The adoption of sound change : synchronic and diachronic processing of regional variation in Dutch

Voeten, C.C.

Citation

Voeten, C. C. (2020, October 13). *The adoption of sound change : synchronic and diachronic processing of regional variation in Dutch*. *LOT dissertation series*. LOT, Amsterdam. Retrieved from https://hdl.handle.net/1887/137723

Author: Voeten, C.C.
Title: The adoption of sound change : synchronic and diachronic processing of regional variation in Dutch
Issue Date: 2020-10-13

The adoption of sound change

Synchronic and diachronic processing of regional variation in Dutch

Published by
LOT
Kloveniersburgwal 48
1012 CX Amsterdam
The Netherlands

phone: +31 20 525 2461

e-mail: lot@uva.nl
http://www.lotschool.nl

Cover illustration: edited version of an unknown painting of the Stratense Molen in Oirschot, probably by Piet Teunissen (1890–1958).

ISBN: 978-94-6093-364-6
NUR: 616

The adoption of sound change
Synchronic and diachronic processing of regional variation in Dutch

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 13 oktober 2020
klokke 13.45 uur

door

Cesko Cis Voeten

geboren op 14 december 1992
te Nijmegen

Promotores: prof. dr. C.C. Levelt
prof. dr. Y. Chen

Promotiecommissie: dr. K. Chládková
(Karelsuniversiteit Praag)
prof. dr. J. Grijzenhout
dr. W.F.L. Heeren
prof. dr. H. Van de Velde
(Fryske Akademy & Universiteit Utrecht)

The research reported here was supported by the Netherlands Organization for Scientific Research under project number PGW-15-15.

Contents

Acknowledgments

While you write a PhD dissertation by yourself, you're never writing it alone. This dissertation and the research carried out as part of it would not have been possible, had it not been for the support, advice, and friendship of many people in and out of LUCL. This starts with Haike Jacobs and Roeland van Hout, who supervised my education as a linguist before this PhD, and with whom I wrote the NWO "PhDs in the Humanities" research proposal that funded my PhD position. It is in part thanks to Haike's and Roeland's curiosity after I had written my BA thesis with them that this book (or PDF[1]) is in front of you now. I owe them—as well as Janine Berns and James McQueen, who also contributed significantly to the development of the proposal—a debt of gratitude. I also thank Marc van Oostendorp for his acting as the project's principal applicant in Leiden. And, of course, I am immensely grateful to my wonderful promotors, Claartje Levelt and 陈轶亚 (Yiya Chen). They probed me with continuous constructive criticism (although sometimes in vain—we still don't agree on the boundary between phonetics and phonology!), helped me give direction to my research, encouraged me to develop my research profile in an ever more quantitative direction, and at times provided emotional support when the going got tough. Thank you for everything.

Within LUCL, I was fortunate enough to be surrounded by many intelligent people. My fellow PhD-student phoneticians: 史濛辉 (Menghui Shi), 杨青 (Qing Yang), 刘敏 (Min Liu), 毕一飞 (Yifei Bi), 王曼 (Man Wang), 李倩 (Qian Li), 邹婷 (Ting Zou), Daan van de Velde, Laura Smorenburg, Meike de Boer.

---

[1] In case you were expecting a probability density function rather than a portable document file, congratulations! You and I think alike. Here is the PDF for the scaled-*t* distribution, which in my experience is much more commonplace than the normal distribution:

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\sigma^2}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$$

I have benefited greatly from your intellectual challenges and counsel. Saskia Lensink, my fellow quantitative linguist, helped me become a better statistician and gave me the idea for the analysis in Chapter 2. My fellow psycholinguists: Bobby "Bobski" Ruijgrok and Olga "Olgie" Kepinska were exceptionally brainy influences on my research, and were also the first users of my R package `permutes`, together with Elly "Elly" Dutton. Thank you for sharing the labs with me, and of course thanks to Jos Pacilly for his work as lab technician (and for not switching the PCs from Windows 7 to Windows 10 in the middle of my longitudinal experiments!). My fellow phonologists: Xander Vertegaal, Rasmus Puggaard, 胡瀚 (Han Hu), and 郑婷婷 (Tingting Zheng—face it, you're a phonologist now!). I have consulted each of you more than once when I was stuck with a hard-to-explain point of data, and all you asked in return was some statistical advice. Without your thoughts and suggestions and those by Sarah von Grebmer zu Wolfsthurn, the explanation of the data in Chapter 5 would not have been as coherent. The speakers who lent their voices for my experiments: Aliza Glasbergen-Plas (Chapters 3, 4, and 6), Lisette Jager (Chapter 5), and Daan van de Velde (instructions in Chapters 3 and 6). Thank you for your contributions. 程航 (Hang Cheng), Rasmus Puggaard, and 吴疆 (Jiang Wu): thank you for your variegated advice (i.e. for tolerating my silly questions) concerning such vital issues as hyphenation and the placement of arrows and tildes inside or outside of phonetic brackets. Μαρίνα Τερκουράφη (Marina Terkourafi), you have freely provided valuable mentorship to me at a few key points along my academic career; I am deeply grateful for that. Finally, thanks to Laura Smorenburg, Meike de Boer, Sanne Ditewig, and Willemijn Heeren for including me in their reading group, even though some of the forensic stuff went a bit over my head sometimes!

Outside of LUCL, I would not have been able to do without the expertise of at least Hans Van de Velde, Silke Hamann, and Klaas Seinhorst. Hans, thank you for your continuous encouragement and feedback the many times we met at conferences, and for the role you have had multiple times in the development of my academic career (including landing me a campus visit at Michigan State University). Silke and Klaas, thank you for the reading group on sound change, and for more than a few pleasant visits to the UvA. Our intellectual exchanges helped me on my way during an uncertain time in my PhD. I am also indebted to Harald Baayen, whom I visited in Tübingen to discuss the analyses in Chapter 2, and to Rob Hartsuiker and Wouter Broos, who helped me get set up in Ghent for the experiment in Chapter 4. And thanks to too many people to name for their helpful advice or challenging discussions at conferences or elsewhere. Finally, of course, thanks to my parents, my grandmother and late grandfather, my brothers Kaz and Victor (especially Kaz for patiently walking me through some math), and our dog Floortje. Your love and support are invaluable.

CHAPTER 1

Introduction

In a seminal paper that determined the agenda of sociolinguistics for decades to come, Weinreich, Labov, & Herzog (1968) identified five major problems that need to be solved in order to explain the phenomenon of linguistic change. The *actuation problem* concerns the initiation of change: why does a certain change occur in a certain language at a certain point in time, but not in another language or at another point in time? The *constraints problem* seeks to identify the sets of possible and impossible changes, and their structural conditions. The *embedding problem* situates an individual change within its larger linguistic and social context, and the closely-related *evaluation problem* discusses the social meaning of a change. The final problem, the *transition problem*, is the focus of this dissertation. In its original formulation, the goal of the transition problem is to identify the pathway through *linguistic structure* by which a change progresses. An example, taken from Scheer (2014), is the change from [l] to [ʁ] in intervocalic position in the Genoni dialect of the Italian peninsula of Sardinia. Synchronically, this change is "crazy" (Scheer 2014), in the sense that it is phonetically unmotivated and hence appears unnatural. The same is true diachronically: a historical change l → ʁ / V_V is "crazy". However, Scheer (2014) argues that this change has actually arisen quite plausibly, via an intricate chain of l > *ɫ > w > gʷ > *ɣʷ > ʁ. The transition problem formulated in Weinreich, Labov, & Herzog (1968) deals with establishing these types of plausible historical derivational chains.

There is, however, a second type of transition that Weinreich, Labov, &

Herzog (1968) do not discuss in detail. Just as the embedding problem is (correctly) split into "embedding in the linguistic structure" and "embedding in the social structure" (Weinreich, Labov, & Herzog 1968:185), the idea I wish to put forward is that the same distinction should be made for the transition problem. That is, a distinction should be made between the transition throughout *linguistic structure* (corresponding to Weinreich, Labov, & Herzog's 1968 original formulation of the transition problem) and the transition throughout *social structure*, i.e. the speech community. The concept of such a split was present in Weinreich, Labov, & Herzog (1968): situated within the context of the 1960s, the transition of a change throughout social structure was in fact wholly incorporated in Weinreich, Labov, & Herzog's (1968) evaluation problem. With the present-day knowledge and (statistical) methods available, however, an addendum is necessary. Linguistics in the 21^{st} century is much more concerned with individual differences than it was in the 1960s, and recent advances in statistics such as mixed-effects models or the more flexible generalized additive mixed models make it possible to explicitly take into account such individual variation at no additional methodological cost. These advances in the field make it possible to look at the transition of novel linguistic structure not just throughout the social community, but also throughout the *individual members* of that community. This individualized version of the transition problem is concerned with the individual language user as a processor of spoken language: how do you pick up a language change, and what is the time course involved? This is the question this dissertation aims to answer.

Throughout the remainder of this introductory chapter, I will refer to this narrowly-scoped variant of the transition problem as the *adoption problem*. This dissertation investigates the adoption problem by taking advantage of a set of sound changes that are currently on-going in Dutch. These are introduced in Section 1.1. Section 1.2 then discusses the origin and actuation of sound change within a single individual and the large-scale propagation of sound change throughout the community, two large issues between which the adoption problem is perfectly sandwiched. Section 1.3 discusses the extent to which these theoretical notions are in line with psycholinguistic and computational models of human speakers and listeners, and also goes into some important methodological innovations that make it possible to perform the psycholinguistic experiments used in this dissertation. Finally, Section 1.4 concludes the introduction of this dissertation with an overview of the remaining chapters.

1.1 The Polder shift: an on-going vowel shift in Dutch

A sound change that has been on-going in Dutch for approximately a hundred years by now is the diphthongization of the tense mid vowels /eː,øː,oː/ towards present-day [ei,øy,ou]. Because this sound change is relatively recent, it has been well-described (Adank, van Hout, & Smits 2004, Adank, van Hout, & Van de Velde 2007, Van de Velde 1996, Van de Velde, van Hout, & Gerritsen 1997, Van de Velde & van Hout 2003, Zwaardemaker & Eijkman 1924). Of particular note are Zwaardemaker & Eijkman (1924), who were the first to notice (and express their disapproval of) a small offglide developing in the vowel /eː/, probably realized as something approaching [eːʲ]. A detailed account of subsequent events that took place between 1935 and 1995 is provided by Van de Velde (1996) on the basis of historical radio recordings. He shows that the adoption of the diphthongization resembles a logistic curve—the S shape that is typical of on-going language change. In 1935, the vowels are diphthongized only negligibly, in 1965, the average diphthongization is about 50%, and in 1995, nearly all realizations are fully diphthongized. Later synchronic measurements by van der Harst (2011) confirm that, as of the twenty-first century, the tense mid vowels have indeed become genuine diphthongs. This is typical for processes of language change: in the beginning of a change, only a few individuals share a linguistic innovation; at some point, these individuals spread the innovation to their peers and, perhaps most importantly, their children, causing a rapid ascent in the adoption of the change; finally, the (usually older) people who have not acquired the change either manage to acquire it, or pass away. This yields an S-curve with exponential growth at the beginning and exponential decay at the end. The fact that the diphthongization of the Dutch tense mid vowels follows exactly such a curve suggests that this is, indeed, a language change in progress.

The diphthongization of the tense mid vowels went hand in hand with a second change, namely the lowering of the original diphthongs /ɛi,œy,ɔu/, particularly /ɛi/ (Blankestein 1994, Gerritsen & Jansen 1980, Gussenhoven & Broeders 1976, Jacobi 2009, Mees & Collins 1983, Stroop 1992, Stroop 1998, Van de Velde 1996, Voortman 1994), commonly referred to as "Polder Dutch" (Stroop 1998). The relatedness of these two changes is obvious, but the causal connection is debatable. Stroop (1998) suggests that the phonetic lowering of /ɛi/ initiated a drag chain that attracted the tense mid vowels' nuclei, causing the intrinsic tendency towards slight diphthongization of these vowels to become extrinsicized. Another option, however, would be to postulate a push chain, where the diphthongization of /eː,øː,oː/ was the original innovation, which then pushed the lax diphthongs /ɛi,œy,ɔu/ out of the way (vi-

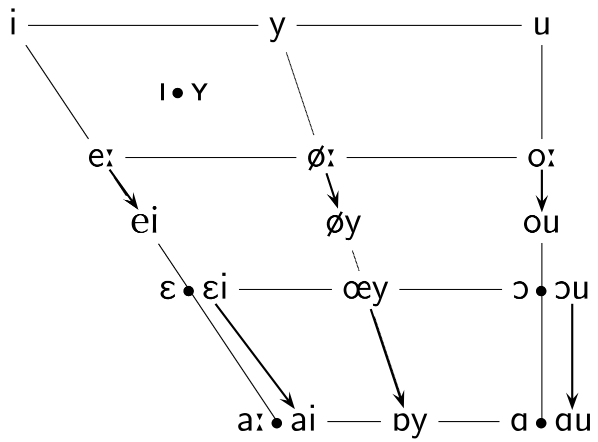

Figure 1.1: Vowel diagram showing the changes constituting the on-going vowel shift. The arrows indicate the diachronic changes; secondary arrows to indicate upgliding diphthongization are not included to prevent cluttering the diagram.

sualized in Figure 1.1). This latter option seems more credible for three reasons. First, the diphthongization of the original /eː,øː,oː/ was first observed in 1924 (Zwaardemaker & Eijkman 1924), while the lowering of the original diphthongs was only put to paper in 1990 (Jacobi 2009). Secondly, diphthongization of tense mid vowels is a natural phonetic development (Labov, Yaeger, & Steiner 1972, Watt 2000), which makes this change a plausible initiator of a chain shift, whereas the lowering of pre-existing diphthongs is more often the consequence of an earlier change in a chain shift (Labov 1994). Finally, Labov (1994:116) notes that "in chain shifts, the nuclei of upgliding diphthongs fall". If the diphthongization of [eː,øː,oː] was the first step in this chain shift, then the second step would be their nuclei falling, which would cause them to merge with the original, lax, diphthongs. Preservation of phonemic contrast would then naturally require these original diphthongs to move "out of the way", resulting in a push chain, as opposed to a drag chain; thus, Labov's observation indirectly makes a push chain more likely than a drag chain.

For as long as they have existed, the original diphthongs /ɛi,œy,ɔu/ have been subject to distributional restrictions. The canonical reference is Booij (1995), who notes that these diphthongs are realized as long monophthongs [ɛː,œː,ɑː] before coda /l/, presumably, Booij argues, for articulatory reasons. It might then come as no surprise that this very same restriction was also noted by Zwaardemaker & Eijkman (1924): according to them, the synchronic process of /eː/ → [eːʲ] was also blocked before coda /l/; later authors added /r/

(Gussenhoven 1993) and /ʋ,j/ (Collins & Mees 1999) to the list, completing the natural class of Dutch approximant consonants. The precise set of distributional restrictions is somewhat complicated; a comprehensive description is given in Voeten (2015). These distributional restrictions complicate the picture of the vowel shift described thus far, by adding a *phonological* dimension to the phonetic changes affecting these vowels.

The resulting vowel shift I will call the "Polder shift", in hommage to Stroop's (1998) "Polder Dutch" term. This shift seems to be an ordinary vowel shift, albeit one with a specific, known, historical track record. This is a rare fortune in historical linguistics. In addition, the Polder shift can be shown to recruit phonological knowledge rather than being a plain phonetic vowel shift (Voeten 2015). A final point, which makes the Polder shift especially suitable for the investigation undertaken in this dissertation, is that the Polder shift has resulted in significant sociolinguistic variation. Previous research has shown significant differences, particularly in the realizations of the tense mid vowels and diphthongs, between the Dutch spoken in the Netherlands and the Dutch spoken in Flanders, the northern half of Belgium; in particular, see Adank, van Hout, & Smits (2004), Adank, van Hout, & Van de Velde (2007), Van de Velde (1996), and Van de Velde, van Hout, & Gerritsen (1997), and Van de Velde & van Hout (2003), among others. Chapter 2 is devoted to a thorough investigation of the way the Polder shift has spread throughout these two communities of spoken Dutch. It will be shown that there are robust synchronic differences between Netherlandic Dutch and Flemish Dutch, which parallel the diachronic differences between modern-day Netherlandic Dutch and its state before the Polder shift. This observation forms the foundation for the remainder of the dissertation, in which these synchronic differences are exploited to perform a *synchronic* investigation of the adoption of *diachronic* change.

This dissertation investigates these on-going sound changes in Dutch on the basis of psycholinguistic experiments with speakers of Netherlandic Dutch and speakers of Flemish Dutch. In doing so, the dissertation aims to describe and explain the phonetic and psycholinguistic mechanisms underlying the adoption of sound change using a variety of methods. The following sections of this introductory chapter show that this is a particularly necessary enterprise: currently, historical phonologists and psycholinguists are at odds with one another in their explanations of how humans process variation, and therefore how sound change can be actuated, adopted, and transmitted.

1.2 Theories about the lifecycle of sound change

1.2.1 Misperception as a source of sound change

Historical approaches to sound change tend to focus on the role of the listener in their interaction with the speaker. Thus, the most oft-cited source of the origin of sound change is *misperception*. Bermúdez-Otero (2007) calls a misperception event a "coordination failure" between a speaker (henceforth "S") and a listener (henceforth "L"). A coordination failure occurs when L perceives S's realization of a hypothetical category /A/ as [A′] rather than [A]; L is then assumed to reanalyze his representation of /A/ as /A′/ when he himself becomes the speaker in future communication events. A couple of remarks on this core proposal are in order. First of all, the present formulation of this proposal leaves ambiguous whether S actually *realized* [A′], or whether this is merely what L *heard* (sensorily) or *perceived* (after parsing the sensory input as a phonetic category). Classic accounts such as Hyman (1976) or Ohala (1981) assume that [A] and [A′] are actually the same speech signal, but that this speech signal is misparsed by L. To reuse the example used in Hyman (1976): if S plans to realize a syllable /bā/, S's F0 will be slightly lowered at the onset of the /ā/ vowel for unavoidable reasons of human anatomy, resulting in [bā̂]; if L incidentally fails to perceptually compensate for this intrinsic effect, they will perceive /bā̂/. The theoretical account, however, does not require that what L perceived is actually what S *produced*; a simple mishearing on the part of L will produce the same result. In fact, there are at least three different types of misperception, as summarized by the three pillars of Blevins's (2004) Evolutionary Phonology framework. Misperception of the Ohalian kind, whereby L mis-partitions the intrinsic and extrinsic sources of variation in the speech signal, is considered "CHANCE" in Evolutionary Phonology. This classic type of misperception is more generally called "hypocorrection", following Ohala (1989) and Lindblom et al. (1995). The second type of misperception in Evolutionary Phonology terminology is "CHANGE", which corresponds to Ohala's (1989) and Lindblom et al.'s (1995) "hypercorrection". This situation is the opposite of CHANCE, in that rather than incorrectly attributing intrinsic variation to an intended gesture by S, L now thinks that part of the intended variation by S was actually *not* intended, thus "overparsing" the speech signal. Finally, sound change due to "CHOICE" takes place when L hears S produce multiple phonetic variants of the same word, and reconstructs a different underlying form for this word than S had intended. This is a type of drift in the statistical distribution of word tokens: the mode of the distribution shifts slightly towards a different variant.

The view that sound change has its basis in misperception is attractive, because it is obvious that a language-acquiring child must perceive before it can produce. If authors like Labov (2007) or Hamann (2009) are correct that non-superficial reanalyses can only be made in first-language acquisition, then the driving force of sound change must be one of two things. On the one hand, it is possible that sound change is due to literal misperception (Ohala 1981), which leads to reanalysis by the language-learning child (Labov 2007). However, because it is this reanalysis that is the critical step in the acquisition, it can also suffice to single out this step of reanalysis. In this case, which is the interpretation by Beddor (2009) and Hamann (2009), the language-learning child does not make an error in perception, but rather uses a grammar with different cue weights than those of the adult speaker, thus arriving at a different phonetic interpretation of the speech signal produced by the adult. Chapter 5 provides evidence that this is more likely to be correct than Ohala's (1981) misperception account, although it will be shown *not* to hinge on children as the actors (as already argued by Bybee & Slobin 1982). This is problematic for generative theories of sound change, since these assume that grammatical restructuring after the critical period is not possible due to the inaccessibility of UG. Recent evidence by Pinget, Kager, & Van de Velde (2019) (also available in Pinget 2015) demonstrates that this assumption is unwarranted. Their results show not only that adults are capable of changing their sound systems just as much as children are (a well-known fact outside of generative linguistics, which will be revisited in Chapter 4), but also that there are different roles for perception and production. On the basis of two on-going mergers in Dutch, Pinget, Kager, & Van de Velde (2019) show that when changes are incipient, perception goes first: individuals need to perceive a change (via, e.g., a coordination failure of the Ohalian kind) before they will produce it. Depending on each individual's perception–production link, they may subsequently continue to spread the change among their peers, thereby pushing the change from the incipient into the on-going stage; see Coetzee et al. (2018) for detailed discussion of this phase of the process. Finally, when the sound change reaches the advanced stage and nears its completion, Pinget, Kager, & Van de Velde (2019) show that the perception–production relationship reverses, such that individual speakers come to produce sound systems in which the change has completed, but are still able to draw on any remaining fine phonetic detail in perception left in place by the old system. Curiously, this last step has also been observed to reverse: Labov, Yaeger, & Steiner (1972) report on the phenomenon of "near-mergers", where individuals fail to perceptually differentiate between two sounds that are involved in an on-going merger, yet consistently produce such differentiation in their own productions.

A corollary of a central role for misperception must be that "mini sound changes" are actuated all the time. In fact, if misperception is the sole mechanism behind sound change, Weinreich, Labov, & Herzog's (1968) actuation problem has been solved: a certain sound change actuates at a certain point in time due to a misparsing by L of a certain kind, which subsequently becomes entrenched in L's grammar and is then spread to other individuals. The bigger question is then that of transmission: how is this incidental reanalysis transmitted, by L turned speaker, to the other members of their community? Yang (2009) provides a mathematical model of what is required, and solidifies his claims on the basis of real-world data by Johnson (2007). His answer is very simple: transmission takes place if at least 21.7% of the input consists of novel forms. However, what is going on between 0 and 21.7%? If the transmission of sound change beyond those 21.7% is reliant on the listener (as Ohala 1981 would claim), it could be that the first 21.7% have to be actuated by the speaker. Section 1.2.2 discusses this.

1.2.2 Speaker-induced sound change and the role of the representation

Research on the speaker as a possible source of sound change is, perhaps surprisingly, rather scarce. This can be understood when viewed in light of the structuralist tradition of phonology in which the lion's share of prior work on sound change has been done: under structuralist views, interlocutors both possess discrete categories, and an incidental pronunciation change by a speaker in continuous ("analog") acoustic space cannot initiate a sound change at the discrete ("digital") level. The listener, on the other hand, can initiate sound change of the Ohalian kind by erroneously creating a novel category or merging existing categories during a serendipitous misperception event. If the speaker is to play a substantive role in sound change, it must be on the continuous level of phonetics, a domain which is implicitly neglected in structuralism-inspired generative views on phonology.

Various attempts have been made to integrate continuous phonetics with discrete phonology. In the context of historical phonology, the theoretical necessity of an integration of both levels into theories of sound change was most clearly argued by Hyman (2013). Later authors, such as Bermúdez-Otero (2015) and Ramsammy (2015) provide explicit models taking this into account. These papers appear to have been implicitly influenced by Boersma's (2011) BiPhon model, as they incorporate the same major building blocks, save for the distinction between the Auditory Form and the Articulatory Form. As a framework for modeling sound change not just by the listener, but also by the speaker, BiPhon is a particularly relevant candidate, because it is *bidirectional*, as opposed to most models of sound change, which are (often implicitly) feed-forward. This is a problem, because there is psycholinguistic evidence (see Section 1.3) that, for example, a misperception by L can be counteracted by L's lexical knowledge of what S is likely to have said given the discourse context. Accomplishing this either requires the ability to move back and forth between different levels of representation in the model, i.e. feedback, or requires that the various sources of information are accumulated and a decision is only made in a single evaluation at the very end of the model. The latter follows directly from the parallel-OT architecture of BiPhon, but by definition cannot be obtained by feedforward models operating on ordered transformational rules.

BiPhon's parallel architecture offers another advantage over traditional feedforward models, because it models the listener and the speaker in exactly the same way: to change the role of the listener into that of the speaker, simply traverse the model in reverse. Thus, where the listener starts from a speech signal as input and interprets the optimal underlying form, the speaker starts from an underlying form and outputs the optimal speech signal, without requiring any alterations to be made to the grammar or its evaluation. As noted by Boersma (2011), this automatically bidirectional architecture generates specific predictions concerning sound change. If a BiPhon listener encounters variation of the Ohalian kind (i.e. variation due to anatomical differences or otherwise speaker- or condition-specific variation), termed "transmission noise" by Boersma (2011), said listener will acquire a broader distribution of possible auditory values for the sound(s) in question. This may over time lead to drift of the means of these distributions towards the novel realizations. However, because the listener will eventually also need to speak, this drift is counteracted by Boersma's (2011) mechanism of *prototype selection*. Prototype selection is the process of finding the optimal auditory form given a certain phonological surface form. As shown in simulations by Boersma & Hamann (2008), the increased variance in auditory values caused by transmission noise is counteracted by the fact that more extreme values, and hence large deviations from an established mean value, are more difficult to realize. The reason is partly in universal anatomical restrictions, but also lies in the fact that an individual's repertoire of motor programs will have been optimized for the original values. This causes the adoption of Ohalian variation to be maladaptive from the perspective of the speaker. Therefore, given that BiPhon models speaking and listening using the same grammar, a sound change can only be obtained if it is acquired by the listener via, e.g., Ohalian misperception *and* does not present problems for the speaker.

1.2.3 Exemplar Theory

An altogether different approach to integrating the speaker and the listener into the same model is to abandon the classic structuralist model of phonology and branch out to a new type of models. This step has been taken by Exemplar Theory (Pierrehumbert 2001). This brings us back to the possible types of sound change discussed in Bermúdez-Otero (2007), who argues that exemplar models are in fact the only theoretical device available that can predict sound change that is both phonetically gradual and lexically gradual. The main claim of exemplar models is that our minds do not contain discrete phonological categories, but rather store (in full detail) the raw acoustic data with which we are provided as listeners. As this information passes through short-term memory, working memory, and long-term memory, it gradually decays (unless reactivated by, for instance, an attempt on the part of the speaker to reproduce the utterance). Decay can be considered the loss of fine phonetic detail, and the subsequent consolidation of different exemplars into a single prototype. Subsequent examples entering into long-term memory become absorbed into this prototype, and influence it by being averaged with the prototype exemplar; thus, many realizations which are accepted as belonging to a certain prototype but are not exactly equal to it will slowly result in drift. When the listener turns into a speaker, this same prototype will be used to generate speech, and thus new examples. Hence, slow drift in perception begets slow drift in production.

Bybee (2002) provides a general overview of how Exemplar Theory can be used to describe a specific type of sound change, viz. *reductive* sound change (reduction of segments or of individual articulatory gestures within segments, leading to lenition). In the case studied by Bybee (2002), /t,d/ deletion in American English, the reductive sound change is both phonetically and lexically gradual, and hence requires a framework such as Exemplar Theory to be modeled successfully. Bybee's (2002) idea is along the following lines. Words that are predictable, i.e. high-frequency words, are particularly good candidates for reduction: their inherent predictability means that for a correct interpretation, they are less dependent on the redundancy that is inherent to a full pronunciation. If, following Levelt, Roelofs, & Meyer (1999) and Wheeldon & Levelt (1995), words are viewed as highly-trained motor programs, and if humans desire to minimize their articulatory effort whenever possible (Passy's 1890 "Economy" principle), then reduction can be expected to take place particularly in these high-frequency words. By the tenets of Exemplar Theory, these words will then have a high proportion of reductions stored within their corresponding exemplar clouds. The average of all exemplars will then naturally shift towards the more reduced variants, until an equilibrium is reached between the desire for articulatory reduction ("Economy") and the need for functional communication (Passy's 1890 "Emphasis"). Note that no separate mechanism is required to spread these exemplars from one speaker to the next: because Exemplar Theory does not assume a unidirectional feedforward model of phonological processing, the more-reducing speaker will naturally provide more-reduced exemplars to his or her peers, without the need for any specific theoretical machinery.

1.2.4 Which comes first: perception or production?

While the preceding sections could be taken to suggest that pure speaker-caused sound change is rare, there are studies that have found change in production before change in perception, in which case it must have been the speakers who have made the first move. One such study is Evans & Iverson (2007), who studied the accents of British-English high-school students who were about to enter university. While at the end of high school these students had distinctly local English accents, after a year in university their productions had measurably changed to be more in line with Standard Southern British English. In perception, however, Evans & Iverson (2007) did not find significant changes, although they did find a correlation between their perception measures and the degrees to which their participants had changed in production. These results are incompatible with a view in which sound change starts in the listener and only later spreads to speakers.

Other work, however, has obtained precisely the reverse findings. In a study of the devoicing of Dutch fricatives and bilabial plosives, Pinget (2015) found strong evidence that it is the listener who has to initiate a sound change, which is in line with classic misperception-based accounts of sound change. Harrington, Kleber, & Reubold (2008), in a study of Standard Southern British English /u/-fronting (whereby /u/ is changing into /ʉ/), come to the same conclusion: their data are compatible with an Ohalian account of sound change, and not compatible with a speaker-initiated account of sound change.

Why do Evans & Iverson (2007) find change starting in production, and do Pinget (2015) and Harrington, Kleber, & Reubold (2008) find change starting in perception? The critical difference between the studies is the level of representation at which the change is playing out. For Evans & Iverson (2007), the sound changes to be acquired by their subjects are changes in surface realizations: a vowel /V/ changes its realization [V_1] to [V_2]. In the case of the other two studies, the sound changes that are on-going are not phonetic, as in the case of Evans & Iverson (2007), but phonological. Harrington, Kleber, & Reubold (2008) make the case that their sound change started by listeners undercompensating for coarticulation of /u+t/ sequences, where the [u] becomes more front due to coarticulation with [t], leading them to a reanalysis of /u/ as /ʉ/. In this case, differently from Evans & Iverson's (2007), the realization [ʉt] was already in existence and the actual sound change is the reanalysis of the underlying form /u/ as /ʉ/. For Pinget (2015), the same is true: her study investigates the merger of the *phonemes* /f/ and /v/ and of the *phonemes* /p/ and /b/. Her conclusion that these sound changes needed to be perceived by listeners before they would become produced by speakers is in line with a hypothesis that change at the underlying level must be initiated by listeners, but change at only the surface level starts with the speakers.

The idea that underlying-form change starts in perception and surface-form change starts in production is supported by evidence from psycholinguistics. Research in psycholinguistics has shown that listeners are exceptionally skilled at compensating for variation in the phonetic input they receive. Thus, a listener presented with a subtle difference in a phonetic realization (as is the object of change in Evans & Iverson 2007) will perceptually compensate for this difference, preventing the actuation of an Ohalian sound change. However, as will be extensively discussed in Section 1.3, this ability to compensate for changes has limits. In particular, Witteman et al. (2015) have shown that listeners fail to accommodate on-line to changes by which a sound is realized as a member of another phoneme category—in this case, Dutch /i/ realized as [ɪ], which also exists as a separate phoneme in Dutch. This suggests that a listener cannot compensate for sound changes affecting underlying forms, and hence, for these sound changes, it has to be the listener who performs the crucial reanalysis. The implications for sound change are that if initiated by a speaker, sound change involves a reanalysis of the concrete realization corresponding to an abstract phonological category, i.e. of the phonology-phonetics mapping. In contrast, if a sound change is initiated by a listener, the reanalysis is one of the abstract category system itself, i.e. the phonetics-phonology mapping.

1.2.5 Types of change

After an individual has come into contact with a sound change, it needs to spread through their linguistic system. This is exactly Weinreich, Labov, & Herzog's (1968) original transition problem; recent literature (e.g. Bermúdez-Otero 2007) prefers to speak of *implementation*. Implementation is generally recognized to take place in one of four ways. The two most-well-known are Neogrammarian change (Osthoff & Brugmann 1878) and change by lexical diffusion (Wang 1969). Neogrammarian changes are those that start out as gradual phonetic innovations, which are then grammaticalized according to the lifecycle discussed previously. It follows that if sound change starts out in the realization of a phonetic category, all words in the lexicon are affected by the change at the same rate at the same time; thus, Neogrammarian change is phonetically gradual but lexically abrupt (Bermúdez-Otero 2007). By contrast, in the case of change by lexical diffusion, the change is a phonetically abrupt substitution of one phonetic category for another, which takes place in individual lexical items. Here, the locus of change does not lie in the realizations of phonetic categories, but rather in the realizations of individual lexical items: some words will have implemented the change, other words will not have. Thus, classic lexical diffusion is lexically gradual, but phonetically abrupt.

As a third option, Bybee (2002) notes that there are some sound changes that appear to be both phonetically and lexically gradual. The dichotomy between Neogrammarian change and lexical diffusion presented above explicitly disallows this third mechanism of spread: a change is Neogrammarian, which is phonetically gradual but lexically abrupt, or it is lexically diffuse, in which case it proceeds through the lexicon slowly, but when a word changes, it immediately does so fully. Bybee's (2002) observation that there *are* changes that can be shown to be lexically diffuse, but where the individual lexical items are not changing at the same rate, shows that a third option is needed, which I provisorily term "change by exemplar", with "exemplar" referring to Pierrehumbert's (2001) Exemplar Theory. Bybee's (2002) analysis of these kinds of sound changes is that the exemplar clouds that ultimately give rise to phonological representations are slowly undergoing a regular sound change. Since exemplar clouds are formed separately for each word in the lexicon, the diffusion throughout the lexicon of this phonetically gradual regular sound change then naturally obtains.

The fourth and final mode of implementation in which a sound change can be implemented is phonetically and lexically abrupt. These changes arise due to (un)conscious choices, such as in accommodation. What changes here is what Janson (1983) calls the *norm*, which in his view is a conscious sociolinguistic allophone choice (such as which of a large variety of rhotics to use for a single /r/ category), which, upon transmission to a new generation of speakers only (Janson 1983), will result in a change in the underlying form for this category. In terms of the lifecycle, these changes follow the same steps as Neogrammarian changes, with the single difference that the original phonetic change is too large to have arisen gradually. In these changes, a central role is to be played by sociolinguistic factors such as accommodation by listeners to speakers they evaluate positively (Auer & Hinskens 2005, Chambers 1992, Janson 1983, Pardo 2006, Sonderegger, Bane, & Graff 2017, Trudgill 1986). However, this reveals a deficit in Janson's (1983) account that it shares with that of generative views on misperception. If adults can only change the underlying forms associated with phonemes and not the number and distribution of these phonemes themselves, as explicitly claimed by Janson (1983), then any categorical reanalysis can be performed only by their children. However, if adults can only change their surface realizations, the only system that they can transmit to their children is one in which changes in surface representations alone will do, and, due to the subset principle (Berwick 1985), no incentive for children to change will *ever* arise.

1.2.6 Summary

The received, generative, view of sound change claims that sound change originates in a coordination failure between a speaker and a listener, which involves over- or underparsing of phonetic cues as phonological and vice versa. While coordination failures may occur at any place and point in time, reanalyses beyond the surface level can be performed by children only (Section 1.2.1). Sound change tends to be caused by the listener, but it can also be caused by the speaker, if and only if this comes at no additional cost for them, i.e. the change does not result in a system that is harder to produce or more difficult to understand. Specifically, the speaker selects prototypes that are optimal according to both production and perception (Section 1.2.2); this may result in the actuation of sound change if the distribution of available tokens or exemplars is skewed towards a better variant that is not currently the norm (Section 1.2.3). This type of speaker-induced sound change may be more likely to start as a physical change in the phonetic realization, whereas listener-induced sound change may be more likely to start as an abstract reanalysis of the underlying form (Section 1.2.4). Change may be categorized along phonetic and lexical abruptness/graduality, and these different modes of implementation may operate on different principles (Section 1.2.5).

This view is well-established, convenient, and largely plausible. However, Section 1.2.5 ended with a critical remark: if adults can only *produce* grammars that are consistent with surface-level changes, they can also only *transmit* grammars that are consistent with surface-level changes. It is then up to their children to reanalyze such phonetic changes as being part of the phonology, but the subset principle predicts that they do not generally do so. More importantly, however, is that neither of these predictions are completely in line with reality. As the Polder shift demonstrates (Voeten 2015), adults *can* in fact perform truly phonological reanalyses and *can* transmit these to their children. In addition, the same is true for children: as part of the normal process of phonological acquisition, children often come up (at least temporarily) with incorrect phonological analyses. What makes all this possible? The answer is probably in the way speakers and listeners cope with variation.

The individual's processing of variation falls under the purview of psycholinguistics, to which due attention is paid throughout the remainder of this dissertation. Empirical research in this field has provided additional challenges for the received view on sound change, and findings by psycholinguists will provide important stepping stones in the synthesis offered in Chapter 7. Section 1.3 provides a brief summary of the ways in which pycholinguistic empirical research is and (mostly) is not compatible with the received view of sound change. This provides the empirical background for an important component of this dissertation: methodology.

1.3 The psycholinguistics of variation in perception

1.3.1 Perceptual learning as the antagonist of misperception

The misperception-based account of sound change faces fundamental obstacles from a field of linguistic research that can from time to time be underappreciated by historical phonologists. Decades of work in psycholinguistics (summarized in major works such as Cutler 2012) indicate that human speakers and listeners are, in fact, extremely skilled at compensating for variation in the speech signal. In fact, psycholinguistic evidence shows that *compensating* for variation is a misnomer: variation is actually *used* in talker-specific processing strategies (Nygaard, Sommers, & Pisoni 1994, Palmeri, Goldinger, & Pisoni 1993). A specific problem for the misperception account of sound change is posed by the existence of lexically-guided perceptual learning (Norris, McQueen, & Cutler 2003): the phenomenon that listeners adapt to persistent deviations from their own expectations by individual speakers.

Perceptual learning is a process that primarily operates on the basis of lexical knowledge. The original Norris, McQueen, & Cutler (2003) paper was based on knowledge of individual words. Their experiment divided Dutch listeners into two groups, presenting them with words containing either /f/ or /s/ phonemes realized as an ambiguous intermediate sound [?]. In a training phase, these words were selected such that only one of these two interpretations was possible (e.g. "witlo[?]" can only be "witlo[f]" meaning "chicory", as there is no Dutch word *"witlo[s]"). In a subsequent categorization task, the /f/-familiarized participants showed expanded /f/ categories (an ambiguous sound needed to be more [s]-like for them than for the other group before they would categorize it as /s/), and the /s/-familiarized participants showed the reverse. In a later study on the same phenomenon, McQueen, Cutler, & Norris (2006) found that this effect generalized not just to phoneme categorization, but also to word recognition. In their experiment, participants were familiarized in the same way as in Norris, McQueen, & Cutler (2003), but then did a lexical-decision task with ambiguous words (e.g. "doo[?]", which can make either "doof" meaning "deaf", or "doos" meaning "box") with a cross-modal priming component. For both groups, the results from this experiment showed facilitatory priming effects for prime–target pairs congruent with training (e.g. auditory "doo[?]" paired with the picture of a box in the /s/-familiarized condition), and inhibitory priming effects for incongruent prime–target pairs. These results show that listeners use lexical information to retune their sound categories, and that this retuning generalizes to new items, and hence takes place at an abstract phonological level.

Further evidence demonstrates that this retuning of categories is not limited to lexical words nor to phonemes, but is also obtained for more surface-like phonetic categories. Cutler et al. (2008) replicated Norris, McQueen, & Cutler's (2003) findings on the [f∼s] continuum, this time based not on lexical knowledge (participants trained on words like "witlof"), but based on phonotactics. In their study, British-English listeners trained on "[?]rul" learned that the ambiguous sound must have been /f/ (*[sr] being an illicit word onset in English), whereas a second group of British-English listeners trained on "[?]nud" learned /s/ (*[fn] being an illegal English word onset). In a later categorization task, /f/-trained participants gave more /f/ responses, and /s/-trained participants gave more /s/ responses along an [f∼s] continuum. These results show that perceptual learning does not necessarily rely on individual words inside the lexicon: phonological knowledge of static lexical patterns also suffices to trigger the process.

For misperception-based sound change to take place, the mechanism for perceptual learning needs to be impaired somehow. An obvious candidate is the amount of exposure. For instance, Maye, Aslin, & Tanenhaus (2008) have shown that listeners can adapt to entire vowel shifts (all vowels lowered by one degree, so "wicked witch" becomes "weckud wetch"), but these participants received twenty minutes of consistent exposure in a laboratory setting. On the other hand, results by Witteman et al. (2015) show that participants can adapt after as little as 3.5 minutes, and that such adaptation is even long-lasting (see also Gaskell & Dumay 2003, who suggest a critical role for sleep in such long-term accommodation). If adaptation to different speakers and their sound systems is this rapid, amount of exposure might not be a viable contender for bypassing perceptual learning. Witteman et al. (2015) suggest one possible failure mode, namely that their participants were unable to adapt to realizations that crossed phoneme boundaries (which probably has a neurolinguistic correlate in the P600; Chapter 6), but since not all sound changes are phonemic mergers, this cannot be a general explanation. Results by Witteman, Weber, & McQueen (2014) implicate the consonant-vowel asymmetry as contributing to sound change, having found that adaptation to vowels (as in Maye, Aslin, & Tanenhaus 2008) is easier than adaptation to consonants, but again, this applies only to a portion of all sound changes.

Neurolinguistic research, mostly centered around the mismatch negativity ("MMN") ERP component, provides some more perspective. In a study of long-distance coarticulation as a possible source for sound change actuation, Grosvald & Corina (2012) found that the brain was sensitive to long-distance coarticulation. That is, a vowel colored by coarticulation from a vowel that was one or two (but not more) syllables away elicited a significant MMN in an oddball task, showing that the brain was capable of detecting the phonetic difference from a stream of non-coarticulated vowels. Other work has shown that the MMN is not necessarily acoustic (and Grosvald & Corina indeed argue that theirs is not), but can also reflect phonological knowledge. Four publications by the same researchers on the same data (Jacobsen 2015, Steinberg, Truckenbrodt, & Jacobsen 2010a, 2010b, 2011) report significant MMNs to the difference between a correct allophone (the German realization [ɛç]) and an incorrect allophone (the German realization *[ɛx]). These results show that the brain does not compensate for all types of variation, even at the more abstract level of phonology. This ties in with results from the field of regional-and-foreign-acccent processing, which report cumulative interference effects in reaction times (Floccia et al. 2009, Floccia et al. 2006) and the N400 (Goslin, Duffy, & Floccia 2012). This suggests that while a human being in a conversation is very adept at compensating for variation, "under the hood" there are problems that the brain needs to actively resolve. It is highly probable that the degree to which this succeeds is subject to a significant degree of individual variation.

As detailed in the introduction to Chapter 4, various studies of adaptation to phonetic differences across the lifespan (e.g. Alshangiti & Evans 2011, Bauer 1985, Carter 2007, Cedergren 1987, Chambers 1992, De Decker 2006, Evans & Iverson 2007, Harrington 2006, Harrington, Palethorpe, & Watson 2000, Hinton 2015, Nahkola & Saanilahti 2004, Nycz 2011, Nycz 2013, van Oostendorp 2008, Prince 1987, Sankoff 2004, Sankoff & Blondeau 2007, Sankoff, Blondeau, & Charity 2001, Trudgill 1988, Wagner 2008, Yaeger-Dror 1994, Ziliak 2012) have found that while some individuals adopt such differences (which include sound change) with relative ease, others do not. Similarly, psycholinguists have warned for a very long time that analyses of psycho- and neurolinguistic data need to properly take into account variation between individual participants, due to obvious variation in psychophysiological makeup leading to equally obvious variation in measures such as response latencies in RT experiments. Psycholinguists have additionally realized that the incorporation of merely participants as a random factor in statistical models of language processing is not enough; language items are equally random, leading Clark (1973) to formulate his "language-as-a-fixed-effect fallacy". While the remedy—the mixed-effects model—had already been formulated by statisticians as far back as 1950 (Henderson), it was only through the effort of authors like Baayen, Davidson, & Bates (2008) that mixed-effects models finally became popular in (psycho-) linguistics. Section 1.3.2 reflects on these and other methodological innovations that made the research in this dissertation possible.

1.3.2 Methodological innovations for psycholinguists

Since Barr et al. (2013), psycholinguists have scrambled to incorporate differences between participants and items into their statistical models to the absolute fullest extent, citing Barr et al.'s (2013) slogan to "keep it maximal". It is nowadays believed that the advice by Barr et al. (2013) to unconditionally fit the maximal random-effects model was somewhat overzealous. Beyond the computational expense to fitting models with random slopes up to the error term, the resulting models often converge to a solution that fails the KKT criteria (Karush 1939, Kuhn & Tucker 1951) or converges to a boundary solution for which these criteria do not even apply. The former failure mode is reported by most statistical software as "failure to converge", whereas boundary solutions are currently only reported by recent versions of R (R Core Team 2020) package `lme4` (Bates et al. 2015b), which reports them as "singular fit"s. The cause for either failure is the same: the model contains too many, probably (multi)collinear, unknown terms, for which no (statistically felicitous) global optimum can be found. Bates et al. (2015a) argue that these problems with such overparameterized models make them unsuitable for routine use, and in some cases even lead researchers to incorrect conclusions. Even if the maximal model can be made to "work", that is, fit nonsingularly without convergence warnings, the inverse relationship between the number of free parameters and statistical power means that such models are costly not just in terms of CPU time, but also in terms of the number of participants and items necessary to be able to detect a true effect (Judd, Westfall, & Kenny 2017, Matuschek et al. 2017).

Bates et al. (2015a) propose to tackle these issues by starting with an infelicitous (nonconverged or singular) maximal model, and then manually identifying the extraneous random effects. They have developed an R function (now incorporated into package `lme4`) called `rePCA` which assists in this process. However, this is a laborious and not necessarily straightforward task, as the θ parameters on which `rePCA` operates do not always correspond directly to individual terms in the researcher's design matrix. For example, one offending θ parameter may correspond to the correlation of two levels of two different categorical variables with many levels—should the researcher then drop all correlations between all of these combinations, or find some way to convince the software to hold only this problematic parameter at a value of zero? Even software that allows more flexible covariance structures than `lme4`, such as R package `glmmTMB` (Brooks et al. 2017), cannot easily accommodate such a request.

An alternative approach is suggested by Matuschek et al. (2017). They propose to use backward stepwise elimination, a well-established technique in the field of psycholinguistics, to identify which of the terms included in a maximal model are truly required. As the backbone of this technique is a simple likelihood-ratio test, this approach to random-effects selection is principally motivated and straightforward to use. The only requirement that may be difficult to meet is a converged maximal model from which to start backward elimination. This procedure thus finds the balance between Type I error rate and power that Barr et al.'s (2013) approach lacks, but still requires a feasible maximal model from which to start. One way to define a feasible maximal model is as the model that includes the *most important* random-effect terms that can still converge. To find such a model, this dissertation relies on R package `buildmer` (Voeten 2020). This is an R package that builds up a mixed-effects model by starting with only the fixed effects, and adding random-effect terms one by one as long as the model is still able to converge. Random effects are added to the model in order of their contribution to the likelihood-ratio-test statistic or an information criterion, such that when the model eventually fails to converge, the most important random effects have made it in. From this maximal *feasible* model, backward elimination is then used to identify which of the included random and fixed effects significantly improve the model fit.

Beyond the mixed-effects model for linear regression (which includes ANOVA) and generalized linear regression, linguists have added another methodological notch onto their toolbelt in the past few years. The generalized additive model, known since Hastie & Tibshirani (1987) but popularized in linguistics by recent papers such as Baayen et al. (2017), makes it possible to perform regression analysis with predictors that are not linearly related to the response variable, but have effects that take arbitrary forms. Chapter 2 of this dissertation uses this type of model to deal with a longstanding and particularly vexing problem in phonetics, namely the problem of segmenting VC sequences where the consonant is very vowel-like. Dutch coda /l/, realized as [ɫ] in the Netherlands and also beginning to vocalize there (van Reenen & Jongkind 2000), is an example of such a problem-creating consonant. The transitions between VC sequences like [eːɫ] are smooth and continuous rather than discrete, and hence the concept of an *a priori* acoustic segmentation simply does not apply. However, it is nonetheless perfectly possible for a phonetician to formulate hypotheses about the temporal dynamics of vowels followed by coda /l/—in fact, Chapter 2 will demonstrate that coda /l/ indeed plays a major role in the Polder shift. Using the second formant as an example, one such hypothesis could be that the F2 will remain relatively high throughout the course of the vowel and fall as the articulation transitions into the [ɫ]. The modern implementation of the generalized additive mixed model (henceforth "GAMM") in R package `mgcv` (Wood 2017) makes it possible to model this nonlinear trajectory, including random effects, without additional methodological cost by the experimenter beyond a powerful computer. By using GAMMs, Chapter 2 does not require explicit segmentations of these highly gradient [Vɫ] transitions, but simply models the entire VC trajectory as a smooth nonlinear function of time. This makes it possible to compare hard-to-segment [Vɫ] sequences to unproblematic sequences of the same vowels followed by a nonapproximant consonant, dispensing with manual segmentation of the former but not the latter.

Other situations where linear mixed-effects models fare poorly are cases involving many categorical predictors. If multiple many-leveled categorical predictors interact to produce meaningful differences, a regression model ends up becoming very complex to interpret due to including all combinations of all factor levels of the highest interaction and all lower-order terms. In these situations, regression *trees* (see Tagliamonte & Baayen 2012, who discuss the closely-related conditional-inference trees) are more appropriate. These models operate on the basis of recursive partitioning, which results in very intuitive tree diagrams of the relative importance and effects of each variable given the variables that were of higher importance. R package `glmertree` (Fokkema et al. 2018) extends the basic principle of the regression tree by making it possible to incorporate random effects, using a very simple quasi-likelihood algorithm that iterates between building the tree given the random effects and estimating the random effects given the built tree until convergence. Chapter 3 relies on this technique to quantify the degree to which individuals adopt the Polder shift in their perception over a period of nine months.

The mixed-effects model has more uses than controlling for differences between participants and items, in which case the random effects are just nuisance terms. It is also possible for these random effects to be of interest in and of themselves. Psycho- and neurolinguistics have recently begun to realize the potential these models have of offering insight into individual differences (Eekhof et al. submitted, Kliegl et al. 2011, Mak & Willems 2019); the same is true of sociolinguistics (Drager & Hay 2012, Tamminga to appear). Chapter 4 uses the by-participants predicted random effects to classify individuals who have been exposed to the Polder shift for varying numbers of years as "adapted" or "non-adapted". It will be shown that the individual-level differences provide a more nuanced view than the aggregate group differences.

The aforementioned statistical techniques all rely on the existence of a null hypothesis that an effect to be tested is equal to zero, a philosophy known as "null-hypothesis significance testing" ("NHST"). A p-value $<.05$ means that the probability of observing the measured outcome variable y, given that this null hypothesis is true, is smaller than 5%. Therefore, either the null hypothesis is false, or the data are improbably unrepresentative. It is strange to think about statistical models in this way, making inferences about the value of a parameter β by arguing that the probability of the *data*, given the parameter being zero, is very low, i.e. $p(y|\beta = 0) < .05$. What we really want to know is the probability of the *parameter* having a certain value taking the data as given, i.e. $p(\beta = \hat{\beta}|y)$. This is the difference between frequentist statistics and Bayesian statistics. The former is an incoherent hybrid of the original views of Fisher (1955, 1956) and Neyman & Pearson (Neyman 1950, 1957; see Gigerenzer 2004 for details), while the latter is philosophically more sensible, but not the standard scientific practice (see Kruschke 2010a, 2010b for commentary).

In practice, null-hypothesis significance testing is a useful tool to have available for testing point hypotheses about model parameters, particularly because the frequentist maximum-likelihood estimates can be computed efficiently, which is not true for the Bayesian maximum-*a-posteriori* estimates except in special cases. It is also safe to use as long as one is aware of the pitfalls, the most important of which are that a p-value <.05 does not constitute absolute, categorical, evidence that the alternative hypothesis can only be true, and that a p-value >.05 does not constitute *any* kind of evidence that the null hypothesis is true. If hypotheses are framed within these constraints, it is unlikely that a Bayesian analysis would result in a different substantive conclusion than a frequentist analysis, unless the data were prepared to be excessively pathological. For most of the studies reported in this dissertation, these caveats are acceptable, except for the EEG experiment reported in Chapter 5. This chapter investigates the mismatch-negativity ERP, henceforth "MMN". The MMN is an automatic brain response that is generated when the brain detects a change in sensory stimulation—in the case of this dissertation, when a syllable like [ei] is replaced by one like [eː]. This ERP is almost always asymmetric (Lahiri & Reetz 2010), which means that a switch such as [ei]→[eː] will generate an MMN, but the reverse switch [eː]→[ei] will not. The presence of an MMN can be argued using NHST, but its absence cannot; thus a Bayesian approach is needed. Chapter 5 takes the approach by Wagenmakers (2007), in which Bayes factors are computed based on the difference in BIC (Schwarz 1978) between two candidate models. If these model comparisons are set up such that a full model is compared to a model in which a single focal term has been removed, the corresponding Bayes factor quantifies the odds of that term being equal to zero. Given the *a priori* assumption that both models are equally likely, this Bayes factor makes it possible to quantify the evidence against this assumption (similar to the NHST p-value) as well as *in favor of* this assumption (thus providing evidence that the models are indeed equivalent, i.e. that the effect being tested *is* zero).

All of the aforementioned statistical methods assume that the researcher knows what they are looking for. For example, when analyzing the data of an EEG experiment, the researcher needs to specify *a priori* what combination of electrodes is of interest, and at which moments in time. This information is not available beforehand when the research is exploratory. In this situation, permutation testing can be used (Maris & Oostenveld 2007), and this is done in Chapter 6 of this dissertation. This chapter reports an exploratory investigation of EEG differences that are related to the Polder shift. The chapter will reveal a phonological P600 modulated by factors unrelated to the Polder shift, of which the precise nature is not yet fully known. The permutation tests made it possible to identify a window of statistically-significant differences that corresponded precisely to a P600, which, combined with the correct direction for the observed difference, corroborate the sparse scientific literature on the phonological P600, and made it possible to take first steps to further qualifying the conditions under which this effect can be obtained. The analysis used in Chapter 6 led to the development of R package `permutes` (Voeten 2019c), the code of which has also seen use outside of the Polder shift by authors such as Ruijgrok (2018).

1.4 This dissertation

The literature discussed thus far paints the picture of a field that is internally divided. On the one hand, we have historical phonologists, who claim that production and perception are especially prone to sound change. In this view, perception is fallible and may result in perceptual reanalyses, and production favors articulations that are more familiar or in other ways "easier", and may therefore lead to speaker-driven sound change. On the other hand, psycholinguistic evidence has shown that both the production and the perception apparatus are extremely flexible, and can cope with incidental variation without any problems. The overarching question of this dissertation is: **what factors influence the adoption of sound change?**

The currently-ongoing Dutch Polder shift provides an opportunity to investigate this major question. There are three properties that make this vowel shift particularly well-suited for this purpose. First and foremost, the Polder shift is currently on-going, which means that, for once, we are *not* too late (cf. Pinget 2015). Secondly, the Polder changes are phonologically conditioned (Voeten 2015). This makes it possible to disentangle phonetic changes from phonological changes, and thus provides a unique test case for the claims from Section 1.2 that adults are not able to adopt phonological changes, but *can* simulate them using more superficial accommodation rules. Finally, the Polder shift is geographically stratified, such that both conservative and innovative individuals are available for psycholinguistic experiments. This brings us to the first research question. For practical reasons of needing to select participants suitable for a psycholinguistic investigation of the on-going Polder shift, a clear picture of the current state of affairs of the specific sound changes subsumed under the Polder shift is necessary. The first research question in this dissertation can therefore only be: **what is the synchronic diatopic diffusion of the sound changes involved in the Polder shift?** This is discussed in Chapter 2.

Based on the findings in Chapter 2, the most suitable participants for further empirical investigation of the adoption of the Polder shift were found to be *sociolinguistic migrants*, specifically speakers of Flemish Dutch who have migrated to the Netherlands, and have hence come into contact with the Polder shift. In order to generally investigate the adoption problem and specifically test the claims of the generative view on sound change outlined in Section 1.2, two related research questions present themselves. This brings us to the second research question—(**how**) **do sociolinguistic migrants adopt the Polder shift?**—as well as the third—**which individuals, after how much time, are more likely to adopt the Polder shift?** These two questions are superficially similar, but call for different methodological approaches. The former question calls for a *longitudinal* investigation at the *group* level. By contrast, the the latter question is more advantageously defined by a *cross-sectional* study at the *individual* level. Naturally, this dissertation takes these considerations into account. Chapter 3 answers the former question using a small-scale group-level investigation, whereas Chapter 4 deals with the latter issue using a large-scale individual-level study.

It was established in Section 1.3 that phonological variation is handled by specific psycho- and neurolinguistic mechanisms. Two specific correlates of phonological variation were identified in ERPs: the MMN and the P600. These ERPs were shown to be sensitive specifically to phonological-rule violations, and hence may inform us about individuals' adoption of the Polder shift. This brings us to the fourth and final research question: (**how**) **is the adoption of the Polder shift reflected in ERPs?** Chapter 5 focuses on the MMN component, using the specific changes in the phonological rules involved in the Polder shift, whereas Chapter 6 focuses on the P600 component and ends with a more general claim about the types of variation in which the P600 is involved.

CHAPTER 2

Regional variation in on-going sound change: the case of the Dutch diphthongs

This chapter has been submitted.

Abstract

This chapter discusses the regional variation in four on-going sound changes in the Dutch vowels /eː,øː,oː,ɛi,œy/ that are conditioned by a following coda /l/. The synchronic diatopic diffusion of these changes is charted using the Dutch teacher corpus, a comprehensive dataset containing word-list data from four regions in the Netherlands and four in Flanders. Comparisons are made of the five vowels preceding nonapproximant consonants and preceding coda /l/. To avoid manually segmenting the oftentimes highly gradient vowel–/l/ boundary, GAMMs are used to model whole formant trajectories. Comparisons are then made of trajectories and of peaks of trajectories. The results are used to classify the nature of the four sound changes in terms of phonetic and lexical abruptness/graduality, and show that the changes are intertwined in such a way that they can only be considered as separate facets of a single, currently on-going, vowel shift.

2.1 Introduction

When we want to study sound change, we are always too late: by the time a successful change can be identified, it has—by definition—already spread beyond the incipient stage, which makes it difficult to study the implementation of such a change (see e.g. Pinget 2015). However, if a change is regionally stratified, then this synchronic variation can be used as a proxy for the diachronic change, yielding a specific case of the apparent-time method to the study of sound change. The present chapter uses this approach to investigate an on-going vowel shift in Dutch that has been covered synchronically (Adank, van Hout, & Smits 2004, Adank, van Hout, & Van de Velde 2007) but has not yet been investigated from the perspective of diachronic change. The aim of the chapter is to make two points. First, the present chapter will show that recent innovations in statistical methods make it possible to analyze challenging phonetic data. Second, these approaches make it possible not only to characterize the changes in Dutch that are currently on-going, but also to retrospectively say something about the *nature* of these sound changes, specifically whether they were originally Neogrammarian, lexically diffuse, or something else.

Dutch is currently undergoing multiple interrelated changes in its vowel system. The vowels /eː,øː,oː/ are changing into upgliding diphthongs [ei,øy,ou] (change 1; van der Harst 2011, van der Harst, Van de Velde, & van Hout 2014, Van de Velde 1996, Zwaardemaker & Eijkman 1924), the vowels /ɛi,œy,ɔu/ are lowering towards [ai,ɒy,ɑu] (change 2; Blankestein 1994, Gerritsen & Jansen 1980, Gussenhoven & Broeders 1976, van Heuven, Van Bezooijen, & Edelman 2005, Jacobi 2009, Mees & Collins 1983, Stroop 1992, Stroop 1998, Van de Velde 1996, Voortman 1994), and both of these sets of vowels are realized as monophthongs when preceding coda /l/ (change 3; Berns & Jacobs 2012, Botma, Sebregts, & Smakman 2012, Voeten 2015), while coda /l/ itself is undergoing a process of vocalization and is causing retraction of the preceding vowel (change 4; Berns & Jacobs 2012, van Reenen & Jongkind 2000). These diachronic changes manifest synchronically as regional variation. Change 1 shows a clear split between the Dutch spoken in the Netherlands versus the Dutch spoken in Flanders (Adank, van Hout, & Smits 2004, Van de Velde 1996), and change 2 is restricted to the Randstad part of the Netherlands (Jacobi 2009, Stroop 1998). Preliminary research on change 3 shows that it is split between the Netherlands and Flanders in the same way as change 1 (Chapter 4). The sociogeographical status of change 4 is well-known, in that the change is restricted to the Netherlands, where coda /l/ is velarized. In Belgian Standard Dutch, coda /l/ is not velarized and change 4 has not taken place. However, there are a few Flemish dialects (mostly in West-Flanders and the west of East-Flanders; De Wulf, Goossens, & Taeldeman 2005:map 176) that have developed coda-/l/-vocalization independently, or have retained an ety-

Table 2.1: Modes of implementation of historical sound changes, after Bermúdez-Otero (2007).

	Lexical dimension	
Phonetic dimension	Abrupt	Gradual
Abrupt	Change in underlying forms (Janson 1983)	Lexical diffusion (Wang 1969)
Gradual	Neogrammarian change (Osthoff & Brugmann 1878)	Change by exemplars (Bybee 2002)

mological vocoid (in words like "geel", which in Proto-Germanic had a second syllable following the /l/, as in the English cognate "yellow"; De Wulf, Goossens, & Taeldeman 2005:map 175).

While the fact *that* these sound changes are on-going or have perhaps already completed in the language is well-known, *how* these sound changes are implemented remains to be established. The present chapter makes use of the same corpus as Adank, van Hout, & Smits (2004) and Adank, van Hout, & Van de Velde (2007) to answer this question. The mode of implementation of a historical sound change is generally classified along two axes: phonetic abruptness versus graduality on the one hand, and lexical abruptness versus graduality on the other (Bermúdez-Otero 2007). Consequently, four possible types of change have been attested in the literature, which vary along phonetic and lexical abruptness vs. graduality. These are changes in underlying forms, Neogrammarian change, change via classic lexical diffusion, and change by exemplars within an Exemplar-Theory framework. Table 2.1 provides an overview of how these modes of implementation map to the phonetic and lexical dimensions. It is not yet known how the sound changes that are currently on-going in Dutch can be classified in these terms. However, this can be measured from the synchronic data available in the corpus. Phonetic graduality versus abruptness can be inferred by looking at the differences between the regions that have been included in the corpus. If the regional differences in the realizations of the same vowel or vowel-/l/ sequence show a smooth trend, then a change is phonetically gradual. If there are sharp categorical differences between the regions, then the change is phonetically abrupt in the synchronic grammar (although synchronic data cannot rule out the possibility that the change was originally of a gradual nature, but has already completed). Similarly, if realizations of the same vowel or vowel-/l/ sequence are very different between the words containing them, there is evidence for lexical graduality. This makes it possible to operationalize the main research question: what types of changes are changes 1–4?

The phonetic data on the basis of which this question can be answered present methodological challenges. While changes 1 and 2 can be investigated with relative ease, changes 3 and 4 are more challenging to operationalize. Change 3 and 4 involve vowels followed by coda /l/. This is a challenging sequence to segment for a phonetician, because the transition between these two segments is phonetically highly gradient, and this problem gets worse if the coda /l/ is strongly vocalized (which is one of the sound changes to be investigated; van Reenen & Jongkind 2000). The acoustic-phonetic transition from a vowel to a coda /l/ is smooth and continuous rather than discrete, and hence these segments cannot be segmented reliably. One might even argue that in cases of such smooth transitions, the concept of a phonetic segmentation does not even make sense in the first place. It may not be surprising, then, that sociolinguistic studies on Dutch normally exclude vowels followed by liquids and glides, because a reliable way to analyze such vowels has up to now been lacking (see e.g. Van de Velde 1996, among many others). However, for the present chapter, ignoring coda /l/ is not an option, as it is an integral part of the research question. The present paper demonstrates a solution to this long-standing problem of analyzing vowel–approximant sequences by making use of generalized additive mixed models, henceforth "GAMMs". These models make it possible dispense with manual segmentation altogether and to instead analyze the entire time course of the vowel plus coda /l/ as-is. This makes it possible to compare hard-to-segment [Vɫ] sequences to unproblematic sequences of the same vowels followed by a nonapproximant consonant, dispensing with manual segmentation of the former but not the latter. The results in Section 2.3 will show that the GAMM-based approach to formant measurements provides new perspectives on the measurement of the four different types of sound change, for which the four changes currently on-going in Dutch are an excellent example. The results also highlight the advantages and limitations of a synchronic approach towards the analysis of diachronic sound change.

2.2 Method

2.2.1 Data and measurements

The regional variation in the on-going sound changes in Dutch is investigated using a large dataset called the "teacher corpus" (Adank 2003, van Hout et al. 1999), a corpus of 5,407 tokens of monosyllabic words sampled from four representative regions in the Netherlands and four in Flanders. The teacher corpus is particularly well-suited to investigating the regional variation in the realizations of the tense mid vowels and diphthongs, for at least three reasons.

The first reason is that it is phonologically comprehensive, in that it contains all the vowels of interest for the research into the particular sound changes, and it specifically distinguishes between coda-/l/ and non-coda-/l/ contexts within these vowels. The second reason is that it is excellently regionally stratified: the corpus consists of four regions in the Netherlands and four regions in Flanders, logically ranging from more central (prestigious) to more peripheral (non-prestigious). Lastly, the corpus is well-suited for studying on-going change in particular because it makes an effort to disentangle regional variation in implementation from *dialectal* variation. The ingenious approach is due to Van de Velde & van Hout (2003): the data in the corpus were collected from *teachers of Dutch*, who serve a role-model function to their students. This makes them representative for their region's interpretation of the standard language (see Delarue 2013, Grondelaers & van Hout 2012, Van Istendael 2008). For more details on the way the data were collected, and an in-depth treatment of the sociolinguistic issues involved, the reader is referred to Van de Velde & van Hout (2003) and Patti Adank's (2003) PhD dissertation. Part of the data presented in this chapter were also analyzed in Sander van der Harst's (2011) PhD dissertation, albeit with different aims and using different methods.

The corpus consists of samples of Dutch taken from four different regions in both the Netherlands and Flanders. For both countries, one "central" region was sampled (NL: Netherlands-Randstad, henceforth "NR"; FL: Flemish-Brabant, "FB"), one intermediate region (NL: the south of Gelderland, henceforth "NM" for "Netherlands-Middle"; FL: East-Flanders, "FE"), and two peripheral regions (NL: Groningen and Dutch Limburg, henceforth "NN" and "NS" for "Netherlands-North" and "Netherlands-South"; FL: Flemish Limburg and West-Flanders, respectively "FL" and "FW"); for details, see Adank (2003). A map of the regions is shown in Figure 2.1, which was created using the DynaSAND website (Barbiers et al. 2006). The corpus is further subcategorized for gender and age, and then has five speakers per cell, yielding a total of 8 (regions) × 2 (genders) × 2 (age groups: young vs. old) × 5 (speakers per cell) = 160 speakers in total. For the age groups, an age between 22 and 40 years was considered "young", whereas an age between 40 and 65 was considered "old".

The data that are relevant to the investigation in this chapter are those containing the vowels /eː,øː,oː,ɛi,œy,ɔu/, followed either by a nonapproximant consonant or by coda /l/. This yields the words "fee" /feː/, "beuk" /bøːk/, "boog" /boːx/, "boten" /boːtən/, "do" /doː/, "pook" /poːk/, "dij" /dɛi/, "meid" /mɛit/, "duin" /dœyn/, "luis" /lœys/, "tuin" /tœyn/, and "saus" /sɔus/ in the non-/l/ condition and the words "keel" /keːl/, "veel" /veːl/, "beul" /bøːl/, "geul" /ɣoːl/, "school" /sxoːl/, "zool" /zoːɫ/, "geil" /ɣɛil/, "heil" /ɦɛil/, "ruil" /rœyl/, and "uil" /œyl/ in the /l/ condition. Using Praat (Boersma & Weenink 2016), each of these words was sampled on F1 and F2 in 10-ms steps, using the

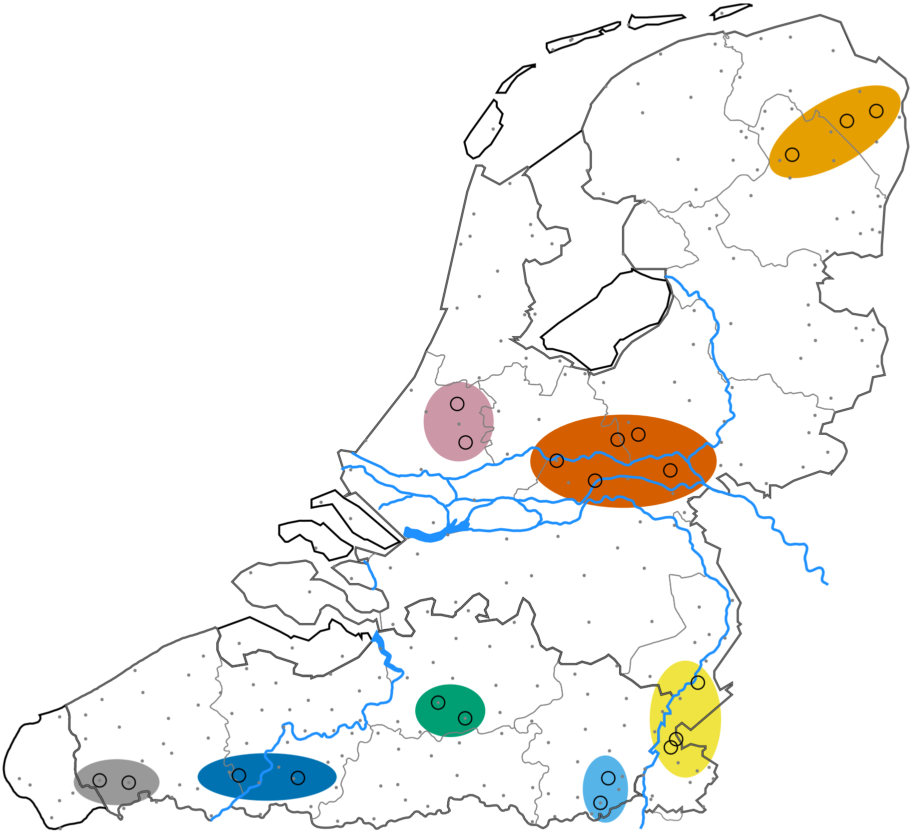

Figure 2.1: Overview of the different cities and towns from which data were sampled. Dots indicate major cities and towns according to Barbiers et al. (2006). Open circles indicate the cities and towns which were sampled for the teacher-corpus data. The overlaid colored circles indicate the corresponding regions, which are summarized in the following table (based on van der Harst 2011:55):

Region	**Color**	**Cities/Towns**
Netherlands-Randstad	🟪	Alphen aan den Rijn, Gouda
Netherlands-Middle	🟫	Tiel, Veenendaal, Ede, Culemborg, Elst
Netherlands-North	🟧	Assen, Veendam, Windschoten
Netherlands-South	🟨	Sittard, Geleen, Roermond
Flemish Brabant	🟩	Lier, Heist-op-den-Berg
Flemish Limburg	🟦	Ieper, Poperinge
Flanders-East	🟦	Oudenaarde, Zottegem
Flanders-West	⬛	Tongeren, Bilzen

same formant settings as in van der Harst (2011). Sampling started at the onset of the vowel (which was segmented manually) and continued to the 10-ms point (rounded down) at either the end of the vowel (for the non-/l/ words) or the end of the vowel+/l/ (for the coda-/l/ words). This resulted in varying numbers of 10-ms point samples, depending on the duration of the vowel, per token. The different token durations were normalized by converting the sample timestamps to percentages of vowel realization, such that each vowel token's duration ranged from 0% to 100% with a duration-dependent number of samples in between. Formant-measurement errors were excluded from the data by removing all samples falling outside the 100–1,000-Hz band for F1, and the 500–3,000-Hz band for F2.[1]

2.2.2 Data analysis

The resulting F1 and F2 trajectories were modeled by running separate GAMMs for each ⟨formant,vowel⟩ pair using function `bam` from R (R Core Team 2020) package `mgcv` (Wood 2017). Models were built up on the principle of parsimony (Bates et al. 2015a), based on visual inspection of the individual tokens and directly incorporating terms hypothesized to contribute to differences between them, until no remaining structure was visible in the by-token residuals.[2] This led to the inclusion of fixed effects for the predictors "Gender" (coded as male or female, sum-coded such that female $= 1$ and male $= -1$), "Region" (the eight regions in the corpus, sum-coded such that Netherlands-Randstad $= -1$ and the others are 1), "Following segment" (treatment-coded as /l/ or non-/l/, such that /l/ $= 1$ and non-/l/ $= 0$), and "Region × Following segment". In addition, random intercepts and slopes by following segment were added by participants. Smooths, defined as thin-plate regression splines with 30 basis functions, were added for the predictor "Time" by following segment; these terms model the nonlinear evolution of the dependent variable over time, for the non-/l/ and /l/ contexts separately. Finally, by-participants random smooths for "Time" by following segment were added to the model, configured in the same way as the regular smooths just described; penalties on the null space and on the first basis function of the thin-plate regression spline were added appropriately. Models were fitted to scaled-t errors including an order-1 autoregressive process with $\rho = .5$. The F1 model for the /ɔu/ vowel was fitted without effects for following segment, as this vowel was only

---

[1] All data were also checked manually by the author for outliers or "suspicious" formant values, but it was difficult to come up with a single, consistent, non-arbitrary set of criteria that was obviously correct for all cases. For this reason, it was decided to use only this formant-band-based criterion and to not use these additional manual corrections. The scaled-t error distribution makes the models robust against any remaining outlying observations; such outliers will be smoothed out without exerting undue influence on the regression estimates.

[2] I thank Harald Baayen for introducing me to the procedure.

present in one word: "saus". No F2 model was fitted for this vowel, as there are no coda-/l/ words available for this vowel, and the only change involving F2 is change 4, which only concerns coda /l/.

Regional variation due to the four on-going sound changes was established by comparing the fitted trajectories between the Netherlands-Randstad region with those in the other regions. The Randstad is chosen as the reference region because this is the region where all four changes are considered to be the most advanced (van der Harst 2011, Stroop 1998). For changes 1, 2, and 4, regional diffusion was assessed by predicting the fitted models' linear-predictor matrices onto a time grid of 101 points, corresponding to 0–100% realization. With eight regions, two following-consonant types, and two genders, this resulted in $101 \times 8 \times 2 \times 2 = 3,232$ linear predictions for each model. These were averaged over the two genders, and the linear predictions for the Randstad region were subtracted from them. The resulting linear differences were finally multiplied by the model's linear coefficients to obtain difference curves. Accompanying 95% Bayesian credible intervals were calculated using the approach in Wood (2017:293–294). Differences along a formant's time course are considered significant if their credible interval excludes zero.

Investigating the regional diffusion of change 3, the blocking of diphthongization before coda /l/, requires comparing the difference in diphthongization between vowels before non-/l/ and before coda /l/. In order to characterize the difference between these different types of trajectories, change 3 looks at the difference in formant *ranges*. Starting from the predicted trajectories provided by the GAMMs, for each vowel and region a trough was found by taking the highest F1 (corresponding to the lowest position of the tongue) within the first 50% realization, and a peak was found by taking the lowest F1 (corresponding to the highest position of the tongue) within the final 50% realization. The range of diphthongization is defined as the range between the trough and the peak, such that a negative range indicates upgliding diphthongization, whereas a range of zero or a small positive range indicates absence of this upglide (see Equation 2.1). Of interest is the *difference* in trough-to-peak ranges between the non-/l/ condition and the /l/ condition (henceforth "ΔTTP", see Equation 2.2). This difference is defined such that negative values indicate that there is more diphthongization in the non-/l/ condition than in the /l/ condition, whereas positive values indicate the reverse. Regional differences were established by subtracting the ΔTTP for the Randstad from that for the other regions ("ΔNR"; see Equation 2.3). Credible intervals for the ΔTTP and ΔNR measures were computed by performing the same steps as for the ΔTTP and ΔNR themselves on the corresponding linear-predictor matrices, and then again following the procedure outlined in Wood (2017:293–294).

$$\text{TTP} = \text{F1}_{\text{peak}} - \text{F1}_{\text{trough}} \tag{2.1}$$

$$\Delta\text{TTP} = \text{TTP}_{\text{non-/l/}} - \text{TTP}_{\text{/l/}} \tag{2.2}$$

$$\Delta\text{NR} = \Delta\text{TTP}_{\text{region}} - \Delta\text{TTP}_{\text{Netherlands-Randstad}} \tag{2.3}$$

The degree of lexical diffusion of the four sound changes is quantified by the between-words variability in each result to be discussed. To calculate a statistic representing this lexical variability, the same GAMMs (without effects for following segment and with the number of basis functions reduced to 10 for computational efficiency) were run for each word separately. For each peak value along the difference trajectory (changes 1, 2, 4) and ΔNR (change 3), the sum of squared differences of this result from the by-words individual estimates was computed and divided by the original result's variance (Equation 2.4). The resulting ratio is a chi-square random variable with $n - 1$ degrees of freedom, where n is the number of words.

$$\chi^2_{n-1} = \frac{\sum_{i=1}^{n}(x_{\text{full model}} - x_{\text{by-words model } i})^2}{\text{Var}(x_{\text{full model}})} \tag{2.4}$$

Section 2.3 discusses the results of the analyses. The data and R code with which these have been produced are available at https://figshare.com/s/48e0afc5dc7b10d24726 as the files `data.csv` and `analysis.R`, respectively.

2.3 Results

The results for the four sound changes are discussed in order, based on the relevant statistics extracted from the fitted GAMMs. One of the models—the F1 model for the /ɔu/ vowel—did not converge successfully. This vowel was therefore excluded from the results. For reference, Figures 2.2–2.4 provide a general overview of the data for the F1 followed by a nonapproximant consonant, the F1 followed by coda /l/, and the F2 followed by coda /l/, respectively. These figures were obtained by predicting from the model in the same way as described in Section 2.2.2, without subsequently calculating any differences between conditions. As such, they are equivalent to smoothed versions of the raw data, for the average participant and the average word. In Figure 2.2, the full vowel trajectories are shown for the vowels followed by a nonapproximant consonant. This consonant itself is not included, but the vowels' trajectories to and from the consonants are clearly visible. This underscores the observations by van der Harst (2011) on the same data that the influence of coarticulation is minimal no earlier than 25% realization and no later than 75% realization. Figures 2.3 and 2.4 show the vowels followed by coda /l/, which is *included* in the depicted trajectories, as it could not be reliably segmented. Thus, in these plots, the ends of the depicted trajectories coincide with the ends of the words.

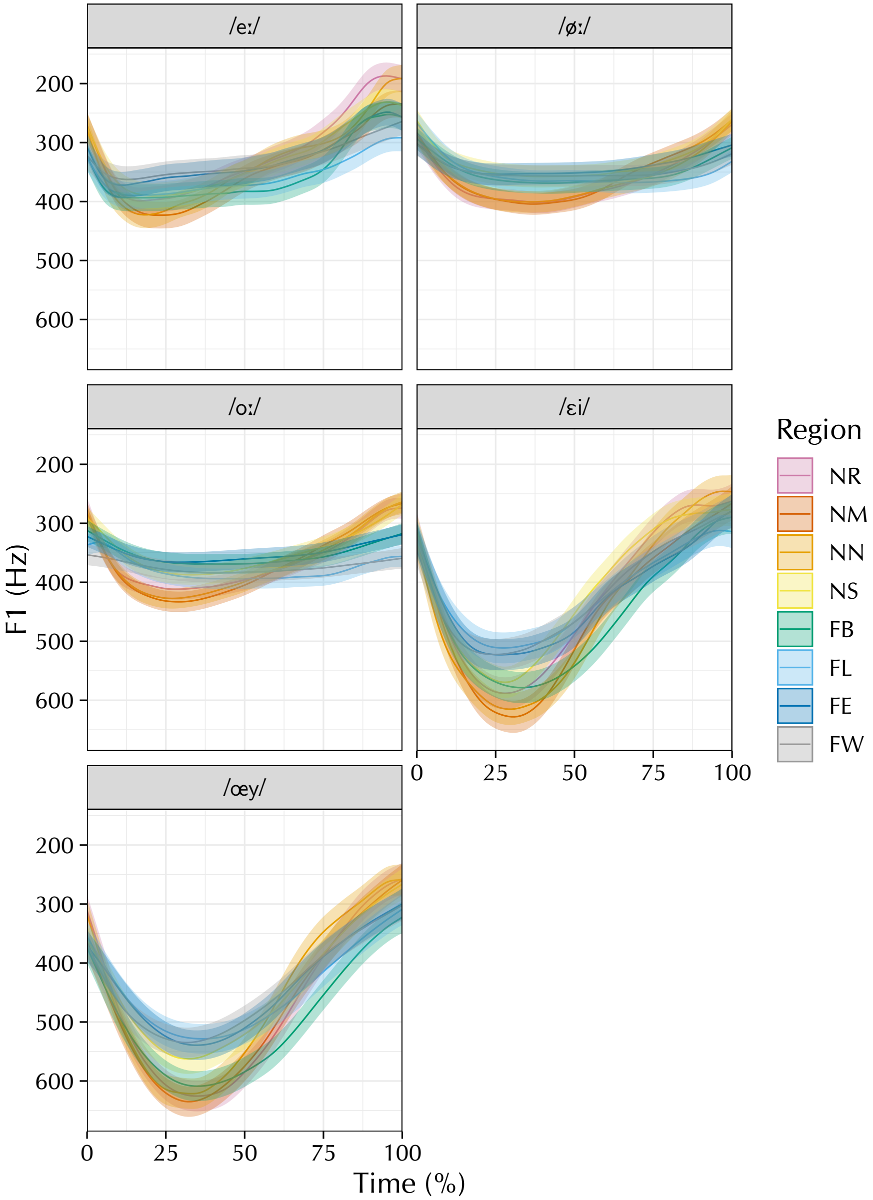

Figure 2.2: Overview of the vowel trajectories as smoothed curves, for the F1 data (closed/open dimension) when followed by a nonapproximant consonant. The following consonant itself is not included. The ribbons around the curves indicate the 95% CI.

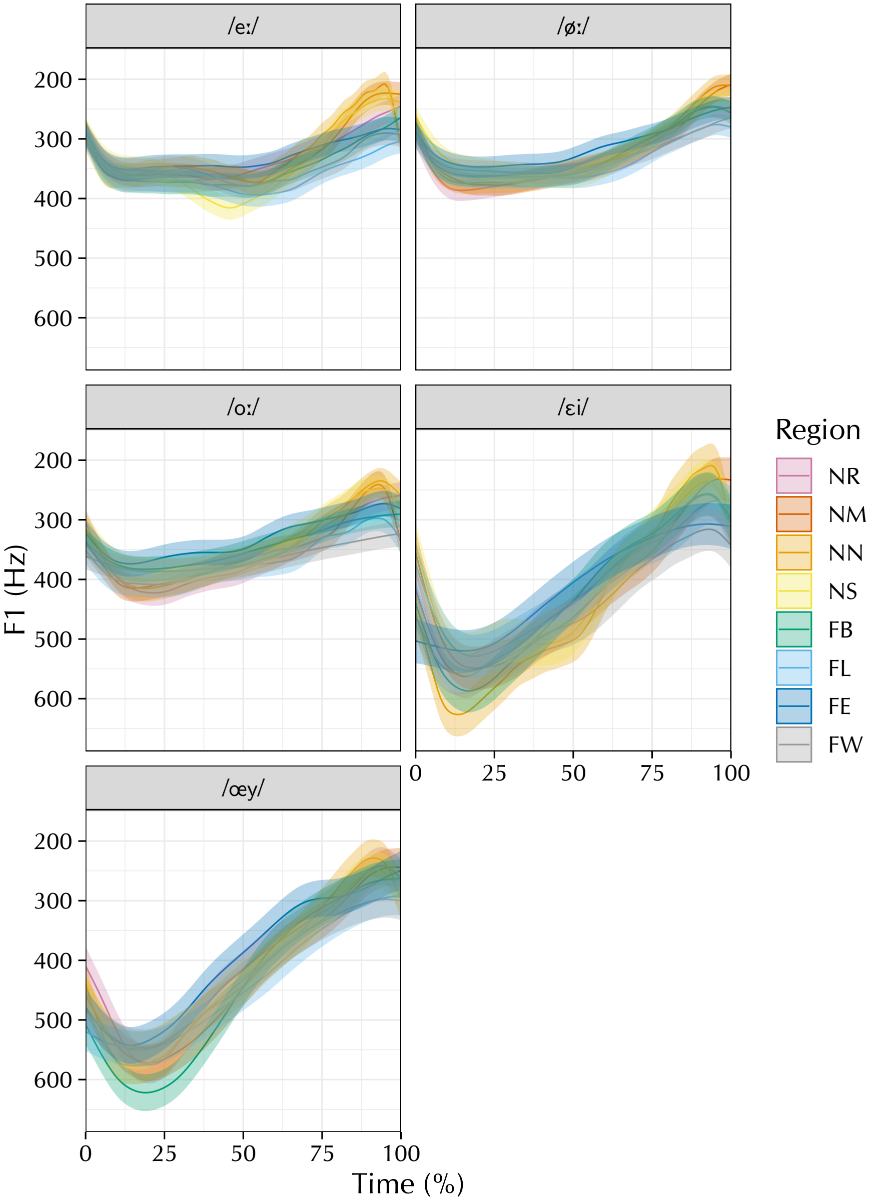

Figure 2.3: Overview of the vowel trajectories as smoothed curves, for the F1 data (closed/open dimension) when followed by coda /l/. The curves include the coda /l/ in its entirety. The ribbons around the curves indicate the 95% CI.

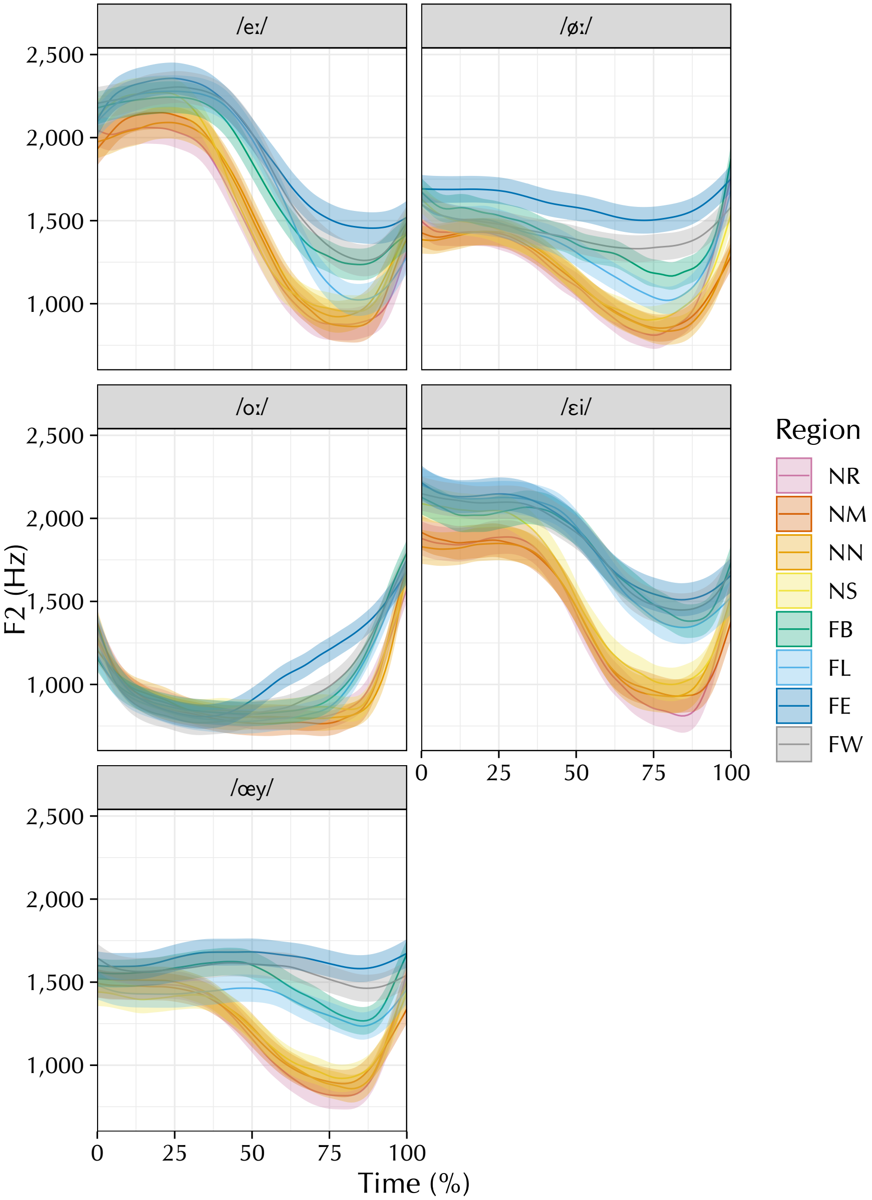

Figure 2.4: Overview of the vowel trajectories as smoothed curves, for the F2 data (front/back dimension) when followed by coda /l/. The curves include the coda /l/ in its entirety. The ribbons around the curves indicate the 95% CI.

2.3.1 Change 1: diphthongization of /eː,øː,oː/

Figure 2.5 shows the difference smooths of the five vowels when not followed by /l/, compared to the Netherlands-Randstad region. As the focus for change 1 is the diphthongization of /eː,øː,oː/, the dependent variable in this plot is the F1. In this figure as well as Figure 2.6, significance of the differences is indicated by the presence of a ribbon around the smooth; the width of the ribbon spans precisely the 95% CI. The peak points of the significant differences are listed in Table 2.2. Because change 1 is about upgliding diphthongization, which only affects the latter half of the vowel, only differences beyond the vowel midpoint (>50% realization) are considered relevant for interpretation. Significant differences that are found *only* in the final 10% of the smooth are excluded, as this part of the signal is strongly influenced by coarticulation (van der Harst 2011), making these differences unreliable.

There is systematicity in the combinations of vowel and region that show significant differences from the Randstad. With the exception of the /øː/ vowel in Flanders-East, all Flemish regions realize all five vowels with significantly less upgliding diphthongization (higher target F1) than the Randstad region. This observation already covers 79% (19/24) of the significant differences that were found. A second major role is played by the vowel /eː/, which the three non-Randstad regions in the Netherlands also diphthongize less strongly than the Randstad region, although the differences are quantitatively smaller than those between the Randstad and the Flemish regions. Thirdly and finally, in the Netherlands-Middle region the vowel /ɛi/ diphthongizes significantly less than the Randstad, and in the Netherlands-North the vowel /œy/ diphthongizes significantly more than in the Randstad.

The χ^2 values in Table 2.2 measure the variability across words of the peak differences presented in the table, and hence provide an index of the degree of lexical diffusion detectable in these data. There is limited evidence for lexical diffusion of this sound change. This is partly due to a shortcoming of the corpus—for /eː/ and /øː/, only a single word was available in the non-/l/ condition ("fee" and "beuk", respectively)—but for the words that are available, the variation did not turn out very large. Significant evidence of lexical diffusion is found in the /oː/ vowel (in regions FB, FE, and FW) and for the /œy/ vowel (in regions NN and FE). Only the former result is geographically contiguous, spanning all Flemish regions minus Flemish Limburg.

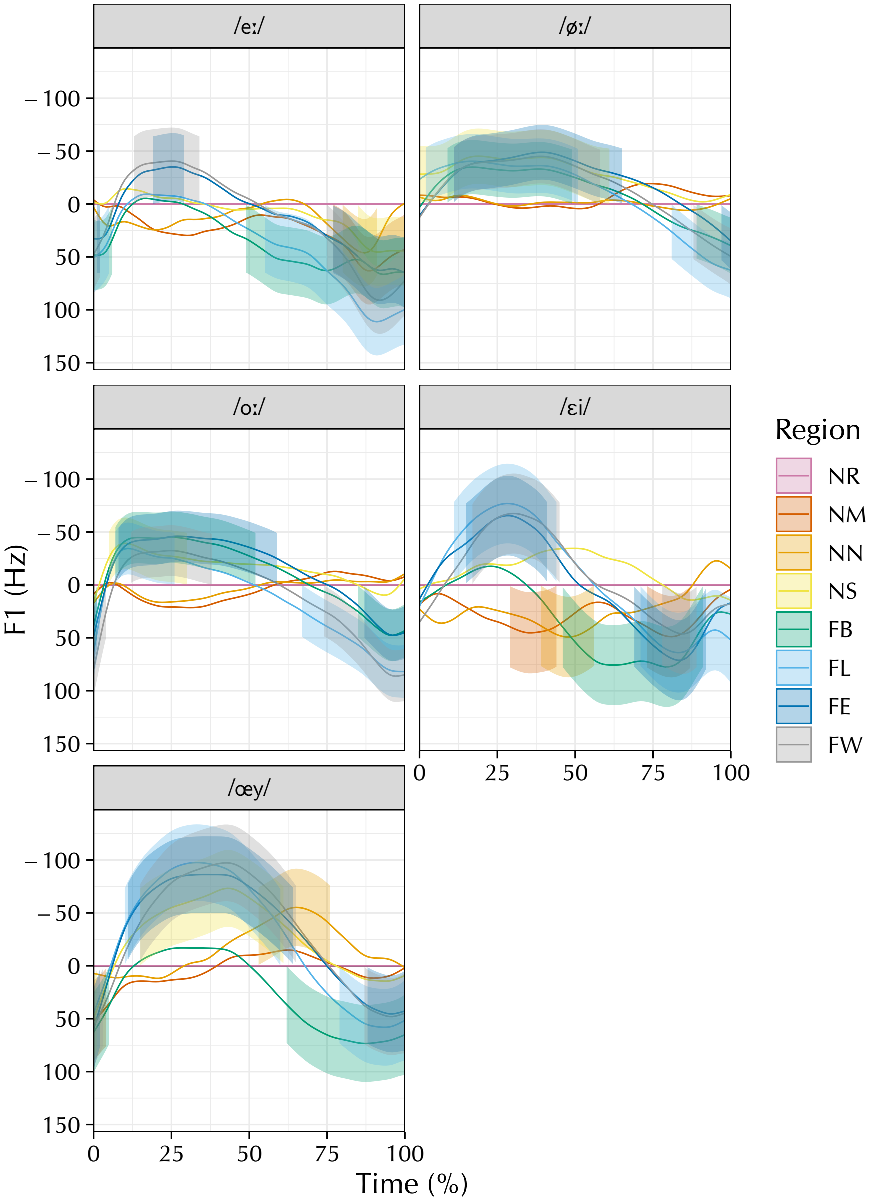

Figure 2.5: Differences in vowel diphthongization before nonapproximant consonants by the separate regions, relative to the Netherlands-Randstad region.

Table 2.2: Regional differences in diphthongization before nonapproximant consonants. The "Timespan" column reflects the start and end point of the consecutive stretch of largest significant differences from the Randstad; the column labeled "95% CI" gives the 95% Bayesian credible interval of this largest difference. Only significant results are shown. The two right-hand columns are the lexical-diffusion measure; the χ^2 measures the amount by which individual words deviate from the peak difference in the middle column.

Vowel	**Region**	**Timespan (%)**	**Peak diff. (Hz)**	**95% CI (Hz)**	χ^2	p
/eː/	NM	77 – 100	63.09	31.21 – 94.97		
/eː/	NN	80 – 91	46.19	14.26 – 78.11		
/eː/	NS	85 – 100	45.38	13.53 – 77.24		
/eː/	FB	49 – 100	66.49	34.35 – 98.63		
/eː/	FL	55 – 100	111.26	79.35 – 143.17		
/eː/	FE	77 – 100	65.03	31.87 – 98.19		
/eː/	FW	75 – 100	90.97	59.18 – 122.77		
/øː/	FB	89 – 100	39.35	12.77 – 65.92		
/øː/	FL	81 – 100	62.51	35.92 – 89.10		
/øː/	FW	88 – 100	49.51	22.94 – 76.09		
/oː/	FB	85 – 100	47.69	23.17 – 72.20	11.19	.01
/oː/	FL	67 – 100	81.86	57.10 – 106.62	2.59	.46
/oː/	FE	87 – 100	47.68	23.29 – 72.06	50.94	<.001
/oː/	FW	74 – 100	86.24	61.86 – 110.62	22.73	<.001
/ɛi/	NM	73 – 87	48.89	10.39 – 87.40	3.49	.06
/ɛi/	FB	46 – 91	77.42	39.50 – 115.35	0.89	.35
/ɛi/	FL	71 – 100	64.02	26.26 – 101.79	2.62	.11
/ɛi/	FE	69 – 92	71.23	33.39 – 109.07	2.80	.09
/ɛi/	FW	78 – 89	46.24	8.26 – 84.22	1.02	.31
/œy/	NN	53 – 76	−37.49	−74.12 – −0.86	14.07	<.001
/œy/	FB	62 – 100	73.40	37.06 – 109.73	3.06	.22
/œy/	FL	79 – 100	58.08	21.86 – 94.31	2.64	.27
/œy/	FE	88 – 100	45.43	9.03 – 81.83	6.37	.04
/œy/	FW	87 – 100	48.00	11.50 – 84.49	5.03	.08

2.3.2 Change 2: lowering of /ɛi,œy,ɔu/

Change 2, the lowering of /ɛi,œy,ɔu/ to [ai,ɒy,ɑu], concerns the same modeled differences as change 1, so the relevant differences can also be observed in Figure 2.5. Because change 2 is concerned with the nuclear vowels of the diphthongs, only differences *before* 50% realization are considered for interpretation. Differences that remain confined to the first 10% are excluded. As before, the differences are relative to the Netherlands-Randstad and are based only on the non-/l/ data. The peaks of the significant differences from the Randstad are listed in Table 2.3.

The significant differences are mostly confined to the Flemish regions, which overall have higher starting points (lower F1s) than in the Randstad. For the vowel /eː/, this is the case for the regions FE and FW, for the vowels /øː/ and /oː/, it is true of all four Flemish regions, and for the diphthongs /ɛi/ and /œy/, it holds for all Flemish regions but Flemish-Brabant. For the vowels /øː,oː,œy/, Netherlands-Limburg goes along with the Flemish regions, in having higher F1s. Finally, for the /ɛi/ vowel, the regions Netherlands-Middle and Netherlands-North have a significantly lower F1 than the Netherlands Randstad.

The lexical-diffusion chi-squares in Table 2.3 again provide limited evidence for lexical diffusion. These indicate the degree to which the peak differences in Table 2.3 are variable between the different words in the corpus. Significant χ^2s are found for the /oː/ (regions NS, FB, FL, and FW) and /ɛi/ vowels (regions NN and FE) and for /ɛi/ in Netherlands-North and Flanders-East. The /oː/ vowel additionally shows marginal signs of lexical diffusion in Flanders-East ($p = .06$), in which case this vowel forms a coherent group: all of Flanders plus Netherlandic Limburg. Of the lexical-diffusion pattern found in the /ɛi/ vowel, the same cannot be said.

Table 2.3: Regional differences in the lowering of /ɛi,œy,ɔu/. Only significant results are shown.

Vowel	**Region**	**Timespan (%)**	**Peak diff. (Hz)**	**95% CI (Hz)**	χ^2	p
/eː/	FE	19 – 29	−35.05	−66.97 – −3.13		
/eː/	FW	13 – 34	−40.52	−72.26 – −8.79		
/øː/	NS	0 – 61	−45.22	−71.35 – −19.10		
/øː/	FB	9 – 49	−34.84	−60.47 – −9.22		
/øː/	FL	2 – 51	−40.42	−66.14 – −14.71		
/øː/	FE	11 – 65	−48.90	−74.78 – −23.02		
/øː/	FW	11 – 58	−45.07	−70.64 – −19.49		
/oː/	NS	5 – 30	−39.13	−63.54 – −14.72	11.04	.01
/oː/	FB	7 – 52	−44.99	−69.30 – −20.68	32.89	<.001
/oː/	FL	7 – 26	−34.48	−58.99 – −9.98	9.38	.02
/oː/	FE	8 – 59	−46.02	−70.22 – −21.83	7.55	.06
/oː/	FW	12 – 38	−32.45	−56.44 – −8.46	17.37	<.001
/ɛi/	NM	29 – 44	39.01	0.69 – 77.33	3.30	.07
/ɛi/	NN	39 – 56	38.71	0.72 – 76.71	4.24	.04
/ɛi/	FL	11 – 44	−76.94	−114.58 – −39.30	0.33	.57
/ɛi/	FE	15 – 41	−65.47	−103.16 – −27.77	9.13	<.01
/ɛi/	FW	17 – 45	−67.50	−105.31 – −29.68	1.28	.26
/œy/	NS	15 – 61	−73.24	−109.58 – −36.91	3.87	.14
/œy/	FL	10 – 60	−97.61	−133.71 – −61.51	1.01	.60
/œy/	FE	11 – 64	−86.28	−122.42 – −50.14	0.76	.68
/œy/	FW	15 – 65	−97.30	−133.56 – −61.04	3.15	.21

2.3.3 Change 3: blocking of diphthongization before /l/

Table 2.4 lists the ΔTTPs between the non-/l/ and the /l/ contexts, and their differences from the Netherlands-Randstad. The lexical-diffusion χ^2s concern the between-word variation in these ΔNR scores.

Similarly to the previous results, it is mostly the Flemish regions where ranges that are significantly different from the Randstad are found. For the /eː/ and /œy/ vowels, three of the four Flemish regions have significantly different ΔTTPs from the Netherlands-Randstad, and for /eː/ so does the Netherlands-South region. For the vowel /øː/ all four of the Flemish regions differ significantly. For the /oː/ vowel, only one region differs significantly from the Randstad: Flemish Limburg; the same also holds for the vowel /ɛi/. The directions of the differences call for some discussion. The hypothesized change 3 was one whereby vowels followed by non-/l/ would be diphthongized more strongly than vowels followed by /l/, but the ΔTTPs in Table 2.4 largely go into the opposite direction (the only exception is the /eː/ vowel). The ΔNRs, by contrast, are exactly as expected: all regions that are significantly different from the Randstad have more positive trough-to-peak differences than the Randstad, indicating a less severe distinction between the non-/l/ and /l/ conditions. Section 2.4 will discuss possible explanations.

The lexical-diffusion χ^2s for the peak differences are large. This is because there is variation from the TTPs for the vowels preceding coda /l/ and also for the vowels preceding non-/l/. The fact that the *combinations* of these two sources of variation are what has to be considered means that the χ^2 values will be larger, but so will their degrees of freedom and hence their p-values. The fact that all of the relevant χ^2 values are significant thus suggests that these differences from the Netherlands-Randstad are quite variable between words, indicating that change 3 is lexically diffuse.

Table 2.4: Differences in the ranges of diphthongization before nonapproximant consonants versus before coda /l/, split out by vowel and region in order to answer RQ 3. Only regions significantly different from the Netherlands-Randstad are shown.

Vowel	Region	ΔTTP	95% CI	ΔNR	95% CI (Hz)	χ^2	p
/eː/	FB	−26.61	−60.69 – 7.48	59.31	10.63 – 107.99	33.96	<.001
/eː/	FL	−10.84	−45.13 – 23.44	75.07	26.25 – 123.90	47.30	<.001
/eː/	FW	−10.82	−44.78 – 23.14	75.09	26.50 – 123.69	40.83	<.001
/eː/	NS	23.67	−8.68 – 56.02	109.59	62.10 – 157.07	76.38	<.001
/øː/	FB	56.67	31.55 – 81.78	47.73	10.94 – 84.52	49.31	<.001
/øː/	FE	50.52	22.29 – 78.75	41.59	2.61 – 80.57	24.40	<.001
/øː/	FL	56.44	31.28 – 81.60	47.51	10.69 – 84.33	8.71	<.01
/øː/	FW	55.85	31.50 – 80.19	46.91	10.65 – 83.18	3.90	.048
/oː/	FL	73.59	46.04 – 101.14	47.89	7.40 – 88.37	52.30	<.001
/ɛi/	FL	61.07	8.63 – 113.50	82.57	6.91 – 158.23	16.50	<.001
/œy/	FB	72.76	21.53 – 124.00	107.27	35.92 – 178.62	106.03	<.001
/œy/	FE	54.07	2.56 – 105.59	88.58	17.02 – 160.13	193.81	<.001
/œy/	FL	37.58	−12.55 – 87.71	72.08	1.52 – 142.64	78.99	<.001

2.3.4 Change 4: vocalization and retracting effect of coda /l/

The fourth change concerns the second formant, rather than the first; Figure 2.6 shows the significant differences from the Netherlands-Randstad. As the coda /l/ was not separated from the vowel, it is included in this figure. Table 2.5 provides a summary of the significant differences relative to the Netherlands-Randstad that are visible. It can already be seen from Figure 2.6 that quite a few significant differences start at the very first few timepoints, which is consistent with coda /l/ having a retracting effect on the quality of the entire vowel.

Table 2.5 shows very clear results, which can be summarized as follows. All vowels in all Flemish regions exhibit significantly less retraction (i.e. higher F2) than the same vowels do in the Netherlands-Randstad. The Netherlands-North (for the vowels /eː/ and /oː/) and the Netherlands-South (for the vowels /ɛi/ and /œy/) also demonstrate some sporadic differences, but these are not very meaningful for interpretation, as they occupy relatively small stretches of signal and the 95% CIs only just exclude zero, neither of which is true for the massive differences from the Randstad region in the Flemish regions. These regions show large effects (median difference = 574 Hz) over, in many cases, nearly the complete vowel-/l/ trajectory.

There is some evidence for lexical diffusion. This is particularly the case for the /oː/ vowel, which is lexically diffuse in all four of the Flemish regions. Lexical-diffusion results for the other vowels are a bit more haphazard. The /øː/ vowel shows significant lexical diffusion in Flemish Brabant and Flanders-West, and the /œy/ vowel does so in Flanders-East.

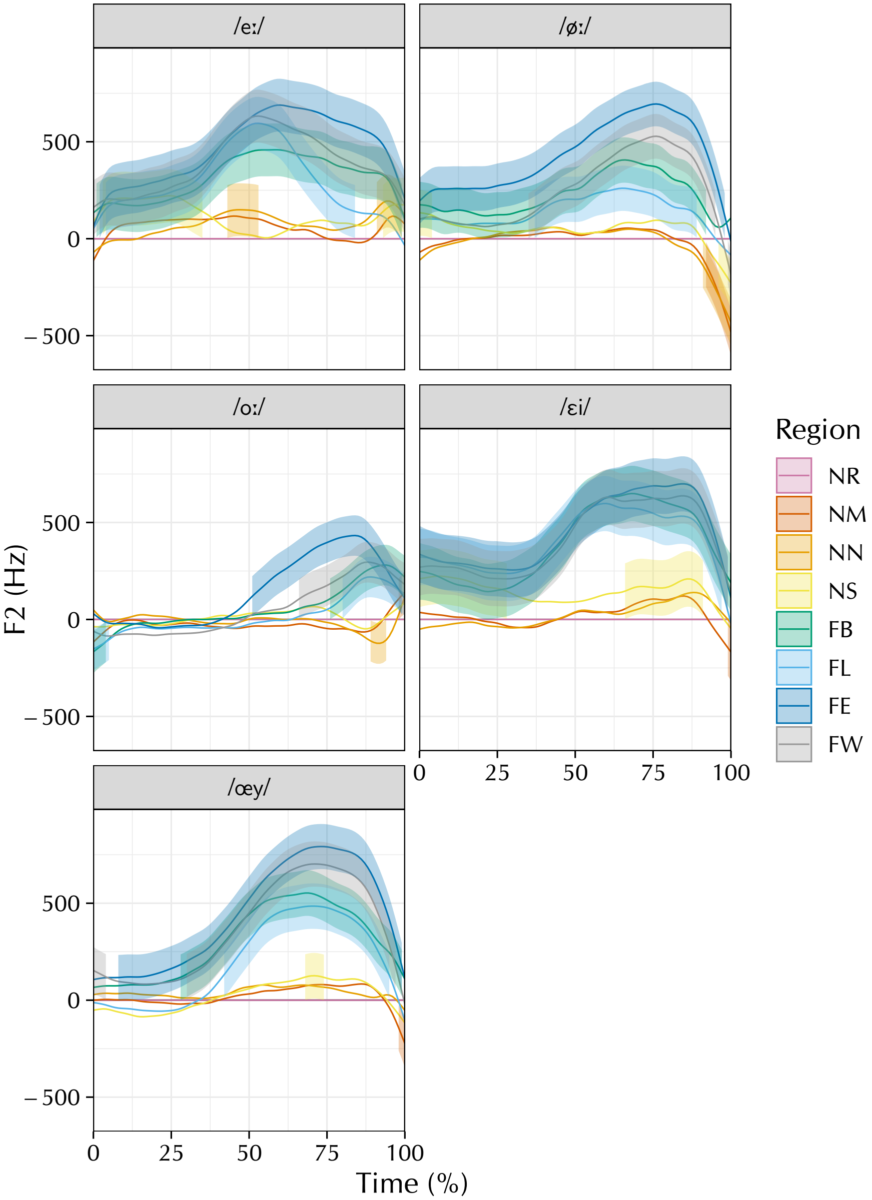

Figure 2.6: Differences from the Randstad in the retraction of vowels including a following coda /l/, averaged over gender.

Table 2.5: Regional differences in the retracting effect of coda /l/.

Vowel	Region	Timespan (%)	Peak diff. (Hz)	95% CI (Hz)	χ^2	p
/eː/	NN	43 – 53	145.88	9.43 – 282.33	1.97	.16
/eː/	FB	1 – 99	458.69	322.60 – 594.79	1.30	.25
/eː/	FL	4 – 84	595.26	458.75 – 731.77	2.49	.11
/eː/	FE	2 – 100	689.18	553.29 – 825.07	0.06	.81
/eː/	FW	0 – 100	632.59	496.38 – 768.79	0.29	.59
/øː/	FB	0 – 92	406.11	289.08 – 523.13	11.21	<.001
/øː/	FL	37 – 89	260.18	143.59 – 376.78	3.27	.07
/øː/	FE	0 – 98	694.84	579.86 – 809.82	0.44	.51
/øː/	FW	35 – 95	528.36	413.00 – 643.71	5.34	.02
/oː/	NN	89 – 94	−110.78	−216.65 – −4.91	2.90	.09
/oː/	FB	76 – 100	262.49	157.64 – 367.35	9.29	<.01
/oː/	FL	80 – 100	219.23	114.69 – 323.76	12.40	<.001
/oː/	FE	51 – 99	432.83	328.59 – 537.07	25.19	<.001
/oː/	FW	66 – 100	295.56	190.87 – 400.25	5.59	.02
/ɛi/	NS	66 – 91	209.11	65.89 – 352.33	0.11	.74
/ɛi/	FB	0 – 100	649.40	506.24 – 792.56	1.59	.21
/ɛi/	FL	0 – 97	597.89	455.55 – 740.23	0.97	.33
/ɛi/	FE	0 – 99	699.22	557.11 – 841.33	2.95	.09
/ɛi/	FW	0 – 99	637.34	494.76 – 779.92	1.97	.16
/œy/	NS	68 – 74	126.21	8.48 – 243.95	3.53	.06
/œy/	FB	28 – 99	552.74	434.87 – 670.62	1.81	.18
/œy/	FL	42 – 95	485.18	367.96 – 602.41	0.08	.78
/œy/	FE	8 – 99	792.47	676.23 – 908.71	7.75	<.01
/œy/	FW	30 – 97	702.30	585.62 – 818.98	0.81	.37

2.4 Discussion

We have discussed four sound changes that are currently on-going in Dutch: (1) the diphthongization of /eː,øː,oː/; (2) the lowering of /ɛi,œy,ɔu/; (3) the monophthongization of diphthongs before coda /l/; (4) the vocalization and retracting influence of coda /l/. The main tenet of the present chapter was that the present-day regional distribution of these sociolinguistic variables could inform us about the current status and the nature of these four sound changes. The results presented in Section 2.3 support this viewpoint.

Change 1, the diphthongization of /eː,øː,oː/, was shown to be subject to significant regional variation. Nearly all Flemish regions were found to diphthongize the five vowels /eː,øː,oː,ɛi,œy/ significantly less than the Netherlands-Randstad. The same is true of the more peripheral vowel–region combinations in the Netherlands. This paints the picture of an on-going sound change that originated in the Randstad and has partially spread towards the other regions in the Netherlands, while affecting very little of Flanders. The between-region effects are phonetically gradual: there are no regions which categorically do not diphthongize their vowels, but there are quantitative differences in the degrees to which they diphthongize. Little evidence was found for lexical diffusion. Thus, according to the data in this corpus, change 1 is phonetically gradual but lexically (mostly) abrupt. Referring back to the typology of sound change reviewed in Table 2.1, this means that this change can be qualified as Neogrammarian. Note that the data show that change 1 is not, in fact, restricted to /eː,øː,oː/ but that /ɛi/ and /œy/ are also involved in the change, in exactly the same way as the tense mid vowels.

Change 2, the lowering of /ɛi,œy/, yielded similar results, mostly concerning the boundary between the Netherlands and Flanders: the Netherlands have undergone the change, but Flanders has not. To a lesser extent (viz. excluding the front vowels), Netherlands-Limburg turned out to be more conservative than the rest of the Netherlands, patterning more with Flanders on this change. This is consistent with change 2 being more recent than change 1, while both originated in the same area (Netherlands-Randstad; Jacobi 2009, Stroop 1998). As with change 1, only weak evidence was found for lexical diffusion while there is substantial intra-country phonetic variation, allowing change 2 to be qualified also as Neogrammarian, as far as the data permit. In addition, the data again reveal that change 2 is not only lowering the nuclei of /ɛi/ and /œy/, but also those of /eː,øː,oː/.

Change 3, the blocking of diphthongs before coda /l/, is a rule change, rather than a change in the locations of the vowels in the articulatory space. This qualitative change has quantitative effects on the difference between a vowel followed by /l/ vs. by another consonant. The results suggest, to some degree, a split between the Netherlands and Flanders, although this split is not perfect and does not hold for all vowels (the vowels /oː/ and /ɛi/ are largely excluded). Most of the Flemish regions were shown to have a significantly less negative difference between vowels in /l/ versus non-/l/ words than the Randstad region. These regions thus make a smaller distinction between these two contexts than the Randstad does. This is in line with the hypothesis that the blocking of diphthongization started out as a sound change in the Randstad area.

Concerning change 3, a final issue remains to be solved. While the differences from the Netherlands-Randstad region were all in the expected direction (ΔNR being positive, indicating less diphthongization in the non-Randstad regions), the range differences between /l/ and non-/l/ themselves were not. The mostly positive ΔTTPs in Table 2.4 suggest that there is *more* upgliding diphthongization before coda /l/, not less. The approach used to extract these scores compared the ranges of diphthongization between F1 peaks and troughs before /l/ vs. non-/l/. In the non-/l/ case, this is not problematic, but in the /l/ case, the F1 will naturally fall at the onset of the /l/, because this consonant requires alveolar occlusion and hence raising of the tongue blade. This will also lower the F1, and will cause it to fall more strongly than an upgliding diphthong, as the latter does not require the tongue tip to make full contact with the alveolar ridge. This contextualizes the findings regarding change 3, in that the positive ΔNRs measured are more likely to have arisen due to differences in the nuclei of the diphthongs (i.e. the first 50% of the trajectory) rather than due to differences in the target positions: the latter are the same for all the /l/ words, hence making these words' contributions to the ΔTTP measure relatively constant. The interpretation of the significant differences in the ΔNRs thus has to be that the Randstad observes a larger distinction in vowel quality between vowels followed by /l/ and vowels followed by non-/l/.

This is in line with change 2, the lowering of /ɛi,œy/. It was shown in Voeten (2015) that, even in the Randstad region of the Netherlands, the realizations of /eː,øː,oː,ɛi,œy/ before coda /l/ were as monophthongs [eː,øː,oː,ɛː,œː], omitting the lowering observed in change 2. It must then be the Netherlands-Randstad, in which area /ɛi,œy/ are lowered the most strongly, where change 3 results in a maximal difference between the /l/ vowel allophone and the non-/l/ vowel allophone. In regions where the lowering of /ɛi,œy/ is less advanced, this difference should be less salient. This is precisely what the results for change 3 show. The results for change 3 thus indicate that the lowering of diphthongs in change 2 is restricted to the non-/l/ condition, where vowels are realized as full upgliding diphthongs. It should also be observed that the degree of lexical diffusion is significant: all but one of the differences from the Randstad that were significant also achieved significant lexical-diffusion χ^2s. Change 3 is thus lexically gradual. Given that it is also phonetically gradual—as demonstrated by the significant inter-region variability—this change can therefore be qualified as a change by exemplars.

Finally, change 4, the vocalization and retracting effect of coda /l/, is the largest of the four changes under investigation. The significant differences are large in magnitude, and in most cases span large sections of the signal, which is consistent with an across-the-board change in vowel quality. The results chiefly suggest a split between the Netherlands and Flanders, with all vowels in all Flemish regions being significantly different from the Randstad for very large stretches of signal. This agrees with observations by van Reenen & Jongkind (2000) that the Flemish varieties of Dutch use a clear coda [l] and the Netherlandic varieties realize a dark [ɫ]. The present results also extend them by showing that this difference is not confined to the realization of coda /l/ itself, but also affects the entire vowel preceding it. There is some evidence for lexical diffusion, although it is really only the /oː/ vowel that stands out in terms of significance, with all four of the Flemish regions obtaining a significant χ^2). In terms of the mode of implementation of change 4, the evidence is thus inconclusive, both in the phonetic dimension and in the lexical dimension. Phonetically, the categorical differences between the Netherlands and Flanders suggest that change 4 is abrupt, but this could also simply reflect a change that has already completed, and lexically, the evidence of lexical diffusion achieves significance only for the /oː/ vowel, but the relatively large χ^2 values do suggest a trend. Future research is necessary.

On the subject of lexical diffusion, the /oː/ vowel presents a noteworthy case. For changes 1, 2, and 4, the evidence for lexical diffusion was haphazard with the exception of this vowel. For some reason, in all of these three otherwise Neogrammarian changes, the /oː/ vowel shows significant between-words variation. This cannot be coincidental, and could be taken to imply that, technically speaking, none of these changes are *truly* Neogrammarian, as there is evidence of systematic lexical diffusion. This point is well-taken, and demonstrates how the categories in Table 2.1 represent only theoretical endpoints of a practical *continuum*. But why is it the vowel /oː/ that consistently shows this high degree of lexical variability between the words in the corpus? The relationship between /oː/ and /ɔu/ may provide an explanation. The lowering of /ɔu/ to /ɑu/ is phonologically complete (as can be seen by the fact that /pɑul/ "Paul" is monophthongized to [pɑːɫ] rather than *[pɔːɫ], at least in Netherlandic Dutch; Voeten 2015) and phonetically more advanced than the lowering of /ɛi,œy/ (Adank, van Hout, & Smits 2004). If the vowels /oː/ and /ɔu/ have become further apart in phonetic space, then this may provide more room for /oː/ to vary due to, for instance, coarticulation, compared to the other vowels under study, which would naturally lead to increased between-word variation. A way to investigate this possibility would be to compute differences between /oː/ and /ɔu/ realizations and to compare their stability across different words using, for instance, the χ^2 statistic. This approach would be analogous to that taken in the investigation of change 3 in the present chapter. Unfortunately, as there was only a single word available for the /ɔu/ vowel in the present corpus and the associated model failed to converge, this must be left to future research using a different dataset or methodology.

A remark on the data used in the present chapter is that the number of words included in the corpus was relatively low. While the data collector (Adank 2003) had made a very deliberate effort to incorporate sociolinguistically and phonologically relevant factors into her design, and the dataset is unique in its thorough representation of the regional variation in Dutch sociophonetics, the number of words per cell in the design ranged from 1 to 5. This made it easy to construct the statistically principled χ^2 measures of lexical diffusion (which would not have been feasible with thousands of words), but generalization of the lexical-diffusion results must be approached with appropriate caution. Despite the clear results found in the present study, which show that the quantitative approach to sound change taken here is promising, the χ^2 measures would have achieved more power if the corpus had contained more words. As such, the lexical-diffusion statistics reported in the present paper represent only a lower bound, constrained by the limited amount of available data.

The results from this study provide new insights into the sociophonetic variation in the Dutch language area. Changes 1 and 2 turned out to be suspectly similar to one another in terms of their behaviors. In particular, change 1 turns out to actually not be restricted to the tense mid vowels /eː,øː,oː/ and change 2 turns out not to be restricted to the diphthongs /ɛi,œy,ɔu/. Instead, both changes demonstrate a more general split between the Netherlands and Flanders, where the latter has generally less open vowels and generally less diphthongization than the former. These observations are consistent with the idea that changes 1 and 2 are actually facets of a single, larger-scale, Neogrammarian change, that originated in the Netherlands and has subsequently not spread uniformly across the two countries. This is briefly touched upon by Jacobi (2009:87), although she does not explicitly posit this theory; the present results suggest that such an integrated account of change 1 and 2 as a single sound change is warranted by the data. The results additionally suggest that change 3 interacts with these two changes, which has been implied before (Voeten 2015), but has only now been shown explicitly. Finally, the results from change 4 highlight the limitations of a synchronic approach to diachronic variation: the observed differences themselves are crystal clear, but these data do not tell us if this is because the change itself is very abrupt, or because it has already completed. A true diachronic investigation would be needed to answer this question.

2.5 Conclusion

The goal of this study was to describe and explain the synchronic regional differences resulting from four diachronic changes in Dutch: the diphthongization of /eː,øː,oː/, the lowering of /ɛi,œy,ɔu/, the blocking of diphthongs before coda /l/, and the increasing vocalization and retraction of coda /l/. These issues were investigated by means of an analysis of the teacher corpus (Adank 2003). The results show that changes 1 and 2 are Neogrammarian, while change 3 was classified as change by exemplars. The evidence suggests that these three changes together constitute different facets of a single on-going vowel shift. Change 4, on the other hand, was of indeterminate status; this change demonstrated the limitations of the synchronic approach to diachronic variation, and is in need of future research. These results show that the approach adopted in the present chapter, when combined with the appropriate statistical tools, can lead to new insights that would not have been obtained with the same efficiency from a real-time diachronic study.

Acknowledgments

I am grateful to my dissertation supervisors, Claartje Levelt and Yiya Chen, for their helpful and insightful advice on various issues, from methodology to writing style, concerning this chapter. I also thank Harald Baayen for warmly receiving me for a one-day stay at Eberhard Karls University in Tübingen and offering his advice on the analyses presented here. I am also indebted to Hans Van de Velde, who provided me with the dataset. Finally, I thank Laura Smorenburg, Meike de Boer, and Willemijn Heeren for helpful discussions on various parts of the data and their analysis.

This work is part of the research programme *Watching Dutch Change* with project number PGW-15-15, which is (partly) financed by the Dutch Research Council (NWO).

CHAPTER 3

How long is “a long term” for sound change? The effect of duration of immersion on the adoption of on-going sound change

This chapter has been submitted.

Abstract

This chapter investigates the adoption of on-going community sound change by individuals, by considering it as an instance of second-dialect acquisition. Four on-going changes in Dutch, all involving the move from one-allophone to two-allophone systems, make this possible: these on-going diachronic changes are simultaneously a source of synchronic variation between Netherlandic Dutch and Flemish Dutch. The chapter investigates the adoption of these differences by “sociolinguistic migrants”: Flemish-Dutch speakers who migrated to the Netherlands to start their university studies. Participants were tracked over the course of nine months, using three sessions of perception and production laboratory-phonological experiments. Results show robust differences from Netherlandic-Dutch controls, which do not diminish over the nine months. While longer-term accommodation to these same changes has been found elsewhere, it appears that nine months is not enough time. The implications of these findings for various subfields of linguistics, particularly sound change and second-dialect acquisition, are discussed.

3.1 Introduction

3.1.1 Investigating the adoption of on-going sound change

This chapter investigates the adoption of on-going community sound change by individual speakers and listeners over the medium term in real time. Research into the processing of such variation and its eventual adoption is made challenging by the fact that the researcher is always "too late": generally speaking, sound changes are so rare to actuate (Stevens & Harrington 2014) that by the time a researcher has identified a certain novel variant as being stable, the sound change has already become well-established (Pinget 2015). Traditionally, sociolinguists and phonologists have therefore had to limit themselves to retrospective studies (often of small groups or even single individuals; Alshangiti & Evans 2011, Bauer 1985, Carter 2007, Cedergren 1987, De Decker 2006, Harrington 2006, 2007, Harrington, Palethorpe, & Watson 2000, 2005, Hinton 2015, Nahkola & Saanilahti 2004, van Oostendorp 2008, Prince 1987, Sankoff 2004, Sankoff & Blondeau 2007, Sankoff, Blondeau, & Charity 2001, Trudgill 1988, Yaeger-Dror 1994) or use a proxy measure or experiment that is assumed to be analogous to genuine sound change (Coetzee et al. 2018, Grosvald & Corina 2012, Pinget 2015, Pinget, Kager, & Van de Velde 2019). Both approaches, while useful, have their limitations. The retrospective approach may be able to consider individuals' speech production if suitable recordings were made (e.g. Labov, Rosenfelder, & Fruehwald 2013, Van de Velde 1996), but cannot also consider the role of their perception. The proxy approach can show experimentally that mechanisms exist that render individuals able to adopt ambient sound changes, but cannot subsequently prove that these are indeed used in real-life situations of on-going sound change. Thus, a gap is left. Do individuals adapt their production and perception to on-going sound change in real time, and if so, in *how much* time?

The present chapter addresses the aforementioned question by studying the production and perception of on-going sound change in a laboratory setting. Four on-going sound changes in Dutch offer an opportunity to do so, thanks to the sociolinguistic situation of the Low Countries. Standard Dutch is spoken in both the Netherlands (henceforth "Netherlandic Dutch", "ND") and the Flemish part of Belgium (henceforth "Flemish Dutch", "FD"). Over the past 100 years or so (at least since Zwaardemaker & Eijkman 1924), the ND tense mid vowels [eː,øː,oː] have changed into upgliding diphthongs [ei,øy,ou] (Van de Velde 1996, Voeten 2015). In tandem with this phonetic change, a phonological change has taken place in these vowels: the diphthongal realizations lose their upglide before, among others, coda /l/ (Berns & Jacobs 2012, Botma, Sebregts, & Smakman 2012, Voeten 2015) and /r/ within the same foot (Gussenhoven 1993). This is true of both the newly-diphthongal tense mid vowels [ei,øy,ou] and the original diphthongs [ɛi,œy,ɔu], which shows that the tense mid vowels have not just changed phonetically, but also phonologically (Voeten 2015). Simultaneously with these two sound changes, the nuclei of [ɛi,œy,ɔu] have begun to lower to [ai,ɒy,ɑu] (Jacobi 2009, Stroop 1998). Finally, the rhotic has developed a novel allophone in coda position, realized [ɹ] and distinct from the other possible rhotic realizations, which are trills, taps, or fricatives (Sebregts 2015). These four sound changes, all of which involve the move from a one-allophone to a two-allophone contrast, are particularly suitable for experimental investigation, because they have all remained confined to Netherlandic Dutch. By contrast, in Flemish Dutch, these sound changes have not taken place at all (Gussenhoven 1999, Sebregts 2015, Van de Velde 1996, Verhoeven 2005, Chapter 2). Thus, the on-going diachronic changes coincide with well-established synchronic variation. This makes these sound changes suitable for synchronic experimental investigation.

3.1.2 Sound change as second-dialect acquisition

This chapter uses the aforementioned synchronic differences to study the adoption of the same diachronic differences. This is done by performing laboratory-phonological experiments with *sociolinguistic migrants*: speakers of Flemish Dutch who have moved to the Netherlands to do their university studies there. A previous large-scale cross-sectional study on the aforementioned four sound changes (Chapter 4) has confirmed that, in the long term (years–multiple decades), these changes are indeed adopted by the sociolinguistic migrants studied there. This (eventual) adoption of the ND sound changes by FD sociolinguistic migrants, used here as a model to investigate the individual adoption of community change, presents a case of second-dialect acquisition. This could be argued of *any* instance of community sound change: if an individual adopts a sound change that has been going on in their environment, then by definition they are adopting a *slightly* different dia- or idiolect. While this chapter's four specific sound changes in isolation might qualify as being "slightly different", it is important to note that these differences are by no means the *only* differences between Netherlandic Dutch and Flemish Dutch. Thus, the adoption of the sound changes of interest is a sub-problem of the larger issue of second-dialect acquisition.

Second-dialect acquisition (henceforth "SDA") is a broad field, and for an extensive overview the reader is referred to books such as Siegel (2010) or specialized reviews such as Nycz (2015). A common theme in SDA research is the low "success rate": the synthesis of many studies given in Siegel (2010) yields an average outcome of 50%, meaning that the odds of a given individual successfully adopting a given second dialect are at chance level. This degree of attainment is influenced by system-internal, individual, and social factors. For example, Siegel (2010) argues that superficial items such as differing lexical items are easier to adopt, and are adopted more often, than abstract features such as grammatical rules. This has also been found in cases of sound change: Sneller (2018) demonstrates how a diachronically innovative simple allophonic rule can rapidly overtake an older, more abstract and irregular, system as a result of contact between the two systems. Individual and social factors similarly overlap between SDA and sound change; those discussed by Siegel (2010) largely overlap with those reported in sound change, such as duration of exposure, cognitive-processing styles, social-network size, and motivation (Beddor 2015, Coetzee et al. 2018, Lev-Ari 2018, Yu 2013, Chapter 4).

SDA has been the subject of substantial experimental investigation, often combining production and perception research, which is also the approach taken in this chapter. Bowie (2000), Evans & Iverson (2007), Nycz (2011) (see also Nycz 2013) and Ziliak (2012) looked at vowel production and perception in American sociolinguistic migrants who had moved out of state post-adolescence. They find heterogenous adoption of the new dialects in production, and next to no adoption in perception—only a small number of individuals in Ziliak (2012) adopted the new dialect's perception, and in Evans & Iverson (2007) no individual perceptual change was found, but there was a stable link between an individual's perception and production. Walker (2014) used production and perception experiments to investigate SDA in sociolinguistic migrants between the US and the UK and found small differences in production as well as perception, that were unidirectional: the American migrants who had moved to the UK had adopted parts of the UK accent, but the reverse was not found. Walker (2014) additionally considered explicit priming effects of conversational topic (i.e. British soccer versus American rugby). These effects were indeed found: after priming with an American topic, sociolinguistic migrants from the US to the UK produced less British variants and performed worse in the British-accented version of the perception task.

The observed similarities between SDA and sound change are brought together in the "change-by-accommodation model" (Auer & Hinskens 2005, Chambers 1992, Trudgill 1986). This model considers an individual's adoption of ambient change (as in sociolinguistic migration and in sound change) to be a long-term extension of the well-known process of phonetic accommodation (Giles, Coupland, & Coupland 1991, Giles & Smith 1979, Giles, Taylor, & Bourhis 1973, Maye, Aslin, & Tanenhaus 2008, Norris, McQueen, & Cutler 2003, Pardo et al. 2012). Under this model, the effects of linguistic priming in SDA (Walker 2014) and sound change (Pinget, Kager, & Van de Velde 2019)—and perhaps also non-linguistic priming of the type in Hay & Drager (2010), although cf. Walker, Szakay, & Cox (2019)—are readily accommodated. However, the evidence for this model has been anecdotal at best (Auer & Hinskens 2005), limited to what Babel, Haber, & Walters (2013:7) call "trends and tendencies". It therefore remains to be seen to what extent this model can deliver. The present study provides an indirect contribution to this question.

3.1.3 The present study

The present study capitalizes on the similarities between second-dialect acquisition and sound change, using the former to investigate the time course of the latter under carefully-controlled laboratory conditions. The on-going changes in Netherlandic Dutch but not in Flemish Dutch make the study possible: the synchronic and diachronic language situation of these varieties provides a unique case in which SDA and on-going sound change coincide in real time. The main question investigated by the chapter is: do individuals adapt their production and perception to on-going sound change in real time? If they do, then how much time is enough? This is studied empirically using experiments in perception and in production, which are performed three times over the course of nine months to investigate the migrant participants' malleability in the medium term. The behavioral experiments reported here are part of a larger battery of behavioral and EEG experiments to study this question. The reader is referred to Chapters 5 and 6 for details on the other tasks that were performed in the experiment sessions reported in the present chapter.

The perception experiment is a rhyme-decision task, used previously in Nycz (2011) and expanded upon here with modifications to the paradigm to test the highly specific coda-/l/ environment, which is important for the sound changes under discussion. The object of investigation is the category boundary between monophthongal and diphthongal vowel phonemes, and glided and non-glided rhotics. Given the sound changes discussed in Section 3.1.1, it is expected that the Netherlandic controls will require a much steeper F1 slope than the Flemish sociolinguistic migrants for them to no longer consider a vowel to be an acceptable monophthong. Conversely, the Flemish sociolinguistic migrants should be used to much less upgliding diphthongization than the Netherlandic controls, and hence be quicker to judge a vowel as diphthongal. However, when a coda /l/ follows, neither group should have an *a priori* expectation of any upgliding diphthongization, and the hypothesized group differences should then become much smaller. The inclusion of the latter condition, in which upgliding diphthongization is always unexpected, makes it possible to separate participants' phonetic interpretation of the degree of upgliding diphthongization present in the stimulus from their knowledge of the phonological differences between Netherlandic Dutch and Flemish Dutch.

The production task is a word-list reading task using real words, commonly used in studies on SDA and on sound change. The task includes a priming component which considers whether a single-vowel manipulation in perception can induce the migrant participants to switch to the Netherlandic realizations. Such single-segment ultra-short-term accommodation has already been established in non-SDA contexts (Zellou, Dahan, & Embick 2017), and would extend the research on short-term accommodation to cross-dialectal linguistic primes (Pinget, Kager, & Van de Velde 2019, Walker 2014). The expectation is that the production experiment will replicate the between-groups dialectal differences known from the literature discussed in Section 3.1.1, but that the Flemish sociolinguistic migrants adapt their realizations in the directions of the Netherlandic controls over the nine months that were measured, as they have been found to do in the long term (years–decades; Chapter 4).

3.2 Experiment 1: rhyme decision

The perception experiment's rhyme-decision task with ambiguous stimuli is similar to the tasks in Nycz (2011) and McQueen (1993). Nycz (2011) successfully used a different kind of rhyme-decision task in an SDA context. McQueen (1993) found that nonsense words were viable for use in rhyme-decision tasks, incurring only an obvious slower RT compared to real words. Being able to use nonsense words in the present task is important, for two reasons. The first is theoretical: real words will be subject to the Ganong effect (Ganong 1980), by which top-down knowledge is used to repair bottom-up ambiguity; this is precisely what we do not want. The second is practical: by using nonsense words one can easily synthesize as many varied tokens as needed. Nonsense words are used in which the critical vowel or rhotic is replaced with an intermediate variant, generated by morphing together two naturally-produced endpoint sounds. Participants are asked whether differently-morphed intermediate stimuli rhyme with orthographically-presented target words that clearly contain a monophthong phoneme or clearly contain a diphthong phoneme.

By presenting the target words orthographically, participants are required to construct their own phonological representations of these words, to which they then need to compare the auditory stimuli. These stimuli are created using holistic morphing (Kawahara et al. 2008), rather than using formant synthesis, because Dutch diphthongs utilize complex trajectorial information, which would be difficult to discretize for synthesis purposes. While the primary cue for the phonological category of diphthong in Dutch is uncontroversially defined as a downward trend in F1 over time (Booij 1995, van Oostendorp 2000), the precise temporal dynamics of the F1 slope are much more complex, and this information is used by listeners as secondary cues (Peeters 1991). By starting from natural speech, the morphing procedure automatically takes these cues into account, and produces a more natural result than attempting to generate the requisite trajectories synthetically.

The experiment was piloted before being performed with the participants reported in this chapter. A particular concern was whether the manipulation would be effective with rhotics. The on-going sound change diphthongizing the tense mid vowels brings them closer to the original diphthongs, which exist as separate phonemes, so that a strongly diphthongized realization of /eː/ might be interpreted as a weakly-diphthongized /ɛi/. For the rhotics, by contrast, Dutch has only one phoneme and there is no possibility for phonemic confusion. However, it had been suggested to the author by multiple departmental colleagues that it is the phonological surface form, not the phonemic representation, that decides whether two words rhyme. The pilot indeed bore out this result, showing a very similar trajectory to the one in Figure 3.4, in that individuals whose own rhotic is glided reported more positive rhyme decisions the more glided the auditorily-presented stimulus was. For the vocalic conditions, results of the pilot were similarly positive.

The hypotheses for the perception task for the Flemish-Dutch sociolinguistic migrants and Netherlandic-Dutch controls are the following. When presented auditorily with a word containing a variable amount of upgliding diphthongization, it is expected that the sociolinguistic migrants are more likely to interpret this word as containing a diphthong phoneme, whereas the Netherlandic controls are more likely to interpret the vowel realization as reflecting a monophthong phoneme. When asked whether this word rhymes with a visually-presented target word containing either a diphthong or a monophthong phoneme, the sociolinguistic migrants should be more likely to say "yes" to the former and "no" to the latter, whereas for the Netherlandic controls, this prediction is reversed. In the case of the coda rhotic, the Netherlandic controls should show a preference for the glided realization, and the sociolinguistic migrants should show a preference for the trilled realization. These predictions constitute the first hypothesis of this experiment. The second hypothesis is that the between-groups differences in the vowels become smaller before coda /l/. In this condition, neither the sociolinguistic migrants nor the Netherlandic controls have an *a priori* reason to expect any diphthongization to be present in the stimulus, so any perceptual compensation for upgliding diphthongization that the groups perform should not apply here, resulting in more diphthong-phoneme percepts. However, even with this phonological knowledge taken out of the picture, participants' phonetic knowledge should still be able to play a role, such that the hypothesis cannot be that the group differences even out before a coda /l/; they should just become smaller. Finally, the third hypothesis in this experiment is that over the course of nine months, the sociolinguistic migrants will become more used to the Netherlandic realizations and hence the group differences will become smaller over the course of time.

3.2.1 Method

Participants

The participants were the same as in Chapter 6: 10 Netherlandic-Dutch speakers who were students at the universities of Leiden (LU) and Amsterdam (UvA) who served as the control group, and 10 Flemish-Dutch-speaking sociolinguistic migrants who had recently migrated from Flanders and were first-year students at the same universities. The sociolinguistic migrants were tested first, as close to the beginning of the academic year as possible (mean number of days past September 1st = 21.5 days, SD = 7.93 days). The control group was tested later (mean number of days past September 1st = 104.30 days; SD = 54.40 days). As in Chapter 6, the experiment was run three times, each time with the same participants barring dropouts, over the course of nine months. The mean interval between the first two sessions was 129.29 days (SD = 23.19 days), and the mean interval between the last two sessions was 112.75 days (SD = 22.94 days). Between the first two sessions, one control participant and two sociolinguistic migrants dropped out; for the final session, a single additional sociolinguistic migrant dropped out. Note that drop-outs were not given special treatment in the data; their followup responses are simply considered censored[1]. Table 3.1, copied from Chapter 6 (with a small change, because on one occasion EEG data collection failed but the behavioral data for the present study were collected successfully), summarizes the final population from which data were obtained. In this table and in the remainder of the text, the sociolinguistic migrants will be referred to as "FDS", for "Flemish-Dutch students", and the control participants will be labeled "NDS", for "Netherlandic-Dutch students".

The experiments followed the Ethics Code for linguistic research in the faculty of Humanities at Leiden University, which approved its implementation. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Stimuli

The auditory stimuli were 238 pseudo-word pairs that contained one of seven phones or phone sequences: [eː∼ɛi], [oː∼ɑu], [ɛ∼ɛi], [aːʀ∼aːɹ], [eːɫ∼ɛiɫ], [oːɫ∼ɑuɫ], [ɛɫ∼ɛiɫ]. These conditions are listed schematically in Table 3.2; from there on, the [aːʀ∼aːɹ] condition will be referenced as [ʀ∼ɹ], as it is only the rhotic that is of interest. All pseudowords were disyllabic according to a [C_1_.C_1ə(ɹ)] template; the final [ɹ] was present in exactly half of the words pre-

---

[1] The inclusion of by-participants random effects in all analyses used in this chapter means that the retention of these participants' data points will not bias the group-level results even at the sessions that contain censored data points.

Table 3.1: Overview of the final population from which data were obtained. "FDS" indicates a Flemish-Dutch speaker (i.e. a sociolinguistic migrant); "NDS" indicates a Netherlandic-Dutch speaker (i.e. a control participant).

	Session		
Participant	1	2	3
FDS-0	✓	✓	✓
FDS-1	✓	✓	
FDS-2	✓		
FDS-3	✓	✓	✓
FDS-4	✓	✓	✓
FDS-5	✓	✓	✓
FDS-6	✓	✓	✓
FDS-7	✓		
FDS-8	✓	✓	
FDS-9	✓	✓	✓
NDS-0	✓	✓	✓
NDS-1	✓	✓	✓
NDS-2	✓	✓	✓
NDS-3	✓	✓	✓
NDS-4	✓	✓	✓
NDS-5	✓	✓	✓
NDS-6	✓	✓	✓
NDS-7	✓		
NDS-8	✓	✓	✓
NDS-9	✓	✓	✓

Table 3.2: Schematic overview of the conditions in the rhyme-decision task. Because the vocalic conditions all appeared twice (with vs. without a following coda /l/), the four rows in the table together make seven conditions for the experiment. Each condition consisted of 34 items.

Left endpoint	Right endpoint	Possible coda /l/
[eː]	[ɛi]	Yes
[oː]	[ɑu]	Yes
[ɛ]	[ɛi]	Yes
[ʀ]	[ɹ]	No

sented. The two consonants were chosen from all possible phonemes available in Dutch, with one exception: for the four contrasts ending in consonants, the second C was fixed to be /d/. Candidate pseudoword pairs were removed from the list if one of their elements turned out to be a real word. The remaining list of pseudowords was sorted to maximize first the combined syllable probability of both syllables for both elements of each pair, and secondly the combined phoneme transitional probabilities for both elements of each pair. For each of the seven conditions, the 34 best pairs were then selected.

Each of the resulting 476 words was recorded in a carrier sentence by a trained female speaker who normally uses a Randstad accent. The phoneme or phoneme sequence of interest was extracted from each of these words, and of the resulting 34 tokens per contrast per condition available, the most prototypical was selected based on formant measurements. The criteria for this prototypicality were as follows: for the non-upgliding vowels, the token that showed the smallest difference in F1 at 25% vs. 75% realization was chosen; for the upgliding vowels, the vowel with the largest difference was chosen. For the [ʀ∼ɹ] contrast, the criterion for the latter realization was that the F3 measured at the midpoint of the rhotic should be as low as possible, whereas for the former it should be as high as possible. Using Tandem-STRAIGHT (Kawahara et al. 2008), the two endpoints for each of the seven phonological contrasts were morphed holistically into four intermediate ambiguous tokens containing either 20%, 40%, 60%, or 80% upgliding diphthongization (for the vowels) or gliding (for the rhotic). These tokens were then cross-spliced into the 34 stimuli, after which any resultant discontinuities in pitch were smoothed out using PSOLA. As an example, Figure 3.1 shows the waveforms, spectrograms, and F1 trajectories (the critical difference between upgliding and non-upgliding realizations) for the [eː∼ɛi] contrast.

For each of the 34 words per contrast, the four possible variants were yolked across participants, as illustrated in Table 3.3. Each of these auditorily-

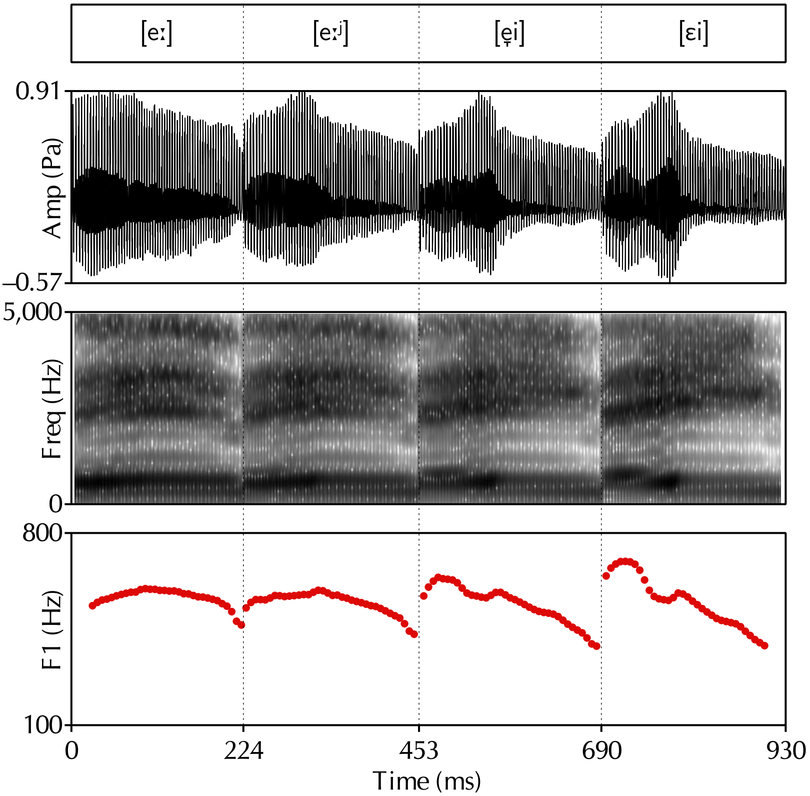

Figure 3.1: Example waveforms, spectrograms, and F1 trajectories for the [eː∼ɛi] contrast. The four tokens shown in this figure correspond to the stimuli containing 20%, 40%, 60%, and 80% morphing, respectively. Note the lowering of the nucleus and the increase in the diphthong's slope over the four figures.

Table 3.3: Example of four words yolked across four participants. The percentages refer to the amount of upgliding diphthongization ([eː]→[ɛi]) present in the speech signal.

	Participant			
Word	1	2	3	4
[deːtə]	20%	40%	60%	80%
[bleːtə]	40%	60%	80%	20%
[χʀeːvəɹ]	60%	80%	20%	40%
[tʋeːdəɹ]	80%	20%	40%	60%

presented tokens was presented twice: once in combination with a visually-presented word that would rhyme only if the participant had auditorily perceived the left-endpoint phoneme, and once again with the same visual word modified to rhyme only if the participant had auditorily perceived the right-endpoint phoneme. A diagram of the structure of each trial is provided in Figure 3.2. The visually-presented words were selected in the same manner as the auditorily-presented words, except with the obvious additional requirement that, where necessary, their C2 should be the same as their auditory counterpart's to be sure that the two would be able to rhyme. In cases where the syllable probabilities and/or transitional probabilities of the auditory and visual elements of the rhyme pair were not the same, the word with the highest syllable/transitional probabilities of the two was selected for the visual word.

In total, each of the participants judged 2×238 = 476 word pairs. To prevent the experiment from becoming too long, no explicit filler items were included. Instead of including fillers, the 7-contrast nature of the design was exploited: each of the seven contrast is considered to be "filled" by the remaining six contrasts. The only difference with a truly-filled design is that, in this case, what are considered fillers with respect to one contrast can simultaneously be analyzed as target items with respect to another contrast. The full list of items is available in Appendix A.

Procedure

This experiment was part of a larger battery of tests which included an EEG component. Participants performed, in order, the perception task reported here, a passive-oddball task (with EEG, which is reported in Chapter 5), and the production task reported in Section 3.3 (with EEG, which is reported in Chapter 6). The testing thus took place in a sound-attenuated EEG booth. Participants were seated in front of a computer screen flanked on both sides by a loudspeaker box. Two buttons had been built into the armrests of their chair;

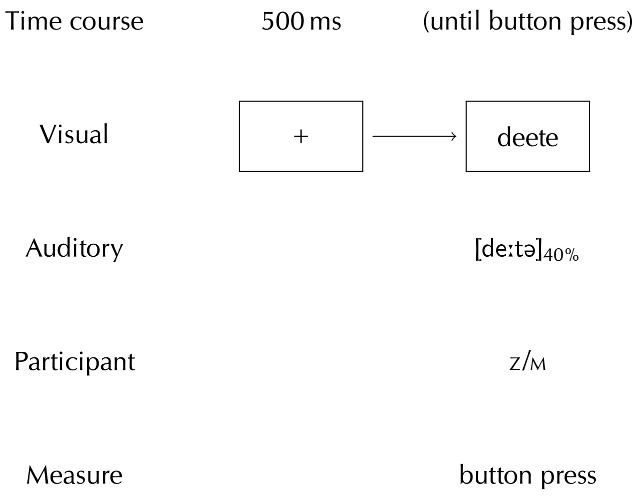

Figure 3.2: Example trial for the rhyme-decision task. At a later trial, the same auditory stimulus will be presented, but the visual target ⟨deete⟩ (/deːtə/) will be replaced by ⟨dijte⟩ (/dɛitə/).

the left armrest's button was labeled "z", and the right armrest's button was labeled "m". In between the computer screen and the participant, a microphone was positioned on a microphone stand. Before the start of the experiment, instructions were presented on the computer screen and were also presented auditorily via a recording of a male speaker of Netherlandic Standard Dutch who read them aloud. Each trial started with the auditory presentation of an ambiguous word, followed by visual presentation of a target whose rhyming with the auditory stimulus depended solely on the phoneme perceived by the participant for the prime. Participants used the armchair buttons to indicate whether they considered the prime or target to rhyme—one button indicated yes, the other indicated no; which of the two meant which was randomized across participants. An example trial is shown in Figure 3.2. Between the trials, a fixation cross was presented for 500 ms.

The whole experiment consisted of 476 trials: each auditory stimulus was presented twice, once matched with a target rhyming with the left-endpoint phoneme, and once again matched with a target rhyming with the right-endpoint phoneme (randomized and counterbalanced across participants). There were four breaks, spaced evenly across the trials. Before the actual experiment began, participants did a practice block consisting of a miniature version of the actual experiment; for each of the seven contrasts, the rhyme pair that had the lowest syllable probability of those selected was used in these practice trials. Given that, just as in the real experiment, all stimuli were presented twice, this made for a total of $7 \times 2 = 14$ practice trials.

Data analysis

Responses with reaction times <100 ms or >5 s were excluded from further processing. The remaining "yes"/"no" rhyme decisions were recoded into "phoneme A"/"phoneme B" decisions for all pairs of primes and targets. Mixed-effects regression trees (see Tagliamonte & Baayen 2012 for an accessible introduction to the closely-related conditional-inference trees in linguistics) were used to determine which factors influenced participants' perceptions of the upgliding realizations [ɛi,ɑu,ɛi] and the glided rhotic [ɹ], relative to the non-upgliding realizations [eː,oː,ɛ] and the trilled rhotic [ʀ]. Function `glmertree` from the eponymous R (R Core Team 2020) package (Fokkema et al. 2018) was used to fit a logistic mixed-effects regression tree for each of these four conditions separately. The trees included fixed effects for "Step", "Following consonant" (/l/ or non-/l/; reported as "Following" in Figs 3.3–3.6) "Group" (FDS or NDS), and "Session". The splitting criterion was Bonferroni-corrected to $\alpha = .0125$. Random intercepts and slopes by all predictors were included for participants and items (rhyme pairs). The random-effects covariance matrix was constrained to be diagonal. Function `buildmertree` from R package `buildmer` (Voeten 2019b) was used to identify the maximal random-effect structure that achieved non-singular convergence, with terms selected for inclusion based on their contribution to the AIC (Akaike 1973) of the tree.

3.2.2 Results

Figures 3.3–3.6 show the four mixed-effects regression trees resulting from the analysis. For the [eː∼ɛi] contrast, the first and hence most important split made by the model is one between the first two steps and the last two steps of the continuum. Starting with the left branch, a distinction is made between the two types of following consonant: if this is not /l/, the tree terminates with a 14.4% probability of the participant reporting a percept consistent with the diphthong phoneme. A following /l/, however, induces a split between the FDS and the NDS group, with the former reporting 16.5% diphthong percepts, but the latter reporting significantly more at 26.7% diphthong percepts. At the later two morphing steps, the largest difference is made by the following consonant: if this is not /l/, the FDS path through the tree terminates at only 19.2% diphthong percepts. This is different for the NDS participants: they continue to divide the morphing continuum into steps 3 and 4, arriving at more diphthong responses than the FDS participants in both steps (27.3% and 34.7%, respectively). If the following consonant is /l/, the same picture is obtained,

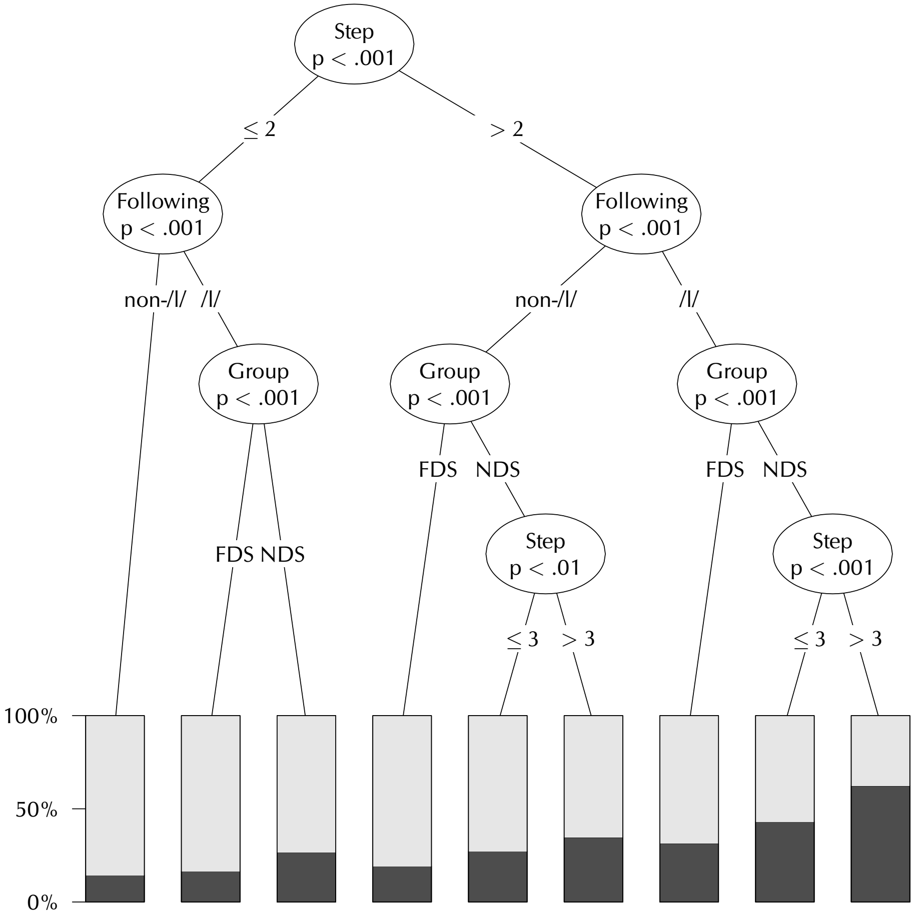

Figure 3.3: Logistic mixed-effects regression tree for the [eː∼ɛi] continuum (20 participants, 68 items). The target variable is the probability of indicating an [ɛi] percept.

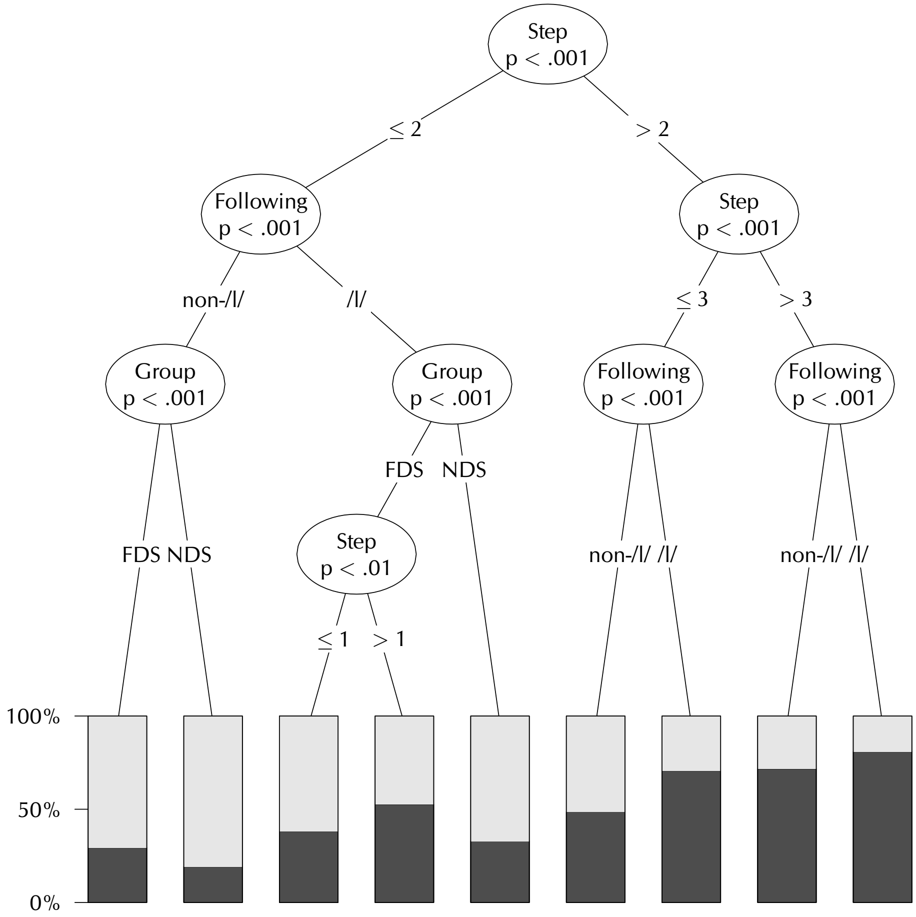

Figure 3.4: Logistic mixed-effects regression tree for the [oː∼ɑu] continuum (20 participants, 68 items). The target variable is the probability of indicating an [ɑu] percept.

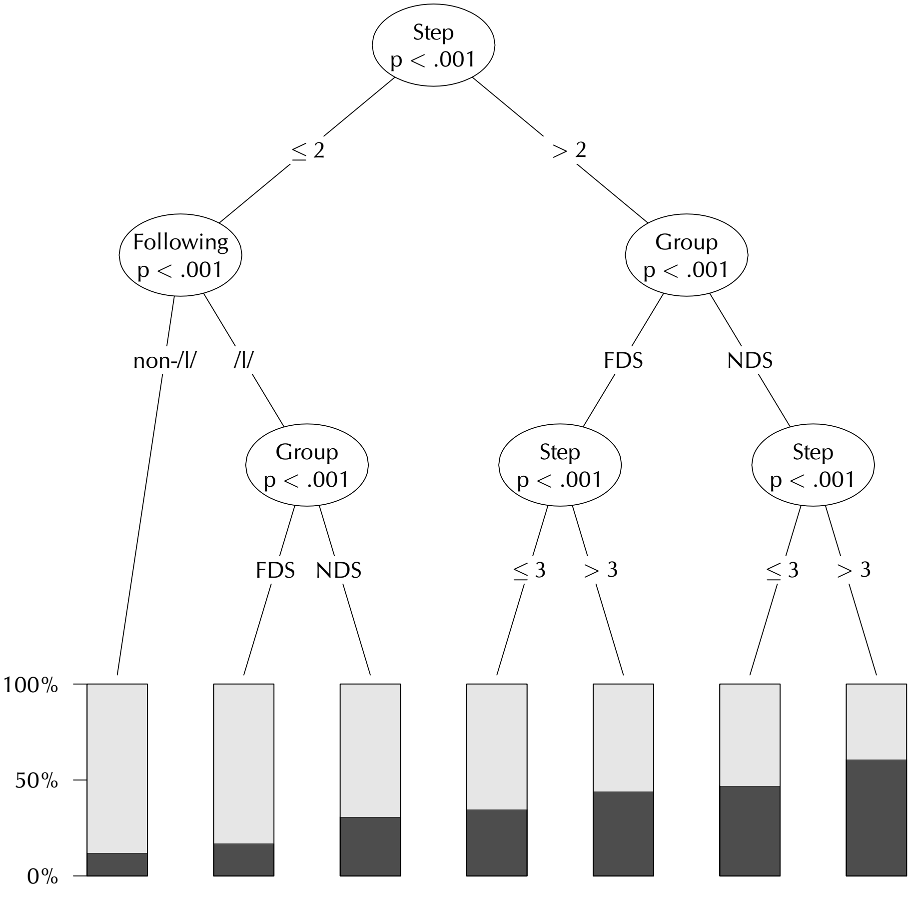

Figure 3.5: Logistic mixed-effects regression tree for the [ɛ∼ɛi] continuum (20 participants, 68 items). The target variable is the probability of indicating an [ɛi] percept.

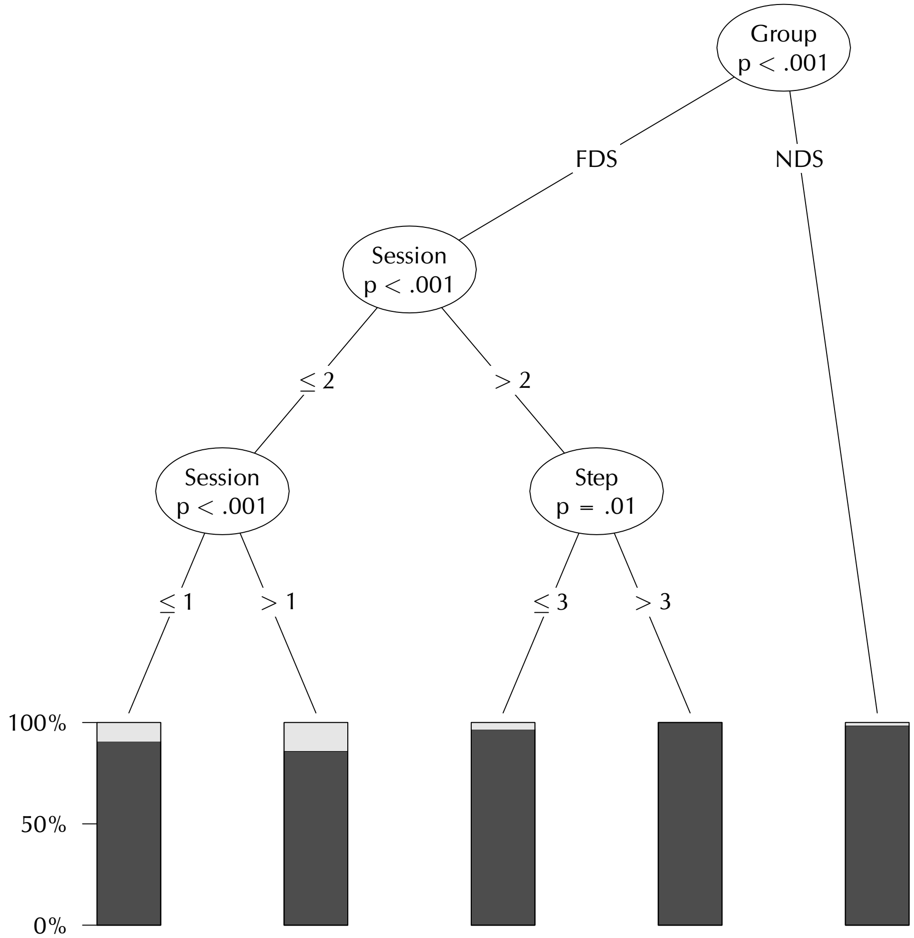

Figure 3.6: Logistic mixed-effects regression tree for the [ʀ∼ɹ] continuum (20 participants, 34 items). The target variable is the probability of indicating an [ɹ] percept.

with an additive effect included for the overall larger probability of reporting a diphthong percept before /l/. Here, the FDS path again terminates directly, this time at 31.5% diphthong responses, whereas the NDS further carve up the continuum into steps 3 and 4, arriving at respectively 43.1% and 62.3% diphthong responses.

In the [oː∼ɑu] condition, the most important predictor is again a split between the first two morphing steps and the latter two. Starting with the <2 branch, the next predictor is the consonant following the vowel. If this is not /l/, the tree terminates with a final split between the two groups, where the FDS report 29.4% diphthong responses and the NDS report 19.2%. If the following consonant was /l/, the next predictor is again a split between the groups. The NDS then terminate directly with 33.0% diphthong responses, whereas the FDS further distinguish between steps 1 and 2. In step 1, they report 38.2% diphthong responses, and in step 2, they report 52.6%. Turning to the second main branch of the tree, the first split along this branch is made between steps 3 and 4. Both steps are subsequently split up by the consonant following the vowel. If this is /l/, participants give more diphthong responses (step 3: 70.7%, step 4: 80.9%) than if it is not (step 3: 48.7%, step 4: 71.7%).

In the [ɛ∼ɛi] condition, the by now familiar effect of the following consonant is obtained only in the first two morphing steps. If the following consonant is /l/, a between-groups difference is obtained here, such that the NDS report more [ɛi] percepts (30.8%) than the FDS (16.8%). In the final two morphing steps, the first split made by the model is one between the FDS and the NDS, such that the NDS again report more diphthong percepts than the FDS. Both groups additionally report more diphthong percepts in step 4 than in step 3.

Finally, the [ʀ∼ɹ] condition shows a pattern that is very different from the vocalic conditions. The results show a very strong preference for the [ɹ] realization, which is only subtly modulated by the predictors entered into the analysis. The first split is one between the groups, after which the NDS branch of the tree immediately terminates with 98.5% [ɹ] responses. The FDS branch, however, continues on, and the next significant predictor is "Session". In session 1, the FDS report 90.1% [ɹ] percepts; in session 2, they report 85.9%. In session 3, their response depends on the morphing step: at the final step, they opt for [ɹ] 100% of the time, whereas in the steps before that, they do so in "only" 96.7% of the cases.

3.2.3 Discussion

The main research question in Section 3.1 was whether, and in how much time, sociolinguistic migrants adopt new variants caused by on-going sound change. For the perception part, this overarching question was broken down into three hypotheses in Section 3.2. The first of these was that the FDS would be more likely to interpret a stimulus containing even mild upgliding diphthongization as reflecting a diphthong phoneme, whereas the NDS would not be. The second hypothesis was that there would be more diphthong responses for both groups before coda /l/, where the presence of upgliding diphthongization is not expected *a priori* and hence more prominent. The third hypothesis was that the differences between the FDS and the NDS would reduce over the measured nine months' time.

The overall results suggest that for the vowels, the manipulation was successful. The [eː∼ɛi] model reports diphthong-phoneme responses ranging from 14.4% to 62.3%. The reason for the deviation from the theoretically-expected 80% is in the choice of continuum, which ranged from [eː] to [ɛi], rather than from [eː] to [ai], and hence did not take into account the lowering of [ɛi,œy,ɔu] to [ai,ɒy,ɑu]. The [oː∼ɑu] condition "repairs" this, sacrificing step granularity to achieve a wider sampling of the *full* monophthong–diphthong continuum. Here, the percentage of diphthong responses ranges from 19.2% to 80.9%, almost precisely as expected. The [ɛ∼ɛi] contrast manipulated sounds using the same morphing technique as the previous two experimental conditions, but along a dimension which is irrelevant for the on-going sound changes (viz. the combination of upgliding diphthongization with duration). Here, diphthong responses range from 12.1% to 60.7%, which mirrors the [eː∼ɛi] condition.

The results from the [eː∼ɛi] tree confirm the second hypothesis, but seem to refute the first. Both the FDS and the NDS report more diphthong responses as the morphing step increases, and report more diphthong responses preceding coda /l/, where upgliding diphthongization would not be expected by either group. However, the between-groups differences are contrary to the first hypothesis: in the [eː]∼[ɛi] condition it is the *NDS* who consistently (and significantly) report more diphthong responses. By contrast, in the [oː∼ɑu] condition, this between-groups difference reverses. Here, again both the FDS and the NDS report more diphthong responses at later morphing steps, but now it is the *FDS* who consistently report more diphthong responses than the NDS. This effect is in line with the hypothesis that these participants expect more upgliding diphthongization to be present in general, for which they compensate only in the non-/l/ condition, where this diphthongization is prescribed by their phonology. FDS phonology, on the other hand, does not prescribe such upgliding diphthongization, and indeed, the FDS have a much smaller difference between their non-/l/ and /l/ diphthong-phoneme percepts (only 2.1% more diphthong responses in the /l/ condition).

In understanding the contradictory effects between the two aforementioned conditions, it is worthwhile to refer to the control condition [ɛ∼ɛi]. Here, we see that the same pattern is obtained as in the [eː∼ɛi] condition. In addition, the expected pattern of increased diphthong-phoneme responses at later morphing steps is obtained. This shows that the pattern of results in the [eː∼ɛi] condition cannot simply be explained by a putative defect in the auditory stimuli. It appears instead that the FDS already have knowledge in place about the realizational differences in the [ɛi] vowel (and, therefore, the [eː] vowel), and, if the NDS behavior is taken as a baseline, are overcompensating. This knowledge could have been obtained via, for example, Netherlandic-media exposure prior to arrival (although cf. Kuhl, Tsao, & Liu 2003 for infants and Romeo et al. 2018 for adults, who both suggest that media do not play a significant role in grammar acquisition). However, system-internal factors provide a more plausible alternative. It is shown in Van de Velde (1996) that the diachronic diphthongization of /eː/ is a more advanced sound change than that of /oː/. If this has also been picked up in perception by the FDS, then it is probable that they have learned to (over)compensate for the presence of diphthongization in the [eː∼ɛi] condition, but have not done so for the [oː∼ɑu] condition.

On the second hypothesis, that a following coda /l/ would increase participants' diphthong responses, all trees agree. The [eː∼ɛi] tree shows that there are significant differences between a following non-/l/ and /l/: the latter always leads to more diphthong responses. The [oː∼ɑu] tree bears out the same result. In the [ɛ∼ɛi] tree, this effect is present only in the first two morphing steps, and then only markedly so in the NDS. This can be explained easily: in the first two steps there is little diphthongization to react to, and we have already seen that the FDS are stronger compensators than the FDS for diphthongization towards [ɛi].

For the rhotics, the results obtained are very different from those seen thus far, including those obtained in the pilot study. The general high proportion of [ɹ] responses is striking. The group difference that was expected as part of the first hypothesis *is* borne out: the FDS, on average, show a lower glide preference than the NDS do. However, while this difference is statistically significant, it should be interpreted with care: the significant difference is one of 91.1% (FDS, averaged over the subsequent splits) vs. 98.5% (NDS). These large and near-categorical preferences for the glided rhotic, even in the FDS group, are perhaps more telling than the significance of the difference between the groups is. It seems that even the FDS are simply very aware that the NDS realization of the rhotic is indeed supposed to be [ɹ]. A possible explanation for this awareness could lie in the strong sociolinguistic salience of the many different varieties of the Dutch rhotic (Sebregts 2015). An explanation in terms of salience is in line with Auer, Barden, & Grosskopf (1998) and opens up a possibility for future research: by repeating the same experiment with different consonants that are less sociolinguistically salient, this explanation of the present results as being due to sociolinguistic salience can be put to the test.

The [ʀ∼ɹ] condition is also the first and only condition where an effect of

"Session" emerges. It is only in this condition that support is found for the third hypothesis. While the NDS controls do not turn out to be significantly influenced by the three experimental sessions, this *was* found for the FDS. In the first session, they report significantly fewer [ɹ] percepts than in the third condition, where they have caught up to (and even numerically exceed) the NDS participants. In the second session, they report slightly *fewer* [ɹ] percepts. This looks like U-shaped development, but given the very high proportion of [ɹ] percepts across the board and hence the small room for meaningful differences, such an interpretation should be considered with the appropriate caution.

Having established that there are significant differences between the two groups in perception, it remains to be seen whether and to what extent these differences transfer to the participants' production. This is investigated next in Experiment 2.

3.3 Experiment 2: word production

Experiment 2 complements Experiment 1 by investigating the *production* of the tense mid vowels, original diphthongs, and rhotic in the same 2 × 10 participants. This experiment has three goals. The first is to establish the realizations used by the sociolinguistic migrants, particulary with respect to the phonological differences between Netherlandic Dutch and Flemish Dutch. *A priori*, the hypothesis is that the sociolinguistic migrants use non-upgliding tense mid vowels, less-upgliding diphthongs, and unglided rhotics, all irrespective of the phonological context. The Netherlandic controls, on the other hand, are expected to alternate between fully-upgliding realizations and non-upgliding realizations for all six vowels, and trills and glides for the rhotic, depending on the phonological environment. The second goal of the production experiment is to establish whether these differences between the sociolinguistic migrants and the Netherlandic controls remain stable over time, or if the sociolinguistic migrants converge towards the Netherlandic controls norms over the course of the nine months. The latter option is the hypothesis to be verified. To investigate this, the production experiment was performed three times, exactly as the perception experiment was. The third question is whether the sociolinguistic migrants can be nudged to adopt more Netherlandic realizations by priming them (as was done in Pinget, Kager, & Van de Velde 2019 and Walker 2014) with a more Netherlandic-like realization or a more Flemish-like realization of a single phoneme; the hypothesis to be tested is that this is indeed the case. This is not required by, but would corroborate, the proposal that the long-term adoption of sound change is the result of repeated short-term accommodation, discussed in Section 3.1.2.

3.3.1 Method

Participants

The participants were the same as in Experiment 1.

Stimuli

Stimuli consisted of 3 × 3 words containing a point vowel /i,u,aː/ (used in the practice trials), 8 × 20 words containing one of the phonemes /eː,øː,øː,ɛi,œy,ɑu, aːʀ,ɛ/, and 8 × 20 words containing one of the phoneme sequences /eːl,øːl,oːl,ɛil, œyl,ɑu,aːʀ,ɛl/. The third set is equal to the second set plus a coda-/l/ phoneme (words were selected so that the /l/ was always coda), with the exception of */aːʀl/ and */ɑul/ conditions: the former of these is phonotactically illegal, and the latter does not occur in the language due to a lexical gap (save for the proper name "Paul"). In principle, words were chosen such that the phoneme (combination) of interest was word-initial. This could not be achieved for the coda-/l/ conditions and for the /øː/ condition, in which cases this requirement was dropped. Given these constraints, for each cell in the design, the 20 words were chosen on the basis of frequency: the 20 highest-frequency words based on CELEX (Baayen, Piepenbrock, & Gulikers 1995) frequency were chosen.

A question additional to the participants' basic formant values was if they would copy realizations that are particularly characteristic of one of the two varieties of Standard Dutch. To investigate this, each of the stimuli was read aloud in a carrier sentence by a the same speaker who had produced the materials for Experiment 1, who produced each stimulus in two different variants, one with upgliding diphthongization and a trilled rhotic, and one without upgliding diphthongization and with a glided rhotic. I will transcribe the speaker's upgliding-diphthong realizations as [ei,øy,ou,ɛi,œy,ɑu] and her non-upgliding-diphthong realizations as [eː,øː,oː,œː,ɑː]. The latter are the transcriptions that one would obtain by removing the upglide, which is the primary cue for diphthongization in Dutch (Booij 1995, van Oostendorp 2000), while keeping all else equal.[2] For the rhotic, I will use [ʀ] to refer to the non-glided variant and [ɹ] to refer to the glide.

Two experimental conditions were constructed out of the two different variants the speaker had produced for each item. In one of these, (the "A" condition), the phoneme (sequence) of interest was realized by means of a typical Netherlandic-Dutch allophone; these are upgliding realizations [ei,øy,ou, ɛi,œy,ɑu], non-upgliding realizations [eːɫ,øɫ,oːɫ,ɛːɫ,œːɫ], a glided coda /r/, and a short [ɛ]. In the other condition (the "B" condition), the segments were

---

[2] These realizations are independently attested in regional dialects of Dutch, spoken in areas like Maastricht (for [eː,øː,oː]; Gussenhoven & Aarts 1999) and The Hague (for [ɛː,œː,ɑː]; Timmerman 2018); both of these are infamous for their monophthongal realizations of the mentioned vowels.

realized with an incorrect allophone according to Netherlandic-Dutch phonology given the context. For the tense mid vowels and diphthongs, this condition consisted of non-upgliding realizations (correct before coda /l/, but not before a nonapproximant consonant). For the rhotic, the "non-Netherlandic-Dutch allophone" condition consisted of using the trilled [ʀ] as opposed to the glide [ɹ], which is the norm in syllable onset but not in the coda. The "non-Netherlandic-Dutch allophone" condition for the vowel /ɛ/ was a realization as [ɛː], which is an illicit realization of /ɛ/ not only in Netherlandic Dutch but also in all Flemish varieties. In this condition, the difference between the two realizations is not correlated with Netherlandic-Dutch–Flemish-Dutch differences, making this vowel suitable as a filler. A crucial property of the experiment is that only the target phoneme (or phoneme sequence) was realized in a specific way; the remainder of the word was produced naturally. This ensures that participants respond, if they do so, only to the phonological differences being investigated, rather than switching between ND and FD accents wholesale. No dual realizations were presented for the point vowels in the practice session (there are no allophonic differences in the realizations of these vowels between Netherlandic Dutch and Flemish Dutch). Table 3.4 provides a schematic overview of the realizations.

The auditory prime words were paired to the visual target words in two ways. In the [eː,ɛi,aːɹ,eːl,ɛil,aːʀ] conditions, the prime and target words were identical. In the other conditions, the prime words were paired to the target words randomly. This makes it possible to separate putative adaptation by copying of the prime realization from putative adaptation by accent switching. Each word was presented twice: once for each recorded variant. To prevent the experiment from becoming too long, only the /ɛ/ condition was included as an explicit filler. However, each of the eight sets of 20 words under investigation can be considered to be filled not only by the true fillers, but also (overlappingly) by the words from the other seven conditions. This means that every condition is immersed in $2 \times (300 - 20) = 560$ stimuli not related to that condition. The full list of items is available in Appendix B.

Procedure

The experiment took place in the same booth as Experiment 1. The experiment consisted of 618 trials with three breaks in between. Instructions were presented in the same way as in Experiment 1. Participants could initiate the experiment by pressing one of the buttons in the armchair. Figure 3.7 presents a diagram of the structure of each of the experimental trials. A trial started with a black screen, after which the auditory prime word was presented. After presentation of the prime, the visual target word was presented, which participants had to read aloud. Between two trials, a fixation cross was presented for

Table 3.4: Overview of the allophone variants used in the experiment (618 trials). For the point vowels /i,u,aː/, both allophone variants are the same.

	Realization used in prime items			
	Before non-/l/		**Before /l/**	
Phoneme	NDS	Non-NDS	NDS	Non-NDS
/eː/	[ei]	[eː]	[eː]	[ei]
/øː/	[øy]	[øː]	[øː]	[øy]
/oː/	[ou]	[oː]	[oː]	[ou]
/ɛi/	[ɛi]	[ɛː]	[ɛː]	[ɛi]
/œy/	[œy]	[œː]	[œː]	[œy]
/ɑu/	[ɑu]	[ɑː]	[ɑː]	[ɑu]
/aːʀ/	[aːɹ]	[aːʀ]		
/ɛ/	[ɛ]	[ɛː]	[ɛ]	[ɛː]
/i/	[i]	[i]	[i]	[i]
/u/	[u]	[u]	[u]	[u]
/aː/	[aː]	[aː]	[aː]	[aː]

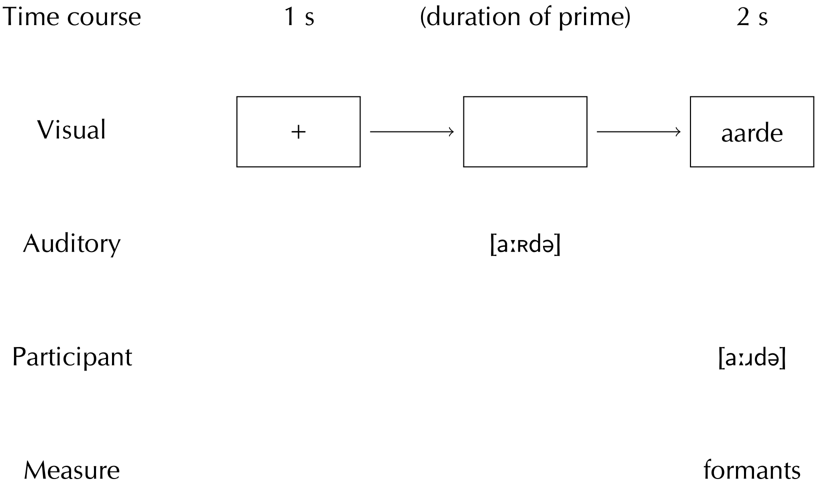

Figure 3.7: Example trial for the production task.

1 s. The experiment started with 2 × 9 practice trials, in which the /i,u,aː/ words were used as target words. After the practice block, participants were able to initiate the 600 remaining trials of the actual experiment by pressing one of the two buttons.

Data analysis

The acquired single-word speech recordings for the experimental items were forcedly aligned[3] to their CELEX reference transcriptions using HTK (Young et al. 2002). Using Praat (Boersma & Weenink 2016), samples of F1 and F2 at both 25% and 75% realization were extracted for the vowels, as were samples of F3 at 50% realization for the rhotic (all using the Burg algorithm, time step 10 ms, 5 formants, cut-off point 5,000 Hz for men and 5,500 Hz for women, window length 25 ms, pre-emphasis from 50 Hz). Outliers (the result of formant-tracking errors or of incorrect forced alignment) were identified using the procedure by van der Harst (2011:82) and were removed from the data. Before outlier removal, the total dataset (including practice trials) consisted of 62,281 observations; after outlier removal, 61,058 observations (98.0%) remained. Figure 3.8 and Table 3.5 provide overviews of the resulting data. In the interest of space, the vowel-space plot collapses the data of the three sessions of the NDS controls into one, and omits the effect of NDS vs. non-NDS prime realizations. Naturally, the analyses (reported in Tables 3.6 and 3.7) are based on the original, uncollapsed, data.

While outlying observations were identified based on their quantiles after Lobanov normalization—as is needed for the procedure by van der Harst—the subsequent data modeling, which is described next, operated on the unnormalized data. Accounting for systematic between-participants differences was left to the random-effect structure of the statistical models. For the vowels, a diphthongization score "ΔF1" was created by subtracting the F1 at 25% realization from the measure at 75% realization, resulting in 24,506 data points. For the rhotic, the sample at the rhotic's midpoint was used directly, resulting in 1,984 data points. These scores were used as dependent variables in two linear mixed-effects models, one for all vowels which modeled the ΔF1 and one for the rhotic which modeled the F3. For the vowel model, fixed effects were included for "Vowel" (sum-coded), "Following consonant" (treatment-coded

---

[3] It is well-known that vowel-approximant transitions are hard to segment consistently by human listeners. The present chapter follows authors like Walker (2014) in trusting the forced aligner's placement of the boundaries in these cases, which will at least be always consistent. In addition, by taking conservative measurement points (25% and 75% for the vowels, based on findings by van der Harst 2011 that these timepoints are reliably separated from the surrounding coarticulation, and 50% for the rhotic), some margin for error in the forced aligner's performance is allowed for. However, I recognize that a true solution to the problem of segmenting vowel-approximant transitions requires a much more sophisticated formant-tracking approach than falls within the scope of the present chapter; I refer to Chapter 2 for a possible approach for future work.

with /l/ = 0 and non-/l/ = 1), "Group" (treatment-coded with FDS = 0 and NDS = 1), "Session" (coded using orthogonal polynomials), and "Prime realization" (treatment-coded with NDS-like = 0 and non-NDS-like = 1). All interactions were considered, as were all legitimate random intercepts and slopes by participants, by target words, and by prime words. The random-effects covariance matrix was constrained to be diagonal. Function `buildmer` from the eponymous R package (Voeten 2019b) was used to identify the maximal model that would still converge without singularities and perform backward stepwise elimination from this maximal model to arrive at a suitably parsimonious final model (following the argument by Matuschek et al. 2017 based in statistical power). For both of these stages of term selection, the BIC (Schwarz 1978) was chosen as the measure of term importance. Degrees of freedom for the final models were calculated using the Kenward-Roger approximation (Kenward & Roger 1997) via R package `lmerTest` (Kuznetsova, Brockhoff, & Christensen 2017) using the implementation in R package `pbkrtest` (Halekoh & Højsgaard 2014).

In light of the possibly large individual differences in speech-shadowing tasks, an anonymous reviewer asks whether any group-level effects of the prime manipulation might be eclipsed by major individual differences. This was explicitly looked into by investigating the by-participants random effects from supplementary mixed-effects models in which the factor "Prime realization" was explicitly included. That is: starting from the final models obtained via the stepwise procedure, fixed-effect terms and by-participants random slopes were added for "Prime realization" plus its interaction with all other fixed-effect terms and by-participants random slopes in the model. This was done regardless of whether any of these "Prime realization" terms had been selected for inclusion into the model by stepwise elimination in the first place. Note that, by sidestepping the stepwise procedure in this way, the following caveat applies: if the variance components of any of these random-effect terms terms truly are zero (or are shrunk to zero, which happens when the explained variance is smaller than the penalty term), their inclusion into the model will—by definition—cause the model to converge to a singular fit. While this would normally be an indication that the model is overfitted and should be reduced (Bates et al. 2015a, Matuschek et al. 2017), for this specific inquiry singular fits must be permitted. Hence, convergence was checked only based on the gradient and the Hessian of the maximized REML criterion, ignoring `lme4`'s singular-fit check. These checks indicated that the two supplementary models had converged without incident. From these models, the by-participants estimated random slopes pertaining to the factor "Prime realization" and its interactions were extracted. Next, the same procedure as in Chapter 4 was applied: a cluster analysis was run on each random-effect term, using function `Mclust` from R package `mclust` based on a one-dimensional variable-variance model. As demonstrated in Chapter 4, this provides an empirical test of the extent of any individual differences: if participants are not significantly different from one another with respect to these factors, the cluster analysis will identify a single cluster, whereas if there is a statistically significant pattern in the between-participant variation, multiple clusters will be identified.

3.3.2 Results

The results from the statistical analyses show that both groups of participants produce significantly more upgliding diphthongization before a following non-approximant consonant ($\hat{\beta} = -71.03$, SE = 8.88, $t_{23.59} = -8.00$, $p < .001$). The various main effects for the factor "Vowel" show that the different vowels have slightly different targets. The interactions of "Following segment = non-/l/ × Vowel" show that the different vowels also have different ranges available for upgliding diphthongization, such that the tense mid vowels /eː,øː,oː/ diphthongize less strongly than the average vowel, and the vowel /œy/ diphthongizes more strongly than the average vowel.

The crucial effect for the primary hypothesis is the significant interaction "Following segment = non-/l/ × NDS" ($\hat{\beta} = -93.21$, SE = 12.10, $t_{20.78} = -7.70$, $p < .001$). This shows that, for the average vowel, the NDS diphthongized more than the FDS did when the vowel was followed by a different consonant than coda /l/. This effect was across the board, insofar that there was a significant two-way interaction of "Following segment × Group", but no significant three-way interaction of these two factors with "Vowel". In point of fact, the three-way interaction "Following segment × Vowel × Group" was selected out of the model (ΔBIC = 17.07, which is larger than zero and hence a worse score than a model not including this interaction).

Table 3.5: Averages of the raw F3 data in Hz (20 participants, 68 items). Note how the NDS consistently have lower F3s than the FDS, and that the FDS do not appear to be moving closer to the NDS over the three sessions.

		Session		
Group	**Prime realization**	1	2	3
FDS	ND allophone	2,944	2,963	2,993
FDS	non-ND allophone	2,971	2,982	2,972
NDS	ND allophone	2,336	2,350	2,227
NDS	non-ND allophone	2,367	2,352	2,205

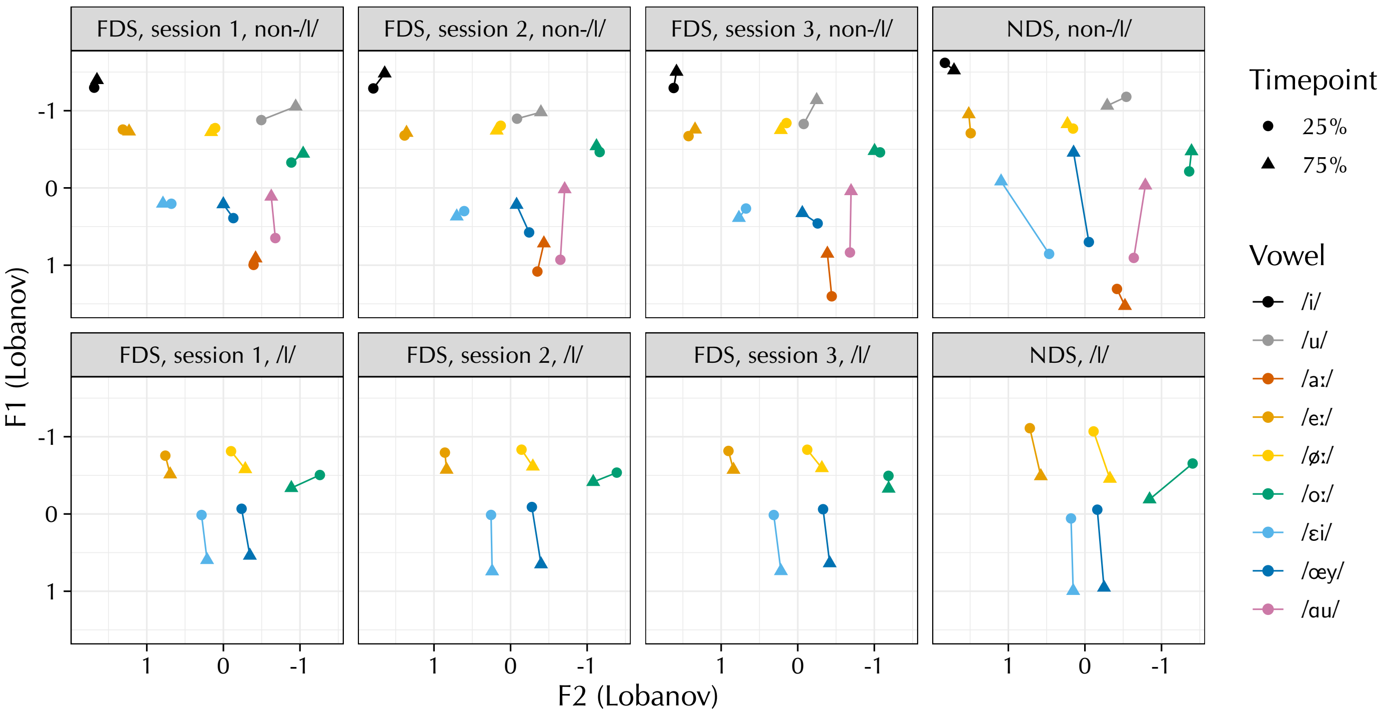

Figure 3.8: Vowel-space plot of the raw F1/F2 data (20 participants, 550 items). For reasons of space, the data for the NDS controls have been collapsed over the three sessions, as has the effect of prime realization. Observe how the NDS have upgliding realizations of the vowels under investigation in the non-/l/ condition, and have non-upgliding (downgliding) realizations in the coda-/l/ condition. The FDS, by contrast, have non-upgliding realizations in both conditions, across all sessions; in the /l/ condition, they exhibit the same downglide as the NDS.

Table 3.6: Results for the F1 analysis (20 participants, 550 items). Observe the significant effect for "Following segment = non-/l/", which shows that the NDS produce more upgliding diphthongization in non-coda-/l/ environments than the FDS do. There are significant per-vowel adjustments to this effect, but they do not obviate this main result. There are differences across the three sessions of the experiment, but they, too, are not of sufficient magnitude to make a meaningful contribution to the bigger picture.

Factor	**Estimate (SE)**	***t***	**df**	***p***	**Sig.**
Intercept	8.04 (5.84)	1.38	25.08	.18	
Following segment = non-/l/	−71.03 (8.88)	−8.00	23.59	<.001	***
Vowel = /eː/	21.00 (5.85)	3.59	111.35	<.001	***
Vowel = /øː/	22.90 (5.87)	3.90	109.64	<.001	***
Vowel = /oː/	−6.69 (6.18)	−1.08	90.06	.28	
Vowel = /œy/	34.87 (6.27)	5.56	85.99	<.001	***
Vowel = /ɑu/	−96.40 (12.22)	−7.89	23.17	<.001	***
Group = NDS	−5.86 (7.99)	−0.73	22.09	.47	
Session (Linear)	−2.84 (1.76)	−1.62	19.06	.12	
Session (Quadratic)	1.78 (1.74)	1.02	20.31	.32	
Following segment = non-/l/ × /eː/	39.18 (8.13)	4.82	113.73	<.001	***
Following segment = non-/l/ × /øː/	59.41 (8.30)	7.16	105.04	<.001	***
Following segment = non-/l/ × /oː/	46.71 (9.17)	5.09	74.07	<.001	***
Following segment = non-/l/ × /œy/	−95.99 (8.59)	−11.18	92.35	<.001	***
Following segment = non-/l/ × NDS	−93.21 (12.10)	−7.70	20.78	<.001	***
Vowel = /eː/ × Session (Linear)	−1.59 (1.97)	−0.81	18,239.26	.42	
Vowel = /eː/ × Session (Quadratic)	−2.34 (1.96)	−1.19	23,420.98	.23	
Vowel = /øː/ × Session (Linear)	8.17 (1.97)	4.15	18,566.60	<.001	***
Vowel = /øː/ × Session (Quadratic)	−5.89 (1.96)	−3.00	23,448.58	<.01	**
Vowel = /oː/ × Session (Linear)	−1.10 (1.98)	−0.55	20,059.96	.58	
Vowel = /oː/ × Session (Quadratic)	−3.37 (1.96)	−1.71	23,591.59	.09	
Vowel = /œy/ × Session (Linear)	3.95 (1.99)	1.99	20,285.82	.047	*
Vowel = /œy/ × Session (Quadratic)	1.63 (1.97)	0.82	23,611.86	.41	
Vowel = /ɑu/ × Session (Linear)	−14.00 (2.15)	−6.51	23,856.46	<.001	***
Vowel = /ɑu/ × Session (Quadratic)	14.36 (2.14)	6.71	23,859.16	<.001	***
Following segment = non-/l/ × Session (Linear)	1.42 (5.54)	0.26	18.78	.80	
Following segment = non-/l/ × Session (Quadratic)	9.84 (4.43)	2.22	19.25	.04	*

Table 3.7: Results for the F3 analysis (20 participants, 68 items). Observe that the NDS have a significantly lower F3 than the FDS do, and that there is no significant evidence that this gap narrows over the three sessions of the experiment.

Factor	Estimate (SE)	*t*	df	*p*	Sig.
Intercept	2,952.15 (80.15)	36.83	21.43	<.001	***
Group = NDS	−634.04 (111.91)	−5.67	20.45	<.001	***
Session (Linear)	14.94 (39.89)	0.37	23.46	.71	
Session (Quadratic)	80.00 (36.10)	0.22	28.26	.83	
Prime realization = non-NDS	8.78 (20.95)	0.42	1,893.13	.68	
Group = NDS × Session (Linear)	−77.70 (53.72)	−1.45	23.15	.16	
Group = NDS × Session (Quadratic)	−70.11 (48.89)	−1.43	27.76	.16	
Group = NDS × non-NDS	0.55 (28.55)	0.02	1,891.52	.98	
Session (Linear) × non-NDS	−30.06 (35.43)	−0.85	1,889.10	.40	
Session (Quadratic) × non-NDS	−4.97 (37.13)	−0.13	1,897.87	.89	
Group = NDS × Session (Linear) × non-NDS	−7.76 (48.47)	−0.16	1,889.12	.87	
Group = NDS × Session (Quadratic) × non-NDS	0.97 (50.44)	0.02	1,895.08	.98	

The second hypothesis, that the FDS would become more Netherlandic-like over the three sessions, bears on the factor "Session". Significant effects of this predictor are found, but the effect sizes are very small (all <15 Hz, which is below the JND for an F1; Kewley-Port 1995) and, more importantly, the effects are neither specific to the groups, nor to the following segments. The crucial interaction, "Following segment × Group × Session", was selected out of the model (ΔBIC = 14.28). The corresponding ΔBIC can be converted into a Bayes factor using the formula given in Wagenmakers (2007). Per Kruschke & Liddell (2018), this makes it possible to say if the lack of an effect is due to a lack of statistical power, or whether there is sufficient evidence in the data to say that the effect is truly absent. The Bayes factor shows that the data are 1,264.40 times as likely under the null model than they are under the alternative model, which is "decisive" (Jeffreys 1961) evidence that the differences in diphthongization between these groups do not decrease over time.

The third hypothesis was that a more vs. less Netherlandic-Dutch realization of the prime words could induce the Flemish-Dutch participants to similarly modify their own production during that trial. This was not found, and again the interaction of interest, "Group × Prime realization" was selected out of the model (ΔBIC = 8.99, BF$_{01}$ = 89.66, "very strong" evidence for the null model). In addition, there was no evidence for significant inter-individual differences in shadowing patterns: the cluster analysis of the by-participants random effects in the supplementary models did not find evidence for more than a single cluster ($\Delta\text{BIC} = \langle 0.74, 5.78, 10.52 \rangle$, $\text{BF}_{01} = \langle 1.44, 17.96, 192.75 \rangle$, "anecdotal"/"strong"/"decisive" evidence for the one-cluster models, per Jeffreys 1961). For reference, these random effects are plotted in Figure 3.9, in order.

The results for the F3 data show that the NDS have a significantly more glided rhotic than the FDS ($\hat{\beta} = -634.04$, SE = 111.91, $t_{20.45} = -5.67$, $p < .001$). This did not change significantly over the three sessions. The interaction "Group × Session" is not significant; computing a Bayes factor as before gives "substantial" evidence for the null model ($\Delta\text{BIC} = -4.29$, $\text{BF}_{01} = 8.56$), indicating that the lack of significance is not due to lack of power. The interaction "Group × Prime realization" is extremely small (0.55 Hz, well below the JND for an F3; Allen, Kraus, & Bradlow 2000) and nonsignificant ($p = 0.98$). Because of its small size, the differences in model fit between a model with and without this predictor come out small as well, and a Bayes factor only provides "anecdotal" evidence ($\Delta\text{BIC} = -0.95$, $\text{BF}_{01} = 1.61$). Per Kruschke & Liddell (2018), this implies that the model has insufficient power to clearly disambiguate between presence and absence of an effect this small in size. However, the real-world relevance of such an effect is negligible. Nonetheless, it is prudent to explore if the small group average for this effect (or any other interactions of the factor "Prime realization") might be due not to a global absence of such an effect, but rather due to large individual differences that perhaps cancel out on average. As for the F1 data, the supplementary models investigating this possibility did not provide significant evidence that some participants were more extreme shadowers than others: no inter-individual clusters could be found in the shadowing-related by-participants random slopes ($\Delta\text{BIC} = \langle 5.43, 12.80, - \rangle$, $\text{BF}_{01} = \langle 15.13, 602.11, - \rangle$, "substantial"/"decisive" evidence for the one-cluster models, per Jeffreys 1961; for the final model in Figure 3.9, `Mclust` could only compute the one-cluster model). These are also included in Figure 3.9.

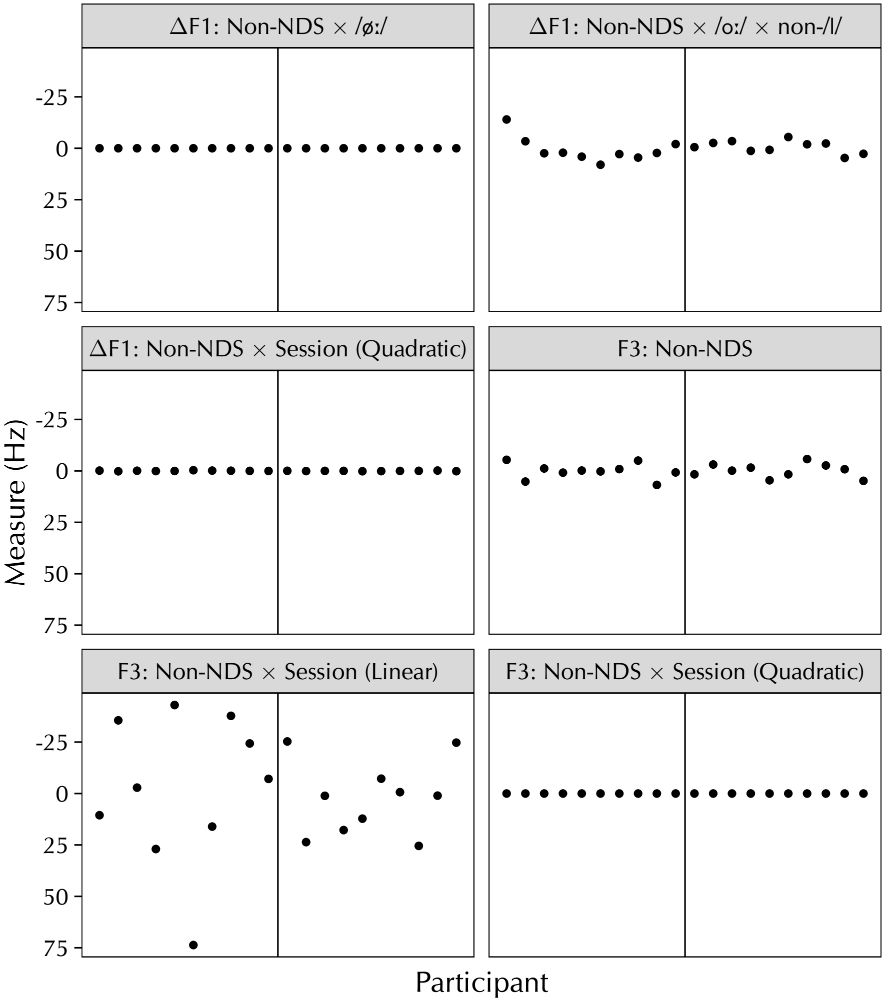

Figure 3.9: The estimated by-participants random effects (also known as "BLUPs", for "best linear unbiased predictors") for the random slopes involving the factor "Prime realization" in the supplementary models. Random-effect vectors that were estimated with zero variance (i.e. all by-participants random-effect coefficients are zero) have been omitted. Each panel represents a single random slope from either the ΔF1 or the F3 model; each dot is a participant. Each panel has been separated into two panes: the left pane corresponds to the ten Flemish participants and the right pane corresponds to the ten Netherlandic participants. Cluster analyses revealed no clusters in any of these six random slopes.

3.3.3 Discussion

For the production part, the main research question—do sociolinguistic migrants adopt the novel variants, and after how much time?—was again broken down into three hypotheses. The first hypothesis was that there would be significant differences between the NDS and the FDS, as per the studies described in Section 3.1.1. The second was that the FDS' productions would become more aligned with those of the NDS during the nine months for which the investigation ran. The final hypothesis was that the FDS could be nudged into a more NDS-like or less NDS-like realization by providing them with a more NDS-like or less NDS-like prime realization.

For the vowels, the NDS turned out to produce significantly more upgliding diphthongization than the FDS, but only when the following consonant is not /l/. In the coda-/l/ condition, Figure 3.8 suggests that the vowels do not have an upglide, but rather show a downgliding realization. An anonymous reviewer offers an interpretation of this effect in terms of centralization, which corresponds to the behavior of these vowels before coda /r/ (Gussenhoven 1993). Under this interpretation of the observed behavior before coda /l/, the tense mid vowels and diphthongs behave uniformly when followed by a liquid in the coda (a proper subset of the environment identified in Gussenhoven 1993, which was the foot). As this denotes a phonological natural class, there may be a more general phonological rule subserving these allophonic patterns, which may inspire future research.

The results obtained for the vowels bear on the first goal of Experiment 2, which was to establish the realizations used by the two groups of participants and differences in phonological knowledge between the groups. The results confirm that the NDS produce more upgliding diphthongization than the FDS do, but only in the non-/l/ context. This difference shows that the NDS employ phonological knowledge that the FDS do not: the NDS implement an allophonic distinction between vowels followed by /l/ and vowels followed by a nonapproximant consonant. The first hypothesis of Experiment 2 is thus confirmed. In the /l/ condition, the FDS and NDS were not found to be significantly different. However, in the non-/l/ condition, the NDS indeed produce a more diphthongal allophone, while the FDS do not. This difference between the NDS and FDS was across the board, and was found not to depend on the specific vowels investigated. For the rhotic, the NDS turned out to produce significantly more gliding than the FDS. These results confirm the NDS–FDS differences in realizations, and show that the NDS have the hypothesized phonological restrictions in their grammars, while the FDS do not.

The second hypothesis, that the FDS would become more NDS-like over the three sessions, must be rejected. For both the vowels and the rhotic, differences between the groups as a function of session were not only not found, but were also more likely to be absent than present. The third hypothesis was that a more FDS-like or more NDS-like prime realization of the critical segments would prime participants to adopt such a more FDS/NDS-like realization for themselves. This hypothesis must also be rejected, as the evidence points against an interaction "Group × Prime realization", or, for that matter, an effect of "Prime realization" at all. Shadowing behavior of more NDS-like or less NDS-like realizations could not be observed in either the FDS or the NDS group, and individual differences could not explain this.

3.4 General discussion and conclusion

The present paper set out to investigate the process by which individual speakers and listeners adopt on-going community sound change. This empirical study of the adoption of sound change was made possible by the fortuitous intersection of diachronic sound change currently on-going in Netherlandic Dutch and well-established synchronic variation between Netherlandic Dutch and Flemish Dutch. This made it possible to investigate individuals' adoption of the on-going diachronic changes in laboratory-controlled circumstances (and with the concomitant methodological precision) by framing the adoption of sound change as a second-dialect-acquisition problem. The main question for this study was: do individuals adapt their production and perception to on-going sound change in real time, and if so, in *how much* time?

The two experiments performed in this study show robust and persistent differences between the FDS and NDS groups. Concerning the vowels, there are significant differences between the NDS and the FDS in production and perception, which are in full agreement with the results from earlier work on production by Gussenhoven (1999), Van de Velde (1996), Verhoeven (2005), and Chapter 2. The same is true for the rhotic; the differences between the two groups of participants are in line with the findings by Sebregts (2015). In perception, the rhotic was also the only segment for which between-groups effects of the experimental session were found, but the differences were very small and did not carry over to production. These perceptual differences were U-shaped, in that participants indicated a slightly increased preference for the non-NDS [ʀ] realization in session 2, which reversed into a stronger [ɹ] preference in session 3. Although the small magnitude of both effects makes their real-world relevance debatable, the observed differences might be indicative of the first steps of long-term accommodation. If the adoption curve of the FDS is indeed U-shaped, this would be in line with other research on the acquisition of novel grammatical structure, such as in infants (Becker & Tessier 2011) and in second-language learners (Trofimovich et al. 2012).

On the basis of this study, the answer to the main research question could be one of two options. The first is that individuals do not adapt their production and perception to on-going sound change in a way that can be detected by the experiments employed in the present study. The second possibility is that individuals *do* adopt the on-going changes, but that nine months is not enough time for this process to take place. The latter is the correct conclusion, as another study using the same changes, setup, and tasks *did* find adoption of these changes in Flemish sociolinguistic migrants who had spent much more time in the Netherlands (years–decades; Chapter 4). It was already known from many prior studies that second-dialect acquisition and the adoption of sound change take time (Alshangiti & Evans 2011, Bauer 1985, Carter 2007, Cedergren 1987, Chambers 1992, De Decker 2006, Evans & Iverson 2007, Harrington 2006, Harrington, Palethorpe, & Watson 2000, Hinton 2015, Nahkola & Saanilahti 2004, Nycz 2011, Nycz 2013, van Oostendorp 2008, Prince 1987, Sankoff 2004, Sankoff & Blondeau 2007, Sankoff, Blondeau, & Charity 2001, Trudgill 1988, Wagner 2008, Yaeger-Dror 1994, Ziliak 2012); the results from the present study contribute towards establishing a lower bound for this timeframe: more than nine months.

Siegel (2010) noted that the outcomes of second-dialect acquisition are highly variable, with the average "success probability" coming out at chance level. Siegel suggested that a reason for some individuals being less successful is that abstract features are harder to learn than surface features. It is possible that this factor contributed to the FDS' non-adoption of the sound changes in the present study. The sound changes reported here are all allophone splits, of which the sociolinguistic migrants have one category already available (non-upgliding vowel, trilled rhotic) but not the other. Hence, the sound changes involve an abstract change in the system of linguistic categories, similar to that reported in Sneller (2018). In fact, the change in Sneller (2018) worked in the reverse direction—a complex phonological rule diachronically changing into a simpler one—and was adopted rapidly, suggesting that the individuals are more likely to adopt systems that are simpler than the one they have, and then indirectly are less likely to adopt new allophone systems that are more complex, as is the case for these Dutch changes. Furthermore, it should be noted that there is no real pressure on the FDS to adopt these changes: the variables are allophonic, not phonemic, and therefore do not impact the FDS's ability to function in everyday life in any way.

The rhyme-decision data for at least the [eː∼ɛi] condition suggested that, nonetheless, the participants have *begun* to alter their perception (and eventually finish this process; Chapter 4). We know that such perceptual adjustments do not directly carry over to production (Pardo 2012), as was also the case for the explicit perceptual priming used in the production task. Corroborating Ziliak (2012) and Evans & Iverson (2007), the present study found that perception was more malleable than production. Walker (2014) had found an exceptionally direct relationship, in that explicit priming in perception similar to the present chapter's was able to nudge a participant to shift their own productions; the present study did not find such direct priming effects. Finally, it has been observed that the adoption of change starts in perception, but ends with production overtaking perception (Pinget 2015, Pinget, Kager, & Van de Velde 2019). The present study's finding of small changes starting in perception could be an incipient instance of the same process.

The findings from the present study have implications for sociolinguistic methodology. The robust differences that were found between the groups show that the rhyme-decision task and the word-list-reading task used in the present study are highly suitable for use in sociolinguistic research. In addition, the results speak positively to the general use of laboratory-phonological methods in sociolinguistic research. The perception results for the rhotic and the [eː∼ɛi] contrast also highlight the importance of knowledge of phonological variables and sociolinguistic salience, which were shown to be probed by the rhyme-decision task. Finally, the results suggest that short-term accommodation and long-term accommodation are separate processes. While the present study was not set up directly to confirm or refute models such as the change-by-accommodation model (Auer & Hinskens 2005, Chambers 1992, Trudgill 1986), the finding that nine months is not long enough, while more time is (Chapter 4), is somewhat problematic for these theories, since phonetic accommodation is known to happen much more rapidly than in nine months (Maye, Aslin, & Tanenhaus 2008, Norris, McQueen, & Cutler 2003, Pardo et al. 2012). It is possible that there are other factors at play which mediate the adoption of community variation by individuals, such as prestige (Labov 2001), salience (Auer, Barden, & Grosskopf 1998), social-network size (Lev-Ari 2018), or heritage (Wagner 2008). The present study's setting of laboratory sociolinguistics is not appropriate to investigate these factors: future research in the form of an ethnographic sociolinguistic study is necessary instead.

Acknowledgments

I thank my dissertation supervisors, Claartje Levelt and Yiya Chen, for their feedback and support from the conception of the experiments to the final manuscript. I am also due thanks to Aliza Glasbergen-Plas, for her willingness to be the speaker for both experiments reported in this chapter. The chapter has benefited greatly from the judicious comments by two anonymous reviewers and their helpful suggestions. Finally, I thank the twenty university students who were willing to take part in the experiments reported here.

This work is part of the research programme *Watching Dutch Change* with project number PGW-15-15, which is (partly) financed by the Dutch Research Council (NWO).

CHAPTER 4

Individual differences in the adoption of sound change

This chapter has been accepted for publication as: Voeten, C. C. (to appear). Individual differences in the adoption of sound change. Language and Speech. *https://doi.org/*10.1177/0023830920959753.

Abstract

It is still unclear whether an individual's adoption of on-going sound change starts in production or in perception, and what the time course of the adoption of sound change is in adult speakers. These issues are investigated by means of a large-scale (106 participants) laboratory study of an on-going vowel shift in Dutch. The shift involves the tense mid vowels /eː,øː,oː/, which are changing into phonologically-conditioned upgliding diphthongs, and the original diphthongs /ɛi,œy,ɔu/, whose nuclei are lowering. These changes are regionally stratified: they have all but completed in the Netherlands, but have not affected the variety of Dutch spoken in neighboring Belgium. The study compares production (word-list reading) and perception (rhyme-decision) data from control groups from each country to those of eighteen "sociolinguistic migrants": Belgian individuals who moved to the Netherlands years ago. Data are analyzed using mixed-effects models, considering not just the group level, but also individual differences. Production results show that at the group level, the migrant group is in between the two control groups, but at the individual level it becomes apparent that some migrants have adopted the Netherlandic norms, but others have not. Perception results are similar to the production results at the group level. Individual-level results do not provide a clear picture for the perception data, but the individual differences in perception correlate with those in production. The results agree with and extend previous findings on the role of individual differences in the individual adoption and eventual community propagation of on-going sound change.

4.1 Introduction

4.1.1 The adoption of sound change

It has been suggested that sound change originates when there is a mismatch (or "coordination failure"; Bermúdez-Otero 2007) between a speaker and a listener. Either they have the same grammar, but one of the pair over- or under-applies rules compensating for intrinsic variation (Ohala 1981), or the speaker and the listener have acquired subtly different grammars in childhood and assign different cue weights to the same auditory information in the phonetic signal (Beddor 2009, Hamann 2009). When the originating individual begins to reproduce their individual grammatical innovation, and hence begins to transmit the sound change to other individuals through the medium of speech production, the sound change is considered to have been actuated. After that point, it will either spread to other members of the community or peter out. The prerequisites for a sound change to originate, actuate, and spread are thus exceedingly rare: one needs a specific type of variation which is conducive to coordination failure at the right place at the right time, one or more specific individuals to initiate a sound change based on this variation, and a specific reception by the speech community (namely one in which the change is copied and again transmitted further). The rarity of this specific combination of individual and community characteristics has been considered both the reason why sound change takes place at all (if such eventualities were commonplace, we would have learned to be robust against them), and is rare to actuate in the first place (Stevens & Harrington 2014).

Following Pinget (2015), we may assume that sound change spreads through a continuous chain of actuations by individual speaker–listener interactions. The present chapter focuses on the individuals who form the links of this chain: how do they adopt an actuated sound change from their interlocutor into their own grammars? This question is positioned squarely in between the issues of actuation and community spread. It is related to Weinreich, Labov, & Herzog's (1968) transmission problem, but in an individualized form; the transmission problem by Weinreich, Labov, & Herzog's (1968) concerns the dissemination of sound change throughout the community (or its grammar) as a whole. The idea that sound change needs to be initiated by individuals follows from classic models of origin and actuation by authors such as Ohala (1981) and Hyman (1976) (who have since also refined their positions to incorporate representations of the phonetic implementation; Hyman 2013, Ohala 2012), and is also fully compatible with the model by Beddor (2009). It is *almost* compatible with the model by Hamann (2009); her model differs from that by Beddor (2009) in the aspect that, like Labov (2007), Hamann posits that the crucial phonological reanalysis can only be made by children, and that adults can only perform superficial phonetic reanalysis. These models can be considered the theoretical-phonological backdrop of this chapter.

4.1.2 Perception, production, and the individual

Recent work on sound change has recognized that the adoption of sound change by individual speakers and listeners relies on the link between their perception and their production. Consider, for instance, Baker, Archangeli, & Mielke (2011), who study American English [s]-retraction in words like "street". Their results show that speaker-specific coarticulatory variation between the [s] and the following [ɹ] leads to inter-speaker variation in the degree of [s]-retraction. Speakers who strongly coarticulate the [s] with the following [ɹ] produce an [s] that is realized similarly to [ʃ]. For speakers who coarticulate only weakly, that sound is considered a distinct articulatory target. When the latter speakers are paired as listeners with strong [s]-retractors, the weak retractors have an opportunity to actuate a sound change, if their percept of [s] as [ʃ] is reanalyzed to /ʃ/ (or to an equivalent phonological rule) and they begin to use that system in their own productions. Beddor et al. (2018) found that the link between production and perception extends also to the *time course* according to which listeners make use of phonetic cues. Their results show that participants' production of coarticulatory nasalization was predictive of the time course of their perception of the same information. As the authors note, this suggests that differences in perception grammars need not be restricted to cue weightings *per se* (cf. Beddor 2009, Hamann 2009), but can also lie in which cues are utilized when. The results by Beddor et al. (2018) also show that this perception–production link remains stable during on-going sound change, i.e. participants who are more advanced along a sound change in perception are also more advanced in production, causing them to spread the sound change further.

There is evidence that the roles of perception and production reverse depending on the degree to which a sound change has progressed at the community level. Pinget, Kager, & Van de Velde (2019) (also in Pinget 2015) show that, for the on-going merger of Dutch /v,z,ɣ/ into /f,s,x/ and the incipient merger of Dutch /b/ into /p/, adoption by individuals starts in perception: one needs to perceive a change before one will produce it. However, as the change progresses, this relationship slowly comes to reverse: an individual who does not produce a contrast anymore may still be able to draw on subtle differences in perception (although cf. Labov, Yaeger, & Steiner's 1972 "near-mergers"). Coetzee et al. (2018) make a similar observation concerning tonogenesis in Afrikaans: while they generally find that speakers' use of VOT vs. F0 in the production of phonologically voiced vs. voiceless plosives correlates with their use in perception, four of their participants did not produce a reliable VOT contrast, but did rely on such cues in perception. Importantly, the reverse was not found, corroborating Pinget, Kager, & Van de Velde (2019) that, for incipient changes, perception precedes production.

Besides formal linguistic variation such as differences in production and perception grammars, there is evidence that variation at the individual level plays a role in the extent to which sound change is actuated and propagated. Studies of this aspect of sound change largely reinforce the stereotype that leaders of change are young, educated women with certain personality attributes (Haeri 1991, Labov 2001, Milroy 1993, Yu 2010, 2013) and large social networks (Denis 2011, Lev-Ari 2018). These characteristics overlap with the individuals who fit Marslen-Wilson's (1973) description of "close shadowers" in speech-shadowing tasks, which suggests that these personality factors are not directly responsible, but rather indirectly affect socio-cognitive processing (see Yu 2013), and that the latter is what causes these individuals to be leaders of sound change as well.

The role of individual-level factors, and hence of all of the factors discussed in this section, may also depend on whether a sound change is system-internal or contact-driven.[1] The theories by Beddor (2009), Hamann (2009), and Ohala (1981) are mainly concerned with system-internal changes, such as those in Coetzee et al. (2018) or Pinget (2015). Following Hamann (2009) and Labov (2007), these changes spread via L1 acquisition. In contrast, contact-driven changes are spread via contact between adult speakers and listeners. The most obvious factor affecting adults' adoption of contact-driven changes is the amount of time for which they have been exposed to the sound changes. Generally speaking, the shorter the timespan, the more heterogeneous individuals are in adopting an ambient phonetic change. This is illustrated by Alshangiti & Evans (2011), Bauer (1985), Carter (2007), Cedergren (1987), Chambers (1992), De Decker (2006), Evans & Iverson (2007), Harrington (2006, 2007), Harrington, Palethorpe, & Watson (2000, 2005), Hinton (2015), Nahkola & Saanilahti (2004), Nycz (2011), Nycz (2013), van Oostendorp (2008), Prince (1987), Sankoff (2004), Sankoff & Blondeau (2007), Sankoff, Blondeau, & Charity (2001), Trudgill (1988), Wagner (2008), Yaeger-Dror (1994), and Ziliak (2012), who all found small adoption effects in small minorities of studied individuals. When such changes become fully stable within a community is not precisely known, although it has been shown that fifteen

---

[1] I thank an anonymous reviewer for pointing out this possible dependency.

years can be enough (Gordon & Maclagan 2001, Trudgill 1999, Yaeger-Dror 1994). The speed with which a change is spread likely depends on its salience (Auer, Barden, & Grosskopf 1998; see e.g. Rácz 2013 for what this could mean) to the listener: individuals who are more attentive to a change in perception are more likely to adopt the change in production, particularly if the change involves social indexation (Beddor 2015).

4.1.3 Phonological change vs. phonetic change

The degree to which a sound change, particularly one that is contact-induced, can be adopted by individuals depends on the type of the sound change. According to Labov (2007) and Hamann (2009), phonological reanalyses are restricted to language-learning infants, with adult speakers only being able to enact superficial phonetic changes. In the case of phonological change, the sound change involves the phonological grammar, either by adding or deleting a phonological rule, or by adding, removing, or substituting a phonemic category. A clear example of such a case is offered by Harrington, Kleber, & Reubold (2008), who studied /uː/-fronting in Standard Southern British English (henceforth "SSBE"). In an apparent-time study of younger and older SSBE participants, they found that the younger group had a more fronted /uː/ category both in production (i.e. [ʉː]) and in perception (i.e. /ʉː/). Since Harrington, Kleber, & Reubold (2008) show that the perception change preceded the production change, they conclude that SSBE /uː/-fronting started with a phonological change in the underlying form (/uː/>/ʉː/). In a different study, Kleber, Harrington, & Reubold (2012) show that the same account holds for SSBE /ʊ/-fronting, except that this change is in an earlier stage of completion.

In contrast to phonological changes, some changes within the grammars of individuals do not involve a representational change but merely change the phonetic implementation of a particular segment. Such changes are reported by, for instance, Evans & Iverson (2007), who followed nineteen British-English high-school students who were about to enter university. After two years, these students' vowel systems had become more aligned with Standard Southern British English in production, but no reliable change was found in perception. The lack of reliable findings in perception suggests that only the phonetic implementation has changed, and not the phonological representation. It also shows that, for these changes, production changed before perception did, in contrast to the aforementioned results. A possible related observation is that, as pointed out by an anonymous reviewer, one of the changes studied by Evans & Iverson (2007) would be a phonological change instead of a phonetic change. Specifically, Evans & Iverson's (2007) Northern-English students would need to split their *cud–could* vowels: in the North, both of these words have /ʊ/, but in SSBE, *cud* takes the vowel /ʌ/ instead, which does not exist in the Northern phoneme inventory. The successful adoption of this difference would constitute a phonological change, namely a phoneme split. However, Evans & Iverson (2007) did not find direct evidence that their participants managed to do this in either production or perception: while some individuals changed their production targets for these vowels more into the SSBE direction, and their perception correlated with their production changes, they did so for both vowels simultaneously.

The studies mentioned in the previous two paragraphs differ in at least two important ways. The (successful) vowel changes in Evans & Iverson (2007) are contact-induced phonetic changes, while SSBE /uː/-fronting is a system-internal phonological change and, as a result, also a phonetic change. The obvious third alternative for study, and the subject of the present chapter, is a contact-driven phonological change. As the interactions between the many factors mentioned in this introduction—perception vs. production, system-internal vs. contact-driven, infant vs. adult, phonological vs. phonetic change—are still very much the topic of on-going research, filling this gap makes a small contribution to the larger puzzle of sound change in general. A currently-ongoing sound change in Dutch offers a unique opportunity to study precisely this type of change.

4.1.4 The present study

The present study investigates the role of individual variation in sound change, using an on-going vowel shift in Dutch which has led to notable sociolinguistic variation. Dutch is spoken both in the Netherlands and in Flanders, the northern part of Belgium. In the Netherlands, a sound change is currently on-going whereby the tense mid vowels /eː,øː,oː/ are becoming diphthongs [ei,øy,ou] (van der Harst 2011, van der Harst, Van de Velde, & van Hout 2014, Van de Velde 1996, Zwaardemaker & Eijkman 1924), except when followed by coda /l/ (realized as [ɫ], with optional vocalization in the Netherlands; van Reenen & Jongkind 2000) or another approximant consonant (/r,ʋ,j/) in specific phonological configurations (Berns & Jacobs 2012, Botma, Sebregts, & Smakman 2012, Voeten 2015). In addition, the original diphthongs /ɛi,œy,ɔu/ have begun to diphthongize more strongly in the Netherlands (Blankestein 1994, Gerritsen & Jansen 1980, Gussenhoven & Broeders 1976, van Heuven, Van Bezooijen, & Edelman 2005, Jacobi 2009, Mees & Collins 1983, Stroop 1992, Stroop 1998, Van de Velde 1996, Voortman 1994) except before coda /l/, while they diphthongize more weakly in Flanders (Van de Velde 1996, Verhoeven 2005). In both varieties, vowels are categorically realized as monophthongs before coda /l/ (Goossens, Taeldeman, & Verleyen 1998:maps 63/66, Voeten 2015). Thus, the Netherlandic vowel system has six diphthongs which have corresponding monophthongal allophones before coda /l/, whereas the Flemish system has three monophthongs and three very light diphthongs, which are also monophthongized before coda /l/. (For extensive discussions of Netherlandic–Flemish vowel-system differences, see Adank, van Hout, & Smits 2004 and Van de Velde et al. 2011, in addition to the aforementioned references.)

The chapter focuses its investigation on the adoption of the Netherlandic diphthongization patterns in *sociolinguistic migrants*: individuals born in Flanders who moved to the Netherlands post-adolescence and have lived there for a significant amount of time. It is investigated whether these sociolinguistic migrants adopt the Netherlandic realizations as well as their phonological conditioning related to coda /l/. Thus, the present chapter fits into the second-dialect acquisition literature (which is also the perspective of many of the studies discussed at the end of Section 4.1.2), and is indirectly related to the second-language acquisition literature (e.g. Flege 1987, Flege & Wayland 2019) as well. A comprehensive treatment of these perspectives is beyond the scope of this chapter; the reader is referred to the aforementioned works instead. The present chapter makes *use* of the situation of sociolinguistic migrants to serve as a model of the acquisition of sound change. This is possible because the synchronic sociolinguistic differences happen to coincide with the diachronic changes in the Netherlandic vowel system, such that the Flemish vowel system can be seen as a model for the Netherlandic system before these changes had taken place. This makes it possible to use laboratory experiments comparing sociolinguistic migrants to two suitable control groups to investigate how such synchronic differences are adopted by individuals on their way to becoming diachronic changes. The hypothesis is that the sociolinguistic migrants are not homogeneous; instead, some participants in this group will have become more Netherlandic-like, whereas others will have remained more Flemish-like. This hypothesis is assessed separately for production and perception, using two laboratory experiments which include suitable control groups. This makes it possible to investigate with precision whether and which of these participants have adopted the on-going sound changes in production and in perception. The link between production and perception is also examined.

Concerning production, we know that Flemish-Dutch speakers do not diphthongize their tense mid vowels /eː,øː,oː/ and only weakly diphthongize their "true" diphthongs /ɛi,œy,ɑu/. In contrast, Netherlandic-Dutch speakers use diphthongal realizations for all six of these vowels, especially strongly so for the diphthongs /ɛi,œy,ɑu/. However, all of this diphthongization is blocked categorically before a following coda /l/. Hence, predictions for the Flemish-Dutch speakers are monophthongal realizations of /eː,øː,oː/ and weakly-diphthongal realizations of /ɛi,œy,ɑu/, whereas predictions for the Netherlandic-Dutch speakers are diphthongal realizations of /eː,øː,oː/ and strongly-diphthongal realizations of /ɛi,œy,ɑu/, but only when there is no following coda /l/. If there is a coda /l/, the Netherlandic group is expected to realize a monophthong, with possible weak diphthongization due to coarticulation, leveling the hypothesized group differences. The migrant group is hypothesized to be heterogeneous: given the background sketched in Section 4.1.1, it is expected that some participants will have been relatively successful at adopting the Netherlandic-Dutch sound changes, whereas others will have done so only minimally. On average, the group should then be in between the Netherlandic and Flemish groups, though with significant intragroup variation.

Turning to perception, the linguistic facts remain the same, but their implications are different. Previous research has shown that listeners of Dutch are sensitive to the specific trajectory of diphthongal vowels in perception (Peeters 1991). Because Flemish speakers of Dutch do not produce diphthongization for the vowels /eː,øː,oː/, it is expected that their category boundary between the tense mid vowels and the diphthongs will be relatively close towards the monophthongal realizations. In concrete terms, even a little diphthongization will be a cue for a Flemish-Dutch speaker to perceive a vowel as /ɛi,œy,ɑu/. Alternatively, a reviewer suggests that it is also possible that the Flemish group does not perceive such light diphthongization at all, precisely because they do not have it in their repertoire. For a Netherlandic-Dutch speaker the same weak diphthongization should be a perfectly regular cue for a percept of /eː,øː,oː/, but only in absence of a following coda /l/. In this situation, it is expected that the Netherlandic group has their category boundary further towards the diphthongal realizations than the Flemish group: the Netherlandic group should require more diphthongization to be present in an ambiguous signal to perceive it as /ɛi,œy,ɑu/. This effect is expected to be quite strong, because in addition to the diphthong–monophthong differences between the two varieties (Adank, van Hout, & Smits 2004, van der Harst 2011, van der Harst, Van de Velde, & van Hout 2014, Van de Velde 1996, Van de Velde et al. 2011), the Netherlandic realizations of /ɛi,œy/ have significantly lower onsets and are more similar to [ai,ɒy] (van Heuven, Van Bezooijen, & Edelman 2005, Jacobi 2009, Stroop 1998). If the vowel is followed by /l/, these between-groups differences are hypothesized to vanish: here, neither group has grounds to expect diphthongization for reasons beyond effects of phonetic implementation. As for production, the migrant group is expected to be in between the two control groups at the group level, but is expected to show significant individual variation, correlated with the individual variation found in the production experiment.

Section 4.2 investigates the hypotheses for the production part, using a simple word-list-reading task eliciting a representative subset of the relevant vowels both before /l/ and before nonapproximant consonants. Due attention is paid not just to variation at the group level, but also to individual differences. Individual differences are investigated by analyzing the predicted random effects of a mixed-effects model, which has only recently gained traction in sociophonetics (e.g. Drager & Hay 2012, Tamminga to appear). The analysis controls for one major factor, namely lexical frequency, which is known to play a role both in sound change (Bybee 2002) and in the adoption of sociolinguistic differences (Nycz 2013).

The perception hypotheses are investigated in Section 4.3. Participants' category boundaries between monophthong and diphthong phonemes are elicited on the basis of a novel experiment based on participants' rhyme decisions to nonsense words. In general, rhyme-decision tasks have been used successfully in the related field of second-dialect acquisition (Nycz 2011) and have the advantage of being less direct than the more traditional task of phoneme decision (as had been used in e.g. Pinget 2015). Compared to phoneme decision, rhyme decision is less obvious about the nature of the experiment and is more linguistic rather than meta-linguistic in nature. Both points serve to reduce the likelihood of participants resorting to explicit cognitive strategies that do not reflect their everyday linguistic processing, and of them letting any overt or covert linguistic norms influence their responses. This is important especially for the migrant participants, who may have become subliminally or supraliminally aware of the relevant accent differences. The rhyme-decision task tests three phonetic continua: [eː∼ɛi] (testing diphthongization), [oː∼ɑu] (testing diphthongization plus marked lowering of the nucleus), and a control contrast [ɛ∼ɛi] (testing diphthongization and duration). Rhyme decisions between an ambiguous auditory prime and an orthographic target are used to obtain covert phoneme decisions: at what point along the continuum do participants stop perceiving /eː,oː/ and start perceiving /ɛi,ɑu/? The condition using [ɛ∼ɛi] is included as a control: these vowels are like the experimental items in that /ɛ/ is a monophthong and /ɛi/ is a diphthong, but the boundary between these otherwise unrelated categories is unaffected by the on-going sound changes studied in this chapter.

Section 4.4 investigates the individual-level correlations between production and perception, which were suggested to exist in the formulated hypotheses. This is done by calculating correlation coefficients between the participants' coefficients from the individual-level analyses of the production and perception tasks. The findings are brought together in the general discussion in Section 4.5. Finally, Section 4.6 concludes the chapter with possible avenues for further study.

4.2 Experiment 1: production

4.2.1 Method

Participants

Three groups of participants were recruited. The first group consisted of 45 Dutch students at Leiden University, the Netherlands. These students were native speakers of Netherlandic Standard Dutch and were born and raised in the Randstad[2], which is the area of the Netherlands where the prestigious variety of Dutch is spoken which forms the basis for Netherlandic Standard Dutch (Grondelaers & van Hout 2011). The second group consisted of 45 Belgian students at Ghent University, Belgium. These students were all native speakers of Belgian Standard Dutch and were born and raised in Flanders. The third group of participants consisted of 18 Belgians who were native speakers of Belgian Standard Dutch and had been living in the Randstad area of the Netherlands for a long time (mean = 18.71 years, SD = 11.18 years). Two participants in the Ghent group were excluded due to technical failures, resulting in a total of 106 participants used.

Table 4.1 details participants' reported regional backgrounds, split out by province. In addition to speaking the standard language (either Netherlandic Dutch or Flemish Dutch), seven of the Ghent participants and seven of the migrant participants reported being proficient in their local Flemish dialects. None of the Leiden participants reported dialect competence. Figure 4.1 shows the variation in the participants' ages and, for the migrant participants, their lengths of stay in the Netherlands. The migrant participants' ages are heterogeneous, but their ages of arrival are not: with two exceptions, participants were in their mid-twenties (range: 18–32 years old) when they migrated to the Netherlands. Of the two exceptional participants, one had just turned 43 when they moved countries, and the other was a few days short of their 60th birthday upon migration.

The experiments followed the Ethics Code for linguistic research in the faculty of Humanities at Leiden University, which approved its implementation. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

---

[2] Five participants attended high school in non-Randstad areas, but as the results show, they were not significantly different from the participants who never left the Randstad after their childhood years.

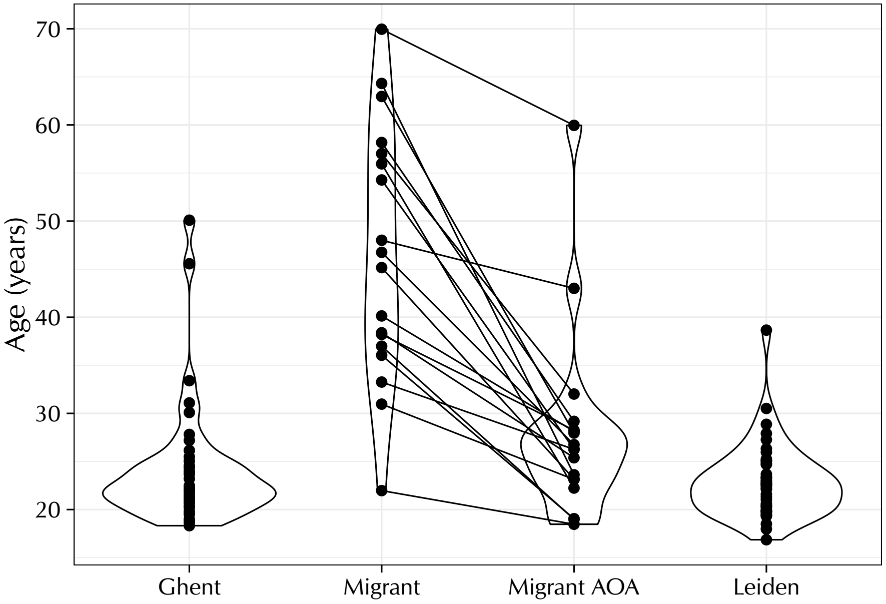

Figure 4.1: Dot and violin plots of the ages of the participants in the Ghent (n = 43, 1 missing age value), migrant (n = 18, and Leiden (n = 45, 2 missing age values) groups. For the migrant group, the chronological ages have been linked to the participants' ages of arrival in the Netherlands ("AOA").

Table 4.1: Regional backgrounds of the participants, defined as the province in which they attended high school. Gelderland–Zeeland are Netherlandic provinces, while the others are Flemish.

	Group		
Region	Ghent	Migrant	Leiden
Gelderland	0	0	2
Netherlandic Limburg	0	0	1
North Holland	0	0	6
South Holland	0	0	28
Utrecht	0	0	5
Zeeland	0	0	2
Antwerp	5	1	0
Brussels	0	1	0
East and West Flanders	0	1	0
East Flanders	23	4	0
Flemish Brabant	6	4	0
Flemish Limburg	0	2	0
West Flanders	14	5	0
(missing)	0	0	1

Stimuli

Words were selected from the combined CELEX (Baayen, Piepenbrock, & Gulikers 1995) and SUBTLEX (Keuleers, Brysbaert, & New 2010) corpora[3] to elicit (possible) diphthongs followed by either coda /l/ or a nonapproximant consonant. Words were selected such that the critical vowel always received primary stress. Words sharing the same lemma were avoided. For the words where the critical vowel was followed by a nonapproximant consonant, a distinction was made between low-frequency and high-frequency (henceforth: "HF") words, defined by the set of words falling in the first and third quartiles, respectively, of the $\log_{10}$ word frequency. This distinction was not made for the coda-/l/ words, due to there not being enough high-frequency V+/l/ words in the corpus. The vowels /øː/ and /ɑu/ were included in the study but excluded from the data analysis, because the former vowel was not frequent enough to fill the design cells with the full 20/40 tokens and the latter cannot be followed by coda /l/ due to a lexical gap (which, as also mentioned by van Reenen & Jongkind 2000 and Peeters 1991, is due to a historical process of /l/-vocalization and diphthongization). Table 4.2 summarizes the stimuli design.

---

[3] CELEX was used as a starting point because it has phonetic transcriptions, but SUBTLEX was used for its more reliable indication of word frequency (Keuleers, Brysbaert, & New 2010).

Table 4.2: Number of words in each cell of the stimuli design of the production experiment, with the column "Analyzed" reflecting which of these were included in the data analysis. In addition to the vowels mentioned here, the practice part of the experiment elicited 30 tokens (5 per combination) of /i,u,aː/ before coda /l/ and before non-/l/.

		Following segment + frequency		
Vowel	**Analyzed**	coda /l/	non-app. + LF	non-app. + HF
/eː/	Yes	20	20	20
/oː/	Yes	20	20	20
/ɛi/	Yes	20	20	20
/œy/	Yes	20	20	20
/øː/	No	11	20	0
/ɑu/	No	0	20	20

In addition to the vowels mentioned there, the practice trials of the experiment were used to elicit the point vowels /i,u,aː/, both preceding /l/ (5 tokens per vowel) and preceding a nonapproximant consonant (5 tokens per vowel), but without regard for frequency. These were also excluded from the analysis, and only serve to provide anchor points for the vowel-space plots in Figure 4.2. The full list of items is available in Appendix C.

Procedure

For the Leiden and migrant participants (who were tested in Leiden), the experiment took place in a dimly-lit, sound-attenuating booth, where participants were seated in front of a computer screen and a studio-quality microphone. For the Ghent participants, the experiment took place in a quiet room, where participants were seated in front of a laptop and wore a studio-quality headset. The participants performed a word-list-reading task, presented using E-Prime (Psychology Software Tools 2012) running on Windows 7 (both setups). The words selected for the experiment were presented on the screen one by one (in pseudorandomized order), and participants were instructed to read these words aloud into the microphone. Each trial had a fixed duration of two seconds and was followed by a fixation cross presented for 500 ms. A total of 2.5 s was thus available for speaking (the presentation of the fixation cross did not terminate the recording).

Data analysis

The acquired single-word recordings were forcedly aligned to their CELEX reference transcriptions using HTK (Young et al. 2002). Using Praat (Boersma & Weenink 2016), the F1s of the vowels of interest were extracted at 25% and 75% realization of the vowel (Burg algorithm, time step 10 ms, 5 formants, cutoff point 5,000 Hz for men and 5,500 Hz for women, window length 25 ms, pre-emphasis from 50 Hz). The choice to use 25% and 75% points is based on findings by van der Harst (2011) that these are the first and last time points, respectively, that are not too strongly influenced by coarticulation. Following van der Harst (2011:82), outliers were identified for male and female speakers separately and for the two measurement points separately, with outliers defined as F1 values whose absolute values exceeded 1.5 times the interquartile range within the relevant combination of gender and timepoint. These outliers' formant values were removed from the data, based on the argument that they constituted either formant-tracking errors or forced-aligner errors. Of the 71,054 measurements in total, 67,283 (94.7%) remained after this procedure. The data were subsequently normalized using the Lobanov method; this was found to be the best-performing method for this type of data in the comparisons by Adank, Smits, & van Hout (2004) and van der Harst (2011).

In order to capture diphthongization in a single variable suitable for statistical modeling, the two time points per token were converted into a single difference score, to which I henceforth refer as "ΔF1". This score was computed for each token by subtracting the measurement at 25% from the measurement at 75% and encodes the amount of diphthongization present in the vowel. Negative values indicate upgliding diphthongization, while values of zero or slightly above it indicate lack of upgliding diphthongization. The resulting data contained 32,861 difference scores, which amounts to 92.5%[4] of the original data. Finally, only the vowels /eː,øː,ɛi,œy/ were selected from these data, yielding a final dataset of 23,393 difference scores.

The F1 difference scores were analyzed using a mixed-effects model in two ways. In both approaches, the dependent variable was ΔF1. Fixed effects were included for "Vowel" (levels /eː,øː,oː,ɛi,œy,ɑu/; sum-coded such that /ɛi/ = −1 and the other vowels = 1; this coding scheme makes the estimated contrasts relative to the grand mean of all vowels), "Following segment" (levels /l/ or non-/l/, treatment-coded such that non-/l/ = 1 and /l/ = 0), frequency (deviation-coded such that High Frequency = 0.5 and other = −0.5; this coding scheme uses the mean as the reference, but estimates the difference between the two frequency types, rather than their difference from the mean), and all appropri-

---

[4] The additional data loss compared to the previous 94.7% is because if *either* element of a pair of ⟨25%, 75%⟩ measures had been removed, their difference is undefined even if the other element of the pair was still present.

ate interactions. In the analysis in Section 4.2.2, a factor "Group" has also been included, coded using Helmert coding such that the first contrast compares the Ghent group and the migrant group, and the second contrast compares the Leiden group to the average of the other two. A maximal model structure was formed including all interactions and full random slopes by participants and by words, but excluding correlations between random slopes. The model was fitted to scaled-*t* errors, using function `bam` from `R` package `mgcv` (Wood 2017). From this maximal model, terms that did not achieve omnibus significance[5] were removed using backward elimination (see Matuschek et al. 2017 for justification) to arrive at a parsimonious final model (per Bates et al. 2015a).

Based on the studies reported in Section 4.1.1, it can be considered questionable to lump the 18 migrant participants together into a single explicit group. For this reason, Section 4.2.2 presents an alternative analysis, where the interest was not in group patterns, but in individual differences. The objective of this second analysis was to find homogeneous groups in the set of participants, in order to identify which participants are more Ghent-like and which participants are more Leiden-like. This analysis did not include the factor "Group" and fitted the full model directly, without performing stepwise elimination. For each of the by-participants random-effect terms estimated by the model, the predicted *b* values were extracted. These values are the individual participants' coefficients for the estimated random slopes. Function `Mclust` from R package `mclust` (Scrucca et al. 2016) was used to perform a cluster analysis on these coefficients, based on a one-dimensional variable-variance cluster model. This analysis provides a quantitative measure of the degree to which individual participants are more Ghent-like or more Leiden-like.

The data and R code for the analyses are available at https://figshare.com/s/731e0a32480e876530e0 as the files `production.csv` and `production.R`, respectively.

4.2.2 Results

Figure 4.2 shows vowel-space diagrams of the collected data, without any prior analysis. The figure shows the four vowels of critical interest, plus the point vowels /i,u,aː/ and the excluded diphthongs /øː,ɑu/ for context.

Analysis by groups

Table 4.3 shows the results of the analysis in which the three groups of participants were categorized into their respective three groups *a priori*. The results show that a following non-/l/ induces significant upgliding diphthongization

---

[5] Likelihood-ratio tests are not applicable to `bam` models with non-Gaussian errors, as these are fitted using penalized quasi-likelihood ("PQL"); see Wood (2017:149–151) for details.

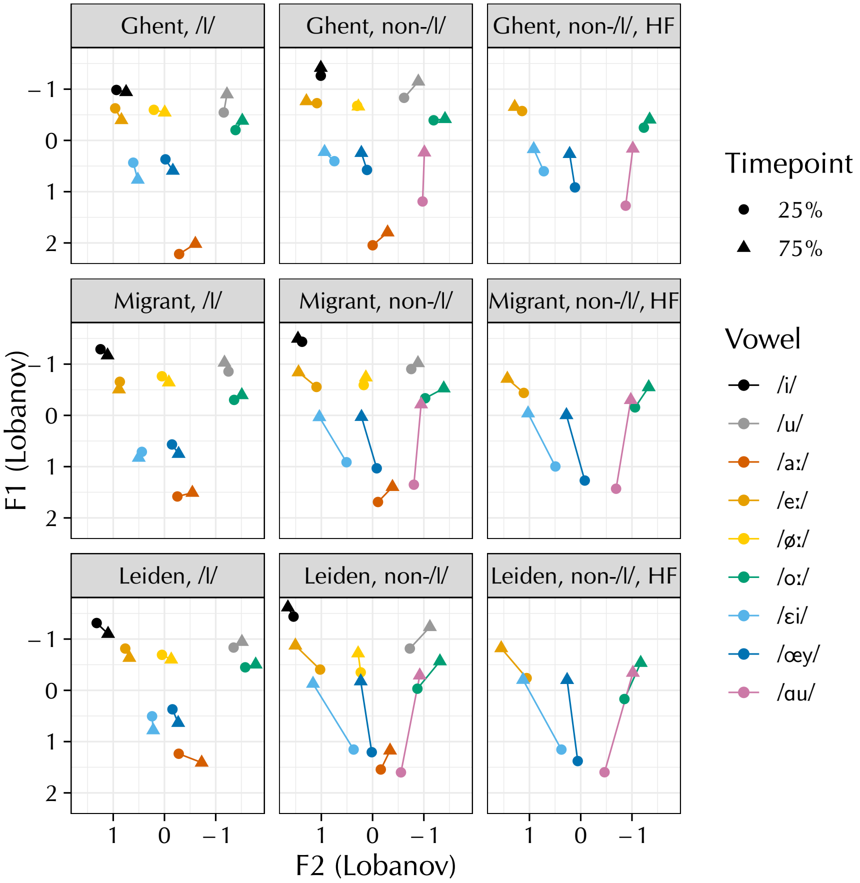

Figure 4.2: Vowel diagrams of the data collected in the study (106 participants, 361 items). The figure is divided into nine panels, to account for the 3 groups × 3 types of following segment in the design. The vowels of experimental interest are /eː,oː,ɛi,œy/; the other vowels are included for context.

Table 4.3: Group-level results for the production data (106 participants, 235 items). Critical factors are "Following segment" and its interaction with the group predictors. The key observations are that the average participant produces significantly more diphthongization before non-/l/ than before coda /l/, and that this additionally varies between the three groups. The migrant group produces significantly more diphthongization in this context than the Ghent group, and the Leiden group produces even more diphthongization in this context than the average of the other two groups.

Factor	**Estimate (SE)**	***t***	***p***	**Sig.**
Intercept	0.20 (0.03)	6.86	<.001	***
Vowel = /eː/	0.02 (0.04)	0.45	.65	
Vowel = /oː/	−0.24 (0.05)	−5.38	<.001	***
Vowel = /œy/	0.11 (0.04)	2.43	.02	*
Following segment = non-/l/	−0.79 (0.04)	−22.12	<.001	***
Group = Migrant–Ghent	−0.06 (0.02)	−2.92	<.01	**
Group = Leiden–Others	0.01 (0.01)	0.46	.64	
Following segment = non-/l/ × Frequency = HF	−0.05 (0.02)	−2.58	<.01	**
Vowel = /eː/ × non-/l/	0.31 (0.06)	5.59	<.001	***
Vowel = /oː/ × non-/l/	0.56 (0.06)	9.91	<.001	***
Vowel = /œy/ × non-/l/	−0.53 (0.05)	−9.68	<.001	***
Following segment = non-/l/ × Migrant–Ghent	−0.19 (0.03)	−7.14	<.001	***
Following segment = non-/l/ × Leiden–Others	−0.20 (0.02)	−12.13	<.001	***
Vowel = /eː/ × non-/l/ × Migrant–Ghent	0.12 (0.03)	4.45	<.001	***
Vowel = /oː/ × non-/l/ × Migrant–Ghent	0.11 (0.03)	3.91	<.001	***
Vowel = /œy/ × non-/l/ × Migrant–Ghent	−0.15 (0.02)	−6.20	<.001	***
Vowel = /eː/ × non-/l/ × Leiden–Others	0.07 (0.02)	4.21	<.001	***
Vowel = /oː/ × non-/l/ × Leiden–Others	0.03 (0.02)	1.81	.07	
Vowel = /œy/ × non-/l/ × Leiden–Others	−0.04 (0.02)	−2.85	<.01	**

($\hat{\beta} = -0.79$, SE = 0.04, $t = -22.12$, $p < .001$), but there are large and highly significant differences between the groups in this respect. The significant "Following segment = non-/l/ × Migrant–Ghent" interaction shows that migrants on average produce −0.19 standard deviations more diphthongization (SE = 0.03, $t = -7.14$, $p < .001$) than the Ghent group does. There is also a significant interaction of "Following segment = non-/l/ × Leiden–Others", indicating that the Leiden participants produce even more diphthongization than the other two groups do: on average, they diphthongize an additional −0.20 standard deviations more (SE = 0.02, $t = -12.13$, $p < .001$) than the Ghent and migrant groups. There are significant per-vowel adjustments to the regression coefficients discussed thus far, but in all three groups these are small enough that they do not rise above the crucial effect of a following non-/l/ consonant.

Analysis by participants

Figure 4.3 shows a plot of the individual participants' random-effect coefficients (henceforth: "BLUPs", for "best linear unbiased predictors") for the "Following segment = non-/l/" term; this is the single predictor of critical interest. The cluster analysis found 2 clusters for this factor, which have a clear interpretation: the cluster analysis managed to completely recover the Leiden and Ghent groups, despite not having been provided with any *a priori* group information. This was not the case for the other random-effect terms, where either only one cluster was found, or where two clusters were found but these failed to coincide with either of the group boundaries. Since these other terms were not of theoretical interest anyway, they have been relegated to Appendix D.

The BLUPs in Figure 4.3 show that one set of participants diphthongizes significantly more than average before a nonapproximant consonant, and one set of participants diphthongizes significantly less than average in this environment. All of the Leiden participants are in the former cluster, and all of the Ghent participants are in the latter cluster. Concerning the migrant group, ten of these participants diphthongize to such an extent that they are classified with the Leiden group, whereas the other eight are still classified with the Ghent group. An anonymous reviewer asks whether the two migrant participants who had arrived to the Netherlands relatively late were among those clustered with the Ghent group. This was indeed the case.

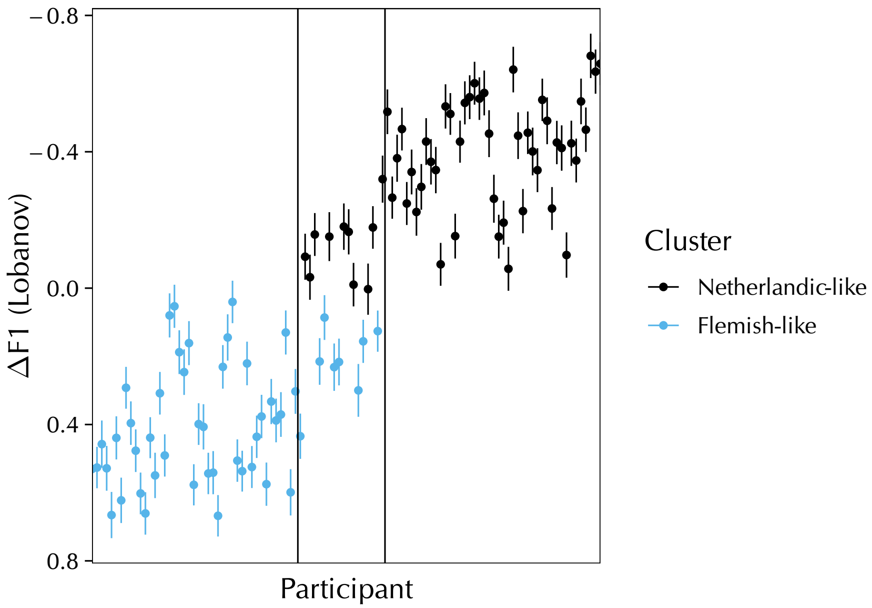

Figure 4.3: BLUPs for the "Following segment = non-/l/" term in the by-individuals (n = 106) model. Each dot is a participant's individual random-effect coefficient; lines indicate the standard errors. The left pane shows the participants from the Ghent group, the middle pane shows the participants from the migrant group, and the right pane shows the participants from the Leiden group.

4.2.3 Discussion

The two largest effects in the group-level analysis are the effect for "Following segment = non-/l/" and its interaction with "Group = Leiden–Others". The former shows that the main difference is between all vowels before non-/l/ versus before /l/. The latter shows, wholly in line with the hypothesis, that the Leiden participants diphthongize significantly more than the Ghent participants. As a group, the migrant participants are in between: they are significantly different from the Ghent participants, but do not diphthongize as much as the Leiden participants (their effect was approximately one third the size of that of the Leiden group).

The results at the individual level confirm and extend these findings. The cluster analysis shows that nearly all Leiden participants produce significantly more diphthongization than nearly all Ghent participants. The migrant participants are in between: some diphthongize to such extent that they are classified with the Leiden participants, some do not and are classified with the Ghent participants. Thus, the cluster analysis makes it possible to identify precisely which individuals make positive or negative contributions to the overall group-level effect. The critical difference in diphthongization was captured by the BLUPs for the "Following segment = non-/l/" random slope, which is the grand mean of all five possibly-diphthongizing vowels. This shows that these differences in diphthongization are across the board, and are not specific to one vowel or one subset of vowels.

The results suggest a role for age of arrival in determining the migrant participants' degrees of sound-change adoption, insofar as the two participants who arrived well past their twenties were clustered with the Flemish group, whereas the migrants who were classified with the Netherlandic group were all in their mid-twenties when they arrived in the Netherlands. The data prohibit a formal test of this observation, both because of the small sample size and because it is confounded with participants' lengths of exposure. However, on purely theoretical grounds it seems likely that individuals with younger AOAs would indeed more readily adopt accent differences such as those discussed here. Such participants are likely more cognitively flexible, and may also have a greater desire to fit into their peer group. They may hence be both able and willing to adopt their peers' accents. The results by Evans & Iverson (2007) corroborate this view, but the results by Nycz (2013)—who found no effect of AOA, despite having tested a relatively comparable participant group[6]—again muddle the picture. Further investigation of a link between age of arrival, length of exposure, and cognitive and social factors influencing the adoption of accent differences and sound change is left to future research.

---

[6] While Nycz (2013) does not report participants' ages of arrival, Table 2.1 in Nycz (2011) shows that they ranged from 21 to 47 years of age, distributed approximately normally with M = 32.26 years and SD = 6.51 years.

4.3 Experiment 2: rhyme decision

4.3.1 Method

Participants

The participants were the same as in Experiment 1, which was performed on the same occasion. Half of the participants had participated in Experiment 1 prior to taking Experiment 2, and for the other half the experiments were performed in the opposite order.

Stimuli

The experiment used vowel pairs of [eː∼ɛi], [oː∼ɑu], and [ɛ∼ɛi]. As detailed in Section 4.1.4, the former two are of experimental interest, whereas the latter served as a control condition. For the orthographic targets, 2 × 192 pseudowords were generated according to a template of [C_1_(ɫ).{t,d}ə(ɹ)]; one pseudoword was generated for each vowel in each pair. The presence/absence of the parenthesized /l/ and /r/ and the choice between /t/ and /d/ were perfectly balanced, leading to 16 items per cell for a total of 384 targets, half of which (viz. those with the word-final [ɹ]) look like plausible Dutch nouns and half of which (viz. those without the word-final [ɹ]) look like plausible Dutch inflected verbs. Pseudoword pairs were generated to maximize (a) syllable frequencies and (b) transition probabilities, for both words in each pair together, so as to maximize the naturalness of the words included in the experiment. Real words (defined as words occurring in CELEX; Baayen, Piepenbrock, & Gulikers 1995) were excluded to prevent the possible confounding of the experiment by this extra factor.

An equal number of auditory prime words were generated in exactly the same way as the orthographic targets. All generated prime words were read aloud in a carrier sentence by a female native speaker of Netherlandic Standard Dutch from the Randstad area of the Netherlands, who read the stimuli using her regular accent. For each vowel, two tokens were selected, one followed by coda /l/ and one followed by a syllable boundary. Using Tandem-STRAIGHT (Kawahara et al. 2008), a continuum of four intermediate steps (20%–40%–60%–80% morphing from the monophthong to the diphthong realization) was generated for each vowel pair using holistic morphing between the two selected vowels. Figure 4.4 shows an example of the resulting waveforms, spectrograms, and F1 trajectories (the critical difference between monophthongal and diphthongal realizations) for the [eː∼ɛi] contrast. These manipulated vowels were spliced into the original prime words. Any F0 discontinuities were smoothed out using the PSOLA algorithm in Praat (Boersma & Weenink 2016).

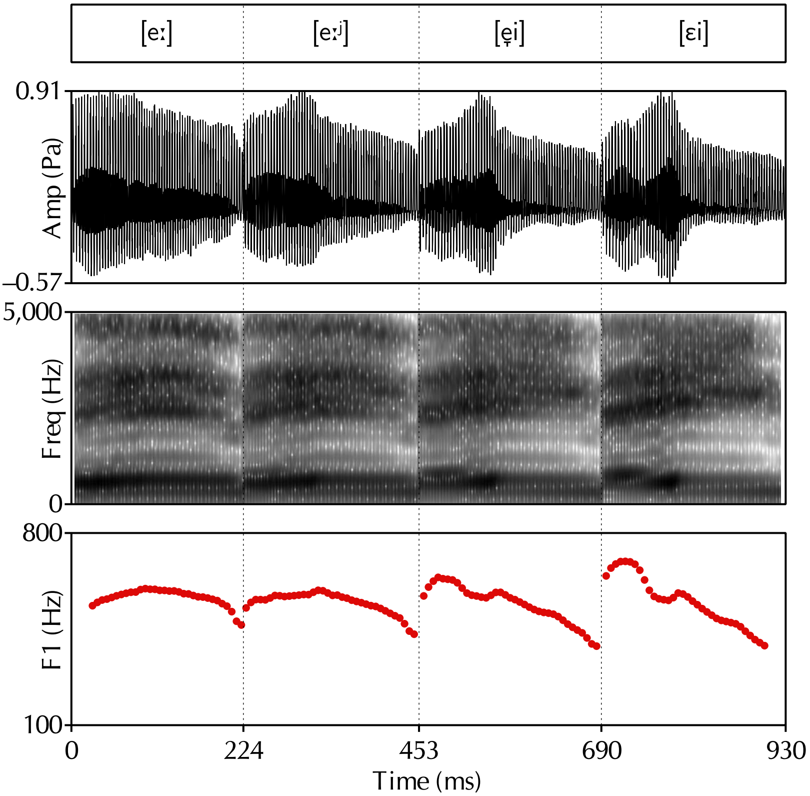

Figure 4.4: Example waveforms, spectrograms, and F1 trajectories for the [eː∼ɛi] contrast. The four tokens shown in this figure correspond to the stimuli containing 20%, 40%, 60%, and 80% morphing, respectively. Note the increase in the diphthong's slope and the lowering of its nucleus over the four figures.

The prepared prime pseudowords were paired to the target pseudowords, such that the prime and target would rhyme if and only if participants perceived the same vowel in the prime word as they read in the target word. As an example, the orthographic target pair ⟨nebe,nijbe⟩ (/neːbə,nɛibə/) was paired with a set of auditory primes which can be rendered approximately as [beːbə,beːʲbə, be̞ibə,bɛibə], which respectively correspond to the 20%–40%–60%–80% steps. The resulting stimuli were randomized across four lists, which were assigned to participants in fixed order (participants 1, 5, and 9 received list A, participants 2, 6, and 10 received list B, etc.). The pairing of target and prime words was yolked across these lists, such that between all of the participants, all combinations of target word and morphing step in the corresponding prime word were represented (list A paired ⟨nebe,nijbe⟩ with [beːbə] and ⟨kede,kijde⟩ with [veːʲdə], list B paired the same two targets with primes [beːʲbə] and [ve̞idə], etc.). This ensured that all participants would be presented with all steps of the continua and with all words in the experiment, without repeating individual pseudowords with a different morphing step (which would be another possible confounding factor). The full list of prime–target pairs is available in Appendix E.

Procedure

Participants were seated in the same booth as in Experiment 1. Instructions about the procedure of the experiment were presented on the computer screen. Participants could start the experiment whenever they were ready by pressing one of the five buttons on the Serial Response box that was in front of them. For each trial in the experiment, the prime word was played while the target word was displayed on the screen. Participants had to indicate, by pressing either the leftmost or the rightmost button on the response box, whether the prime and target word rhymed or not. For half of the participants, the leftmost button indicated "yes" and the rightmost button indicated "no"; for the other half of the participants, this was swapped to counterbalance for any possible left–right response bias to the stimuli. All experimental items were presented twice: once for each of the two orthographic targets assigned to the auditory prime for each participant. After completion of a trial, a fixation cross appeared for 500 ms, after which the next trial was presented. There were three breaks spaced evenly throughout the experiment. Before the real experiment began, participants were presented with a few practice trials. These consisted of a total of 12 prime–target combinations, generated and administered according to exactly the same procedure as the main experiment.

While the general approach of rhyme decision has been used before (Nycz 2011), the paradigm proposed in the present study, which uses nonsense words, is novel. Therefore, the experiment was subjected to extensive piloting, on two occasions: once with a group of colleagues in Leiden University's linguistics department, to solicit comments on anything from the general principles behind the experiment to subjective experiences of individual stimuli, and once with a group of Netherlandic-Dutch speakers and a control group of Flemish sociolinguistic migrants (different from the group tested in the present chapter) to validate that the expected effects were indeed borne out. The lessons learned from the first pilot were used to improve the second pilot, which validated that this experimental paradigm could indeed capture participants' perceptual categories in sufficient detail. These results, reported in more detail in Chapter 3, showed the expected, approximately linear, increase from a very small towards a very large probability of reporting a diphthong percept as the morphing step increased.

Data analysis

To directly test the effect of morphing step on participants' vowel perceptions, participants' yes/no responses to the prime–target combinations were recoded into vowel percepts associated with each auditory prime. Responses with reaction times <100 ms or >5,000 ms were excluded from the dataset. The remaining data were analyzed using a similar approach to the one used in Section 4.2.1. Separate[7] mixed-effects logistic-regression models were fitted for the three conditions [eː∼ɛi], [oː∼ɑu], and [ɛ∼ɛi]. The dependent variable was "Phoneme decision", which coded whether the participant's judgment was consistent with the monophthongal phoneme (/eː,oː,ɛ/, coded as 0) or the diphthongal phoneme (/ɛi,ɑu,ɛi/, coded as 1). Fixed effects were added for "Step" (coded for linear, quadratic, and cubic trends using orthogonal polynomials) and "Following segment" (deviation-coded such that non-/l/ $= -0.5$ and /l/ $= 0.5$; this coding scheme tests for the difference between the two stimuli while using their average as the reference). Random intercepts and slopes for all predictors by participants were included, as was a random intercept by items (the prime-target pair presented in each trial). As in Section 4.2.1, models were run both with and without an explicit factor "Group". When included, the factor "Group" was coded in the same way as in Section 4.2.2, with all fixed-effect interactions and a random slope by items. For the by-groups model, function `buildglmmTMB` from R package `buildmer` (Voeten 2019b) was used to identify the maximal random-effect structure that still converged, and terms were eliminated using backward stepwise elimination based on the likelihood-ratio

---

[7] The reason for fitting three separate models as opposed to one model with "Condition" as a factor is that the three models differ not only on the initial and final endpoints of the continuum, but also on the acoustic range spanned by the four intermediary steps. As such, quantitative differences between the three conditions are not interpretable: they could be due to linguistic differences or due to differences in the acoustic endpoints.

test. The data and R code for the analyses are available at https://figshare.com/s/731e0a32480e876530e0 as the files `rhyme.csv` and `rhyme.R`, respectively.

4.3.2 Results

Cluster analyses on the by-participants model revealed no robust groupings: all analyses yielded only one cluster. For this reason, only the results from the three by-groups models are reported here. These results are shown in Table 4.4. This table only shows results that achieved significance according to a Bonferroni-corrected α of .017; significance stars have been corrected to reflect two-tailed *p*-values of .017 ($*$), .0033 ($**$), and .00033 ($***$). Appendix F presents also the results that did not reach significance.

To aid interpretation of the model coefficients, Figure 4.5 plots the raw data on which these models have been based. The three panels of this figure correspond to the three fitted models, which will now be discussed in turn.

The model for the [eː∼ɛi] condition reveals significant linear and quadratic effects of "Step" ($\hat{\beta} = 1.24$, SE = 0.13, OR = 3.47 : 1, $z = 9.87$, $p <.001$; $\hat{\beta} = 0.33$, SE = 0.08, OR = 1.39 : 1, $z = 4.22$, $p <.001$), indicating a steeper-than-linear trend of obtaining a diphthong percept at later morphing steps. A following /l/ increased the odds of a participant choosing the diphthongal target, both at the baseline ($\hat{\beta} = 0.83$, SE = 0.11, OR = 2.29 : 1, $z = 7.59$, $p <.001$) and as a (linear) function of morphing step ($\hat{\beta} = 1.68$, SE = 0.21, OR = 5.37 : 1, $z = 8.14$, $p <.001$). The Ghent group had lower baseline odds of obtaining a diphthong percept than the other two groups ($\hat{\beta} = -0.51$, SE = 0.13, OR = 1 : 1.66, $z = -3.95$, $p <.001$), but their odds also increased more steeply as a linear function of the morphing step ($\hat{\beta} = 0.63$, SE = 0.15, OR = 1.87 : 1, $z = 4.06$, $p <.001$). Taken together, these results paint a picture where participants become more likely to opt for the diphthong target at later morphing steps, which is exactly what was expected based on the manipulation. In the coda-/l/ condition, participants were already more likely to obtain a diphthong percept, and became even more so at the later morphing steps, doing so more rapidly than in the non-/l/ condition. The differences between the groups revolved around the difference between the Leiden and migrant groups on the one hand and the Ghent group on the other—the latter initially showed a preference for the monophthongal targets, but went for the diphthongal targets more rapidly than the other groups at the later morphing steps.

For the [oː∼ɑu] model, the results show the same effect of participants becoming more likely to select the diphthong target at later morphing steps, which again developed according to a combined linear and quadratic trend ($\hat{\beta} = 1.73$, SE = 0.15, OR = 5.65 : 1, $z = 11.48$, $p <.001$; $\hat{\beta} = 0.45$, SE = 0.11, OR = 1.57 : 1, $z = 4.17$, $p <.001$). There were again significant effects of a following /l/ ($\hat{\beta} = 0.95$, SE = 0.12, OR = 2.59 : 1, $z = 7.83$, $p <.001$) and its interac-

Table 4.4: Results of the rhyme-decision task (106 participants, 1,536 items). Only significant results are shown; the reader is referred to Appendix E for the full result set. The key results are (1) the significant linear trends of participants indicating more diphthong percepts at later morphing steps; (2) participants becoming more likely to give diphthong responses to a following coda /l/, demonstrating perceptual compensation in the non-/l/ words; (3) significant between-groups differences in the effect of morphing step in the [eː∼ɛi] and [ɛ∼ɛi] models.

Factor	Estimate (SE)	Odds ratio	*z*	*p*	Sig.
Model = [eː∼ɛi]					
Intercept	−1.46 (0.10)	1 : 4.30	−14.49	<.001	∗∗∗
Step (Linear)	1.24 (0.13)	3.47 : 1	9.87	<.001	∗∗∗
Step (Quadratic)	0.33 (0.08)	1.39 : 1	4.22	<.001	∗∗∗
Following segment = /l/	0.83 (0.11)	2.29 : 1	7.59	<.001	∗∗∗
Group = Migrant–Ghent	−0.51 (0.13)	1 : 1.66	−3.95	<.001	∗∗∗
Step (Linear) × /l/	1.68 (0.21)	5.37 : 1	8.14	<.001	∗∗∗
Step (Linear) × Migrant–Ghent	0.63 (0.15)	1.87 : 1	4.06	<.001	∗∗∗
Model = [oː∼ɑu]					
Intercept	0.20 (0.07)	1.22 : 1	2.76	.01	∗
Step (Linear)	1.73 (0.15)	5.65 : 1	11.48	<.001	∗∗∗
Step (Quadratic)	0.45 (0.11)	1.57 : 1	4.17	<.001	∗∗∗
Following segment = /l/	0.95 (0.12)	2.59 : 1	7.83	<.001	∗∗∗
Step (Linear) × /l/	−0.66 (0.25)	1 : 1.93	−2.67	.01	∗
Step (Quadratic) × /l/	−0.74 (0.22)	1 : 2.09	−3.32	<.001	∗∗
Model = [ɛ∼ɛi]					
Intercept	−0.94 (0.07)	1 : 2.55	−13.12	<.001	∗∗∗
Step (Linear)	2.49 (0.15)	12.06 : 1	16.44	<.001	∗∗∗
Step (Quadratic)	0.27 (0.10)	1.31 : 1	2.59	.01	∗
Step (Cubic)	−0.74 (0.11)	1 : 2.09	−6.99	<.001	∗∗∗
Following segment = /l/	0.96 (0.12)	2.60 : 1	7.90	<.001	∗∗∗
Group = Migrant–Ghent	−0.21 (0.08)	1 : 1.23	−2.64	.01	∗
Group = Leiden–Others	0.09 (0.04)	1.10 : 1	2.50	.01	∗
Step (Linear) × /l/	−1.38 (0.24)	1 : 3.98	−5.84	<.001	∗∗∗
Step (Cubic) × /l/	0.87 (0.21)	2.38 : 1	4.13	<.001	∗∗∗
Step (Linear) × Migrant–Ghent	0.43 (0.17)	1.53 : 1	2.52	.01	∗
Step (Linear) × Leiden–Others	−0.23 (0.08)	1 : 1.26	−2.91	<.01	∗∗
Step (Linear) × /l/ × Leiden–Others	−0.31 (0.11)	1 : 1.36	−2.95	<.01	∗∗

tion with "Step" both linearly and quadratically ($\hat{\beta} = -0.66$, SE = 0.25, OR = 1 : 1.93, $z = -2.67$, $p = .01$; $\hat{\beta} = -0.74$, SE = 0.22, OR = 1 : 2.09, $z = -2.32$, $p < .001$). Differences between the groups are not borne out in this model. These results show that all groups of participants again became more likely to obtain a diphthong percept at later morpheme steps. If a coda /l/ followed, they became even more likely to opt for the diphthong, but the gap between the two following segments narrowed at the later morphing steps. The most important effects are those that are not found: the hypothesized group differences do not appear to be borne out in the [oː∼ɑu] condition.

Finally, the [ɛ∼ɛi] model is the model for the control condition, where the diphthong [ɛi] was morphed together with the—as far as the relevant sound changes are concerned, arbitrary—vowel [ɛ]. Reassuringly, the same linear and quadratic effects for "Step" are obtained ($\hat{\beta} = 2.49$, SE = 0.15, OR = 12.06 : 1, $z = 16.44$, $p < .001$; $\hat{\beta} = 0.27$, SE = 0.10, OR = 1.31 : 1, $z = 2.59$, $p = .01$). An additional cubic effect is also observed ($\hat{\beta} = -0.74$, SE = 0.11, OR = 1 : 2.09, $z = -6.99$, $p < .001$). These effects together create a curve that that has a sharp increase between steps 2 and 3, and much lower increases between the first two steps and between the last two steps. A following coda /l/ again increases the odds of participants choosing the diphthong target ($\hat{\beta} = 0.96$, SE = 0.12, OR = 2.60 : 1, $z = 7.90$, $p < .001$). As for the [oː∼ɑu] model, the interaction terms of "Step" by "Following segment = /l/" show that the gap between the two following consonants closes towards the later morphing steps, with evidence for both a linear trend and a cubic trend ($\hat{\beta} = -1.38$, SE = 0.24, OR = 1 : 3.98, $z = -5.84$, $p < .001$; $\hat{\beta} = 0.87$, SE = 0.21, OR = 2.38 : 1, $z = 4.13$, $p < .001$). The cubic trend corresponds to what can be observed happening in Figure 4.5 at the 60% step, where the non-/l/ condition briefly overtakes the /l/ condition. There are significant differences between all three participant groups. The migrant participants are less likely to opt for the diphthong target ($\hat{\beta} = -0.21$, SE = 0.08, OR = 1 : 1.23, $z = -2.64$, $p = .01$), but become significantly more likely to do so at later morphing steps ($\hat{\beta} = 0.43$, SE = 0.17, OR = 1.53 : 1, $z = 2.52$, $p = .01$). This simply means that their decision boundary between steps 2 and 3 is steeper. The Leiden participants are in between, both at the baseline ($\hat{\beta} = 0.09$, SE = 0.04, OR = 1.10 : 1, $z = 2.50$, $p = .01$) and in interaction with "Step (Linear)" ($\hat{\beta} = -0.23$, SE = 0.08, OR = 1 : 1.26, $z = -2.91$, $p < .001$). The latter effect becomes stronger in the presence of a following /l/; because this is a three-way interaction, Figure 4.6 provides a visualization to ease interpretation. It can be observed in this figure that the difference between the Leiden group and the other two groups follows a steeper slope in the non-/l/ than in the /l/ condition, with the non-/l/ condition eliciting much fewer diphthong responses in the first three morphing steps. However, by the final morphing step, these effects have crossed over, such that there the non-/l/ condition elicits slightly more diphthong responses than the /l/ condition than the /l/

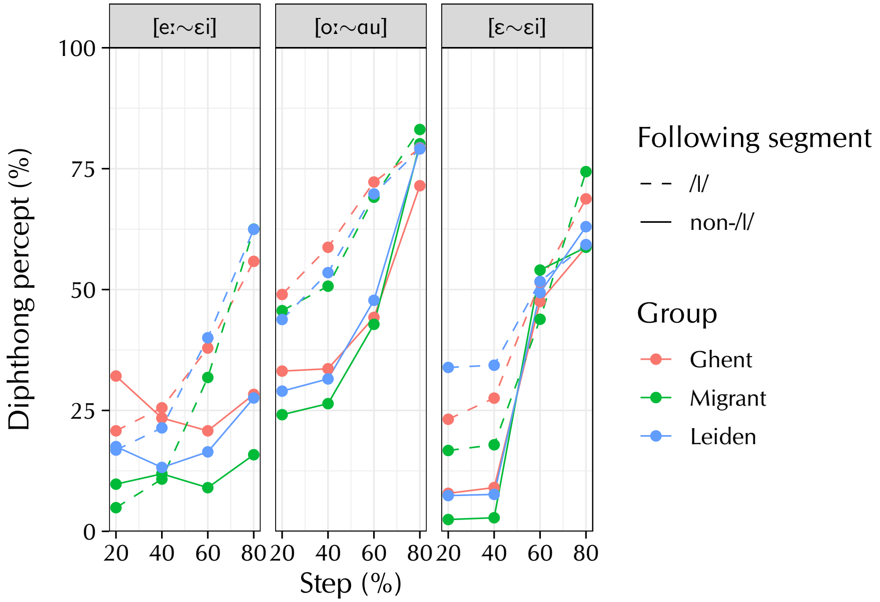

Figure 4.5: Averaged raw data from the rhyme-decision task (106 participants, 1,536 items). The general trends are that (1) participants become more likely to indicate a diphthong percept at later morphing steps; (2) this effect is larger before coda /l/ than before non-/l/, indicating that participants are perceptually compensating in the latter but not the former condition; (3) there are differences between the groups both at the baseline and as a function of the morphing step.

condition does between these participant groups. Thus, the group differences are such that the S-curve patterns visible in Figure 4.5 are slightly steeper for the Leiden participants, and even more steep for the Ghent participants. However, as the [ɛ~ɛi] condition was a control condition, this does not matter all that much: these differences must be ascribed solely to the differences in the [ɛi] phone, which were already covered in a much more meaningful way in the [eː~ɛi] model. Rather, the [ɛ~ɛi] model serves to show that a classic S-curve pattern is obtained when two arbitrary sounds are morphed together in a rhyme-decision experiment, providing additional validation of the experimental method itself.

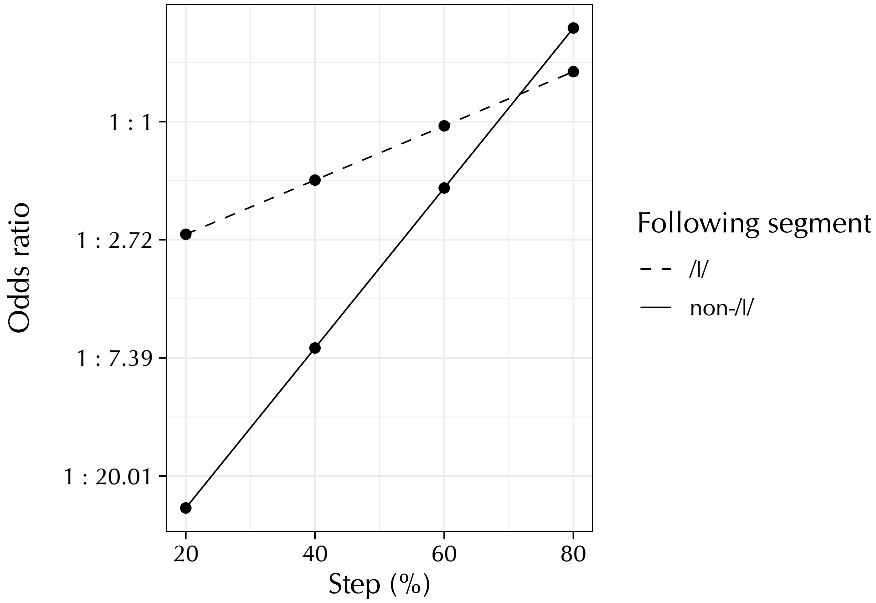

Figure 4.6: Partial-effect plot of the three-way interaction in Table 4.4. The plot shows the difference between the Leiden group and the others, in their interaction of "Step" (on the *x* axis) and "Following segment" (as separate lines), in the [ɛ∼ɛi] condition. The *y* axis is on a logarithmic scale, as this is the scale on which the partial effects in the logistic-regression analysis are linear. Observe that the Leiden group has much lower odds of reporting a diphthong percept than the other two groups in the first three steps, but at the fourth morphing step this preference reverses and the Leiden group has slightly higher odds of reporting a diphthong percept than the two other groups. Finally, note that this effect is much more pronounced, in having a steeper linear slope leading to larger group differences in the earlier steps, in the non-/l/ case than in the /l/ case.

4.3.3 Discussion

Contrary to what would be expected based on the results from Experiment 1, Experiment 2 did not reveal significant differences at the level of the individual in the cluster analysis. The group-level analysis, however, showed the expected effects of "Step" and its interaction with "Following segment = non-/l/" for all three models.

For the [eː∼ɛi] model, the results show that a following coda /l/ makes participants more likely to opt for the diphthong target as a function of the morphing step. In other words, a following coda /l/ shifts participants' perceptual category boundary further towards /eː/. This is in line with the prediction that participants allow for some diphthongization to be present in an /eː/ realization, but only when it is not followed by coda /l/. This result shows that participants perceptually compensate for some diphthongization in the speech signal, but only in the phonological context where such diphthongization is allowed, demonstrating phonotactic knowledge. A second finding was that participants' category boundaries are located at different positions. The Leiden group initially shows smaller odds of perceiving the diphthong target, but at later morphing steps catches up to the baseline, implying that the Leiden participants' category boundary lies closer to /ɛi/. This agrees with the prediction that this participant group is willing to tolerate more diphthongization in the speech signal before switching from a slightly-diphthongized-/eː/ percept to a slightly-monophthongized-/ɛi/ percept. The same is true for the migrant participants, but only in the /l/ condition, showing that these participants compensate more strongly for diphthongization in a context in which it is unexpected, mirroring what they do in production.

The [eː∼ɛi] condition does not sample the *entire* continuum of possible realizations of /eː/ and /ɛi/, due to the on-going lowering of the latter diphthong. This is reflected in the results: even in the final step of the continuum, the proportion of /ɛi/ responses is still low. Compare this to the [oː∼ɑu] condition, in which the proportion of diphthong responses is much higher in all four steps of the continuum. This condition sampled the *full* range of the diphthong phoneme, including the lowered [ɑu] realization, and for this reason reaches a much higher proportion of diphthong percepts at the final stage of the realization, which is 80% on the full [oː∼ɑu] continuum. Note that sampling a wider continuum also implies using larger step sizes along the four intermediate points: the first step of 80% [oː] morphed with 20% [ɑu] includes more diphthongization than the first step of the [eː∼ɛi] continuum would have. This is also reflected in the results: the proportion of diphthong responses at all four steps is higher for the [oː∼ɑu] condition than it is for the [eː∼ɛi] condition. The general trends of more diphthong responses at later morphing steps, and of more diphthong responses when there is a following coda

/l/, are similar between the two conditions. Note, however, that the group differences are different: the statistical analysis revealed significant group differences for the [eː∼ɛi] condition, but did not do so for the [oː∼ɑu] condition. This suggests that the group differences are robust chiefly in the former half of the monophthong–diphthong continuum. As this is the part pertaining to the diphthongization of /eː,øː,oː/, which is a much older sound change (first mentioned by Zwaardemaker & Eijkman 1924) than the lowering of the original diphthongs (of which the earliest reference in Section 4.1.4 is Gussenhoven & Broeders 1976), this result is not wholly surprising.

The [ɛ∼ɛi] model served as a control condition; here, the different effects add up to produce a classic S-curve pattern, which is expected if two arbitrary sounds are morphed together. This curve has significantly sharper edges in the migrant group. This is likely to be due to the smaller sample size of this group, which provides less opportunity for sharper edges to be smoothed out by many observations.

4.4 Link between production and perception

The results from Section 4.2.2 found significant inter-individual differences in their adoption of the sound changes in production, but the same individual variation was not found in perception, where only group-level results were found. Following Evans & Iverson (2007), however, it is possible that the individual results for the perception data are correlated with those for the production data. As explained in the Introduction, the existence of such a production–perception link is of major importance for the individual adoption and community propagation of sound change. Section 4.2.2 showed that the individual variation in production is represented well by the "Following segment = non-/l/" BLUPs; the present section investigates whether the variation in these BLUPs can be (partly) explained by the BLUPs from the individual-level analysis of the perception data.

4.4.1 Method

Running 24 correlation tests (3 models for the perception task × 8 random-effect vectors each) would be improper for reasons of multiple comparisons. However, since BLUPs are Gaussian random variables, it is possible to test each of these 24 correlations simultaneously by simply performing a linear regression analysis of these 106 × 24 data points onto the 106 BLUPs obtained from the analysis of the production experiment. Thus, a linear-regression analysis was performed with the "Following segment = non-/l/" BLUPs from the production experiment as the dependent variable, and the 24 sets of BLUPs from the rhyme-decision task as covariates. All variables included were standardized (i.e. z-transformed), so that the estimated regression coefficients are exactly equal to Pearson product-moment correlation coefficients. As comparisons *between* these 24 predictors are not of interest, neither an intercept term nor any interactions were included in the model. The R code for this analysis is available at https://figshare.com/s/731e0a32480e876530e0 as the file `correlation.R`.

4.4.2 Results

Figure 4.7 provides a visualization of the correlations that reached significance in the analysis. These are partial-effect plots, meaning that the plots show the effect for each correlation term while controlling for the other 23 terms present in the linear-regression model. The standardization has been reverted in this figure, such that the visualized correlations are on the same scale as the original BLUPs and are therefore directly interpretable as relationships between the individual differences in ΔF1 in production and the log-odds of the diphthong percept in perception. In total and after adjusting for multiple testing, the individual differences in the perception data were able to account for 34% of the variance in the individual differences in the production data.

Participants who diphthongized less strongly (higher ΔF1) in the production task were also less likely to indicate a diphthong percept in the [eː∼ɛi] condition of the perception task ($r = -.27$). Similarly, if this condition in the perception experiment contained a following coda /l/, participants became more likely to indicate a diphthong percept if they produced more diphthongization themselves ($r = -.23$). In the [oː∼ɑu] condition, this effect was reversed: participants were more likely to indicate a diphthong percept if they themselves produced *less* diphthongization ($r = .28$). Finally, in the [ɛ∼ɛi] condition, participants who diphthongized less strongly in production were quicker to perceive a diphthong as a function of the morphing step ($r = .43$), and became so even more in the /l/ context ($r = .31$).

Table 4.5: Correlations of the various random slopes for the rhyme-decision task with the "Following segment = non-/l/" random slope from the production task (n = 106). $F_{24,82}$ = 3.23, p <.001, R^2 = .49, R^2_{adj} = .34. The correlations are visualized in Figure 4.7.

Factor	Estimate (SE)	t	p	Sig.
Model = [eː∼ɛi], Intercept	.24 (.13)	1.80	.08	
Model = [eː∼ɛi], Step (Linear)	−.27 (.12)	−2.30	.02	*
Model = [eː∼ɛi], Step (Quadratic)	.02 (.09)	0.16	.87	
Model = [eː∼ɛi], Step (Cubic)	−.12 (.09)	−1.41	.16	
Model = [eː∼ɛi], /l/	−.23 (.09)	−2.49	.01	*
Model = [eː∼ɛi], Step (Linear) × /l/	−.08 (.10)	−0.77	.44	
Model = [eː∼ɛi], Step (Quadratic) × /l/	.12 (.09)	1.40	.17	
Model = [eː∼ɛi], Step (Cubic) × /l/	−.02 (.09)	−0.22	.83	
Model = [oː∼ɑu], Intercept	.13 (.10)	1.33	.19	
Model = [oː∼ɑu], Step (Linear)	.01 (.15)	0.08	.94	
Model = [oː∼ɑu], Step (Quadratic)	−.17 (.11)	−1.54	.13	
Model = [oː∼ɑu], Step (Cubic)	−.06 (.09)	−0.60	.55	
Model = [oː∼ɑu], /l/	.28 (.11)	2.53	.01	*
Model = [oː∼ɑu], Step (Linear) × /l/	.16 (.12)	1.36	.18	
Model = [oː∼ɑu], Step (Quadratic) × /l/	.01 (.09)	0.12	.91	
Model = [oː∼ɑu], Step (Cubic) ×/ l/	−.09 (.10)	−0.86	.39	
Model = [ɛ∼ɛi], Intercept	−.12 (.13)	−0.91	.36	
Model = [ɛ∼ɛi], Step (Linear)	.43 (.13)	3.31	<.01	**
Model = [ɛ∼ɛi], Step (Quadratic)	−.02 (.10)	−0.20	.84	
Model = [ɛ∼ɛi], Step (Cubic)	.02 (.10)	0.23	.82	
Model = [ɛ∼ɛi], /l/	−.09 (.11)	−0.86	.39	
Model = [ɛ∼ɛi], Step (Linear) × /l/	.31 (.11)	2.87	<.01	**
Model = [ɛ∼ɛi], Step (Quadratic) × /l/	.08 (.11)	0.73	.46	
Model = [ɛ∼ɛi], Step (Cubic) ×/ l/	.08 (.11)	0.77	.44	

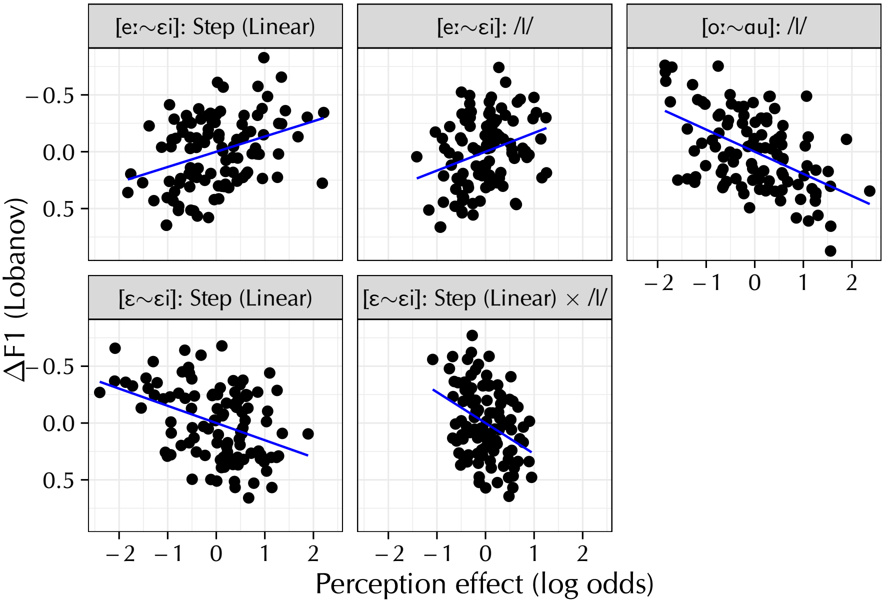

Figure 4.7: The partial correlations that reached significance in the analysis ($n = 106$), backtransformed to the original linear-predictor scales. Participants who produce more diphthongization are more likely to indicate a diphthong percept in the [eː∼ɛi] perception model at later morphing steps as well as when a coda /l/ followed the vowel. In the [oː∼ɑu] model, participants who produce more diphthongization are less likely to indicate a diphthong percept when a coda /l/ follows. Finally, in the control model [ɛ∼ɛi], participants who diphthongize more in production are less likely to indicate a diphthong percept as a function of the morphing step in both the non-/l/ and /l/ conditions.

4.4.3 Discussion

The correlations show that participants who realize vowels such as /eː/ with less diphthongization are also less likely to perceive slightly-diphthongized realizations of this vowel as realizations of /ɛi/. These participants are thus less advanced on the sound change diphthongizing [eː] to [ei]: they diphthongize less themselves, and perceptually allow for more diphthongization in the speech signal before switching their percept to the diphthongal target. For participants who are further along the sound changes, the presence of a following coda /l/ makes an important difference. In this situation, these participants have no reason to expect diphthongization based on the phonological context. This is why, at the group level, this condition resulted in significantly more diphthong percepts. For participants who are less far along on the sound change, i.e. who diphthongize less in production, the difference a following coda /l/ makes is much smaller, as these participants have no need for the phonological rule blocking diphthongization before /l/. The [eː∼ɛi] correlations also reflect this. These results corroborate the findings by Beddor et al. (2018) and Coetzee et al. (2018).

In the [oː∼ɑu] condition of the perception experiment, the latter effect reverses. Recall that this is also the condition that sampled a more complete continuum of the diphthong phoneme, and the condition in which between-groups differences were not borne out. The latter suggests that the differences between /l/ and non-/l/ observed in this condition are not driven by between-groups differences in phonological rules. If participants do not assign differential weight to the effect of a following coda /l/ (as the previous paragraph argued for the [eː∼ɛi] condition), the observed correlation follows naturally. Participants who are further along the sound changes then allow for more intrinsic diphthongization, and will thus indicate more monophthong percepts even in the presence of a following coda /l/, whereas participants who are less far along the sound changes are more likely to indicate a diphthong percept.

In the control condition [ɛ∼ɛi], participants who diphthongize less in production are *more* likely to perceive slightly-diphthongized realizations as /ɛi/. This condition is not affected by the on-going sound changes, and hence there is no reason for participants to expect any intrinsic diphthongization to be present in the monophthongal endpoint of the perception continuum. In this case, participants who produce less diphthongization also allow for less diphthongization in the speech signal before switching to an /ɛi/ percept, so this vowel's category distinctions in perception and production directly mirror one another. Similarly, in the [eː∼ɛi] condition when followed by coda /l/, participants also have no reason to expect any intrinsic diphthongization and indeed show the same behavior.

4.5 General discussion

The main goals of this chapter were to further advance our understanding of sound change by investigating in detail a contact-driven phonological change and by also taking into account variation at the individual level. At the group level, both the production and the perception results showed significant influences of the distinction between a following coda /l/ vs. another consonant as a function of the participant group, demonstrating between-groups differences in phonological knowledge. This was also borne out in a particularly clear way by the individual-level production results. While the group-level results in production simply placed the migrant group in between the two control groups, the individual-level results revealed a more nuanced picture, by showing that the migrant group was not homogeneous: *some* individuals had adapted so much that they were classified with the Leiden participants, but other individuals had not and were classified with the Ghent participants.

The results for the rhyme-decision task were quite different: at the group level, the migrant group showed a systematic shift in one of the two critical conditions (the boundary between [eː∼ɛi]) and in the control condition (the boundary between [ɛ∼ɛi]). This pattern of results suggests that the migrant participants' perception of the /ɛi/ category shifted more towards the Netherlandic system. Contrary to the results for the production data, these findings were only observed at the level of the whole group; in the variation between individual participants, no systematic patterns were observed. However, significant and meaningful relationships were found between the individual differences in perception and those in production.

The production results and their correlation with the perception results corroborate the results by Evans & Iverson (2007), and also agree with findings from the field of L2 acquisition, which show that L2 learners change their production over long periods of time, but not their perception (Flege 1987, Flege & Wayland 2019). The production results and their substantial inter-individual differences are also in line with Nycz (2013) and Evans & Iverson (2007). Although Evans & Iverson (2007) do not actually discuss it, the production results in their Table 1 (p. 3,817) show that some speakers changed their phonetic implementations to larger or smaller degrees, and some did not at all. When considered as a single group, their results show a small but systematic change across the board. The production findings from the present experiment paint exactly the same picture: some migrant individuals have changed their Flemish-Dutch vowels to conform to the Netherlandic-Dutch system, some have not, and at the group level, these individual effects are large enough to quantitatively push the whole group towards a more Netherlandic vowel system. The individual differences between perception and production fit right into the picture painted by Beddor et al. (2018) and Coetzee et al. (2018), in that participants who are more advanced in production are, generally speaking, also more advanced in perception. It additionally appeared that, while Section 4.3's results for the [eː∼ɛi] condition could be explained by differences in phonological knowledge between the participants, the results for the [oː∼ɑu] condition were driven more by phonetic expectations than by phonological knowledge. Following Baker, Archangeli, & Mielke (2011), Pinget (2015), and Pinget, Kager, & Van de Velde (2019), this is in line with the [oː∼ɑu] data reflecting a sound change that is in an earlier stage of completion, in which phonetic variation has not yet been fully encoded into a complete sound change. As this continuum incorporates not just the diphthongization of the tense mid vowels but also the much more recent lowering of the original diphthongs, this is a possibility, although the present set of experiments cannot prove this conjecture.

On the question if adoption starts in perception or in production, the results from the present chapter are in line with Evans & Iverson (2007): change in production was easily detected, change in perception was not. Specifically, while the sociolinguistic migrants as a group had shifted the category boundary of at least their /ɛi/ phoneme to be more like the Netherlandic group, it was not possible to single out a discrete set of specific individuals who were uniquely responsible for this group-level effect, although individual-level correlations between perception and production were found (which was also true for Evans & Iverson). These correspondences make it plausible that the changes in these socoiolinguistic migrants started out in production, and hence that the contact-driven phonological change studied here is wholly similar to Evans & Iverson's (2007) contact-driven phonetic change. These results are compatible with the observation by Pinget (2015) and Pinget, Kager, & Van de Velde (2019) that sound changes become production-driven when they have almost come to completion. This follows from the idea that sociolinguistic migrants are comparable to individuals who have remained conservative while their environment has adopted a novel sound change.

4.6 Conclusion

The present chapter investigated the role of the individual in the adoption of sound change. The focus of investigation was sociolinguistic migrants, in this case Flemish-Dutch speakers who moved to the Netherlands multiple years to decades ago. The results are partially in line with the findings by Evans & Iverson (2007). On the one hand, in agreement with Evans & Iverson, group-level adoption of the sound changes was found in production; specifically, the group as a whole had undergone a quantitative shift to be more Netherlandic-Dutch like, and ten of the eighteen participants had changed to such a degree that, in a cluster analysis, they were classified as having become qualitatively Netherlandic. The present study also found similar effects in perception in a group-level analysis, but the same effects were not borne out in an individual-level analysis of the perception data. However, individual-level correlations were found between perception and production. These results are in line with previous findings on individual differences by Beddor et al. (2018) and Coetzee et al. (2018). They also fall in line with findings on the individual level by Baker, Archangeli, & Mielke (2011), Pinget (2015), and Pinget, Kager, & Van de Velde (2019) inasmuch as they suggest that younger sound changes are more reliant on superficial phonetic variation than on structural phonological variation. Taken together, the results from the present study contribute to our knowledge of phonological change, and also provide another demonstration how individual differences can provide a richer view of the adoption of sound change than could have been obtained by considering only patterns at the level of the group, precisely as Stevens & Harrington (2014) had foreshadowed.

The present study is not without its limitations, of which I highlight one which could inspire future research. The migrant group of participants was quite small ($n = 18$), which limited the individual-level analyses reported in this chapter. While the results were very clear for purposes of the present chapter, in showing that the productions of the migrant group could be classified into the expected two groups with sufficient statistical power, the migrant group was too small to determine what factors drove this classification in the first place. For instance, do participants' degrees of adoption correlate with the amount of time they have lived in the Netherlands? If it does, does it do so still after controlling for participant age—in other words, do the participants adopt lifespan changes or are these instances of age-grading (Wagner 2012)? To further tease apart these two possibilities, it would be particularly fortuitous if future research recruited control participants matched in age to the migrant group. However, such evidence could also be gathered from other sources, such as the Dutch teacher corpus (van Hout et al. 1999), which maps the regional variation in the sound changes discussed in this chapter in great detail, and in which age was explicitly taken into account as a variable during the data collection. The combination and integration of these different sources of knowledge into a single larger picture of the on-going Dutch sound changes would be a welcome continuation of the research presented in this chapter.

Acknowledgments

I am grateful to my dissertation supervisors, Claartje Levelt and Yiya Chen, for their feedback and discussions from the set-up of the experiments to the final version of this chapter. I also thank three reviewers of *Language & Speech* as well as its editors for their insightful and comprehensive comments. I am indebted to Katja Lubina for her enthusiastic help in recruiting the migrant participants, and to Sjef Barbiers, who also got me in touch with a number of suitable individuals. I wish to thank Rob Hartsuiker and Wouter Broos for allowing me to use the lab facilities at Ghent University, and I thank Aliza Glasbergen-Plas for her willingness to record the stimuli used in the rhyme-decision task. Finally, of course, I extend my gratitude to the 106 participants who were willing to take part in the experiments reported here.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is part of the research programme *Watching Dutch Change* with project number PGW-15-15, which is (partly) financed by the Dutch Research Council (NWO).

Declaration of Conflicting Interest

The Author(s) declare(s) that there is no conflict of interest.

CHAPTER 5

Noticing the change: misrepresentation, not misperception, of allophonic variants in sound change

This chapter has been submitted.

Abstract

Linguists have posited that phonological change arises through misperception. This is evaluated using a longitudinal EEG experiment. Three on-going changes in Dutch are studied: the diphthongization of /eː,øː,oː/ to [ei,øy,ou], the blocking of diphthongs before coda /l/, and the gliding of coda /r/ to [ɹ]. These changes have essentially completed in the Netherlands, but have not taken place in Flanders, the Dutch-speaking part of Belgium. A passive-oddball task was performed with Flemish-Dutch speakers (plus Netherlandic controls) who migrated to the Netherlands to start their university studies. Previous work has shown that such sociolinguistic migrants readily adopt the local phonology, and hence on-going changes. Participants did the experiment four months after arrival and again four months later. Results show that, initially, the Flemish participants have stronger mismatch negativities to a deviant [ei] within a stream of standard [eː]s, but four months later this difference has disappeared: they have learned the vowel shift. The gliding of /r/ continues to elude them: they show an MMN, but with a less frontal topographical distribution—they find the glide less salient. This is interpreted as showing successful phonological learning, but not yet sociolinguistic learning. The results argue against misperception, favoring misrepresentation instead.

5.1 Introduction

It was proposed by Ohala (1981) that sound change originates when a listener under- or overapplies rules compensating for coarticulation during the transmission of the speech signal. In this scenario, the listener has processed the incoming speech signal incorrectly, either by phonologically encoding accidental details ("intrinsic variation", Wang & Fillmore 1961) or by failing to encode information that was linguistically relevant ("extrinsic variation", ibid.). Thus, while the listener hears the speech signal without problem, they make an error in *processing* it. This type of categorization error is commonly referred to as "misperception", and is said to lead to sound change if the listener subsequently adjusts their phonological representation to match (Hyman 1976). Theoretical analyses of historical sound changes have found (indirect) support for this mechanism of sound change, claiming it the most likely scenario in various case studies (e.g. the infamous [k]>[ʧ]>[s] change in Proto-Romance, giving Latin *caelum*, Italian *cielo*, and French *ciel*; Guion 1998).

An alternative account by Hamann (2009) suggests that these types of sound changes do not in fact take place in the listener's processing of the speech signal, but rather in the grammar they use to perform this processing with. In this view, the speaker and the listener have acquired slightly different mappings of phonetic cues to phonological categories, and as a result the listener understands a different phonological category than the speaker intended, because they attach different cue weights to the same auditory information. A similar proposal was made by Beddor (2009), which differs from Hamann (2009) only in not requiring that the grammatical innovation take place in childhood. In these accounts, the critical difference between speaker and listener is not located in the listener's *perception* of the speech signal, but rather in their grammatical *representation* of the relevant phonological and phonetic features.

The present chapter aims to contribute to the debate surrounding these two alternatives—broadly speaking, the misperception account by Ohala (1981) and others versus the misrepresentation account by Hamann (2009) and Beddor (2009)—using neurolinguistic evidence. The chapter draws strongly on Grosvald & Corina (2012), who used a mismatch-negativity paradigm to show that listeners are able to perceive and *encode* sound change—in their case, long-distance vowel-to-vowel coarticulation of up to three syllables away—automatically. Grosvald & Corina (2012) demonstrated this encoding using a mismatch-negativity ("MMN") experiment, which is also the paradigm used in the present paper. MMN studies have a long history of use in phonology (see e.g. Cornell, Lahiri, & Eulitz 2011, Grosvald & Corina 2012, Hestvik & Durvasula 2016, Lahiri & Reetz 2010, Lanwermeyer et al. 2016, Mitterer & Blomert 2003, Scharinger & Lahiri 2010; for accessible introductions to the

MMN in general, see Näätänen 1990, Näätänen & Alho 1997, Sussman et al. 2014). Generally speaking, the MMN is an event-related potential evoked in EEG experiments using the passive-oddball task. In these experiments, participants listen to a stream of "standard" sounds, which is sometimes interrupted by a "deviant" sound. If the standard has a neurally-encoded feature that the deviant does not, this difference triggers an MMN (but the reverse difference does not; Cornell, Lahiri, & Eulitz 2011, Lahiri & Reetz 2010). To disentangle phonological encoding from phonetic encoding (i.e. to ensure that the MMN probes a phonological feature rather than the obvious acoustic differences between standards and deviants), MMN studies typically use multiple, different, tokens of the same surface allophone, called the "varying standards" paradigm (Hestvik & Durvasula 2016). In this paradigm, the presence of an MMN reflects the phonological encoding of the standard stimulus, with the size of the MMN in microvolts proportional to the phonological distance between the standards and the deviant (Näätänen et al. 2007).

MMNs have been used to study sound change, ranging from phonemic mergers to allophonic shifts. Lanwermeyer et al. (2016) show that a phonemic merger resulting in lexical confusion elicits an MMN that is much earlier (100–200 ms) than the allophonic MMN found by Grosvald & Corina's (2012) (275–325 ms). In addition, they find a P600, which reflects the semantic reintegration and reevaluation of an initially misanalyzed phoneme category (see also Kung, Chwilla, & Schriefers 2014 and Chapter 6). However, in a contrasting condition where only allophonic differences were manipulated (similar to Grosvald & Corina 2012), the MMN was reduced and temporally shifted and the P600 disappeared. The absence of the P600 in this condition is not surprising, as an allophonic difference cannot result in lexical overlap, and hence no reanalyis was necessary. Similarly, the reduction of the MMN is logically explained as allophonic switches being less salient than phonemic switches, as is argued by Kazanina, Phillips, & Idsardi (2006), who failed to find an allophonic MMN. However, these authors presented their allophonic condition (Korean [t,d]; these are allophones of the same phoneme, with /t/ becoming [d] intervocalically) without providing the requisite phonological context (both variants were presented word-initially, which does not trigger the allophony). This may explain why they did not find an allophonic MMN, whereas other studies (Jacobsen 2015, Lanwermeyer et al. 2016, Steinberg, Truckenbrodt, & Jacobsen 2010a, 2010b, 2011) did. The temporal shifting of the allophonic MMN observed by Lanwermeyer et al. (2016) brings it exactly in line with the window where Grosvald & Corina (2012) found their strongest effect (Lanwermeyer et al.: 250–350 ms, Grosvald & Corina: 275–325 ms), which inspires confidence that the allophonic MMN is indeed later than the phonemic MMN. Note that the mentioned allophonic MMNs are really responses to the phonological allophone, and do not simply reflect acoustic differences in the phonetic signal: both Grosvald & Corina (2012) and Lanwermeyer et al. (2016) used the varying-standards paradigm.

More specific than research into allophonic variation is research into allophonic *violations*. In this strand of research, one does not investigate allophonic differences in realization *per se*, as in Kazanina, Phillips, & Idsardi (2006) and Lanwermeyer et al. (2016), but rather the grammatical knowledge that is a prerequisite for processing such differences. An example, and the phenomenon studied in this chapter, is the rise of a new allophonic rule due to phonotactically-conditioned regular sound change. Previous studies (Jacobsen 2015, Steinberg, Truckenbrodt, & Jacobsen 2010a, 2010b, 2011) have shown that phonotactic violations result in MMNs. However, these papers are about violations of well-established allophonic rules in the standard variety of a language. It might be the case that novel allophonic rules involved in on-going sound change are less salient (and hence encoded less strongly) than well-established allophonic rules of the type studied in Jacobsen (2015) and Steinberg, Truckenbrodt, & Jacobsen (2010a, 2010b, 2011). Hence, the present study integrates and extends the aforementioned findings by studying a currently-on-going sound change that involves the genesis of a new allophone distinction. The study uses a combined cross-sectional and longitudinal design, aimed at providing a window into the processing of sound change as it unfolds in real time.

The language used for the investigation is Dutch, in which the tense mid vowels /eː,øː,oː/ have changed into upgliding diphthongs [ei,øy,ou]. This regular sound change is blocked before coda /l/, leading to novel allophone pairs: monophthongs before coda /l/, diphthongs elsewhere. Independently of these changes, Dutch has also undergone an allophone split in the rhotic, changing /r/[1] to [ɹ] in coda position. These three distributional changes are regionally stratified. They have all but completed in the Netherlands, but the Dutch spoken in Flanders (the northern part of Belgium) has not undergone them, leading to salient sociolinguistic differences between Netherlandic Dutch and Flemish Dutch (Sebregts 2015, Van de Velde 1996). This is particularly true for the rhotic, which is perhaps the most-well-known sociolinguistic variable in the Netherlands and Flanders (Sebregts 2015). Table 5.1 provides a complete overview of the relevant allophonic rules in both varieties. The present study

---

[1] The phonetic implementation of /r/ is highly variable between different regions of Dutch (see Sebregts 2015 for a thorough overview), including alveolar as well as uvular trills and fricatives. However, the phonological allophone split between onset and coda variants is restricted to Netherlandic Dutch, and is also only implemented by means of the [ɹ] realization; in addition, this realization can never occur in onset position in either Netherlandic Dutch or Flemish Dutch. The phonetic implementation of the onset allophone in this experiment was the alveolar trill, as this is the standard variant in Flemish Dutch and is one of the major variants in Netherlandic Dutch, and shares its place of articulation with the [ɹ] allophone, which helps keep the difference between standards and deviants to the minimum required.

Table 5.1: The relevant allophonic rules involved in the on-going sound changes and their regional differences.

Underlying form	Netherlandic realization	Flemish realization
/eː/ followed by coda /l/	[eː]	[eː]
/eː/ elsewhere	[ei]	[eː]
/ər/	[əɹ]	[ər]

uses an MMN experiment to investigate the processing of these allophonic rules in two populations: a control group of Netherlandic students, and an experimental group of Flemish students in their first year of study at a university in the Netherlands. It is expected, and has been shown experimentally in similar research (Evans & Iverson 2007), that this experimental group will adapt to the Netherlandic realizations as part of the normal process of accent accommodation, paralleling the adoption of these historical sound changes. To investigate such adaptation over time, the cross-sectional comparison is performed two times, with four months in between.

The task is an oddball task, the same task used by Grosvald & Corina (2012) and Lanwermeyer et al. (2016). The two accounts of sound change under discussion make different predictions on the degree of encoding of the allophones in question, and hence on the degree to which the oddball task should yield MMN ERPs. Under the misperception account, the prediction would be that the Flemish participants will not grammatically encode the difference between the allophones for each of [eːɫ∼eiɫ], [eː∼ei], and [ər∼əɹ], as these differences are not relevant in their own grammars and hence fall under the purview of intrinsic variation. This would preclude MMN effects from showing up. In turn, the misrepresentation account predicts that the Flemish participants do *encode* the allophonic distinctions, but subsequently *evaluate* them in a different way (e.g. through the P600, as observed by Lanwermeyer et al. 2016 and Chapter 6; a separate experimental paradigm would be required to evaluate this possible mechanism). In this case, the participants will perceive a mismatch between the deviant and the standards on an extrinsic property, which is visible as the MMN. A second prediction for the present experiment, which holds for both accounts of historical sound change equally, is that in the four months between the two sessions of the experiment, the Flemish participants have begun to adopt the Netherlandic system, such that the differences between the groups will have become smaller. There is evidence for this type of post-adolescent adjustment of the perception of vowel categories from both sociolinguistics (e.g. Bowie 2000, Evans & Iverson 2007, Nycz 2011, Ziliak 2012, Chapter 4) and the related field of second-language acquisition (Flege 1987, Flege & Wayland

2019, *inter alia*); it seems reasonable to hypothesize that those findings reflect a general ability that is also relevant here.

The pool of participants suitable for this experiment is small, because the experiment calls for a specific and special population: participants in the experimental group must be Flemish, must have migrated to the Netherlands post-adolescence (and not before), must be measured relatively shortly after arrival (no later than a couple of months; compare Chapter 3 and Chapter 4), and must be willing to commit to a two-part experiment with some time in between. Because the GDPR was not yet in effect when recruitment for this experiment was initiated, it was possible to obtain a list of recently-arrived Flemish individuals who had just begun their studies at two universities in the Netherlands: Leiden University and the University of Amsterdam. This made it possible to recruit eight participants, resulting in fourteen obtained repeated-measures datasets. Both are typical sample sizes for similar sociolinguistic studies on the adoption of phonetic variation along the lifespan (e.g. Alshangiti & Evans 2011, Bauer 1985, Carter 2007, Cedergren 1987, Chambers 1992, De Decker 2006, Evans & Iverson 2007, Harrington 2006, Harrington, Palethorpe, & Watson 2000, Hinton 2015, Nahkola & Saanilahti 2004, Nycz 2011, Nycz 2013, van Oostendorp 2008, Prince 1987, Sankoff 2004, Sankoff & Blondeau 2007, Sankoff, Blondeau, & Charity 2001, Trudgill 1988, Wagner 2008, Yaeger-Dror 1994, Ziliak 2012). However, small sample sizes raise concerns about the power of the experiment. This issue of power is explicitly taken into account in the present chapter by using appropriate statistical methods, particularly the generalized additive mixed model ("GAMM"; Wood 2017). Contrary to ANOVA, GAMMs do not require data from experimental trials to be averaged over both the time and space dimensions, thus achieving more precision. At the same time, the GAMM analyses used in this chapter make it possible to model the topographical distribution of the MMN (as was the focus in Grosvald & Corina 2012) without incurring the "curse of dimensionality", by not requiring that electrodes be coded using many-leveled factors for hemisphere and anteriority, thus permitting parsimonious models. The Bayesian approach adopted in this paper, which is explained below, provides a natural measure of the power of the analysis, by defining power as the degree to which the experimental goal of rejecting the null or the alternative hypothesis was reached (Kruschke & Liddell 2018). The Bayes factors used in this chapter provide a direct measure of the degree to which this goal was attained, and hence whether the sample size was sufficient to detect the presence or absence of group differences.

5.2 Method

5.2.1 Participants

Participants were eight Flemish-Dutch first-year students in the Netherlands and nine Netherlandic-Dutch controls. The participants were measured in two sessions: one approximately four months after the start of the academic year, and once again approximately four months later; with the exception of two Flemish students, all participants took part in the second session. This yielded 14 datasets for the Flemish-Dutch students and 18 datasets for the Netherlandic-Dutch students.

The experiments followed the Ethics Code for linguistic research in the faculty of Humanities at Leiden University, which approved its implementation. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

5.2.2 Stimuli

Stimuli were realizations of [eː] versus [ei], [eiɫ] versus [eːɫ], and [ər] versus [əɹ]. The stimuli were produced in a carrier word and sentence by a trained phonetician. Of all stimuli, five different tokens were selected. Praat (Boersma & Weenink 2016) was used to extract the relevant segment(s), to equalize all F0s to the average of all tokens used, to equalize all amplitudes to 60 dB SPL, and to fix the durations of the stimuli [eː,ei,ər,əɹ] at 200 ms and those of the stimuli [eːɫ,eiɫ] at 300 ms. For each of the six stimulus types, all five tokens were included as varying standards (68 presentations each, constituting 85% of the experiment when taken together) and one token was included as the deviant (60 presentations, or 15%. This resulted in a total of six experimental conditions, which are summarized in Table 5.2. In the remainder of the chapter, these conditions will be referred to by the corresponding deviant stimulus, such that "the [əɹ] condition" is the condition where [əɹ] was the deviant and [ər] were five varying standards. As an illustration of the stimuli, Figure 5.1 shows waveforms, spectrograms, and F3 trajectories[2] of the stimuli used in this condition; note how the [əɹ] stimulus has a much lower F3 than the others.

---

[2] The jittery F3 track in the trilled part of the [r] is not erroneous; this is inherent to the nature of this trill.

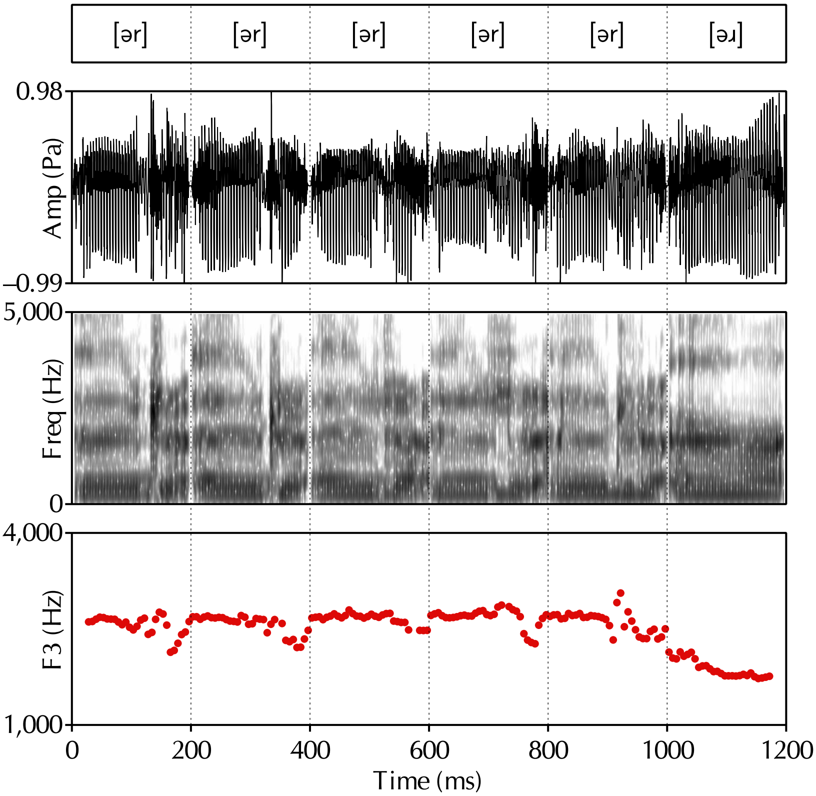

Figure 5.1: Example waveforms, spectrograms, and F3 trajectories (the critical difference between the two types of rhotic realization) of all five tokens of [ər] used as standards and one token of [əɹ] used as deviant, which together make up Table 5.2's [əɹ] condition.

Table 5.2: Design of the six conditions used in the experiment.

Standard (68 × 5 tokens)	**Deviant (60 × 1 token)**
[eː]	[ei]
[ei]	[eː]
[eiɫ]	[eːɫ]
[eːɫ]	[eiɫ]
[ər]	[əɹ]
[əɹ]	[ər]

5.2.3 Procedure

The experiment took place in a dimly-lit sound-attenuated booth. Participants were seated in a chair in the center of the room in front of a computer monitor, which was located behind an electrically-shielding glass pane. Two loudspeaker boxes were placed in the corners of the room at a distance of approximately 80 cm from the participant. During the experiment, the computer monitor was used to display a silent movie, so as to occupy the participant's attention. The sound stimuli corresponding to the six experimental conditions were presented over the loudspeaker boxes at a volume that was comfortable to the participant. The experiment was administered using PsychoPy (Peirce 2007) on a PC running Windows 7. EEG activity was recorded using a BioSemi ActiveTwo system with a sampling rate of 512 Hz. 32 AgCl electrodes were used, arranged according to the 20/10 system. Six additional electrodes recorded the left and right mastoids and the left and right horizontal and vertical extra-oculograms. Raw data were collected and were referenced off-line to the linked mastoids. Previous research has shown that this is the optimal reference choice for the auditory MMN (Mahajan, Peter, & Sharma 2017), because this is where the MMN effect achieves the highest signal-to-noise ratio even when it is small in magnitude (Kujala, Tervaniemi, & Schröger 2007, Mahajan, Peter, & Sharma 2017, Picton et al. 2000).

Before the start of the experiment, participants were instructed by the researcher to try to sit still and to try to keep blinking to a minimum. When the researcher started the experiment, an audio file spoken by a male speaker of Netherlandic Standard Dutch was played, which provided the participant with instructions. Participants were instructed to focus their attention on the silent movie and to ignore the auditory stimuli, and again to try to sit still and keep blinking to a minimum. A transcript of the spoken instructions was also shown on the screen. Participants could initiate the experiment of their own accord by pressing any one of two buttons located on either armrest of the chair. The six conditions were then presented to the participants in pseudorandomized order. There were no breaks in the experiment, which lasted exactly 28 minutes.

5.2.4 Data analysis

The raw EEG data were processed using R (R Core Team 2020) package `eegUtils` (Craddock 2019) by detrending the referenced data and applying a band-pass filter with a low cut-off of 1 Hz and a high cut-off of 30 Hz. Data were epoched from −100 ms to 450 ms post-stimulus-onset, where the first 100 ms served as baseline. Eyeblinks were removed from the epochs using least-squares regression. Trials contaminated by artifacts were rejected automatically. The data analysis focused on the six sounds used as deviants, compared to when these same six sounds were used as one of five varying standards. As such, trials of standards that were not also used as deviants were removed from the data.

The temporal window for the analysis was set at 275–325 ms post-stimulus-onset. This is the same window that was used by Grosvald & Corina (2012), and a narrower version of the 250–350-ms window used by Lanwermeyer et al. (2016). Other 50-ms windows were also investigated, but results from other possible MMN windows (e.g. 175–225 ms) were qualitatively similar enough that arbitrarily selecting a different window from the established 275–325 ms was not warranted. The data were averaged over time within this interval. Following the approach by Grosvald & Corina (2012), the data were not subsequently averaged over a specific region of electrodes, but the 32 electrodes were instead explicitly involved in the analysis. Compared to Grosvald & Corina (2012), a slightly more modern approach is used by modeling the electrodes as measurement sites on the surface of a 3D sphere. This removes the need to fit complex interaction terms of the "Hemisphere × Anteriority × Electrode" type, while retaining their advantages of specifying a precise model that achieves sufficient statistical power despite the relatively small sample size.

The statistical analysis was implemented by means of generalized additive mixed models, using function `gam` from R package `mgcv` (Wood 2017). The EEG amplitude was modeled using a smooth spline of the 32 EEG electrodes, which were mapped to a sphere based on their latitude and longitude coordinates using at most fifteen basis functions. This "spline on the sphere" informs the model that the data sampled from nearby electrodes are correlated to one another in a way that corresponds to data collected from the surface of a sphere. Difference smooths were included by the factors "Group" (coded such that Netherlandic = 0 and Flemish = 1), "Session" (coded such that the first session = 0 and the second session = 1), and "Deviant" (coded such that the sound used as standard = 0 and the sound used as deviant = 1) and all interactions. Random smooths by participants were added for the reference condition and by the factors "Deviant", "Session", and "Deviant × Session". Thus, each model contains a reference smooth, which models the topographical distribution of the EEG activity of the Netherlandic listeners, in session 1, presented with standards. Separate terms then model the difference in activity between this reference condition and the various factors manipulated in the experiment. Separate models were run for each of the six vowels. All models were fitted to scaled-t errors.

To test for possible asymmetrical effects, significance was not established via p-values but rather using Bayes factors. These make it possible to argue not just that an MMN is *present*, but also that it is *absent*, which is expected if the MMNs to be obtained are indeed asymmetrical (Cornell, Lahiri, & Eulitz 2011, Lahiri & Reetz 2010). For each of the eight smooth terms present in the model, a model with this term removed was fitted using maximum likelihood. The BIC (Schwarz 1978) of this model was compared to that of the full model (refitted using maximum likelihood). The difference between the two BICs was converted into a Bayes factor using equation (5.1), which is due to Wagenmakers (2007).

$$\mathrm{BF}_{10} = \exp\left(-\frac{1}{2}(\mathrm{BIC}_{\text{full model}} - \mathrm{BIC}_{\text{reduced model}})\right) \tag{5.1}$$

Bayes factors larger than 1 indicate support for the alternative hypothesis (the full model providing a better fit than the reduced model) and values smaller than 1 indicate support support for the null hypothesis. Section 5.3 reports these on the $\log_{10}$ scale instead, in which case the interpretation is symmetrical around zero: a $\log_{10}$ Bayes factor of zero indicates no support, positive values indicate support for the alternative hypothesis, and negative values indicate support for the null hypothesis.

5.3 Results

Table 5.3 shows the $\log_{10}$ Bayes factors corresponding to the the terms included in the statistical analyses. These are considered to be significant if their magnitude exceeds 0.5; this corresponds to what Jeffreys (1961) calls "substantial" evidence. Bayes factors with smaller magnitudes than this criterion indicate that there was insufficient evidence to be confident in a conclusion; this is indicative of an insufficiency in statistical power (Kruschke & Liddell 2018). Within each vowel, the critical effect is the difference between the vowel used as standard and the same vowel used as deviant, encoded by the factor "Deviant" and its interactions with the other factors in the design. Hence, of interest for the hypotheses are the effects for "Deviant", "Deviant × Group", and "Deviant ×

Factor	Condition [ei]	 [eː]	 [eː˧˨]	 [ei˧˨]	 [əɹ]	 [ər]
Reference smooth	0.88	1.34	−0.01	2.75	0.34	31.28
Deviant	0.58	−0.26	−0.11	−0.10	34.70	24.46
Group	0.01	−0.08	2.94	7.09	−1.71	0.17
Session	−0.07	15.61	3.42	7.41	3.72	−1.18
Deviant × Group	0.52	2.05	−0.67	−1.42	20.73	−12.30
Deviant × Session	11.18	−0.41	4.31	−0.05	5.94	−0.39
Group × Session	−0.04	0.00	−0.04	0.92	3.16	−12.29
Deviant × Group × Session	0.53	15.14	−0.42	0.43	−0.07	−3.56

Table 5.3: Results of the statistical analyses, reported as Bayes factors on the $\log_{10}$ scale.

Group × Session". Where these terms' Bayes factors provide significant support for the alternative hypothesis, Figure 5.2 visualizes the marginal effect mapped onto a stereographic projection of the head. The topographical plots in this figure thus directly correspond to the significant differences in EEG amplitude across the scalp.

In the [ei] condition, where [ei] is the deviant and varying [eː]s are the standards, there is substantial evidence for a difference between the [ei] used as standards versus used as deviant, i.e. an MMN. Figure 5.2 shows that this corresponds to a very small MMN, which reaches a minimum of −0.21 µV at Fp1/Fp2 (compared to a maximum of +0.01 µV near PO3). There is substantial evidence that this MMN differs for the Flemish students: their MMN is more negative by −0.67 µV frontally (to −0.47 µV near PO3). There is also substantial evidence that this between-groups difference changes over the sessions: the decrease in Session 1 of the experiment is counteracted by at least +0.53 µV near F7 and at most +1.01 µV around PO4, bringing their MMN in line with that of the Netherlandic controls.

In the reverse condition, with [eː] as the deviant and [ei] as standards, there is insufficient evidence to warrant claims about differences between [eː] used as standard versus as deviant. Following Kruschke & Liddell (2018), this can be rephrased as a lack of statistical power. There is, however, "decisive" (Jeffreys 1961) evidence for a different response to the deviants by the Flemish students, as well as decisive evidence that this difference changes over the two sessions. The second row of plots in Figure 5.2 shows that the Flemish students have a less negative MMN ERP to the [eː] deviants than the Netherlandic controls, by as much as +0.83 µV around Fz. In contrast, in the second session they have attained a strong MMN, which differs from the first session's MMN by

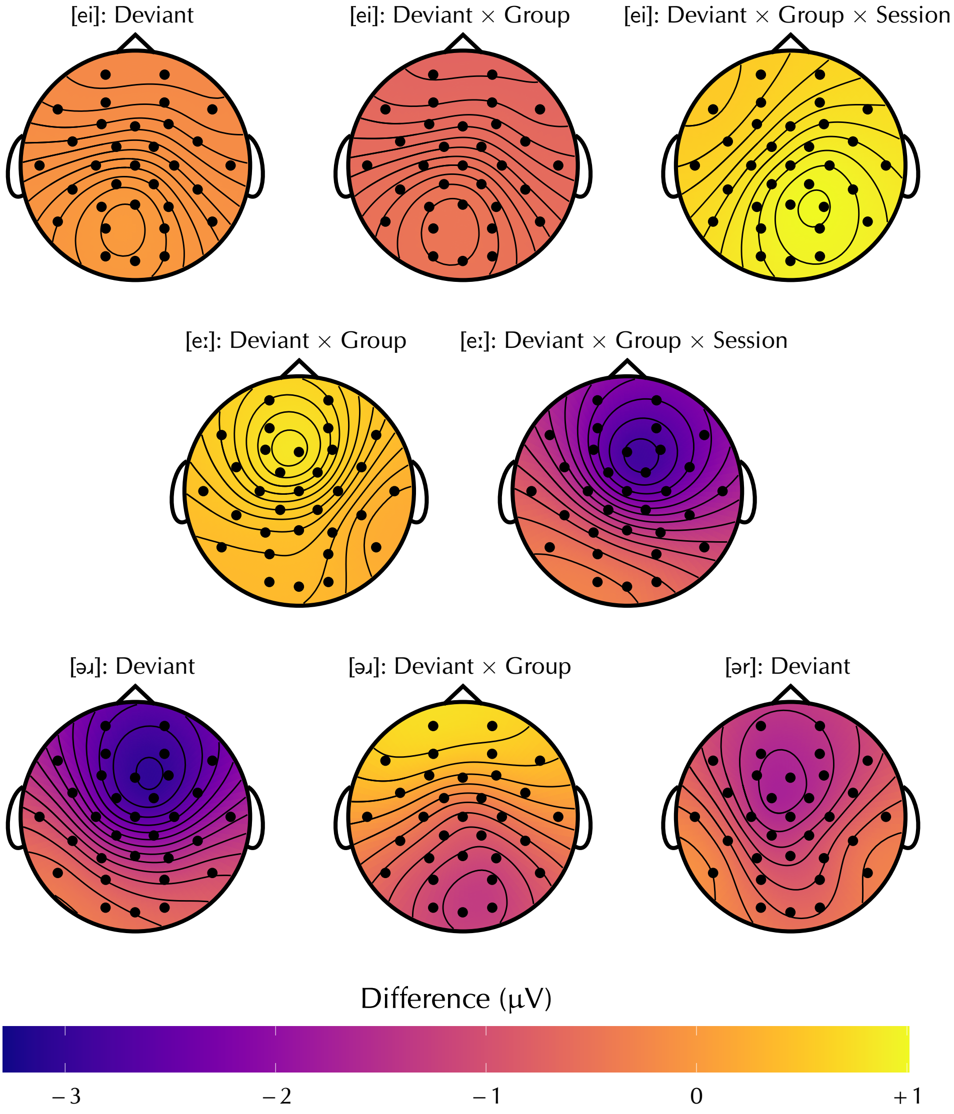

Figure 5.2: Topographical plots of the marginal effects of interest whose Bayes factors indicated at least substantial support for the alternative hypothesis. The baseline is the sounds used as standards, heard by the Netherlandic controls, in the first session of the experiment.

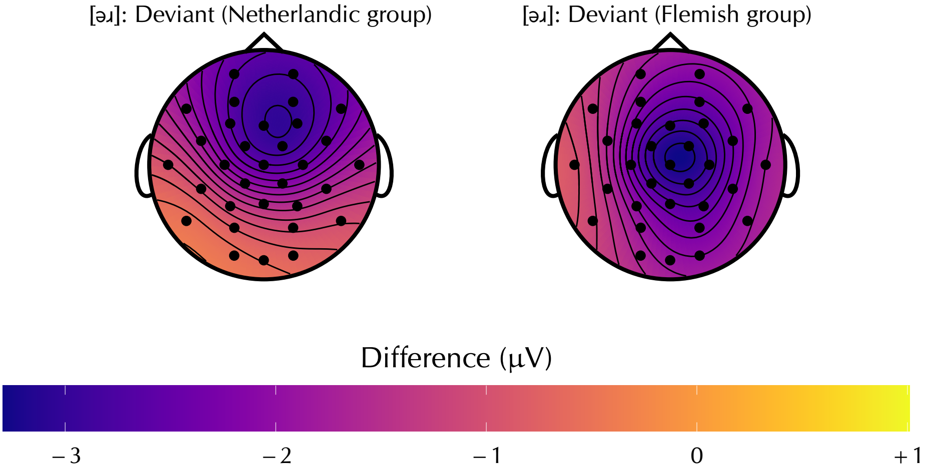

Figure 5.3: Side-by-side comparison of the "[əɹ]: Deviant" effect (left) and the sum of this effect and the "[əɹ]: Deviant × Group" effect (i.e. the MMN difference between the baseline Netherlandic and the contrasting Flemish group; right). The magnitudes of the MMNs are very similar between the groups, but the Flemish group has the effect shifted significantly towards the midpoint of the scalp.

up to −2.88 μV near Fz.

Finally, in the [əɹ] condition, decisive evidence is found for an MMN response to the [əɹ] deviants. The effect reaches a magnitude of up to −3.04 μV between Fz and F4. The evidence for a between-groups difference in this MMN is also decisive, such that the Flemish group's MMN is significantly less pronounced at the frontal pole (with a maximal difference of +0.76 μV) and more negative near occipital sites (by up to −1.31 μV). When this effect is added on top of the main effect for "Deviant" (see Figure 5.3), the result is an MMN similar in magnitude to the one for the Netherlandic group (with a largest negativity of −3.29 μV), but shifted away from the frontal pole and closer towards the center of the head for the Flemish group. No evidence was found that this shifting of the MMN in the Flemish group changed over the two sessions. In the reverse condition, where [əɹ] formed the standards and [ər] was the deviant, there is again decisive evidence for an MMN response to the [ər] deviants, but this MMN is smaller in magnitude (−1.70 μV around Fz). There is decisive evidence that there are no differences between the groups in this MMN, and that this did not change for the Flemish group over the two sessions.

5.4 Discussion

The results from the present oddball experiment show differences in aspects of phonotactic knowledge of Netherlandic Dutch between the Flemish experimental group and the Netherlandic control group. The sound change diphthongizing [eː] to [ei] has left its mark, in that both the Flemish students and the Netherlandic students exhibit a small MMN ERP when the monophthongal allophone is changed to a diphthongal one. This MMN is much larger in the Flemish group, where it reaches a peak negativity of −0.90 μV. The sizes of these MMNs are on the same order of magnitude as the one reported by Lanwermeyer et al. (2016) in their allophonic condition, making the results credible as reflections of allophonic knowledge related to sound change. The [ei] condition additionally shows that the Flemish students are learning over the course of their stay in the Netherlands: approximately four months after the first session, their MMN to the [ei] allophone has reduced in size, bringing it to the same small level as the Netherlandic controls. The reverse condition, where [eː] is the deviant amidst [ei] standards, shows a similar learning effect: in the first session, the Flemish group has a significantly attenuated MMN response compared to the Netherlandic controls, but in the second session they attain a strong MMN at the expected topographical location.

The Flemish students' increased MMN to the [ei] realization in the first session of the experiment shows that, already in the first session, this difference is encoded by the Flemish students, and is in fact represented more strongly than it is by the Netherlandic controls. After the Flemish students have spent more time in the Netherlands, and have become more used to the diphthongal allophone, they are observed to attenuate their MMN response, coming in line with the Netherlandic controls. The same is observed in the reverse condition, where [eː] is the deviant. Here, the Flemish students' MMN is attenuated in the first session, indicating that they do not find the [eː] realization as salient as the Netherlandic controls do, but they reverse this difference in the second session of the experiment.

The results for the [eː∼ei] allophones provide evidence that sets apart the misperception and misrepresentation accounts of sound change. The pattern of results for the [ei] allophone is incompatible with misperception: not only did the Flemish participants perceive the difference at all, they encoded it even more strongly than the Netherlandic controls did, already in the first session of the experiment. The former result can be explained by either account, but the latter cannot be explained by making reference only to sound perception. The misrepresentation account has no problem with this result, and might speculatively attribute the stronger response in the first session of the experiment to an on-going learning effect (analogous to that found in second-language and second-dialect acquisition). The return of this between-groups difference to the baseline, Netherlandic, level in the second session is fully in line with the second prediction made in the Introduction: by the second session of the experiment, the Flemish participants have acquired the Netherlandic pattern. The result for the [eː] allophone can be explained by either theory. The Flemish participants find the switch from [ei] to [eː] less noteworthy in the first session, but have gotten more attentive to it by the second session; under the misperception account, this is because in the first session, they have not yet learned to neurally encode the difference between these stimuli as strongly as the controls, but by the second session they have. On the other hand, under the misrepresentation account, the reason is that after they correctly perceive the [eː] sound, they impart less sociolinguistic salience to this switch, which by the second session of the experiment they have managed to acquire. The results for the rhotic, described further down, will lend more credence to the latter interpretation.

The results for the [eːɫ] and [eiɫ] realizations are quite different from those for their single-vowel counterparts. No MMN-related effects were found, and for the most important term "Deviant × Group", there was (very) strong evidence that there was no difference between the groups. This result is surprising, given the positive findings in the single-vowel conditions and the findings by Jacobsen (2015) and Steinberg, Truckenbrodt, & Jacobsen (2010a, 2010b, 2011), who also used vowel-consonant sequences to demonstrate allophonic knowledge in an oddball task. The major phonological difference between the latter authors' experiment and the present one is the type of allophonic rule: in their publications the vowel determined the realization of the following consonant, whereas in the present experiment the opposite was true. However, this cannot be the full explanation, as this was not the case for the comparable null findings by Kazanina, Phillips, & Idsardi (2006). One possible scenario is that there are in fact MMNs in the baseline condition, but that the present sample was not sufficient to detect them: in both the [eːɫ] and the [eiɫ] conditions, the Bayes factors indicated that the statistical power was too low to draw any firm conclusion one way or the other. Further research is necessary.

The results for the rhotics are partially similar to those for the single vowels. Both [əɹ] and [ər] generate MMNs when presented as deviants, but the MMN to [əɹ] is twice the size of that to [ər]. The difference between these two sounds compared to the two vowel conditions is twofold: first, the critical allophone difference is in the consonant rather than in the vowel; second, the rhotic condition is significantly more salient sociolinguistically. The [ɹ] realization, though Netherlandic-Dutch, is a highly salient sociolinguistic variable in both the Netherlands and Flanders (Sebregts 2015), whereas the [r] realization is just as sociolinguistically demarcative, but does not come with the strong public awareness of its counterpart. With prior research not making a strong case for an explanation of the rhotic results in terms of the C/V distinction (recall the results by Jacobsen 2015 and Steinberg, Truckenbrodt, & Jacobsen 2010a, 2010b, 2011), the sociolinguistic explanation remains. The MMNs show that both [əɹ] and [ər] deviants elicit a mismatch, and that the [əɹ] elicits a larger MMN, which shows that this sound is more salient (per Scharinger, Monahan, & Idsardi 2016).

The [əɹ] deviant additionally elicits a difference between the Flemish group and the Netherlandic controls. This difference is topographical in nature: the lowest MMN value is approximately the same for both groups, but the Flemish group shows less activity at frontal sites (for comparison, Figure 5.3 shows the groups side by side, with the effect in the Netherlandic group on the left and the difference with the Flemish group added on top of it on the right). Due to the inverse-mapping problem (computing how electrical signals transmitted from a certain dipole in the brain are distorted by the surrounding brain areas, the skull, and the skin tissue is straightforward; computing the reverse starting from the voltage measured at the scalp is an unsolved problem), the difference in EEG signal at this location does not necessarily reflect differential activity of specifically the frontal brain areas in the Flemish participants. However, we know from prior literature that, among others, frontal areas are involved in MMN generation (e.g. Baldeweg et al. 2002, Giard et al. 1990, Rinne et al. 2000) and that these areas are also responsible for attention (Deouell 2007, Rinne et al. 2000), which is the primary component of sociolinguistic salience (Rácz 2013). As the only remaining difference between the two rhotic allophones is the increased sociolinguistic salience of the [ɹ] allophone, a speculative explanation in sociolinguistic terms could be as follows: the Flemish students did not grow up with this [ɹ] allophone, and hence do not have its sociolinguistic salience ingrained in their representations, hence the reduced contribution from frontal areas to the MMN. However, this needs to be tested thoroughly by future research; as sociolinguistic salience was not the primary manipulation in this study, any explanation in terms of salience can only be exploratory here. Future research should investigate effects of salience on the MMN directly.

In conclusion, the results found in the present chapter do not support Ohala's (1981) account of the actuation of sound change, and do support the views by Hamann (2009) and Beddor (2009). The group differences in the MMNs show that the Flemish students are perfectly able to perceive and categorize the diphthongal [ei], monophthongal [eː, and gliding [əɹ] allophones, but process them differently. In the [ei] case, they even have a *stronger* MMN than the Netherlandic control group. These results do not make sense if the Flemish participants were not able to appropriately perceive or encode these sounds. The results *do* make sense, however, with reference to phonological and sociolinguistic knowledge (the latter by process of elimination, although a neurophysiological basis was suggested). Such knowledge operates on a higher and more abstract level than bottom-up phonetic processing, and effects of such knowledge in the process of sound change are therefore incompatible with Ohala's (1981) view. If, however, sound change happens not during the transmission of the speech signal but during its grammatical evaluation, the results follow naturally as a result of differences in the phonological and sociolinguistic representation of the stimuli in the present study.

5.5 Conclusion

The present study built on previous work by Grosvald & Corina (2012) and others to investigate the listener's role in sound change: is sound change due to differences in low-level perceptual processing (Ohala 1981) or due to differences in higher-level representation in the grammar (Beddor 2009, Hamann 2009)? The results showed evidence against the former but in favor of the latter: the experimental group in this experiment perceived the diphthongal [ei], monophthongal [eː], and glided coda [ɹ] just fine, but processed them differently compared to the control group. While this in and of itself does not speak against Ohala (1981), and in fact would be predicted by him, the specific differences that were found are not easily amenable to an explanation in terms of misperception. The Flemish participants' MMN to the [ei] vowel implies that their perception of it is more than adequate, and is in fact even stronger than it is for the Netherlandic controls. On the rhotics, the Flemish group displayed a less frontal MMN to the [ɹ] allophone, which shows that their perception of this sound is again fine, but that they do not find this sound as salient as the Netherlandic controls do (a sociolinguistic observation which has support in neurophysiological findings, but should be subjected to future research). A possible explanation of the group difference in the [eː] allophone, which was present in the first session and inverted in the second session, was along similar lines. The aforementioned findings reflect sources of grammatical knowledge that are of a higher order than Ohala's (1981) distinction between intrinsic and extrinsic variation (Wang & Fillmore 1961), on which his proposal is founded. In contrast, the results follow naturally if the necessary information for the processing of sound change is evaluated as a normal component of the grammatical system as a whole, and thus if sound change corresponds to a change in the linguistic *grammar*. This is the proposal by Hamann (2009) and Beddor (2009). The observation that the Flemish participants became more Netherlandic-like in their perception of the [ei] vowel after four months' time is thus an observation of grammar learning, not of changes in perception.

Acknowledgments

I thank my dissertation supervisors, Claartje Levelt and Yiya Chen, for steering me into the direction of the experiment reported in this chapter. I additionally thank Daan van Renswoude, who introduced me to Wagenmakers's (2007) method for calculating Bayes factors, and Janet Grijzenhout, for helping me make the link with the work by Flege. I am also grateful for the helpful feedback I received at the University of Amsterdam and at the fifth Sociolinguistics Circle conference, where I presented early versions of this chapter.

CHAPTER 6

ERP responses to regional accent reflect two distinct processes of perceptual compensation

This chapter has been published as: Voeten, C. C., & Levelt, C. C. (2019). ERP responses to regional accent reflect two distinct processes of perceptual compensation. Frontiers in Neuroscience, *13, 546. https://doi.org/10.3389/fnins.2019.00546.*

Abstract

Humans possess a robust speech-perception apparatus that is able to cope with variation in spoken language. However, linguists have often claimed that this coping ability must be limited, since otherwise there is no way for such variation to lead to language change and regional accents. Previous research has shown that the presence or absence of perceptual compensation is indexed by the N400 and P600 components, where the N400 reflects the general awareness of accented speech input, and the P600 responds to phonological-rule violations. The present exploratory chapter investigates the hypothesis that these same components are involved in the accommodation to sound change, and that their amplitudes reduce as a sound change becomes accepted by an individual. This is investigated on the basis of a vowel shift in Dutch that has occurred in the Netherlands but not in Flanders (the Dutch-speaking part of Belgium). Netherlandic and Flemish participants were presented auditorily with words containing either conservative or novel vowel realizations, plus two control conditions. Exploratory analyses found no significant differences in ERPs to these realizations, but did uncover two systematic differences. Over 9 months, the N400 response became less negative for both groups of participants, but this effect was significantly smaller for the Flemish participants, a finding in line with earlier results on accent processing. Additionally, in one control condition where a "novel" realization was produced based on vowel lengthening, which cannot be achieved by any rule of either Netherlandic or Flemish Dutch and changes the vowel's phonemic identity, a P600 was obtained in the Netherlandic participants, but not in the Flemish participants. This P600 corroborates a small number of other studies which found phonological P600s, and provides ERP validation of earlier behavioral results that adaptation to variation in speech is possible, until the variation crosses a phoneme boundary. The results of this exploratory study thus reveal two types of perceptual-compensation (dys)function: on-line accent processing, visible as N400 amplitude, and failure to recover from an ungrammatical realization that crosses a phoneme boundary, visible as a P600. These results provide further insight on how these two ERPs reflect the processing of variation.

6.1 Introduction

It has been successfully argued by many historical linguists that one of the key factors responsible for language variation and change, particularly when it relates to phonetics and phonology, is a poor ability of the human perceptual system to deal with unintentional variation in the speech signal, leading to *misperception* of a speaker by a listener (e.g. Bermúdez-Otero 2015, Blevins 2004, Guion 1998, Hyman 1976, 2013, Ohala 1981, 2012). However, multiple decades of research on speech processing by psycholinguists show that, in fact, the human speech system is very capable of handling non-meaningful variation, such as variation due to anatomical differences between speakers or the use of a regional accent. Processes such as perceptual learning (Norris, McQueen, & Cutler 2003), rate normalization (Bosker & Reinisch 2015), compensation for coarticulation (Mann & Repp 1980), and many other innate or acquired perceptual skills (see Cutler 2012) enable the listener to accurately make the link between the forms speakers intend vs. the sounds they actually produce. If historical linguists are correct that the driving force behind linguistic change (and particularly sound change) is misperception, then the question is when and how these perceptual-compensation processes found in psycholinguistics "give way", i.e. fail to correctly compensate for variation, thus enabling historical sound change to actuate. In empirical terms: under what conditions can we detect psycho- or neurolinguistic correlates of *unsuccessful* perceptual compensation for variation? The present chapter provides a starting point in answering this question using evidence from ERP data.

Evans & Iverson (2007) have shown that it is possible to detect long-term accommodation to variation in speech, by investigating the speech production and perception of 19 English first-year university students. These students hailed from different dialect regions of the United Kingdom, and were shown behaviorally to adapt their speech production to the Standard Southern British English university norms. In addition, a correlation was found with participants' perception of accented speech, but the latter did not reliably change on its own over time. The present study takes a similar approach, but focuses on the processing aspect. The language used for the investigation is Dutch. Dutch is spoken both in the Netherlands and in the northern part of Belgium (henceforth: Flanders), but due to thorough standardization processes that took place in the Netherlands but not in Flanders (Grondelaers & van Hout 2011), there are significant differences in the phonological systems of these two varieties. The Netherlandic variety (henceforth: Netherlandic Dutch), has undergone changes in its distribution of the tense mid vowels (/eː,øː,oː/), diphthongs (/ɛi,œy,ɑu/), and rhotic (/r/). Specifically, Netherlandic Dutch has diphthongal realizations of /eː,øː,oː,ɛi,œy,ɑu/ (thus realizing these vowels as [ei,øy,ou,ɛi,œy,ɑu]) and a glided coda /r/ (realized [ɹ]), whereas the Belgian variety (henceforth "Flemish Dutch") has monophthongal realizations of /eː,øː,oː/ (yielding realizations [eː,øː,oː]), markedly less diphthongization in /ɛi,œy,ɑu/, and does not glide the coda rhotic (realizing it as [r]). These differences have all arisen via sound changes that have taken place in Netherlandic Dutch but not in Flemish Dutch (Sebregts 2015, Van de Velde 1996). This makes the present-day variation between Netherlandic Dutch and Flemish Dutch a useful proxy for historical sound change.

The study reported here investigates the perception of these speech sounds in speakers of Flemish Dutch who have migrated to the Netherlands. Ten Flemish-Dutch speakers (henceforth: "FDS"), all first-year university students who migrated to the Netherlands, are compared to 10 Netherlandic-Dutch speakers (henceforth: "NDS"). Participants are tested multiple times to test for possible longitudinal adaptation on the part of the Flemish students. Using an exploratory extension of the violation paradigm (Friederici, Pfeifer, & Hahne 1993) to phonological processing, the objective of the investigation is to find behavioral and electrophysiological correlates of the processing of the type of variation under discussion. While it will turn out that this endeavor will be unsuccessful, two robust differences between the two groups will provide new information about the types of phonological variation whose processing neurolinguistic methods can detect. It will be shown that the FDS are less able to "take in" NDS speech, which is reflected by a smaller N400 decrease over two repetitions of the same experiment compared to the NDS (in line with behavioral findings by Floccia et al. 2009). In addition, it will be shown that there are ERP-detectable differences in the processing of NDS speech between the two groups. Specifically, in words where the vowel /ɛ/ is realized as [ɛː] —an ungrammatical realization that cannot be achieved by any known phonological rule of Standard Dutch—a P600 is obtained in the NDS, but not in the FDS. This is in line with previous research (Domahs et al. 2009, Pater et al. submitted) about the role of the P600 in the processing of phonological rules.

The remainder of the chapter is structured as follows. Section 6.2 discusses the psycholinguistic and neurolinguistic correlates of accent (violation) processing that have been identified in the prior literature. Due attention is paid to the well-known N400 component, and to the P600, which is a relatively new, but not unknown, component in this field. Section 6.3 details the methodology used in the present experiment. Section 6.4 provides the results, which are discussed in Section 6.5, first seperately for the two findings (N400 and P600) and then together. Finally, Section 6.6 concludes the chapter.

6.2 Accent processing

At its core, the present study is about accent accommodation, a subfield intersecting psycholinguistics and phonetics. Previous research in this field has shown that listeners are very adept at compensating for linguistic variation, particularly in the vowel system. Maye, Aslin, & Tanenhaus (2008), for instance, show that listeners are able to accommodate to a completely novel vowel shift in English (all vowels lowered by one degree, so "wicked witch" becomes "weckud wetch") after only a few minutes of exposure. At the same time, however, Floccia et al. (2006) found that a notable regional accent incurs a slowdown in lexical-decision tasks of about 30 ms. This effect accumulates over time, i.e. with longer words this delay increases more than proportionally. This suggests an interference effect starting from the very beginning of lexical processing, which persists even as the listener receives more exposure to the accented speech (Floccia et al. 2009).

These results suggest that while participants in behavioral tasks are able to accommodate successfully in order to fulfill the task, their processing is still somehow impaired when confronted with accentual variation. This processing difficulty has been measured directly in ERP investigations. Goslin, Duffy, & Floccia (2012) found that accented realizations reduced the amplitude of both the phonological-mapping negativity (otherwise known as the N280) and the N400. The relationship between accent and the phonological-mapping negativity is obvious, but the involvement of the N400 might be considered surprising, given that this component is normally connected to semantic processing, or more generally to lexical predictions (Dambacher et al. 2006). The results to be presented in Section 6.4 will give reason to postulate two different ways in which the N400 may be modulated by accentual variation. On the one hand, the findings by Goslin, Duffy, & Floccia (2012) suggest that a persistent regional accent reduces the strength of the lexical predictions made by a listener, resulting in a *reduced* overall N400 amplitude due to simple parsing difficulty, which causes the listener to predict more cautiously. On the other hand, the results from the present chapter will give reason to postulate an N400-*increasing* effect for regionally-accented words, due to their decreased consolidation in the lexicon over time.

Recent studies have intimated another ERP component in accented-speech processing: the P600. Originally known from syntax (Osterhout & Holcomb 1992), it was observed by Liu et al. (2011) that a P600 could also be elicited by phonology. Specifically, Liu et al. (2011) observe a P600 in Chinese participants who read well-known poems in which some words were replaced with synonyms that only differed in orthographic and phonological form. They argue that the P600 that was elicited by these "deviant" words must be due to phonological processing, since other sources of integration difficulties such as semantic violations were absent. Kung, Chwilla, & Schriefers (2014) found a similar effect in a lexical-decision task in Chinese. In this task, Chinese words with a low lexical tone were embedded as the final word in a sentence that carried the intonation pattern of a question. Because questions in Chinese end with rising intonation, the pitch contour of such words is very similar to the pitch contour of words with a high lexical tone in a regular statement sentence. The resulting processing difficulty ("is this a word with a low lexical tone that rises because the sentence is a question, or is the sentence a statement and does the word simply carry a high lexical tone?") manifested as a P600, which Kung, Chwilla, & Schriefers (2014) interpret as being caused by reanalysis when the listener resolves this conflict by choosing (in these cases) for the question interpretation.

Phonological P600 effects have also been found beyond prosody, viz. in the domain of segmental phonology. Domahs et al. (2009) obtained a P600 in a lexical-decision task, and Pater et al. (submitted) found a P600 in a phonological artificial-language-learning task. Both of these studies investigated a specific subset of accented speech, viz. violations of allophonic rules: Pater et al. (submitted) violated an artificially-learned voicing-agreement rule, and the Domahs et al. (2009) study found a P600 when two stop consonants followed each other in a way that violated the phonotactics of German (the native language of their participants). It is important to mention that both violations were neutralizing: the violating consonant was already present on its own in the phoneme inventory. Behavioral findings by Witteman et al. (2015) suggest that this matters: they show that adaptation to an accent in general is possible, but not to individual sounds that cross a phoneme boundary. Furthermore, there is more interference when a single sound in a word is replaced by a realization *that does not occur normally* (e.g. Dutch /œy/ replaced by German [ɔɪ], which does not exist in Dutch; Witteman, Weber, & McQueen 2014).

It thus appears that, even though adaptation is possible, there are multiple behavioral and electrophysiological correlates of problems faced by listeners when processing accented speech. In reaction-time experiments, they are slower. In ERP studies, the N280, N400, and P600 play a role. Two of the four phonological P600 studies discussed found the effect in the context of allophonic-rule violations. The present study integrates these results and attempts to extend them by investigating reaction times and ERP responses to Netherlandic-Dutch speech by Flemish-Dutch students. The approach is similar to that taken by Witteman et al. (2015) and used in the P600 studies by Domahs et al. (2009) and Pater et al. (submitted): only a single sound is manipulated in an otherwise normal Netherlandic-Dutch word.

Given the above, the aim of the present study is to investigate two things. Behaviorally, it can be expected that the well-known effect of identity priming (participants being faster to read a word aloud if they have just been presented the same word auditorily) will be less strong for realizations that do not conform to participants' phonological grammars. Specifically, the expectation is that unmanipulated identity primes facilitate word reading, but manipulated identity primes incur the same RT slowdown reported by Floccia et al. (2009) and Floccia et al. (2006) on top of this identity-priming effect. Electrophysiologically, the manipulations are expected to specifically elicit a P600 ERP when they result in ungrammatical allophones, and possible across-the-board N280 and N400 effects may arise in general.

The hypotheses presented above are evaluated for 10 NDS and 10 FDS who have only just moved to the Netherlands to start their university studies there (paralleling Evans & Iverson's 2007 study). The expectation is that there will be differences, but that they will reduce over months of time as the FDS participants receive more exposure; the present experiment takes 9 months divided into three sessions. Differences are expected in terms of RTs and in terms of ERPs. Concerning RTs, the expectation is that the hypothesized difference in identity-priming effects will be smaller for the FDS than for the NDS, as the FDS will have remaining difficulty parsing also the non-manipulated segments of the words. In terms of ERPs, it is expected that the FDS have different N400 responses (regardless of task or type of violation), in line with findings by Goslin, Duffy, & Floccia (2012). In addition, the P600 effect in response to allophonic violations (Domahs et al. 2009, Pater et al. submitted) is expected to be smaller for the FDS than for the NDS, as the FDS should have less robust prior expectations due to having had less exposure to Netherlandic-Dutch speech.

The task used in the present study, explained in Section 6.3, has not been used in a violation paradigm before, and the research is therefore of an exploratory nature. The results will show that the task is sensitive to phonological violations in individual speech sounds, but that this effect must have a deeper source than surface allophones: it will be shown that the task cannot detect contextual violations, but is instead sensitive only to realizations that lie outside the set of possible realizations of a phoneme, further precising the behavioral findings by Witteman et al. (2015) and Witteman, Weber, & McQueen (2014).

6.3 Materials and method

6.3.1 Participants

Participants consisted of 10 FDS participants (seven female, three left-handed; mean age = 22.71 years, SD = 3.54 years) and 10 Dutch controls (seven female, one left-handed; mean age = 20.53 years, SD = 2.48 years). The FDS participants were all in their first year of study at a Dutch university in the Randstad (either Leiden University or the University of Amsterdam) and were speakers of a variety of Flemish Dutch. The control participants were Netherlandic-Dutch students, not necessarily in their first year, who were also studying in the Randstad and had grown up in a Randstad-Dutch environment. The FDS participants were tested as soon as possible after the beginning of the academic year (mean number of days past September 1^{st} = 21.5 days; SD = 7.93 days). This restriction was not applied to the control group (mean number of days past September 1^{st} = 104.30 days; SD = 54.40 days).

To find out about possible longitudinal adaptation processes, participants were tested over the course of three sessions. The mean interval between the sessions was 129.29 days (SD = 23.19 days) for session 1–2, and 112.75 days (SD = 22.94 days) for session 2–3. The experimental procedure and tasks, which are described below, were the same for all three sessions. In the end, 23 FDS datasets were collected and 28 NDS datasets; the discrepancy with the expected $3 \times 10 = 30$ datasets per group is due to drop-outs (session 1–2: two left-handed FDS and one right-handed NDS; session 2–3: one additional right-handed FDS) and equipment failure (one left-handed FDS in session 1 and one right-handed FDS in session 2). Table 6.1 lists the final set of participants present in the sample.

Table 6.1: Overview of the final population from which data were obtained.

	Session		
Participant	1	2	3
FDS-0		✓	✓
FDS-1	✓	✓	
FDS-2	✓		
FDS-3	✓	✓	✓
FDS-4	✓	✓	✓
FDS-5	✓	✓	✓
FDS-6	✓	✓	✓
FDS-7	✓		
FDS-8	✓	✓	
FDS-9	✓	✓	✓
NDS-0	✓	✓	✓
NDS-1	✓	✓	✓
NDS-2	✓	✓	✓
NDS-3	✓	✓	✓
NDS-4	✓	✓	✓
NDS-5	✓	✓	✓
NDS-6	✓	✓	✓
NDS-7	✓		
NDS-8	✓	✓	✓
NDS-9	✓	✓	✓

6.3.2 Stimuli

A phonetically-trained female speaker from the Randstad area of the Netherlands produced 309 prime words embedded in a carrier sentence. The experiment was an exploratory part of a larger battery of both neurolinguistic and non-neurolinguistic tests; for purposes of the present chapter, 160 of these words are relevant and the remainder are fillers. The 160 experimental words comprised 8 groups of 20 words containing one of the phonemes /eː,øː,oː,ɛi,œy,ɑu,aːʀ,ɛ/ (the last vowel does not differ between Netherlandic Dutch and Flemish Dutch and was included as a control) in stressed[1] position. An equal number of fillers was used, containing the same phonemes separated into the same conditions, but positioned in a different phonotactic environment, namely preceding coda /l/. In this environment, the Netherlandic-Dutch realization of these vowels is the same as the Flemish-Dutch realization. An obvious exception was made for consonantal control /aːʀ/, which cannot be followed by coda /l/; this condition was simply included as a target twice, in order to retain the balance of the conditions presented to the participants. The same held for the /ɑul/ condition, as a lexical gap in Dutch prevents the vowel /ɑu/ from being followed by coda /l/. For reasons of convenience, 3 × 3 words beginning with one of the point vowel phonemes /i,u,aː/ were also included as fillers, both before /l/ and before non-/l/.

The 309 prime words thus present in the design were selected on the basis of frequency: for each cell, the 20 words selected are the 20 most frequent words according to CELEX (Baayen, Piepenbrock, & Gulikers 1995) starting with the relevant phoneme(s) (mean log frequency = 6.41; SD = 2.07). The critical phonemes were always located at the beginning of the words to maximize any possible priming effects on participants' reaction times, and to enable time-locking of ERPs to the onset of the critical manipulations. There were two exceptions: the requirement of word-initiality was dropped for the vowel /øː/, as no words beginning with stressed /øː/ were available in the corpus, presumably as a result of this vowel's general low frequency in Dutch. The initiality requirement was also rescinded for the filler items.

All primes, both target and filler with the exception of the nine point-vowel tokens, were recorded in two different variants. One of these variants was typical for NDS phonology and one of these variants was atypical of NDS phonology (and, in all non-filler non-control items, typical of FDS phonology). These two variants will henceforth be referred to as "NDS realizations" and "non-NDS realizations". Table 6.2 summarizes the design. For the NDS realizations, the vowels /eː,øː,oː,ɛi,œy,ɑu/ were realized as [ei,øy,ou,ɛi,œy,ɑu] in the experimental items. This is typical of NDS phonology, but atypical for FDS

---

[1] Unstressed vowels in Dutch may optionally undergo reduction (Booij 1995), which could result in the elision of the crucial upgliding diphthongization.

phonology. For the filler items where these vowels were followed by a coda /l/, they were realized as [eː,øː,oː,ɛː,œː,ɑː], which is typical for both NDS and FDS phonology. For the /ɛ/ control, the vowel was realized as [ɛ] (typical for both NDS and FDS), and for the /aːʀ/ control, the sequence was realized [aːɹ] (typical of NDS only). For the non-NDS realizations of the experimental items, the vowels /eː,øː,oː,ɛi,œy,ɑu/ were realized as [eː,øː,oː,ɛː,œː,ɑː], which is typical of FDS but not permissible in NDS given the lack of a following coda /l/ in the experimental items. For the filler items where these vowels were followed by a coda /l/, they were realized as [ei,øy,ou,ɛi,œy,ɑu], which is not grammatical in FDS (where these realizations simply do not occur) or NDS (because these realizations are not permitted before coda /l/). The non-NDS /ɛ/ control was realized as [ɛː], which is a phonologically illicit realization in either NDS or FDS speech, and the /aːʀ/ control was realized as [aːʀ], which does not apply the NDS-typical but FDS-atypical rule gliding /ʀ/ to [ɹ] in coda position.

Figures 6.1 and 6.2 show spectrograms of the word /eːn/ realized as [ein] and [eːn] that demonstrate the difference under discussion. A crucial property of the experiment is that only the target phoneme (or phoneme sequence) was realized in a specific way; the remainder of the word was produced naturally. This prevents confounding the specific effect of allophone pronunciation with a global effect of regional accent, which is precisely the distinction that this study aims to tease apart. A reviewer additionally asks if there are orthographic differences between Netherlandic Dutch and Flemish Dutch that could influence FDS performance on the task; this is not the case.

Each prime was presented auditorily followed by a target presented visually. The targets were selected from the same set of 309 words as the primes. The pairing of targets to primes is as follows. In three conditions (19.5% of the experiment), viz. /eː,ɛi,aːʀ/, the prime word and target word were the same; in the other conditions (80.5% of the experiment), the word on the screen was a random selection (without replacement) of the non-/eː,ɛi,aːʀ/ words in the experiment.

Table 6.2: Overview of the allophone variants used in the experimental items.

	Realization (NDS versus non-NDS) used in prime items							
	Before non-/l/ (Target)				**Before /l/ (Filler)**			
Phoneme	NDS	Typical for	Non-NDS	Typical for	NDS	Typical for	Non-NDS	Typical for
/eː/	[ei]	NDS	[eː]	FDS	[eː]	FDS	[ei]	neither
/øː/	[øy]	NDS	[øː]	FDS	[øː]	FDS	[øy]	neither
/oː/	[ou]	NDS	[oː]	FDS	[oː]	FDS	[ou]	neither
/ɛi/	[ɛi]	NDS	[ɛː]	FDS	[ɛː]	FDS	[ɛi]	neither
/œy/	[œy]	NDS	[œː]	FDS	[œː]	FDS	[œy]	neither
/ɑu/	[ɑu]	NDS	[ɑː]	FDS	[ɑː]	FDS	[ɑu]	neither
/aːʀ/	[aːɹ]	NDS	[aːʀ]	FDS				
/ɛ/	[ɛ]	NDS + FDS	[ɛː]	neither	[ɛ]	NDS + FDS	[ɛː]	neither
/i/	[i]	NDS + FDS	[i]	NDS + FDS	[i]	NDS + FDS	[i]	NDS + FDS
/u/	[u]	NDS + FDS	[u]	NDS + FDS	[u]	NDS + FDS	[u]	NDS + FDS
/aː/	[aː]	NDS + FDS	[aː]	NDS + FDS	[aː]	NDS + FDS	[aː]	NDS + FDS

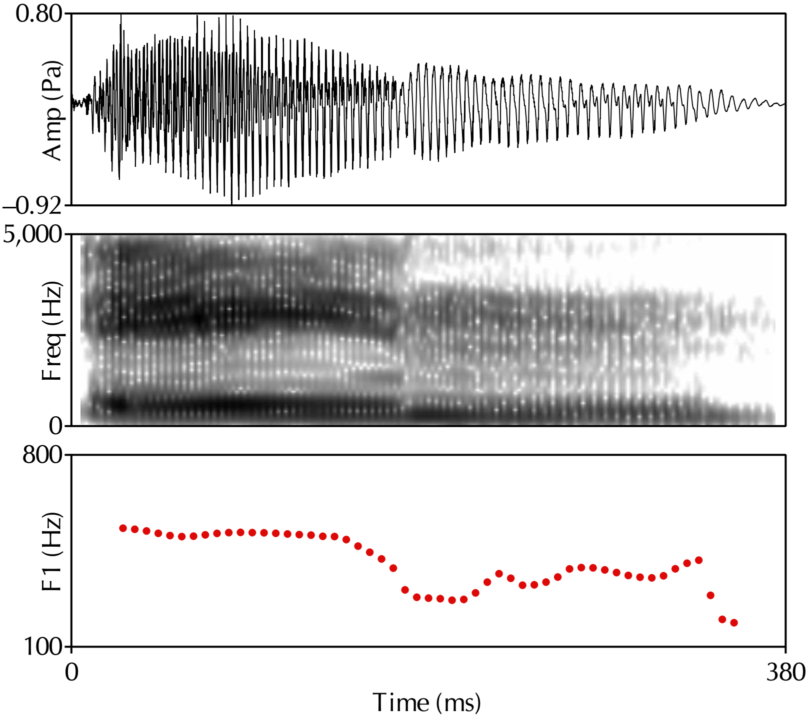

Figure 6.1: Example waveform, spectrogram, and F1 trajectory (the critical difference between diphthongal and monophthongal realizations) for the NDS realization of /eːn/ as [ein]. Towards the end of the vowel, the F1 falls.

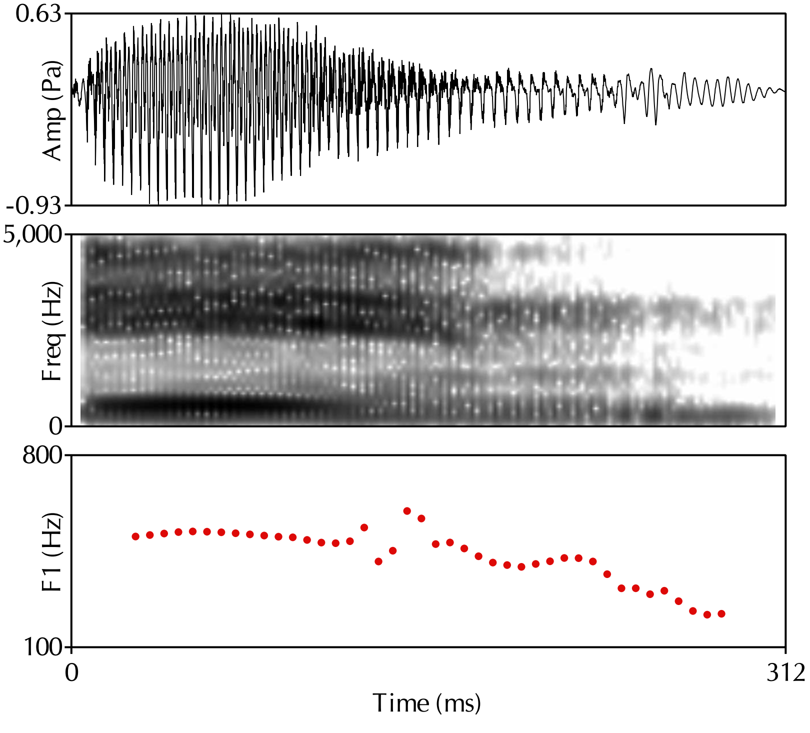

Figure 6.2: Example waveform, spectrogram, and F1 trajectory (the critical difference between diphthongal and monophthongal realizations) for the non-NDS realization of /eːn/ as [eːn]. The F1 stays stable throughout the vowel.

6.3.3 Procedure and data acquisition

Participants were seated in front of a computer screen, loudspeakers, and a microphone, in a sound-attenuated and electrically-shielded booth. The experiment consisted of 618 trials with three breaks in between, spaced evenly throughout the experiment. All trials were presented on the computer using PsychoPy version 1.83.04. Before the experiment started, participants were presented instructions on the computer screen, which were also read aloud by a male Standard-Dutch speaker with a Netherlandic-Dutch accent.

Each trial started with a black screen, followed by auditory presentation of the prime. When this prime had finished playing, the target word appeared on the screen (presented orthographically), which participants had been instructed to read aloud. Between two trials, a fixation cross was presented for 1 s. The data collected from the task are ERP responses to the prime words and vocal reaction times (i.e. speaking latencies) to the target words. A diagram showing an example trial and the data recorded from it is shown in Figure 6.3.

The manipulation took place in the primes: for each of the 309 words recorded, participants heard both the Netherlandic-Dutch variant and, in a different trial, the non-Netherlandic-Dutch variant. Which of these two variants was presented first for each word was randomized and counterbalanced.

During the whole task, continuous-time EEG activity was recorded using 32 Ag/AgCl electrodes with a sampling rate of 512 Hz. Two flat electrodes at the mastoids provided a reference signal, which was subtracted from the EEG signal in off-line processing; an additional four flat electrodes recorded horizontal and vertical extra-oculograms. In addition to the EEG activity, the speech of the participant was recorded while they realized the target word. Recording was started immediately after the prime word was presented, right when the target word became visible on the screen.

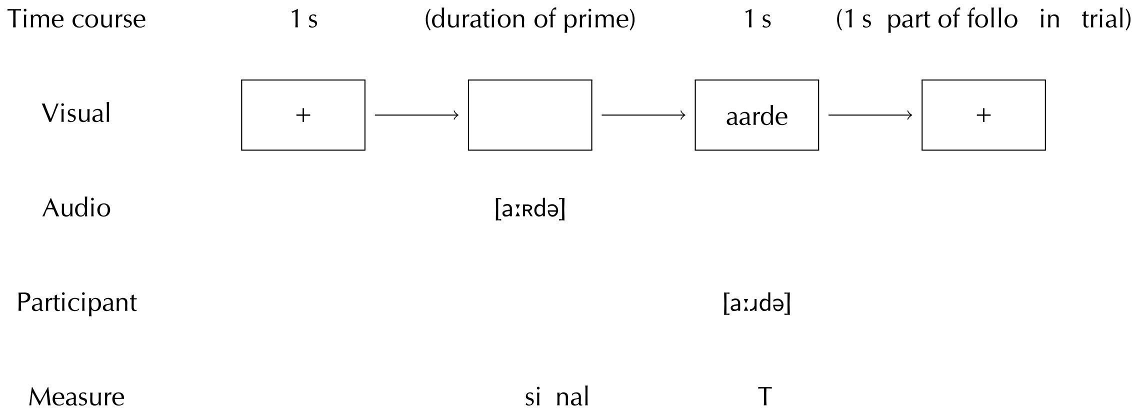

Figure 6.3: Example trial for the production task.

6.4 Results

6.4.1 Reaction times

The words realized by the participants were aligned to their phonetic transcriptions (which were obtained from CELEX) using the Viterbi forced aligner present in HTK (Young et al. 2002). Forced alignment of speech sounds to phonemes is a more principled measure of speech onset time than thresholding raw acoustic energy, as the procedure uses speech-specific information in the signal and can hence do a better job at separating speech from background noise. For every word, the program produced a list of start and end points of the individual consonants and vowels present in the speech stream. Vocal RTs were obtained by extracting the time index of the first phoneme following the word-initial silence. RTs were obtained with a granularity of 10 ms.

The effects of the various factors in the design on these reaction times were analyzed by means of a generalized linear mixed-effects model with identity link and gamma errors[2], following Lo & Andrews (2015). The fitting engine used for the model was function `glmer` from R (R Core Team 2020) package `lme4` (Bates et al. 2015b). Fixed effects were added for "Group" (treatment-coded: 0 = NDS, 1 = FDS), "Allophone" (treatment-coded: 0 = Netherlandic-Dutch; 1 = non-Netherlandic-Dutch), "Session" (coded for linear and quadratic trends using orthogonal polynomials), "Identity" (treatment-coded: 0 = the prime and target words differed; 1 = the prime and target word were the same, which was the case in the /aːʀ,eː,ɛi/ conditions), "Condition" (sum-coded), and all possible interactions. Using R package `buildmer` (Voeten 2019a), random slopes by participants and words were included over all terms as long as the model would still converge; these terms were entered in the order of their contribution to the log-likelihood, such that when the model eventually failed to converge, the most information-rich random slopes had been included. From this maximal model, terms were excluded in backward stepwise order based on the change in BIC (Schwarz 1978). (Given the large number of interaction parameters present in the maximal model, BIC is a more natural elimination criterion than the likelihood-ratio test.)

Results of the analysis are shown in Table 6.3. Because the model used an identity link, the resulting model coefficients are directly interpretable as milliseconds of response latency. The intercept is placed at 827 ms ($\hat{\beta} = 827.16$, $SE = 2.22$, $t = 372.06$, $p < .001$). This reflects the temporal onset of the first phoneme in the participant's response, for the Dutch control participants when they were presented with non-identity, Netherlandic-Dutch targets. Participants became slightly slower over the three sessions ($\hat{\beta} = 54.22$, $SE = 1.86$,

---

[2] Inverse-Gaussian (Wald) errors were also considered, but provided a worse fit to the data (higher AIC) compared to the gamma-errors model.

Table 6.3: Fixed-effect coefficients of the reaction-times analysis. The model additionally includes random intercepts by participants and by words, and random slopes for the factor "Session" by participants and by words.

Factor	**Estimate (SE)**	***t***	***p***	**Sig.**
(Intercept)	827.16 (2.22)	372.06	<.001	***
Session (Linear)	54.22 (1.86)	29.13	<.001	***
Session (Quadratic)	−10.59 (2.90)	−3.65	<.001	***
Group = FDS	41.32 (1.93)	21.37	<.001	***
Prime = Identity	−46.64 (2.03)	−23.02	<.001	***
Allophone = Non-NDS	−0.15 (1.71)	−0.09	.93	
Group = FDS × Identity	11.61 (2.28)	5.10	<.001	***
Prime = Identity × Non-NDS	−6.00 (1.67)	−3.59	<.001	***

$t = 29.13$, $p < .001$), although the speed loss between sessions 2 and 3 was not as large as the speed loss between sessions 1 and 2 ($\hat{\beta} = -10.59$, $SE = 2.90$, $t = -3.65$, $p < .001$). Overall, the FDS were slower than the NDS ($\hat{\beta} = 41.32$, $SE = 1.93$, $t = 21.37$, $p < .001$). Identity primes incurred faster RTs than non-identity primes ($\hat{\beta} = -46.64$, $SE = 2.03$, $t = -3.02$, $p < .001$), although this advantage was smaller for the FDS ($\hat{\beta} = 11.61$, $SE = 2.28$, $t = 5.10$, $p < .001$). Overall, the non-Netherlandic-Dutch allophones incurred slightly faster RTs than the Netherlandic-Dutch allophones, although the size of this effect was smaller than the 10-ms granularity with which HTK had provided the reaction times ($\hat{\beta} = -6.00$, $SE = 1.67$, $t = -3.59$, $p < .001$).

6.4.2 ERP results

The ERP data were detrended and an off-line bandpass filter was applied passing a frequency domain of 1–30 Hz. Epochs were time-locked to the onset of the prime words and were extracted 0–800 ms post-stimulus-onset, after subtracting a −100 ms baseline[3]. Epochs contaminated by eyeblinks or other movement-related artifacts were rejected. The resulting grand-average waveforms (averaged over all participants, items, and electrodes), are shown in Figure 6.4.

In order to determine the precise temporal window and ROI to be used in the statistical analyses and data plots, permutation testing (Maris & Oostenveld 2007) was applied to identify the locations of significant differences between the conditions. Repeated ANOVAs were run on each ⟨timepoint,electrode⟩ pair in the data using R package `permutes` (Voeten 2018). The design of the test

---

[3] The rather short span of this baseline was necessary due to divergence of the baselines earlier than −100 ms.

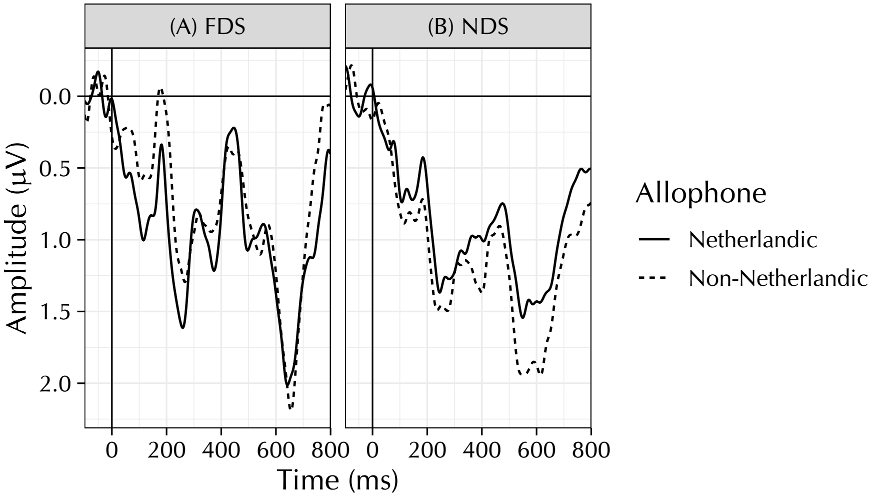

Figure 6.4: Grand-average waveforms calculated over the full dataset, averaged over all participants, electrodes, and the three sessions.

was a 2 × 2 ANOVA with fixed factors for "Allophone" (encoding the type of allophone presented to the participant), "Group" (FDS or NDS), and the interaction between the two. Because it was conceivable that the experimental items /eː,øː,oː,ɛi,œy,ɑu/ and the control items /ɛ,aːʀ/ might be differentially sensitive to the manipulation of the prime allophones, the permutation tests were run twice: once on the full dataset (to identify global differences between the groups of participants and allophones) and once for each of the eight conditions separately (to identify possible differences between the experimental items, the control vowel /ɛ/, and the control consonant condition /aːʀ/). The analysis of the whole dataset, plotted in Figure 6.5, identified a global effect of "Group", ranging from 390 to 470 ms at frontal, central, and parietal sites. Figure 6.7 shows the grand-average waveforms corresponding to this global difference between the groups. The analyses of the individual vowels failed to identify meaningful windows in the allophonic conditions [ei∼eː], [øy∼øː], [ou∼oː], [ɛi∼ɛː], [œy∼œː], [ɑu∼ɑː], and [aːɹ∼aːʀ], but for the [ɛ∼ɛː] condition, an effect of "Allophone × Group" was observed at essentially all electrode sites within a temporal window of 560 to 660 ms. This effect is plotted in Figure 6.6; Figure 6.8 shows the corresponding grand-average waveforms. Reasons why only this condition elicited a significant ERP are discussed in Section 6.5.

The effects found in Figures 6.5 and 6.6 appear to correspond, respectively, to the classic N400 and P600 effects. To analyze the N400 effect, each response was averaged over the 390–470 ms window and over the frontal and central

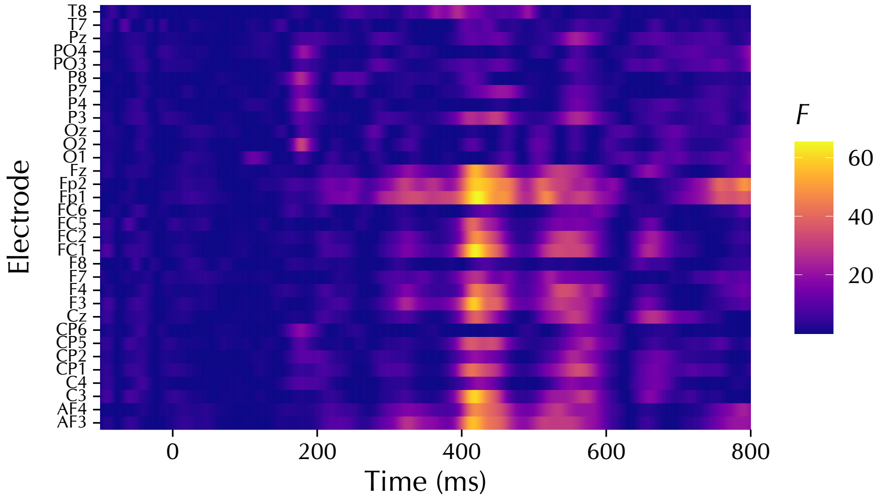

Figure 6.5: Permutation tests performed on the whole dataset, showing the effect of the factor "Group". A significant difference can be observed, which reaches permutation-based significance between 390–470 ms.

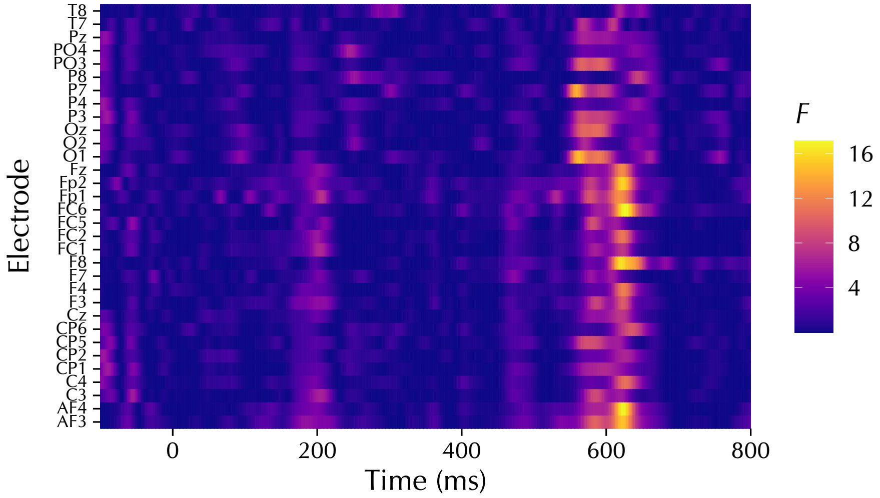

Figure 6.6: Permutation tests performed on the data for the [ɛ∼ɛː] contrast, showing the effect of the factor "Allophone × Group". A significant difference can be observed, which reaches permutation-based significance between 560–660 ms.

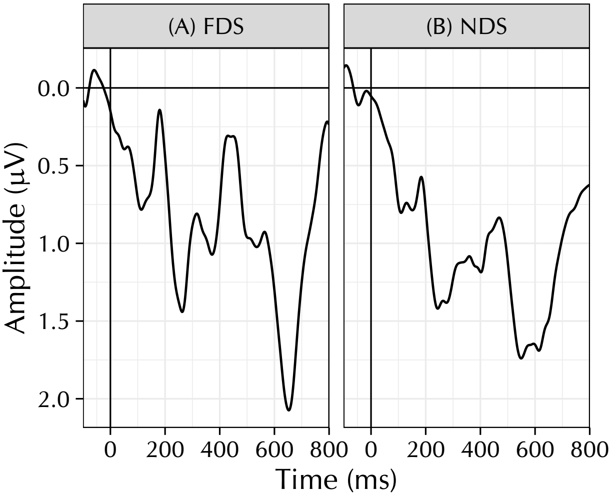

Figure 6.7: Grand-average waveforms calculated over the full dataset, averaged over the three sessions and the two allophone conditions. A difference in amplitude can be observed between the FDS and the controls, which reaches permutation-based significance in the 390–470 ms window.

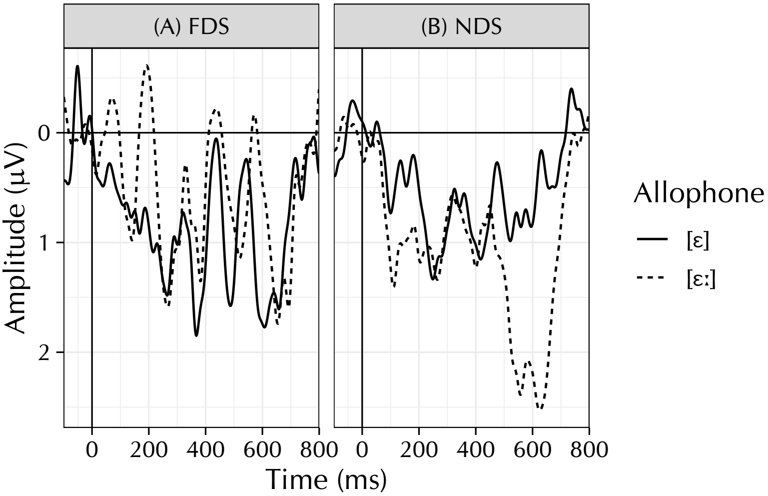

Figure 6.8: Grand-average waveforms calculated over the [ɛ∼ɛː] condition only, averaged over all participants, electrodes, and the three sessions. A difference in amplitude can be observed between the two allophone conditions, in the NDS group only, which reaches permutation-based significance in the 560–660 ms window.

electrodes. The resulting data were analyzed by means of a linear mixed model with fixed effects for "Group", "Allophone", and "Session", and a complete random-effect structure for participants and items using the same procedure as in Section 6.4.1. Terms were selected for inclusion in the model by means of backward stepwise elimination, using the significance of the change in log-likelihood as the inclusion criterion. The resulting model is shown in Table 6.4. The results show significant effects for "Session (Linear)" ($\hat{\beta} = 0.45$, SE = 0.18, $t_{12,650.91} = 2.55$, $p = .01$) and "Session (Quadratic)" ($\hat{\beta} = 0.79$, SE = 0.17, $t_{12,672.79} = 4.55$, $p < .001$). The linear component shows that the N400 became less pronounced over the three sessions, whereas the quadratic component implies that the N400 shrinkage between sessions 2 and 3 was larger than the reduction between sessions 1 and 2. The linear component additionally entered into a significant interaction with "Group = FDS" ($\hat{\beta} = -1.17$, SE = 0.29, $t_{12,250.75} = -3.97$, p .001). This suggests that the linear trend of decreasing N400 amplitudes was significantly less pronounced for the FDS than it was for the NDS.

To analyze the P600 effect, the data were averaged over the 560–660 ms window and all electrodes. The data were analyzed in the same way as for the N400 effect. The model containing the terms that remained after stepwise elimination is shown in Table 6.5.The results show a significant effect for "Session (Linear)" ($\hat{\beta} = -0.76$, SE = 0.31, $t_{1,179.68} = -2.47$, $p = 0.01$). This suggests that the average amplitude in this window became slightly smaller in magnitude over the three sessions. A significant main effect was found for "Allophone = Non-NDS" ($\hat{\beta} = 1.70$, SE = 0.44, $t_{1,394.70} = 3.86$, $p < .001$), indicating that the non-NDS allophone elicited a much larger P600 response than the NDS allophone. However, this factor interacted significantly with "Group = FDS" ($\hat{\beta} = -2.51$, SE = 0.67, $t_{1,395.54} = -3.76$, $p < .001$), such that the P600 was completely negated in the FDS and in fact only showed up in the NDS group.

Table 6.4: Fixed-effect coefficients for the N400 effect.

Factor	Estimate (SE)	*t*	df	*p*	Sig.
Intercept	1.43 (0.74)	1.94	17.78	.07	
Group = FDS	−0.89 (1.04)	−0.85	17.97	.41	
Session (Linear)	0.45 (0.18)	2.55	12,650.91	.01	*
Session (Quadratic)	0.79 (0.17)	4.55	12,672.79	<.001	***
Group = FDS × Session (Linear)	−1.17 (0.29)	−3.97	12,250.75	<.001	***
Group = FDS × Session (Quadratic)	0.21 (0.27)	0.78	12,665.72	.44	

Table 6.5: Fixed-effect coefficients for the P600 effect.

Factor	Estimate (SE)	*t*	df	*p*	Sig.
Intercept	0.81 (0.51)	1.60	24.60	.12	
Session (Linear)	−0.76 (0.31)	−2.47	1,179.68	.01	*
Session (Quadratic)	−0.17 (0.29)	−0.59	1,400.81	.56	
Allophone = Non-NDS	1.70 (0.44)	3.86	1,394.70	<.001	***
Group = FDS	0.49 (0.74)	0.66	27.79	.51	
Allophone = Non-NDS × FDS	−2.51 (0.67)	−3.76	1,395.54	<.001	***

6.4.3 Topographical distribution

The topographical distribution for the two effects is shown in Figure 6.9 for the N400 and Figure 6.10 for the P600 effect. For the N400, both the FDS and the NDS showed the lowest activity in central-parietal areas. The difference between the two was that the FDS's activity was lower than the NDS's at especially the frontal and frontal-central sites. Since this between-groups difference was present throughout the whole experiment, it also shows up in the topographical plots of the P600 effect. In those plots, the FDS do not show any interpretable differences between the [ɛ] and the [ɛː] allophones, other than the aforementioned frontal activity being more negative for the latter allophone. The NDS, however, show significantly more activity in parietal-occipital areas for the [ɛː] allophone compared to the [ɛ] allophone, which corresponds to the classic ROI of the P600.

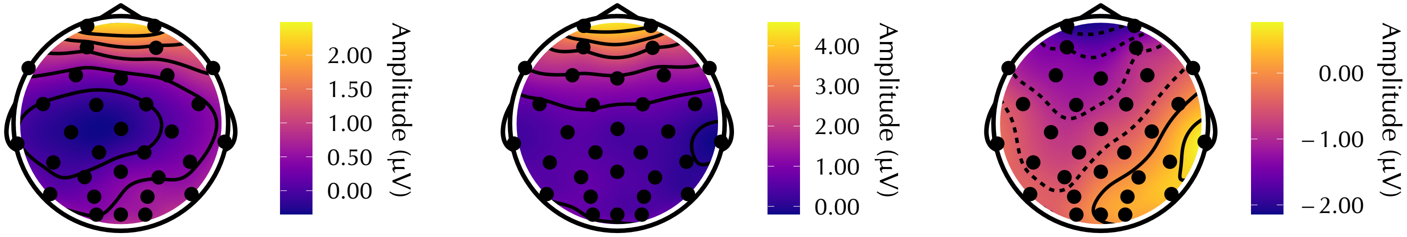

Figure 6.9: Topographical distribution of the N400 effect. The left head shows the FDS, the middle head shows the NDS, and the right head shows the difference between the two.

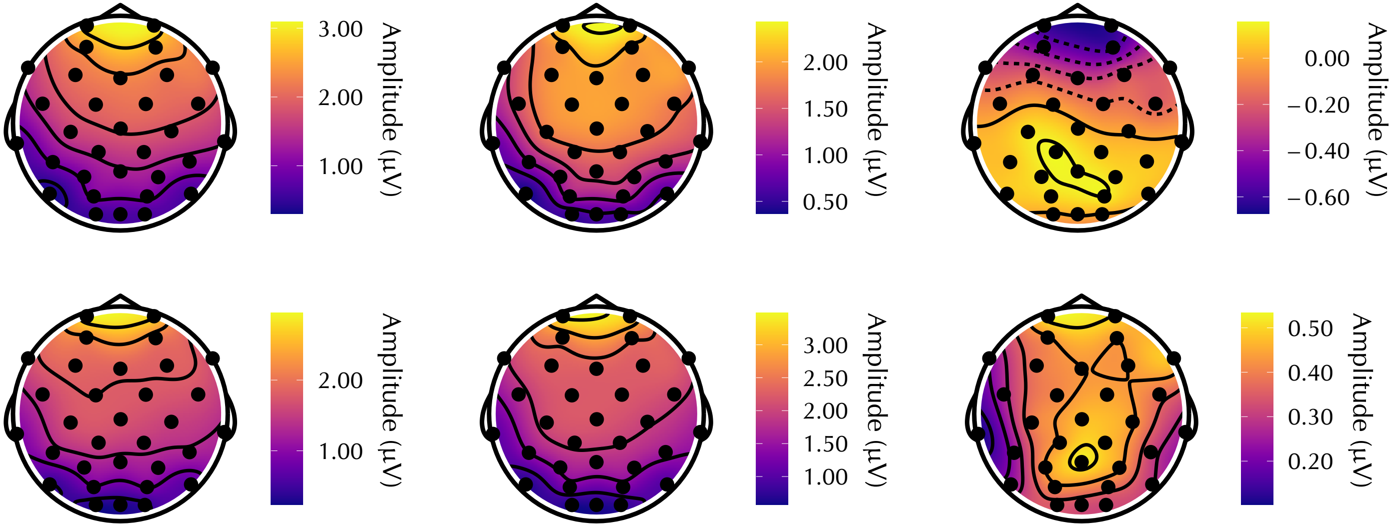

Figure 6.10: Topographical distribution of the P600 effect. The top three heads show the FDS, and the bottom three heads show the NDS. From left to right, both rows display, respectively: the NDS-allophone words; the non-NDS-allophone words; the difference between the two.

6.5 Discussion

The reaction-time data do not show any meaningful results for the research question. An expected effect of identity priming, with a plausible effect size of 47 ms facilitation, was found, but no significant difference in this facilitatory effect was found between the NDS and non-NDS allophones. What was found, however, was that the FDS were in general slower responders than the NDS, by approximately 41 ms, an effect that is of similar magnitude to the identity-prime effect. This is in line with similar findings on accented-speech perception by Floccia et al. (2009) and Floccia et al. (2006).

One of the two main findings of this study is the difference in N400 amplitudes between the FDS and the NDS over the three sessions of the experiment. The magnitude of the N400 decreased in both groups over the three sessions, but did less so for the FDS than it did for the NDS, with an effect size of $-1.17\,\mu V$. The aforementioned findings by Dambacher et al. (2006) relating the N400 to general familiarity can explain this result: the FDS had less experience with Netherlandic-Dutch speech than the NDS did, and therefore were not as strongly facilitated in sessions 2 and 3 by their previous experiences with session 1. This result mirrors the behavioral findings by Floccia et al. (2009), who show that the processing impairment incurred by accented stimuli does not improve with more exposure. The present study extends this finding, by showing that it has an electrophysiological correlate in the N400.

The second main finding of this study was the P600 found when the phoneme /ɛ/ was realized as [ɛː], which is an impossible realization of this phoneme (cf. Witteman et al. 2015). This phonological P600 is in line with recent papers, particularly those by Domahs et al. (2009) and Pater et al. (submitted). The effect was only found in the [ɛ∼ɛː] condition, which differed from the other conditions in one way, namely that the [ɛː] is not just *phonologically* illegal, but also does not exist as an allophone of *any* phoneme in either NDS or FDS, making this condition most similar to Witteman et al. (2015). This sheds new light on the phonological P600 found by Domahs et al. (2009) and Pater et al. (submitted). They obtained P600s for allophonic violations, but their critical conditions were phonologically neutralizing. The artificial rule violated in Pater et al. (submitted) was a voicing-agreement rule between stop consonants; the Domahs et al. (2009) study investigated phonotactic violations by, again, stop consonants. In both studies, participants' native languages (English and German, respectively) contained a full stop system, making these specific violations cross phoneme boundaries. The present study's finding of the same P600 effect in the [ɛ∼ɛː] condition, but not the allophonic-violation conditions, implies that the phonological P600 is restricted to these neutralizing violations.

What remains to be discussed is the finding that the P600 was only observed in the NDS, and not in the FDS. The most likely explanation is that this is due to the FDS being less familiar with Netherlandic Dutch, and therefore being less disturbed (or not significantly more than their baseline levels) by the [ɛː] realizations. This interpretation is supported by the finding that their N400 amplitudes decreased less steeply over the course of the three sessions. Note however that, while it was smaller, the FDS group still showed a decrease in N400, just as the NDS group did. This suggests that long-term accommodation is possible, and that the FDS may simply require more exposure. The present study cannot shed any light on possible future long-term accommodation by the FDS to the Netherlandic-Dutch accent in general, because the present set-up cannot distinguish between accommodation to the differences in accent and accommodation to this specific experiment. Further research, with diverse stimuli over the multiple sessions, is necessary.

The differences between the NDS and the FDS have implications for our knowledge of the neural processing of on-going historical phonological change. The FDS, who serve as a proxy for a more conservative stage of Dutch, showed increased N400 amplitude by the three sessions and did not show significant P600 modulation in the condition where it was found for the NDS. These findings suggest that the present study successfully managed to elude the robust perceptual compensation mechanisms discussed in the Introduction. It additionally supports the logical assumption that this elusion is not permanent: while the FDS' N400s did not shrink as much as the NDS's did, they did shrink nonetheless. It is conceivable that eventually, the two groups' N400s would come to coincide, which may be the point at which an on-going change can be considered to have been adopted. The finding of the P600 in the [ɛ∼ɛː] condition for the NDS only further specifies the conditions under which this phonological P600 can be elicited. Of historical phonological change, this result implies that on-going sound change in the form of new allophonic variation is processed more subtly by the human perceptual apparatus than a phoneme merger or split.

The present study is not without its limitations. One difference between the manipulation in the present study vs. the manipulation used by Witteman et al. (2015) is that they used cross-spliced speech, while the present study used natural speech produced by a trained phonetician. A small but critical difference between the present study and Witteman et al.'s (2015) means that this study is not as critically reliant on splicing as theirs was. This is the fact that, in contrast to Witteman et al. (2015), the present study used entirely native Dutch material: even the [ɛː] realizations correspond to perfectly fine Dutch phones, found in normal Dutch speech as the realizations of /ɛi/ before coda /l/. Nonetheless, the gain in naturalness of the speech material due to the absence of splicing artifacts might have been offset by a loss in naturalness of the manipulated stimuli realized by the speaker. While none of the participants commented on certain stimuli sounding artificial, and the speaker was phonetically trained and hence used to the task of realizing particular stimuli, this can be seen as a point of criticism in the design. In addition, the paradigm used in the present experiment—listening to single words and reading words aloud as a cover task—was one chosen out of convenience, this study being an exploratory part of a larger project for which this set-up was advantageous. Finally, the sample size—23 vs. 26 EEG recordings, but only 2 × 10 participants—in the present experiment was comparatively small. In sum, the effects found in this pilot experiment are in need of independent replication. I recommend that these findings be re-investigated in different languages using different paradigms, to ascertain whether the effects are cross-linguistic reflections of the processing of phonological differences between accentual varieties, or whether they are specific to these varieties of Dutch, or even to this specific task.

6.6 Conclusion

This study identified two electrophysiological correlates of accent processing and the processing of on-going phonological change in Netherlandic-Dutch and Flemish-Dutch listeners to (experimentally-controlled) Netherlandic-Dutch speech. The amplitude of the N400, measuring listeners' familiarity with the general accent in which the experimental stimuli were spoken, was decreased for the Flemish-Dutch group compared to the Netherlandic-Dutch group, indicating that they were more cautious in applying their predictive-processing abilities to the unfamiliar accent (*pace* Goslin, Duffy, & Floccia 2012). This effect decreased over the three measurement sessions, indicating that over the course of nine months, the Flemish-Dutch participants became more familiar with Netherlandic-Dutch speech, or at least the Netherlandic-Dutch speech of these experimental stimuli. In addition, a P600 effect was found for a very specific violation, viz. the realization of /ɛ/ as the illicit realization [ɛː], in the Netherlandic-Dutch group of listeners only. This shows that the brain is capable of detecting this specific type of violations (viz. violations that cross a phoneme boundary, *pace* Witteman et al. 2015), but only after sufficient familiarity with the general accent is achieved (as implied by the significant difference in sensitivity to this violation between the Flemish-Dutch group and the Netherlandic-Dutch group).

Inherent limitations to this exploratory study, particularly concerning the number of participants and the way in which the stimuli were created, mean that the results of this experiment need to be subjected to independent replication using different languages and paradigms before any definitive conclusions should be drawn. The present pilot experiment, however, has taken the first steps toward an electrophysiological investigation of the processing of on-going historical phonological change.

Data availability

The datasets analyzed for this study are available on request to the corresponding author. The datasets used in the statistical analyses are also included in the Supplementary Materials.

Ethics statement

This study was carried out in accordance with the recommendations of the Ethics Code for linguistic research in the faculty of Humanities at Leiden University with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Leiden University Center for Linguistics.

Author contributions

CV and CL conceived the experiment, designed the stimuli and edited the manuscript, CV performed the experiments, analyzed the data and wrote the manuscript.

Funding

This work is part of the research programme *Watching Dutch Change* with project number PGW-15-15, which is (partly) financed by the Dutch Research Council (NWO).

Acknowledgments

We thank Dr. Yiya Chen for valuable discussions on the design of the experiment and comments on the manuscript.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2019.00546/full#supplementary-material

CHAPTER 7

Conclusion

7.1 The Polder shift and its adoption

The goal of this dissertation was to find out what factors influence the adoption of sound change, using the Polder shift as a particularly suitable case study of on-going sound change. Chapter 2 investigated the diatopic diffusion of the Polder shift, which was the subject of the first research question in this dissertation: **what is the synchronic diatopic diffusion of the sound changes involved in the Polder shift?** This question was important for practical reasons—to be able to select representative participants for the experiments in the following chapters—but also for theoretical reasons, namely to get a clearer picture of the natures of the four sound changes that are involved in the Polder shift. The corpus study revealed that the phonetic changes diphthongizing /eː,øː,oː/ and lowering /ɛi,œy,ɔu/ are Neogrammarian, whereas the phonological change blocking diphthongization before coda /l/ is based on exemplars. The fourth change, the vocalization and retraction of coda /l/, is of indeterminate status. The synchronic diatopic variation of these changes was found to be mostly homogeneous, with the four changes having all but completed in the Netherlands, but having reached very little of Flanders.

These results paved the way for Chapter 3, the purpose of which was to answer the second research question: **(how) do sociolinguistic migrants adopt the Polder shift?** The results from the teacher-corpus study suggested that the ideal participants would be sociolinguistic migrants from Flanders to the

Polder-shift area. Chapter 3 followed ten such sociolinguistic migrants for a total duration of nine months after they had just moved to the Netherlands, and used psycholinguistic experiments to investigate whether and how much they adopted the Polder-shift changes and a control sound change (the realization of coda /r/ as [ɹ]). The results reproduced the differences found in the teacher-corpus data, but were not found to diminish over time; in other words, adoption of the Polder shift by the sociolinguistic migrants was not found.

Chapter 4 followed up on Chapter 3 by investigating the hypothesis that the negative findings in Chapter 3 were not due to failure of the experiment, but because nine months' time may simply not have been long enough to adopt the Polder-shift changes. Previous research has shown that similar changes can be adopted in approximately the same time frame (Evans & Iverson 2007), and hence a larger-scale followup was strongly warranted. The research question answered in Chapter 4 is: **which individuals, after how much time, are more likely to adopt the Polder shift?** This research qustion was answered by means of a large-scale cross-sectional comparison. For this chapter, 18 sociolinguistic migrants were found, who had lived in the Netherlands for various amounts of time ranging from three years to more than two decades. They were compared to suitable control groups of 45 individuals who had lived in the Randstad area of the Netherlands their whole lives and 43 individuals who had lived in Belgium their whole lives. Mixed results were found between production and perception. In production, the migrant group as a whole had moved to be positioned precisely in between the Netherlandic and Flemish control groups. An analysis at the level of the individual showed that this effect was driven the most strongly by ten of the eighteen migrants, who had adopted the Polder shift to such an extent that a cluster analysis grouped them with the Netherlandic control group, rather than the Flemish one. These findings did not directly carry over to the perception data: in perception, expected group differences were found similar to those in production, but an analysis at the individual level did not yield clear results, although its results were partially correlated with the individual-level results in production. This agreed with findings by Evans & Iverson (2007), and suggested that it is not so much the *type* of sound change (phonological change vs. phonetic change) that determines its adoption and subsequent propagation by individuals, but rather its mode of transmission (system-internal vs. contact-driven).

Chapters 5 and 6 probed more specific aspects of the (non)adoption of the Polder shift by the same ten sociolinguistic migrants from Chapter 3 by means of two ERP experiments. Together, they answer the fourth and final research question: **(how) is the adoption of the Polder shift reflected in ERPs?** Chapter 5 resolved an open problem regarding the relative roles of perception and representation in the adoption of sound change: is sound change based in misperception (e.g. Ohala 1981) or in misrepresentation (Beddor

2009, Hamann 2009)? The results of this chapter show that the sociolinguistic migrants start out having weaker knowledge about the phonological distribution of the [eː∼ei] allophone difference, although they do catch up to the Netherlandic level between the two sessions of the experiment, respectively corresponding to their fourth and their ninth months in the Netherlands. In addition, the Flemish participants did not find the [ɹ] realization of the rhotic as attention-grabbing as the Netherlandic participants did. It was argued that these results reflect phonological and sociolinguistic knowledge, rather than differences in phonetic processing. Thus, they do not lend support to the misperception account of sound change (which is characterized by differences in phonetic processing) but can be explained more readily in terms of the accounts by Beddor (2009) and Hamann (2009) based on differences in cue weighting (in this case, differences in the weighting of phonological and sociolinguistic information).

Chapter 6 used a more exploratory ERP experiment. As with Chapter 5, this experiment compared the perception of Netherlandic and Flemish realizations of the vowels involved in the Polder shift, plus two control conditions consisting of the rhotic and a phonologically (as opposed to just sociolinguistically) illicit vowel realization (/ɛ/ realized [ɛː]). Again, the Polder-shift changes produced no significant ERP differences, nor did the rhotic, but the /ɛ/ control condition resulted in a P600, in the control group only. The P600 was argued to be a logical extension of previous work, particularly by Witteman et al. (2015), who found a behavioral slowdown in similar conditions involving regional accents where a vowel's accented realization crossed a phoneme boundary. If the P600 indeed represents the electrophysiological precursor to this behavioral result, this automatically explains why the same P600 was not found in the Polder-shift conditions or in the rhotic: these all involved changes in *allophones*, not phonemes.

7.2 From compensation to adoption

The results from Chapter 3 and Chapter 4 portray an important contrast. While Chapter 3 failed to show any credible adoption of the Polder shift by the Flemish group in nine months' time, Chapter 4 showed that after multiple decades, some of the Flemish participants in that study had adapted. Taken together, Chapters 3, 5, and 6 provide a chain of evidence that explains what steps are to be taken in the process of adopting an on-going sound change. These chapters discuss the same participants in different experiments. The longitudinal experiments in Chapters 3 and 6 were performed after roughly one, five, and nine months after arrival of the Flemish participants in the Netherlands. The experiment in Chapter 5 was performed during the latter two of these three occasions: around month five and around month nine. Chapter 5 provided evidence that suggests that there are three independent sources of knowledge to be acquired: listening competence, phonological knowledge, and knowledge about sociolinguistic evaluation. The findings suggest that these three types of knowledge are acquired in a particular order: listening competence first, phonological knowledge secondly, and sociolinguistic evaluation thirdly. This corresponds to these three skills being located at increasing levels of grammatical abstraction. The presence of MMNs indicates that listening competence is already in place by four months after arrival; since these are sound changes and not second languages, it is not unreasonable to assume that this was never an issue to start with for the Flemish students. Specifically concerning the [eː∼ei] allophone pair, passive (receptive) phonological knowledge is acquired sometime between months four and nine. For the rhotic, the required allophonic knowledge appeared to be already in place (as evidenced by the Flemish group reaching the same mismatch negativities as the Netherlandic group in Chapter 5, but also by their high proportion of [ɹ] responses in the perception task in Chapter 3). The Flemish group's sociolinguistic knowledge of the rhotic, however, was not yet completely Netherlandic-like: the rhotic captured less attention (evidenced by a shift of the MMN away from the frontal pole compared to the Netherlandic controls), suggesting that this sound may not have imparted the same sociolinguistic salience in the Flemish group as it did in the controls.

The EEG task in Chapter 5 found that the grammatical knowledge of the Polder shift was largely in place for the Flemish participants, and improved even further by their ninth month of living in the Netherlands. It is telling, then, that this adoption of the Polder shift found electrophysiologically was in no way reflected in the behavioral experiments in Chapter 3, where the differences between the two participant groups were found to be robust and persistent. It was only after much more time than nine months, viz. the timespan covered by Chapter 4, that the same adoption of the Polder shift was also demonstrated behaviorally. The results from Chapters 4 and 5 have shown that these participants are certainly able to acquire these changes eventually, so then there must be competence-extrinsic factors at play which cause the behavioral adoption of the Polder-shift changes to proceed more slowly than the changes in, for example, Evans & Iverson (2007), which departed from a very similar situation and did find adoption of relevant accent differences. Chapter 3 argued that the reason can only be that nine months is simply not enough time to adopt these language changes in a behaviorally-detectable way.

It is possible that the reason for this lies in the phonological status of the changes of the Polder shift. The Polder-shift changes do not result in phonemic mergers or splits, which could well have reduced the pressure on the Flemish participants to adopt them behaviorally. Chapter 6 provides concrete evidence of this type. This exploratory study did not reveal long-term changes, but this

Table 7.1: Log_{10} Bayes factors for the hypothesis that the P600 difference between the groups equals or exceeds that found in the [ɛ∼ɛː] condition from Chapter 6. Negative values indicate evidence against this hypothesis.

Condition	**Bayes factor ($\log_{10}$)**
[ei∼eː]	−1.93
[øy∼øː]	−1.00
[ou∼oː]	−2.18
[ɛi∼ɛː]	−2.03
[œy∼œː]	−1.57
[ɑu∼ɑː]	−3.10
[aːɹ∼aːʀ]	−2.62
[ɛ∼ɛː]	0.00

was because it turned out that the task used in the experiment (which was simply the production task from Chapter 3) was not sensitive to allophonic violations. This was concluded because a control condition with a phonemic violation, viz. the realization of /ɛ/ as *[ɛː], elicited a clear P600, but the experimental conditions involving the Polder-shift changes did not.

Note that this null result does not yet prove that the manipulation used in Chapter 6 does not elicit a P600 for non-phonemic changes; it only shows that this was not found in the experiment reported in that chapter. However, a reanalysis of that chapter's results in a Bayesian framework confirms that the other conditions indeed *did not* elicit a P600, rather than simply having failed to do so. For this analysis, each condition was averaged over the 560–660 ms window, just as the [ɛ∼ɛː] condition was. R package `brms` (Bürkner 2017, 2018) was used to fit a model containing the same terms as in Chapter 6's model for the [ɛ∼ɛː] condition, with fixed-effect priors set to the same values as those obtained from that model. Table 7.1 shows $\log_{10}$ Bayes factors for the hypothesis of a P600 difference between the Flemish-Dutch and Netherlandic-Dutch students of at least −2.51 µV. For the [ɛ∼ɛː] condition, the evidence for this hypothesis is exactly as strong as the prior (which is obvious, given that the priors were set to precisely this condition's Chapter-6 results), but all of the other conditions provide "strong" (Jeffreys 1961) to "decisive" (Jeffreys 1961) evidence that there is *no* P600 of at least this magnitude to be found.

Thus, the results from Chapter 6 show that participants compensate for allophonic variation in on-line auditory speech processing, and fail to do so when this variation crosses a phoneme boundary (this was the case for the [ɛ∼ɛː] condition, which elicited the P600). For the allophonic changes that constitute the Polder shift, the P600 is not informative: it was absent in both the

NDS and the FDS, despite participants' ability to perceive these changes (Chapter 5) and them not (yet) having had adapted to them behaviorally (Chapter 3). However, the P600 may prove to be informative for sound changes involving phonemic mergers or splits. The significant NDS–FDS difference in P600 amplitude in the [ɛ∼ɛː] condition shows that the FDS were not yet familiar enough with NDS speech in general to perceive the phonemic violation. The FDS' diminished N400 amplitudes compared to the NDS's point in the same direction. This identifies an avenue for future research: replicate the experiment using ongoing sound changes involving phonemic mergers or splits, instead of the allophonic changes that are central to the Polder shift. For those kinds of changes, the N400 and P600 could be informative of the degree to which the phonemic change has been adopted.

7.3 Salience

Summarizing the previous section, the results of this dissertation imply an important role for *salience*. This is a well-known term within sociolinguistics—in fact, it is the foundation of Trudgill's (1986) change-by-accommodation model (cf. the results in Chapter 3, which are somewhat problematic for this model)—but it has not been well-defined: see Rácz (2013) for a discussion. Auer, Barden, & Grosskopf (1998) observe why salience is so hard to define: salience is a highly subjective attribute, driven by personal-evaluative factors rather than objective phonetic or phonological parameters. Accepting this, this dissertation has oftentimes stated that the rhotic changes are more salient than the Polder-shift changes, referencing Sebregts (2015) when doing so. The results from this dissertation support this distinction, and also support an explanation in terms of different types of salience.

Chapter 5 found that the Flemish participants' perception of the [eː∼ei] allophone distinction changed over time, but their perception of the rhotic distinction did not. In Chapter 6, the non-NDS realization of /ɛ/ as [ɛː] triggered a P600, but the equally un-NDS-like realization of the diphthongal Polder-shift vowels as monophthongs before non-/l/ did not. These results show that there are two types of salience to be taken into account. The first is *sociolinguistic* salience, the well-known type which is the focus of authors like Auer, Barden, & Grosskopf (1998). This was observed in Chapter 5, where it was shown that the [əɹ] deviants impart less attentional importance for the Flemish group than they do for the Netherlandic group. This is in line with the well-known observation that, for Netherlandic-Dutch speakers, this realization of /r/ is an extremely salient sociolinguistic marker (Sebregts 2015). The results from this dissertation also identify a second type of salience: *phonological* salience. Chapter 6 showed that the P600 is sensitive to phonemic status: violations that cross phoneme boundaries elicit the P600, violations at the allophonic level do not. This may explain why Chapter 3 failed to find behavioral adoption of the Polder-shift changes: if these changes are not salient enough to the brain, why would the Flemish participants need to adopt them? It must be noted that Chapter 5 did find changes in brain responses to the [eː∼ei] allophone distinction, but this used a mismatch-negativity paradigm, which is pre-attentive, whereas the P600 is post-attentive. Thus, the P600 observed for phonemic violations in Chapter 6 may be indicative of a type of salience that is not primarily sociolinguistic, but rather phonological.

Is it correct to consider phonological category status a type of *salience*, as posited in the previous paragraph? The answer is most likely: yes, in the context of sound change a categorical phonological change may be salient. The argument is provided, indirectly, by Janson (1983). In his study of the sound change from [r] to [ʀ] in Norwegian, Janson argues that this change must have been a change in the underlying form. Note that such a change is not necessarily phonemic: in Janson's (1983) case, it is not a change in the meaning-distinctive-category system of the language, but rather a rule inversion of the type discussed by Hyman (1976). Such changes need to be sufficiently disruptive to the phonological system in order to be adopted, i.e. a change is more likely to be adopted if it is phonologically salient. The realization of /ɛ/ as [ɛː] in Chapter 6, which elicited the aforementioned P600, is exactly the type of change discussed by Janson (1983). Of the Polder shift, however, we know that it does not meet this criterion of phonological salience: Chapter 2 showed that these changes are either Neogrammarian or exemplar-based, not underlying-form changes. It is possible that this is why these changes did not elicit a P600 in Chapter 6, and why they could not be shown to be adopted within nine months' time in Chapter 3. Their eventual adoption after a longer amount of time has passed (Chapter 4) may then be due to some of the sociolinguistic migrants eventually noticing the differences between their productions and the input they receive (for which the cognitive machinery is already in place; Chapter 5). In this case, these individuals adopt the sound change not because it is intrinsically salient (as was the case for the realization of /ɛ/ as [ɛː] in Chapter 6), but because it has become salient *for them*.

7.4 Methodology

Section 1.3.2 discussed how psycholinguists have profited from methodological innovations. This dissertation has demonstrated how the study of ongoing sound change can reap the same benefits. Specifically, this dissertation has made three types of advances: new methods were developed, old methods were given new uses, and—where necessary—this dissertation used methods beyond the traditional linear-regression model (which underlies, e.g., ANOVA). It is worth reflecting on how these innovations have aided the interpretation of the data collected in this dissertation, and how they could be applied in other settings than the adoption of the Polder shift.

The only *new* method in this dissertation was Chapter 2's use of the generalized additive mixed model (henceforth "GAMM") to avoid having to segment the gradient boundary between a vowel and a following coda /l/. The backbone of segmental acoustic phonetics has always been the ability to isolate the segment under investigation from the surrounding context, a requirement which is impossible to meet when the segment transitions are fully gradient. This is the case when the coda /l/ is strongly vocalized, which was found to be the case for the Netherlandic varieties of Dutch, but not the Flemish ones (see Section 2.3.4). Chapter 2 used GAMMs to model these VC trajectories as they are, without requiring an *a priori* manual segmentation.

The second type of methodological innovation in this dissertation was the use of well-established methods, but in novel ways. The most significant example is given in Chapter 4. Chapter 4 made use of the mixed-effects model—a completely uncontroversial statistical technique—to capture individual variation. This is an unconventional use of the mixed-effects model: traditionally, random-effect parameters are considered nuisance variables, to be incorporated into a model to absorb variation between participants and items that may interfere with the group-level patterns which are of primary interest. Chapter 4 demonstrates that, for these data, that approach would have been naïve: when the groups are not perfectly homogeneous, as was the case for the sociolinguistic-migrant group in Chapter 4, an analysis that collapses each group into a single β value will misrepresent the data. The data in Chapter 4 were analyzed more appropriately by excluding the (misleading) fixed factor for "Group" from the model, and reconstructing the groups *a posteriori* on the basis of the empirical BLUPs from a full random-slope model. A cluster analysis on the resulting by-participants b coefficients revealed that the sociolinguistic-migrant group was split between more and less innovative participants, which *together* caused the group as a whole to move in between the Flemish and Netherlandic control groups. The degree of adoption of the Polder shift was quantitative, i.e. the sociolinguistic migrants did not differ in whether or not they had adopted the Polder shift categorically, but in the *degree* to which they had done so. If this degree exceeded 0.07 Lobanov units, participants had adopted the Polder shift to a sufficient extent that they were classified as Netherlandic rather than Flemish. Within sociolinguistics, this BLUP-based approach to studying individual variation has slowly begun to take off (Drager & Hay 2012, Tamminga to appear); this dissertation's use of cluster techniques to shed more light on the meaning of these individual differences is only an additional step in these on-going methodological developments.

The interpretation of the data discussed in this dissertation has been made significantly more feasible by the adoption of statistical methods beyond the simple linear-regression model. This was the case in Chapters 3, 5, and 6. Chapters 3 and 5 both dealt with situations where an extremely large number of all-categorical predictors would be required if these chapters' perception data were to be analyzed using traditional regression models. In Chapter 3, the maximal model for a logistic-regression analysis would have contained 48 regression coefficients for each condition, including a four-way interaction "Group (2 levels) × Session (3 levels) × Following consonant (2 levels) × Step (4 levels)". In Chapter 5, the maximal model for an ANOVA like the one in Grosvald & Corina (2012) would have contained 72 coefficients for each of the six conditions, among which is a five-way interaction "Deviant (2 levels) × Group (2 levels) × Session (2 levels) × Hemisphere (3 levels) × Anteriority (3 levels)". Such interactions are impossible to interpret. Section 3 resolved this by making use of mixed-effects regression trees to model exactly this type of data in a much more interpretable way. Similarly, Section 5 resorted to GAMMs, which not only removed the need for "Hemisphere" and "Anteriority" factors, but also provided a much more fine-grained overview of the data in the first place. To the author's knowledge, this has not been done in linguistics before.

Methodological innovations can take place not just in the actual modeling process itself, but also in the reasoning about models that are possible for a given dataset. In Chapter 5, six separate GAMMs were run corresponding to the six different conditions in the experiment, but the MMN ERP was only expected in three of them, given the normally asymmetric nature of this ERP component (Cornell, Lahiri, & Eulitz 2011, Lahiri & Reetz 2010). Therefore, a way was needed to be able to argue not just *against* the null hypothesis, but also in its favor. For this reason, models were evaluated in a Bayesian framework, though still using maximum-likelihood estimates for reasons of computational feasibility, using the approach by Wagenmakers (2007). Bayesian reasoning brings with it a completely different way of thinking—evaluating the likelihood, $p(\beta|y)$, rather than the p-value, $p(y|\beta)$—which is not yet commonplace within linguistics. As another example, Chapter 6 was an exploratory ERP investigation, and hence it was not known *a priori* what ERP component, if any, would be obtained. This is a solved problem in the field of cognitive neuroscience, where permutation testing (Maris & Oostenveld 2007) is the canonical answer. This nonparametric statistical test made it possible to identify both the temporal window and the spatial ROI in which robust differences due to the various manipulated factors arose. This led to the identification of a new putative marker of phonological status, viz. the P600, which can be triggered by phonological violations that cross phoneme boundaries, but could not be detected for within-category violations.

7.5 Conclusions

This dissertation has investigated the adoption of sound change, and the role played by synchronic and diachronic processing of variation within that process. The investigation focused on the Polder shift, an on-going vowel shift in Dutch that has all but completed in the Netherlands, but has not taken place in Flanders (Chapter 2). The adoption of the Polder shift by Flemish sociolinguistic migrants proved to be difficult to detect in the medium term (nine months) using behavioral methods (Chapter 3), but was detected behaviorally in the long term (multiple decades; Chapter 4). Despite the lack of reliable medium-term behavioral evidence, Chapter 5 found robust evidence for the start of sound-change adoption using an EEG experiment. Chapter 6 attempted to extend these findings using a different, novel, paradigm, and found that this was successful, but only for changes that were large enough to cross a phoneme boundary, i.e. not the Polder shift.

The future of sociolinguistics must be sought in the continuing integration of the five fields of historical phonology, sociolinguistics, psycholinguistics, neurolinguistics, and statistics. The different findings from this dissertation demonstrate how the fundamentally sociolinguistic phenomenon of historical sound change can be studied empirically using psycho- and neurolinguistics. The dissertation has additionally demonstrated on several occasions how much the study of on-going sound change can profit from the continuous innovations in the field of methodology and statistics. These have made it possible to analyze data that would previously have been considered unanalyzable (Chapter 2), and have additionally made it possible to draw new conclusions (Chapters 3, 4, 5, and 6). From this dissertation alone, a few possible avenues for future research present themselves. Chapter 2 concluded with the remark that synchronic evidence cannot distinguish between a sound change that is phonetically abrupt and a sound change that has been phonetically gradual but has already completed. The clear Netherlandic–Flemish split on the effect of coda /l/ on the preceding vowel is one such case: future diachronic research is needed to chart exactly how the Netherlandic F2 retraction before coda /l/ developed. An additional remark that was made on the data in Chapter 2 was the low number of words available in the corpus, which might have caused the lexical diffusion of the Polder-shift changes to have been underestimated. Ample future options for synchronic research like Chapter 2's corpus study present themselves here.

As mentioned before, Chapters 3 and 4 supplement each other, in that the former chapter did not find adoption of the Polder shift after nine months, but the latter chapter did find it after multiple years. The obvious research question following from this discrepancy is: when *does* adoption take place? This is not an easy question, especially as Chapter 4 showed that adoption, at the group level, is gradient, rather than categorical. In addition, this dissertation demonstrated that whether or not one finds adoption of a sound change strongly determines on how one defines "adoption": behaviorally, Chapter 3 did not find any clear adoption of the Polder shift after nine months, but electrophysiologically, Chapter 5 did reveal significant changes. For these two reasons, I would discourage future researchers from devoting their time to the pursuit of broad questions such as "when does adoption take place?". Instead, a more focused and more-thoroughly-operationalized question such as "what is the earliest point in time at which a single sociolinguistic migrant adopts the Polder shift in single-word production?" is more likely to result in positive research outcomes; in the case of this example question, it would identify the empirical amount of time that is *minimally* necessary to adopt the Polder shift. Another point which would be interesting for future study is the question "what is the earliest point in time at which 21.7% of sociolinguistic migrants have adopted the Polder shift in single-word production?", where 21.7% is the critical mass calculated by Yang (2009) for an individual-level sound change to secure its evolution into a group-level sound change.

The neurolinguistics of sociolinguistic variation also deserve further exploration. Chapter 6 revealed a P600 for variation that was particularly salient, viz. crossing the boundary of a phonemic category. This could prove to be a new method for detecting the status of a sound change in progress, with phonemic mergers or splits putatively eliciting a P600, but this needs to be established by research specifically looking into this ERP component as an indicator of phonological status. In addition, this ERP may be the starting point for an objective definition (cf. Auer, Barden, & Grosskopf 1998) of the vague notion of "salience". Future research could explore this further. The same is true for the MMN, of which the topographical distribution was argued in Chapter 5 to index a type of sociolinguistic salience. This, too, needs to be investigated in a more specific manner.

Future research could also proceed from this dissertation in the direction of new methods for investigating linguistic variation. Chapter 2's application of the generalized additive model could be combined with dynamic time warping (Shi et al. 2015) to develop new tools that could aid phoneticians in determining empirical boundaries for speech sounds that are difficult to segment. The method used to investigate individual differences in Chapter 4 could also be extended to GAMMs to directly study individual variation in more complex signals than Chapter 4's point measures. Steps in this direction have already been taken by, for instance, Tamminga, Ahern, & Ecay (2016). It should go without saying that GAMMs in general offer new ways of analyzing data that would have been challenging to analyze in a more traditional way; Chapter 5's use of GAMMs to avoid fitting 6-way interaction models, by smoothing over the entire topographical area present in the data (rather than including many-leveled factors for "Hemisphere" and "Anteriority"), is a prime example.

APPENDIX A

Prime–target list for Experiment 1 from Chapter 3

baube – zaube
baube – zobe
bauke – zauke
bauke – zoke
bebbe – webbe
bebbe – wijbe
begge – wegge
begge – wijge
bekker – hekker
bekker – hijker
bemme – zemme
bemme – zijme
benne – henne
benne – hijne
bezze – hezze
bezze – hijze
blaalder – zaalder
blaude – graude
blaude – grode
blaulde – saulde
blaulde – soolde
blaulder – waulder
blaulder – woolder
blauver – schauver
blauver – schover
blebe – zebe
blebe – zijbe
blege – bege
blege – bijge
blelder – zelder
blelder – zijlder
blene – hene
blene – hijne
blete – lete
blete – lijte
bleter – leter
bleter – lijter
braalder – schaalder
brauver – vauver
brauver – vover
bredde – stedde
bredde – stijde
bredder – tedder
bredder – tijder
breelder – veelder
breelder – vijlder
brelder – welder
brelder – wijlder
brette – tette
brette – tijte
bretter – letter
bretter – lijter
brever – tever
brever – tijver
brevver – hevver
brevver – hijver
bwelde – slelde
bwelde – slijlde
dauver – trauver
dauver – trover
debe – webe
debe – wijbe
dete – zete

dete – zijte
deter – zeter
deter – zijter
devver – levver
devver – lijver
drever – rever
drever – rijver
dweelder – bleelder
dweelder – blijlder
dwever – krever
dwever – krijver
faalder – haalder
fauver – jauver
fauver – jover
faver – taver
fedde – sedde
fedde – sijde
fedder – kedder
fedder – kijder
fevver – wevver
fevver – wijver
flaalder – waalder
flaulde – maulde
flaulde – moolde
flaulder – braulder
flaulder – broolder
fleelder – seelder
fleelder – sijlder
flelder – selder
flelder – sijlder
fraalder – saalder
freelder – weelder
freelder – wijlder
gader – tader
gauver – nauver
gauver – nover
gleelde – kreelde
gleelde – krijlde
glelde – flelde
glelde – flijlde
graalder – taalder
grauver – hauver
grauver – hover
greelde – treelde
greelde – trijlde
greelder – teelder
greelder – tijlder
grelde – trelde
grelde – trijlde
grelder – telder
grelder – tijlder
grever – hever
grever – hijver
grevver – revver
grevver – rijver
gweelde – fleelde
gweelde – flijlde
hader – sader
jaube – haube
jaube – hobe
jauge – nauge
jauge – noge
jauter – nauter
jauter – noter
jauze – nauze
jauze – noze
jede – krede
jede – krijde
jevver – sevver
jevver – sijver
kebe – tebe
kebe – tijbe
keder – zeder
keder – zijder
kegge – begge
kegge – bijge
klaalder – draalder
klaulde – faulde
klaulde – foolde
klaulder – laulder
klaulder – loolder
klauver – vrauver
klauver – vrover
kleelder – geelder
kleelder – gijlder
klelder – kwelder
klelder – kwijlder
kneelde – pleelde
kneelde – plijlde
kraalder – raalder
kraulder – naulder
kraulder – noolder
kraver – paver
kreder – deder
kreder – dijder
kreelder – feelder
kreelder – fijlder
krelder – nelder
krelder – nijlder
kwaulde – gaulde
kwaulde – goolde
kwaulder – fraulder
kwaulder – froolder
kwaver – maver
kweelder – zeelder
kweelder – zijlder
laalder – kaalder
lauder – zauder
lauder – zoder
lebbe – zebbe
lebbe – zijbe
mauder – jauder
mauder – joder
meker – heker
meker – hijker
nauve – zauve
nauve – zove
nebe – bebe
nebe – bijbe
neke – heke
neke – hijke
nete – tete
nete – tijte
neter – teter
neter – tijter
pader – jader

pauder – bauder
pauder – boder
pevver – tevver
pevver – tijver
pfelde – klelde
pfelde – klijlde
pjaalder – staalder
pjelde – xelde
pjelde – xijlde
plaalder – vaalder
plaulde – vaulde
plaulde – voolde
plaulder – praulder
plaulder – proolder
plauver – mauver
plauver – mover
pleelder – jeelder
pleelder – jijlder
plelde – nelde
plelde – nijlde
plelder – jelder
plelder – jijlder
plever – pever
plever – pijver
praalder – paalder
praver – saver
preelder – steelder
preelder – stijlder
prelder – stelder
prelder – stijlder
prever – twever
prever – twijver
psaulder – draulder
psaulder – droolder
pselde – lelde
pselde – lijlde
rebe – hebe
rebe – hijbe
sauver – bauver
sauver – bover
schaube – kaube
schaube – kobe
schauge – kauge
schauge – koge
schauke – nauke
schauke – noke
schaulder – jaulder
schaulder – joolder
schaver – kaver
schevver – zevver
schevver – zijver
schraalder – jaalder
sedder – zedder
sedder – zijder
sfaulde – praulde
sfaulde – proolde
sfeelde – preelde
sfeelde – prijlde
sfelde – blelde
sfelde – blijlde
sjeelde – vreelde
sjeelde – vrijlde
skeelde – freelde
skeelde – frijlde
skelde – twelde
skelde – twijlde
skelder – twelder
skelder – twijlder
slaalder – kwaalder
slaulde – baulde
slaulde – boolde
slaulder – taulder
slaulder – toolder
slaver – waver
sleelder – peelder
sleelder – pijlder
slelder – dwelder
slelder – dwijlder
smaalder – traalder
smeelde – bleelde
smeelde – blijlde
smelde – drelde
smelde – drijlde
smelder – drelder
smelder – drijlder
snaulde – draulde
snaulde – droolde
snaulder – graulder
snaulder – groolder
sneelde – sleelde
sneelde – slijlde
snever – jever
snever – jijver
spaalder – gaalder
spaulde – naulde
spaulde – noolde
spaulder – gaulder
spaulder – goolder
speelder – leelder
speelder – lijlder
spever – dever
spever – dijver
spleelde – kleelde
spleelde – klijlde
splelde – krelde
splelde – krijlde
spraulde – braulde
spraulde – broolde
spraver – zaver
spreelde – breelde
spreelde – brijlde
sprelde – brelde
sprelde – brijlde
sprever – mever
sprever – mijver
sraulde – xaulde
sraulde – xoolde
sreelde – greelde
sreelde – grijlde
srelde – grelde
srelde – grijlde
staude – gaude
staude – gode
staulde – jaulde
staulde – joolde
staulder – raulder

staulder – roolder
stauver – lauver
stauver – lover
staver – naver
stedder – hedder
stedder – hijder
stette – hette
stette – hijte
stetter – wetter
stetter – wijter
straulde – fraulde
straulde – froolde
straver – vaver
strelde – frelde
strelde – frijlde
strelder – frelder
strelder – frijlder
strevver – mevver
strevver – mijver
swaulde – kraulde
swaulde – kroolde
sweelde – leelde
sweelde – lijlde
sweelder – dreelder
sweelder – drijlder
swelde – delde
swelde – dijlde
swelder – relder
swelder – rijlder
tater – jater
taude – paude
taude – pode
tegge – wegge
tegge – wijge
tetter – hetter
tetter – hijter
tezze – zezze
tezze – zijze
tjaulder – baulder
tjaulder – boolder
tjeelde – dreelde
tjeelde – drijlde
tjelde – jelde
tjelde – jijlde
trauder – nauder
trauder – noder
trauge – kauge
trauge – koge
traulder – haulder
traulder – hoolder
traute – jaute
traute – jote
traver – javer
treelder – neelder
treelder – nijlder
trelder – belder
trelder – bijlder
trever – sever
trever – sijver
twaalder – naalder
twaulde – zaulde
twaulde – zoolde
twaulder – maulder
twaulder – moolder
twebe – lebe
twebe – lijbe
tweder – heder
tweder – hijder
tweelder – beelder
tweelder – bijlder
tweke – meke
tweke – mijke
vaube – naube
vaube – nobe
vauge – nauge
vauge – noge
vauke – hauke
vauke – hoke
vedde – kedde
vedde – kijde
vedder – ledder
vedder – lijder
vede – kede
vede – kijde
vleelde – tweelde
vleelde – twijlde
vlelde – prelde
vlelde – prijlde
vrauder – vauder
vrauder – voder
vraulde – waulde
vraulde – woolde
vraulder – vaulder
vraulder – voolder
vraver – raver
vrebe – mebe
vrebe – mijbe
vreelder – deelder
vreelder – dijlder
vrelder – delder
vrelder – dijlder
wauver – kauver
wauver – kover
weme – heme
weme – hijme
wemme – hemme
wemme – hijme
weppe – heppe
weppe – hijpe
xaulder – saulder
xaulder – soolder
xeelde – neelde
xeelde – nijlde
xelder – pelder
xelder – pijlder
zader – mader
zevve – hevve
zevve – hijve
zwaulde – graulde
zwaulde – groolde
zwaver – baver
zweelde – jeelde
zweelde – jijlde
zweelder – reelder
zweelder – rijlder
zwelder – lelder
zwelder – lijlder

APPENDIX B

Word list for Experiment 2 from Chapter 3

Australische
Engelse
Engelsen
Keuls
Keulse
L-vormig
Poolse
aan
aandacht
aantal
aard
aardappelen
aardbeien
aardbeving
aardbodem
aardbol
aarde
aardedonker
aardewerk
aardgas
aardig
aardigheid
aardkorst
aardoppervlak
aardrijkskunde
aardse
aars
aartsbisschop
aarzelde
aarzeling
afgebeuld
afgebeulde
audiovisuele
augustus
aula
auteur
authenticiteit
authentieke
autisme
autistische
auto
autobiografie
autobiografische
autobus
automaat
automatisch
automatisering
automatisme
automobiel
automobilisten
autonome
autonomie
autoritaire
autoriteiten
autoweg
basisschool
beeld
beulsknechten
beulswerk
bijl
bijvoorbeeld
deel
doolhof

echt
echter
echtgenoot
echtgenote
echtpaar
economische
een
eenheid
eenmaal
eens
eenvoudig
effect
effectief
eigen
eigenaar
eigenaardige
eigenbelang
eigendom
eigenlijk
eigenschappen
eiland
eind
einde
eindelijk
eindeloos
eindigde
eis
el
elders
eldorado
elementen
elf
elfduizend
elfenkoningin
elfhonderd
elfstedentocht
elft
elftal
elkaar
elkander
elke
elkeen
ellende
elpee
elpenbenen
els
elzehout
elzestruiken
en
energie
enerzijds
engel
enige
enigszins
enkele
enorme
enthousiasme
enthousiast
enzovoort
essentieel
eten
even
evenals
eveneens
evenmin
eventueel
evenwel
evenwicht
exemplaren
experiment
extra
gedeelte
gedeeltelijk
geheel
geheugen
gehuil
geil
geschoolde
geul
geultje
gezeuld
grotendeels
heel
heil
heilzame
heul
heus
heuvel
hoeveel
hoeveelheid
hogeschool
houtskool
huilbui
huilde
iedereen
iemand
iets
ijdelheid
ijl
ijs
ijver
ijverig
ijzer
ijzeren
ingeruild
integendeel
jeugd
keuken
keus
keuze
kleuterschool
kool
kuil
leuk
leunde
levensstijl
machinepistool
middenschool
mijl
milieu
monsieur
muil
muilezels
nerveus
neus
oefenen

oefening
oever
ogen
ogenblik
olie
onderdeel
onderwijl
onheil
onheilspellend
onverwijld
ook
oom
oordeel
oosten
openbare
opende
operatie
oude
oudejaarsavond
ouden
ouderdom
ouderen
ouderlijke
ouderling
ouders
ouderschap
ouderwetse
oudheid
oudjes
oudsher
oudste
out
output
over
overal
overeenkomst
overeenstemming
overeind
overheid
overige
overigens
overkant
overtuigd
overtuiging
parool
peil
personeel
peul
peultjes
peulvruchten
pijl
pistool
pool
puilden
religieuze
reusachtige
ruil
school
schoolmeester
schooltijd
schoolwezen
schuilging
schuilkelder
schuilnaam
schuilplaats
schuilt
serieus
sleutel
smeulde
smeulden
smeult
speelde
steenkool
steil
steun
stijl
symbool
teil
terwijl
teveel
uil
uit
uitbreiding
uitdrukking
uiteen
uiteindelijk
uiteraard
uiterst
uitgangspunt
uitgenodigd
uitgesproken
uitgevoerd
uiting
uitleg
uitsluitend
uitspraak
uitvoerig
uitvoering
uitwerking
uitzicht
uitzondering
vaargeul
veel
veelal
verzeild
viool
vleugels
voorbeeld
voordeel
vreugde
vuil
vuilnisbak
vuilnisbelt
vuiltje
zeil
zeildoek
zeulde
zeulden
zeult
zuil

APPENDIX C

Word list for Experiment 1 from Chapter 4

achteruitkijkspiegel
adieu
afstotelijk
alleenstaand
auto-ongevallen
autokerkhoven
autoritje
autosleutels
bakkerij
bedruipende
beenbeschermers
beenbreuken
beloninkje
bereid
bereikbaar
bespeel
beu
beugel
beul
beult
bevrijdingsactie
bewegingsmogelijkheden
bezopener
bijgepunt
blasfemie
boerderij
bonis
breakdancing
breedgeschouderde
brekebeen
buidelrat
buil
buis
buiswater
buitenland
buitenschools
bullshit
bureaublad
cacao
cacaoboom
degusteren
deug
deugdzaamheid
dialogen
dichtstbijzijnde
doodstraf
dreun
dromen
droogkamer
droogshampoo
dweep
eindig
eindresultaten
enigszins
entertainer
eureka
evenwicht
existentieel
familiariteit
filosoof
financieel
foyer
fundamenteel

gebruiksmogelijkheden
geelst
geestesleven
gefoezel
geilaard
geilheid
gelegenheden
gelouterde
genereus
geneugten
geul
geweigerd
gonorroe
goud
grootbrengen
heethoofd
heil
heildronk
heul
heulde
heult
heup
heus
heuvel
hogelijk
holocaust
hoofdrol
hoog
hoogmoed
houder
hout
houtbewerker
houthandelaren
houtig
houtindustrie
houtluis
houtsnijder
houtsnijwerken
houtwerker
houweel
huichelaar
huiskamerbijeenkomst
huisnummer
huisschilders
iel
ijlkoorts
ijsbeer
immoreel
kaneel
kannibaals
kleuter
kleverigheid
kloostermuren
knuisten
kokosvlees
komen
koolblad
koolsoep
koolzuurgas
koopziek
kreek
kwartaal
kwartaalblad
kwijl
langszijde
leid
leidinggevende
lepelaar
leugenaar
levendig
lijk
lijntrekkers
lou
louter
luistervinkte
makreel
manipulatie
mantilla
medisch
meegesleept
meegevoerd
meekwamen
meervoudige
meinedig
meineed
mitotisch
morfologische
necrofiel
nijlpaard
nijpender
nobelere
noordpoolcirkel
officieus
omgeving
onderkoeld
ondersteun
onophoudelijk
onpeilbaar
ontboden
ontleedtafel
ontsteking
onverdeeld
onverwijld
oostelijk
oosterse
ouden
ouderdomsproces
ouderlijk
ouderling
ouderschapsverlof
oudste
oudtante
oudtantes
overgehaald
overgestapt
overgezonden
overgrootvader
peil
peilsignaal
peilstift
peilstok
personeelsbestand
personeelsgegevens
personenwagen
pieptoon
pijlinktvis

pijlsnel
pijn
pijpkruid
pistoolgreep
polio
pruikje
pruil
pruimentaart
puilde
raap
rancuneus
religieus
reserveren
rijnstenen
rijzige
rioolrat
rioolstelsel
risico-kinderen
ritueel
rovertje
ruil
ruilbeurs
ruilhandel
ruilmiddel
saus
schadelijk
schapenfarm
schoolagenda
schoolfrik
schoolgebied
schooljuffrouw
schoolkameraad
schoolkamp
schoolkleuren
schoolreis
schoolteam
schoolvoorstelling
schoolvriend
schoolziek
schroothoop
schuiladres
schuilgaan
schuilgehouden
schuilhut
schuilkelder
schuilnaam
schuilplaats
sein
smeul
smeult
smoel
speelbal
speelbare
speelgoedwinkel
speelhal
speelkaart
speelmakker
speelster
staal
staalplaat
stoomstrijkijzers
strategen
stuiptrekkende
stuiter
stuiver
subtiel
suikerraffinaderij
suikerziek
tegenstrijdig
tegenvoeter
terpentine
terwijl
textielindustrie
thuishoorden
thuismarkt
tijdregistratie
tijdschema
tijdsprobleem
toegespitst
toveraar
uil
uilskuiken
uitbarsten
uitbazuinen
uitbuiten
uitdrukking
uitgehuwelijkt
uitgekeken
uitgerekend
uitgroeide
uithoren
uitladen
uitnamen
uitpuilend
uitreiking
uitscheld
uitspraak
uitzwaaien
veiligheidsredenen
veiligheidsscheermes
veiligheidssituatie
verbeid
verdraaglijk
verdrievoudigd
vereenzaamd
vereisen
vergeet
vergeving
verhouding
verijs
verkoelde
vermeende
verpauperd
verpersoonlijk
verruil
verschuil
verstuikt
vertonen
vervloek
vervuil
verzeild
verzwijg
vleesetende
vleeshouwer
vleugel
vloekwoord

waarheidlievend
wede
wezenfonds
wijdvertakte
wijnhuis
wijnkleurige
wijziging
woonark
zedelijkheidsgevoel
zeilt
zeiltocht
zeilwedstrijd
zeug
zeul
zeulde
zeult
zintuiglijk
zodanig
zoem
zotternij
zouden
zoutarm
zoutjes
zoutwinning
zuil
zwoel

APPENDIX D

Full BLUPs for Experiment 1 from Chapter 4

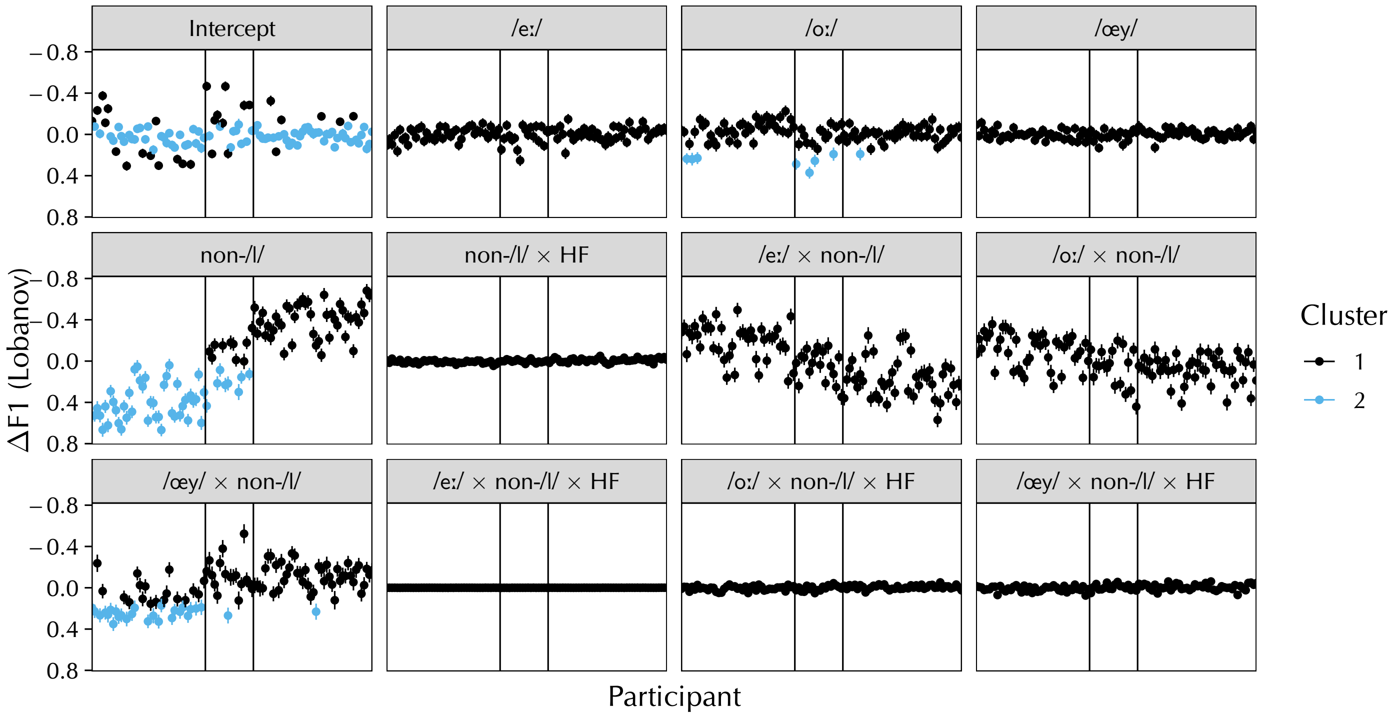

Figure D.1: Separate BLUPs for the each of the random effects present in the by-individuals ($n = 106$) model for the production data, with separate cluster analyses for each panel. As in Figure 4.3, each panel is separated into three panes; the left pane shows the participants from the Ghent group, the middle pane shows the participants from the migrant group, and the right pane shows the participants from the Leiden group.

APPENDIX E

Prime–target list for Experiment 2 from Chapter 4

bauke – zauke
bauke – zoke
bebbe – webbe
bebbe – wijbe
begge – wegge
begge – wijge
bekker – hekker
bekker – hijker
bemme – zemme
bemme – zijme
benne – henne
benne – hijne
bezze – hezze
bezze – hijze
blaude – graude
blaude – grode
blaulder – waulder
blaulder – woolder
blauver – schauver
blauver – schover
blebe – zebe
blebe – zijbe
blege – bege
blege – bijge
blelder – zelder
blelder – zijlder
blene – hene
blene – hijne
blete – lete
blete – lijte
bleter – leter
bleter – lijter
brauver – vauver
brauver – vover
bredde – stedde
bredde – stijde
bredder – tedder
bredder – tijder
breelder – veelder
breelder – vijlder
brette – tette
brette – tijte
bretter – letter
bretter – lijter
brever – tever
brever – tijver
bwelde – slelde
bwelde – slijlde
dauver – trauver
dauver – trover
debe – webe
debe – wijbe
deter – zeter
deter – zijter
devver – levver
devver – lijver
drever – rever
drever – rijver
dweelder – bleelder
dweelder – blijlder
dwever – krever
dwever – krijver
fauver – jauver

fauver – jover
fedde – sedde
fedde – sijde
fedder – kedder
fedder – kijder
fevver – wevver
fevver – wijver
flaulde – maulde
flaulde – moolde
flaulder – braulder
flaulder – broolder
fleelder – seelder
fleelder – sijlder
flelder – selder
flelder – sijlder
gauver – nauver
gauver – nover
gleelde – kreelde
gleelde – krijlde
glelde – flelde
glelde – flijlde
greelde – treelde
greelde – trijlde
greelder – teelder
greelder – tijlder
grelde – trelde
grelde – trijlde
grelder – telder
grelder – tijlder
grevver – revver
grevver – rijver
gweelde – fleelde
gweelde – flijlde
jaube – haube
jaube – hobe
jauge – nauge
jauge – noge
jauter – nauter
jauter – noter
jauze – nauze
jauze – noze
jede – krede
jede – krijde
jevver – sevver
jevver – sijver
kebe – tebe
kebe – tijbe
keder – zeder
keder – zijder
kegge – begge
kegge – bijge
klaulde – faulde
klaulde – foolde
klaulder – laulder
klaulder – loolder
klauver – vrauver
klauver – vrover
kleelder – geelder
kleelder – gijlder
klelder – kwelder
klelder – kwijlder
kneelde – pleelde
kneelde – plijlde
kraulder – naulder
kraulder – noolder
kreder – deder
kreder – dijder
kreelder – feelder
kreelder – fijlder
krelder – nelder
krelder – nijlder
kwaulde – gaulde
kwaulde – goolde
kwaulder – fraulder
kwaulder – froolder
kweelder – zeelder
kweelder – zijlder
lauder – zauder
lauder – zoder
lebbe – zebbe
lebbe – zijbe
mauder – jauder
mauder – joder
meker – heker
meker – hijker
nauve – zauve
nauve – zove
nebe – bebe
nebe – bijbe
neke – heke
neke – hijke
nete – tete
nete – tijte
neter – teter
neter – tijter
pauder – bauder
pauder – boder
pevver – tevver
pevver – tijver
pfelde – klelde
pfelde – klijlde
pjelde – xelde
pjelde – xijlde
plaulde – vaulde
plaulde – voolde
plaulder – praulder
plaulder – proolder
plauver – mauver
plauver – mover
pleelder – jeelder
pleelder – jijlder
plelder – jelder
plelder – jijlder
plever – pever
plever – pijver
preelder – steelder
preelder – stijlder
prelder – stelder
prelder – stijlder
prever – twever
prever – twijver
psaulder – draulder
psaulder – droolder
pselde – lelde
pselde – lijlde
rebe – hebe

rebe – hijbe
sauver – bauver
sauver – bover
schaube – kaube
schaube – kobe
schauge – kauge
schauge – koge
schauke – nauke
schauke – noke
schaulder – jaulder
schaulder – joolder
schevver – zevver
schevver – zijver
sedder – zedder
sedder – zijder
sfaulde – praulde
sfaulde – proolde
sfeelde – preelde
sfeelde – prijlde
sfelde – blelde
sfelde – blijlde
sjeelde – vreelde
sjeelde – vrijlde
skeelde – freelde
skeelde – frijlde
skelde – twelde
skelde – twijlde
skelder – twelder
skelder – twijlder
slaulde – baulde
slaulde – boolde
slaulder – taulder
slaulder – toolder
sleelder – peelder
sleelder – pijlder
slelder – dwelder
slelder – dwijlder
smeelde – bleelde
smeelde – blijlde
smelde – drelde
smelde – drijlde
smelder – drelder
smelder – drijlder
snaulde – draulde
snaulde – droolde
snaulder – graulder
snaulder – groolder
sneelde – sleelde
sneelde – slijlde
snever – jever
snever – jijver
spaulde – naulde
spaulde – noolde
spaulder – gaulder
spaulder – goolder
speelder – leelder
speelder – lijlder
spever – dever
spever – dijver
spleelde – kleelde
spleelde – klijlde
splelde – krelde
splelde – krijlde
spraulde – braulde
spraulde – broolde
spreelde – breelde
spreelde – brijlde
sprelde – brelde
sprelde – brijlde
sprever – mever
sprever – mijver
sraulde – xaulde
sraulde – xoolde
sreelde – greelde
sreelde – grijlde
srelde – grelde
srelde – grijlde
staude – gaude
staude – gode
staulde – jaulde
staulde – joolde
staulder – raulder
staulder – roolder
stauver – lauver
stauver – lover
stedder – hedder
stedder – hijder
stetter – wetter
stetter – wijter
straulde – fraulde
straulde – froolde
strelde – frelde
strelde – frijlde
strelder – frelder
strelder – frijlder
strevver – mevver
strevver – mijver
swaulde – kraulde
swaulde – kroolde
sweelde – leelde
sweelde – lijlde
sweelder – dreelder
sweelder – drijlder
swelde – delde
swelde – dijlde
swelder – relder
swelder – rijlder
taude – paude
taude – pode
tegge – wegge
tegge – wijge
tetter – hetter
tetter – hijter
tezze – zezze
tezze – zijze
tjaulder – baulder
tjaulder – boolder
tjeelde – dreelde
tjeelde – drijlde
tjelde – jelde
tjelde – jijlde
trauder – nauder
trauder – noder
trauge – kauge
trauge – koge
traute – jaute

traute – jote
treelder – neelder
treelder – nijlder
trelder – belder
trelder – bijlder
trever – sever
trever – sijver
twaulde – zaulde
twaulde – zoolde
twaulder – maulder
twaulder – moolder
twebe – lebe
twebe – lijbe
tweder – heder
tweder – hijder
tweelder – beelder
tweelder – bijlder
tweke – meke
tweke – mijke
vaube – naube
vaube – nobe
vauge – nauge
vauge – noge
vauke – hauke
vauke – hoke
vedde – kedde
vedde – kijde
vedder – ledder
vedder – lijder
vede – kede
vede – kijde
vleelde – tweelde
vleelde – twijlde
vlelde – prelde
vlelde – prijlde
vrauder – vauder
vrauder – voder
vraulde – waulde
vraulde – woolde
vraulder – vaulder
vraulder – voolder
vrebe – mebe
vrebe – mijbe
vreelder – deelder
vreelder – dijlder
vrelder – delder
vrelder – dijlder
wauver – kauver
wauver – kover
weme – heme
weme – hijme
wemme – hemme
wemme – hijme
weppe – heppe
weppe – hijpe
xaulder – saulder
xaulder – soolder
xelder – pelder
xelder – pijlder
zevve – hevve
zevve – hijve
zwaulde – graulde
zwaulde – groolde
zweelde – jeelde
zweelde – jijlde
zweelder – reelder
zweelder – rijlder
zwelder – lelder
zwelder – lijlder

APPENDIX F

Full results of Experiment 2 from Chapter 4

Factor	**Estimate (SE)**	**Odds ratio**	***z***	***p***	**Sig.**
Model = [eː∼ɛi]					
Intercept	−1.46 (0.10)	1 :4.30	−14.49	<.001	***
Step (Linear)	1.24 (0.13)	3.47 :1	9.87	<.001	***
Step (Quadratic)	0.33 (0.08)	1.39 :1	4.22	<.001	***
Step (Cubic)	0.03 (0.08)	1.03 :1	0.34	.74	
Following segment = /l/	0.83 (0.11)	2.29 :1	7.59	<.001	***
Group = Migrant–Ghent	−0.51 (0.13)	1 :1.66	−3.95	<.001	***
Group = Leiden–Others	0.07 (0.06)	1.07 :1	1.06	.29	
Step (Linear) × /l/	1.68 (0.21)	5.37 :1	8.14	<.001	***
Step (Quadratic) × /l/	−0.15 (0.14)	1 :1.16	−1.07	.29	
Step (Cubic) × /l/	−0.13 (0.14)	1 :1.14	−0.92	.36	
Step (Linear) × Migrant–Ghent	0.63 (0.15)	1.87 :1	4.06	<.001	***
Step (Quadratic) × Migrant–Ghent	−0.04 (0.08)	1 :1.04	−0.47	.64	
Step (Cubic) × Migrant–Ghent	0.05 (0.07)	1.05 :1	0.66	.51	
Step (Linear) × Leiden–Others	0.01 (0.07)	1.01 :1	0.17	.87	
Step (Quadratic) × Leiden–Others	0.04 (0.04)	1.04 :1	1.11	.27	
Step (Cubic) × Leiden–Others	−0.05 (0.03)	1 :1.05	−1.49	.14	
Following segment = /l/ × Migrant–Ghent	0.26 (0.13)	1.30 :1	2.05	.04	
Following segment = /l/ × Leiden–Others	0.08 (0.06)	1.08 :1	1.31	.19	
Model = [oː∼ɑu]					
Intercept	0.20 (0.07)	1.22 :1	2.76	.01	*
Step (Linear)	1.73 (0.15)	5.65 :1	11.48	<.001	***
Step (Quadratic)	0.45 (0.11)	1.57 :1	4.17	<.001	***
Step (Cubic)	−0.03 (0.11)	1 :1.03	−0.28	.78	

Factor	Estimate (SE)	Odds ratio	z	p	Sig.
Following segment = /l/	0.95 (0.12)	2.59 : 1	7.83	<.001	***
Step (Linear) × /l/	−0.66 (0.25)	1 : 1.93	−2.67	.01	*
Step (Quadratic) × /l/	−0.74 (0.22)	1 : 2.09	−3.32	<.001	**
Step (Cubic) × /l/	−0.19 (0.22)	1 : 1.21	−0.88	.38	
Model = [ε∼εi]					
Intercept	−0.94 (0.07)	1 : 2.55	−13.12	<.001	***
Step (Linear)	2.49 (0.15)	12.06 : 1	16.44	<.001	***
Step (Quadratic)	0.27 (0.10)	1.31 : 1	2.59	.01	*
Step (Cubic)	−0.74 (0.11)	1 : 2.09	−6.99	<.001	***
Following segment = /l/	0.96 (0.12)	2.60 : 1	7.90	<.001	***
Group = Migrant–Ghent	−0.21 (0.08)	1 : 1.23	−2.64	.01	*
Group = Leiden–Others	0.09 (0.04)	1.10 : 1	2.50	.01	*
Step (Linear) × /l/	−1.38 (0.24)	1 : 3.98	−5.84	<.001	***
Step (Quadratic) × /l/	0.25 (0.21)	1.29 : 1	1.22	.22	
Step (Cubic) × /l/	0.87 (0.21)	2.38 : 1	4.13	<.001	***
Step (Linear) × Migrant–Ghent	0.43 (0.17)	1.53 : 1	2.52	.01	*
Step (Quadratic) × Migrant–Ghent	0.09 (0.09)	1.09 : 1	0.99	.32	
Step (Cubic) × Migrant–Ghent	−0.20 (0.09)	1 : 1.22	−2.25	.02	
Step (Linear) × Leiden–Others	−0.23 (0.08)	1 : 1.26	−2.91	<.01	*
Step (Quadratic) × Leiden–Others	−0.03 (0.04)	1 : 1.03	−0.67	.50	
Step (Cubic) × Leiden–Others	0.04 (0.04)	1.04 : 1	1.08	.28	
Following segment = /l/ × Migrant–Ghent	0.11 (0.12)	1.12 : 1	0.92	.36	
Following segment = /l/ × Leiden–Others	0.01 (0.06)	1.01 : 1	0.15	.88	
Step (Linear) × /l/ × Migrant–Ghent	−0.24 (0.23)	1 : 1.27	−1.04	.30	

Factor	Estimate (SE)	Odds ratio	*z*	*p*	Sig.
Step (Quadratic) × /l/ × Migrant–Ghent	0.28 (0.18)	1.32 :1	1.57	.12	
Step (Cubic) × /l/ × Migrant–Ghent	0.39 (0.18)	1.47 :1	2.15	.03	
Step (Linear) × /l/ × Leiden–Others	−0.31 (0.11)	1 :1.36	−2.95	<.01	**
Step (Quadratic) × /l/ × Leiden–Others	−0.18 (0.08)	1 :1.19	−2.26	.02	
Step (Cubic) × /l/ × Leiden–Others	−0.09 (0.08)	1 :1.10	−1.14	.25	

Bibliography

Adank, P. M. (2003). *Vowel normalization: A perceptual-acoustic study of Dutch vowels* (Doctoral dissertation). Katholieke Universiteit Nijmegen. https://doi.org/2066/19286

Adank, P. M., van Hout, R. W. N. M., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America*, *116*, 1729–1738. https://doi.org/10.1121/1.1779271

Adank, P. M., van Hout, R. W. N. M., & Van de Velde, H. (2007). An acoustic description of the vowels of Northern and Southern Standard Dutch II: Regional varieties. *The Journal of the Acoustical Society of America*, *121*, 1130–1141. https://doi.org/10.1121/1.2409492

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*, 3099–3107. https://doi.org/10.1121/1.1795335

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: B. N. Petrov & F. Csáki (Eds.). *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2–8, 1971*, Akadémiai Kiadó.

Allen, J., Kraus, N., & Bradlow, A. (2000). Neural representation of consciously imperceptible speech sound differences. *Perception & Psychophysics*, *62*, 1383–1393. https://doi.org/10.3758/BF03212140

Alshangiti, W., & Evans, B. G. (2011). Regional accent accommodation in spontaneous speech: Evidence for long-term accent change? In: W. S. Lee & E. Zee (Eds.). *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS), Hong Kong, 17–21 August*.

Auer, P., Barden, B., & Grosskopf, B. (1998). Subjective and objective parameters determining 'salience' in long-term dialect accommodation. *Jour-*

nal of Sociolinguistics, *2*, 163–187. https://doi.org/10.1111/1467-9481.00039

Auer, P., & Hinskens, F. (2005). The role of interpersonal accommodation in a theory of language change. *Dialect Change: Convergence and Divergence in European Languages*, 335. https://doi.org/10.1017/CBO9780511486623.015

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

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

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2). *Linguistic Data Consortium*.

Babel, M., Haber, G., & Walters, S. (2013). Individual-level connections between perceptual adaptation and phonetic imitation. *Proceedings of the workshop on sound change actuation, University of Chicago*.

Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, *23*, 347–374. https://doi.org/10.1017/S0954394511000135

Baldeweg, T., Klugman, A., Gruzelier, J. H., & Hirsch, S. R. (2002). Impairment in frontal but not temporal components of mismatch negativity in schizophrenia. *International Journal of Psychophysiology*, *43*, 111–122. https://doi.org/10.1016/s0167-8760(01)00183-0

Barbiers, S. et al. (2006). *Dynamische Syntactische Atlas van de Nederlandse Dialecten (DynaSAND)*. http://www.meertens.knaw.nl/sand/

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, L. (1985). Tracing phonetic change in the received pronunciation of British English. *Journal of Phonetics*, *13*, 61–81. https://doi.org/10.1016/S0095-4470(19)30726-0

Becker, M., & Tessier, A.-M. (2011). Trajectories of faithfulness in child-specific phonology. *Phonology*, *28*, 163–196. https : / / doi . org / 10 . 1017 / S0952675711000133

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 785–821. https://doi.org/10.1353/lan.0.0165

Beddor, P. S. (2015). The relation between language users' perception and production repertoires. In: The Scottish Consortium for ICPhS 2015 (Ed.). *Proceedings of the 18th International Congress of Phonetic Sciences* (*ICPhS XVIII*). International Phonetic Association.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, *94*, 931–968. https://doi.org/10.1353/lan.2018.0051

Bermúdez-Otero, R. (2007). Diachronic phonology. In: P. de Lacy (Ed.). *The Cambridge handbook of phonology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486371

Bermúdez-Otero, R. (2015). Amphichronic explanation and the life cycle of phonological processes. In: P. Honeybone & J. Salmons (Eds.). *The Oxford handbook of historical phonology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199232819.013.014

Berns, J., & Jacobs, H. (2012). A first glance at the role of length in production and perception of diphthongs before Dutch coda l. *Linguistics in the Netherlands*, *29*, 15–26. https://doi.org/10.1075/avt.29.02ber

Berwick, R. C. (1985). *The acquisition of syntactic knowledge*. MIT Press.

Blankestein, H. (1994). Wij, hai en zaai. *NRC Handelsblad*, (15 December 1994), 32.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. *Bidirectional Optimality Theory*, *180*, 33. https://doi.org/10.1075/la.180.02boe

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, *25*, 217–270. https://doi.org/10.1017/S0952675708001147

Boersma, P., & Weenink, D. (2016, September 26). *Praat: Doing phonetics by computer* (Version 6.0.21). http://www.praat.org/

Booij, E. E. (1995). *The phonology of Dutch*. Oxford University Press.

Bosker, H. R., & Reinisch, E. (2015). Normalization for speechrate in native and nonnative speech. *18th International Congress of Phonetic Sciences* (*ICPhS 2015*). https://doi.org/11858/00-001M-0000-0028-5DFD-D

Botma, B., Sebregts, K., & Smakman, D. (2012). The phonetics and phonology of Dutch mid vowels before /l/. *Laboratory Phonology*, *3*, 273–297. https://doi.org/10.1515/lp-2012-0015

Bowie, D. (2000). *The effect of geographic mobility on the retention of a local dialect* (Doctoral dissertation). University of Pennsylvania.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*, 395–411. https://doi.org/10.32614/RJ-2018-017

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*, 261–290. https://doi.org/10.10170S0954394502143018

Bybee, J. L., & Slobin, D. I. (1982). Why small children cannot change language on their own: Suggestions from the English past tense. In: A. Alqvist (Ed.). *Papers from the Fifth International Conference on Historical Linguistics*. John Benjamins Publishing Company.

Carter, P. (2007). Phonetic variation and speaker agency: Mexicana identity in a North Carolina middle school. *University of Pennsylvania Working Papers in Linguistics*, *13*(2), 1–14.

Cedergren, H. (1987). The spread of language change: Verifying inferences of linguistic diffusion. *Language Spread and Language Policy: Issues, Implications, and Case Studies*, 45–60. https://doi.org/10822/555480

Chambers, J. K. (1992). Dialect acquisition. *Language*, *68*, 673–705. https://doi.org/10.1353/lan.1992.0060

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, *66*, 185–216. https://doi.org/10.1016/j.wocn.2017.09.009

Collins, B., & Mees, I. M. (1999). *The phonetics of English and Dutch*. Brill Academic Publishers Incorporated.

Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). "What you encode is not necessarily what you store": Evidence for sparse feature representations from mismatch negativity. *Brain Research*, *1394*, 79–89. https://doi.org/10.1016/j.brainres.2011.04.001

Craddock, M. (2019). *eegUtils: A collection of utilities for EEG analysis*. R package version 0.4.0. R package version 0.4.0. https://github.com/craddm/eegUtils

Cutler, A. (2012). *Native listening*. MIT Press.

Cutler, A., McQueen, J. M., Butterfield, S., & Norris, D. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. In: The International Speech Communications Association (Ed.). *Interspeech 2008*. https://doi.org/2066/72799

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*, 89–103. https://doi.org/10.1016/j.brainres.2006.02.010

De Decker, P. (2006). A real-time investigation of social and phonetic changes in post-adolescence. *University of Pennsylvania Working Papers in Linguistics*, *12*(2), 7.

De Wulf, C., Goossens, J., & Taeldeman, J. (2005). *Fonologische atlas van de Nederlandse dialecten. Deel IV: De consonanten* (Vol. 1). Koninklijke Academie voor Nederlandse Taal-en Letterkunde.

Delarue, S. (2013). Teachers' Dutch in Flanders: The last guardians of the standard? In: T. Kristiansen & S. Grondelaers (Eds.). *Language (de)standardisation in Late Modern Europe: Experimental studies*. Novus Forlag. https://doi.org/1854/LU-4175602

Denis, D. (2011). Innovators and innovation: Tracking the innovators of and stuff in York English. *University of Pennsylvania Working Papers in Linguistics*, *17*(2), 8.

Deouell, L. Y. (2007). The frontal generator of the mismatch negativity revisited. *Journal of Psychophysiology*, *21*, 188–203. https://doi.org/10.1027/0269-8803.21.34.188

Domahs, U., Kehrein, W., Knaus, J., Wiese, R., & Schlesewsky, M. (2009). Event-related potentials reflecting the processing of phonological constraint violations. *Language and Speech*, *52*, 415–435. https://doi.org/10.1177/0023830909336581

Drager, K., & Hay, J. (2012). Exploiting random intercepts: Two case studies in sociophonetics. *Language Variation and Change*, *24*, 59–78. https://doi.org/10.1017/S0954394512000014

Eekhof, L. S., Kuijpers, M., Faber, M., Gao, X., Mak, M., van den Hoven, E., & Willems, R. M. (submitted). Lost in a story, detached from the words:

Absorbed readers are less sensitive to word characteristics during narrative reading. *Discourse Processes*.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, *121*, 3814–3826. https://doi.org/10.1121/1.2722209

Fisher, R. A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society: Series B (Methodological)*, *17*(1), 69–78. https://www.jstor.org/stable/2983785

Fisher, R. A. (1956). *Statistical methods and scientific inference*. Oliver & Boyd.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. E., & Wayland, R. (2019). The role of input in native Spanish late learners' production and perception of English phonetic segments. *Journal of Second Language Studies*, *2*, 1–44. https://doi.org/10.1075/jsls.00004.fle

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research*, *38*, 379–412. https://doi.org/10.1007/s10936-008-9097-8

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1276. https://doi.org/10.1037/0096-1523.32.5.1276

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods*, *50*, 2016–2034. https://doi.org/10.3758/s13428-017-0971-x

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*, 183–192. https://doi.org/10.1016/0926-6410(93)90026-2

Ganong III, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*, 105–132. https://doi.org/10.1016/s0010-0277(03)00070-2

Gerritsen, M., & Jansen, F. (1980). Word frequency and lexical diffusion in dialect borrowing and phonological change. In: W. Zonneveld, F. van Coetsem, & O. W. Robinson (Eds.). *Dutch studies 4: Studies in Dutch* *phonology*. Martinus Nijhoff. https://doi.org/10.1007/978-94-009-8855-2_3

Giard, M.-H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology*, *27*, 627–640. https://doi.org/10.1111/j.1469-8986.1990.tb03184.x

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*, 587–606. https://doi.org/10.1016/j.socec.2004.09.033

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In: H. Giles, J. Coupland, & N. Coupland (Eds.). *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press. https://doi.org/10.1017/CBO9780511663673.001

Giles, H., & Smith, P. M. (1979). Accommodation theory: Optimal levels of convergence. In: H. Giles & R. St Clair (Eds.). *Language and social psychology* (pp. 45–65). Blackwell.

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, *2*, 177–192. https://doi.org/10.1017/S0047404500000701

Goossens, J., Taeldeman, J., & Verleyen, G. (1998). *Fonologische atlas van de Nederlandse dialecten. Deel I: Het korte vocalisme*. Koninklijke Academie voor Nederlandse Taal-en Letterkunde.

Gordon, E., & Maclagan, M. (2001). 'Capturing a sound change': A real time study over 15 years of the NEAR/SQUARE diphthong merger in New Zealand English. *Australian Journal of Linguistics*, *21*, 215–238. https://doi.org/10.1080/07268600120080578

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*, 92–102. https://doi.org/10.1016/j.bandl.2012.04.017

Grondelaers, S., & van Hout, R. (2011). The standard language situation in the low countries: Top-down and bottom-up variations on a diaglossic theme. *Journal of Germanic Linguistics*, *23*, 199–243. https://doi.org/10.1017/S1470542711000110

Grondelaers, S., & van Hout, R. (2012). Where is Dutch (really) heading?: The classroom consequences of destandardization. *Dutch Journal of Applied Linguistics*, *1*, 41–58. https://doi.org/10.1075/dujal.1.1.05gro

Grosvald, M., & Corina, D. (2012). The production and perception of subphonemic vowel contrasts and the role of the listener in sound change. In: M.-J. Solé & D. Recasens (Eds.). *The initiation of sound change: Perception, production, and social factors*. John Benjamins Publishing Company. https://doi.org/10.1075/cilt.323.08gro

Guion, S. G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica*, *55*, 18–52. https://doi.org/10.1159/000028423

Gussenhoven, C. (1993). The Dutch foot and the chanted call. *Journal of Linguistics*, *29*, 37–63. https://doi.org/10.1017/S0022226700000049

Gussenhoven, C. (1999). Dutch. In: International Phonetic Association (Ed.). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Gussenhoven, C., & Aarts, F. (1999). The dialect of Maastricht. *Journal of the International Phonetic Association*, *29*, 155–166. https://doi.org/10.1017/S0025100300006526

Gussenhoven, C., & Broeders, A. (1976). *The pronunciation of English: A course for Dutch learners*. Wolters-Noordhoff-Longman.

Haeri, N. (1991). *Sociolinguistic variation in Cairene Arabic: Palatalization and the Qaf in the speech of men and women* (Doctoral dissertation). University of Pennsylvania.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – the R package pbkrtest. *Journal of Statistical Software*, *59*, 1–30. https://doi.org/10.18637/jss.v059.i09

Hamann, S. (2009). The learner of a perception grammar as a source of sound change. In: P. Boersma & S. Hamann (Eds.). *Phonology in perception*. Mouton de Gruyter.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, *34*, 439–457. https://doi.org/10.1016/j.wocn.2005.08.001

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. *Laboratory Phonology*, *9*, 125–143.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America*, *123*, 2825–2835. https://doi.org/10.1121/1.2897042

Harrington, J., Palethorpe, S., & Watson, C. (2000). Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, *30*, 63–78. https://doi.org/10.1017/S0025100300006666

Harrington, J., Palethorpe, S., & Watson, C. (2005). Deepening or lessening the divide between diphthongs? An analysis of the Queen's annual Christmas broadcasts. In: W. J. Harcastle & J. M. Beck (Eds.). *A figure of speech: A festschrift for John Laver*. Lawrence Erlbaum.

van der Harst, S. (2011). *The vowel space paradox: A sociophonetic study on Dutch* (Doctoral dissertation). Radboud Universiteit Nijmegen. LOT. https://doi.org/2066/86656

van der Harst, S., Van de Velde, H., & van Hout, R. (2014). Variation in Standard Dutch vowels: The impact of formant measurement methods on identifying the speaker's regional origin. *Language Variation and Change*, *26*, 247–272. https://doi.org/10.1017/S0954394514000040

Hastie, T., & Tibshirani, R. (1987). Generalized additive models: Some applications. *Journal of the American Statistical Association*, *82*(398), 371–386. https://www.jstor.org/stable/2289439

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*, 865–892. https://doi.org/10.1515/ling.2010.027

Henderson, C. R. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics*, *21*, 309–310.

Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language*, *152*, 28–43. https://doi.org/10.1016/j.bandl.2015.10.007

van Heuven, V. J., Van Bezooijen, R., & Edelman, L. J. (2005). Pronunciation of /ei/ in avant-garde Dutch: A cross-sex acoustic study. In: M. Filppula, J. Klemola, M. Palander, & E. Penttilä (Eds.). *Dialects across borders*. John Benjamins Publishing Company. https://hdl.handle.net/11245/1.246311

Hinton, M. (2015). Changes in Received Pronunciation: Diachronic case studies. *Research in Language*, *13*, 21–37. https://doi.org/10.1515/rela-2015-0010

van Hout, R. W. N. M., de Schutter, G., de Crom, E., Huinck, W., Kloots, H., & Van de Velde, H. (1999). De uitspraak van het Standaard-Nederlands. Variatie en varianten in Vlaanderen en Nederland. In: E. Huls & B. Weltens (Eds.). *Artikelen van de derde sociolinguïstische conferentie*. Eburon.

Hyman, L. M. (1976). Phonologization. In: A. Juilland (Ed.). *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*. Anima Libri.

Hyman, L. M. (2013). Enlarging the scope of phonologization. In: A. C. L. Yu (Ed.). *Origins of sound change: Approaches to phonologization*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199573745.001.0001

Jacobi, I. (2009). *On variation and change in diphthongs and long vowels of spoken Dutch* (Doctoral dissertation). Universiteit van Amsterdam. https://doi.org/11245/1.316951

Jacobsen, T. K. (2015). *Preattentive phonotactic processing* (Doctoral dissertation). Helmut-Schmidt-Universität.

Janson, T. (1983). Sound change in perception and production. *Language*, 18–34. https://doi.org/10.2307/414059

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Johnson, D. E. (2007). Stability and change along a dialect boundary: The low vowels of southeastern New England. *University of Pennsylvania Working Papers in Linguistics*, *13*(1), 7.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Karush, W. (1939). *Minima of functions of several variables with inequalities as side constraints* (Master's thesis). University of Chicago.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. https://doi.org/10.1109/ICASSP.2008.4518514

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, *103*, 11381–11386. https://doi.org/10.1073/pnas.0604821103

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 983–997. https://doi.org/10.2307/2533558

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643–650. https://doi.org/10.3758/BRM.42.3.643

Kewley-Port, D. (1995). Thresholds for formant-frequency discrimination of vowels in consonantal context. *The Journal of the Acoustical Society of America*, *97*, 3139–3146. https://doi.org/10.1121/1.413106

Kleber, F., Harrington, J., & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, *55*, 383–405. https://doi.org/10.1177/0023830911422194

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, *1*, 238. https://doi.org/10.3389/fpsyg.2010.00238

Kruschke, J. K. (2010a). Bayesian data analysis. *WIREs Cognitive Science*, *1*, 658–676. https://doi.org/10.1002/wcs.72

Kruschke, J. K. (2010b). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*, 293–300. https://doi.org/10.1016/j.tics.2010.05.001

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*, 178–206. https://doi.org/10.3758/s13423-016-1221-4

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*, 9096–9101. https://doi.org/10.1073/pnas.1532872100

Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In: J. Neyman (Ed.). *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, University of California Press.

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, *74*, 1–19. https://doi.org/10.1016/j.biopsycho.2006.06.001

Kung, C., Chwilla, D. J., & Schriefers, H. (2014). The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia*, *53*, 293–309. https://doi.org/10.1016/j.neuropsychologia.2013.11.020

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Labov, W. (1994). *Principles of linguistic change. Volume I: Internal factors*. Blackwell.

Labov, W. (2001). *Principles of linguistic change. Volume II: Social factors*. Blackwell.

Labov, W. (2007). Transmission and diffusion. *Language*, *83*, 344–387. https://doi.org/10.1353/lan.2007.0082

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, 30–65. https://doi.org/10.1353/lan.2013.0015

Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress*. University of Philadelphia.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*, 44–59. https://doi.org/10.1016/j.wocn.2010.01.002

Lanwermeyer, M., Henrich, K., Rocholl, M. J., Schnell, H. T., Werth, A., Herrgen, J., & Schmidt, J. E. (2016). Dialect variation influences the phonological and lexical-semantic word processing in sentences: Electrophysiological evidence from a cross-dialectal comprehension study. *Frontiers in Psychology*, *7*, 739. https://doi.org/0.3389/fpsyg.2016.00739

Lev-Ari, S. (2018). Social network size can influence linguistic malleability and the propagation of linguistic change. *Cognition*, *176*, 31–39. https://doi.org/10.1016/j.cognition.2018.03.003

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38. https://doi.org/10.1017/S0140525X99001776

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di linguistica*, *7*, 5–36.

Liu, B., Jin, Z., Wang, Z., & Xin, S. (2011). An ERP study on whether the P600 can reflect the presence of unexpected phonology. *Experimental Brain Research*, *212*, 399–408. https://doi.org/10.1007/s00221-011-2739-3

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171/full

Mahajan, Y., Peter, V., & Sharma, M. (2017). Effect of EEG referencing methods on auditory mismatch negativity. *Frontiers in Neuroscience*, *11*, 560. https://doi.org/10.3389/fnins.2017.00560

Mak, M., & Willems, R. M. (2019). Mental simulation during literary reading: Individual differences revealed with eye-tracking. *Language, Cognition and Neuroscience*, 1–25. https://doi.org/10.1080/23273798.2018.1552007

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*, 213–228. https://doi.org/10.3758/BF03204377

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522. https://doi.org/10.1038/244522a0

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*, 543–562. https://doi.org/10.1080/03640210802035357

McQueen, J. M. (1993). Rhyme decisions to spoken words and nonwords. *Memory & Cognition*, *21*(2), 210–222.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*, 1113–1126. https://doi.org/10.1207/s15516709cog0000_79

Mees, I., & Collins, B. (1983). A phonetic description of the vowel system of Standard Dutch (ABN). *Journal of the International Phonetic Association, 13*, 64–75. https://doi.org/10.1017/S0025100300002565

Milroy, J. (1993). On the social origins of language change. *Historical Linguistics: Problems and Perspectives, 215*, 236. https://doi.org/10.4324/9781315845814

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics, 65*, 956–969. https://doi.org/10.3758/BF03194826

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences, 13*, 201–233. https://doi.org/10.1017/S0140525X00078407

Näätänen, R., & Alho, K. (1997). Mismatch negativity–the measure for central sound representation accuracy. *Audiology and Neurotology, 2*, 341–353. https://doi.org/10.1159/000259255

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*, 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Nahkola, K., & Saanilahti, M. (2004). Mapping language changes in real time: A panel study on Finnish. *Language Variation and Change, 16*, 75–92. https://doi.org/10.1017/S0954394504162017

Neyman, J. (1950). *First course in probability and statistics*. Henry Holt & Company.

Neyman, J. (1957). "Inductive behavior" as a basic concept of philosophy of science. *Revue de l'Institut International de Statistique*, 7–22. https://www.jstor.org/stable/1401671

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Nycz, J. (2011). *Second dialect acquisition: Implications for theories of phonological representation* (Doctoral dissertation). New York University.

Nycz, J. (2015). Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass, 9*, 469–482. https://doi.org/10.1111/lnc3.12163

Nycz, J. (2013). Changing words or changing rules? Second dialect acquisition and phonological representation. *Journal of Pragmatics, 52*, 49–62. https://doi.org/10.1016/j.pragma.2012.12.014

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42–46. https://doi.org/10.1111/j.1467-9280.1994.tb00612.x

Ohala, J. J. (1981). The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.). *Proceedings of the Chicago Linguistics Society 17: Papers from the parasession on language and behaviour*. University of Chicago Press.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In: L. E. Breivik & E. H. Jahr (Eds.). *Language change: Contributions to the study of its causes*. Mouton de Gruyter. https://doi.org/10.1515/9783110853063.173

Ohala, J. J. (2012). The listener as a source of sound change: An update. In: M.-J. Solé & D. Recasens (Eds.). *The initiation of sound change: Perception, production, and social factors*. John Benjamins Publishing Company. https://doi.org/10.1075/cilt.323.05oha

van Oostendorp, M. (2000). *Phonological projection: A theory of feature content and prosodic structure*. Walter de Gruyter.

van Oostendorp, M. (2008). Hoe de slot-r verdween uit het Nederlands: Ook koningin Beatrix volgt uitspraakverandering. *Onze Taal*, *2*(3), 53–55.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Osthoff, H., & Brugmann, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*. S. Hirzel.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309. https://doi.org/10.1037/0278-7393.19.2.309

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

Pardo, J. S. (2012). Reflections on phonetic convergence: Speech perception does not mirror speech production. *Language and Linguistics Compass*, *6*, 753–767. https://doi.org/10.1002/lnc3.367

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*, 190–197. https://doi.org/10.1016/j.wocn.2011.10.001

Passy, P. (1890). *Etude sur les changements phonétiques et leurs caractères généraux*. Librairie Firmin-Didot.

Pater, J., Sanders, L., Moore-Cantwell, C., Staubs, R., & Zobel, B. H. (submitted). *Event-related potential evidence of abstract phonological learning in the laboratory*. Preprint available at http://www.clairemoorecantwell.org/phonolearn/moore-cantwell-et-al-2018-ERP-ALL.pdf. Preprint available at http://www.clairemoorecantwell.org/phonolearn/moore-cantwell-et-al-2018-ERP-ALL.pdf.

Peeters, W. J. M. (1991). *Diphthong dynamics: A cross-linguistic perceptual analysis of temporal patterns in Dutch, English, and German*. Mondiss.

Peirce, J. W. (2007). Psychopy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*, 127–152. https://doi.org/10.1111/1469-8986.3720127

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In: J. Bybee & P. Hopper (Eds.). *Frequency and the emergence of linguistic structure*. John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.08pie

Pinget, A.-F. C. H. (2015). *The actuation of sound change* (Doctoral dissertation). Universiteit Utrecht. LOT. https://doi.org/1874/318102

Pinget, A.-F. C. H., Kager, R., & Van de Velde, H. (2019). Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. *Language and Speech*, 1–26. https://doi.org/10.1177/0023830919880206

Prince, E. F. (1987). Sarah Gorby, Yiddish folksinger: A case study of dialect shift. *International Journal of the Sociology of Language*, 83–116. https://doi.org/10.1515/ijsl.1987.67.83

Psychology Software Tools (2012). *E-Prime* (Version 2.0). https://www.pstnet.com

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

Rácz, P. (2013). *Salience in sociolinguistics: A quantitative approach*. Walter de Gruyter.

Ramsammy, M. (2015). The life cycle of phonological processes: Accounting for dialectal microtypologies. *Language and Linguistics Compass*, *9*(1), 33–54. https://doi.org/10.1111/lnc3.12102

van Reenen, P., & Jongkind, A. (2000). De vocalisering van de /l/ in het Standaard-Nederlands. *Taal en Tongval*, *52*(1), 189–199.

Rinne, T., Alho, K., Ilmoniemi, R. J., Virtanen, J., & Näätänen, R. (2000). Separate time behaviors of the temporal and frontal mismatch negativity sources. *Neuroimage*, *12*, 14–19. https://doi.org/10.1006/nimg.2000.0591

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological science*, *29*, 700–710. https://doi.org/10.1177/0956797617742725

Ruijgrok, B. J. (2018). *Tapping into semantic recovery: An event-related potential study on the processing of gapping and stripping* (Doctoral dissertation). Universiteit Leiden. LOT. https://doi.org/1887/62457

Sankoff, G. (2004). Adolescents, young adults and the critical period: Two case studies from 'Seven Up'. *Sociolinguistic variation: Critical reflections*, 121–39.

Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, *83*, 560–588. https://www.jstor.org/stable/40070902

Sankoff, G., Blondeau, H., & Charity, A. (2001). Individual roles in a real-time change. Montreal (r→ʀ) 1947–1995. In: H. Van de Velde & R. van Hout (Eds.). *'r-atics: Sociolinguistic, phonetic and phonological characteristics of /r/*, IVLP.

Scharinger, M., & Lahiri, A. (2010). Height differences in English dialects: Consequences for processing and representation. *Language and Speech*, *53*(2), 245–272.

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2016). Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations. *NeuroImage*, *128*, 293–301. https://doi.org/10.1016/j.neuroimage.2016.01.003

Scheer, T. (2014). How diachronic is synchronic grammar? Crazy rules, regularity and naturalness. In: P. Honeybone & J. C. Salmons (Eds.). *The handbook of historical phonology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199232819.013.003

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. https://doi.org/10.1214/aos/1176344136

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, *8*(1), 205–233. https://journal.r-project.org/archive/2016-1/scrucca-fop-murphy-etal.pdf

Sebregts, K. (2015). *The sociophonetics and phonology of Dutch /r/* (Doctoral dissertation). Universiteit Utrecht. LOT. https://doi.org/1874/306415

Shi, T., Kasahara, S., Pongkittiphan, T., Minematsu, N., Saito, D., & Hirose, K. (2015). A measure of phonetic similarity to quantify pronunciation variation by using ASR technology. In: The Scottish Consortium for ICPhS 2015 (Ed.). *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)*. International Phonetic Association.

Siegel, J. (2010). *Second dialect acquisition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511777820

Sneller, B. (2018). *Mechanisms of phonological change* (Doctoral dissertation). University of Pennsylvania.

Sonderegger, M., Bane, M., & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language*, *93*, 598–640. https://doi.org/10.1353/lan.2017.0038

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2010a). Activation and application of an obligatory phonotactic constraint in German during automatic speech processing is revealed by human event-related potentials. *International Journal of Psychophysiology*, *77*, 13–20. https://doi.org/10.1016/j.ijpsycho.2010.03.011

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2010b). Preattentive phonotactic processing as indexed by the mismatch negativity. *Journal of Cognitive Neuroscience*, *22*, 2174–2185. https://doi.org/10.1162/jocn.2009.21408

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2011). Phonotactic constraint violations in German grammar are detected automatically in auditory speech processing: A human event-related potentials study. *Psychophysiology*, *48*, 1208–1216. https://doi.org/10.1111/j.1469-8986.2011.01200.x

Stevens, M., & Harrington, J. (2014). The individual and the actuation of sound change. *Loquens*, *1*, e003. https://doi.org/10.3989/loquens.2014.003

Stroop, J. (1992). Weg standaardtaal. de nieuwe koers van het Nederlands. *Onze Taal*, *6*(9), 179–182.

Stroop, J. (1998). *Poldernederlands. Waardoor het ABN verdwijnt*. Bert Bakker.

Sussman, E. S., Chen, S., Sussman-Fort, J., & Dinces, E. (2014). The five myths of MMN: Redefining how to use MMN in basic and clinical research. *Brain Topography*, *27*(4), 553–564.

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, *24*, 135–178. https://doi.org/10.1017/S0954394512000129

Tamminga, M. (to appear). Leaders of language change: Macro and micro perspectives. In: H. Van de Velde, N. H. Hilton, & R. Knooihuizen (Eds.). *Language variation — European perspectives VIII: Selected papers from the tenth International Conference on Language Variation in Europe (ICLaVE 10), Leeuwarden, June 2019*, John Benjamins Publishing Company.

Tamminga, M., Ahern, C., & Ecay, A. (2016). Generalized additive mixed models for intraspeaker variation. *Linguistics Vanguard*, *2*. https://doi.org/10.1515/lingvan-2016-0030

Timmerman, J. H. (2018). *Ken mèn 't rotte hoe of jenut spelt!! Een vergelijking van de orthografische diepte van het Standaardnederlands en het Haagse stadsdialect* (Bachelor's thesis). Leiden University. https://doi.org/1887/65913

Trofimovich, P., Collins, L., Cardoso, W., White, J., & Horst, M. (2012). A frequency-based approach to L2 phonological learning: Teacher input and student output in an intensive ESL context. *TESOL Quarterly*, *46*, 176–187. https://doi.org/10.1002/tesq.9

Trudgill, P. (1986). *Dialects in contact*. Blackwell.

Trudgill, P. (1988). Norwich revisited: Recent linguistic changes in an English urban dialect. *English World-wide*, *9*, 33–49. https://doi.org/10.1075/eww.9.1.03tru

Trudgill, P. (1999). Norwich: Endogenous and exogenous linguistic change. *Urban voices: Accent studies in the British Isles*, 124–40.

Van de Velde, H. (1996). *Variatie en verandering in het gesproken Standaard-Nederlands* (Doctoral dissertation). Katholieke Universiteit Nijmegen. https://doi.org/2066/146159

Van de Velde, H., Kissine, M., Tops, E., van der Harst, S., & van Hout, R. (2011). Will Dutch become Flemish? Autonomous developments in Belgian Dutch. *Multilingua*, *3-4*, 385–416. https://doi.org/10.1515/mult.2010.019

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

Van de Velde, H., & van Hout, R. W. N. M. (2003). Diftongering in het Standaard-Nederlands. In: T. Koole, J. Nortier, & B. Tahitu (Eds.). *Artikelen van de vierde sociolinguïstische conferentie*. Eburon.

Van Istendael, G. (2008). Plezier en pijn van het Nederlands. *Samen beleid maken in het GO! Onderwijs van de Vlaamse Gemeenschap. Beleidscahier*, *3*, 29–33.

Verhoeven, J. (2005). Belgian Standard Dutch. *Journal of the International Phonetic Association*, *35*, 243–247. https://doi.org/10.1017/S0025100305002173

Voeten, C. C. (2015). *The interaction between the segmental and the prosodic phonology in the midst of an on-going sound change: Resolving a contradiction in the synchronic phonology of Dutch* (Master's thesis). Radboud University Nijmegen. https://theses.ubn.ru.nl/handle/123456789/628

Voeten, C. C. (2018). *permutes: Permutation tests for time series data*. R package version 0.1. R package version 0.1. https://CRAN.R-project.org/package=permutes

Voeten, C. C. (2019a). *buildmer: Stepwise elimination and term reordering for mixed-effects regression*. R package version 1.1. R package version 1.1. https://CRAN.R-project.org/package=buildmer

Voeten, C. C. (2019b). *buildmer: Stepwise elimination and term reordering for mixed-effects regression*. R package version 1.4. R package version 1.4. https://CRAN.R-project.org/package=buildmer

Voeten, C. C. (2019c). *permutes: Permutation tests for time series data*. R package version 1.0. R package version 1.0. https://CRAN.R-project.org/package=permutes

Voeten, C. C. (2020). *buildmer: Stepwise elimination and term reordering for mixed-effects regression*. R package version 1.6. R package version 1.6. https://CRAN.R-project.org/package=buildmer

Voeten, C. C. (to appear). Individual differences in the adoption of sound change. *Language and Speech*. https://doi.org/10.1177/0023830920959753

Voeten, C. C., & Levelt, C. C. (2019). ERP responses to regional accent reflect two distinct processes of perceptual compensation. *Frontiers in Neuroscience*, *13*, 546. https://doi.org/10.3389/fnins.2019.00546

Voortman, B. (1994). *Regionale variatie in het taalgebruik van notabelen. Een sociolinguïstisch onderzoek in Middelburg, Roermond en Zutphen*. IFOTT.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779–804. https://doi.org/10.3758/BF03194105

Wagner, S. E. (2008). *Linguistic change and stabilization in the transition from adolescence to adulthood* (Doctoral dissertation). University of Pennsylvania.

Wagner, S. E. (2012). Age grading in sociolinguistic theory. *Language and Linguistics Compass*, *6*, 371–382. https://doi.org/10.1002/lnc3.343

Walker, A. J. (2014). *Crossing oceans with voices and ears: Second dialect acquisition and topic-based shifting in production and perception* (Doctoral dissertation). Ohio State University.

Walker, M., Szakay, A., & Cox, F. (2019). Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners? *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*. https://doi.org/10.5334/labphon.90

Wang, W. S.-Y. (1969). Competing changes as a cause of residue. *Language*, 9–25. https://doi.org/10.2307/411748

Wang, W. S.-Y., & Fillmore, C. J. (1961). Intrinsic cues and consonant perception. *Journal of Speech and Hearing Research*, *4*, 130–136. https://doi.org/10.1044/jshr.0402.130

Watt, D. J. L. (2000). Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Language Variation and Change*, *12*, 69–101. https://doi.org/10.1017/S0954394500121040

Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of language change. In: W. Lehmann & Y. Malkiel (Eds.). *Directions for historical linguistics*. University of Texas Press.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*, 311–334. https://doi.org/10.1006/jmla.1995.1014

Witteman, M. J., Bardhan, N. P., Weber, A., & McQueen, J. M. (2015). Automaticity and stability of adaptation to a foreign-accented speaker. *Language and Speech*, *58*, 168–189. https://doi.org/10.1177/0023830914528102

Witteman, M. J., Weber, A., & McQueen, J. M. (2014). Tolerance for inconsistency in foreign-accented speech. *Psychonomic bulletin & review*, *21*, 512–519. https://doi.org/10.3758/s13423-013-0519-8

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman & Hall/CRC.

Yaeger-Dror, M. (1994). Phonetic evidence for sound change in Quebec French. *Phonological structure and phonetic form: Papers in laboratory phonology*, *3*, 267–292. https://doi.org/10.1017/CBO9780511659461.016

Yang, C. (2009). Population structure and language change. *Ms.*

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D. et al. (2002). *The HTK book*. Cambridge University Engineering Department.

Yu, A. C. L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PloS One*, *5*, e11950. https://doi.org/10.1371/journal.pone.0011950

Yu, A. C. L. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In: A. C. L. Yu (Ed.). *Origins of sound change: Approaches to phonologization*. Oxford University Press.

Zellou, G., Dahan, D., & Embick, D. (2017). Imitation of coarticulatory vowel nasality across words and time. *Language, Cognition and Neuroscience*, *32*, 776–791. https://doi.org/10.1080/23273798.2016.1275710

Ziliak, Z. L. (2012). *The relationship between perception and production in adult acquisition of a new dialect's phonetic system* (Doctoral dissertation). University of Florida.

Zwaardemaker, H., & Eijkman, C. L. P. H. (1924). *Leerboek der Phonetiek. Inzonderheid met betrekking tot het Standaard-Nederlands*. De erven F. Bohn.

Samenvatting in het Nederlands

Dit proefschrift heeft als doel te beschrijven hoe klankveranderingen worden overgenomen door sprekers en luisteraars, en welke factoren daar invloed op hebben. Dit wordt bestudeerd aan de hand van de “Polderverschuiving”: een klinkerverschuiving die vandaag de dag gaande is in het Nederlands, die bestaat uit vier onderling-verwante klankveranderingen. Dit zijn (1) de diftongering van de lange middenvocalen /eː,øː,oː/, (2) de verlaging van /ɛi,œy,ɔu/, (3) de blokkering van diftongering voor coda /l/, en (4) de vocalisering van coda /l/. Het proefschrift bestudeert deze veranderingen door meerdere deelgebieden van de taalwetenschap met elkaar te verbinden, namelijk de historische fonologie (wat weten we over patronen van klankveranderingen in het algemeen?), de sociofonetiek (welke hedendaagse variatie en verandering speelt er rondom de Polderverschuiving?), de psycholinguïstiek en neurolinguïstiek (hoe raakt synchrone variatie verankerd in het taalsysteem van een individu?), en de kwantitatieve taalwetenschap (hoe kunnen ontwikkelingen in statistische methodologie taalkundige problemen inzichtelijker maken?). Met uitzondering van laatsgenoemde staan deze deelgebieden momenteel op gespannen voet met elkaar. Psycho- en neurolinguïstisch onderzoek toont consequent aan dat sprekers en luisteraars zeer goed in staat zijn om te compenseren voor variatie in de uitspraak, terwijl de historisch-fonologische literatuur het er juist nagenoeg over eens is dat uitspraakpatronen veranderen omdat sprekers niet goed kunnen compenseren voor zulke variatie. De sociolinguïstische literatuur neemt op zijn beurt aan dat sprekers en luisteraars zich bovendien uitermate bewust zijn van verschillen in de uitspraak, en zich actief aanpassen, hetgeen uiteindelijk tot individuele verandering op de lange termijn zou moeten leiden. De kernvraag van dit proefschrift is hoe deze tegenstrijdige opvattingen van verschillende deelgebieden van de taalwetenschap met elkaar verenigd kunnen worden.

De verschillende hoofdstukken van het proefschrift hebben betrekking op verschillende aspecten van de onderzoeksvraag. Hoofdstuk 2 beantwoordt de eerste deelvraag: wat is de synchrone geografische verspreiding van de Polderverschuiving? Deze vraag wordt onderzocht door de bestudering van een omvattend corpus genaamd het "lerarencorpus". Dit corpus levert synchrone metingen van het gesproken Nederlands in vier representatieve gebieden in Nederland en vier in Vlaanderen. Een moderne statistische analyse van de data in het corpus levert een helder beeld op van de hedendaagse variatie in de Polderverschuiving: in Nederland heeft de gehele verschuiving zich zo goed als voltrokken, en in Vlaanderen is hier zo goed als geen sprake van. Hiernaast laat hoofdstuk 2 zien dat het mogelijk is om de vier veranderingen te kwalificeren volgens de vier grote typen klankveranderingen (fonetisch gradueel of abrupt en lexicaal gradueel of abrupt). Op basis hiervan laat hoofdstuk 2 zien dat de veranderingen in de Polderverschuiving inderdaad onderling verwant zijn, maar ook van elkaar verschillen: de fonetische veranderingen (veranderingen 1 en 2) lijken Neogrammariaans, maar de fonologische conditionering (verandering 3) lijkt eerder gebaseerd op *exemplars*. Voor verandering 4 zijn de data ontoereikend voor een duidelijke conclusie.

Het belangrijkste resultaat uit hoofdstuk 2 is dat er inderdaad sprake is van een klinkerverschuiving, die zowel grammaticaal (afhankelijk van een al-dan-niet-volgende coda /l/) als geografisch (Nederland tegenover Vlaanderen) geconditioneerd is. De overige hoofdstukken in het proefschrift onderzoeken hoe dit zo heeft kunnen komen. Dit wordt gedaan door middel van longitudinale psycholinguïstische experimenten onder *sociolinguïstische migranten*: Vlamingen die na de adolescentie naar Nederland verhuisd zijn. De deelvraag in hoofdstuk 3 is: nemen zij de Polderverschuiving over? Het hoofdstuk onderzoekt de relatie tussen fonetische accommodatie op de korte termijn en klankverandering op de lange termijn door tien sociolinguïstische migranten negen maanden te volgen. De resultaten van de experimenten bieden geen evidentie dat deze proefpersonen zich aanpassen aan de veranderingen in de Polderverschuiving in hun productie of in hun perceptie. Deze bevindingen suggereren, in tegenstelling tot eerdere literatuur, dat er een fundamenteel verschil is tussen accommodatie op de korte termijn en op de lange termijn.

Er zijn meerdere redenen denkbaar waarom er in hoofdstuk 3 geen evidentie is gevonden dat de sociolinguïstische migranten zich hebben aangepast aan de Polderverschuiving. Zo is het mogelijk dat negen maanden niet genoeg tijd is om veranderingen door te voeren, of zou het kunnen dat de proefpersonen zich simpelweg niet wílden aanpassen. Hoofdstuk 4 onderzoekt de verschillende mogelijkheden nader: wat is het tijdpad van de overname van klankveranderingen door volwassenen, en wat zijn de bijdragen van perceptie en productie daaraan? Hoofdstuk 4 onderzoekt deze deelvraag door dezelfde experimenten als in hoofdstuk 3 in te zetten, maar in een veel grotere steekproef over veel meer tijd. De opzet van dit experiment is noodzakelijkerwijs crosssectioneel, en het experiment maakt dan ook gebruik van twee grote Nederlandse en Vlaamse controlegroepen. De resultaten laten zien dat van achttien sociolinguïstische migranten die al tientallen jaren in Nederland wonen, tien zich zodanig hebben aangepast dat ze in een clusteranalyse bij de Nederlanders worden geclassificeerd, en de acht anderen niet. Dit resultaat suggereert dat de reden dat er in hoofdstuk 3 geen aanpassing gevonden is, is dat negen maanden gewoonweg onvoldoende tijd is om een klankverandering over te nemen. Er blijkt ook een betekenisvol verschil te zijn tussen perceptie en productie: waar er in de productie evidentie is dat de sociolinguïstische migranten zodanig heterogeen zijn dat een analyse op groepsniveau misleidend zou zijn, is er in de perceptie juist evidentie dat de groep homogeen is. Wel is het zo dat sommige van de individuele verschillen in perceptie significant correleren met de individuele verschillen in productie.

De laatste deelvraag in dit proefschrift gaat over de neurolinguïstische verwerking van de Polderverschuiving: is er ook verandering in het brein te zien van sociolinguïstische migranten? Hoofdstuk 5 beslecht een oud debat in de klankveranderingsliteratuur, namelijk of klankverandering te wijten is aan verschillen in *perceptie* of aan verschillen in *representatie* van spraakklanken. Dit wordt uitgezocht aan de hand van een EEG-taak, die laat zien of proefpersonen de gediftongeerde en de niet-gediftongeerde varianten van de bij-de-Polderverschuiving-betrokken klanken van elkaar kunnen onderscheiden. Tevens wordt er een controleconditie getoetst waarvan bekend is dat er grote verschillen in (sociolinguïstische) interpretatie spelen: de Gooische /r/. De proefpersonen zijn gelijk aan die in hoofdstuk 3. De resultaten van hoofdstuk 5 laten zien dat de sociolinguïstische migranten geen enkel probleem hebben de "nieuwe" klanken te horen, en er zelfs gevoeliger voor zijn dan de Nederlandse controlegroep. Dit verschil is vier maanden later echter verdwenen, wat laat zien dat de sociolinguïstische migranten de Polderverschuiving in die tijd zijn begonnen te leren. Bij de Gooische /r/ wordt een aanhoudend verschil tussen de groepen gevonden in de locatie in het brein waar de getrilde [ʀ] en de glijdende [ɹ] van elkaar onderscheiden worden: de sociolinguïstische migranten rekruteren andere hersengebieden dan de controlegroep. Dit wordt geïnterpreteerd als evidentie dat de Gooische /r/ voor hen minder de aandacht trekt. Beide bevindingen samen bieden evidentie dat klankverandering te wijten moet zijn aan het *interpretatie*proces, en niet aan het *perceptie*proces.

Hoofdstuk 6 biedt een tweede perspectief op de rol van het brein bij klankveranderingsprocessen. Dit hoofdstuk biedt empirisch bewijs voor het *on-line* verschil tussen compensatie voor fonologische variatie enerzijds, en acceptatie ervan anderzijds, door middel van een exploratieve EEG-studie op dezelfde groep sociolinguïstische migranten. De resultaten laten effecten zien die interessant kunnen zijn voor fonologen in het algemeen, maar die niet direct betrekking hebben op de Polderverschuiving. Er wordt een algemeen N400-effect gevonden van spraak met een regionaal accent, maar in de omgekeerde richting van eerder onderzoek: de sociolinguïstische migranten hebben *kleinere* N400-effecten wanneer ze naar Nederlandse spraak uit Nederland luisteren. In een controleconditie waarin /ɛ/ wordt uitgesproken als [ɛː] wordt bovendien een P600-effect gevonden, maar alleen in de Nederlandse controlegroep. Deze controleconditie was de enige conditie waarin er sprake was van een *fonemische* schending. De P600-bevinding in alleen deze conditie in alleen deze groep biedt verduidelijking aan het kleine hoekje van de wetenschappelijke literatuur dat eerder fonologische P600's gevonden heeft, door aanvullend inzicht te geven in het type schendingen waarin deze wel en niet gevonden kunnen worden.

Curriculum vitae

Cesko Voeten was born in Nijmegen in 1992. After attending his secondary education at the Titus-Brandsmalyceum in Oss, he obtained his BA (with honors) in Linguistics from Radboud University Nijmegen in 2013 and his research MA (with honors) in Language and Communication from Radboud University Nijmegen and Tilburg University in 2015. In that same year, he started his PhD at Leiden University, thanks to a "PhDs in the Humanities" grant from the Dutch Science Foundation ("NWO"). After finishing his PhD dissertation, he accepted a position as lecturer in sociolinguistics at Leiden University.