------|----------|-----------|---------|---------|
| 1 (intercept)            | -1.097   | 0.375     | -2.925  | < .01 **|
| 2 CongrAccent- Incongruent Frisian speaker | 0.124 | 0.392 | 0.315 | .75 |
| 3 CongrAccent – Congruent Dutch speaker | 0.039 | 0.408 | 0.096 | .92 |
| 4 CongrAccent – Incongruent Dutch speaker | 0.078 | 0.281 | 0.277 | .78 |
| Smooth terms: EdF | Ref.DF | F | P-value |
| 5 s(Time): CongrAccent-Congruent Frisian | 3.971 | 4.915 | 2.032 | .07 |
| 6 s(Time): CongrAccent- Incongruent Frisian | 3.069 | 3.8 | 5.106 | < .001 *** |
| 7 s(Time): CongrAccent-Congruent Dutch | 31.06 | 3.847 | 5.929 | < .001 *** |
| 8 s(Time): CongrAccent-Incongruent Dutch | 3.981 | 4.927 | 3.867 | < .01 ** |
| Random effects | |
| 9 s(Time, Subject) | 55.478 | 180 | 2.07 | < .001 *** |
| 10 s(Time, Target) | 396.613 | 2088 | 2.104 | < .001 *** |

* p < .05
** p < .01
*** p < .001

R-Squared (adjusted): 0.122
Deviance explained: 9.06%
It seemed that the differences between levels were very small. There also seemed to be a negativity for all levels of the interaction, but no level differed much from any of the others. This was also suggested in Table 2, where none of the levels differed significantly from the reference level (lines 2 through 4). The fixed effects, however, did not take into account changes over time. Therefore, to see whether any of these differences were significant over time, difference plots were made for Congruent – Incongruent per accent (Figure 10), as well as the difference between Frisian and Dutch per direction (Figure 11).

Figure 9: Estimated effect of Congruency x Accent over time for all Status items over the electrodes in the central ROI (measured in mV).
Firstly, in **Figure 10** no significant differences were found between congruent and incongruent sentences for the Frisian speaker (left pane). There was a small negativity effect for incongruent sentences around the 1000ms mark, but this did not meet the significance threshold. Statistically, therefore, there was no significant difference between expected (congruent) and unexpected (incongruent) sentences for the Frisian speaker. Similarly, for the Dutch speaker the difference congruent – incongruent was modelled (right pane). Again there was no significant difference between congruent and incongruent sentences. Instead, a small but insignificant positivity effect for the congruent sentences could be seen in the 1000-1500ms range.

![Figure 10: Difference plots showing congruent minus incongruent status sentences per accent for the central ROI (measured in mV). In the left pane the difference between Congruent-Frisian (i.e. low-status) and Incongruent-Frisian (i.e. high-status) is modelled. In the right pane the difference between Congruent-Dutch (i.e. high-status) and Incongruent-Dutch (i.e. low-status) is modelled. There are no significant differences between congruent and incongruent for either accent.](image)
Subsequently, difference waves per direction (high – high and low – low) were plotted for both accents (Figure 11). These difference waves still modelled the difference between congruent – incongruent sentences, but now the response per accent to the same type of sentences was compared. In the pane on the left the difference between high-status Dutch (congruent) and high-status Frisian (incongruent) was modelled. A small positivity effect for high-status Frisian (incongruent) items appeared to arise from the start of the trigger word until 600ms, but this difference was not significant. On the right, the difference between low-status Frisian (congruent) and low-status Dutch (incongruent) was modelled. Again, there were some differences. Mainly a positivity effect for low-status Dutch (incongruent) items before the onset of the stimulus could be seen. Additionally, a small positivity effect for low-status Dutch items appeared at the end of the measured time-window. Again, however, at no point in time were there significant differences between low-status Frisian and low-status Dutch.

Figure 11: Difference plots for high-status sentences per accent (left) and low-status sentences per accent (right). All signal is measured over the Central ROI (measured in mV). In the left pane the difference between high-status Dutch (congruent) minus high-status Frisian (incongruent) is modelled. On the right, the difference between low-status Frisian (congruent) minus low-status Dutch (incongruent) is plotted. There are no significant differences in either plot.
5.1.4. Status: Congruency x Accent x Group

Disclaimer: Before discussing this model, it is imperative to stress that there are only 4 Frisian subjects. Any results regarding group membership, therefore, should be seen as completely exploratory.

In the last model for the status sentences, the previous interaction Congruency x Accent was expanded to include information regarding the subject’s background. More specifically, information was added about whether the subject was Frisian or Dutch. This interaction was investigated to see what differences group-membership (i.e. ingroup/outgroup) would make in the response. The summary for this model can be seen in Table 3.

For the parametric coefficients, there were significant differences between the reference level (Congruent x Frisian speaker x Dutch subject) and Incongruent x Dutch speaker x Frisian subject ($T = 2.29$, $p < .05$), in that the latter displayed a significantly more positive signal (line 7). There were no significant differences between the reference level and any of the other levels in the fixed effects structure.

For the smooth terms, the effects in lines 9 through 12 differed significantly from 0. Furthermore, like in the previous models, there was a significant effect of subject and item. To see the effect of Congruency x Accent per Group (Frisian or Dutch), please refer to Figure 12.
Table 3: Summary of the results for our GAMM model, modelling the effects of Congruency x Accent x Group over Time for Status sentences.

| Parametric coefficients: | Estimate | Std.Error | t-value | Pr(>|t|) |
|--------------------------|----------|-----------|---------|----------|
| 1 (intercept)            | -1.325   | 0.4       | -3.309  | < .001 *** |
| 2 Incongruent - Frisian speaker – Dutch subj. | 0.649 | 0.405 | 1.604 | .11  |
| 3 Congruent - Dutch speaker – Dutch subj. | 0.027 | 0.421 | -0.064 | .95   |
| 4 Incongruent - Dutch speaker – Dutch subj. | 0.336 | 0.31 | 1.083 | .28   |
| 5 Congruent - Frisian speaker – Frisian subj. | 0.711 | 0.708 | 1.004 | .32   |
| 6 Incongruent - Frisian speaker – Frisian subj. | 0.31 | 0.783 | 0.396 | .69   |
| 7 Congruent - Dutch speaker – Frisian subj. | 1.773 | 0.774 | 2.292 | < .05 * |
| 8 Incongruent - Dutch speaker – Frisian subj. | 0.186 | 0.708 | 0.262 | .79   |

<table>
<thead>
<tr>
<th>Smooth terms:</th>
<th>EDF</th>
<th>Ref.DF</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>s(Time): Congr. x Accent x Subj. group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Congruent – Frisian speaker – Dutch subj.</td>
<td>4.016</td>
<td>4.977</td>
<td>2.672</td>
<td>&lt; .05 *</td>
</tr>
<tr>
<td>10 Incongruent - Frisian speaker – Dutch subj.</td>
<td>2.692</td>
<td>3.331</td>
<td>3.146</td>
<td>&lt; .05 *</td>
</tr>
<tr>
<td>11 Congruent – Dutch speaker – Dutch subj.</td>
<td>3.293</td>
<td>4.086</td>
<td>4.916</td>
<td>&lt; .001 ***</td>
</tr>
<tr>
<td>12 Incongruent - Dutch speaker – Dutch subj.</td>
<td>3.844</td>
<td>4.767</td>
<td>6.081</td>
<td>&lt; .001 ***</td>
</tr>
<tr>
<td>13 Congruent - Frisian speaker – Frisian subj.</td>
<td>1.025</td>
<td>1.047</td>
<td>1.281</td>
<td>.42</td>
</tr>
<tr>
<td>14 Incongruent - Frisian speaker – Frisian subj.</td>
<td>2.211</td>
<td>2.749</td>
<td>2.747</td>
<td>.06</td>
</tr>
<tr>
<td>15 Congruent - Dutch speaker – Frisian subj.</td>
<td>1.005</td>
<td>1.01</td>
<td>0.45</td>
<td>.64</td>
</tr>
<tr>
<td>16 Incongruent - Dutch speaker – Frisian subj.</td>
<td>1.004</td>
<td>1.008</td>
<td>1.065</td>
<td>.3</td>
</tr>
</tbody>
</table>

Random effects

| 17 s(Time, Subject) | 53.342 | 180 | 1.592 | < .001 *** |
| 18 s(Time, Target)  | 378.446 | 2087 | 1.039 | < .001 *** |

* p <.05
** p <.01
*** p <.001

R-Squared (adjusted): 0.106
Deviance explained: 8.19%
As can be seen in Figure 12, there was a substantial degree of variation between the levels of this three-way interaction. Most notably, Frisian subjects displayed much less variation in signal over time than Dutch subjects. In order to see whether group-membership had any effect on response, difference plots were created (Figures 13 through 16).

**Figure 6:** Estimated effect of Congruency x Accent x Group over time for all Status sentences (signal from central ROI measured in mV). NL/FR denotes the accent of the speaker, whereas no/yes denotes whether or not the subject was Frisian themselves. **Please note:** the standard error bars have been removed in order to increase readability of the estimates.
The first comparison plotted here is congruent – incongruent per group (Frisian and Dutch) for the Frisian speaker (Figure 13). A large difference between ingroup and outgroup members could be seen, immediately. For ingroup members (Frisian), the items that were incongruent with the Frisian speaker (i.e. high-status) elicited a negativity relative to the congruent (low-status items). This started around the onset of the item, peaked around 600ms, and ended at the end of the sample. Conversely, for outgroup members (Dutch), the congruent items (low-status) elicited a small negativity effect relative to the incongruent items (high-status). This difference first peaked around 500ms, and then became larger again at the end of the complete time-window. Neither of these differences were significant, however.

Figure 13: Difference in signal for congruent minus incongruent Status items for the Frisian speaker per subject group (measured over central ROI in mV). The plot on the left shows the difference for ingroup (Frisian) members, while the plot on the right shows that of outgroup (Dutch) members. No significant differences were found.
The second comparison for the Frisian speaker was the difference between each group per level of congruency. This way it could be investigated whether ingroup and outgroup members responded differently to the same type of sentences. These difference plots can be seen in Figure 14. Again, the difference plots were made in such a way that the level largest negativity was expected from (i.e. the most surprising sentence for that particular group) was subtracted from the one the smallest negativity was expected from (i.e. the least surprising sentence for that group).

Frisian (ingroup) subjects displayed a larger negativity when they heard the Frisian speaker utter a high-status (incongruent) sentence than the Dutch (outgroup) subjects did, as can be seen on plot on the left. For the low-status (congruent) sentences, on the other hand the Dutch (outgroup) subjects displayed a larger negativity relative to the Frisian (ingroup) subjects. Neither of these differences was significant, however.

*Figure 14: Difference in signal between ingroup and outgroup per direction of Status items uttered by the Frisian speaker (over the central ROI, measured in mV). On the left, Dutch (outgroup) minus Frisian (ingroup) can be seen for congruent (low-status) sentences. On the right, Frisian (ingroup) minus Dutch (outgroup) can be seen for incongruent (high-status) sentences. No significant differences were found.*
Subsequently the same procedure as in Figure 13 was repeated for participants listening to the Dutch speaker. This time, the Dutch subjects were the ingroup, whereas the Frisian subjects were the outgroup. First, like before, two difference plots were made for congruent – incongruent for both the ingroup and the outgroup. These difference plots are modelled in Figure 15. Dutch (ingroup) subjects displayed a slight positivity effect for low-status (incongruent) sentences uttered by the Dutch speaker. This difference was not significant, however. Frisian (outgroup) subjects, on the other hand, showed a significant negativity effect for low-status (incongruent) sentences when listening to the Dutch speaker. This difference became significant at 183ms and was sustained almost until the end of the recording (1490ms).

**Figure 15:** Difference in mV for congruent minus incongruent Status items for the Dutch speaker (measured over the central ROI). The plot on the left shows the difference for ingroup (Dutch) subjects, while the plot on the right shows that of outgroup (Frisian) subjects. No significant differences were found for Dutch subjects. Frisian subjects showed a significantly larger negativity effect for incongruent items. This difference is significant from 183-1490ms.
Lastly, a comparison was made for the Dutch speaker, in which the difference between each group per level of congruency was modelled. Again, this was done to inspect if ingroup and outgroup members responded differently to the same type of sentences. These difference plots are visible in Figure 16. On the left, the signal of Frisian (outgroup) subjects was subtracted from that of the Dutch (ingroup) subjects. Dutch subjects displayed a significantly more negative signal than Frisian subjects over the 143-1289ms range for high-status Dutch items. On the right, the signal of Dutch (ingroup) subjects listening to low-status Dutch sentences was subtracted from that of Frisian (outgroup) doing the same. Dutch subjects displayed a slight negativity relative to the Frisian subjects, peaking around 500ms. However, this difference was not significant.

**Figure 16:** Difference in mV between ingroup and outgroup per direction for the Dutch speaker (over the central region of interest). On the left, Dutch (ingroup) minus Frisian (outgroup) can be seen for congruent (high-status) sentences. Dutch subjects display a significantly larger negativity from 143-1289ms. On the right, Frisian (outgroup) minus Dutch (ingroup) can be seen for incongruent (low-status) sentences.
5.2. Solidarity

5.2.1. Solidarity: Speaker ratings

Each participant was also asked to rate the speakers they heard in the experiment on a number of traits related to solidarity: kindness (unkind-kind), normalness (weird-normal), socialness (asocial-social), trustworthiness (untrustworthy-trustworthy), and warmth (cold-warm). An average solidarity rating was then obtained by calculating the mean score for these traits.

Interestingly, a paired two-sided Wilcoxon-Signed rank test showed that the Frisian accent was deemed significantly lower in solidarity (*Median* = 3.4) than the Dutch accent (*Median* = 4), *Z* = -2.73, *p* < .01. This difference can be seen in Figure 17. Additionally, Frisian participants rated the Frisian accent somewhat lower on solidarity (*Median* = 3.9, *SD* = 0.57) than they did the Dutch accent (*Median* = 3.8, *SD* = 0.77). Frisians, however, did rate the Frisian accent slightly higher on solidarity traits than Dutch participants did (*Median* = 3.2, *SD* = 0.68). An unpaired Wilcoxon-Signed rank test, however, showed that this difference was not significant (*Z* = -.28, *p* = .39). Additionally, Dutch participants rated the Dutch speaker somewhat higher on solidarity traits (*Median* = 4.2, *SD* = 0.6) than Frisian participants did. Again, this difference was not significant (*Z* = -.004, *p* = .5)

![Figure 17: Boxplots with solidarity ratings per accent and subject group (FR = Frisian subject, NL = Dutch subject)](image-url)
5.2.2. Solidarity: Congruency

Like for the Status models, the first model for solidarity only investigated Congruency. Estimates for this model can be seen in Table 4. The estimates for the parametric coefficients can be seen in lines 1 and 2. The difference between congruent and incongruent sentences, when time is not taken into account, was not significant ($T = -0.195, p = .83$). In lines 4 through 6 of the table, the estimates for our smooth terms can be seen. Both the smooths for incongruent and congruent items over time were significant. Again, since a significant result for smooths only indicates that a variable differs from 0, this was not deemed very informative. Therefore, plots were made to visualize the signal amplitude over time for both Congruent and Incongruent items (Figure 18). Lastly, like in all the previous models, significant effects were found for both subject (line 5) and target (line 6).

Table 4: Summary of the results for our GAMM model, modelling the effects of congruency over Time for Solidarity sentences

| Parametric coefficients: | Estimate | Std.Error | t-value | Pr(>|t|) |
|---------------------------|----------|-----------|---------|----------|
| 1 (intercept)             | -0.501   | 0.377     | -1.331  | .18      |
| 2 Congruency-Incongruent  | -0.0037  | 0.188     | -0.195  | .85      |
| Smooth terms:             |          |           |         |          |
| 3 s(Time): Congruent-Congruent | 2.03   | 2.667     | 3.704   | < .05    |
| 4 s(Time): Congruent-Incongruent | 2.805 | 3.437     | 5.195   | < .01    |
| Random effects            |          |           |         |          |
| 5 s(Time, Subject)        | 64.195   | 180       | 2.983   | < .001   |
| 6 s(Time, Target)         | 321.17   | 1782      | 2.21    | < .001   |

* $p < .05$  
** $p < .01$  
*** $p < .001$

R-Squared (adjusted): 0.116
Deviance explained: 8.87%
As can be seen in Figure 18, the signal for incongruent items was slightly more positive before the onset of the stimulus (i.e. at the point where both sentence pairs were still identical). After this, incongruent items display a negativity around 200ms. Lastly a slightly bigger negativity for incongruent items appeared towards the end of our sample. None of these effects, however, were significant.

**Figure 18:** Estimated effect of congruency over time for all solidarity sentences (left panel) with a difference plot (right panel) over the electrodes in the central ROI (measured in mV). There were no significant differences between Congruent and Incongruent sentences in the present sample.
5.2.3. Solidarity: Congruency x Accent

A second model was specified that expanded the original model to include a level for the different accents. This model enabled us to see if the same congruent – incongruent difference would be significant when looking at the speakers separately. Furthermore, this model investigated whether any significant differences could be found between items that were congruent for the Frisian-accented speaker and incongruent for the standard Dutch speaker (i.e. high-solidarity sentences), or vice versa (i.e. low-solidarity sentences: congruent for Dutch, incongruent for Frisian). In Table 5 we can see that for the parametric coefficients Incongruent-Dutch (line 4) was slightly more negative than the reference level (Congruent-Frisian). As for the smooth terms, Incongruent-Frisian, Congruent-Dutch, and Incongruent-Dutch all were significantly different from 0 (lines 6 through 8). Lastly, significant effects were found for both of the random effects (lines 9 and 10).

Table 5: Summary of the results for our GAMM model, modelling the effects of Congruency x Accent over Time for Status sentences

| Parametric coefficients: | Estimate | Std.Error | t-value | Pr(>|t|) |
|--------------------------|----------|-----------|---------|----------|
| 1 (intercept)            | -0.164   | 0.41      | -0.401  | .69      |
| 2 CongrAccent- Incongruent Frisian speaker | 0.016 | 0.367 | 0.044 | .96      |
| 3 CongrAccent – Congruent Dutch speaker | -0.525 | 0.38 | -1.368 | .17      |
| 4 CongrAccent – Incongruent Dutch speaker | -0.612 | 0.285 | -2.147 | < .05 * |

<table>
<thead>
<tr>
<th>Smooth terms:</th>
<th>EDF</th>
<th>Ref.DF</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 s(Time): CongrAccent-Congruent Frisian</td>
<td>1.006</td>
<td>1.012</td>
<td>3.648</td>
<td>.11</td>
</tr>
<tr>
<td>6 s(Time): CongrAccent- Incongruent Frisian</td>
<td>1.461</td>
<td>1.749</td>
<td>5.822</td>
<td>&lt; .05 *</td>
</tr>
<tr>
<td>7 s(Time): CongrAccent-Congruent Dutch</td>
<td>1.492</td>
<td>1.793</td>
<td>7.683</td>
<td>&lt; .01 **</td>
</tr>
<tr>
<td>8 s(Time): CongrAccent-Incongruent Dutch</td>
<td>1.005</td>
<td>1.009</td>
<td>10.356</td>
<td>&lt; .01 **</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9 s(Time, Subject)</td>
<td>65.885</td>
</tr>
<tr>
<td>10 s(Time, Target)</td>
<td>314.923</td>
</tr>
</tbody>
</table>

* p <.05
** p <.01
*** p <.001

R-Squared (adjusted): 0.106
Deviance explained: 8.36%
In order to visualize the smooth Congruency x Accent over time, a plot was created showing each level in this interaction (Figure 19). The difference between the levels generally seemed very small. In order to see whether participants responded differently to incongruent items, a difference plot was made showing congruent – incongruent per speaker (Figure 20). Another difference plot modelled the differences between high-solidarity sentences per accent and low-solidarity sentences per accent (Figure 21).

Figure 19: Estimated effect of Congruency x Accent over time for all Solidarity sentences (measured over the central ROI in mV)
In Figure 20, difference waves were made for congruent – incongruent per speaker. On the left, we see that incongruent status items uttered by the Frisian speaker elicited a slightly larger negativity starting around 500ms and persisting until the end of the sample. This difference is not significant. Conversely, for the Dutch speaker, incongruent items elicit a slightly larger positivity from around 500 to 1000ms. Again, this difference is not significant.

**Figure 20:** Difference plots showing congruent minus incongruent solidarity sentences per accent (over the central ROI measured in mV). In the left pane Congruent-Frisian (i.e. high-solidarity) minus Incongruent-Frisian (i.e. low-solidarity) is plotted. In the right pane the difference Congruent-Dutch (i.e. low-solidarity) minus Incongruent-Dutch (i.e. high-solidarity) is plotted. There were no significant differences between congruent and incongruent for either accent.
Figure 21, on the other hand, modelled the differences between accents for the same type of sentences. On the left, the difference in signal between Dutch and Frisian is displayed for the low-solidarity items. The Dutch speaker, from whom a low-solidarity sentence should be expected, elicited a significant negativity effect that was sustained from 505ms to 987ms. One the right, the same difference was plotted for the high-solidarity items. Again, a significant difference between the two types of sentences was found. In this case, a large negativity was elicited by the high-solidarity Dutch (incongruent) items which lasted from 244ms to 1490ms.

Figure 21: Difference plots for low-solidarity sentences per accent (left) and high-solidarity sentences per accent (right). All signal is measured over the Central ROI (measured in mV). In the left pane the difference between low-solidarity Dutch (congruent) minus low-solidarity Frisian (incongruent) is modelled. On the right, the difference between high-solidarity Frisian (congruent) minus high-solidarity Dutch (incongruent) is plotted. In the left plot, the Dutch speaker elicited a significant negativity from 505-987ms. In the right plot, the Dutch speaker elicited a significant negativity from 244-1490ms.
5.2.4. Solidarity: Congruency x Accent x Group

Disclaimer: Before discussing this model, it is imperative to stress that there are only 4 Frisian subjects. Any results regarding group membership, therefore, should be seen as completely exploratory.

Lastly, the interaction Congruency x Accent was expanded to include the subject Group (i.e. Frisian or Dutch participant). The estimates for this model can be seen in Table 6.

Table 6: Summary of the results for our GAMM model, modelling the effects of Congruency x Accent x Group over Time for Solidarity sentences.

| Parametric coefficients: | Estimate | Std.Error | t-value | Pr(>|t|) |
|--------------------------|----------|-----------|---------|----------|
| 1 (intercept)            | -0.332   | 0.443     | -0.75   | .75      |
| 2 Incongruent - Frisian speaker – Dutch subj. | 0.078 | 0.383 | 0.203 | .84 |
| 3 Congruent -Dutch speaker – Dutch subj. | -0.524 | 0.398 | -1.32 | .19 |
| 4 Incongruent - Dutch speaker – Dutch subj. | -0.414 | 0.313 | 0.931 | .19 |
| 5 Congruent - Frisian speaker – Frisian subj. | 0.796 | 0.854 | 0.913 | .35 |
| 6 Incongruent - Frisian speaker – Frisian subj. | 0.819 | 0.897 | 0.593 | .36 |
| 7 Congruent - Dutch speaker – Frisian subj. | 0.538 | 0.907 | -0.207 | .55 |
| 8 Incongruent - Dutch speaker – Frisian subj. | -0.177 | 0.857 | 0.207 | .55 |

<table>
<thead>
<tr>
<th>Smooth terms:</th>
<th>EDF</th>
<th>Ref.DF</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>s(Time): Congr. x Accent x Subj. group</td>
<td>1.434</td>
<td>1.716</td>
<td>4.823</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>10 Incongruent - Frisian speaker – Dutch subj.</td>
<td>1.435</td>
<td>1.718</td>
<td>5.656</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>11 Congruent – Dutch speaker – Dutch subj.</td>
<td>1.011</td>
<td>1.021</td>
<td>11.284</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>12 Incongruent - Dutch speaker – Dutch subj.</td>
<td>1.004</td>
<td>1.007</td>
<td>7.774</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>13 Congruent - Frisian speaker – Frisian subj.</td>
<td>1.007</td>
<td>1.013</td>
<td>0.746</td>
<td>.47</td>
</tr>
<tr>
<td>14 Incongruent - Frisian speaker – Frisian subj.</td>
<td>1.006</td>
<td>1.012</td>
<td>0.489</td>
<td>0.51</td>
</tr>
<tr>
<td>15 Congruent - Dutch speaker – Frisian subj.</td>
<td>1.413</td>
<td>1.695</td>
<td>0.572</td>
<td>0.55</td>
</tr>
<tr>
<td>16 Incongruent - Dutch speaker – Frisian subj.</td>
<td>1.003</td>
<td>1.006</td>
<td>1.214</td>
<td>.41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th>EDF</th>
<th>Ref.DF</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 s(Time, Subject)</td>
<td>63.644</td>
<td>178</td>
<td>2.434</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>18 s(Time, Target)</td>
<td>3.7155</td>
<td>1781</td>
<td>0.917</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

* p < .05  
** p < .01  
*** p < .001

R-Squared (adjusted): 0.097  
Deviance explained: 7.82%
None of the parametric coefficients differed significantly from the reference level Congruent-Frisian speaker-Frisian subject. Of the smooths over time, all the levels representing Dutch subjects were significantly different from 0 (lines 9 through 12). On the other hand, none of the smooths for Frisian subjects were significant. Each level of the smooth s(Time, Congruency-Accent-Group) was then plotted in Figure 22.

The largest negativity seemed to occur in Dutch subjects hearing Congruent-Dutch sentences. After this, followed Dutch subjects listening to Incongruent-Dutch sentences. Generally speaking, the Frisian subjects showed a much more positive signal. In order to see if the subject groups varied significantly, difference plots were made. These can be seen in Figures 23 through 26.

**Figure 22:** Estimated effect of Congruency x Accent x Group over time for all Solidarity sentences (signal from central ROI measured in mV). NL/FR denotes the accent of the speaker, whereas no/yes denotes whether or not the subject was Frisian themselves. Please note: the standard error bars have been removed in order to increase readability of the estimates.
The first set of difference plots in Figure 23 modelled the difference between congruent and incongruent items for the Frisian speaker. The first difference plot in the left pane of the figure modeled Congruent – Incongruent for Frisian (ingroup) subjects. Before the onset of the stimulus a positivity seemed to be present for participants who listened to the incongruent items. After approximately 500ms, however, the incongruent items elicited a noticeable, but non-significant, negativity that lasted until the end of the measurement. On the right, the same difference congruent – incongruent was modelled, but for the Dutch (outgroup) subjects. In the outgroup there seemed to be hardly any difference between congruent and incongruent sentences, except for a very small positivity for incongruent items. Again, however, this difference was not significant.

**Figure 23:** Difference in signal for congruent minus incongruent Solidarity items for the Frisian speaker per subject group (measured over central ROI in mV). The plot on the left shows the difference for ingroup (Frisian) members, while the plot on the right shows that of outgroup (Dutch) members. No significant differences were found.
The second set of difference plots modelled the difference between groups listening to the same type of solidarity sentence spoken by the Frisian speaker (Figure 24). In the left pane the difference ingroup (Frisian) – outgroup (Dutch) was modelled for the congruent items. The outgroup displayed a negativity for congruent items (high-solidarity). This negativity started around the onset of the stimulus and increased in amplitude until the end of the measurement. This difference was not significant, however. On the right, a difference plot was made that modelled outgroup (Dutch) – Ingroup (Frisian), for incongruent (low-solidarity) items. Frisian participants displayed a negativity starting before the onset of the stimulus, which increased in amplitude until the end of the sample. Again, however, this difference was not significant.

Figure 24: Difference in signal between ingroup and outgroup per direction of Solidarity items uttered by the Frisian speaker (over the central ROI, measured in mV). On the left, Frisian (ingroup) minus Dutch (outgroup) can be seen for congruent (high-solidarity) sentences. On the right, Dutch (outgroup) minus Frisian (ingroup) can be seen for incongruent (low-solidarity) sentences. No significant differences were found.
The third set of difference plots in Figure 25 modelled the difference between congruent and incongruent solidarity items for the Dutch speaker. The first difference plot in the left pane modelled congruent – incongruent for Dutch (ingroup) subjects. Incongruent (high-solidarity) items elicited a slightly bigger positivity that lasted from around 100ms after the onset of the stimulus until the end of the measurement. This difference was not significant, however. On the right, the same difference plot is made for Frisian subjects (outgroup). Frisians, on the other hand, displayed a negativity for incongruent (high-solidarity) items that increased in amplitude over the course of the measurement. Again, however, this difference did not reach significance at any point during the measurement.

Figure 25: Difference in signal for congruent minus incongruent Solidarity items for the Dutch speaker per subject group (measured over central ROI in mV). The plot on the left shows the difference for ingroup (Dutch) members, while the plot on the right shows that of outgroup (Frisian) members. No significant differences were found.
The fourth, and last, set of difference plots modelled the difference between the ingroup and outgroup for both types of sentences (Figure 26). The plot on the left modelled the difference in signal for Frisian (outgroup) – Dutch (ingroup) for congruent (low-solidarity) sentences. Dutch subjects displayed a negativity effect when listening to a fellow Dutch speaker utter a low-solidarity stimulus. This negativity increased in amplitude until the end of the time-window. This difference became significant at 1309ms where it remained significant until 1490ms. On the right, the difference in signal was plotted for Dutch (ingroup) – Frisian (outgroup) for incongruent Dutch (high-solidarity) solidarity sentences. Frisian subjects displayed a positivity for high-solidarity sentences uttered by a Dutch speaker. This positivity increased in amplitude throughout the entire duration of the time-window. However, the difference between Frisian and Dutch subjects was not significant in this comparison.

Figure 2: Difference in signal between ingroup and outgroup per direction of Solidarity items uttered by the Dutch speaker (over the central ROI, measured in mV). On the left, Frisian (outgroup) minus Dutch (ingroup) can be seen for congruent (low-solidarity) sentences. On the right, Dutch (ingroup) minus Frisian (outgroup) can be seen for incongruent (high-solidarity) sentences. On the left low-status Dutch elicited a significant negativity in Dutch subjects (1309-1490ms). On the right, no significant differences were found.
5.3. Cloze probability

Lastly, a model was built for Cloze probability. This last model mainly served as a control, because the N400 effect in sentences with low cloze-probability has been reproduced in many studies (Loerts et al., 2013; Delong, Urbach, Groppe, & Kutas, 2011; Block & Baldwin, 2010). This last model consisted of a fixed effect for direction (high or low Cloze probability), a smooth term for direction over time, and two random slopes: one for subject and one for target (or word). Results of this model can be seen in Table 7.

**Table 7: Estimates of Cloze sentences**

| Parametric coefficients: | Estimate | Std.Error | t-value | Pr(>|t|) |
|--------------------------|----------|-----------|---------|----------|
| 1 (intercept)            | -1.196   | 0.377     | -3.174  | <.01     |
| 2 DirectionLow           | 0.078    | 0.223     | 0.349   | 0.727    |

<table>
<thead>
<tr>
<th>Smooth terms:</th>
<th>EDF</th>
<th>Ref.DF</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 S(DirectionHigh)</td>
<td>1.003</td>
<td>1.005</td>
<td>26.927</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2 S(DirectionLow)</td>
<td>1.004</td>
<td>1.007</td>
<td>13.56</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3 S(Time, Subject)</td>
<td>27.622</td>
<td>117</td>
<td>1.193</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4 S(Time, Target)</td>
<td>162.943</td>
<td>1125</td>
<td>1.075</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

* p < .05
** p < .01
*** p < .001

R-Squared (adjusted): 0.077
Deviance explained: 6.37%

We can see in Table 7 that there was no significant difference between levels in the parametric coefficient structure (line 2). On the other hand, the smooths for both high and low direction sentences displayed significant differences (both p < 0.001). Again, this entailed that signal in both and low direction sentences differed significantly from 0, not necessarily that either differed significantly from the other. To see whether high and low Cloze items elicited different responses, their estimates and the difference between them was plotted (Figure 27).
Figure 3: Estimated effect of congruency over time for all Cloze sentences (left panel) with a difference plot (right panel) over the electrodes in the central ROI (measured in mV). There were no significant differences between Congruent and Incongruent sentences in the present sample.

As is visible in Figure 27, a slight negativity appeared for high Cloze items. This started around 500ms, and lasted until the end of the sample. This difference, however, was not significant.
6. Discussion

The present study investigated whether a mismatch between speaker and message would lead to an N400 effect (RQ1). Additionally, it was investigated if there were any differences in this incongruency effect between the Dutch and Frisian speaker utilized in our stimuli (RQ2). Lastly, this study aimed to investigate to what extent group membership (ingroup or outgroup) would influence the direction of this effect (RQ3). We expected that an incongruent utterance would lead to an N400 effect (Hypothesis 1).

Furthermore, we expected that this incongruency effect would also be visible between speakers (Hypothesis 2). For instance, we hypothesized that a high-status sentence uttered by a Frisian speaker (i.e. an incongruent item) would elicit a significantly larger negativity compared to the signal elicited by a Dutch speaker uttering a high-status sentence.

Lastly, we expected that this incongruency effect would be partially reversed in ingroup members (Hypotheses 3a & 3b). More specifically, for Hypothesis 3a we hypothesized that ingroup members would display a larger N400 amplitude for accent-congruent sentences that reflected negatively on their group (i.e. low-status Frisian for Frisian subjects, low-solidarity Dutch for Dutch subjects). For Hypothesis 3b, on the other hand, we expected that accent-incongruent stimuli that were positive about a particular group would elicit only a larger N400 amplitude in the outgroup (i.e. high-status Frisian for Dutch subjects, high-solidarity Dutch for Frisian subjects).
The raw signal over the central ROI seemed to show a slightly more negative N400 effect for incongruent status stimuli, as well as incongruent solidarity stimuli. However, when the models were built, no statistical evidence to support Hypothesis 1 was found, as there were no significant differences between congruent and incongruent items. The same was the case for solidarity items.

Evidence in favor of Hypothesis 2 was also extremely limited. For the status stimuli, a small (insignificant) negativity was found for incongruent Frisian items (high-status) relative to congruent Frisian items (low-status) that peaked around 1000ms. For solidarity stimuli, on the other hand, the incongruent Frisian items (low-solidarity) elicited a significantly larger sustained positivity compared to the congruent Dutch items (low-solidarity). This effect lasted from 500-1000ms. This was unexpected, as neither the sustained time-window, the latency, or the polarity match that of a typical N400 (e.g. Loerts et al., 2013). In the comparison congruent Frisian (high-solidarity) minus incongruent Dutch (high-solidarity), however, a sustained effect for accent-incongruent items was found: the signal for the Dutch speaker (i.e. the incongruent stimuli) was significantly more negative from around 250-1500ms. This window of significance does not seem to match with that typical of an N400. It seems then that the significant differences between the two speakers were not always modulated by congruency in the way that was expected. Indeed, while in one comparison the incongruent stimuli elicited a negativity effect, while in the other a positivity effect was found for incongruent items.
When looking at the three-way interaction Congruency x Accent x Group (Frisian/Dutch subj.), only limited evidence can be found for **Hypothesis 3a**. Frisian subjects who listened to low-status stimuli from a fellow Frisian speaker displayed a slight positivity effect around 500ms, although this was not significant. Dutch speakers hearing low-solidarity stimuli from an ingroup member, however, displayed a sustained negativity when compared to Frisian subjects. This difference became significant around 1300ms. Again, this sustained and late effect was far outside the 300-500ms window associated with the N400 (e.g. Balconi & Pozzoli, 2005; Kmiecik & Morrison, 2013).

Evidence for **Hypothesis 3b** was also non-existent. Moreover, contrary to expectations, Frisian subjects displayed a small (non-significant) and sustained negativity compared to Dutch subjects when listening to high-status (incongruent) sentences uttered by a Frisian speaker. Similarly, Dutch subjects listening to high-solidarity (incongruent) sentences uttered by a Dutch speaker displayed a sustained (non-significant) negativity compared to Frisian subjects.

Lastly, evidence for **Hypothesis 4**, our control model, was non-existent. Only a small (but again non-significant) positivity effect was found for low Cloze predictability items.

These results, to say the least, were unexpected. Van Berkum et al. (2008), whose study was similar to ours, reported a small but significant N400 for speaker-incongruent utterances. Before attempting to reinterpret the present study’s results, therefore, a more in-depth comparison to Van Berkum et al.’s study is in order. Firstly, van Berkum and colleagues manipulated three traits to measure difficulties in speaker-message integration: young versus old, upper versus lower class, and young versus old. Stimuli for these three traits were recorded by 16 speakers, and 5 additional speakers were also used for filler items. The present study, on the other hand, investigated two traits, high vs low-status and high vs low-solidarity, using two accents. The only difference in terms of procedure was that the present study asked participants to answer true/false questions regarding some of the stimuli. Additionally, the main difference in ‘other’ stimuli was that our filler-items were Cloze sentences, while for van Berkum and colleagues’ study the fillers consisted of sentences that were either semantically congruent or incongruent. Furthermore, they added a small number of filler items that were either congruent or incongruent with real-world knowledge (the same items used in Hagoort et al., 2004).

Van Berkum and colleagues only reported explicitly on one comparison: speaker-congruent versus speaker-incongruent items. However, they did compare the effects between the three traits they used in their study. No differences were found in this comparison, although incongruent items for male/female elicited a late posterior positivity. This was tentatively attributed to the fact that some items were not biologically possible (e.g. a man talking about having his period).
As we will come back to at a later point, this is likely a large attributor to our lack of results.

As mentioned before, a big difference between the present study and van Berkum et al.’s is the number of speakers used. It could very well be the case that the repetitive use of accent-incongruent items uttered by only four speakers in our sample (including the two accents not analyzed in the present study) reduced the N400 effect to such an extent that they became insignificant. Indeed, it has been reported before that repetition of the same type of items causes shortened and reduced N400 effects (e.g. Penney, Mecklinger, & Nessler, 2001; Besson, Kutas, & Petten, 1992).

Another important difference is that the speaker-inconsistencies in van Berkum et al. were much more rigid, and not as dependent on personal opinion. The stimuli used by them, for instance, included a young child talking about drinking a glass of wine. This utterance would be strange for any listener. While the materials for the present study were pre-tested to ensure that they were exemplary of the category they represented, the results still hinged on the participants’ a priori association with the speakers: if a participant personally did not think Frisian-accented speakers were nicer than Dutch ones, then they would not be surprised when they heard the Frisian speaker utter a low-solidarity stimulus. Indeed, if the (explicit) post-test ratings for the speakers are at least somewhat indicative of the participants’ implicit attitudes towards the speaker, participants did not perceive the speakers as was envisioned (i.e. Frisian as low-status but high-solidarity, and vice versa for Dutch). Lastly, because only one
speaker per accent was used, it would have been impossible to conclusively state that effects were caused by accent, as opposed to other speaker-characteristics.

6.2. Early negativity effects: an artifact of auditory stimuli?

An interesting finding we reported was that a long-lasting negativity was present for incongruent items in some comparisons. High-solidarity Dutch (incongruent) stimuli, for instance, elicited a significantly larger negativity than high-solidarity Frisian (Congruent) items from 244-1490ms, while low-Status (incongruent) Dutch items elicited a larger negativity than high-status (congruent) Dutch items in Frisian subjects (183-1490ms). This negativity increased until the end of the sample. This is not in line with typical N400 effects (e.g. Loerts et al., 2013). Similar results, however, have been reported for semantically anomalous stimuli in studies looking at spoken language (Hagoort & Brown, 2000; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003). Van Berkum and colleagues (2003) argued that the early onset of this effect, which they said was still functionally an N400, was caused by the presence of a broader context. Indeed, when an incongruent target was presented in isolation, the negativity appeared later. Hagoort and Brown, however, did not report an early N400 effect, but rather they distinguished a negativity around 200ms and a separate N400, following each other. They suggested that this early effect could reflect a pre-lexical selection tool that was activated when context was lacking. While this gives a potential explanation for the early onset of an effect, neither Hagoort and Brown or Van Berkum et al. (2003) discuss a possible reason for the sustained effects.
Applying this reasoning to our data, however, turns out to be impossible from the onset. Dutch subjects, for instance, were also found to display a larger negativity than Frisian subjects for high-status (congruent) Dutch stimuli (143-1289ms). This is unexpected, as we believed that Dutch would be seen as the high-status variant by default (e.g. Grondelaers et al., 2011). Therefore it was believed that no difference in signal should have been present in this comparison. Since this direction is the opposite of what was expected, another explanation for this sustained negativity is in order.

6.3. The sustained negativity as an measure of stereotype congruency?

Another study focused specifically on this sustained negativity, which they called the late posterior negativity (LPN), in the context of gender stereotypes (Leynes, Crawford, Radebaugh, & Taranto, 2013). The researchers utilized a male and female speaker who uttered typically masculine and feminine words. In the stereotype consistent condition (i.e. speaker and utterance match expectations), subjects displayed a much larger negativity than in the stereotype inconsistent condition. This LPN effect grew in amplitude until approximately 1400ms post stimulus. The researchers argued that these late effects reflected more conscious access to stereotypes upon activation, as opposed to immediate unconscious activation from the semantic memory. This finding was later reproduced, with a manipulation of focus encoding. Participants who were asked to focus on how the word made them feel showed a large sustained negativity, while participants who were asked to focus on how the speaker felt when they uttered the word displayed a much smaller effect (Leynes & Nagovsky, 2016). The authors argued that “self-focused encoding increased dependence on
stereotype information” (p. 171). On the surface, then, this theory seems to be a more appropriate explanation of the present findings. Indeed, high-status (congruent) Dutch items elicited an insignificant but visible sustained negativity compared to low-status Dutch items in Dutch subjects. This could potentially reflect self-focused encoding that was invoked involuntarily. Similarly, low-solidarity (congruent) Dutch items also elicited a significantly larger negativity than low-solidarity (incongruent) Frisian items that lasted from 505-987ms.

Again, though, this does not explain why the effect was only found in a limited number of comparisons. Furthermore, this explanation cannot explain why high-solidarity (incongruent) Dutch items elicited a significantly larger negativity than high-solidarity (congruent) Frisian items in this time-window. Overall, our data seems to contain conflicting results to the point where they cannot be unified into a fitting theory.

6.4. A more likely explanation

A much more probable explanation for our results, however, is that the present sample had a lack of data and too many variables. In order to validate our data, we added a control condition consisting of Cloze sentences. The N400 component in these types of sentences is reported often and seems to be at least somewhat more robust than the effect that was expected from accent-incongruent items (Loerts et al., 2013; Delong et al., 2011; Block & Baldwin, 2010). Therefore, we expected an N400 effect from the low Cloze condition, regardless of accent. No such effect was found in our data, however. In fact, no significant differences in signal were found between high and low Cloze probability items. This strongly suggests that our data was simply too noisy. Indeed, looking back at Figure 7, a
large amount of variation can be seen between participants. Some clearly displayed N400 effects for incongruent items (e.g. subject 18), others showed a sustained negativity for incongruent items (e.g. subjects 12 and 23), whereas yet others displayed the opposite pattern of what was expected (e.g. subject 21). When comparing the combined raw data, however (Figure 4), a small but visible negativity can be seen for incongruent items in both the status and solidarity conditions.

One could argue, based on the fact that a tiny difference was found only in the raw data, that ERPs are not suitable for measuring implicit associations altogether. Multiple studies, however, have reported N400 components for stereotype-incongruent stimuli (e.g. Van Berkum et al., 2008; White et al., 2009; Hehman et al., 2013). We believe that the lack of effects arose because the previously mentioned studies differed from ours in two ways.

Firstly, the speakers who recorded the stimuli were chosen because we believed they represented their groups well. We assumed that as a result of this, implicit associations would be present for both the Dutch (Heijmer & Vonk, 2002) and Frisian accent (Hilton & Gooskens, 2013). It turned out, however, that subjects had considerable difficulty with correctly identifying the Frisian speaker (6/25 participants guessed Frisian correctly in the post-test, half of whom were Frisian themselves). Evidently, then, a specific regional accent is not as easily identifiable as gender (White et al., 2009) or race (Hehman et al., 2013). As a result, the associations a participant had with that particular accent could have been weak or non-existent, simply due to a lack of familiarity. In turn, this lack of associations could have been reflected in the lack of N400 effects. Indeed,
Hehman and colleagues found that participants who were negatively biased against African Americans displayed significantly larger N400 effects for black-incongruent items compared to those with relatively little bias.

Lastly, the present study differed from the studies mentioned above in that a comparison between ingroup and outgroup members was made. This was done based on the assumption that ingroup members would respond differently to their own stereotypes than outgroup members would, and vice versa (e.g. Bishop et al., 2005; Tajfel, 1974). Since we only expected a small effect size for incongruent stimuli to begin with (cf. Van Berkum et al., 2008), a larger sample of Frisian subjects would have made our study more powerful (VanVoorhis & Morgan, 2007). In this case, however, the Frisian participants acted more as a confounding variable, because they were the few people who recognized the Frisian accent.
7. Limitations

Our results were very inconsistent, and mostly not in line with the hypotheses that we posed. Therefore, it is even more important than normal to discuss the shortcomings of the present study. As mentioned before, the most important limitation is the small sample size (N=25, 5 participants discarded, 4 Frisian participants, 16 Dutch participants). Like many EEG studies, the present research was highly underpowered. The study by van Berkum et al. (2008), which this study was heavily based on, reported an N400 effect for speaker-incongruent stimuli, but this negativity was much smaller than for semantically-incongruent items. This lack of participants is likely the largest contributor to the lack of consistent effects (Boudewyn, Luck, Farrens, & Kappenman, 2018; Vanvoorhis & Morgan, 2007). This became especially clear in the Cloze sentences. Since this effect is typically more robust than the expected effect of speaker-incongruency, (e.g. Delong et al., 2011; Block & Baldwin, 2010), we expected a clear difference between high and low Cloze sentences. More specifically, the stimuli that were used for the Cloze stimuli were taken from a study that did find an N400 effect for low Cloze sentences (Loerts et al., 2013). However, no such effect was visible in the present dataset. Again, this suggests that our data is too noisy at the moment to make any meaningful assumptions regarding speaker-message integration.

Another, admittedly unexpected, limitation to our present dataset, was that the regional Frisian accent was rated significantly lower on solidarity traits than the standard Dutch accent. Based on previous studies, it was taken as a given that the regionally flavored accent would be rated higher than standard
Dutch (Grondelaers et al., 2010) or at least equally high as standard Dutch (Hilton & Gooskens, 2013) on solidarity traits. On the one hand, this could mean that Wilkinson's (1965) framework needs to be re-evaluated for the Netherlands. On the other hand, it could merely imply that the Frisian speaker utilized in the creation of stimuli simply did not sound like a 'nice' person. Since only one speaker was utilized per accent, however, it is not feasible to make any meaningful speculations regarding this difference.

In a different vein, a large number of potentially confounding variables were not taken into account during the analysis. While random effects were included for both speaker and item, no such controlling variables were added in our models for word-frequency, word-length, speaker ratings per participant, sentence ratings from our pre-test, stimulus number, etcetera. It would have been entirely possible to add these to our models; GAMMs, however, are as computationally expensive as they are sophisticated. Furthermore, doubling the number of cores used during the analysis does not necessarily mean that computation time is halved. In our experience, every random effect that was added to the analysis increased the computation time by a factor of 10. The models utilized in our method took ten to thirty minutes to run, and they included every possible measure to decrease computation time. It was therefore decided that model complexity should be kept to a minimum to reduce computation time. Subsequently, a number of confounding variables were not taken into account. As a result, a number of possibly interesting effects, such as those of word-length and word-frequency were not investigated at all.
Lastly, for the present study it was assumed that all Frisian subjects unequivocally deemed themselves members of their ingroup, and that all Dutch speakers saw Frisian speakers are outgroup members. A look at the post-test data, however, shows that a portion of non-Frisian participants deemed themselves non-standard Dutch speakers (see Appendix D). As this was out of the scope of the present study, this was not investigated in our analyses. However, considering the strong association of standard language ideology and Randstad-Dutch (e.g. Grondelaers et al., 2011), one could entertain the idea that all participants with a non-standard accent may consider Randstad-Dutch an outgroup.

8. Suggestions for further research

While our data were mostly not in line with our hypotheses, the present study should not be discredited, altogether. As discussed in the limitations section, our biggest shortcoming was the small sample size. Therefore, we propose that any follow-up study increases the sample size. Additionally, adding an interaction for group membership may have been too ambitious, considering the lack of literature replicating an N400 in speaker-message incongruities. A better practice for further research, therefore, would be to reduce variation in the subject pool by only testing those from a specific region of the Netherlands. Alternatively, a more easily recognizable and more heavily stigmatized accent, such as Moroccan-Dutch (e.g. Nortier & Dorleijn, 2008), could be used.

Additionally, in order to control for speaker-characteristics that are not related to accent, multiple speakers should be recorded per accent (as per Van Berkum et al., 2008). Even more ideal would be to index associations regarding
several accents in the Netherlands on a larger scale. While accent-evaluation studies up until now (e.g. Hilton & Gooskens, 2013) have shown attitudes towards some regional or non-standard accents, a successful mega-study on this subject (cf. Coupland & Bishop, 2007) has not been done.

Another shortcoming of the present study was the lack of control variables due to computation times. A later analysis should at the very least include measures for word-frequency, as this could simply not be kept constant due to the large number of stimuli utilized. Previous studies have indeed shown that low-frequency words elicit significantly larger N400 components than high-frequency ones (Van Petten & Kutas, 1990; Dambacher, Kliegl, Hofmann, & Jacobs, 2006).

Furthermore, EEG studies are notoriously time-consuming, as it can take a considerable amount of time to fit the cap, fill the 72 electrodes, and get signal impedance down to acceptable levels. Therefore, other possibilities could be considered. Eye-tracking, for instance, offers a similar ‘window into the mind’ as EEG, while having a preparation time of barely five minutes. For instance, multiple studies reported a positive correlation between surprisal and pupil-size (Frank & Thompson, 2012; Alamia, VanRullen, Pasqualotto, Mouraux, & Zenon, 2019). Furthermore, previous work has already found that pupil-size can reflect own-race bias (Wu, Laeng, & Magnussen, 2011), and racial bias in a threat-assessment situation (Fleming, Bandy, & Kimble, 2010). While no studies have investigated the effects of mismatches between speaker and message on pupil size, a similar effect could potentially be found, considering that surprisal effects can easily be measured with this method.
9. Conclusion

The present study aimed to investigate if implicit attitudes towards accented speakers could be measured using the N400 component. Furthermore, we were interested in whether ingroup members would respond differently to their own stereotypes than outgroup members. While some differences between accents and groups were found, these were often conflicting with one another. Therefore, the present study did not find any evidence in favor of using the N400 component to measure these implicit attitudes. However, it must be stressed that the present study was heavily underpowered due to a lack of participants. ERPs, therefore, should not be discredited immediately as a potential method for measuring implicit associations.
References


Audacity Team (2019). Audacity: Free Audio Editor and Recorder (version 2.3.2) [Computer software]. Retrieved from https://audacityteam.org


Fleming, K. K., Bandy, C. L., & Kimble, M. O. (2010). Decisions to shoot in a weapon identification task: The influence of cultural stereotypes and perceived threat on false positive errors. *Social neuroscience, 5*(2), 201-220.


Skidmore, S. T., & Thompson, B. (2013). Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behavior research methods, 45*(2), 536-546.


*Social psychology quarterly, 63*(3), 224-237


*Cognitive, Affective, & Behavioral Neuroscience, 17*(3), 475-490.


Appendices

Since a large number of stimuli were used for the present stimuli, the supplemental materials have been uploaded online. These can be accessed at:

https://drive.google.com/open?id=1Zxoe2fQ90Ein5j4vGmHJZKxmHHeBYsMO

Alternatively, a shortened link can be found here:


The following appendices are available:

Appendix A: Overview of the stimuli
Appendix B: Ratings per sentence
Appendix C: Overview pre-test questionnaire
Appendix D: Overview and results post-test questionnaire
Appendix E: Consent form
Appendix F: Information letter
Appendix G: R Scripts for GAMMs
Appendix H: Other supplemental materials used in analysis

- Data per condition (Central Cloze, Central Solidarity, Central Status, All ROIs Solidarity, All ROIs Status, Subject data)
- Preprocessing R script
- Loading data in a single dataframe
- Exploratory plots on raw data per subject/ROI (for status and solidarity)
- Wilcoxon-signed rank tests on speaker ratings