The actuation of sound change

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111
e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: Anne-France Pinget

ISBN: 978-94-6093-182-6
NUR: 616

The actuation of sound change

Hoe klankverandering op gang komt

(met een samenvatting in het Nederlands)

Le déclenchement du changement linguistique

(avec résumé en Français)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 4 september 2015 des middags te 12.45 uur

door

Anne-France Claire Hélène Pinget

geboren 30 juli 1987
te Charleroi, België

Promotor:	Prof. dr. R. Kager
Copromotor:	Dr. H. Van de Velde

Contents

My thank-you
or the art of baking a chocolate cake

Even though chocolate cake is my favorite, I need to confess that it is not always the easiest to bake. In the case of this one, it took me 4 years to prepare. I needed to be cautious with the ingredients, adding enough of sociolinguistics, not too much of phonetics and a well-balanced dose of statistics so that it can rise. Today, I am happy you can have some cake with me.

Before you eat the cake (and have it too), I want to go with you through a good recipe based on my experience in baking this one.

➔ First, don't think you can manage all by yourself (I am sometimes tempted to do so). No, you need good chefs to supervise you. I had two: Hans Van de Velde and René Kager. They both confidently guided me in this process. René and Hans, you offered me all the expertise, support and the encouragement I hoped for. René challenged me with the right questions at the right moment, keeping me (and probably Hans too) on the right track. He is the most enthusiastic researcher I have ever met. Hans managed to always be there for me, literally in Utrecht, Erps-Kwerps or Bologna, virtually from China, Indonesia or Leeuwarden (geographically less far, but evenly remote), and figuratively. He was not only a mentor for my research, but also a confidant for more private matters (which we usually discussed first) and a protector against the difficulties in Academia.

➔ Don't start the cake without a solid preliminary training. Mine started at the University of Namur (yes, you can get into research, even if you started at a French-speaking university). Namur will remain the best place ever to study languages (believe me, I have tested many). A special thank you to Laurence Mettewie who was the very first of a long series of professors who managed to arouse my interest in Linguistics. I will never forget the late evenings we spent at the fifth floor of the university building looking at my first experimental results. My training continued at the University of Gent where I met more professors of this kind, among them Johan De

Caluwe, who helped me show that it was possible for a French-speaking Belgian to contribute to the study of the language situation in Flanders.

- It is important to evaluate the cake once it is done and to get good feedback. So, have a bunch of tasters! In this matter, I would like to thank the committee members: Pam Beddor, Didier Demolin, Roeland van Hout, Hugo Quené and Wim Zonneveld. I appreciate your insightful comments a lot, and if you can't find all your comments incorporated in the text, I promise it will come in the next cake.
- You need nice colleagues and a good working atmosphere. I want to thank mine at the Utrecht University! There are too many to cite here and I am afraid I would forget some. But they will recognize themselves and know how much they meant to me. Along this line, it is very important that you have a good kitchen to work in. I spend much of my cooking time in the office with Marko Hladnik, Fang Li and Anna Chernilovskaya who I want to thank here. Our office was probably the quietest of the whole building, but at least we got some work done by the end of the day! Besides, I would like to give a heartfelt thanks to Hans Rutger Bosker, Aoju Chen, Sander van der Harst and Brigitta Keij who gave me advice, tools or feedback in the process, Hannah Hassig, Bert Le Bruyn and Julien-Paul Remy who accepted to go through the language use of this book, Theo Veenker who gave my the experimental and technical tools to bake the cake. A special thank to Martin Everaert who once introduced himself to me as the most Belgian professor in Utrecht (since he had − I cite − an <e> in the spelling of his name to form a long vowel). Martin, you gave me a lot of support in dealing with matters unrelated to the baking of the cake, but highly politically loaded.
- Besides, it is also good to have a bunch of nice colleagues from other universities, with whom you can go to conferences, to summer schools or just have a good beer in a Brussels park (or a meal in Zürich) when discussing research. Thank you for that!
- Crucially (and I wonder why I did not start with that), you need participants. They are the ingredients of your cake. A big thanks to all of them. They are 100 men and women in total (don't underestimate the importance of the elegancy of numbers, even in Linguistics!) from five different regions of Belgium and the Netherlands. The testing itself was often very boring (for them and for me), but somehow they managed to make it more exciting that you might think, like when the fire alarm went of in Ghent when I was testing a participant's imitation capacities, or when a 2 meter high guy from Groningen fainted in my arms, because he had not eaten before the experiment.
- Moreover, you need places to test the participants. It is like the mold of your cake. In sociolinguistics, this often means travelling. I am very grateful to the colleagues and friends of the five universities in which I tested, for helping me to find the right lab, the many participants and sometimes

even providing me with a bed and a good movie (you know which one, Anne-Sophie) after the many hours in the lab. At the Rijksuniversiteit Groningen, I thank Charlotte Gooskens and Stephanie Voigt, at the Radboud Universiteit Nijmegen, Margret van Beuningen, Roeland van Hout and Annemarie Weerman, at Leiden Universiteit, Jos Pacilly and Willemijn Heeren, at the Vrije Universiteit Brussel, the colleagues of the 'Slaaplabo', Elisabeth De Schutter and Marie Schellen and at Universiteit Gent, Anne-Sophie Ghyselen, Steven Delarue, Emmie Gunst, Valerie Rouquaert and Johan de Caluwe.

➔ I am very grateful to my paranymphs, Anne van Leeuwen and Pauline Pinget. I keep trying to explain to people outside Academia what the function of paranymph exactly involves. Well, Anne, Pauline, you do know this more than anyone else. Thank you for being there and helping me!

➔ You might think that cooking is a business thing, and in an ideal world, you want to keep business separate from your private life. At least that's what I always intend to do. But the further I got into cooking business, the more I realized how it was difficult to separate these things (No, research is not just like any other job). This is why I want to thank the people who have little or nothing to do with the baking of this cake (fortunately), but by whom I always can find fun, support, friendship, and love.

(This is the point where multilingualism gets relevant for sociolinguistics. If you don't understand the language, don't worry, this probably does not directly concern you).

Dank aan de vrienden in Utrecht voor de toffe momenten in deze stad en elders! Merci aux ami(e)s belges que je vois bien trop peu, mais qui me remplissent de joie chaque fois que nos chemins se croisent (une pensée spéciale pour les goumiches et les moins goumich(e)s). Merci aux amis du ski pour les supers moments sur et en dehors des pistes. Danke schön an D24 für die wunderschöne Zeit, die wir in vielen Teilen Europas gehabt haben und haben werden. Ihr seit sehr wichtig für mich! Una parte per miei coinquilini di Bologna, con qui ho passato una fine di tesi così bella. Mi avete fatto vedere la città, mangare e bere buonissomo e mi avete imparato l'italiano. Grazie mille a voi 5!

Les derniers seront les premiers. Et cette dernière pensée est pour mes parents et ma soeur Pauline, eux qui me connaissent mieux que n'importe qui. Vous m'avez toujours soutenue, à chaque moment, peu importe la route que je prenne! Merci d'être toujours là pour moi! C'est à vous que je dédie mon travail!

I hope you will enjoy eating this chocolate cake in the coming chapters. The chocolate may be more or less to your taste, but at least it's Belgian (I guarantee!).

CHAPTER 1

Introduction

VARIABILITY is a defining characteristic of human speech. No two voices are identical. No two utterances are the same. The presence of variability in everyday speech was already pointed out in the nineteenth century by Paul (1880). He observed – as many linguists after him – that there is massive variation in pronunciation between speakers, but also in the speech of a single speaker.

Part of this variation can be explained by articulatory constraints and natural laws of aerodynamics operating within the vocal tract (Ohala, 1983; Shadle, 2012; Stevens, 2000). Moreover, due to the inherent lack of precision in neuromuscular control of the articulatory organs, speakers are physically unable to hit the same target every time. An additional part of this variation appears to be language-specific and needs to be acquired. There is variation due to idiosyncratic features (e.g., voice, speech rate, styles), to sociogeographic features (e.g., origin, age, gender), and to linguistic contextual features (e.g., prosody, coarticulation).

While speakers are highly variable in their production, listeners tend to be highly robust to all these sorts of variation. Despite all the variation present in speech, individuals can easily juggle the multitude of speakers and still interpret sounds. Our perception device is highly flexible and adapts quickly to new voices, new accents, new dialects, new coarticulation patterns, etc. (Clopper & Pisoni, 2008). The paradox in speech perception is that – although the physical properties of stimuli vary continuously – people's perceptions of those stimuli are remarkably constant. This has been called the *invariance problem* (e.g., Perkell & Klatt, 2014), implying that there is no one-to-one relationship between the physical properties of speech sounds and their perception.

However, this well-working production-perception system as just described – with large variability in production but highly robust perception – is not stable. It is subject to change, another defining characteristic of human speech. As Labov argued, followed by many others, it is exactly this great variability in speech that provides the raw material for change (Hruschka et al., 2009; Labov, 2001; Labov, Yaeger, & Steiner, 1972; Ohala, 1989). Processes of speech production and perception generate what Ohala (1989) describes as a 'pool of variation', from which new phonological patterns emerge.

If change is a process intrinsic to language, and if it relies on variability omnipresent in the language, why isn't there change all the time? Conversely, why do languages sometimes change, when they could just as well not change? This question has been called the *actuation problem*.

The *actuation problem* of sound change was first defined by Weinreich, Labov, and Herzog (1968) as follows:

> Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times? (Weinreich et al., 1968, p.102)

The question raised is thus simple: why does sound change occur? And conversely, why does sound change not occur? In other words, actuation has to explain both change *and* stability.

This dissertation aims at contributing to a solution for the actuation problem by looking at the role of different aspects of the linguistic competence in sound change. In this chapter, we firstly situate the actuation question in the study of sound change (in Section 1.1), and secondly explain how our approach differs from previous research and how our study can contribute to a better understanding of the phenomenon of sound change (in Section 1.2).

1.1 The actuation problem in the study of sound change

Interest in sound change is as old as the study of linguistics itself. Garrett and Johnson (2013, p.10) have summarized the three main long-standing questions in the study of sound change. These questions are related to typology, conditioning and actuation matters.

- Typology: Why are some sound changes common while others are rare or nonexistent?
- Conditioning: What role do lexical and morphological factors play in sound change?
- Actuation: What triggers a particular sound change at a particular time and place?

Actuation is thus one of the central questions in the study of sound change in progress and − even five decades of research after its formulation by Weinreich et al. (1968) − still remains largely unsolved. Labov conceded that 'the question of why a change is initiated at a particular time and place [remains] the most recalcitrant' (Labov, 1982, p.81). Milroy was even less optimistic, characterizing the actuation problem as 'insoluble', because it 'implies the capacity to predict not only what particular change will happen, but also when and where it will happen' (Milroy, 1992, p.20).

It is commonly thought that the process of sound change cannot be studied straightforwardly due to its slow nature. Therefore, sociolinguistics have relied on extensive longitudinal documentation or on cross-sectional studies with age or social stratification (Hopper & Traugott, 2003; Labov et al., 1972). It is assumed that sound change follows a characteristic sigmoidal progression (Bailey & Haggard, 1973; Kroch, 1989; Labov, 1994, 2001). Change begins slowly, accelerates to a peak rate-of-change midcourse, and then slows down as the change is reaching completion. This suggests that the solution to the actuation problem must lie in the very earliest initiators of sound change. Baker (2008) for instance proposed that the leaders in the change occasionally misidentify linguistic variables and attribute significance to an accidental correlation of phonetic and social variables.

These ideas are supported by findings within the field of phonetics. Ohala's work (Ohala, 1981, 1983, 1989, 1993a, 1993b) developed the idea that sound changes originate in listeners' misperceptions. In this theory, sound change originates when a listener fails to compensate for coarticulatory effects. For example, a fronted realization of the vowel /u/ due to consonantal influences in a context such as /tut/ can be perceived as /y/ and not as /u/. The listener fails to attribute the fronting to coarticulatory effects and therefore misperceives the sound. It appeared attractive to base sound changes on misperceptions, making their occurrence accidental, and thus non-teleological.

A few years later, Lindblom's Hyper- and Hypo-articulation theory of speaker-listener interactions took a similar approach to coarticulation and its role in sound change (Lindblom, 1990; Lindblom, Guion, Hura, Moon, & Willerman, 1995). The Hyper and Hypo-articulation theory proposed that speakers vary their articulatory clarity according to the informational requirements of the listener. Speakers hyper-articulate when listeners require maximum acoustic information and reduce articulatory effort when listeners can supplement the acoustic input with information from other sources (Lindblom, 1990).

In conclusion, the view has prevailed in the field that sound changes are initiated by a phonetic mechanism, while the spread of these changes, however, is done by social means (Ohala, 1981).

1.2 Studying sound change differently

The goal of this section is to explain how the approach of sound change taken up in this dissertation differs from previous research. The specificities of our approach are described by answering three crucial questions:

- What is actuation?
- At which level do we study sound change?
- At which sound changes do we look?

1.2.1 What is actuation?

Many linguists have assumed that sound change relies on two events that are distinct in nature and occur sequentially: first, the creation of a new variant in speech variation called *innovation*, and second its propagation, *spread* in the speech community. This conceptualization of the actuation of sound change in terms of a two-step process of variation and selection draws inspiration from biological evolution (Yu, 2013). As pointed out by Labov (1994) however, Weinreich et al. (1968) did not distinguish between these two processes.

In this dissertation, we intend to step away from the concepts of innovation and spread, as it is peculiar to think of sound change relying on two distinct activities: a sort of creative act of deviating from the existing use (of which it is unclear whether it is even observable (Croft, 2010)), followed by the imitation of this innovation by other speakers. We will reconsider the process of actuation as an iterative process at the *individual level* in which minimal changes incrementally accumulate in a speaker's system every time he speaks to a listener. In this way, sound change can be redefined as a purely synchronic process, residing in the variability inherent to production and perception. The propagation of a change perceptible when comparing age groups or diachronic time slices is then merely a series of incremental acts of actuation in a defined direction. Actuation exists at every step. Change and propagation are the additive effects of such small steps.

1.2.2 At which level do we study sound change?

Despite the obvious close relationship between sociolinguistic research on production of variation and research on speech perception, sociolinguists and speech perception researchers have often been working in almost complete isolation from one another (Clopper & Pisoni, 2008).

Speech researchers typically use experimental methods developed in cognitive psychology to explore the perception of variation with discrimination, matching, identification or categorization experiments (Clopper & Pisoni, 2008; Coupland, Peter, & Nikolas, 1999). They aim at understanding and modeling how humans perceive and encode spoken language, and are faced with questions about the invariance problem and the role of variability in language processing.

In contrast to this psycholinguistic approach, sociolinguists typically describe natural variation as it occurs on social, regional, and ethnic levels. They examine variability in production and its social implications. As far as sound change is concerned, sociolinguists look at how the realization of a phonetic category changes over time for the language community as a whole (e.g., Labov, 1994). Some insights in the social characteristics of leaders in linguistic change have been provided in the work of Labov (2001) (e.g., leaders in change are typically female, members of the highest status group, upwardly mobile, with dense network connections, with a history of nonconformity in other respects, etc.). These insights however often remain qualitative observations.

Few systematic quantitative studies of change in adult individuals have been carried out (Foulkes, Scobbie, & Watt, 2010). This is most likely due to logistic difficulties in performing longitudinal studies on the one hand, and to the widespread assumption that linguistic patterns are essentially fixed once a speaker reaches adulthood on the other hand. However, there is evidence from the few available studies that linguistic changes can and do take place across the lifespan (e.g., Bauer, 1985; Nahkola & Saanilahti, 2004; Sankoff & Blondeau, 2007; Trudgill, 1988; Yaeger-Dror, 1994). A particularly famous example is the work of Harrington and colleagues on the speech of Queen Elisabeth II (Harrington, 2006, 2007; Harrington, Palethorpe, & Watson, 2005, 2000). They analyzed the Queen's speech over a period of 50 years through evidence from her annual Christmas broadcasts, and found that her speech has in some respects shifted in line with ongoing changes in Standard British English. More studies of this type on intra-speaker variability is required.

Overall, it seems that the study of sound change traditionally do not offer enough place for the individual. Weinreich et al. (1968) for instance critized theories that situate language change in the individual. They proposed that 'a language change begins when one of the many features characteristic of speech variation spreads throughout a specific subgroup of the speech community' (Weinreich et al., 1968, p.186). With a few exceptions (e.g., Abbs, 1986; Allen & Herron, 2003; Johnson, Flemming, & Wright, 1993; Mees & Collins, 1999; Syrdal, 1996), inter-speaker differences have received too little attention in the sociolinguistic and psycholinguistic studies on sound change. Inter-speaker differences are frequently treated as undesirable noise in the data or faded out by pooling or averaging data over speaker groups (Foulkes, Scobbie, & Watt, 2010).

In conclusion, we aim in this dissertation to give more weight to individual speakers and their linguistic background. The latest developments in the fields of statistics (e.g., the development of mixed-effects modeling (e.g., Baayen, Davidson, & Bates, 2008; Barr, 2013; Barr, Levy, Scheepers, & Tily, 2013; Quené & van den Bergh, 2004, 2008)) give us the tools to achieve this goal. We will focus on intra- and inter-speaker variability and bridge the existing gaps between the *individual* and the *group* and between *sociolinguistic, phonetic* and *psycholinguistic approaches.*

1.2.3 At which sound changes do we look?

Labov (1981) argued that sound change may be implemented either discretely or gradually. However, we observed that even so-called 'discrete' changes proceed somehow gradually (f.i., by means of lexical diffusion), so that change is always characterized by an incremental probabilistic shift of some sort. Phonetic changes, including those studied here do not proceed in large, discrete steps, but rather along continuous scatters of variability in individual systems. The gradual character of sound change is at the heart of this study.

Furthermore, sociophonetic research to date has been skewed towards English vowels (Foulkes & Docherty, 2006). Vowels are typically seen as carriers of sociolinguistic variation, and consequently too little attention has been paid to consonants. Moreover, consonantal variables have often been analyzed auditorily rather than acoustically (Foulkes, Scobbie, & Watt, 2010). By investigating fricatives and stops sociophonetically, our study therefore falls in with a recent line of studies which have begun to apply sophisticated analytic techniques to large consonantal data sets (e.g., Foulkes, Docherty, and Jones (2010); Stuart-Smith (2007) for English, Livijn (2002) for Swedish, and Kissine, Van de Velde, and van Hout (2003); Sebregts (2015); Tops (2009) for Dutch).

Finally, sociolinguists traditionally use longitudinal or cross-sectional designs in order to investigate phonetic contrasts they believe are undergoing sound change. Proceeding this way, there is a high risk that most described sound changes have already reached a high degree of completion upon their description and analysis. The current study remedy this by investigating sound changes with different degrees of completion: on the one hand a change already known to be in an advanced stage, and on the other hand new variation patterns which point at an incipient sound change. In this way, we aim to observe the iterative process of actuation at each step of completion.

In conclusion, this dissertation contributes to the actuation problem by improving shortcomings in the study of sound change which are related to the way actuation is defined, to the level at which sound change is studied (individual vs. group) and to the types of sound changes which are looked at.

The implementation of these research aims leads to the development of a comprehensive experimental design. The specific aim of bridging the gap between the individual and the group levels required an experimental design in which the crucial aspects of the linguistic competence involved in sound change are tested on the *same pool of participants.* Four aspects of the linguistic competence are chosen to be studied in the context of sound change. First, the processes of *speech perception* and *speech production* are investigated since each individual is in turn speaker and listener. In addition to the analysis of these two main processes, we look at two key mechanisms which are usually thought to play an important role in sound change: first, the fact that linguistic variables are constantly evaluated through *language attitudes* and second, the capacity of each individual speaker to imitate the pronunciation of others speakers through *phonetic imitation.*

1.3 A reader's guide

Chapter 2 presents a concise state of the art in the study of the four mentioned aspects of the linguistic competence involved in sound change: speech perception, speech production, language attitudes and phonetic imitation. This chapter formulates the main research questions of this study. *Chapter 3* describes the study in terms of its general design, the linguistic variables undergoing sound change and the pool of participants taking part in the series of experiments reported in the following chapters. This chapter is crucial for the understanding of the structure of the study. *Chapters 4, 5, 6* and *7* can each be read as separate reports of the conducted experiments in the study. Each chapter treats a separate aspect of the linguistic competence: speech production, speech perception, language attitudes and phonetic imitation, respectively. *Chapter 8* brings these four main themes together and establishing relationships which contribute to a better understanding of sound change as a whole. *Chapter 9* summarizes our findings and concludes the dissertation.

CHAPTER 2

Current state of the art

As explained in Chapter 1, we will tackle sound change and the actuation problem by looking at four aspects of the linguistic competence related to sound change: the speech perception, the speech production, the evaluation of linguistic variables through language attitudes, and the phonetic imitation capacities.

The current chapter presents the state of the art in the fields of speech perception (Section 2.1), speech production (Section 2.2), the link between perception and production (Section 2.3), language attitudes (Section 2.4), and phonetic imitation (Section 2.5). Each section leads to the formulation of the main research questions of the study.

2.1 Perception of variation

In the current section, different conceptions of speech perception are reviewed chronologically, from the traditional accounts (in Section 2.1.1) to more recently developed exemplar-based models (in Section 2.1.2).

2.1.1 First conceptions of speech perception

Traditional accounts of speech perception assume that the phonological representation is abstract, and that this representation needs to be extracted from the variability of the speech signals. Some of the earliest experimental work in the study of speech perception was conducted by Liberman, Harris, Hoffman, and Griffith (1957). They showed that listeners grouped sounds into discrete categories despite the continuous variation of sounds. Based on these results, they proposed the key notion of *categorical perception.* Speech perception was considered as a process of mapping highly variable inputs onto invariant abstract representations. Evidence for this position consists mainly of the fact that listeners are less likely to notice small phonetic variations within a phonetic category, while the same amount of variation is much more noticeable for sounds near a category boundary (Liberman et al., 1957).

Despite all the variation in speech, our perception is quite robust, making few miscategorizations. Within traditional accounts of speech perception, this was explained by the mechanism of *normalization.* Ladefoged and Broadbent (1957) showed that listeners adjust to speakers' anatomical differences. Normalization was therefore proposed as an early and automatic process of the speech perception device in which variation was removed, so that the meaningful content could be recognized (Pisoni, 1997).

In the nineties, Kuhl (1991) proposed an alternative speech perception model, the *native language magnet theory*, in which the phonetic categories of one's native language are organized in terms of prototypes or 'best instances' of specific speech sounds. These prototypes function as perceptual magnets since they attract or assimilate phonetically similar members. Consequently, the perceptual space is 'distorted' (Iverson & Kuhl, 1995): the best exemplars of a category pull neighboring tokens closer in the perceptual space.

Recently, researchers started to theorise that speech perception is much more than a normalization device, since it is used to identify voices, speakers, groups of speakers, etc. Variation in the speech signal is more than irrelevant noise that needs to be filtered out. Exemplar-based models were subsequently developed in this context.

2.1.2 The revolution of exemplar-based models

Exemplar clouds

Exemplar-based models, also called *epistimic models*, (Goldinger, 1997, 1998; Hawkins, 2003; Johnson, 1997; Pierrehumbert, 2001, 2002, 2003; Pisoni, 1997) take a different point of departure than most previous models of speech perception by not assuming that linguistic representations are stored solely in an abstract and invariant form. Instead, knowledge of linguistic structure is built up by representing in memory the totality of linguistic experiences that an individual has encountered (Foulkes & Docherty, 2006).

Linguistic knowledge thus consists of a detailed record of all of the exemplars of the words and the segments that an individual speaker has already heard. Each exemplar is simultaneously encoded with non-linguistic information about, for instance, who is speaking, what does the speaker's voice sound like, where does the speaker come from, what is his age, gender, ethnicity, etc. Consequently, each category is represented in memory by a large cloud, a distribution of remembered tokens of that category together with indexical information. Exemplars are organized in a cognitive map, so that exemplars of highly similar instances are close to each other and exemplars of dissimilar instances are far apart.

When a new stimulus is encountered, all past exemplars of the same category are activated in proportion to their similarity to the incoming utterance. Perceiving a sound thus means that the traces of that sound that are most activated come to consciousness and allow us for recognition (Goldinger, 1997). A perceptual category is thus defined as the set of all exemplars of a category and categorization is based on the similarity between exemplars within and between categories.

Unlike other previously mentioned theories of speech perception, which rely on the process of normalization, exemplar theory allows idiosyncratic representations to be stored in order to form a general representation of a stimulus (Goldinger, 1997). Social and linguistic information are stored in the mind in such a way that it can be accessed during speech perception.

The founding principles of exemplar-based models have been confirmed by a large pool of evidence. The models can explain why listeners perceive speech faster and more accurately when it is repeated by the same voice (Goldinger, 1996), by familiar voices (Nygaard, Sommers, & Pisoni, 1994), if it is repeated in the same intonation contour (Church & Schacter, 1994) or in the same speech style (McLennan, Luce, & Charles-Luce, 2003). Exemplar theory also provides a natural way of interpreting frequency effects (Tagliamonte, 2011), since each category is conceived as a distribution of exemplars.

Role of memory decay and individual experience

Importantly, the model proposes that all linguistic information is stored initially, but decays with time (Lacerda, 1995; Pierrehumbert, 2002). Because memories fade over time, recently encountered exemplars have more relative weight than exemplars that have not been encountered for some time. This decay is slowed down by frequent or recent activity (Pierrehumbert, 2001).

Consequently, exemplar representations continue to be updated throughout the lifespan of an individual (Foulkes & Docherty, 2006). Generalizations over exemplars depend on statistical learning experience (Pierrehumbert, 2002). Given that no two speakers are ever exposed to the same language input throughout their life course, perceptual patterns are likely to vary slightly across individuals. Exemplar-based models succeeded in modeling the influence of the ambient language and individual differences in speech perception.

Sound change in exemplar-based

Exemplar models allow the accumulation of phonetically-biased clouds of exemplars that permits sound change. The key aspect of exemplar-based models is that the representation of such a category, thus a cloud of exemplars, includes variants (Garrett & Johnson, 2013; Pierrehumbert, 2002). In this way, the cloud of exemplars may gradually shift as new variants are encountered.

Exemplars are stored together with fine phonetic details, so gradual phonetic drift is possible. Other models of the mapping between phonetic detail and linguistic categorization consider phonetic detail as 'noise' and discard it during speech perception. These theories therefore fail to explain how phonetic detail comes to play a role in sound change (Garrett & Johnson, 2013).

Perception of phonetic detail is dynamic and listener-specific, since it depends on the listener's experience. Exemplar-based models constitute a first step in the direction of accounting for this variability theoretically. Moreover, it offers a stronger theoretical basis for the investigation of the social patterns underlying linguistic use. However, additional work is required in order to fully understand the role played by sociolinguistic variation in speech perception, and especially how this might lead in some cases to sound change (Casserly, 2010; Thomas, 2002). Therefore, this study tries to answer:

RQ 1: What is the role of speech perception in sound change?

In Chapter 5, we investigate whether differences can be found in the perception of a changing sound within the same language area between and within groups (at the *group level*) and to what extent individual listeners differ in their perception of a changing sound (at the *individual level*). In Chapter 8, perception results are compared with other results in order to get more insights into the role played by perceptual variation in sound change.

2.2 Production of variation

The exemplar-based model has largely arisen from work on perception and categorization. The question of how exemplar representations relate to speech production has been explored relatively little (Foulkes & Docherty, 2006).

Pierrehumbert (2002) tried to extend the exemplar processing approach by including speech production. She proposed that production targets are randomly selected from the appropriate exemplar distribution. The probability that a particular exemplar is chosen for production is relative to principles of frequency and recency of activation. In contrast, others (e.g., Babel, 2009; Goldinger, 1997) have proposed that the produced exemplars are not chosen at random from the full set. Instead, all similar activated traces create a 'generic echo', regressing towards the mean of the category, and this echo is selected for production.

The main reason why it is still unclear how speech production functions in exemplar models is because there is still an ongoing debate about how speech perception relates to speech production and to what extent the two systems are intrinsically linked. The link between speech perception and speech production therefore constitutes the topic of the next section.

2.3 The relationship between speech perception and speech production

In this section, different conceptions of the relationship between speech perception and speech production are reviewed in two steps: the first models before exemplar-based theory (Section 2.3.1) and the dual representation model developed within the frame of exemplar-based theory (Section 2.3.2). Section 2.3.3 demonstrates the lack of and the contradiction of existing evidence related to the relationship between speech perception and speech production.

2.3.1 Before exemplar-based theory

A theoretical position held by many researchers is that an important link exists between the process of perception and production in the human cognitive system (Fox, 1982). One can imagine that perception and production are based on the same underlying phonetic categories, and that any changes in production would be accompanied by changes in perception, and the other way around.

This position has traditionally been adopted in theories preceding exemplar models, such as *Motor Theory* (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman et al., 1957; Liberman & Mattingly, 1985; Liberman & Whalen, 2000) or *Direct-Perception Theory* (Fowler, 1986; Goldstein & Fowler, 2003). Motor theory argues that adults perceive speech by making reference to their own articulation. They decode the speech signal by recovering the movements that have produced it. In the same way, the Direct-Perception theory suggests that listeners directly perceive the gestures or productions of the speaker. According to Liberman and Mattingly (1985), 'perception and production are only different sides of the same coin' (p. 30).

Later on, these ideas were further developed outside the speech domain (e.g., psychology and neuroscience) in which it was argued that perception and action are intimately linked (Hommel, Müsseler, Aschersleben, & Prinz, 2001; Rizzolatti & Craighero, 2004). These approaches are based on the existence of mirror neurons thanks to which the observation of another person's actions activates the same motor neural circuit as is activated when the action itself is performed (Iacoboni & Dapretto, 2006).

2.3.2 A dual representation?

With the development of exemplar theories arose the idea that speech perception and speech production might be intrinsically different. Speech perception is highly malleable and this great malleability was shown to be essential for comprehension. Production, on the contrary, is less flexible. Listeners may perfectly understand variants they cannot produce themselves (Thomas, 2000).

Garrett and Johnson (2013) therefore recently felt the need to propose a *dual representation model* in which perception and production are based on different sets of exemplars. Since the space of familiar exemplars for speech perception is necessarily larger and more diverse than the space of exemplars used in speech production, they distinguished between the two, proposing one phonetic space for listening and another one for speaking.

2.3.3 Insufficient and conflicting evidence

Although different schools have maintained different theoretical positions, there is still not a great deal of direct, empirical evidence bearing on the existence of a production-perception link (Frieda, Walley, Flege, & Sloane, 2000). For the most part, speech perception and speech production have been investigated independently. Accordingly, the existence and the nature of the link between the two activities has not been frequently addressed.

To date, only a small number of studies have attempted to examine the link between perception and production (Ainsworth & Paliwal, 1984; Bailey & Haggard, 1973; Bell-Berti, Raphael, Pisoni, & Sawusch, 1979; Evans & Iverson, 2007; Fox, 1982; Fridland & Kendall, 2012; Frieda et al., 2000; Grosvald, 2009; Grosvald & Corina, 2012; Harrington, Kleber, & Reubold, 2008, 2012; Janson, 1983; Johnson et al., 1993; Kendall & Fridland, 2012; Kleber, Harrington, & Reubold, 2011; Mitterer & Ernestus, 2008; Newman, 1997, 2003; Nygaard et al., 1994; Paliwal, Lindsay, & Ainsworth, 1983; Perkell et al., 2004; Sumner & Samuel, 2009). Moreover, these studies highly differ in their methodological approach of the question. Some studies focus on evidence gained from coarticulation processes (e.g., Harrington et al., 2012; Kleber et al., 2011). Others look at the relationship between listeners' judgments on category goodness and their own production (e.g., Frieda et al., 2000; Johnson et al., 1993; Newman, 2003). An alternative type of evidence comes from studies showing that different groups of speakers employ different articulatory strategies in achieving the same phonological goals (e.g., Bell-Berti et al., 1979; Ferguson & Quené, 2014; Fox, 1982).

Crucially, studies do not only differ in their methodologies, but also in their results. Additionally, most of these results have been prone to alternative interpretations or have been unreplicated in subsequent experiments. Some studies found evidence for a clear link between speech production and speech perception (e.g., Bell-Berti et al., 1979; Evans & Iverson, 2007; Fox, 1982; Fridland & Kendall, 2012; Harrington et al., 2008; Hay, Warren, & Drager, 2006; Janson, 1983; Kendall & Fridland, 2012; Kleber et al., 2011; Perkell et al., 2004). Bell-Berti et al. (1979) demonstrated that differences in production strategies for the /i/ vowel were significantly correlated with differences in perception. Looking at coarticulation, Harrington et al. (2008) and Kleber et al. (2011) found that younger speakers who produced vowels with fewer coarticulatory influences do also perceptually adjust less for these influences. In the same way, Perkell et al. (2004) found that speakers who articulated phonemically contrasting vowels more distinctly also showed a greater ability to distinguish vowel contrasts.

In contrast to this evidence for a clear link, other studies have shown evidence for a link only at a very abstract level (Mitterer & Ernestus, 2008) or even no link at all between perception and production (Ainsworth & Paliwal, 1984; Bailey & Haggard, 1973; Grosvald, 2009; Grosvald & Corina, 2012; Paliwal et al., 1983). Bailey and Haggard (1973) tested correlations between production and perception measures of English initial stops. They found no correlation between average VOT's produced in voiced and voiceless consonants and listeners' perceptual category boundaries for these consonants. Similarly, Ainsworth and Paliwal (1984) and Frieda et al. (2000) found no correlation between performance on production and perception tasks for English glides and the English /i/ vowel respectively. Grosvald (2009) showed that participants who produce more extensive coarticulation do not perceive differences between coarticulated variants more accurately.

Exploring to what degree a speaker's own productive system mirrors or interacts with his perceptual system is quite complex. Evidence is insufficient and contradictory. Therefore, the existence and the nature of this link are not entirely straightforward and require additional research in which it can be studied at both the group and the individual level. Hence, the central question in this study with regards to the relationship between speech perception and speech production is:

> **RQ 2: What is the link between variation in speech perception and variation in speech production?**

In Chapter 4, we investigate to what extent groups of speakers differ in their realizations of changing sound (at the *group level*) and to what extent production patterns differ individually (at the *individual level*). In Chapter 8, the relationship between the variability found in speech perception (from Chapter 5) and in speech production (from Chapter 4) is investigated.

If we assume that there necessarily exists some kind of relationship between perception and production (even at a very abstract level), it becomes important in the context of sound change to investigate which of the two processes is the first to be involved in the change.

Some researchers have put forward the idea that in the context of sound change, alternations in speech perception precede alternations in production. The reasoning is that the first step towards adopting a new form must be to hear it and identify it. In this way, the language user must first incorporate the new form in his perception and only later, if at all, he will start to use the form in production. Harrington et al. (2012) and Kleber et al. (2011) for example found results consistent with a model of sound change in which change in the perception of coarticulatory effects precedes change in production. They even imply a causal relation between changes in perception and changes in production:

> During a sound change in progress, the association between the perception and production of coarticulation passes through an unstable state during which the two modalities are out of alignment and in which changes to the coarticulatory relationships in perception lead those in production (Kleber et al., 2011, p.401).

However, others have argued for the opposite process: alternations of speech production precede alternations in perception. Janson (1983) found in a study of the /a:/-/o:/ merger in Swedish that change in production precedes change in perception:

> What seems to be clear is that, for an individual in a situation of change, perception seems to lag behind production. The old pattern of perception is still needed, even if one's own production is changed (Janson, 1983, p.31).

In the same line of thought, some have suggested that speech perception may not always be affected when there is a change in production (e.g., Evans & Iverson, 2007; Kraljic, Brennan, & Samuel, 2008). Listeners might also use phonetic cues in perception that they do not use in their own production, but that are used by other speakers in the community (Hay, Warren, & Drager, 2006; Thomas, 2000).

Faced with these contradictory conclusions, we feel the need to find additional evidence in order to answer the question:

> **RQ 3: Does change in speech perception precede change in speech production?**

This question is examined in Chapter 8 after having established the nature of the relationship between speech perception and speech production.

2.4 Language attitudes

The current section focuses on a first key mechanism in sound change: the evaluation of linguistic variables through languages attitudes. In Section 2.4.1, language attitudes are first defined. Section 2.4.2 concentrates on the link between language attitudes and sound change, and Section 2.4.3 situates language attitudes in the frame of exemplar-based theories.

2.4.1 Definition

In the field of social psychology, *attitudes* are defined as 'psychological tendencies that are expressed by evaluating a particular entity with some degree of favor or disfavor' (Eagly & Chaiken, 2007). The definition of attitude as a 'tendency to evaluate' currently tends to replace the definition of Allport (1935) of attitude as 'a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's responses to attitude objects or situations with which it is related'.

Inspired by the developments within social psychology, sociolinguists have acknowledged that *language attitudes* (i.e. 'attitudes towards languages, language varieties and language variants') (Knops & van Hout, 1988, p.1) are crucial, because they influence language behaviors and behaviors towards language. In fact, the social-indexical value of linguistic features appears to be inherent to language variation: 'No natural human utterance offers information without indexing some social factor' (Foulkes & Docherty, 2006, p.26).

Subjective evaluations of linguistic objects have been investigated in a number of sociolinguistic studies. As Thomas (2002) noted, the assessment of listeners' interpretations of linguistic variation dates back at least to Pear (1931). Choices of languages or language varieties, choice of allophones, and variation in phonetic parameters may all affect the way listeners judge a speaker (Foulkes, Scobbie, & Watt, 2010). Various techniques have been employed, including the *matched guise paradigm* (Lambert, Hodgson, Gardner, & Fillenbaum, 1960).

In the processing of results from attitudinal research, it is often found that ratings on linguistic variables correlate into two basic dimensions: *power* and *solidarity*. They are often found to be the two primary evaluative dimensions of language attitudes, as Brown and Gilman (1960) first demonstrated.

2.4.2 Language attitudes and sound change

Sociolinguists, with Labov as pioneer and leading figure in the field, have demonstrated that it is essential to investigate human communities and human interactions in order to fully understand how and why sound change takes place. Languages change, because linguistic communities are heterogeneous, and because speaker-listeners constantly *evaluate* competing linguistic forms (Foulkes, Scobbie, & Watt, 2010). Since variants have indexical meanings, their use may be more or less attractive, appropriate or valuable in particular social circumstances. Positively-evaluated variants generally spread at the expense of less positively evaluated competing forms. Nettle (1999) argued that language change can only occur when the social conditions are suitable.

The sociolinguistic notion of *prestige* is tied to the idea that individuals who lead linguistic change are those of higher status in their community (Labov, 2001; Labov et al., 1972; Milroy & Milroy, 1985). Social significance is associated with new variants via the influence of socially-relevant innovators within the speech community (Labov, 2001).

Labov has attempted to make sense of the apparent arbitrariness of many sound changes by determining the degrees of conscious awareness on the part of speaker-listeners (Foulkes, Scobbie, & Watt, 2010). Variables were categorized as *stereotypes*, *markers*, or *indicators*, in decreasing order of awareness. Different types of change may affect these different types of variables. A distinction was proposed between changes *from above* and changes *from below* (Labov, 1966; Meyerhoff, 2011). These terms refer to the degree to which the speakers in a speech community are consciously aware of the linguistic change affecting a particular linguistic variable. Change *from above* denotes a linguistic change above the level of a speaker's conscious awareness. Speakers are consciously aware of an ongoing change and they can comment on this linguistic innovation. On the contrary, a change *from below* denotes a linguistic change below the level of a speaker's conscious awareness. Trudgill (1972) introduced the notions of *overt* and *covert prestige*. *Overt prestige* is associated with the conscious use of linguistic forms that are recognized as standard and official. Speakers use them in order to be considered well-educated and intelligent. In contrast, *covert prestige* is associated with the use of non-standard forms. It is this prestige that comes with group solidarity and indicates a form of disobedience to authority.

Despite these attempts, the extent to which the development of awareness of the social-indexical value of phonetic variation is an explicit or implicit process remains a problematic matter (Foulkes & Docherty, 2006). In a similar way, it remains questionable whether phonetic forms can be shown to have language-specific and universal degrees of *salience* (Docherty, 2007).

2.4.3 Language attitudes within exemplar-based theory

Currently, the exemplar-based model of representation is the first theoretical framework in which indexicality is centrally embedded within phonological knowledge. As explained in the section above, the relationship between phonetic and social information during speech perception is central to the model and bidirectional (Drager, 2010). Perceived phonetic variants can affect what characteristics are attributed to a speaker, and the characteristics attributed to the speaker can influence how sounds are perceived.

However, exemplar theory still lacks a clear account of how sociophonetic information is represented cognitively. In particular, it is not clear how this information is stored and processed in relation to other types of information (Foulkes & Docherty, 2006). Moreover, it is still uncertain how the awareness and the salience of a linguistic variable can be accounted for. Some suggestions have been made by Drager (2011). She proposed that 'different individuals may consider different sociophonetic relationships to be more salient: they are more explicitly aware that individuals with certain social characteristics tend to speak a certain way' (Drager, 2011, p.118). Indeed, it would be logical that variants undergoing change were particularly salient to those people who have actually observed the progress of the change. When the sociophonetic relationship is more salient to an individual speaker, this speaker would have a greater amount of weight on the encoded relationship between social information and phonetic information. In any case, more research is required on the awareness and salience of linguistic variables and their implementation in exemplar theory.

More importantly, research on the origins of variation is often pursued without even considering the sociolinguistic aspects of change (Yu, 2013). The influence of social factors on language change in general has rarely been tested experimentally (Foulkes & Docherty, 2006; Lev-Ari & Peperkamp, 2014). While some understanding is provided by existing studies (e.g., Clopper & Pisoni, 2004; Evans & Iverson, 2004; Purnell, Idsardi, & Baugh, 1999; van Bezooijen & Gooskens, 1999), relatively little is known about how, and to what extent, listeners are able to make judgments about the indexical properties of linguistic features. Considering sound change in isolation from its social context through the study of articulatory, acoustic, or perceptual data solely, is likely to be less revealing than research incorporating the indexical meanings of variants and their role in the transmission of change (Foulkes, Scobbie, & Watt, 2010).

Therefore, the central question with regards to language attitudes is:

RQ 4: What is the role of language attitudes in sound change?

In Chapter 6, we investigate the extent to which listeners are able to make judgments about the indexical properties of a sound change in progress. It is examined to what extent a changing sound triggers prestige in the different groups of speakers (at the *group level*) and among individual listeners (at the *individual level*). In Chapter 8, the attitude results are brought into relationship with insights on the link between speech perception and speech production.

2.5 Phonetic imitation

The current section focuses on the second key mechanism in sound change: phonetic imitation capacities. In Section 2.5.1, the mechanism of phonetic imitation is defined. Section 2.5.2 concentrates on the link between phonetic imitation and sound change. Sections 2.5.3 and 2.5.4 situate the study of imitation in the frame of two theories: Communication Accommodation Theory (CAT) developed in the seventies within social psychology and exemplar-based theories recently developed within linguistics.

2.5.1 Definition

Imitation is the process by which individuals adjust to one another in social interaction. Imitation is central to human behavior (Meltzoff, Kuhl, Movellan, & Sejnowski, 2009), and has been observed at many levels, including postures, gestures, and facial expressions (Dijksterhuis & Bargh, 2001).

Within the domain of speech, imitation has been observed with respect to a great number of linguistic features. Individual speakers converge in lexical and syntactic alignment (Branigan, Pickering, & Cleland, 2000; Pickering & Garrod, 2004), speech rate (Giles, Coupland, & Coupland, 1991; Giles & Smith, 1979), pause and utterance duration (Jaffe & Feldstein, 1970), subvocal frequency and amplitude contour (Gregory, 1990), vocal intensity (Natale, 1975), etc.

Phonetic imitation, also known as *phonetic convergence* or *phonetic accommodation* (see Section 2.5.3 for the definition within social psychology), is the specific process by which speakers tend to make their speech more similar-sounding to that of the speaker they are interacting with (Babel, 2012). In this dissertation, we focus on the study of phonetic imitation, even if it is for convenience sometimes referred to as solely 'imitation'.

For a long time, studies on phonetic imitation have focused only on broad acoustic measures and non-contrastive phonetic traits like speech rate, voice quality or intonation (Babel, 2012), whereas linguistically-meaningful phonetic contrasts have largely been ignored. The results of recent studies however suggest that speakers also imitate and converge to meaningful phonetic properties such as VOT (Nielsen, 2007, 2011; Shockley, Sabadini, & Fowler, 2004), and vowel quality (Babel, 2009, 2010; Boves, 1992; Delvaux & Soquet, 2007).

Research has also revealed that phonetic imitation can occur in both socially minimal situations where talkers are simply producing isolated words (e.g., Babel, 2010, 2012; Babel & Bulatov, 2012; Goldinger, 1997, 1998; Goldinger & Azuma, 2004; Namy, Nygaard, & Sauerteig, 2002; Nielsen, 2011; Shockley et al., 2004) and in cooperative, socially-rich interactions (e.g., Pardo, 2006; Pardo, Gibbons, Suppes, & Krauss, 2012).

2.5.2 Phonetic imitation and sound change

The idea that imitation plays a central role in sound change can be traced back to the end of the nineteenth century in Sievers (1901) (Delvaux & Soquet, 2007). Trudgill (2008) claimed that phonetic accommodation is one of the main mechanisms by which a change is disseminated through a community. According to this idea, interacting individuals accommodate to changes, thus spreading innovations across the community.

Moreover, it has been recently claimed that accommodative processes can also provide the innovations themselves, as put forth by Babel, McGuire, Walters, and Nicholls (2014); Delvaux, Demolin, and Soquet (2004); Delvaux and Soquet (2007); Garrett and Johnson (2013); Pardo (2006). Sonderegger and Adviser-Goldsmith (2012) refer to this hypothesis as the *change-by-accommodation model*. In this model, the imperfect accommodation to the signal, together with constraints on perception and production which result in mutations to the signal, get passed on. While this account is highly appealing in order to explain sound change, it is difficult to test empirically.

Delvaux and Soquet (2007) described the role of imitation in sound change by linking the individual level and the group level:

> We think that this mechanism [*phonetic imitation*], occurring at the (inter)-individual level, may account for two apparently paradoxical features of the system: (i) the stability of phonetic realizations within a speech community, and (ii) their potential for sound change. Indeed, if speakers mutually imitate

> the productions they are confronted with, there is a wide range of potential variation in the individual phonetic realisations of phonological representations. [...] When speakers are interacting, their speech productions automatically tend to converge on the acoustic-phonetic level. The resulting phonetic variant is a hybrid realization, which is achieved by consensus among interlocutors. On the community level, the accumulation of such consensuses results in the particular set of speech habits, allophones, and control regimes that characterize the particular way of speaking of the speech community. The consensuses about phonetic realizations are constantly renewed, the corresponding mental representations are updated, and this on-going process is one of the starting points of sound change. Indeed, sound change occurs when there is an agreement within a speech community to use a phonetic variant different from the one used previously. (Delvaux & Soquet, 2007, p.146)

2.5.3 Phonetic imitation within Communication Accommodation Theory

While the phenomenon of phonetic imitation is currently typically approached from a sociophonetic perspective, it is important to note that it was previously extensively analyzed in its social function. In social psychology, there is an entire framework, *Communication Accommodation Theory* (CAT), that views imitation behavior as social acts that talkers use to modulate social distances in communication. CAT was initiated in the 1970's by Giles and his colleagues (Giles & Smith, 1979; Giles, Taylor, & Bourhis, 1973). This framework deals with the ways how individuals adjust their communication behaviors to one another, either to minimize the social differences (convergence) or to emphasize them (divergence).

The central assumption is thus that individuals modify their communicative behavior in order to achieve specific intrapersonal goals, such as social approval or more effective communication (Giles et al., 1991). Within this framework, production studies on style-shifts were widely conducted in the 1980's. Bell (1984) further explored potential causes for speech accommodation. His theory of *audience design* proposes that talkers shift speech styles in accordance with the audience (i.e., interlocutor, overhearers, eavesdroppers, referees, etc.), such that a talker's speech style reflects the demands of the setting.

Among others, Trudgill (1986) used CAT to account for change in social and regional dialects in contact. He proposed that individual short-term accommodation might become a long-term accommodation, which might spread throughout the community.

2.5.4 Phonetic imitation within exemplar theory

With its roots in the study of style-shifting, phonetic imitation is currently most often analyzed in a sociophonetic framework. The recent exemplar-based models of speech perception can account for imitative processes in the following way. Within an exemplar-based model, each word in the mental lexicon is associated with many auditory examples. Upon hearing a word, all associated episodic traces are activated in memory. The more familiar a voice or the higher frequency of a word, the higher the number of activated traces. This activation creates a 'generic echo' (see Section 2.2), regressing towards the mean of the activated set (Goldinger, 1997, p.46). It is the mean of the activated set that is selected for production. In terms of phonetic imitation, auditory exposure to words produced by a model talker shift a participant's productions towards those of the model talker.

In imitation studies, larger convergence effects have been observed among low-frequency words than high-frequency words (e.g., Goldinger, 1998; Nielsen, 2007). This is predicted by exemplar-based theories, because the smaller the number of exemplars associated with a given word, the larger the weight of each new perceived exemplar.

Importantly, imitation is not a systematic consequence of auditory exposure. In some cases, speakers do imitate, while in other cases there is no trace of imitation. Therefore, it is still unclear how and when linguistic variables are transmitted from the set of perception exemplars to that of production exemplars, especially in the case of a changing sound. This problem is intrinsically linked with the debate on the relationship between speech production and speech perception. Therefore, this study aims at answering the questions:

> **RQ 5: What is the role of phonetic imitation in sound change?**

In Chapter 7, it is investigated to what extent individuals imitate changing sounds (at the *individual level*) and whether this ability differs across group (at the *group level*). In Chapter 8, the imitation results are examined in relationship with insights on the link between speech perception and speech production.

2.6 Conclusion and research questions

In this chapter, we presented the state of the art in the fields of speech perception, speech production, the link between perception and production, language attitudes and phonetic imitation. Each of these themes was brought into relationship with sound change and with the theoretical framework of exemplar-based models. This resulted in the formation of five main research questions:

RQ 1: What is the role of speech perception in sound change?

RQ 2: What is the link between variation in speech perception and variation in speech production?

RQ 3: Does change in speech perception precede change in speech production?

RQ 4: What is the role of language attitudes in sound change?

RQ 5: What is the role of phonetic imitation in sound change?

Each of these main questions was formulated on the basis of theoretical matters. In the following chapters, these questions will be operationalized in more concrete subquestions tailored to the design of this study. The general design of the study is presented in the next chapter.

CHAPTER 3

The study

This chapter describes the methodology of the study. The study was designed in order to answer the research questions presented in Chapter 2. Section 3.1 describes the study in terms of its general design. Section 3.2 presents the linguistic variables selected for the study of sound change in progress. Sections 3.3 and 3.4 respectively define the *group* and *individual levels* by presenting the participants who took part in the series of experiments. Section 3.5 describes the data collection.

3.1 Design

In order to answer the research questions presented in Chapter 2, a *multi-experimental one-shot cross-sectional study* was designed. The study is multi-experimental, because it consists of a series of eight different types of experiments. The one-shot character of the design refers to the fact that the different experiments of this study were conducted with the same set of participants. The research questions required such a design in which the speech production, speech perception, imitation capacities and language attitudes of each individual participant could be linked. Finally, the study is cross-sectional as it looks at sound change synchronically by making comparisons between population groups which represent different stages of sound change. In this case, the population groups are five geographical regions described in Section 3.3.

3.1.1 A series of experiments

The study consisted of eight experiments conducted with all participants. The different experiments were spread over three experimental sessions. Participants took an intake questionnaire at the beginning of the first session and an exit questionnaire at the end of the third session. The type and the order of the experiments can be found in Table 3.1.

Table 3.1: The eight experiments, their order, the chapter in which they are reported and average duration.

Sessions		Experiments	Chapters	Duration
Session 1	1	Speech production recording	Production (4)	50 min
	2	Spontaneous imitation task	Imitation (7)	30 min
Session 2	3	Matched guise attitude task	Attitudes (6)	20 min
	4	Identification task (1)	Production (4)	15 min
	5	Similarity judgment task (1)	Perception (5)	25 min
Session 3	6	Identification task (2)	Perception (5)	15 min
	7	Similarity judgment task (2)	Perception (5)	25 min
	8	Forced imitation task	Imitation (7)	5 min

In the first session, participants conducted all production experiments followed by the spontaneous imitation task. Session 2 consisted of the attitude experiment and the two perception tasks related to the first target variable (see Section 3.2.1). In the last session, participants conducted the two perception tasks related to the second target variable (see Section 3.2.2) and the forced imitation task. The experiments were ordered by function of the degree of meta-linguistic awareness towards the variables required for the success of the experiments. During the first three experiments, participants were not aware of the test variables. In the four perception tasks, participants were directly presented with the variables. The last imitation task was designed to be inherently meta-linguistic, since participants were explicitly asked to imitate the variables.

The three sessions were planned on different days. For the majority of participants, sessions took place on consecutive days and in all cases the three sessions were completed within maximally one week. In total, participants spent approximately 3 hours and 15 minutes in the lab including questionnaires and instructions. They were paid 40 euro for their participation. To avoid any participants dropping out, they were paid after the third session.

3.1.2 Advantages and disadvantages of a one-shot design

In order to investigate the relationships between speech production, speech perception, imitation capacities and attitudes patterns at the individual level, a one-shot design was required. This means that the series of experiments described in the previous session was designed at once, and subsequently conducted at once by all participants of the study.

To our knowledge, no previous study managed to gather experimental results on these four aspects for the same group of participants. This innovation permits an extensive study of individual patterns and the role of these patterns in sound change.

Since all experiments were designed at once, some experimental risks are associated with this type of design. First, experiments cannot be adjusted to the results of previous ones in the course of the research. Moreover, if an experiment would fail for some reason, it would not be possible to re-run it. Participants could only be tested once, since they would be too much aware of the variables tested at the issue of the experimental series. Using new participants would not be possible either, as all experiments were conducted on the same group of participants. These pitfalls were avoided thanks to an expanded design phase and extensive piloting of all experiments.

3.1.3 A cross-sectional design

Traditionally, sociolinguistic studies on sound change with cross-sectional designs are based on the *apparent time method*. This method relies on the assumption that most linguistic features are acquired during childhood and remain relatively unchanged throughout an individual's lifetime. Therefore, the magnitude and direction of diachronic change can be inferred through the comparison of two or more age groups with the same linguistic background.

In this study, however, geographical regions were selected instead of age groups in order to represent the stages of sound change. The reasons for this choice are twofold. First, both linguistic variables undergoing sound change were shown to display regional variation (see Section 3.2). Second, as explained in Chapter 1, in this type of study we want to avoid making the assumption that speakers' speech remains constant across the lifespan.

The few previous cross-sectional studies on sound change used groups of speakers from either different age groups (e.g., Janson, 1983), or different geographical regions (e.g., Fridland & Kendall, 2012; Kendall & Fridland, 2012), and they proved to represent the stages of sound change successfully.

3.2 Variables

All experiments were designed to test two linguistic variables in Dutch. These variables were chosen to represent different stages of sound change in progress within the Dutch language area. The two variables under investigation are: the *devoicing of labiodental fricatives in word onset position* described in Section 3.2.1, and the *devoicing of bilabial stops in word onset position* described in Section 3.2.2.

Crucially, these two sound changes were chosen because they illustrate different degrees of sound change completion (see Section 1.2). The devoicing of labiodental fricatives illustrates a sound change in its advanced phase, whereas bilabial stops show devoicing patterns that can be associated with an incipient sound change.

3.2.1 The devoicing of labiodental fricatives in word onset position

Standard Dutch is traditionally described as having a phonological distinction between voiced and voiceless fricatives. The major cue for the voiced/voiceless distinction is the presence or the absence of vocal cord vibration in the fricative (Slis & Cohen, 1969; van den Berg, 1988). Moreover, voiceless fricatives tend to be longer (Debrock, 1977; Slis & Cohen, 1969; Slis & van Heugten, 1989) and louder (Kissine et al., 2003) than their voiced counterparts.

During the last decennia it was frequently observed that word-initial voiced fricatives were produced more and more often as voiceless (Cassier & Van de Craen, 1986; Cohen, Ebeling, Fokkema, & Holk, 1961; Donaldson, 1983; Goossens, 1974; Gussenhoven, 1999; Gussenhoven & Bremmer, 1983; Hamann & Sennema, 2005; Kissine et al., 2003; Kissine, Van de Velde, & van Hout, 2005; Mees & Collins, 1982; van der Wal & van Bree, 1992; Van de Velde, 1996; Van de Velde, Gerritsen, & van Hout, 1996). They observed that not all fricatives are equally advanced in the process of change. The devoicing of labiodental fricatives is not as advanced as back fricatives, but further along than alveolar fricatives (Van de Velde, 1996). In this study, labiodental fricatives were chosen for two reasons. First, labiodental fricatives hold an intermediate position within the fricatives undergoing devoicing, and second their devoicing is a well-described phenomenon in production.

These production studies reported significant geographical variation in fricatives. Slis and van Heugten (1989) recorded fricatives in two regions of the Dutch language area: the West of the Netherlands (the central area around the main cities) and the South of the Netherlands (Limburg), and found regional differences. The frequency of initial voicing was lower in the Western region than in the Southern. Van de Velde (1996) showed in a real-time study that fricative devoicing is a change in progress over the entire Dutch language area, but that it is more advanced in the Netherlands than in Flanders. Finally, Kissine et al. (2003, 2005) refined these insights by providing an overview of the differences between four regions in Flanders and four regions in the Netherlands.

In this study, Flemish realizations of the voiced fricatives differed from the Netherlandic ones in duration, noise and pitch. The voiced/voiceless labiodental contrast appeared to be almost completely absent in the Northern part of the Netherlands[1], since the speakers of this region hardly made any distinction in

[1] In some dialects of that area, words starting with voiced fricatives were originally realized with voiceless fricatives in Middle Dutch (Gussenhoven & Bremmer, 1983; van Reenen, 1994). It is thus possible that initial voiced fricatives in some places have never been realized as such.

pitch, duration and noise between /v/ and /f/. The region where the contrast is best maintained is West-Flanders.

In conclusion, there is large consensus and ample evidence that Dutch labiodental fricatives in onset position are undergoing a sound change. This change shows important regional variation, but overall it can be considered as *advanced*, but not completed at this moment.

3.2.2 The devoicing of bilabial stops in word onset position

Cross-linguistically, the main acoustic cue to distinguish between voiced and voiceless stops is Voice Onset Time (VOT). VOT indicates the difference in time between the release of a plosive and the onset of vocal cord vibration, and was first introduced by Lisker and Abramson (1964). The value of VOT is negative when voicing starts before the release of the stop. This is generally referred to as *prevoicing*. If the phonation starts at or after the moment of stop release, VOT is positive. A further distinction is often made between unaspirated and aspirated stops depending on the length of the positive VOT.

Dutch differs from most other Germanic languages in the way that the voicing distinction is phonetically implemented. Whereas other Germanic languages contrast voiceless unaspirated and voiceless aspirated plosives, the contrast in Dutch is made between voiced and voiceless unaspirated plosives. Voiced plosives are produced with a negative VOT and voiceless plosives are produced with little or no aspiration (Cohen et al., 1961; Keating, 1984). In a phonetic study of production and perception, Smits and van Alphen (2004) investigated the voicing distinction in Dutch initial bilabial and alveolar plosives. They collected read speech material of monosyllabic Dutch words and showed that indeed speakers generally implemented the contrast in terms of the presence or absence of *prevoicing* (negative VOT). Perceptually, prevoicing is also clearly the strongest cue that listeners used to classify plosives as voiced or voiceless. Besides prevoicing, a range of secondary cues (e.g., the burst duration, the intensity of the burst) appeared to play some role in the contrast. Even if Smits and van Alphen (2004) found clear differences in prevoicing in their speakers' plosives production, they noticed that prevoicing was absent in 25 % of the voiced plosives. Smits and van Alphen (2004) suggested that this pattern might be explained by sound change.

Since the study of Smits and van Alphen (2004), more evidence has become available to confirm that Dutch voiced stops in onset position are indeed undergoing devoicing. Ziliak and Van de Velde (2008) designed a sociolinguistic study in order to investigate this phenomenon. They looked at different styles (read and spontaneous) in the speech of Dutch language teachers. They found that the patterns of devoicing (voiced plosives realized without prevoicing) showed some geographic variation: devoicing was found in Flanders and not in the Netherlands, which points in the direction of an incipient sound change. They showed that plosives involved in these devoicing patterns mainly vary along two dimensions: duration of prevoicing and duration of the stop (both absolute and relative).

In conclusion, there is evidence that Dutch initial labiodental stops have recently entered a process of change. The change seems to show regional variation and will be called an *incipient* change in this dissertation.

Both phenomena (the devoicing of labiodental fricatives in word onset position in Section 3.2.1, and the devoicing of bilabial stops in word onset position in Section 3.2.2) are said to be *mergers* or *near-mergers*, in which sound change leads to the collapse of a phonemic contrast, so that two previously distinct phonemes come to a single phoneme (Labov, 1994). Interestingly, there are significant differences in functional load of these contrasts. Dutch only has a few minimal pairs with initial /v/ and /f/ (approximately 10 in total see Chapter 4) and some of these have a very low frequency in the Dutch language. Dutch bilabial initial stops, in contrast, show many more minimal pairs.

It is interesting to note that both phenomena seem to originate in different regions of the Dutch language area (central and Northern part of the Netherlands vs. central part of Flanders). For the sake of the present study, we would like to assume that the two phenomena are − from a sociolinguistic perspective − unrelated sound changes since they originate in different areas and are representing different stages of change (advanced vs. incipient) and − from a phonetic perspective − highly comparable phenomena. Hence they can be measured in similar ways and discussed together.

3.3 Regions

As explained in Section 3.1.3, this cross-sectional study is based on different geographical regions which are chosen to represent different stages of sound change. The aforementioned studies on regional variation patterns in the two chosen sound changes helped to target the right participants for the experiments. In Section 3.3.1, a short description of the Dutch language area is given. Section 3.3.2 describes the five regions within this area chosen for this study.

3.3.1 The Dutch language area

Dutch is a language spoken in Belgium, the Netherlands, and Surinam. However, this study is limited to the Dutch-speaking regions of the Low Countries (Flanders and The Netherlands). Traditionally, the Dutch language is described as having two pronunciation standards: one for the Netherlands and one for Flanders (Van de Velde, Kissine, Tops, van der Harst, & van Hout, 2012; Van de Velde, van Hout, & Gerritsen, 1997). This pluricentric nature is not ignored in this study. Nevertheless, we intend to consider the Dutch language area as one single language area as far as the variables chosen for this study are concerned. Although dialects of Dutch are still widely spoken in some regions, the aim of the study is to investigate patterns of variation solely in the pronunciation of Standard Dutch.

3.3.2 Five regions

Five regions within the Dutch language area were chosen to represent different stages of sound change: Groningen, South-Holland, Netherlands Limburg, Flemish-Brabant and West-Flanders. These regions are represented on the map in Figure 3.1 and described in Table 3.2.

Table 3.2: The five selected regions and their phase of the sound change for the two variables under consideration (based on the production data described in Section 3.2).

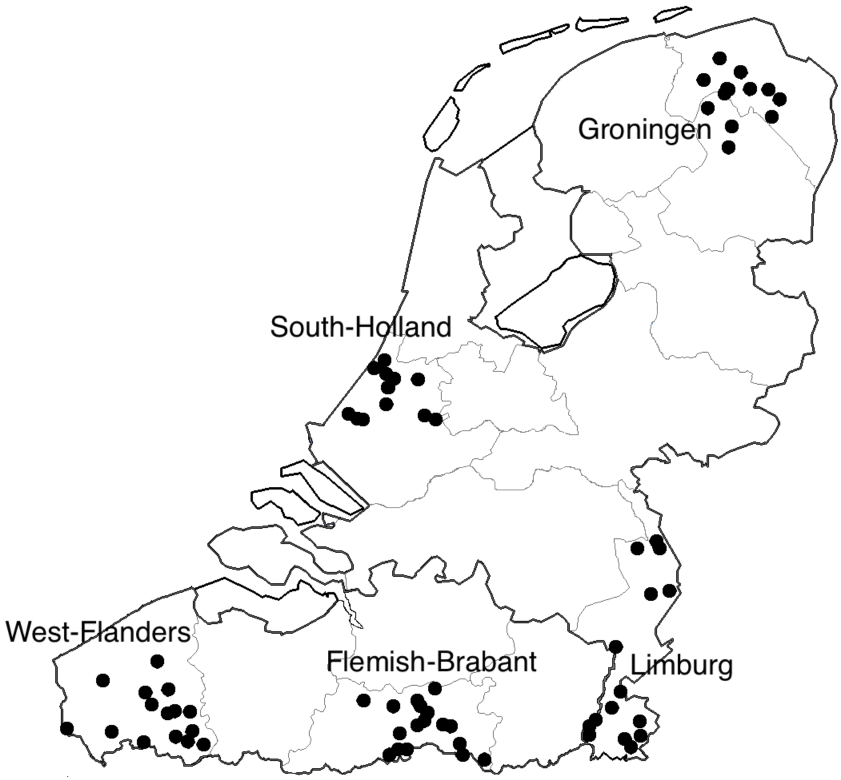
Regions	Country	Region within the country
West-Flanders	Belgium (Flanders)	Peripheral
Flemish-Brabant	Belgium (Flanders)	Central area
Limburg	The Netherlands	Peripheral South
South-Holland	The Netherlands	Central area
Groningen	The Netherlands	Peripheral North
Regions	**Phase of change (v)**	**Phase of change (b)**
West-Flanders	± completely voiced	incipient devoicing
Flemish-Brabant	incipient devoicing	incipient devoicing
Limburg	weak devoicing	± completely voiced
South-Holland	strong devoicing	± completely voiced
Groningen	±complete devoicing	± completely voiced

A region is central when it forms the economic and cultural centre of the community. Regions are peripheral when they are geographically maximally distant from the central region (see Van de Velde et al. (2012)). South-Holland is considered as the cultural and economic centre of the Netherlands and Flemish-Brabant the one in Flanders.

1. **West-Flanders (WF)** is a peripheral region to the West of Flanders along the North Sea. The chosen area was situated to the South of the province West-Flanders around the towns of Kortrijk and Roeselare. This region is known to be the most conservative in terms of fricative devoicing (Kissine et al., 2003, 2005). It is also showing the first traces of stops devoicing (Ziliak & Van de Velde, in progress).

2. **Flemish-Brabant (FB)** is the central area in Flanders, having a comparable economical, cultural and political status in Flanders to South-Holland in the Netherlands. First traces of fricative devoicing in Flanders have been found in this region (Kissine et al., 2003, 2005), as well as the first traces of stops devoicing (Ziliak & Van de Velde, in progress).

Figure 3.1: Map of the Dutch language area (The Netherlands and Flanders) and of the five selected regions. Each dot represents the origin of one or more participants.



3. Netherlands **Limburg (LI)** is a peripheral region situated to the South of the Netherlands. It essentially covers the province of the same name, stretching from Venlo to Maastricht. The fricative devoicing process is described as weak in this region (Kissine et al., 2003, 2005), and no trace of stop devoicing has been shown yet (Ziliak & Van de Velde, in progress).

4. **South-Holland (SH)** is the region representative of the so-called Randstad, the central area in the Netherlands consisting of the urban zone in the western provinces North-Holland, South-Holland and Utrecht. The chosen region in this study centers around the towns of Leiden and Delft. Production studies cited above have shown advanced processes of fricatives devoicing in this region (Kissine et al., 2003, 2005). No trace of stop devoicing has been shown in the region yet (Ziliak & Van de Velde, in progress).

5. **Groningen (GR)** is a peripheral region in the Northern part of the Netherlands. It is the region around the city of Groningen situated in the province of the same name and around Assen, situated in the province of Drenthe. It is known to be a region where the voiced/voiceless fricative contrast has almost disappeared (Kissine et al., 2003, 2005). No trace of stop devoicing has been shown in the region yet (Ziliak & Van de Velde, in progress).

3.4 Participants

The participants were one hundred native speakers of Dutch born and raised in these five regions (see Figure 3.1 where each dot represents the origin of one or more participants and see Appendix A). Within each region, 10 males and 10 females participated. The factors *age* and *educational background* were kept constant. The participants were all highly educated young adults. All of them were attending or recently graduated from a university or a 'hogeschool' (non-universitary higher education). They were aged between 18 and 28 years on the test day (mean=22.03) (see Table 3.3). The factor *sex* was taken in the analysis. Participant will be referred to with an ID number consisting of the abbreviation of the region to which they belong and a number from 1 to 20 (e.g., #FB02 is the second participant of Flemish-Brabant).

Table 3.3: Gender and age of the participants split up by region (N=100).

Regions	Gender		Age	
	Men	Women	Mean	Range
West-Flanders	10	10	22.45	20-25
Flemish-Brabant	10	10	21.40	18-26
Limburg	10	10	21.25	18-25
South-Holland	10	10	22	18-28
Groningen	10	10	23.05	20-28
Total	50	50	22.03	18-28

There was no significant difference in *age* between regions (ANOVA: df=4;95, F=2.161, p=.079). No participant reported having any reading or hearing difficulties. 83 participants were right-handed, and 17 left-handed. Nine participants (9 %) reported to have light dyslexia (West-Flemish 0/20, Flemish-Brabant 3/20, Limburg 2/20, South-Holland 2/20, Groningen 2/20). These participants were not excluded from the pool, since we aimed for a representative sample of the population in the Dutch language area and – in this population – dyslexia has been shown to have prevalence rates ranging from 5 to 10 % (Blomert, 2005).

Next to the factors *region* and *sex*, a range of other participant-related factors were obtained through the intake and exit questionnaires (presented in Appendix B). We distinguished between – on the one hand – factors meant to explain variance in the production/perception experiments (see Chapters 4 and 6): *dialect proficiency*, *regional accent*, *mobility* and *contact with the other country within the Dutch language area*, and on the other hand factors meant to explain variance in imitation capacities (see Chapter 7): *imitation ability*, *musical abilities*, *estimated musical abilities*, *estimated foreign languages proficiency*, *foreign languages use*, *conformism*, *sociability*, *talkativeness*, *egocentrism* and *conformism*. A definition of these factors and an explanation on how they were measured can be found in Appendix C.

Some regional differences appeared in the factors *dialect proficiency*, *regional accent* and *mobility*. These differences are presented in Table 3.4. In total, forty participants reported having active knowledge of a dialect. The proportion of the dialect users differed across regions (CHISQUARE: df=4, χ^2 =52.083, p<.001) and reflected accurately the current dialect use situation in Flanders and in the Netherlands with West-Flanders and Netherlandic Limburg showing high proportions of dialect speakers (e.g., Janssens & Marynissen, 2005; Vandekerckhove, 2009). Moreover, 64 % of the participants reported having a regionally marked accent when they speak Standard Dutch. The proportion of participants reporting a regionally marked accent also differed across regions (CHISQUARE: df=4, χ^2 =15.364, p=.004). The Flemish regions, together with Dutch Limburg, showed the highest rates. Moreover, participants differed across regions in their mobility score. The mobility scores ranged between 0 (not mobile) to 3 (very mobile) and were significantly lower for participants from Limburg and West-Flanders (ANOVA: df=4;95, F=4.186, p=.003).

Table 3.4: Dialect proficiency, regional accent and mobility score split up by region (N=100).

Regions	Dialect proficiency n	Regional accent n	Mobility score mean
West-Flanders	16/20	15/20	1.05
Flemish-Brabant	4/20	14/20	1.65
Limburg	17/20	18/20	0.90
South-Holland	0/20	8/20	1.40
Groningen	3/20	9/20	1.75
Total	40/100	64/100	1.35

Furthermore, a difference between language areas (Flanders vs. The Netherlands) appeared in the factor *foreign language proficiency*. The foreign language proficiency scores ranged from 0 (no proficiency in any foreign language) to 50 (very advanced proficiency in different foreign languages). Flemish participants achieved a higher score in general foreign language proficiency than Dutch participants (means: 22.18 vs. 19.02) (ANOVA: df=1;98, F=7.385, p=.008). This can partly be explained by the good language proficiency in French in Flanders in addition to the English proficiency.

3.5 Data collection

Participants were recruited at the university campuses of Groningen, Leiden, Nijmegen, Ghent and Brussels. They were recorded and tested in an experimental lab of these different universities between April 24, 2013 and November 22, 2013 (see Table 3.5).

Table 3.5: Dates and places of data collection split up by region.

Regions	Dates	Places
West-Flanders	24 April−13 May 2013	Ghent Universiteit Gent
Flemish-Brabant	5−22 November 2013	Brussels Vrije Universiteit Brussel
Limburg	9−25 September 2013	Nijmegen Radboud Universiteit
South-Holland	21 May−6 June 2013	Leiden Leiden Universiteit
Groningen	17 June−4 July 2013	Groningen Groningen Universiteit

All experiments were conducted using the ZEP experiment software (Veenker, 2012) on a laptop operating with Linux, a headphone Beyerdynamic DT 250 and a C420 head-mounted microphone. This equipment was designed for portability, while still providing excellent recordings. Since the exact same recording and computer equipment was used in the five regions, no apparent differences in the quality of the recorded speech signal and no difference in the subjects' performance related to the testing locations were expected.

3.6 Conclusion

In this chapter, we described the general design of the study. The study is a cross-sectional study articulated around a *group level* represented by five different regions and an *individual level* represented by hundred participants nested under these five regions. The regions represent different stages of two sound changes in progress (the devoicing of labiodental fricatives and the devoicing of bilabial stops in onset position). These variables were specifically chosen because they show regional variation and different degrees of sound change completion (advanced vs. incipient sound change). Twenty participants of all five regions took part in a series of experiments on language perception, production, imitation and attitudes. The results of each of these experiments will be presented in the next chapters.

CHAPTER 4

Speech production

This chapter reports on the production experiments of this study. The main goal of these experiments is to determine the stages of sound change by investigating the production patterns of the concerned variables. At the *group level*, the production experiments aim to replicate and refine the regional patterns of change described in Chapter 3. At the *individual level*, the aim is to collect the full range of phonetic realizations that individual speakers produce in different styles and to investigate individual differences in the use of the phonetic cues. The chapter concludes with a comparison of production variation patterns between the two sound changes.

4.1 Introduction and hypotheses

Our analysis of the production patterns of Dutch fricatives and stops is anchored in a number of studies described in Section 3.2, with which our results will be compared.

Fricatives Kissine et al. (2003, 2005) is the most recent detailed study of Dutch fricatives and constitutes our reference for labiodental fricatives. They investigated fricatives phonetically in four regions in Flanders and four regions in the Netherlands. They analyzed the read speech of 160 Dutch language teachers, stratified for community (Flanders vs. The Netherlands), region, sex and age. Their phonetic measurements of fricatives were very comparable to the ones in our study (see Section 4.2.2). As far as voicing and duration are concerned, Kissine et al. (2005) found that [v] was significantly shorter than

[f] (mean= 140.1 ms vs. 175.7 ms) and contained more voicing than [f] (mean= 57.5 % vs. 18.8 %).

As far as regional differences are concerned, Kissine et al. (2003) analyzed eight different regions; five of these regions roughly correspond to the five regions in this study. They reported that West-Flanders was the most voiced region with values around 80 % voicing for [v] and the North of the Netherlands was the most devoiced region with around 30 % voicing for [v]. The other regions, Flemish-Brabant, Limburg and Randstad took an in-between position in this order. For duration, they found that the Flemish regions showed shorter [v] (around 130ms) than the Dutch regions (around 150ms) and that the difference in duration between [v] and [f] was the biggest in West-Flanders and the smallest in the North. The three other regions followed the expected order. In this study, comparable regional differences in voicing and duration are thus expected. Moreover, we expect that the devoicing patterns are further advanced, since Kissine et al. (2003)'s data collection is already a decade old, and since we are testing the youngest adult generation.

Women are often considered to be the active agents of sound change (Labov, 2001). Slis and van Heugten (1989) observed a sex difference in both duration and voicing of fricatives: fricatives were much more frequently devoiced by women than by men. They struggled to find a consistent explanation and did not want to assume that it was caused by anatomical differences as Slis (1985, 1986) proposed. Kissine et al. (2003) found small sex differences in the production of fricatives, but did not discuss them in the paper.

Stops For Dutch stops, Lisker and Abramson (1964) is often cited as the reference with VOT measures for voiced and voiceless bilabial stops (-85 ms vs. 10 ms respectively). More recent production studies (Kager, van der Feest, Fikkert, Kerkhoff, & Zamuner, 2007; Simon, 2009) reported very similar values: mean VOT for [b]= -83 ms and -93 ms vs. VOT for [p]=13 ms and 12 ms.

The devoicing of stops in production is a much less well-studied phenomenon. Only Ziliak and Van de Velde (2008, in progress) preliminarily investigated regional differences. They used the same speech material as Kissine et al. (2003): read speech of Dutch language teachers. They measured 90 % voicing in voiced stops for Dutch regions and only 60 % voicing for the Flemish regions. No VOT measures were reported. Flemish-Brabant, which appeared as the leading region in the change, showed longer [b] than the other regions.

Regarding sex difference in stops, Smits and van Alphen (2004, p.464) reported that 'males produced prevoicing more often than females did' and that 'the effect of sex of the speaker is probably due to differences in the size of the vocal tract between men and women'. In conclusion, we expect to find sex differences in this study. However, we predict that these differences might rather be attributed to the fact that women often lead sound change than to anatomical differences.

In conclusion, we expect to find – at the *group level* – a range of differences in the production of fricatives and stops related to both the sex and region of origin of the participants. At the *individual level*, the aim is to collect the full range of phonetic realizations that individual speakers produce in order to investigate the use of the phonetic cues and in order to compare it to perception and imitation capacities.

4.2 Method

4.2.1 The production tasks

Our production tasks were designed in the tradition of the sociolinguistic interview as created by Labov (1966), who elicited reading styles and spontaneous speech assuming that styles can be ranged along a continuum of attention paid to speech (or level of monitoring). Participants were asked to produce the concerned variables in five different production tasks described below (word reading, carrier sentence reading, sentence reading, semi-spontaneous speech, spontaneous speech).

In the word reading task, the level of speech monitoring was meant to be the highest. In both the carrier sentence reading and the sentence reading tasks, the level of monitoring was also very high, but slightly lower than in the word reading task, since speakers also devoted some attention to content and prosodic patterns. In the semi-spontaneous and spontaneous tasks, speakers were mainly involved in *what* they were telling and not so much in *how* they were telling. As they had to decide on every aspects of the message they wanted to convey, the amount of attention was said to be the lowest. As explained by Blaauw (1995), read and spontaneous speech form the two extremes of the continuum: read speech has the highest level of preparedness, and spontaneous speech the lowest. The major methodological advantage of Labov's approach is that it provides a quick way to elicit a range of styles in a short amount of time, which are highly comparable between speakers.

The interview as it was designed in this study, however, differed from the Labovian approach in three important ways. First, the tasks were all designed to elicit the standard variety; it thus did not aim at triggering variables in sub-standard varieties as it is often the case in this kind of interview. Secondly, our aim was not to directly investigate stylistic variation, but the selection of styles was meant to represent the range of phonetic realizations within standard language. Thirdly, our speakers were – at no point during the production experiments – aware of the tested contrasts. In the three reading tasks, initial stop and fricative realizations were carefully balanced with filler words starting with other phonemes. It is only after the perception tasks (in the second session) that participants became aware of which contrasts were tested. Crucially, all production tasks were completed in the first session, whereas perception tasks were conducted in the two last sessions (see Section 3.1).

Although we did not aim to investigate stylistic variation (i.e., within speaker variation), we are aware of the difficulties posed by including both reading and speaking tasks in the same interview. Romaine (1988, p.1461) for instance remarked that reading is not necessarily the same type of elicitation because it is a direct representation of speech. Indeed, the reader is more aware of the correspondence between letters and phonemes. Orthography might influence pronunciation, especially in our study where the target variants are distinguished in spelling. However, because participants were not aware of the contrasts being tested, we assume that they did not rely more on spelling for the pronunciation of fricatives and stops than for other phonemes. Many phonetic studies used read speech only, which is often pejoratively called 'lab speech'. In contrast, this production study aims at eliciting speech that is phonetically measurable and that gives a good representation of the whole range of variation speakers produce.

It is also important to note that speakers were informed before the interview that the aim of the research was the pronunciation of Standard Dutch. It is very likely that speakers were co-operative, given that they were willing to join the interview and that it probably concerned a topic of their interest. Therefore they are expected to show pronunciation patterns that − according to them − approach the norm as closely as possible.

The five production tasks are described here:

Word reading Participants were asked to read aloud Dutch words. In total 160 words were presented one by one on a computer screen in a random order (see Appendix D). 78 words were test words chosen, containing the target variables in initial position followed by a vowel or a liquid (/r/ or /l/). 82 were fillers, either taken from the list for the spontaneous imitation task (n=43 see Chapter 7) or randomly chosen words beginning with others consonants than the target variables (n=39).

Only test words were analyzed. 74 words remained for the analysis (4 test words were removed): 37 beginning with a labiodental fricative and 37 beginning with a biliabial stop.

Carrier sentence reading This task was aimed at eliciting realizations of the target variables in a linguistic context that resembles the most the context proposed in the perception tasks (see Section 5.2.1). Target variables were therefore embedded in a monosyllabic non-word with a /CVC/ structure. The stressed vowel was /i:/, a vowel with an extreme position (high and front) in the vowel space. Moreover, /i:/ was shown to display least regional variation in Standard Dutch (van der Harst, 2011, p.159). The final consonants were /χ/, /t/, /k/, /p/, /f/, /m/, /n/, /s/ or /l/ (see Appendix E). The carrier sentence was 'ik neem de *target word*' [*I take the 'target word'*]. Participants were asked to read aloud these carrier sentences. In total 36 sentences were presented individually in a random order on a computer screen, 18 eliciting a labiodental fricative and 18 eliciting a bilabial stop. This task also serves as baseline production task for the imitation experiment (see Chapter 7).

Sentence reading Participants were asked to read aloud Dutch sentences presented in a random order on a computer screen. All sentences were declarative and contained one or more target variables (see Appendix F). 56 words were chosen containing the target variables in initial position followed by a vowel or a liquid (/r/ or /l/). They were placed in focus position in the sentences and in non-assimilation contexts. 28 words had a labiodental fricative and 28 words a biliabial stop in the onset.

Semi-spontaneous speech The semi-spontaneous speech production task consisted of two parts (see Appendix G). In Part 1, participants were given a picture of a big house with many different rooms and objects. Participants were asked to imagine that this house was their own and to describe it in as much detail as possible. In the description, they were required to name some objects colored in orange (e.g. een bal [*a ball*], een vaas [*a vase*]). In this way, we made sure that participants used at least a number of words containing the target variables in onset position. In Part 2, participants were given the picture of a labyrinth in which a bee [*een bij*] and a butterfly [*een vlinder*] try to find their way. The task was to describe the routes taken by both the bee and the butterfly. In order to orientate in the labyrinth, some small objects (e.g. fiets [*bike*], bus [*bus*]) were posted at key places in the labyrinth (at crossroads and intersections). Participants could easily describe the routes by referring to these objects. In this way, we made sure that they named at least some of these objects triggering one of the target variables in initial position. Both tasks were designed to trigger a maximum of stops and fricatives realizations. Due to time constraints, a maximum of 5 realizations per participant and per variant were analyzed for each task, resulting in 20 realizations of labiodental fricatives and 20 realizations of biliabial stops.

Spontaneous speech Spontaneous speech was elicited during an interview. Participants were asked to speak about some topics related to their daily life, such as studies, work, friends, hobbies, plans for future, etc. The interview lasted for approximately 15 minutes and was carried out by the researcher herself. The researcher is a female L2 speaker of Dutch with near-native level of proficiency. She belonged to the same age group as the participants. Since she had lived both in the Flemish and the Netherlandic community, she switched to either a more Flemish-sounding standard variety or to a more Northern Dutch standard variety, so that participants of any region could feel comfortable with the variety spoken by the interviewer and experience a small inter-speaker distance. As they possibly perceived that the researcher was a non-native speaker or as they did not perceive a clear regional background within the Dutch language area in the speech of the researcher, we did not expect any clear divergence or convergence effects in the interviewees' speech. The variety used by participants for this task was exclusively Standard Dutch. Dialect speakers realized the necessity of speaking a fairly standard variety in the context of this kind of formal situation. In this task, a maximum of 5 realizations per participant and per variant were analyzed, resulting in 10 realizations of labiodental fricatives and 10 realizations of biliabial stops.

Participants were seated in a sound-attenuated booth and conducted these five production tasks recorded through a C420 head-mounted microphone (more details see Section 3.5). For the read speech tasks, the words or sentences were presented on the computer screen placed in front of the participants. The stimuli remained on the screen until participants clicked on the space bar.

4.2.2 Phonetic measures of variables

All recordings were sampled at 48 kHz (24 bits). All realizations of the target variables were segmented and labeled into phonetic segments within the `Praat` speech-analysis software package (Boersma & Weenink, 2014). In order to achieve a high degree of consistency and a low error rate, the segmentation was done by the main researcher.

Fricatives Fricatives were segmented by assessing their centre of gravity (Gordon, Barthmaier, & Sands, 2002; Jassem, 1979; van Son & Pols, 1996). The center of gravity (CoG) is the average frequency over the entire frequency domain (calculated in semitones). CoG values are characteristically high for stops and fricatives, low for nasals and vowels. Since fricatives are characterized by the presence of noise, the onset of fricatives can be determined by the start of noise (rising CoG values) and the offset of fricatives by the end of the noise (falling CoG values).

Based on a Praat script, CoG values were determined for all fricatives. Subsequently, onset boundaries were placed manually where the rising slope in CoG started and the offset boundaries where the falling slope in CoG ended. Voiced fricatives usually showed a lower CoG than their voiceless counterparts, since they often are less loud. Despite this difference, boundaries of voiced fricatives were easily detectable with CoG patterns.

Two phonetic measures were calculated on segmented fricative realizations: VOICING and DURATION. As explained in Section 3.2, voicing is the major cue for the voiced/voiceless fricatives distinction in Dutch. Voicing values are expected to be higher in voiced than in voiceless realizations. Following Kissine et al. (2005), voicing was calculated by measuring pitch value (in Hertz) with intervals of 10 ms in the segment. The presence of voicing was evaluated between a minimum of 50 Hertz and a maximum of 400 Hertz. The number of measurements with presence of pitch was divided by the total number of measurements and multiplied by 100. This resulted in a relative voicing index, ranging from 0 (no voicing throughout the fricative) to 100 (voicing throughout the entire fricative). Secondly, the total duration of the fricatives were measured. Duration (in ms) was computed from the onset to the offset. Since Kissine et al. (2005) found strong correlations between relative duration measures (fricative/syllable and fricative/utterance) and absolute duration (respectively $r = .844$ and .771), we decided to use the absolute durations only.

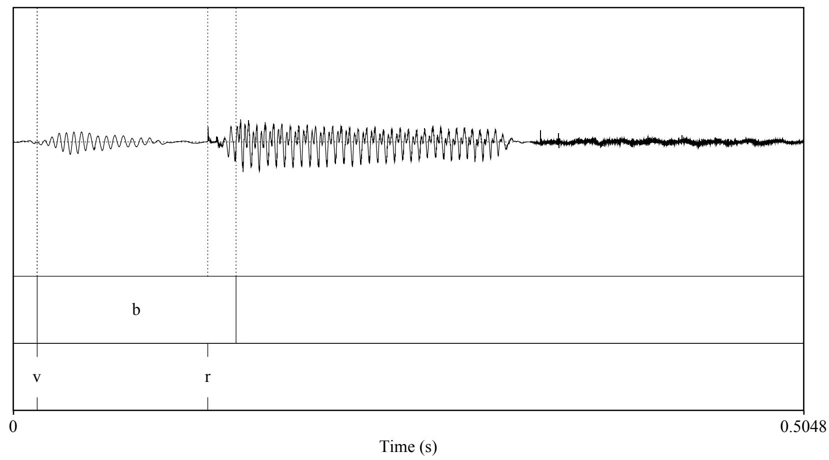
Stops Stops were segmented by defining their onset and their offset with the help of spectrographic display and waveforms. The criterion for indicating the onset of the stop was the point where the second formant (F2) disappears, which is basically the offset of the preceding vowel (Cho & McQueen, 2005). The criterion for indicating the offset of the stop was the F2 onset of the next segment, following Rojczyk (2011, p.117). Moreover, the peak of the stop release burst was determined. In case the stop showed two or more bursts, we followed Cho and Ladefoged (1999, p.215) who advised to measure VOT from the last burst. All measurements were taken at zero crossings.

Two phonetic measures were calculated on segmented stop realizations: VOT and DURATION. As defined in Chapter 3, VOT represents the interval of time in milliseconds (ms) between the stop burst and the onset of vocal fold pulsing. VOT was measured by subtracting the measurement of the onset of vocal pulsing from the measure of the release burst. Secondly, the total duration of the stops (in ms) was measured from the onset to the offset. Measuring the duration of voiceless stops is impossible when the stop is utterance-initial, because there is no cue about the onset of closure (Foulkes, Docherty, & Jones, 2010, p.67). This was the case in our word reading task, as words were read in isolation. Duration values for voiceless stops realizations were thus missing in that task.

As the segmentation was done by hand, the research had to determine the stop onset, offset and release burst by visual analysis of the spectrogram and waveform. For voiced stops, determining the end of the prevoicing was difficult in some cases, and the end of prevoicing did not always coincide with the burst release. In other words, there were cases where voicing faded away before the release or cases where voicing paused in the middle of the prevoiced part and started again before the burst. An example of such a voiced stop where voicing faded away before the release can be found in Figure 4.1.

Foulkes, Docherty, and Jones (2010, p.66) have described such a phenomenon in which the cessation of phonation leads to a short period of voicelessness in the prevoicing. They advised to treat these few cases as a separate category or omit them from further analysis. In view of the great number of these cases in our data and the presumption of sound change, we can only interpret this as a type of devoicing of the voiced stops. This phenomenon is not as strange as it might appear, as it corresponds to devoicing patterns in fricatives where pitch is absent in the middle of the segment, whereas the onset and offset are fully voiced. Moreover, it was already observed by Ziliak and Van de Velde (2008). A major consequence of this finding is that VOT turned out to be an unreliable measure for the concerned stops. When a [b] realization happened to show devoicing in the middle of the prevoicing, it became unclear whether it had to be considered as prevoiced with the onset of voicing at the end of the preceding vowel or whether it should be considered as a voiceless realization with an onset of voicing at the beginning of the following vowel. VOT as a measure lacked the flexibility to attest this gradual phenomenon.

Figure 4.1: Wave form of the word 'beer' (*bear*) produced by participant #FB04 in the word reading task. *b* stands for the target variant, *v* for the onset of voicing and *r* for the release.



The important number of the devoiced cases led to the need of introducing another measure for stops that could capture the gradience of the devoicing phenomenon: VOICING. Voicing was measured in stops in the exact same way it was measured in fricatives. Pitch values (in Hertz) were calculated with intervals of 10 ms between the onset and offset of the stops. The number of measurements with presence of pitch was divided by the total number of measurements and multiplied by 100, resulting in a relative voicing index, ranging between 0 (no voicing throughout the stop) and 100 (voicing throughout the whole stop)[1].

4.3 Results

A first general remark for both fricatives and stops in our data is that there was a great amount of idiosyncrasy. Phonetic production details were often peculiar

[1] For both stops and fricatives, we obtained raw percentage data along the voicing dimension. These percentage data were not normally distributed, but were analyzed without applying any transformation. The arcsin or logit transformations which are traditionally used to transform percentage data did not result in normal distributions. However, the distributions of the measures computed in Section 4.3.2 on the aggregated data, which will be used further in the coming chapters satisfactorily approached normality.

to an individual and were consistently used by this individual in the majority of his realizations. For instance, some participants only devoiced fricatives in the middle of the fricative segment or others systematically produced multiple bursts in voiceless stops.

For both fricatives and stops, a conservative way of removing outliers was handled. All observations greater than four standard deviations from the mean of the measurements were removed (instead one or two standard deviations which is standardly used in this type of analysis). In this way, the procedure managed to remove extremely deviant observations, errors in measurements and mistakes participants made. Very long, very short or other atypical realizations are a normal part of speech and were conserved in the data. The remaining number of observations can be found in Table 4.1.

Table 4.1: The number (n) of fricative (in the upper part) and plosive (in the lower part) observations split up by speech style.

		Fricatives	
Speech styles	n per participant	Total n	n without outliers
Word reading	37	3700	3686
Carrier sent. reading	18	1800	1793
Sentence reading	28	2800	2790
Semi-spontan. speech	20	2000	1157
Spontaneous speech	10	1000	835
Total	113	11300	10261
		Stops	
Speech styles	n per participant	Total n	n without outliers
Word reading	37	3700	3700
Carrier sent. reading	18	1800	1790
Sentence reading	28	2800	2772
Semi-spontan. speech	20	2000	1656
Spontaneous speech	10	1000	886
Total	113	11300	10804

In total, 10261 observations were analyzed for fricatives and 10804 for stops. Each participant recorded a maximum of 113 fricative realizations and 113 plosives. About one third of the realizations was provided by the word reading task. Since the processing of semi-spontaneous and spontaneous speech is more time-consuming than the processing of read speech, the number of observations provided by these two tasks was limited. The semi-spontaneous and spontaneous speech tasks contained more outliers than the other tasks.

The frequency of occurrence of each word in the word and sentence reading tasks was obtained from SUBTLEX-NL[2]. A linear regression model was computed

[2]SUBTLEX-NL is a database of Dutch word frequencies based on 44 million words from film and television subtitles (Keuleers, Brysbaert, & New, 2010). In previous linguistic research, this database proved to be an efficient and easily manipulable tool in order to measure Dutch word frequencies (e.g., Bosker, 2014).

in order to predict the different phonetic measurement of fricatives and stops (voicing and duration) by the log frequency of each word. The model did not reveal any significant effect of frequency. Therefore, we concluded that there was no lexical diffusion effect for the target variables. In this study, Dutch words were selected on the basis of their phonetic features and not of their frequency. More research with a better distribution of the log frequencies of the stimuli is required to confirm the absence of a lexical diffusion effect.

4.3.1 A detour through within-individual differences: differences in speech styles

Even if within-individual differences are not the focus of this study, we first have a look at the results of the five different speech styles before turning to the analyses at the individual and group levels.

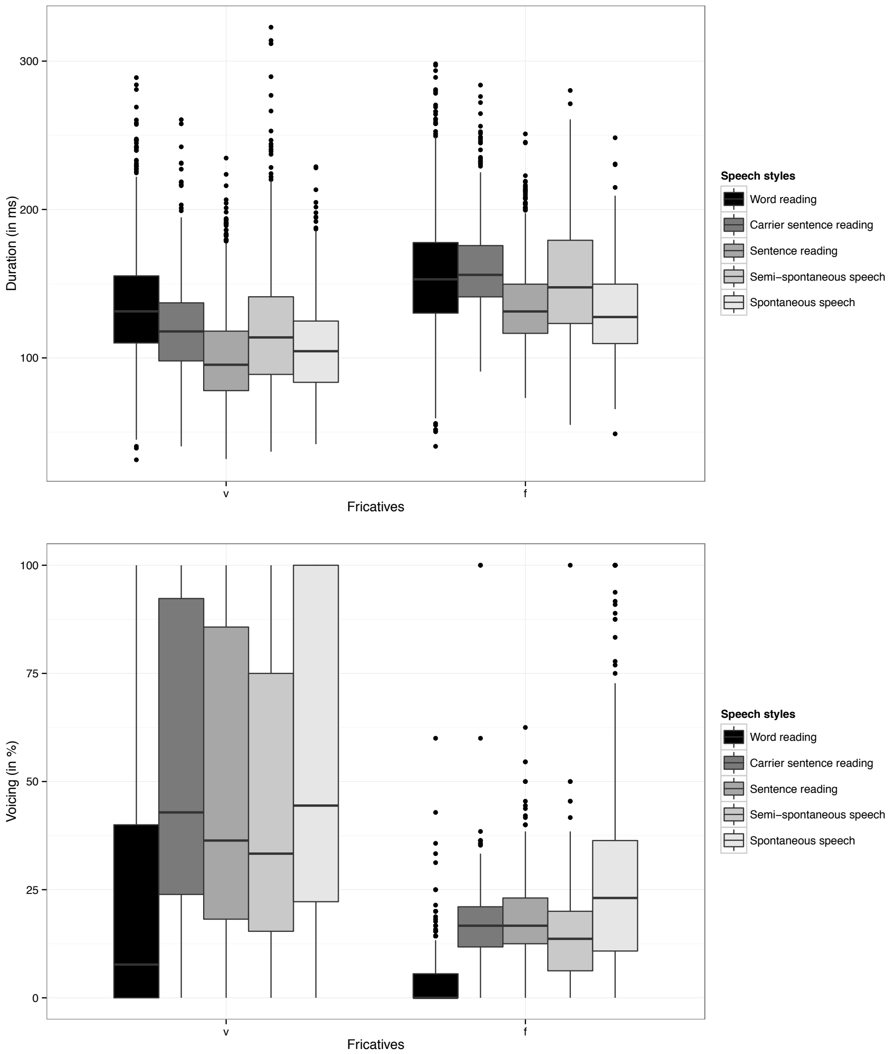
Fricatives The results for fricatives (n=10261) divided by speech styles are found in Figure 4.2 on the next page and in Appendix H. Overall, it turned out that target voiced fricatives were systematically shorter (116 ms vs. 147 ms) and more voiced (43.6 % vs. 15.6 %) than their voiceless target counterparts. These values roughly correspond to the values reported by Kissine et al. (2003), even if voiced fricatives seemed to have devoiced even more in ten years time (43.6 % in our study vs. 57.5 % in Kissine et al. (2003)).

As far as speech styles are concerned, both voiced and unvoiced fricatives are shorter in more spontaneous speech styles. This is probably due to the fact that we measured absolute durations and – in this way – did not normalize for speech rate. Voiced fricatives showed large variation in voicing measurements. All styles triggered a comparable amount of voicing in voiced fricatives, except the word reading task in which voiced fricatives were usually highly devoiced. In this task, words were presented in isolation and it is therefore articulatory difficult to produce voicing from the very beginning of the fricatives onwards. In the same way, voiceless fricatives were more voiceless in the word reading task. In spontaneous speech, voiceless fricatives were produced with more voicing than in the other tasks. Producing fully voiceless fricatives intervocalically appeared to be challenging in fast speech due to coarticulation.

Stops The results for stops (n=10804) split up by speech style can be found in Figure 4.3 and in Appendix H. Voiced stops were systematically shorter (102 ms vs. 136 ms) and more voiced (75.8 % vs. 14.9 %) than their voiceless counterparts[3].

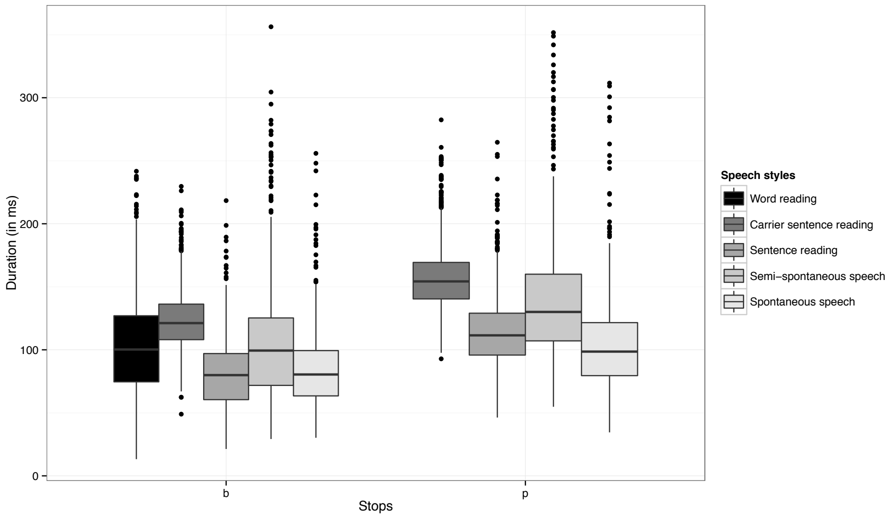
[3]Although not reported here (see explanation in Section 4.2.2), VOT values of measurable tokens were very similar to the reference values reported in Section 4.1. Voiced stops showed negative VOT (prevoicing), whereas voiceless stops had a negative VOT (-77 ms vs. 19 ms).

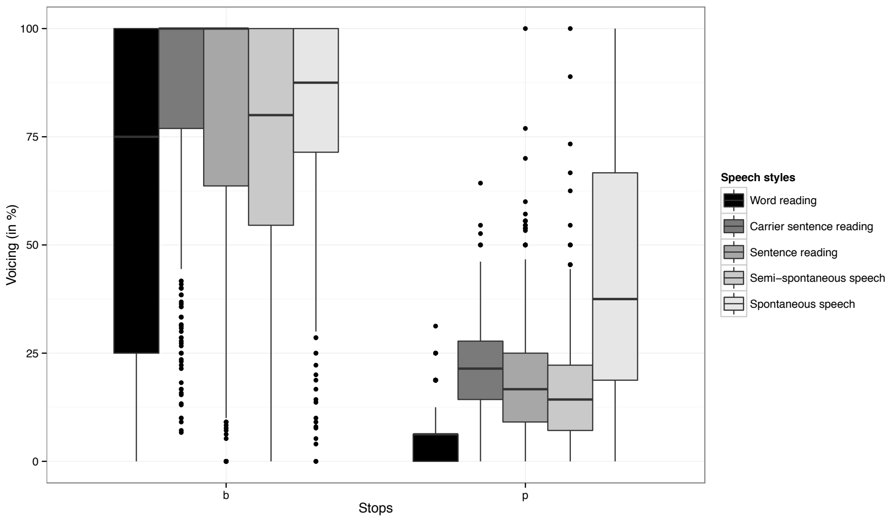
Figure 4.2: Boxplot of fricative duration (in the upper panel) and voicing (in the lower panel) split up by speech style (n=10261).



Note. In boxplots, the median is shown by the horizontal line, the first and third quartiles by the box and the 95 % confidence interval of the median by the vertical line.

Figure 4.3: Boxplot of stop duration (in ms) (upper panel) and voicing (in %) (lower panel) split up by speech style (n=10804).





Both voiced and unvoiced stops got shorter in more spontaneous speech styles, similar to the fricatives. The sentence reading task, however, patterned more with the word reading tasks than expected. In the word reading task, there was no result for duration in voiceless stops, since it was impossible to measure, as explained in Section 4.2.2. Like voiced fricatives, voiced stops showed large variation in voicing. All styles triggered comparable amount of voicing in voiced stops, except the word reading task in which voiced stops were more devoiced. Also voiceless stops were more voiceless in the word reading task. In the spontaneous speech however, voiceless stops were produced with more voicing than in the other tasks, probably for the same reason as fricatives (i.e., the difficulty to maintain voicelessness intervocalically).

In conclusion, some differences appeared between phonetic realizations of fricatives and stops across different speech styles. Most notably the word reading task elicited realizations slightly diverging from the realizations in the other tasks. Yet, realizations across all speech styles were taken together for the further analysis, a choice which can be justified by two reasons. First, this study does not investigate within-speaker differences (i.e., speech style differences). In contrast, the goal is oriented towards the measure of between-group differences and the measure of the complete range of phonetic realizations an individual is able to handle across styles. Second, correlations between styles turned out to be high or very high (ranging between r=.70 and .97 in voicing), which validates the aggregation across speech styles. Hence, for further analysis, observations were aggregated, resulting in one value per individual and per variant (and referred to as *aggregated data*).

4.3.2 Three key measures

In order to look at group and individual differences, three measures were computed on both the fricatives and stops data for both voicing and duration:

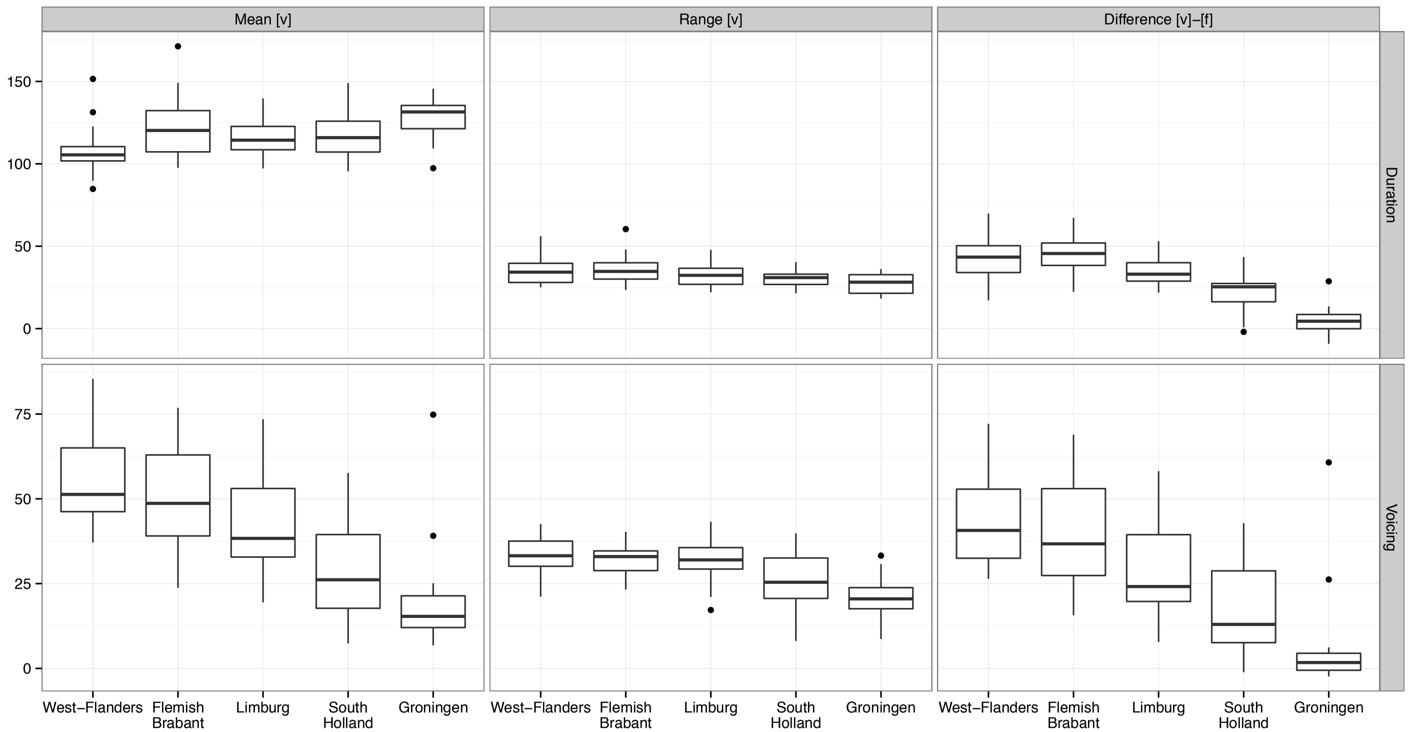
- *The Mean [v]* and *Mean [b]* is the average measurement of a voiced realization in the aggregated data.
- *The Range [v]* and *Range [b]* represents a measure of the range of the voiced category. It was measured as one standard deviation on the non-aggregated data.
- *The Difference [v]-[f]* and *Difference [b]-[p]* is the difference between the average realization of the voiced variant and the average realization of the voiceless variant in the aggregated data.

4.3.3 Effects between groups

Fricatives Regional differences between these three measures for fricative realizations are presented in Figure 4.4. It is a boxplot with duration measures in the upper panel and voicing measures in the lower panel.

As far as duration is concerned, slight regional differences appeared first in the mean [v] duration. Participants from Groningen produced the longest realizations, whereas in West-Flanders they were the shortest (with the other regions in between).

Figure 4.4: Boxplot of fricative duration (in ms) (in the upper panel) and voicing (in %) (in the lower panel) split up by region. These measures are 1) the mean and 2) the range of voiced labiodental fricatives, 3) the difference between voiced and voiceless labiodental fricatives (N=100).



The range of [v] durations was relatively constant across regions (around 40ms). Regarding the difference in duration between [v] and [f] realizations, the difference was the largest in West-Flanders and Flemish-Brabant and tended to decrease in the Dutch regions to reach almost 0 ms in Groningen. This means that participants from Groningen – on average – did not use duration to contrast between voiced and voiceless fricatives.

Voicing measurements showed large regional differences. First, there were differences in mean voicing of [v]. West-Flemish participants produced [v] with the most voicing (average 56 %), already showing a high degree of devoicing. The amount of voicing in [v] realizations decreased across regions and reached less than 20 % voicing in Groningen.

Similarly, the range of voicing in [v] showed small differences across regions. In West-Flanders, Flemish-Brabant and Limburg, ranges came to more than 25 %, whereas South-Holland and Groningen presented smaller ranges. The difference in voicing between [v] and [f] realizations is the largest in West-Flanders and the smallest in Groningen (with the other regions in between). Interestingly, the difference almost reached 0 % in Groningen, indicating that these participants hardly make any voicing distinction between [v] and [f].

A traditional analysis of variance[4] was fitted to the data of voiced fricatives. This analysis revealed a main effect of *region* (ANOVA: df=4;95, F=18.780, p>.001)[5]. The Tukey post-hoc test showed that all regions significantly differed from each other, except the combinations West-Flanders/Flemish-Brabant and South-Holland/Groningen (West-Flanders: 57 %, Flemish-Brabant: 50 %, Limburg: 43 %, South-Holland: 30 %, Groningen: 20 % in voicing for [v]).

Stops The measures for stops realizations are presented in Figure 4.5. The figure is a boxplot with duration measures in the upper panel and voicing measures in the lower panel.

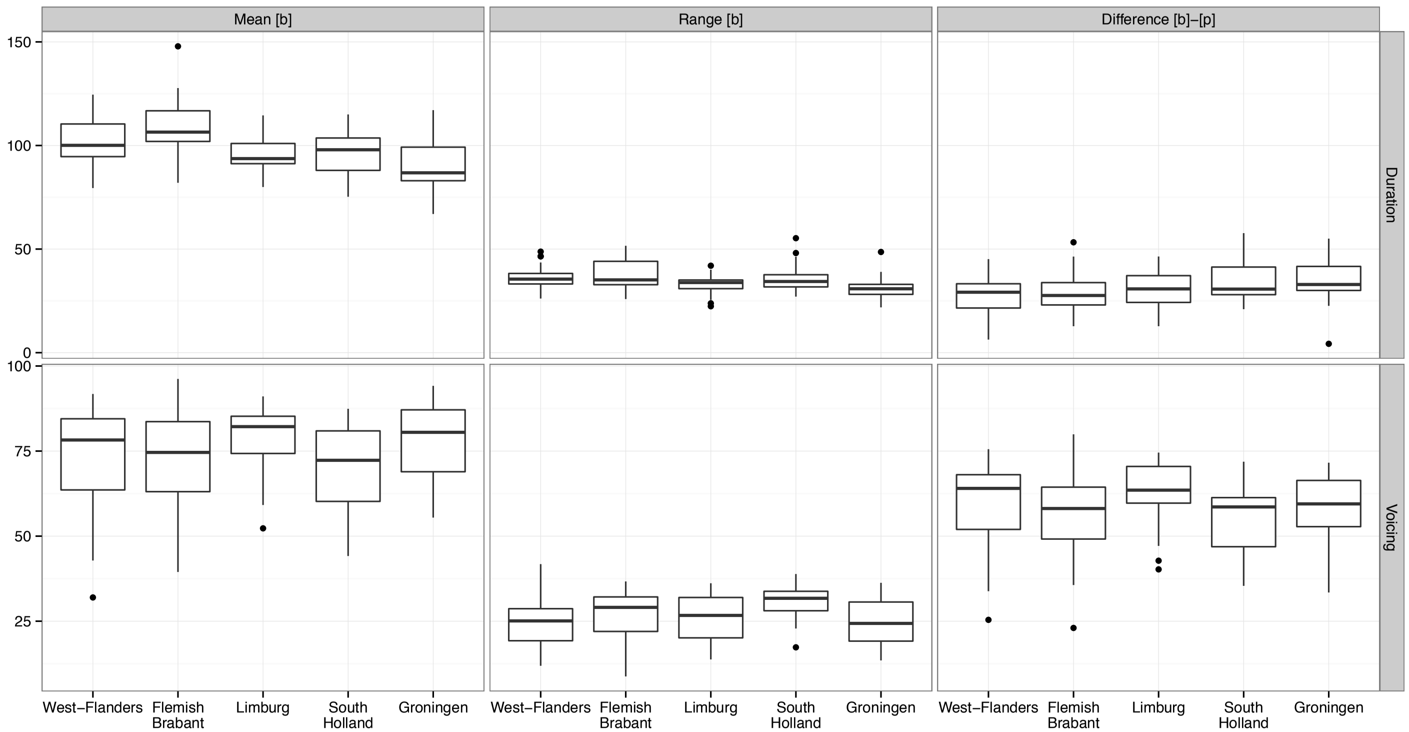
As far as duration is concerned, subtle regional differences appeared in the mean [b] duration. Realizations from West-Flanders and Flemish-Brabant appeared slightly longer than realizations from the Dutch regions. Both the range of [b] durations and difference in duration between [b] and [p] realizations appeared to be relatively constant across regions (around 40 ms and around 35 ms respectively).

Finally, voicing measurements also were rather consistent across regions. [b] realizations were produced with – on average – a lot of voicing (around 75 %). The range of voicing in [b] in all regions was measured around 25 %. Both the difference in voicing and the border between [b] and [p] realizations were relatively constant across regions (around 60 % and around 45 % respectively).

[4]Mixed model regressions were not required for this analysis, since there was only one observation per participant in the aggregated data.

[5]The other participant-related factors *sex*, *dialect proficiency*, *regional accent*, *mobility* and *contact with the other country within the Dutch language area* did not turned out significant and were therefore excluded in the model.

Figure 4.5: Boxplot of stop duration (in ms) (in the upper panel) and voicing (in %) (in the lower panel) split up by region. These measures are 1) the mean and 2) the range of voiced bilabial stops, 3) the difference between voiced and voiceless bilabial stops (N=100).



An analysis of variance was run on the voiced bilabial stops, but none of the external factors turned out significant. Thus – in contrast to fricatives – no difference between regions could be found as all regions showed a similar amount of voicing in voiced stops (West-Flanders: 73 %, Flemish-Brabant: 72 %, Limburg: 78 %, South-Holland: 70 %, Groningen: 78 %). In order to find differences in the progression of stop devoicing, we need to look at the individual level.

4.3.4 Cue weighting at the individual level

The fact that phonetic dimensions do not need to contribute equally to category identity has been referred to as *cue weighting* (Holt & Lotto, 2006). Phonetic dimensions appear to be weighted in the sense that some are more important than others in order to produce contrasts. Since we have measured several phonetic cues in the production of fricatives and stops, it is interesting to investigate which cue weighting strategies are used by individuals.

The cue weighting strategies in these data are explored with the help of Figures 4.6 and 4.7. In these graphs, voicing is plotted along the x-axis and duration along the y-axis. The mean voiced realization is represented – for each participant – by a filled circle and the mean voiceless realization by a cross symbol. A solid line was drawn from the voiced to the voiceless mean measure. The length and the inclination of the lines give insights in the cue weighting patterns of each individual. The longer the line, the more contrast made between the two categories. A more horizontal line means that the participant mostly used voicing to make the contrast. A more vertical line means that the participant mostly used the duration cue. A diagonal line shows that the participant used both voicing and duration in order to make the voiced-voiceless contrast.

Fricatives Figure 4.6 represents the cue weighting patterns for fricatives within individuals grouped by region. First, it appeared that in the five regions voiceless realizations were nicely grouped, whereas averaged voiced realizations showed a lot of variation. In West-Flanders, Flemish-Brabant and Limburg, all participants showed diagonal lines, meaning that all used both voicing and duration in the contrast. The lines of Limburgian participants tended to be shorter, indicating the contrast is fading. In South-Holland, around half of the participants presented relatively short diagonal lines. The other half showed very short and sometimes almost vertical lines, which illustrates the loss of the contrast or – when the contrast is still maintain to some extent – the use of duration solely to make the contrast. Finally, in Groningen most participants showed no line at all, or a very short vertical line. Only two participants stood out by the use of both dimensions to a much larger extent than their peers.

Figure 4.6: Cue weighting in fricatives production split up by individual and grouped by region with voicing (in %) on the x-axis and duration (in ms) on the y-axis. Mean target voiced fricatives are represented by filled circles and target voiceless fricatives by crosses (N=100).

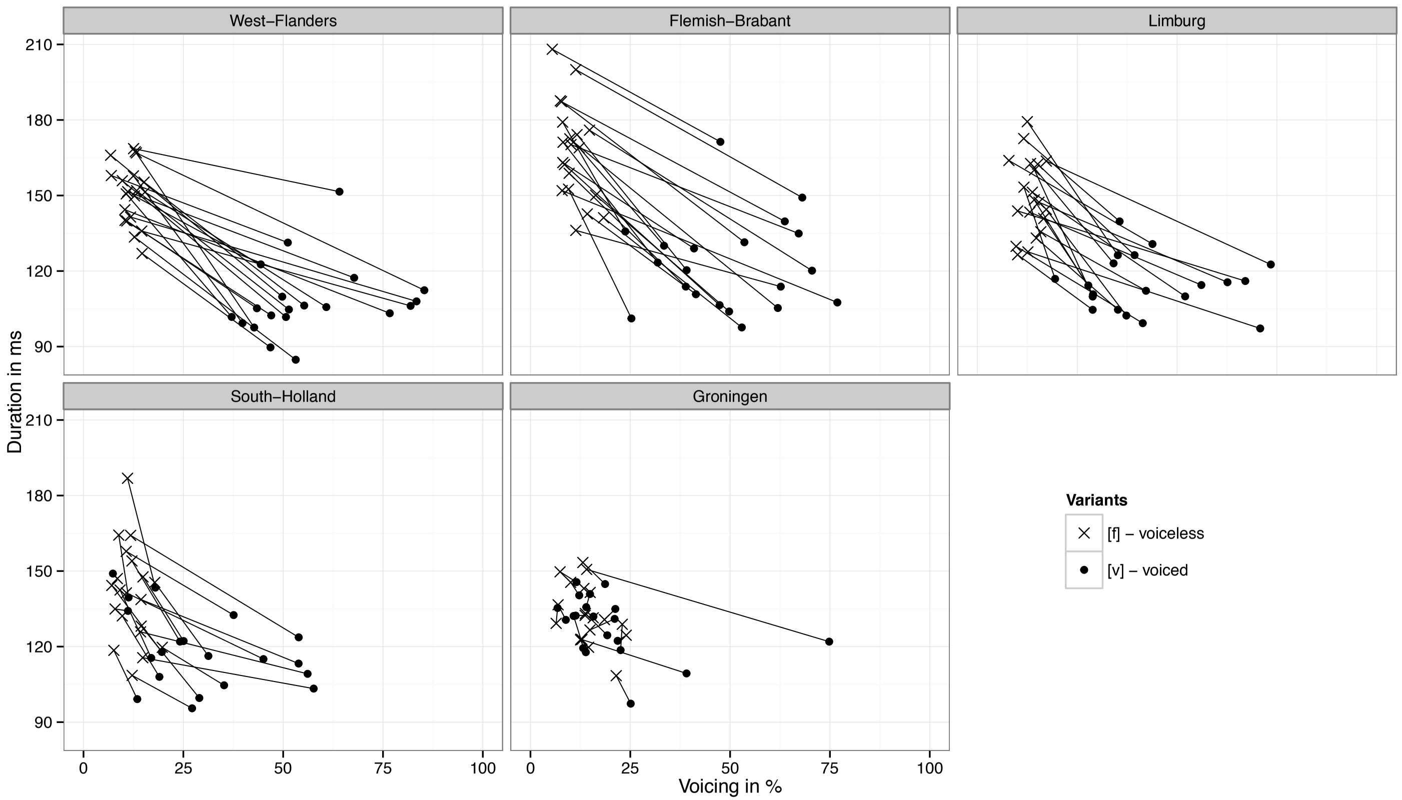
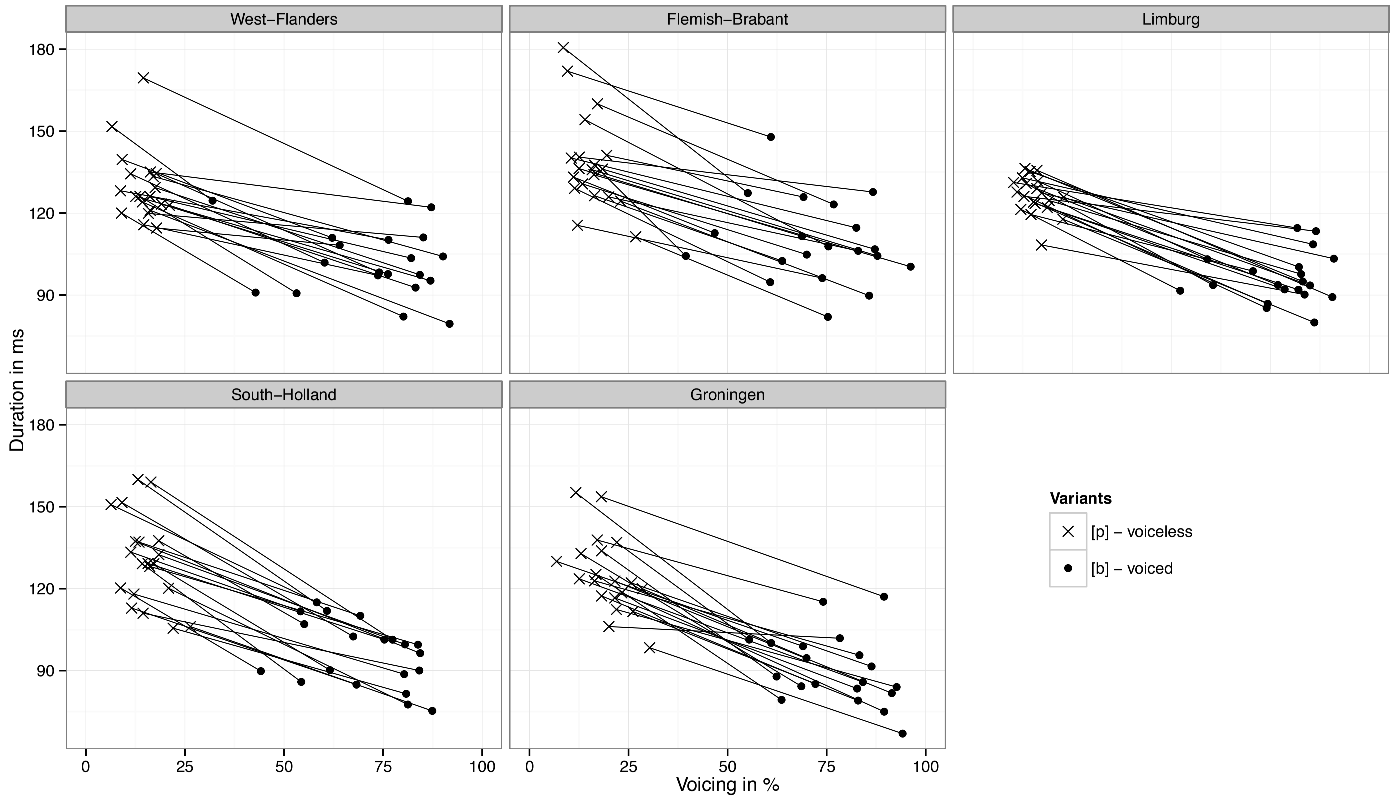


Figure 4.7: Cue weighting in stops production split up by individual and grouped by region with voicing (in %) on the x-axis and duration (in ms) on the y-axis. Mean voiced are represented by filled circles and voiceless fricatives by crosses (N=100).



Stops Figure 4.7 represents the cue weighting patterns for stops within individuals grouped by region. As with fricatives, it appeared in the five regions that voiced realizations showed more variation than their voiceless counterparts. Participants of Flemish-Brabant showed a large range. Participants from all regions showed diagonal lines, indicating cue weighting between voicing and duration. There were a limited number of participants who voiced realizations with less than 50 % voicing (two in West-Flanders, two in Flemish-Brabant and one in South-Holland). These five speakers were the most advanced in the stop devoicing process. Their lines were shorter and more vertical than the ones of the other participants. In all regions, there seemed to be a trend towards more verticality, showing a greater reliance on the duration dimension.

4.3.5 Putting both variables together

Up until this point, we have looked into both regional and individual differences in the production of fricatives and stops separately. However, the goal of choosing these two variables was to be able to test sound changes that have different degrees of completion (an advanced sound change vs. an incipient sound change). Therefore, it is interesting to put the production results of both fricatives and stops together.

First, we found a significant correlation between the mean voicing of the voiced stop and the voiced fricatives (r=.30, p=.002). The more participants devoiced their fricatives, the more they devoiced stops.

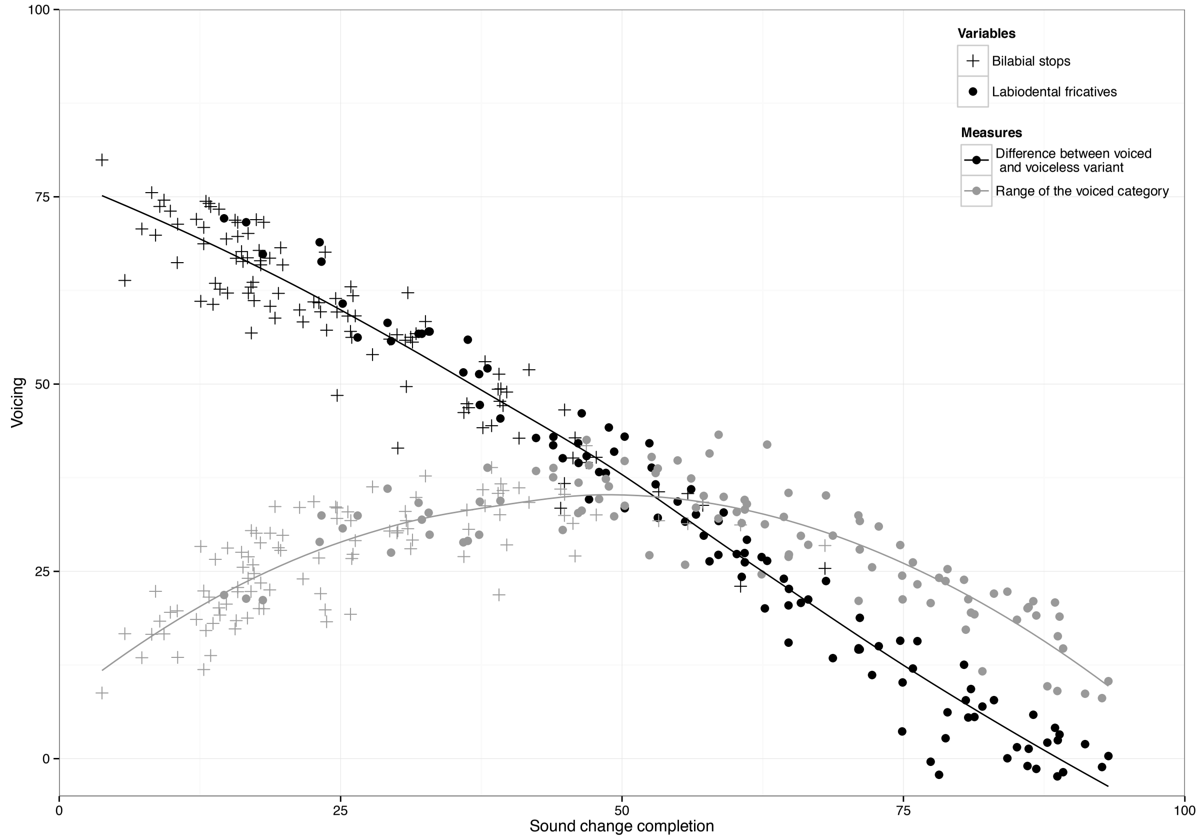
Secondly, when putting fricatives and stops together, we found higher-order relationships between the mean voicing of the voiced category, its range and the difference between the voiced and voiceless variants, measures that were presented in Section 4.3.2. The relationships between these measures are presented in Figure 4.8.

In this figure, the x-axis represents the completion of the sound changes. The *sound change completion* (SC completion) was computed in the following way:

$$SCcompletion = 100 - meanvoicing_{voicedvariant} \tag{4.1}$$

For instance, a participant who produces an mean [v] with 40 % voicing would get completion score of 60. The SC completion score ranged from 0 (no change yet) to 100 (fully completed change), thus the more to the right, the further in the sound change. On the y-axis, the measure of voicing in percentage is plotted. Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot. Each symbol stands for a participant. There are − colored in black − 200 measures of the difference between voiced and voiceless variants (100 participants producing both fricative realizations and stops realizations). Colored in grey, there are 200 observations of the range voiced category (n=100*2). Fitted lines for differences and ranges are presented in the same colors.

Figure 4.8: Scatterplot of the range of the voiced category and the difference between the voiced and voiceless variants (both indicated in percentage of voicing) in function of the completion of the sound change, split up by variable (N=100 participants, n=200 observations).



It turned out that both the range of the voiced category and the difference between the voiced and voiceless variants showed strong relationships with the sound change completion score (i.e., with the mean voicing of the voiced category). On the one hand, the difference between the voiced and voiceless variants appeared to decrease in a linear manner as the change is proceeding. The difference scores approached 80 % at the beginning of the sound change and decreased to 0 % when the change is completed. On the other hand, the range of the voiced category increased in the first phase of the sound change (until around 35 %), but crucially decreased in the second phase to form a curve.

These relationships were tested statistically with the help of linear regressions. The difference between the voiced and voiceless variants was significantly predicted by the completion score (LINEAR REGRESSION t=75.27, p<.001), which could account for 96.62 % of the variance in the difference score. This very strong relationship can partly be explained by the fact that the completion score (measured as the mean voiced variant) and the difference between the voiced and the voiceless variants were by definition intrinsically linked, as they both used the voiced variant in their calculation. Nevertheless, the strength of the relationship indicated that it is especially the voiced category that is moving during sound change, whereas the voiceless category basically remains the same.

The range of the voiced category was significantly predicted by the completion score tested with an additional quadratic component (i.e., completion score^2) (LINEAR REGRESSION: completion score^2: t=-22.713, p<.001 and completion score: t=-1.272, p<.001) and could account for 72.37 % of the variance. Since the completion score and the range of the voiced category were computed independently, this strong relationship is very meaningful.

In conclusion, the range of the voiced category and the difference between the voiced and voiceless variants turned out to be highly predictable from the degree of sound change completion.

4.4 Conclusion

All participants took part in a sociolinguistic interview consisting of five tasks aimed at triggering fricatives and stops realizations in five speech styles (word reading, carrier sentence reading, sentence reading, semi-spontaneous speech and spontaneous speech). Patterns of variation in fricatives and stops production were investigated by means of the phonetic analysis of these realizations. A total of 10261 fricatives and 10804 stops realizations were collected and measured along a set of phonetic dimensions (DURATION and VOICING). VOT turned out to be an unreliable measure in our data. For both fricatives and stops, it turned out that voicing was the phonetic dimension which grasped the devoicing phenomena the best.

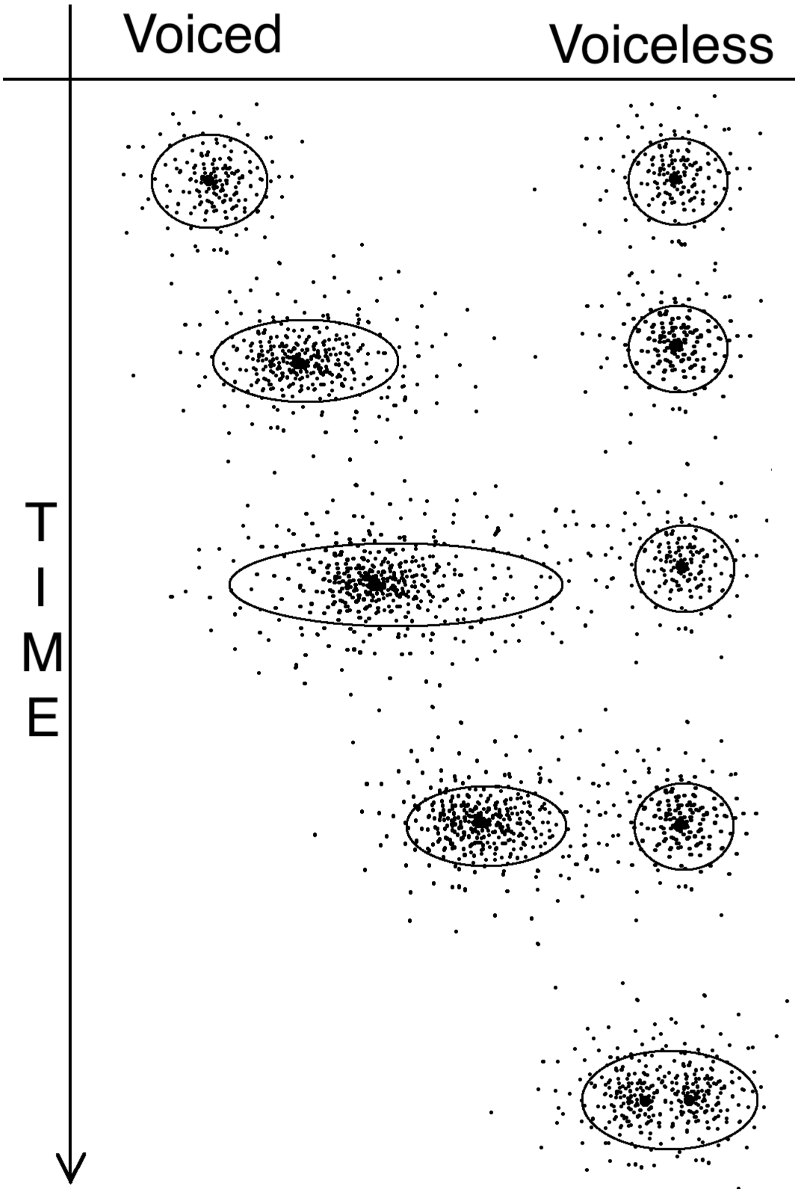
First, the inspection of within-speakers differences (although this was not directly the aim of this experiment) revealed that phonetic realizations of fricatives and stops differed across speech styles. The word reading task elicited realizations slightly diverging from the realizations in the other tasks with smaller amount of voicing. Despite these patterns, all observations were aggregated resulting in one observation per individual and per variant. On the aggregated and non-aggregated data, several measures were calculated: the mean voiced realization, the range of the voiced category, the difference between the voiced and the voiceless variant.

As predicted from previous production studies, differences were found between regions (at the *group level*) in the amount of devoicing in fricatives. In all regions, starting with the Netherlandic regions, the contrast in initial labiodental fricatives is fading. This sound change appeared to be even further advanced than reported ten years ago by Kissine et al. (2003, 2005). Stops however did not directly show regional differences, as expected from previous studies. Patterns of stop devoicing only appeared when looking at the *individual level*. In several regions, a minority of individuals showed clear patterns of devoicing, mostly along the voicing dimension.

No difference in sex could be found in the fricatives nor stops. These results contrasted to the findings of Kissine et al. (2003) for fricatives and of Smits and van Alphen (2004) for stops who found small significant differences between men and women.

Finally, the juxtaposition of the production patterns of both variables gave us precise insights into the relationships between differences between categories and range of these categories on the one hand, and the development of sound change in the other hand. These insights allow us to redefine the mechanism of category merging in production. This mechanism is schematized in Figure 4.9. Before the change, the situation is stable with two separated categories of comparable range. At the change onset, the merging category (in this case the voiced category) increases its range, while the difference between categories decreases in a linear manner. Further in the change however, the range of the merging category starts to decrease, whereas the difference between categories keeps shrinking. At the change completion, both categories are merged (no or very small difference between categories) and have a comparably small range.

Figure 4.9: Schematic representation of devoicing in production categories. The arrow indicates time.



CHAPTER 5

Speech perception

In Chapter 2, it was pointed out that speech perception and the way in which it is affected by social information is best explained within the framework of exemplar theory. This chapter aims at investigating the central question of the role of speech perception in sound change within this framework. The chapter begins with an introduction on the perception of labiodental fricatives and bilabial stops (in Section 5.1). The research question is tackled by gathering two types of perceptual data: first a forced-choice identification task providing quantitative data on perception (reported in Section 5.2), and second a similarity judgment task providing more qualitative perceptual data (reported in Section 5.3). Both tasks are compared in Section 5.4 of this chapter. In these experiments, we examine whether differences in the perception of labiodental fricatives and bilabial stops can be found between regions within the Dutch language area (at the *group level*) and between individuals within these regions (at the *individual level*).

5.1 Introduction and hypotheses

In this section, we first look at previous experimental work done on phonetic cues in the perception of labiodental fricatives and bilabial stops (Section 5.1.1) and on regional differences in speech perception (Section 5.1.2).

5.1.1 Phonetic cues in the perception of labiodental fricatives and bilabial stops

After having obtained more insights in the production patterns of the two sound changes (see Chapter 4), we examine which phonetic cues are known to play a role in the perception of labiodental fricatives and bilabial stops. From experimental phonetic research (e.g., Slis, 1985; Slis & Cohen, 1969; van den Berg, 1988), it has been shown that a large range of perceptual cues may be used in the categorization of consonants (e.g., voicing of the consonant, duration of the consonant, duration of the preceding vowel, intensity of frication noise, pitch pattern of the adjoining vowels).

For stops, early work in speech perception has verified the role of VOT in the perception of utterance-initial stops for speakers of English and other languages (Liberman et al., 1957; Lisker & Abramson, 1970). In these studies, listeners have been found to categorize stops as voiced or voiceless depending on the VOT pattern present in their language. Moreover, other durational measures were shown to play a role in stop perception (e.g., duration of the prevoicing, duration of the burst) (Ali, Van der Spiegel, & Mueller, 2001b; Cho & Ladefoged, 1999; Slis & Cohen, 1969; Smits & van Alphen, 2004). Based on these studies, two main phonetic dimensions were chosen to investigate the perception of bilabial stops: VOT and total DURATION of the plosive.

The perception of fricatives is less well-studied than the perception of stops (Ali, Van der Spiegel, & Mueller, 2001a). The major cue in the perception of fricatives seems to be the presence or absence of vocal cord vibration in the fricative (Slis & Cohen, 1969; van den Berg, 1988). Ali et al. (2001a) showed that the duration of the unvoiced portion plays a significant role. The role of frication noise intensity was shown by Debrock (1977) to be rather inconsistent. Hence, two main phonetic dimensions were chosen to investigate the perception of labiodental fricatives: the VOICING and the total DURATION of the fricative.

5.1.2 Regional differences in speech perception

The majority of previous work on regional differences in perception focused on the perception of social meaning rather than linguistic form. Work within the field of *perceptual dialectology* has shown that speakers are highly aware of regional differences and they associate social meaning to linguistic variation (e.g., Fridland, Bartlett, & Kreuz, 2005; Hay & Drager, 2010; Plichta & Preston, 2005; Preston, 1989).

Beyond studies on social perceptions, there has also been recent work investigating the perception of linguistic variables. Most of these studies aimed to show that listeners are able to attune their perception to the speaker they (think they) are listening to and that they utilize their knowledge of speaker differences in order to help interpret the variation (e.g., Drager, 2005; Foulkes & Docherty, 2006; Hay & Drager, 2010; Hay, Nolan, & Drager, 2006; Hay, Warren, & Drager, 2006; Johnson, Strand, & D'Imperio, 1999; Niedzielski, 1999; Rubin & Smith, 1990; Strand, 1999; Strand & Johnson, 1996). These studies proved that speech perception is highly flexible and provided empirical support for exemplar models in which previously encountered exemplars are socially indexed (see Section 2.1). They have documented the effect of social and linguistic expectation on speech perception.

The studies just mentioned give insight into the extent listeners' perception is attuned to the speaker they are listening. They showed that our perception is able to take many factors into account when perceiving speech (the age of the speaker, his regional background, gender, ethnicity, etc). However, much less is known about differences related to the listeners' own perception, regardless of the speaker they are listening to. Depending on their regional background, listeners are thought to differ in the way they perceive the same speaker. The question raised is thus whether individuals within a community show differences in speech perception depending on their exposure to different regional norms.

Only a few studies, such as Janson (1983), Evans and Iverson (2004), Sumner and Samuel (2009), Kendall and Fridland (2012) and Fridland and Kendall (2012) have investigated dialectal and regional differences in the perception of changing sounds. Janson (1983) examined the perception of Swedish vowels by listeners of two different Swedish dialects. One dialect had four vowels and the other one three (as the result of a merger). They found that – while most of the listeners from the unmerged dialect could discriminate well between the vowels – other listeners from the same dialect could not. Evans and Iverson (2004) asked subjects with Northern and Southern English backgrounds to select best exemplars of vowels. Their results showed that different formant frequencies were selected as best exemplars depending on participants' background. In Sumner and Samuel (2009), three groups of subjects were tested on their perception of non-rhotic forms. Participants differed in whether they produced non-rhotic variants themselves, regularly heard such variants but did not exhibit them productively, or never used nor regularly heard non-rhotics. These results suggested that both the individual production and the local community norm affect individuals' perception. Kendall and Fridland (2012) and Fridland and Kendall (2012) explored how speakers from three different regions in the U.S. performed on a vowel identification task (contrast between /e/ and /ɛ/). The results showed that regional affiliation played a significant role in the perception of the vowel contrast.

Additionally, some studies have found that individuals showed different perceptual patterns based on another factor: their *age* (e.g., Janson, 1979; Malderez, 1995). As explained in Section 3.1.3, the main focus in this study is the investigation of regional differences. However, it is worth noting that these studies also found age-related differences in speech perception.

In conclusion, only a few studies have begun to look at the relationship between regional variation and speech perception. These studies have mostly looked at vowel contrasts, so that less is known about regional differences in the perception of consonants. The current study of regional differences in the perception of changing consonants aims at filling this gap and answering the question whether participants from the five different regions of this study have different perceptual tendencies with regards to labiodental fricatives and bilabial stops.

5.2 The forced-choice identification task

Forced-choice identification experiments are often used in speech perception research (Strange, 1995; Thomas, 2002). In the present study, a forced-choice identification task was conducted to gather quantitative insights in the categorical perception of fricatives and stops. In Section 5.2.1, the method used in the forced-choice identification experiment is described. The results of the experiment are reported in Section 5.2.2 and discussed in Section 5.2.3.

5.2.1 Method

In this section, the stimuli used in the forced-choice identification task are described together with the experimental procedure. Subsequently, the analysis of binary responses is explained and hypotheses are formulated.

Stimuli

For both fricative and stop variables, a bi-dimensional continuum was created by resynthesing natural speech tokens.

Fricatives In order to test the perception of fricatives, a bi-dimensional speech continuum was generated by manipulating the acoustic dimensions VOICING and DURATION, as they were shown to be the two main acoustic cues along which labiodental fricatives are categorized (see Section 5.1.1).

Fricative stimuli were presented in a CV syllable with the /i/ vowel. As explained in Section 4.2, /i/ is a point with an extreme position (high and forward) in the vowel space and can therefore be identified with a high rate of accuracy (e.g., Fox, 1982). As this vowel shows least regional variation in Standard Dutch (van der Harst, 2011, p.159), it was chosen to avoid any bias that could be caused by regional differences in the perception of the vowel.

Naturally produced syllables /vi/ and /fi/ were used to create the bi-dimensional continuum. Speech material of a male native speaker of Dutch (a trained phonetician of 25 years old from South-Holland[1]) was digitally recorded with a sample frequency of 44.1 kHz in a sound-attenuated cabin. The speaker pronounced a series of one-syllable Dutch non-words, among which the target combinations /vi/ and /fi/. A fully-voiced [v] token and a fully-voiceless [f] were selected for the creation of the continuum. The duration of the original [v] was 184 ms and the original [f] 176 ms. The vowel [i] had a duration of 251 ms in the

[1]Following Clopper and Pisoni (2004) we chose a young male speaker with no noticeable regional features.

/vi/ realization and 239 ms in the /fi/ realization. The mean F0 value of [v] was 131 Hz.

The fricatives of the source recordings were extracted from their original context, lengthened to a duration of 196ms, and used as the extremes of the continuum along the voicing dimension. The nine steps along the voicing dimension were generated by spectral linear interpolation, using the PSOLA (Pitch-Synchronous-Linear-OverLap-and-Add) algorithm of `Praat` speech-analysis software package (Boersma & Weenink, 2014) (based on the script of Mitterer (2009)). Besides the two extremes of the continuum with respectively 0 % and 100 % voicing, the interpolation provided seven intermediate realizations characterized by approximately 12.5 %, 25 %, 37.5 %, 50 %, 62.5 %, 75 % and 87.5 % voicing. In this way, the first nine sound realizations of the continuum with a constant duration of 196 ms and varying degrees of voicing (ranging from 0 % to 100 % of voicing) were obtained. The remaining realizations were created by manipulating the duration by means of the 'lengthen' function in `Praat`.

The final stimuli were obtained by concatenating the /v/-/f/ realizations with the [i] produced in the original /vi/ context shortened to a constant duration of 110 ms. Importantly, there is evidence that the pitch contour of the vowels following a fricative depends on the voice quality of that fricative. Vowels following voiced fricatives show a F0 contour that slightly increases, whereas vowels following voiceless fricatives show a decreasing F0 contour (Kissine et al., 2005; Slis & Cohen, 1969). Moreover, the F0 value in the onset of a vowel following a voiceless fricative is higher than in the onset of a vowel following a voiced fricative (Slis & Cohen, 1969). In order to avoid any perceptual bias that would be due to these facts, the pitch contour of the original vowel was manipulated through the PSOLA pitch manipulation and LPC resynthesis functions of `Praat`. In the vowel transition, the pitch contour was flattened to values ranging between 130 and 135 Hz. From 60 ms after the beginning of the vowel onwards, the pitch contour was gradually reduced to 120 Hz in order to get a falling contour. Finally, all stimuli were scaled to an intensity of 70 dB.

This resulted in a two-dimensional grid consisting of nine steps for each dimension, thus a /vi/−/fi/ continuum with 81 realizations which is presented in Figure 5.1 (in the upper part). The voicing dimension represented the degree of voicing present in the fricatives. This dimension ranged from 100 %, a fully voiced fricative, to 0 %, a fully voiceless fricative (9 steps). Along the duration dimension, stimuli ranged from 60 to 196 ms, with equal intervals of 17 ms (9 steps).

Stops For the perception tasks of stops, a bi-dimensional speech continuum was generated by manipulating the acoustic dimensions VOT and DURATION, as they were shown to be the two main acoustic cues along which bilabial stops are categorized (see Section 5.1.1).

Figure 5.1: Schematic representation of the two bi-dimensional fricative continuum, varying voicing horizontally and duration vertically.

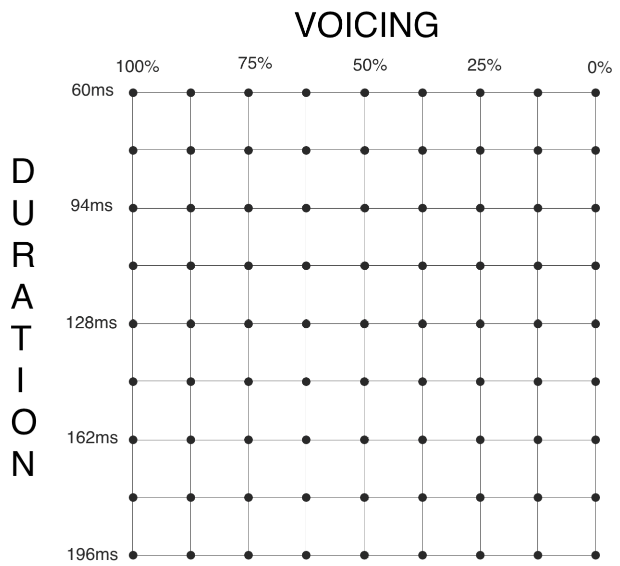
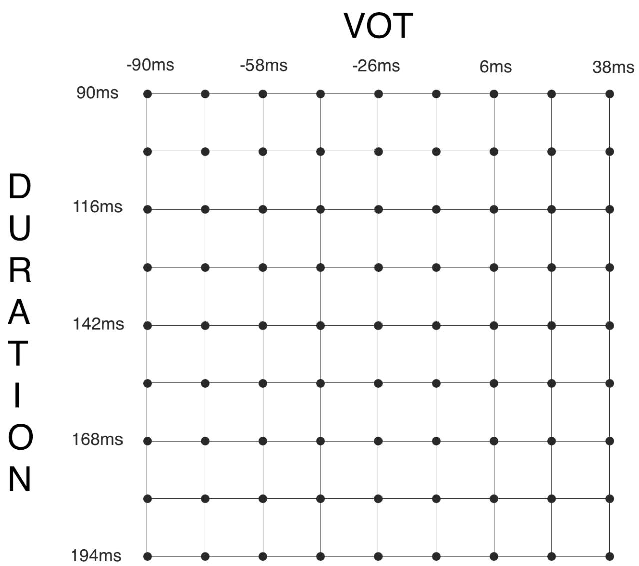


Figure 5.2: Schematic representation of the two bi-dimensional stop continuum, varying VOT horizontally and duration vertically.



Stops stimuli were also presented in a CV syllable with the /i/ vowel for the same reason as explained for fricative stimuli. Naturally-produced syllables /bi/ and /pi/ were used to create the bi-dimensional continuum. Speech material of a second male native speaker of Dutch (34 years old from South-Holland) was digitally recorded with a sample frequency of 44.1 kHz in a sound-attenuated cabin.

The speaker pronounced a series of one-syllable Dutch non-words, among which the target combinations /bi/ and /pi/. A fully voiced [b] token and a fully voiceless [p] were selected for the creation of the continuum. The duration of the selected [b] was 105 ms and the vowel [i] had a duration of 104 ms in the /bi/ realization. The /p/ realization could not be measured since the syllables were produced in isolation. Hence there was no cue to indicate the /p/ onset. The mean F0 value of [b] was 147 Hz.

The voiced and voiceless stops were extracted from the original context. To construct the VOT continuum, parts of the prevoicing were gradually removed from the reference voiced stop in order to shorten the negative VOT. For the steps with positive VOT, parts of the burst from the voiceless stops were added resulting in extended positive VOT.

Besides the two extremes of the continuum with VOT of respectively -90 ms and 38 ms, the manipulation provided seven intermediate realizations characterized by a VOT of -74 ms, -58 ms, -42 ms, -26 ms, -10 ms, 6 ms and 22 ms. In this manner, the first nine realizations of the continuum with a constant duration of 90 ms and varying VOT's were obtained. The remaining realizations were obtained by adding duration before the burst in `Praat`. For prevoiced realizations, this resulted in a prevoicing that is not totally voiced when the duration gets longer, which is very comparable to the type of devoicing realizations we found in production (see Section 4.2.2). For unvoiced realizations, this resulted in a longer unvoiced part before the burst.

The final stimuli were obtained by concatenating the /b/-/p/ realizations with the [i] produced in the original /bi/ context lengthened to a constant duration of 150ms. Like fricatives, there is evidence that the pitch contour of the vowels following a stop depends on the voice quality of that stop (Rietveld & Van Heuven, 2009; Slis & Cohen, 1969). In order to avoid any perceptual bias that would be due to this fact, the pitch contour of the original vowel was manipulated through the PSOLA pitch manipulation and LPC resynthesis functions of `Praat`. In the vowel transition, the pitch contour was flattened to values ranging between 145 and 150 Hz. From 60 ms after the beginning of the vowel onwards, the pitch contour was gradually reduced to 130 Hz in order to get a falling contour. Finally, all stimuli were scaled to an intensity of 70dB.

This resulted in a two-dimensional grid consisting of nine steps for each dimension, thus a /bi/−/pi/ continuum with 81 realizations which is presented in Figure 5.2 (in the lower part). The VOT dimension ranged from -90 ms to 38 ms with nine equal intervals of 16 ms (9 steps). The duration dimension ranged from 90 ms to 194 ms, with nine equal intervals of 13 ms (9 steps).

Procedure

Participants were seated in a sound-attenuated booth and listened to the stimuli via a headphone Beyerdynamic DT 250 (more details see Section 3.5). They were asked to categorize the realizations as being either /v/ or /f/ (or /b/ or /p/) by pressing the red or blue button of a button box labelled with the corresponding sounds. The order of presentation of the consonants on the button box (i.e., voiced-voiceless or voiceless-voiced) was balanced between participants. The task was auto-paced. Reaction times (RT) were recorded from the beginning of the stimuli. Participants had a time window of 800 ms after the end of the stimulus to give their response. Responses given after this time window were excluded in order to avoid responses resulting from second guesses. In these cases, participants automatically went over to a new trial. Twelve stimuli were presented in the practice session to familiarize the participant with the task. After the practice session, the test phase started. The set of 81 stimuli from the continuum was repeated five times. The 405 items were presented in a random order.

Analyzing binary responses

In order to obtain parameters that are more interpretable within the statistical analyses, the nine continuum steps along each dimension were centralized, so that the central intermediate step 5 would be equal to 0. The continua ranged from -4 to 4 instead of from step 1 to step 9. Along the voicing dimension, the leftmost part (negative values) refers to the most voiced realizations and the rightmost part (positive values) refers to the most voiceless realizations. Along the VOT dimension, the leftmost part (negative values) refers to the negative VOT realizations and the rightmost part (positive values) refers to the positive VOT realizations. Along the duration dimensions, the negative values refer to the shortest realizations and the positive values refer to the longest realizations.

The binary responses obtained in this forced-choice identification task were analyzed with Generalized Mixed Effects Logistic regression (GLMM, Jaeger (2008), Quené and van den Bergh (2008)) as implemented in the `lme4` library (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2014). The dependent variable was the percentage of identification as voiced (/v/ or /b/). In logistic regression, the probability of x (P(x)) − in this case the probability of a /v/ or /b/ response − is predicted by the following equation:

$$P(x) = \frac{e(\beta_0 + \beta_1 x)}{1 + e(\beta_0 + \beta_1 x)} \tag{5.1}$$

β_0 is the estimate of the intercept and β_1 is the estimate of the slope of the logistic regression line. The higher the absolute value of β_1, the steeper the *slope of the regression line*. The median (i.e. the point where the probability P(x) is equal to .5) can be calculated on basis of these two estimates with the following formula:

$$Median = \frac{-\beta_0}{\beta_1} \tag{5.2}$$

As commonly used in the analysis of categorical perception (e.g., Kendall & Fridland, 2012) logistic regressions provide psychometric curves that are defined by 1) their *slopes*, which indicate how categorical the judgment is (the steeper the curve, the more categorical the judgment), and 2) their medians, which represent the *cut-off point between the two categories* (also called 50 % crossover point).

Formulating hypotheses

In view of what it is known about the target sound changes (see Section 3.2), we formulate the following hypotheses.

At the *group level*, we expected regional differences in the perception of fricatives both regarding the slope and the cut-off point of the psychometric curves. The less devoicing of /v/ in production in a region, the more categorical the stimuli will be perceived by participants from this region. The following order for the steepness of the slope for both the voicing and duration dimensions is expected: West-Flanders > Flemish-Brabant > Limburg > South-Holland > Groningen. Moreover, the less devoicing of /v/ in production, the closer the cut-off point between categories will be towards /v/ with the same expected order as just mentioned.

Even if the devoicing of stops as a sound change is still in an incipient phase and did not really show regional differences in production (see Chapter 5), we expected some preliminary differences in the perception results of stops. We hypothesize − like for fricatives − regional differences in slope and cut-off point in which both Flemish regions (West-Flanders and Flemish-Brabant) will stand out and show less categorical perceptual patterns.

At the *individual level*, we expected − for both fricative and stop perception − differences in the use of phonetic cues. It is hypothesized that individuals differ in the extent to which they rely on different cues when perceiving the contrasts.

5.2.2 Results

Trials with response times shorter than the duration of the consonants were excluded from the analysis, because they occurred before the participants heard the entire first segment. Hearing the entire first segment was crucial here, since duration was a manipulated cue in both fricatives and stops. Moreover, due to the limited time window, responses given too late (800 ms after the end of the stimuli) were not recorded and resulted in missing values. Excluded data (too short latency) and missing data (too long latency) are described in Table 5.1. In the fricative experiment, 4.76 % of the data were excluded and missing data. 37678 observations remained for the analysis. In the stop experiment, excluded and missing data represented 3.19 % of the data, so 38803 observations remained for further analysis[2].

Table 5.1 shows that the number of excluded data was very low for both fricatives and stops (117 vs. 56 in total). Missing data, however, represented a sizeable proportion of the data. Furthermore, there were significant differences in the amount of missing fricative data across regions (ANOVA: df=4;95, F=3.582, p=.009): the number of missing fricative data was the highest in Groningen and the lowest in West-Flanders (12.27 % vs. 3.33 %). The other three regions took an intermediate position (around 6 % of missing data). This first difference between regions can be interpreted as perceptual insecurity: the more devoicing in a region, the more perceptual insecurity and difficulty to decide between the voiced and voiceless category, which leads to too long latencies and thus missing data.

Table 5.1: Number of missing and excluded data in the identification task of fricatives (in the upper part) and stops (in the lower part).

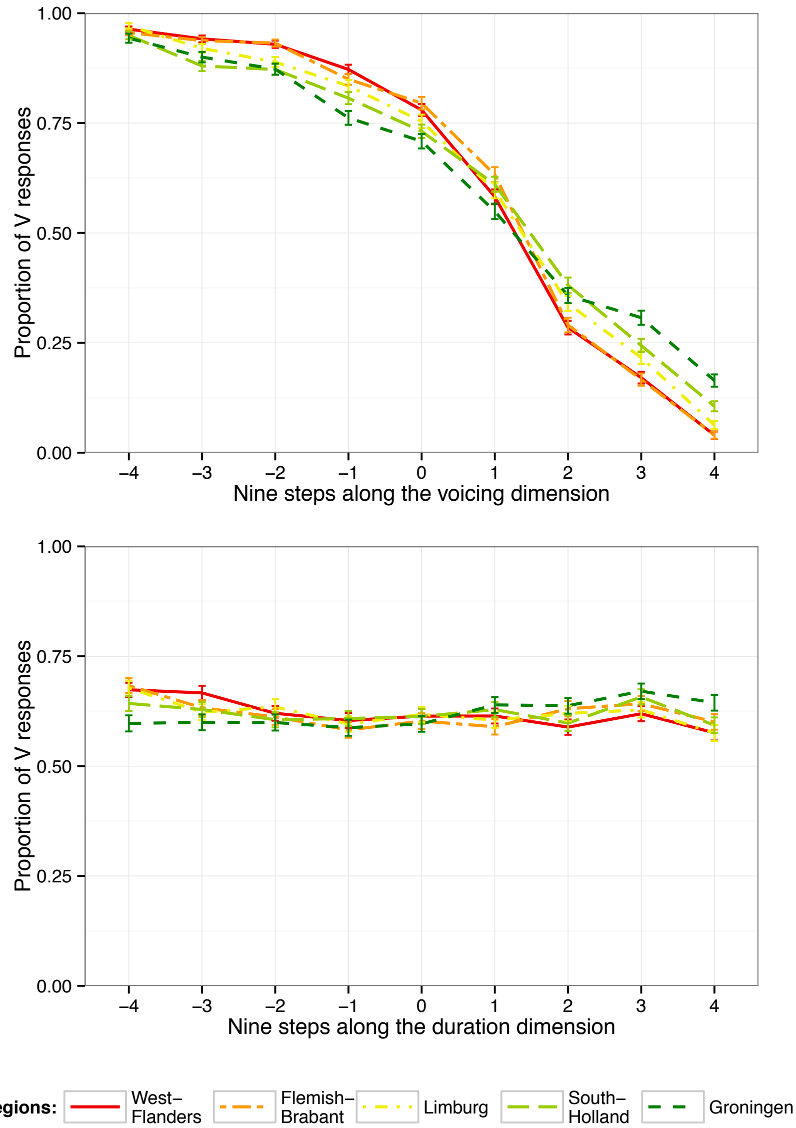
Fricatives	Excluded data		Missing data		Total remaining
	n	%	n	%	n
West-Flanders	13	0.22	270	3.33	7817
Flemish-Brabant	8	0.10	481	5.94	7611
Limburg	37	0.46	435	5.37	7628
South-Holland	22	0.27	530	6.54	7548
Groningen	32	0.40	994	12.27	7074
Total	112		2710		37678
Stops	Excluded data		Missing data		Total remaining
	n	%	n	%	n
West-Flanders	17	0.21	178	2.20	7500
Flemish-Brabant	6	0.07	289	3.57	7805
Limburg	6	0.07	243	3.00	7851
South-Holland	4	0.05	199	2.46	7897
Groningen	23	0.28	327	4.04	7750
Total	56		1236		38803

Effects between regions

Fricatives The results of the identification task for fricatives are shown in Figure 5.3. In the upper graph, the results along the voicing continuum are presented. The centralized nine-steps continuum are presented on the x-axis and the proportion of /v/ responses on the y-axis. The psychometric curves were drawn separately for each region.

[2] Due to a technical problem, the data of the stop identification task were lost for one participant (#WF12).

Figure 5.3: Results of the identification task for fricatives split up by region. The centralized nine-steps continuum along the voicing dimension (in the upper panel) and along the duration dimension (in the lower panel) are presented on the x-axis and the proportion of /v/ responses on the y-axis (n=37678). Error bars represent ±1 standard error.



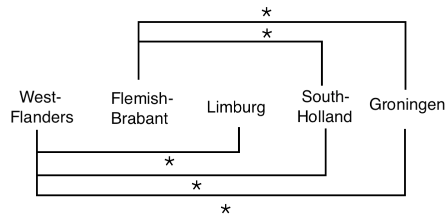
We observed that the extremes were almost unanimously judged as instances of either /f/ or /v/. Listeners of all regions generally heard the extremes of the continuum categorically or near-categorically. For stimuli in the middle of the continuum, the decision was ambiguous and the cut-off point between the categories seemed to be around step +1.

There were differences between regions in the categorization along the voicing dimensions. The slope of the psychometric curves was the steepest for West-Flanders and Flemish-Brabant, a bit less steep for Limburg, and even less steep for South-Holland and Groningen. For the duration continuum in the lower graph, it looks quite different. We observed that for all regions the curves were almost flat and that there are no cut-off points at 50 %.

These data were fitted with a generalized mixed effects logistic regression (n=37678). The random effects in this model consisted of the factor *participants* (N=100), testing for individual differences between participants; the factor *voicing by participant*, testing for individual differences in voicing slope; and the *factor duration by participant*, testing for individual differences in duration slope. The fixed part of the model consisted of the factor *region*, continuum steps (both *voicing* and *duration*) and the interactions between these three as fixed factors.

Voicing and duration were added as numerical predictors. The estimates of the model and their significance are presented in Table 5.2. The logistic regression showed that there was a significant effect of voicing. This slope was negative (-1.124): the more to the left (the more voicing), the more /v/ responses. Furthermore, regions significantly differed from each other. Differences in slopes between regions are shown by the significant interactions between voicing and regions and are presented visually in Figure 5.4. West-Flanders had the steepest slope (-1.124), followed by Flemish-Brabant (-1.019). Limburg (-0.880) and South-Holland (-0.802) had a less steep slope and Groningen had the gentlest slope of all regions (-0.717). All regions except Flemish-Brabant significantly differed from West-Flanders, which is the reference region[3]. Flemish-Brabant differed from South-Holland and Groningen.

Figure 5.4: Schematic representation of the significant differences in voicing slopes between regions (α =.05).



[3]West-Flanders was taken as reference level in the intercept, since it is the region which was predicted to show the most conservative perceptual patterns (see Section 5.2.1).

Table 5.2: Estimated parameters of mixed-effects modelling on the fricative perception data (n=37678).

Random effects	Variance	Standard Deviation			
Participant (Intercept)	0.832	0.912			
Voicing by participant (Slope)	0.093	0.308			
Duration by participant (Slope)	0.016	0.127			
Fixed effects	Estimates	Standard Error	z values	p values	
Voicing	-1.124	0.073	-15.353	<.001	*
Duration	-0.031	0.033	-0.919	.358	
West-Flanders	1.380	0.210	6.563	<.001	*
Flemish-Brabant	1.321	0.209	6.318	<.001	*
Limburg	0.987	0.208	4.754	<.001	*
South-Holland	0.983	0.208	4.725	<.001	*
Groningen	0.848	0.208	4.067	<.001	*
Voicing * Flemish-Brabant	0.104	0.103	1.014	.310	
Voicing * Limburg	0.258	0.102	2.528	.011	*
Voicing * South-Holland	0.335	0.102	3.280	<.001	*
Voicing * Groningen	0.424	0.102	4.162	<.001	*
Duration * Flemish-Brabant	0.064	0.047	1.377	.168	
Duration * Limburg	0.008	0.046	0.171	.864	
Duration * South-Holland	0.047	0.046	1.032	.302	
Duration* Groningen	0.120	0.046	2.614	.009	*
Voicing * Duration	-0.077	0.009	-8.618	<.001	*
Voicing * Duration * Flemish-Brabant	0.004	0.012	0.318	.751	
Voicing * Duration * Limburg	0.034	0.011	3.027	.002	*
Voicing * Duration * South-Holland	0.037	0.011	3.362	<.001	*
Voicing * Duration * Groningen	0.050	0.011	4.729	<.001	*

Note. * The asterisk indicates significance (α =.05).

The model showed that there is no significant effect of duration. Participants did not use duration as a cue when categorizing the sounds. However, Groningen significantly differed from West-Flanders (the reference level) in their use of the duration cue, as shown by the significant interaction between duration and Groningen (small effect). Groningen (when taken as reference level) showed a slope that is significantly higher than 0 which means that these listeners used the duration cue in the opposite way as expected: the longer the duration of the sound, the higher the chance to categorize it as /v/. Listeners from other regions did not make use of the duration cue.

Finally, we observed a significant but weak interaction between voicing and duration (-0.077). The conjunction of both dimensions strengthened the regional effect that we found for voicing. With regards to this interaction, all regions except Flemish-Brabant, significantly differed from West-Flanders in the following order: Limburg followed by South-Holland and Groningen with the gentlest slope.

An analysis of variance performed on the estimated 50 % cut-off points in voicing revealed no statistical difference between regions (ANOVA: df=4;95, F=0.622, p=.648). Participants from all regions appeared to put the categorical boundary between /v/ and /f/ between step +1 and +2.

Stops The results of the identification task for stops are shown in Figure 5.5. In the upper graph representing the results along the VOT continuum, the centralized nine-steps continuum are presented on the x-axis and the proportion of /b/ responses on the y-axis. We observed that the categorization of bilabial stops was very categorical, since most stimuli were unanimously judged as instances of either /b/ or /p/. Only for steps +0 and +1 was the decision truly ambiguous and the cut-off point between the categories seemed to lie around step +1. The five regions seemed to be highly comparable in their categorization along the VOT dimension. For the duration continuum in the lower graph, it looked similar to the graph for durations in fricatives. We observed that for all regions the curves were nearly flat and that there are no cut-off points at 50 %.

These data were fitted with a generalized mixed effects logistic regression (n=38803). The random effects in this model consisted of the factor *participants* (N=99)[4], testing for individual differences between participants; the factor *VOT by participant*, testing for individual differences in VOT slope, and the factor *duration by participant*, testing for individual differences in duration slope. The fixed part of the model consisted of the factor *region*, continuum steps (both *VOT* and *duration*) and the interactions between these three as fixed factors.

VOT and duration were added as a numeric variable. The estimates of the model and their significance are presented in Table 5.3.

[4]The data of the stop identification task were lost for participant #WF12.

Figure 5.5: Results of the identification task for stops split up by region. The centralized nine-steps continuum along the VOT dimension (in the upper panel) and along the duration dimension (in the lower panel) are presented on the x-axis and the proportion of /b/ responses on the y-axis (n=38803). Error bars represent ±1 standard error.

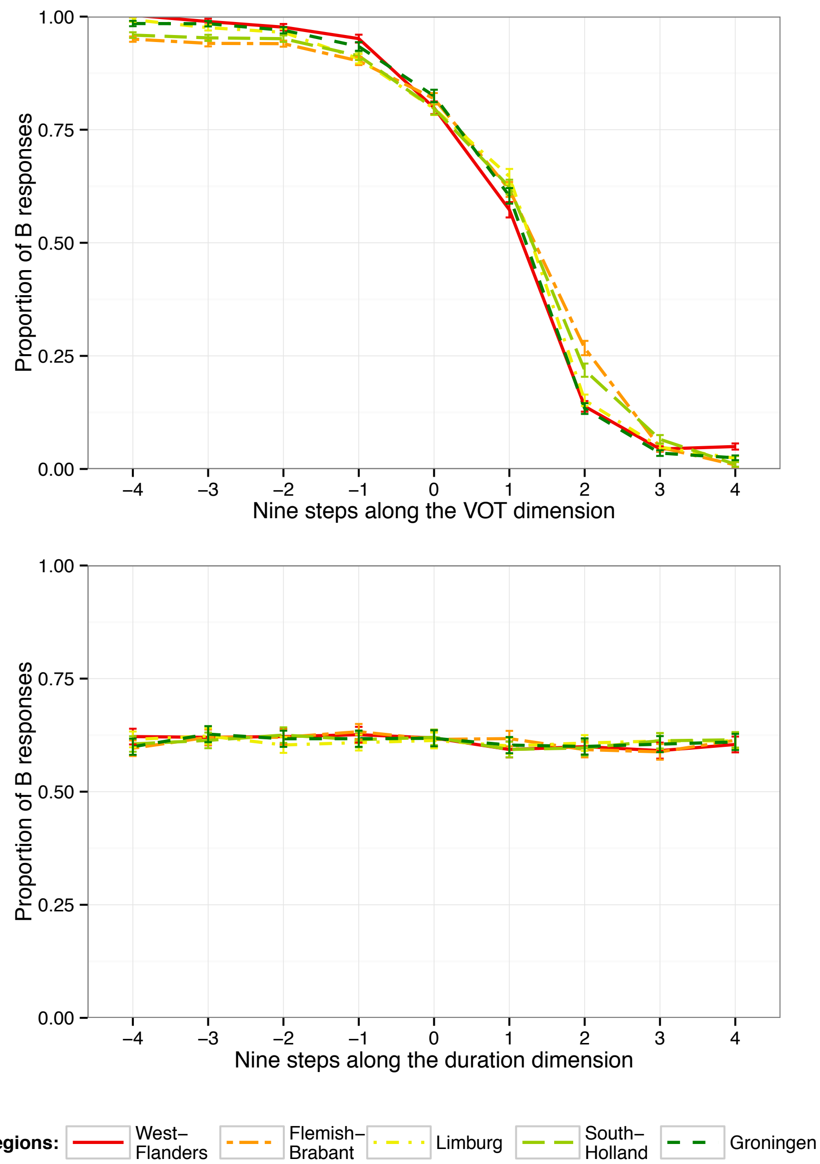


Table 5.3: Estimated parameters of mixed-effects modelling on the stop perception data (n=38803).

Random effects	Variance	Standard Deviation			
Participant (Intercept)	1.739	1.319			
VOT by participant (Slope)	0.327	0.572			
Duration by participant (Slope)	0.002	0.040			
Fixed effects	Estimates	Standard Error	z values	p values	
VOT	-1.613	0.138	-11.694	<.001	*
Duration	-0.015	0.021	-0.714	.475	
West-Flanders	1.611	0.310	5.199	<.001	*
Flemish-Brabant	1.596	0.303	5.277	<.001	*
Limburg	1.538	0.301	5.106	<.001	*
South-Holland	1.668	0.301	5.533	<.001	*
Groningen	2.001	0.303	6.586	<.001	*
VOT * Flemish-Brabant	-0.020	0.193	-0.103	.918	
VOT * Limburg	0.106	0.192	0.552	.581	
VOT * South-Holland	0.173	0.192	0.901	.367	
VOT * Groningen	0.056	0.193	0.289	.773	
Duration * Flemish-Brabant	-0.050	0.030	-1.689	.091	
Duration * Limburg	-0.012	0.029	-0.417	.676	
Duration * South-Holland	-0.015	0.029	-0.507	.612	
Duration* Groningen	-0.004	0.030	-1.455	.146	
VOT *Duration	-0.004	0.012	-0.326	.745	
VOT * Duration * Flemish-Brabant	0.028	0.016	1.747	.081	
VOT * Duration * Limburg	0.023	0.016	1.497	.135	
VOT * Duration * South-Holland	0.025	0.015	1.651	.099	
VOT * Duration * Groningen	0.037	0.015	2.458	.014	*

Note. * The asterisk indicates significance (α =.05).

The logistic regression showed that there is a significant effect of VOT. This slope is negative (-1.613) since the more to the left (the more voicing), the more /b/ responses. Importantly, there was no significant difference between regions in the effect played by VOT, as can be seen from the interactions between VOT and regions. Participants from all regions showed highly comparable perceptual patterns along the VOT dimension. It can also be seen that there is no significant effect of duration. Participants did not seem to use duration as a cue when categorizing the stops.

An analysis of variance performed on these estimated 50 % cut-off points in VOT revealed no statistical difference between regions (ANOVA: df=4;94, F=1.038, p=.392).

Effects within regions

In a second step of the analysis, separate logistic regressions were fitted per participant. Each of these regressions was run on 405 observations (i.e., the number of stimuli per participant) and contained the *continuum steps* as numerical predicting factor (either voicing/VOT or duration)[5]. As shown in the previous section, the cut-off points between categories did not vary much across participants. Hence, in the following analysis participants are compared in terms of their categorization slopes only. The values of the slopes along both dimensions were calculated for each participant from the estimates obtained in the separate logistic regressions.

The following section reports on the use of phonetic cues by individual listeners, revealed by the significance of slope estimates in the logistic regressions. Subsequently, differences in slope values between individuals are investigated.

Fricatives The use of the phonetic cues in the categorization of fricatives are presented in Table 5.4 (in the upper part). For each phonetic dimension, the number of participants for which the slope turned out to be significantly different from 0 is reported. A significant positive or negative slope means that participants used the dimension during categorization. As far as the use of voicing is concerned, all participants except one from Limburg did significantly use it for perception. In contrast, merely one quarter of the participants (N=25) used duration as a phonetic cue. The number between brackets represents the number of participants showing a negative duration slope (N=18). The others significantly used duration, but showed a positive slope (N=7), which means that they categorized – against expectations – longer sounds more often as voiced.

[5]This procedure was preferred to calculating the random slopes per participant on the basis of the full model, since we sought to obtain an indication about the significant use of the phonetic cues at the individual level. The risk of shrinkage (i.e. calculating wrong slope estimations) for which Baayen et al. (2008, p.225) warned is very limited in these data, since there are very strong correlations between slopes obtained in separate models and random slopes from the full model (r= .95 for fricatives and .90 for stops).

A traditional analysis of variance[6] run on the values of the voicing slopes (N=100) revealed a main effect of *region* (ANOVA: df=4;94, F=3.355, p=.013)[7] which matched with the regional effects described in the previous section. The factor *sex* turned out to be marginally significant (ANOVA: df=1;94, F=3.693, p=.058), as women have a slightly more negative slope than men (-0.952 vs. -0.808), thus more categorical perception. The analysis of variance run on the duration slopes (N=100) revealed a main effect of *regional accent*[8] (ANOVA: df=1;98, F=7.556, p=.007). Participants who claimed to have a regional accent showed a slightly negative duration slope (-0.035), whereas participants who did not, showed a slightly positive slope (0.016). This can be explained by the fact that participants who claimed to have a regional accent were mainly participants from West-Flanders, Flemish-Brabant and Limburg (see Chapter 3), which are the regions with more categorical perception. Regional accent was therefore a confounding variable with the factor region.

Table 5.4: Individual use of phonetic cues in the identification task of fricatives and stops: number of participants who used the different phonetic cues split up by region. The number of participants showing a negative duration slope is written in parentheses.

Fricatives	Use of voicing N	Use of duration N
West-Flanders	20	4(4)
Flemish-Brabant	20	2(2)
Limburg	20	6(5)
South-Holland	19	7(3)
Groningen	20	6(4)
Total	99/100	25(18)
Stops	Use of VOT N	Use of duration N
West-Flanders	19	0
Flemish-Brabant	20	1(1)
Limburg	20	1(1)
South-Holland	20	0
Groningen	20	0
Total	99/99	2(2)

Stops The use of the phonetic cues in the categorization of stops is presented in Table 5.4 (in the lower part). All participants significantly used VOT when

[6]Mixed models were not required for this analysis, since there was only one observation per participant in the aggregated data.

[7]The other factors *dialect proficiency*, *regional accent*, *mobility* and *contact with the other country within the Dutch language area* did not turn out significant and were therefore excluded from the model.

[8]The other factors *region*, *sex*, *dialect proficiency*, *mobility* and *contact with the other country within the Dutch language area* did not turn out significant and were therefore excluded in the model.

categorizing bilabial stops. In contrast, only two participants used duration as a phonetic cue. Both of them used duration in the expected direction, namely with a negative slope, categorizing longer sounds more as voiceless. The analyses of variance run on the VOT slopes (N=99) and on the duration slopes (N=99) did not reveal any effect of social factors (sex, dialect proficiency, etc.).

5.2.3 Discussion

The first perception experiment was an identification task in which each participant categorized fricative and stop stimuli as being either voiced or voiceless. The experiment aimed at obtaining quantitative insights in the regional differences in the categorization of these phonemes and in individual perceptual patterns.

A first goal was to determine which phonetic cues the listeners rely on when categorizing fricatives and stops. Our results confirm the findings of previous perceptual studies (Ali et al., 2001a; Slis & Cohen, 1969; van den Berg, 1988), as voicing turned to be the major dimension along which fricatives were categorized. As expected, participants from all regions consistently used them in the identification task: the more voicing in the fricative, the more voiced it is perceived. Following previous perception studies on stops (Ali et al., 2001b; Cho & Ladefoged, 1999; Smits & van Alphen, 2004), we found that VOT turned to be the major dimensions along which stops were categorized. The more negative the VOT in the plosive, the more voiced it is perceived. Duration however did not play a role in the categorization of stops. The differences in fricatives and stops appeared in the calculation of categorization slopes, but not in the cut-off points between categories. Hence, devoicing in perception appeared to show up as a less categorical perception, and not so much as a difference in the placement of the perceptual boundary between categories.

However, duration turned out to play a minor role in the perception of fricatives, as there was no significant effect of duration and only a weak interaction with voicing in fricatives, and no role of duration at all in the perception of stops. This absence of the duration effect might mean either that listeners effectively did not use duration as a phonetic cue to categorize fricatives and stops, or that this result is an artefact of the type of experiment. In this kind of phonetic experiments, it is indeed possible that participants did not rely on the duration cues. Since the CV stimuli were presented in isolation, participants might not have been able to normalize for duration against other surrounding syllables, as it is normally done in spontaneous speech processing. Consequently, they might have shut down the use of the duration cue. As a result, it might not entirely be clear from this experiment what role duration plays in the perception of fricatives and stops.

Furthermore, the experiment aimed at revealing regional differences in the perception of the fricative and stop contrasts. The results showed that region was a significant predictor of perceptual behavior in the fricative data. Listeners showed significant regional differences in the way they used the main cue, voicing. The perceptual patterns in West-Flanders were characterized by a steep psychometric curve (thus a more categorical perception) and Groningen by a gentle slope (thus a less categorical contrast and more merged categories), and the other regions taking a position in between these two extremes. These patterns highly coincide with our expectations based on previous production studies and with the production results in Chapter 5, describing the geographical spread of the sound changes.

In contrast to fricatives, no regional differences could be found in the categorization of stops. Participants from all regions displayed a very similar use of VOT as a perceptual cue. The absence of regional variation in perception in the stop data could be attributed to the fact that the sound change in bilabial stops is still in an incipient phase. Regional perceptual differences might still be too subtle to be revealed in this kind of task. Moreover, it raises the question of the link between speech perception and speech production at the beginning of a sound change. In a model where changes in speech production precede changes in perception, this results could be justified by the fact that stops might already showed devoicing in production, but not yet in perception.

Regarding later and final stages of a sound change, it is interesting to look at the regions showing most fricative devoicing. In the regions Groningen and South-Holland, where the devoicing is very advanced in production with almost fully merged categories, the perception slopes were not totally flattened. Thus, listeners from these regions maintain traces of categorical perception, a fact also observed by Hay, Warren, and Drager (2006) in New Zealand English for merging diphthongs. It thus seems that the end stage in perception could be different from the end stage in production. This matter will be discussed extensively in Chapter 8.

5.3 The similarity judgement task

After the above-reported forced-choice identification experiment, a second perception experiment was conducted: the similarity judgment experiment. In contrast to the identification experiment, the similarity judgment experiment investigates the gradient aspect of speech perception. It aims to represent the perceptual space of participants qualitatively. In Section 5.3.1, the method used in the similarity judgment experiment is described. The results of the experiment are reported in Section 5.3.2 and discussed in Section 5.3.3.

5.3.1 Method

The second perception experiment is a similarity judgment task. Like the forced-choice identification task, it aims at investigating perceptual differences in fricatives and stops. Previous studies having the same goal (e.g., Evans & Iverson, 2004; Kendall & Fridland, 2012; Sumner & Samuel, 2009) often used either identification tasks or discrimination tasks. Similarity rating tasks are less standard than discrimination tasks in speech perception research. However, like Johnson and Babel (2010), we believe that a similarity judgment task is preferable to an AXB discrimination task, because it is less memory intensive and more efficient, since listeners provide a gradient and informative perceptual response. In this section, the stimuli used in the similarity judgment task are described together with the experimental procedure. Subsequently, the analysis of the results with multidimensional scaling is explained.

Stimuli

The same phonetic continua of fricatives and stops were used as in the identification tasks (see Section 5.2.1). However, in order to reduce the length and the cognitive load of the experiment, a subset of 25 sounds was selected from the 81-sounds original continua, namely every other step along both nine-steps dimensions. The chosen sounds are those which are labeled in Figures 5.1 and 5.2. The other sounds were disregarded for this task. For both the fricative and the stop continuum, all 25 sounds were paired with each other with 250 ms of silence in-between. In this way, a total of 325 stimulus pairs for both fricatives and stops were obtained.

Procedure

Participants were seated in a sound-attenuated booth and listened to the 325 pairs of stimuli via a headphone Beyerdynamic DT 250 (for more details see Section 3.5). For each trial, they were asked to rate the similarity between the two sounds forming the pair. The similarity was rated on a continuous scale on the computer screen ranging from *same* [gelijk] to *different* [verschillend]. The response was measured in pixels on the screen, and then converted to a scale from 0 to 100. Traditionally, researchers using MDS used equally-appearing interval similarity scales with seven or nine points (e.g., Fox, Flege, & Munro, 1995; Gandour & Harshman, 1977; Iverson et al., 2003). In this study however, a continuous scale was used, since this kind of scale better represents the gradient aspect of perception. Reaction times (RT) were recorded from the beginning of the stimuli. Each participant completed a practice block of twelve trials. After the practice, participants took the experimental session with all pairs presented in a randomized order.

Multidimensional scaling

Results of a similarity judgment task are traditionally analyzed with *multidimensional scaling* (MDS), which has proven to be the most suitable method to geometrically model the perceptual space. MDS has extensively been used to study the psychological space of vowel and tone systems (Casserly, 2010; Fox, 1982, 1983, 1985; Fox et al., 1995; Gandour & Harshman, 1977; Kewley-Port & Atal, 1989; Mayo, Clark, & King, 2005; Mayo et al., 2005; Pols, Van der Kamp,

& Plomp, 1969; Rakerd, 1984; Rakerd & Verbrugge, 1985; Shepard, 1972; Singh & Woods, 1971; Sinnott, Brown, Malik, & Kressley, 1997; Terbeek, 1977) and to a smaller extent of consonants (Iverson & Kuhl, 1995; Iverson et al., 2003).

MDS is a mathematical tool that enables us to represent the similarities of perceptual objects spatially in a map. It models the degree of perceived similarity or dissimilarity between stimuli to derive a stimulus space, in which the distances between stimuli in the space correspond to the similarity values. Similar stimuli are placed closer and dissimilar stimuli further apart from each other. Additionally, the dimensions that make up the stimulus space correspond to the dimensions used most heavily by the listeners to make their proximity judgments. It is assumed that the dimensions of the space correspond to perceptual dimensions underlying the subjects ratings of similarity (Gandour & Harshman, 1977). Subsequent analysis of these dimensions reveals the physical or psychophysical characteristics of the stimuli on which proximity judgments are made (Mayo et al., 2005).

MDS was used in the nineties by Kuhl and her colleagues as a tool to test and demonstrate the *native language magnet theory* (Kuhl, 1991). According to this theory, the phonetic categories of ones native language are organized in terms of prototypes or 'best instances' of specific speech sounds (see also Section 2.1.1). These prototypes function as 'perceptual magnets', since they attract or assimilate phonetically similar members and thus facilitate the processing of the variability of speech input. Consequently, the perceptual space is 'distorted' (Iverson & Kuhl, 1995): the best exemplars of a category pull neighboring tokens closer in the perceptual space. Perceptual distances are shrunken near excellent exemplars of a category (reduced acoustic sensitivity to phonetic differences), whereas distances are stretched between poor exemplars (increased acoustic sensitivity to phonetic differences). Meanwhile exemplar-based models have proven to account better for perceptual variation (as explained in Chapter 2). The magnet effect has been shown to emerge from this exemplar-based organization (Lacerda, 1995) and to be a consequence of optimally solving the statistical problem of perception (Feldman, Griffiths, & Morgan, 2009; Guenther & Gjaja, 1996). It is however interesting to examine in our data whether some traces of this effect are found: the distortion of the perceptual space with a shrinkage around the best exemplars of a category.

5.3.2 Results

Each listener's responses were compiled into dissimilarity triangular matrixes (N=100), in which each cell in the matrix represented the score given for a specific pair. In order to investigate group differences, matrixes for each region (n=5) were obtained by averaging the ratings given by all participants of the same region for each stimulus pair. Multidimensional scalings were carried out on the similarity matrixes (both the *individual matrixes* and the *regional matrixes*) using the `MASS` package in R (R Core Team, 2014). The chosen MDS solutions were bi-dimensional.

MDS solutions per region

Fricatives The bi-dimensional MDS solutions for fricatives split up by region are presented in Figure 5.6. In this figure, the schematic representation of the phonetic continuum is presented in a box, and allows to visualize the differences between phonetic distances and psycho-acoustic distances. The five steps along the voicing dimension are presented horizontally and the five steps along the duration dimension vertically.

For all regions, the ordering of the stimuli on the horizontal axes corresponds to the voicing dimension and the vertical axes correspond to the duration dimension, which demonstrates a strong correlation between similarity scores and acoustic distances. Moreover, we observed that the perceptual spaces are shrunk around the extremes, showing evidence for the magnet effect. Perceptual distances between fully voiced stimuli of different durations are reduced. In the same way, voiceless stimuli of different durations were highly concentrated in the perceptual space, so that they are perceptually almost equal.

Additionally, the voiced segments of the three rightmost columns of the continuum are in all regions overall closer to each other. This shrinkage in the most voiced fricatives appeared to be stronger than the shrinkage at the side of the voiceless fricatives, which suggests that the devoicing of fricatives in perception concerns voiced stimuli in the first place. This findings is in line with findings in production data (see Chapter 4), in which it was shown that voiced fricatives are devoicing. /v/ is produced more and more like [f], and not the other way around.

As far as regional differences are concerned, MDS solutions seemed highly similar across regions. Especially at the rightmost side (i.e., the side of voiceless fricatives), very few differences between regions can be found. At the leftmost side, West-Flanders and Limburg showed a dense aggregation of the stimuli, whereas the density in that part of the continuum seemed a bit less in the other regions.

Stops The bi-dimensional MDS solutions for stops are presented split up by region in Figure 5.7, together with the schematic representation of the phonetic continuum. The five steps along the VOT dimension are presented horizontally and the five steps along the duration dimension vertically.

For all regions, the MDS solutions for stops seemed to contrast strongly with the ones of fricatives. The order of the stimuli on the horizontal axes clearly corresponds to the VOT dimension, but the stimuli along the vertical axes are highly shrunken so that there is hardly any effect of the duration dimension. The correlation between similarity scores and acoustic distances in the perception of stops is thus less straightforward. Like for fricatives, we observed that the perceptual spaces are shrunk around the extremes.

Figure 5.6: Bi-dimensional MDS solutions of the fricative data split up by region and the reference phonetic continuum. The voicing dimension is presented horizontally (the more to the right, the more voicing) and the duration dimension vertically (the more to the bottom, the longer the duration).

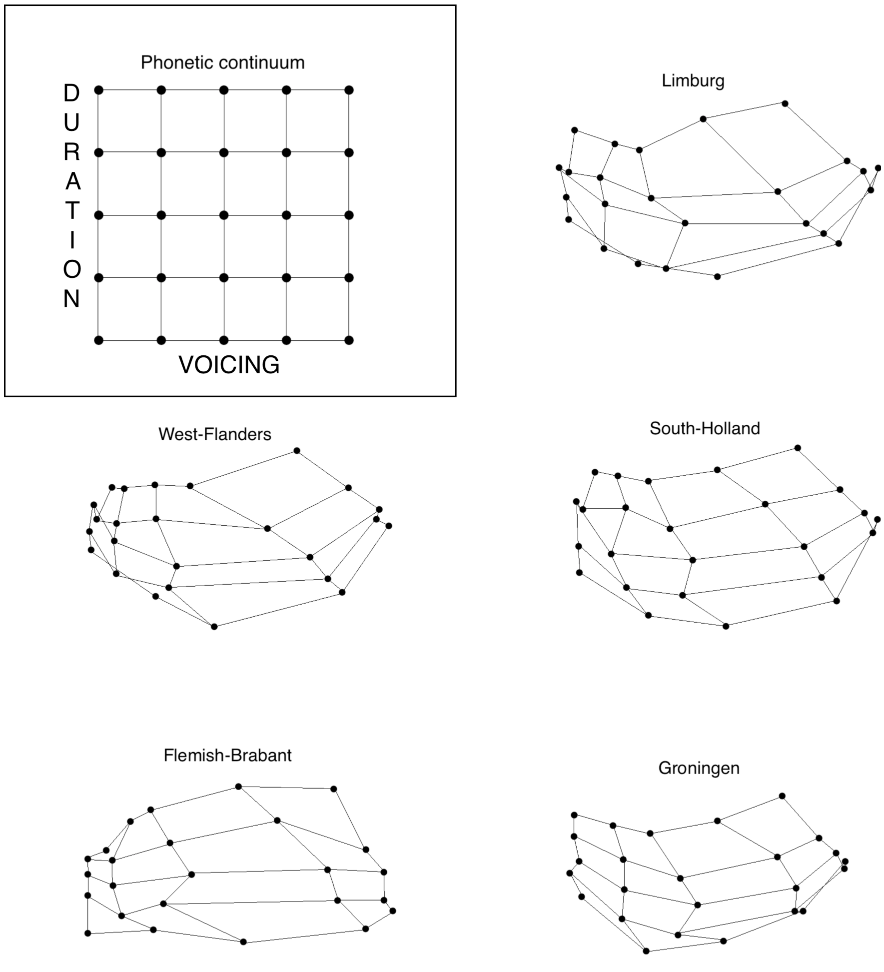
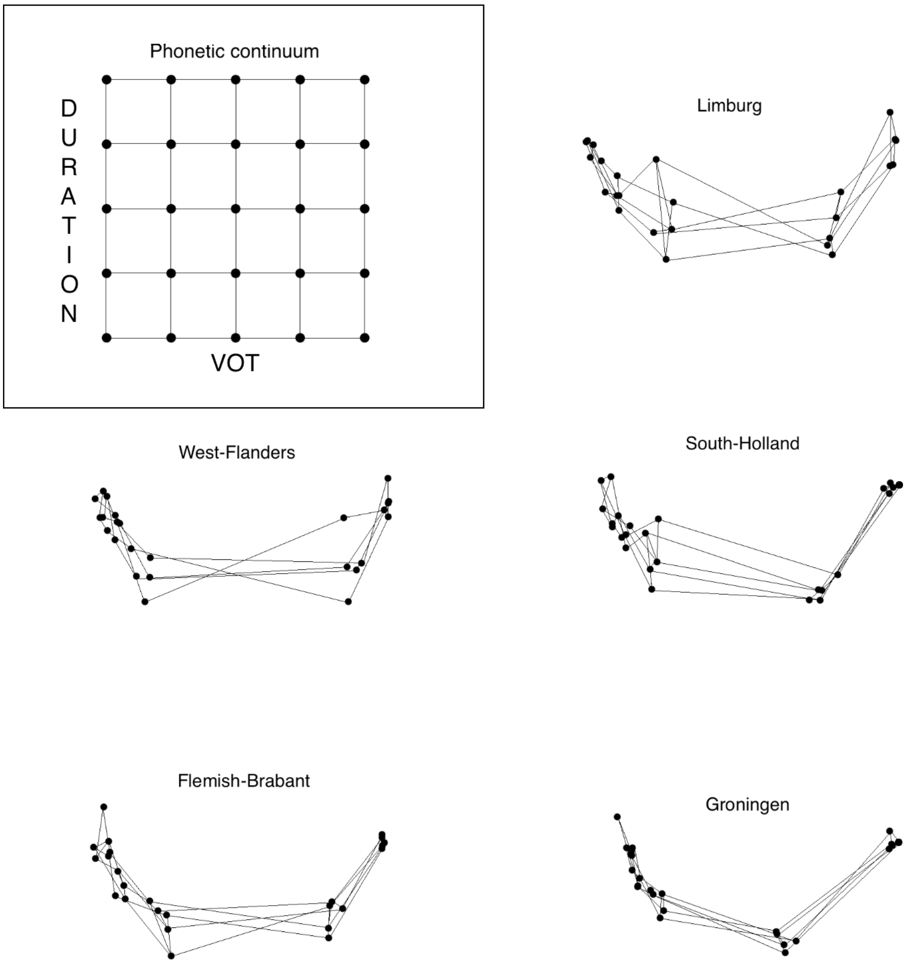


Figure 5.7: Bi-dimensional MDS solutions of the stop data split up per region and the reference phonetic continuum. The VOT dimension is presented horizontally (the more to the right, the more positive the VOT) and the duration dimension vertically (the more to the bottom, the longer the duration).



Perceptual distances between the three leftmost columns (voiced stimuli) are almost equal to 0, implying that they are perceptually equivalent. In the same way, voiceless stimuli of different durations are highly aggregated. In all regions, the perception boundary between voiced and voiceless bilabial stops appeared between step 3 and 4 along the voicing dimensions[9].

As far as regional differences are concerned, MDS solutions seemed highly similar across regions. At the leftmost side (i.e., the side of voiced stops), very few differences between regions can be found except Groningen, which showed the strongest shrinkage of stimuli in duration. At the rightmost side, South-Holland and Groningen tended to show some perceptual differences between the last but one and last column along the VOT dimension, but these are minor differences.

Individual MDS solutions

The individual MDS solutions can be considered as a representation of the perceptual space present in participants' psycho-acoustic system. A visual analysis of each individual configuration was carried out by the author, who established from the MDS solution which phonetic cues had been used by participants. It was qualitatively determined whether the participant clearly used a cue, clearly did not use the cue or whether it was unclear from the MDS solution if the participant had used the cue or not. In Figure 3 of Appendix I, examples of the categorization of individual MDS solutions are presented. In this figure, Part (a) is the schematic representation of the phonetic continuum. Part (b) is an example in which a participant did clearly use the voicing dimension, but did not use duration (participant #FB02 for the perception of fricatives). Part (c) is an example in which a participant clearly used both dimensions (participant #GR13 for the perception of fricatives).

Fricatives The use of the phonetic cues in fricative similarity judgments are presented in Table 5.5 (in the upper part). For each phonetic dimension, the number of participants for which the visual analysis of the MDS solution revealed that they had used that dimension, is reported.

All participants turned out to use voicing when rating similarities between fricatives. 46 participants showed a significant use of duration as a cue. These participants were equally spread across regions.

Stops The use of the phonetic cues in similarity judgments of stops are presented in Table 5.5 (in the lower part). All participants significantly used VOT when categorizing bilabial stops. Only one participant also used duration as a phonetic cue. This participant (#FB05) turned out to be one of the two

[9]These steps correspond to the steps 0 and +2 of the full continuum used in the identification.

participants who also showed a significant use of duration in the identification task.

Table 5.5: Individual use of phonetic cues in individual MDS solutions: number of participants who used the different phonetic cues split up by region.

Fricatives	Use of voicing n	Use of duration n
West-Flanders	20	11
Flemish-Brabant	20	9
Limburg	20	10
South-Holland	20	9
Groningen	20	7
Total	100/100	46/100
Stops	Use of VOT n	Use of duration n
West-Flanders	20	0
Flemish-Brabant	20	1
Limburg	20	0
South-Holland	20	0
Groningen	20	0
Total	100/100	1/100

5.3.3 Discussion

The second perception experiment was a similarity judgment task in which each participant rated the similarity of fricative and stop pairs. The experiment was aimed at gaining qualitative insights in both the regional and individual perceptual patterns.

A first goal was to determine which phonetic cues the listeners rely on when perceiving fricatives and stops. As in the identification task, voicing and VOT were the major dimensions along which fricatives and stops respectively were categorized. Participants from all regions consistently used them. Duration in this task turned out to be a phonetic cue on which half of the listeners rely for the perception of fricatives. In contrast, there was no effect of duration in the perception of stops.

The analyses of the MDS solutions clearly showed that the perceptual space both in the regional and in the individual solutions were distorted in such a way that perceptual distances near the extremes of a category were shrunken (reduced acoustic sensitivity to phonetic differences) and distances between categories were stretched (increased acoustic sensitivity to phonetic differences). This effect was even stronger for stops than for fricatives. Traditionally, this effect was interpreted as evidence for the *magnet effect* (e.g., Iverson & Kuhl, 1996).

Furthermore, the experiment aimed at revealing regional differences in the perception of the fricative and stop contrasts. The results for both fricatives and stops showed that the MDS solutions were very similar across regions, revealing only subtle perceptual differences. As Iverson and Kuhl (1995) explained, the MDS technique does not provide an absolute measure of perceptual distances, since MDS solutions are normalized on the basis of the distribution of tokens. Hence the distortion of the perceptual space can qualitatively be compared across regions, yet the magnitude of the distortion cannot really be measured.

In order to obtain regional MDS solutions, we aggregated the similarity judgments of all participants within a region. This way of processing is common in MDS studies (e.g., Iverson & Kuhl, 1995, 1996; Sinnott et al., 1997). Gandour and Harshman (1977) have claimed that such total sample combined-group solutions are more stable than the individual solutions, because of the larger sample size and because it allows direct comparison between groups. Even if we have conducted and reported such an analysis on group averages, we want to express some scepticism about the ecological validity of this type of analysis. As MDS is a tool aimed at providing a direct representation of an individual's perceptual space, the question raised is what researchers do when they average this space and its distances across participants. The obtained result is a kind of abstraction of regional patterns, but in fact cannot be considered as a direct representation of an individual perceptual space any longer.

5.4 Comparing the results: identification vs. similarity judgement

In this section, the results obtained from both perception tasks are summarized, discussed and compared.

5.4.1 The relationship between phonetic distances and perceptual distances

Following the method proposed by Iverson and Kuhl (1996), a direct comparison is made between phonetic distances between sounds of the continua, and perceptual distances as examined in the two perception tasks (i.e., the identification task and the similarity judgment task). The goal of this comparison is to evaluate both perception tasks in terms of the insights they provided. Like Iverson and Kuhl (1996), our comparison involves three types of perceptual distances: *phonetic distances*, *distances obtained from the identification task* and *distances obtained from the similarity judgment task*:

1. **Phonetic distances** were calculated by measuring the Euclidian distances between the original stimuli from the bi-dimensional grid. Each step represented a difference of a value 1, both in voicing/VOT and duration. According to the Euclidean distance formula, the distance between two stimuli (S1 and S2) in the plane with coordinates (x, y) and (a, b) is

given by:

$$dist((x, y), (a, b)) = \sqrt{(x-a)^2 + (y-b)^2} \qquad (5.3)$$

For example, the distance between the stimuli S1 (0,0) and S8 (2,1) differed with a value of 2.24 ($\sqrt{(0-2)^2 + (0-1)^2} = \sqrt{5} = 2.24$). Phonetic distances ranged from 0 (no distance) to 5.66 (which is the distance for instance between the left-most stimulus in the upper left corner and the right-most stimulus in the lower right corner, see reference phonetic continuum in Figure 5.7)

2. **Distances obtained from the identification task** were calculated as identification z scores following Macmillan and Creelman (1991) and Iverson and Kuhl (1996). Z-transformed identification probabilities were calculated for each stimulus[10]. This measure indicates the distance between each stimulus and the category boundary in standard deviations units. The sign of the measure indicates whether the stimulus is within (positive sign) or outside (negative sign) the category. For example, a z score of 0 means that the stimuli was identified as a member of the category on 50 % of the trials, a negative z score means that the stimuli was identified more often as another category and a positive z score means that the stimuli was identified as member of the category in the majority of cases. Subsequently, the perceptual distance between stimuli could be calculated as the absolute value of the subtraction between the two z scores.

3. **Distances obtained from the similarity judgment task** are the distances between each pair of stimuli obtained from the individual MDS solutions. They represent Eucledian distances within the MDS space.

Correlations were run between the three types of distance measures and are reported in Table 5.6. The strength of correlations is defined as follows: r=.20-.40 is a weak correlation, r=.40-.70 is a moderate correlation, r=.70-.90 is a strong correlation and r=.90-1 is a very strong correlation (Dancy & Reidy, 2007).

For fricatives, there were moderate correlations between the phonetic distances on the one hand, and the identification and MDS distances on the other hand (r=.51 and r=.58, respectively). These correlations were mostly triggered by the strong correlation along the voicing continuum (i.e., the first correlation between brackets). Phonetic distances in duration only weakly correlated with the duration distances in the perception tasks (r=.08 with identification distances and r=.24 with MDS distances). Although both correlations were weak, the correlation between phonetic distances and MDS distances was slightly stronger, showing again the use of the duration cue in the similarity judgment task. Moreover, there was a moderate correlation between the identification distances and the MDS distances (r=.57). Since both tasks were designed to

[10] Only the subset of stimuli which were used in the similarity judgment task were taken for this analysis (n=25 from the original 81).

give insights in the same perceptual patterns, we expected this correlation to be higher. The reason why it is not as high as expected probably lies in the fact that the tasks quite differed in the role attributed to the duration dimension.

For stops, correlations were also moderate between the phonetic distances on the one hand, and the identification and MDS distances on the other hand (r=.53 and r=.49, respectively). In this case also, correlations were triggered by the strong correlation along the VOT continuum. Phonetic distances in duration only weakly correlated with duration distances in the perception tasks (r=.06 with identification distances and r=.07 with MDS distances). Moreover, there was a relatively strong correlation between identification distances and MDS distances (r=.70). As both perception tasks on the stops provided similar results, this strong correlation was expected.

Table 5.6: Correlations between the three types of distances measures for fricative data (in the upper part) and stop data (in the lower part). Correlations for separate dimensions are presented between brackets (voicing/VOT & duration). All reported correlations are significant.

	Phonetic distances	Distances ident.	Distances MDS
Fricatives	Phonetic distances	r=.51 (.65 & .08)	r=.58 (.59 & .24)
		Distances ident.	r=.57
			Distances MDS
	Phonetic distances	Distances ident.	Distances MDS
Stops	Phonetic distances	r=.53 (.72 & .06)	r=.49 (.66 & .07)
		Distances ident.	r=.70
			Distances MDS

Multiple regression analyses were ran on these distance measures in `R` (R Core Team, 2014) in order to assess the relative contribution of phonetic distances and identification distances to MDS distances. MDS distances were thus added to the models as dependent variable. Phonetic distances and identification distances were stepwise added as dependent variables. For fricatives, the model with phonetic distances could only predict 33 % of the variance in MDS distances[11].

[11] The percentage of accounted variance is given by the adjusted R^2 of the linear regression.

When identification distances were added to the model, the accounted variance rose to 43 %. For stops, the model with phonetic distances could predict 24 % of the variance and the full model with both the phonetic distances and the identification distances accounted for 51 % of the variance.

Overall, it appeared that phonetic distances could account for a relatively small proportion of the variance in MDS distances. This proportion was slightly larger for fricatives than for stops, which shows that fricatives were perceived more phonetically than stops. The addition of identification distances to the model of both fricatives and stops significantly helped to increase the proportion of explained variance. The total proportion of explained variance however remained quite low (43 % and 51 % respectively), especially when compared to the results of Iverson and Kuhl (1996). With the same procedure, they could account for 84 % of the variance in MDS distances with phonetic distances and identification distances as predictors. However, they looked to a contrast in liquids which was not involved in sound change, and might explain this result. In conclusion, an important part of the variance in MDS distances in our data still remains to be explained, for instance by factors related to the individual listeners or to the stimuli.

5.4.2 Combined insights from the perception tasks

The perception of labiodental fricatives and bilabial stops was investigated through two perception experiments: an identification task and a similarity judgment task. The identification task was aimed at gaining quantitative insights in the regional and individual perceptual patterns. The similarity judgment task aimed to study phonetic perception in a more qualitative manner without requiring that listeners identify sounds. Previous research (Boomershine, Hall, Hume, & Johnson, 2008; Huang, 2004; Johnson & Babel, 2010) has shown that these tasks are able to reveal cross-linguistic differences in perception. In this study, they were however used for the investigation of perceptual differences within the same language.

In this study, both the identification task and the similarity judgment task clearly succeeded in providing valuable insights in differences in speech perception. These insights concerns first the phonetic cues along which fricatives and stops are perceived, and second regional and individual perceptual differences. Nonetheless, these tasks have shown to display distinct strengths. The identification task was very useful in revealing quantitative results on regional and individual differences. The similarity judgment task appeared less suited at showing regional variation, but more effective at determining the precise use of each phonetic cue.

Looking at the phonetic cues used in the perception of fricatives and stops, voicing (for fricatives) and VOT (for stops) in both tasks turned out to be the major dimensions along which participants categorized. Participants from all regions consistently used these dimensions. Yet, the tasks differed in the role attributed to the duration dimensions. In the identification task, duration played no role in the perception of neither fricatives nor stops. In the similarity judgment task, there was a clear effect of duration in the perception of fricatives. The non-result of duration in the identification task can therefore rather be interpreted as an artefact of this type of experiment. In the identification task, participants did not rely on the duration cues since CV syllables were presented in isolation, and therefore participants were not able to normalize for duration against other surrounding syllables. In contrast, in the similarity judgment task, stimuli were presented in pairs, which allowed participants to compare the duration of one fricative to the one of the other fricative and use duration in a more natural way. Regardless of this difference between the two experiments, it is clear that duration only played a minor role in the perception of both fricatives and stops. Therefore, the analyses related to speech perception provided in the next chapters will mainly focus on voicing (for fricatives) and VOT (for stops) as the main phonetic cues used in the perception of these contrasts.

It is interesting to note that the loss of categorical perception in some regions does not directly imply a merger in perception. It might be the case, for instance, that the voicing dimension is merely replaced by one or more other phonetic cues. In our perception tasks the manipulation was limited to two phonetic dimensions. However, adding more phonetic dimensions would help to establish whether there are new dimensions replacing the older ones helping the listener to maintain contrasts in perception.

The other main research questions concerned the regional and individual differences in speech perception. Like Huang (2004), Boomershine et al. (2008) and Johnson and Babel (2010), we succeeded in showing regional differences in the perception of phonemes undergoing sound change. This type of differences is thus not limited to vowel variation, but also shows up in consonants. In our data, regional differences were found for fricatives, but not for stops. As previously explained, the absence of regional variation in the perception of stops could possibly be attributed to the fact that sound change in bilabial stops is still in an incipient phase. Regional perceptual differences may still be too subtle to be revealed.

The regional differences shown in the perception of fricatives established the first link between speech production and speech perception. Indeed, group of listeners presented differences in perception which are in line with speech production patterns (see Chapter 5). As a result, it is clear that variation involved in sound change is not limited to production, but is also a perceptual phenomenon. The relationship between variation in perception and variation in production will be investigated further in Chapter 8.

CHAPTER 6

Language attitudes

As explained in Chapter 2, speaker-listeners constantly evaluate competing linguistic forms through language attitudes (Foulkes, Scobbie, & Watt, 2010). This chapter aims at investigating to what extent language attitudes play a role in sound change. More concretely, we examine whether the devoicing of labiodental fricatives and bilabial stops is socially indexed and triggers prestige in the different regions (at the *group level*) and among individual listeners (at the *individual level*).

This question is investigated through a matched-guise attitude experiment. The chapter begins with an introduction on language attitudes and the formulation of hypotheses in Section 6.1. In Section 6.2, the method used to test language attitudes is described. In Sections 6.3 and 6.4, the results of the attitude experiment are respectively reported and discussed.

6.1 Introduction and hypotheses

The goal of the current chapter is to determine the sociolinguistic meaning attributed to the devoicing of labiodental fricatives and to the devoicing of bilabial stops in Dutch, a goal which – to our knowledge – has not been attempted in any previous study. It is investigated whether listeners' attitudes differ across regions (at the *group level*) and across listeners (at the *individual level*).

As the language attitudes towards these phenomena have not been investigated quantitatively previously, it is difficult to formulate precise hypotheses. Nevertheless, in view of the regional differences in the production and perception (see Chapters 4 and 5), we hypothesize that language attitudes towards our variables, especially the devoicing of fricatives, will show regional differences. We expect to find more positive attitudes in regions where the devoicing phenomena are most advanced (e.g., Groningen and South-Holland for fricatives), as the change seems already to be above the level of consciousness in these regions. Since bilabial stops did not show clear regional variation in production and perception and since this sound change is still in an incipient phase (see Chapters 4 and 5), we do not expect that this variable triggers clear-cut language attitudes. The devoicing of stops is a variable that might be described as *below the level of consciousness* and might enjoy some *covert prestige*.

As explained in Section 2.4, it is often found in the processing of results from attitudinal research that ratings on linguistic variables correlate into two basic dimensions: *power* and *solidarity*. They are often found to be the two primary evaluative dimensions of language attitudes, as Brown and Gilman (1960) first demonstrated. Grondelaers, Van Hout, and Steegs (2010) proposed a third possible dimension of language attitudes articulated around notions of *personal integrity*. We expect to find that our ratings will be articulated around these dimensions as well.

6.2 Method

In this section, the matched-guise technique used in this study is first described in Section 6.2.1. In Section 6.2.2, the speech stimuli used in the matched-guise experiment are described, followed by the description of the experimental procedure (in Section 6.2.3).

6.2.1 The matched-guise technique

The *matched-guise technique* was first used by Lambert et al. (1960) to elicit listeners' attitudes towards different languages. The technique traditionally employs speakers capable of producing two *guises* (two languages or varieties) fluently, so that a native participant of either language would assume that the speaker of the stimulus heard was native. Listeners are required to listen to the recordings and asked to rate the apparently different speakers' character traits on subjective scales including intelligence, friendliness, socio-economic status, etc. The language attitudes are determined by the analysis of the difference scores in the ratings of the same speaker using different guises. By using a single talker to produce speech samples, variation in ratings can safely be attributed to different attitudes towards the guises and not to inherent differences between speakers of different varieties (Clopper & Pisoni, 2004; Garrett, 2005).

Lambert's work was pioneering in providing evidence that attitude tests based on the matched-guise technique were reliable in eliciting – in an indirect manner – information about the social evaluation of language use in a speech community. Since the sixties, the matched-guise technique has often been used in the study of language attitudes (see Garrett (2005) for an overview). Experimental evidence by Purnell et al. (1999) showed the validity of this technique. The results clearly indicated that speakers were able to recognize segmental cues and that the matched-guise is a valuable tool for language attitude studies.

Importantly, the matched-guise technique constitutes an *indirect* way of testing languages attitudes. It is indirect in that the attitudes towards a language or a variety are not tested directly, but through the attribution of characteristics to a speaker of that language or that variety. It is assumed in attitude research that the indirect approach triggers 'deeper' perceptions than when participants are asked directly, where there is a risk to collect conscious stereotypes rather than actual attitudes. Kristiansen (2009) claimed that there is a clear difference between direct and indirect techniques, because 'the first return attitudes, which are consciously offered while the second return attitudes subconsciously offered'. However, Grondelaers and Van Hout (2010) showed in their study that the use of direct or indirect scales did not affect the nature of the investigated attitudes, countering Kristiansen (2009)'s suggestion that the resulting perceptions correlates with the respondents awareness of the attitudinal object.

Traditionally the matched-guise technique has been used in the study of language attitude towards 'global stimuli' such as languages, dialects or other kinds of 'lects' (Preston, 2013). This practice has not commonly been used to assess attitudes towards a single linguistic variable. Only a few studies have attempted to apply this method to the investigation of specific (morpho-)syntactic (e.g., Corvalan, 1984; Grondelaers & van Hout, 2013) or phonological (e.g., Campbell-Kibler, 2007; Díaz-Campos & Killam, 2012) features. Corvalan (1984) examined the use of verbal forms in Spanish conditional clauses. The matched-guise recordings consisted of one short anecdote with the conditional variant and another with the imperfect subjunctive variant. The results revealed that the imperfect subjunctive variant was rated higher on marks for education and prestige. Grondelaers and van Hout (2013) examined the use of the object pronoun 'hun' in subject positions in Dutch through a matched-guise experiment. They showed that this variable was associated with positive attitudes on scales of dynamism, and that it was preferred by male listeners. Campbell-Kibler (2007) studied listeners' perceptions towards the variable (ing) in American English. Using the speech of speakers from the South and the West of the United States, she created stimuli where the only differing feature was the realization of (ing). The results showed that variants of (ing) were associated with categories such as education, articulateness, formality, and urban origin. Díaz-Campos and Killam (2012)'s paper contributed to the study of attitudes towards consonantal deletion in Venezualan Spanish. The results of their matched-guise task supported the hypothesis that the retention of syllable-final consonants was perceived as a prestigious variant. In conclusion, these studies examined the evaluation of speech based on allosyntactic or allophonic differences as opposed to a holistic speech evaluation and provided more evidence on the sociolinguistic meaning of phonological/syntactic variation phenomena.

In our study, we applied the matched-guise technique for the investigation of two specific phonological variables: the devoicing of labiodental fricatives and bilabial stops. The attitudes towards these two changing variables were tested within one single experiment, but the results related to fricatives and stops are discussed separately.

6.2.2 Stimuli

The speakers Twenty-eight young adult native speakers of Dutch (aged between 21 and 36 years) from different regions of the Dutch language area were recruited as speakers for the series of stimuli. All speakers were linguists, students in linguistics or language teachers. In Appendix J, they are listed with their regional origin, age and gender.

Speakers were balanced across *sex* and across four *regions*: three regions within the Netherlands (North, Middle and South) and one in Flanders (see Table 7 in Appendix J). The speakers' regions roughly corresponded to our listeners' regions. Indeed, the three Dutch regions corresponded to the participants' regions Groningen, South-Holland and Limburg. In Flanders, only one region is distinguished here, whereas listeners come from two different Flemish regions (Flemish-Brabant and West-Flanders)[1]. As noted by Clopper, Conrey, and Pisoni (2005), perceptual studies often tend to use only male speakers (Clopper et al., 2005), but by adding both *sex* and *region* as speaker-related factors, we succeeded in adequately mirroring the existing variation.

Recordings and manipulation The speech materials were recorded with a sample frequency of 44.1 kHz in a sound-attenuated cabin of the phonetic lab of Utrecht University. The 28 speakers were given some time to read the text described below carefully and were instructed of the purpose of the recordings. They were recorded producing the text in a quasi-spontaneous speech manner (i.e., with intonation, pausing on the natural places in the sentences and etc.). The speakers recorded the target text in three guises:

- *A baseline version:* participants' own spontaneous way of pronouncing the text
- *An extra voiced version:* a version comparable to the baseline version (same intonation and speech rate), but with a fully voiced pronunciation of the three initial voiced labiodental fricatives or of the three initial voiced bilabial stops
- *A devoiced version:* a version comparable to the baseline version (same intonation and speech rate), but with a devoiced pronunciation of the three initial voiced labiodental fricatives or of the three initial voiced bilabial stops

[1]This was due to the difficult of finding Flemish linguists who were able to make recordings at the phonetics lab in Utrecht (The Netherlands).

The recording in these three guises was repeated several times until each guise was pronounced successfully and in a natural way.

Subsequently, either the labiodental fricatives or the bilabial stops of the baseline versions were phonetically manipulated in order to obtain − for each variable − two experimental conditions: a *fully voiced condition* and a *devoiced condition*. The manipulation consisted of replacing segments. Either the baseline version or the fully voiced version was used as the first condition. The second condition was obtained by replacing segments from the first condition by devoiced segments taken from the devoiced recording. When manipulating segments, the whole C(C)V syllables were replaced in order to avoid unnatural coarticulation effects. In this way, it was controlled for many extraneous variables as possible, both linguistic and extralinguistic, so that the single phonemes in question (either fricatives or stops) were the only influencing factors (Díaz-Campos & Killam, 2012).

The text The excerpt contained three word initial realizations of each variable (underlined). The realizations of the variables were placed in content words in non-assimilative contexts in order to avoid this effect of voicing assimilation (Zonneveld, 1983):

> Het is nog heel vroeg. Ik heb slecht geslapen. Dat gebeurt nogal vaak de laatste tijd. Ik drink een beetje koffie en ik eet een stukje brood. Zo voel ik me beter om de dag te starten.
>
> [*It is still very early. I slept badly last night. That happens very often lately. I drink some coffee and I eat a piece of bread. Now I feel better to start the day.*]

6.2.3 Procedure

Participants were seated in a sound-attenuated booth and listened to the stimuli via a headphone Beyerdynamic DT 250 (for more details, see Section 3.5).

Participants were instructed that they would hear 44 different speakers from different regions of Flanders and the Netherlands. Their task was to rate the speakers along seven personality scales. As can be seen from Table 6.1, the 44 trials (2 exercise trials + 32 test trials + 10 fillers) actually came from the 28 different speakers (2 exercise speaker + 16 test speakers + 10 filler speakers), but the speech of each test speaker was presented in two conditions (voiced vs. devoiced), resulting in 32 test trials. In Table 6.1, we see that the 32 test trials consisted of eight trials in each condition (voiced vs. devoiced) and for each variable (fricatives vs. stops).

Two strategies were used to increase the likelihood that participants would not realize that some voices were recurrent in the experiment. First, the order of test trials was randomized and they were surrounded by filler items. Second, participants were told that the researcher had difficulties recording so many different speakers, because they all had to come to the lab in the week before the experiment.

Table 6.1: Number of trials and speakers in the attitude experiment, split up by experiment part and by condition (n=44 trials, N=28 speakers).

Part	Trials		Speakers
Exercise	2		2
Voicing conditions:	voiced	devoiced	
Fricatives	8	8	8
Stops	8	8	8
Fillers	10		10
Total	44		28

Participants were asked to rate all speakers on a 7-point interval scale along the seven character traits presented in Table 6.2.

Table 6.2: Schematic representations of the scales used in the matched-guise experiment.

Ik vind dat deze spreker is... [*In my opinion, this speaker is*]									
Modernity	ouderwets *old-fashioned*	*	*	*	*	*	*	*	modern *modern*
Power	dom *stupid*	*	*	*	*	*	*	*	intelligent *intelligent*
	arm *poor*	*	*	*	*	*	*	*	rijk *rich*
	onzeker *unsure*	*	*	*	*	*	*	*	zelfverzekerd *self-confident*
Solidarity	lelijk *ugly*	*	*	*	*	*	*	*	mooi *beautiful*
	onvriendelijk *unfriendly*	*	*	*	*	*	*	*	vriendelijk *friendly*
	koud *cold*	*	*	*	*	*	*	*	warm *warm*

The scales were presented in the form of semantic differentials with the positive differential always on the right side of the screen. Based on previously attitude studies (e.g., Grondelaers & van Hout, 2013; Grondelaers et al., 2010), three scales were chosen to represent the *power* dimension (*intelligent, rich and self-confident*), three other scales represented the *solidarity* dimension (*mooi, vriendelijk, warm*), and the last scale was related to *modernity* (*modern*). The scales associated with the power and solidarity dimensions are commonly used in matched-guise designs. However, the modernity dimension is an innovation of this study and is aimed at grasping attitudes specifically related to a sound change in progress. The order of the scales was randomized between participants, but fixed within participants. Participants had the possibility to listen more than once to a stimulus if they desired, but they could not go back to previously heard stimuli, once the rating was completed.

6.3 Results

We present the analysis of 32000 ratings on every scale (N=100 participants, n=32 trials). Since the matched-guise design was double in the sense that both fricatives and stops were manipulated within the same experiment, 16000 of these ratings reflected attitudes towards the devoicing of fricatives, and 16000 ratings reflected attitudes towards the devoicing of stops. Hence, results on fricatives and stops are discussed separately. The results of the exercise and filler trials were not analyzed. There were no missing value.

Each scale was tested for an effect of phonetic manipulation in ANOVA's. There was no effect of phonetic manipulation in the test items, so that there was no significant difference in ratings between the items that have been phonetically manipulated and the items without manipulation. In this way, it appeared that raters were not reacting to the unnaturalness of the speech rather than to the manipulated parameters.

6.3.1 Reducing scales to three attitudinal dimensions

In order to reduce the seven scales to a smaller number of attitudinal dimensions, Principle Component Analyses (PCA) as implemented in the `psych` library in R (R Core Team, 2014) were carried out. Analyses were run separately for fricative ratings and stop ratings.

Fricatives The varimax rotation computed on the fricative ratings and the factor selection eigenvalue yielded a three-component solution, which accounted for 67 % of the variance. The factor loadings and the eigen values of the PCA solution is presented in Table 6.3.

Table 6.3: Factor loadings and the eigen values of the Principal Component Analysis of ratings for fricatives (n=1600, 7 scales). Loadings higher than .40 are shaded in grey.

	Component 1 *Modernity*	Component 2 *Power*	Component 3 *Solidarity*
modern	.93	.14	.08
intelligent	.09	.73	.24
rich	.25	.76	.00
self-confident	.06	.77	.11
beautiful	.52	.31	.47
friendly	.09	.06	.86
warm	.10	.17	.84
Eigen value	1.85	1.75	1.23

Stops The varimax rotation computed on the stop ratings and the factor selection eigenvalue yielded a three-component solution, which accounted for 68 % of the variance. The factor loadings and the eigen values of the PCA solution is presented in Table 6.4.

Table 6.4: Factor loadings and eigen values of the Principal Component Analysis of ratings for stops (n=1600, 7 scales). Loadings higher than .40 are shaded in grey.

	Component 1 *Modernity*	Component 2 *Power*	Component 3 *Solidarity*
modern	.93	.19	.10
intelligent	-.01	.79	.24
rich	.26	.74	-.02
self-confident	.20	.71	.11
beautiful	.46	.47	.38
friendly	.08	.04	.86
warm	.12	.20	.81
Eigen value	1.97	1.61	1.20

PCA solutions for fricatives and stops turned out to be highly similar. In both cases, a three-component solution yielded the best results. First, the scales *modern* and to a smaller extent the scales *beautiful* loaded on a first component. Second, the three status-related scales *intelligent*, *rich* and *self-confident* consistently loaded on a second component. Finally, the three scales related to social attractiveness *beautiful*, *friendly* and *warm* loaded on a third component. The loadings of the scale *beautiful* on the third component turned out to be weaker than the loading of the two other scales, especially in the stops data. In view of the high loadings on the three classical attitudinal dimensions, the three PCA components were labeled *modernity*, *power*, and *solidarity*.

The similarity between the solutions for fricatives and stops can to some extent be explained by the facts that the variables tested are very similar in nature and that they were included within the same experiment. Both analyses also resulted in the highlighting of the three same attitudinal dimensions. Hence, the importance of these three dimensions in the study of language attitudes, which was also reported in previous studies (e.g., Grondelaers & van Hout, 2013; Grondelaers et al., 2010), is confirmed.

The scale *beautiful* appeared to load on the three components equally well. PCA solutions with four components instead of three did not help isolating the scale *beautiful* as a separate dimension. As explained in Latour, Grondelaers, and van Hout (2012), this scale is often problematic in attitude research. In Grondelaers and Van Hout (2010) and Grondelaers et al. (2010) for instance, it was left aside, because it loaded on all components of the PCA and could therefore not be associated with one component. For statisticians, the fact that a scale is ambivalent and loads on different components is a good reason to leave it aside. However, we followed Latour et al. (2012, p.247) who spoke in favor of keeping the scale *beautiful* just for the reason that it is a crucial scale in attitudinal research. It brings the different components together and fulfills a central position in the social embedding of a linguistic feature.

After the identification of the underlying components, factors scores were computed in order to investigate the research questions of interest. Factors scores represent each individual's placement on the identified components (Di Stefano, Zhu, & Mindrila, 2009). Factor scores are standardized to a mean of zero. If a participant obtains a positive factor score on a dimension, this means that he evaluated the speech more positively than the mean ratings. If his factor score is negative, his ratings were more negative than the mean.

Moreover, factor scores were aggregated across speakers. As explained in Section 6.2.2, eight different speakers were used for each variable, and these speakers were balanced across gender and region. Including these two speaker-related factors in the design allowed us to adequately incorporate enough variation to the task. An analysis of the role of these speaker-related factors however goes beyond the scope of this study. In order to concentrate on the listener-related effects, which constitute the focus of this study, factor scores were therefore aggregated across the eight speakers of each variable.

The aggregated factor scores for each dimension are presented in Figure 6.1. In this barplot, participants' factor scores (N=100) are split up by dimension (*modernity*, *power*, and *solidarity*) and by condition (*voiced* vs. *devoiced*).

This figure gives us the first insights into the difference in attitudes between conditions. For the fricative ratings (on the left), it turned out that the devoiced condition systematically yielded higher ratings than the voiced condition. The difference between the conditions is the biggest along the solidarity dimension. For the stop ratings (on the right), ratings along the power and solidarity dimensions did not really differ across conditions. Along the modernity dimension in contrast, the devoiced condition yielded higher ratings than the voiced condition.

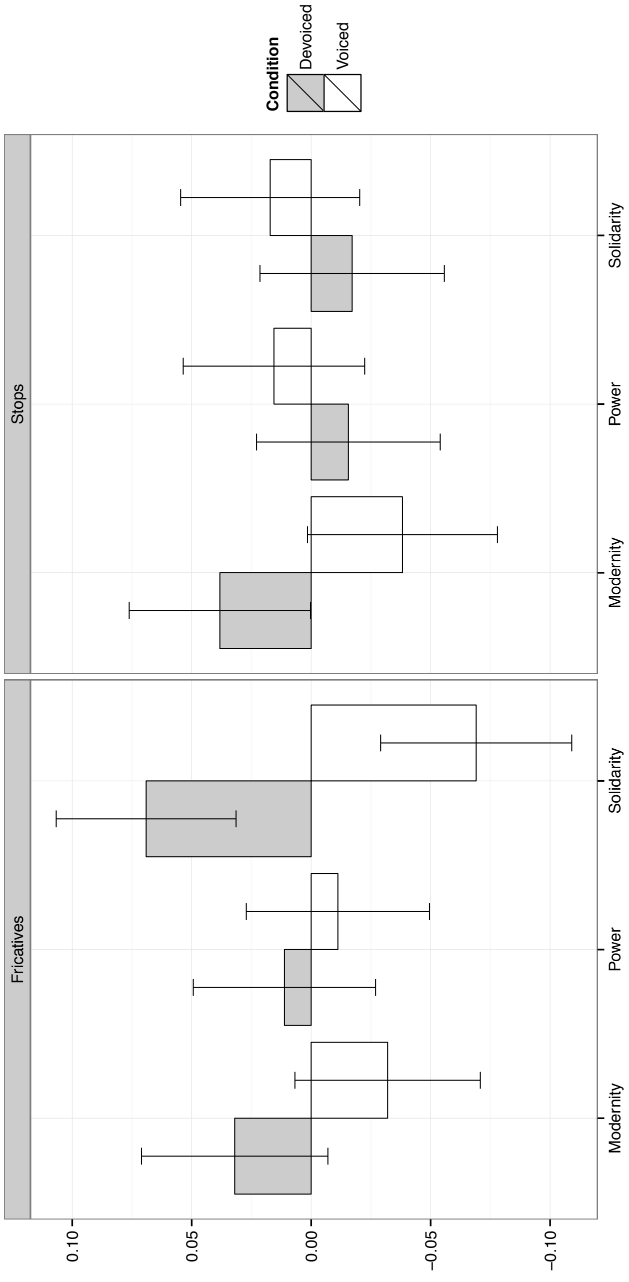


Figure 6.1: Barplot of factor scores for fricatives (on the left) and stops (on the right) along the three dimensions split up by condition (N=100). Error bars represent ±1 standard error.

6.3.2 Effects between regions

In order to investigate regional differences, the aggregated ratings were analyzed with Mixed Effects Linear regressions (LMM, Baayen et al. (2008)) as implemented in the `lme4` library (Bates et al., 2012) in R (R Core Team, 2014). A model was fitted with the ratings of each variable (fricatives and stops) and each dimension (modernity, power and solidarity) separately. In all models, the factor participants (N=100), testing for individual differences between participants was added as a random effect, and the factors *region* and *condition* and the interaction between *region* and *condition* were added as fixed effects[2].

The full regression models are presented in Appendix K. A summary of the significant fixed factors is presented in Table 6.5.

Table 6.5: Significance of fixed factors in the mixed models regressions.

	Fricatives			Stops		
	Modern.	*Power*	*Solid.*	*Modern.*	*Power*	*Solid.*
Voicing cond.			*			
Region			*			
Cond. * Region	*	*			*	
In Figure	6.2a	6.2b	6.2c		6.2d	

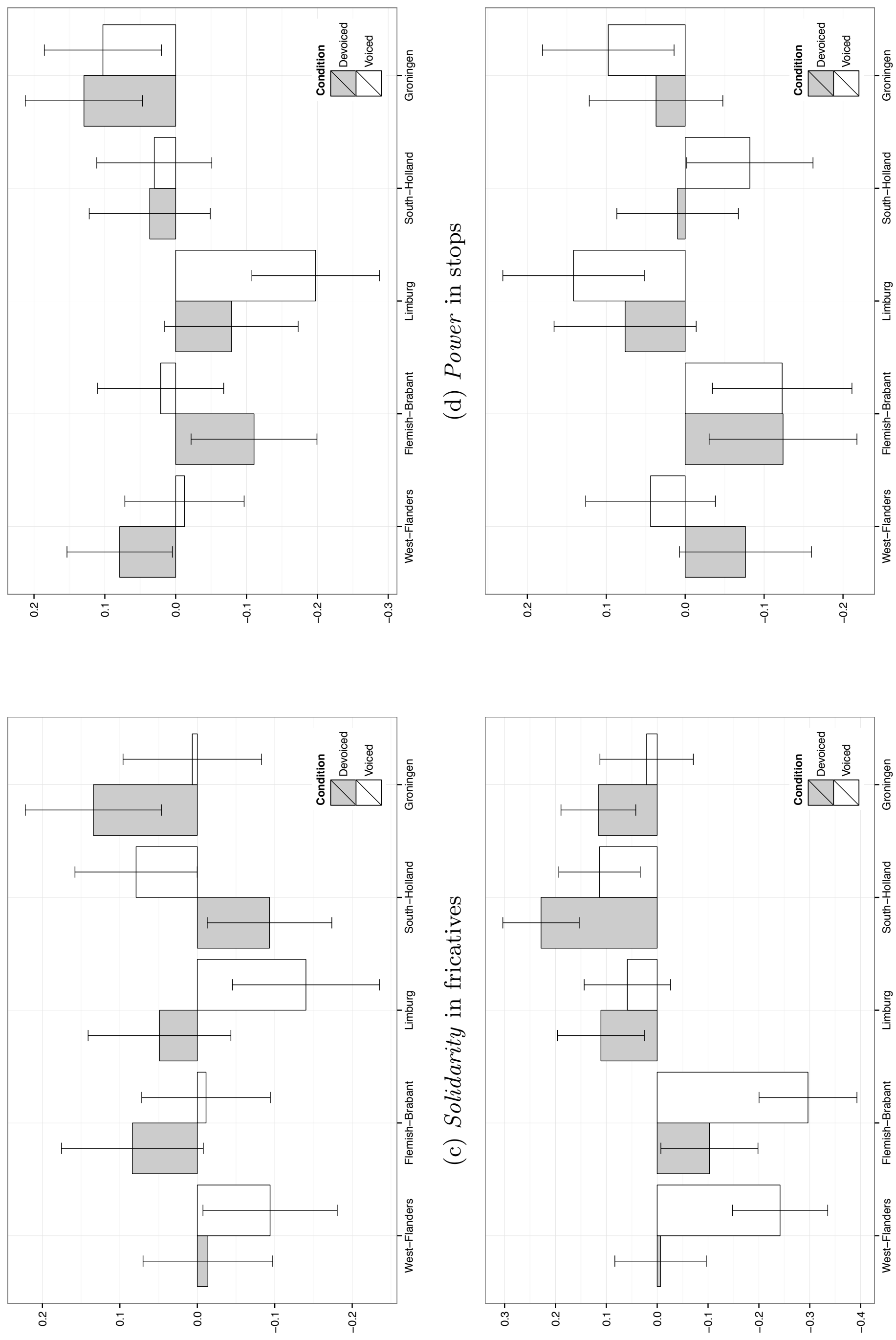
Note. Degrees of freedom (*df*) required for statistical significance testing of t values was given by *df* = *J*-*m*-1 (Hox, 2010), where J is the most conservative number of second-level units (J = 100 speakers) and m is the total number of explanatory variables in the model (m = 3) resulting in *df* = 94.

Barplots of the dimensions which yielded at least one significant effect are plotted in Figure 6.2. The factor scores presented along the y-axes are split up by region and by condition.

Fricatives The model on fricative factor scores for the modernity dimension revealed a significant interaction between condition and region. This is visualized in Figure 6.2a. The interaction is triggered by participants from South-Holland who behave differently from all other region and rated the voiced condition as more modern than the devoiced condition. In all the other regions, ratings in the devoiced condition did not significantly differ from the ratings in the voiced condition.

The model on fricative factor scores for the power dimension revealed that there was a significant interaction between condition and region. This interaction can be visualized in Figure 6.2b. The interaction appears to be triggered by the region Flemish-Brabant where the voiced condition was rated higher on the power scales than the devoiced condition.

[2]The other participant-related factor *sex* did not turn out significant and was therefore excluded from the model.



(a) *Modernity* in fricatives

(b) *Power* in fricatives

(c) *Solidarity* in fricatives

(d) *Power* in stops

Figure 6.2: Barplots of factor scores split up by region and by condition (N=100). Error bars represent ±1 standard error.

In the other regions, the devoiced condition was rated either the same or slightly higher along the power dimension than the voiced condition.

The model on fricative factor scores for the solidarity dimension revealed effects of condition and region, but no interaction between condition and region. In Figure 6.2c, it appeared that in all regions the devoiced condition was rated higher along the solidarity dimension than the voiced condition. Moreover, the effect of region showed that the ratings of solidarity were overall lower in the Flemish regions, regardless of the condition, than in the three Dutch regions. This is probably due to the fact that a majority of speakers in the task were Dutch. Flemish listeners have the tendency to rate these speakers more negatively on the solidarity dimension, since their speech is clearly marked as Dutch.

Stops The model on stop factor scores for the power dimension revealed that there was a significant interaction between condition and region. In Figure 6.2d, this interaction can be visualized. It appeared that in three regions (West-Flanders, Limburg, Groningen) the voiced condition was rated higher along the power dimension than the devoiced condition. In Flemish-Brabant, there was no difference between conditions. In South-Holland however, it was the devoiced condition which was rated higher than the voiced condition. The other dimensions did no yield any significant effect.

6.3.3 Effects within regions

In a second step of the analysis, language attitudes are investigated at the individual level. Therefore, difference scores between the devoiced and the voiced condition were computed for each participant along the three dimensions, giving insight in the size of the effect at the individual level. The difference scores were calculated with the following formula:

$$\text{difference score} = -\,(\text{voiced condition} - \text{devoiced condition}) \tag{6.1}$$

If the difference score has a positive sign, it means that the devoiced condition was rated higher than the voiced condition. If the difference score is negative, it is the other way around: the voiced condition was rated higher than the devoiced. Each participant was thus associated with six difference scores: one for each of the three dimensions for both fricative and stop ratings. In this section, individual differences related to these difference scores are investigated for both fricatives and stops.

Fricatives There appeared to be no significant correlation in difference scores between the three dimensions in fricative ratings. This means that voiced and devoiced fricatives were rated quite differently depending on the dimension.

For fricative ratings, the number of participants who showed a positive difference score was computed. These are individuals who rated the devoiced condition higher than the voiced condition and thus who associated positive attitudes with the devoicing of fricatives. In the upper part of Table 6.6, it is observed that the majority of participants turned out to have a difference score higher than 0 in all three dimensions. Moreover, in the lower part of Table 6.6, a total of 17 participants showed a positive difference score along all three dimensions. Most of these individuals were West-Flemish.

Table 6.6: Number of participants with a preference for the devoiced condition (in the upper part), split up by region (in the middle part) and number of participants with a preference for the devoiced condition along the three dimensions (in the lower part) (N=100).

Number of participants with a preference for the devoiced condition						
		Fricatives			Stops	
Modernity		57/100			54/100	
Power		55/100			48/100	
Solidarity		68/100			48/100	
Number of participants with a preference for the devoiced condition split up by region						
		Fricatives			Stops	
	Mod.	Power	Solid.	Mod.	Power	Solid.
West-Flanders	12/20	15/20	17/20	10/20	7/20	12/20
Flemish-Brabant	13/20	5/20	14/20	9/20	11/20	8/20
Limburg	14/20	14/20	11/20	11/20	8/20	9/20
South-Holland	6/20	10/20	13/20	10/20	11/20	7/20
Groningen	12/20	11/20	13/20	14/20	11/20	12/20
Total	57/100	55/100	68/100	54/100	48/100	48/100
Number of participants with a preference for the devoiced condition along the three dimensions, split up by region						
		Fricatives			Stops	
West-Flanders		7/20			2/20	
Flemish-Brabant		1/20			1/20	
Limburg		4/20			0/20	
South-Holland		3/20			1/20	
Groningen		2/20			3/20	
Total		17/100			7/100	

Stops For stops, there were significant correlations in difference scores between the power dimension on the one hand, and the dimensions modernity and solidarity on the other hand (r=-.22 and r=-.25, respectively). The bigger the difference score between conditions in stops along the power dimension, the smaller the difference score along the modernity dimension and the smaller the difference score along the solidarity dimension. Since these correlations were negative, it seemed that the dimension power in stops behaved differently than the two other (solidarity and modernity).

The number of participants who showed a positive difference score for stops was computed along the three dimensions. In the upper part of Table 6.6, we observed that approximately half of the participants rated the devoiced condition higher than the voiced condition along each dimension. Moreover, in the lower part of Table 6.6, a total of 7 participants showed a positive difference score along all three dimensions.

Finally, there was no significant correlation in the difference scores between fricatives and stops, showing no relationship between fricatives and stops ratings, even if there were collected within the same experimental design.

6.4 Discussion and conclusion

All participants took part in a matched-guise experiment aimed at triggering ratings towards the devoicing of labiodental fricatives and bilabial stops. This experiment sought to investigate language attitudes towards these variables by phonetically manipulating them in the experimental design. This adapted version of the matched-guise experiment in which solely one single linguistic variable is manipulated (instead of a variety or a language) has not often been used, but proved to efficiently contribute to the study of language attitudes.

In the experimental design, the speech of 16 speakers was phonetically manipulated in order to obtain for each of them a voiced and a devoiced 'guise' based on the same baseline recording. Consequently, if participants rated the two conditions differently, this difference had to be explained by the manipulation of the variable, since all the other factors (e.g., gender, voice, content, intonation) were held constant. Listeners were aware of the fact that it was an attitude-rating task (a minority of participants even realized that some voices were recurrent), but crucially the approach was *indirect*, because participants rated speakers instead of linguistic entities (e.g., accents, dialects, etc.).

The Principal Component Analyses ran on ratings towards fricative and stop devoicing turned out to be highly similar, and yielded solutions with three attitudinal dimensions: *modernity*, *power* and *solidarity*. This confirms the importance of these three dimensions for the study of language attitudes, as reported in previous studies (e.g., Grondelaers & van Hout, 2013; Grondelaers et al., 2010).

The experiment examined whether the changing variables were associated with positive language attitudes. If so, it was investigated whether these attitudes differ across regions (at the *group level*) and across listeners (at the *individual level*).

For the devoicing of fricatives, there was clear evidence that the devoicing of labiodental fricatives is considered to be the prestigious variant, especially on the solidarity dimension. At the *group level*, some regional differences were revealed in the ratings of voiced and devoiced fricatives. Especially participants from West-Flanders (i.e., the region where the fricative change is the least advanced) showed positive attitudes towards the devoiced condition. This goes again expectations, because West-Flanders is the region with the least fricative devoicing which showed more positive attitudes, and not regions where the devoicing phenomenon is the most advanced. It thus seems that participants coming from regions where the change is not so advanced are more inclined to associate the new form with positive attitudes. In that sense, having positive attitudes towards a variable seems to be a sign for potential change towards this variable. At the *individual level*, it was shown that the majority of participants rated devoiced fricatives higher than voiced fricatives along one or more dimensions, especially West-Flanders again.

For the devoicing of stops, the evidence for positive attitudes towards this incipient change was less clear, as it was expected. Results of the mixed-models regressions only revealed one interaction between regions and conditions explained by the fact that devoiced stops were rated more positively in South-Holland along the power dimension. As shown in Chapters 4 and 5, bilabial stops did not show clear regional variation in production and perception, it is therefore still unclear why participants from South-Holland rated devoiced stops more positively in this task. In the other regions, it was rather the voiced stop condition, which triggered positive attitudes. Also at the *individual level*, we showed that half of the participants rated devoiced stops higher than voiced stops, which means that the other half in fact still preferred the voiced condition. It thus seemed that participants did not show preferences when rating stops.

In conclusion, it was shown that the devoicing of fricatives is indeed associated with positive language attitudes along different dimensions. Moreover, it appeared that participants coming from regions where the change is not so advanced were more inclined to associated the new form with positive attitudes. In contrast, more evidence is needed to claim that the devoicing of stops is also socially indexed. Since the devoicing of stops is an incipient change, it is possible that the change is not yet far away, so that participants would have encoded a social index together with the phonetic information.

As explained by Thomas (2002), it is crucial to confront the perception results with the results obtained from the production studies. The relationship between variation in perception, variation in production and the described language attitudes will therefore be investigated in Chapter 8.

CHAPTER 7

Phonetic imitation

In Chapter 2, the phenomenon of phonetic imitation was described. In this chapter, imitation capacities are investigated through two separate imitation experiments: a forced imitation task and a spontaneous imitation task. The main goals of these experiments is to test at the *group level* and at the *individual level* whether participants are capable of imitating the devoicing of labiodental fricatives and bilabial stops, and to investigate what role imitation capacities play in sound change.

7.1 Introduction and hypotheses

The introduction of this chapter is articulated around three themes: the selective character of the imitation process described in Section 7.1.1, the imitation paradigm discussed in Section 7.1.2 and the difference between forced versus spontaneous imitation in Section 7.1.3.

7.1.1 Imitation is selective

The phenomenon of phonetic imitation was described in Chapter 2 as crucial to speech communication. Nevertheless, it is not an entirely automatic or unrestricted process (Dijksterhuis & Bargh, 2001; Gentilucci & Bernardis, 2007), but rather a selective process which can be modulated by a range linguistic and social factors.

The selective nature of phonetic imitation was recently showed in terms of the limitation to certain linguistic features. Nielsen (2007) for instance used the imitation paradigm (see Section 7.1.2) to investigate the imitation of VOT. He found that American English participants would imitate lengthened VOT's but not shortened ones. Babel (2009) investigated phonetic imitation of English vowels and found that certain vowels were imitated, but not others. In the same way, Kraljic et al. (2008) showed that English-speaking participants did not imitate alveolar fricatives.

Imitation capacities also turned out to be constrained by a range of social factors. First, phonetic imitation was shown to be varying in function of gender of participants (Namy et al., 2002; Pardo, 2006; Yu et al., 2011, e.g.,). The results are however conflicting, regarding which gender is more likely to imitate. Pardo (2006) and Yu et al. (2011) for instance, found that men were more likely to imitate in a map task than women, yet Namy et al. (2002) found female participants converged more than male participants in a shadowing experiment.

Furthermore, imitation is constrained by individual's evaluative judgment towards the interlocutor. Different studies suggested that the judgment towards the interlocutor plays a significant role in affecting the likelihood and the directionality of phonetic accommodation (Babel, 2012; Babel et al., 2014; Yu et al., 2011). Phonetic convergence is not guaranteed with all model speakers. Phonetic divergence is likely when the speaker is negatively disposed towards the interlocutor. These data align with predictions from the Communication Accommodation Theory (CAT), which suggests that when a talker is viewed positively, there is a desire to decrease social distance, which can be accomplished by imitating.

Finally, not all participants in an imitation task show convergence to the same extent. As noted by Babel and Bulatov (2012) in the context of a shadowing task, large individual differences appeared in the degree of pitch imitation. Nielsen (2011), Miller, Sanchez, and Rosenblum (2010) and Pardo (2006) also found large inter-participant variability in the patterns of imitation.

The first research question raised in the frame of this study is whether we are able to replicate phonetic imitation with the variables under consideration: the devoicing of labiodental fricatives and bilabial stops. Intuitively, it seems to be the case that some speakers are more inclined to imitate, or better at imitating than others. Therefore, the second goal of these experiments is to test for regional and individual differences in imitation capacities.

7.1.2 The imitation paradigm

Phonetic imitation is typically investigated within a paradigm where participants' speech is compared before and after they are exposed to a target speech pattern.

Concretely, participants' baseline productions are first recorded. Next, participants are presented with auditory input from a model talker and asked to repeat the words heard. Subsequently, baseline and repeated productions are compared with each other. An increased similarity between repeated productions and productions of the model talker is used as evidence for phonetic imitation, also called *convergence* in this chapter.

Within this paradigm, imitation can be triggered in the frame of a shadowing task or a non-shadowing task. In the shadowing version which was first used by Goldinger (1998) and which is by far the most commonly used, participants listen to the production of the model speakers and repeat them immediately after. In a non-shadowing version in contrast, participants first listen to the target list entirely, and then read aloud the list of stimulus items once again.

In most imitation studies (Goldinger, 1998; Namy et al., 2002; Pardo, 2006, e.g.), a perceptual method is used subsequently in order to quantify phonetic imitation. The degree of imitation is measured through an AXB discrimination task. In this task, listeners determine whether a talker's repeated or baseline production was more similar to the model production for a given word. If repeated tokens are selected as being more similar to the model than baseline tokens, it is inferred that the talker has imitated the model. If more baseline tokens are selected, the talker has not imitated or has diverged.

The disadvantage of the AXB method is, however, that it is unclear which aspect of the stimulus is imitated (Mitterer & Müsseler, 2013). Like Nielsen (2011) and Fowler, Brown, Sabadini, and Weihing (2003), we argue that acoustic measurements of one or more phonetic dimensions provide a more objective and precise measure of the imitation as well as a detailed, quantitative information about the variability among speakers. For these reasons, the degree of imitation will be assessed in this study by means of acoustic measurements.

7.1.3 Forced versus spontaneous imitation

In the above-described imitation paradigm, individuals' convergence patterns are studied while these individuals are at no point instructed to imitate. This type of imitation capacity is an automatic process of which speakers are unaware, and is therefore often called *spontaneaous phonetic imitation*. This contrasts with the fact that speakers are also capable of voluntary imitation as already reported by Chistovich, Fant, de Serpa-Leitao, and Tjernlund (1966).

To the best of our knowledge, only one study (Dufour & Nguyen, 2013) has yet examined the difference between *forced* vs. *spontaneous imitation* capacities. They compared the phonetic convergence effect observed in – on the one hand – a shadowing task which was meant to trigger unintentional attuning to ambient speech, and on the other hand – a forced imitation task in which participants were explicitly instructed to imitate the productions they were exposed to. They found that – although the phonetic convergence effect was greater when participants intentionally imitated the speaker's productions – shadowing and imitation instructions led to the same degree of convergence in a post-exposure task. Dufour and Nguyen (2013) explained the difference in terms of attentional factors.

In this study, these insights were combined into two imitation tasks: a *forced imitation* task reported in Section 7.2 and a *spontaneous imitation* task reported in Section 7.3. In the forced imitation task, participants were explicitly instructed to imitate, thus to repeat the stimuli in exactly the same way as it has been pronounced, whereas in the spontaneous imitation task, participants were at no point aware that they are conducting a task meant to trigger imitation.

For the sake of the arguments, the forced imitation task is presented before the spontaneous imitation task in this chapter. In the study however, participants conducted the spontaneous imitation task in the first session and the forced imitation task in the last session, because the latter was inherently metalinguistic and risked to influence all the others tasks if it was presented before.

7.2 Forced phonetic imitation

Following Dufour and Nguyen (2013), a forced imitation experiment was conducted in this study. This experiment aimed at gathering insights in participants' capacities at producing the whole range of voicing in labiodental fricatives and in bilabial stops. Since these variables are devoicing, the question raised is whether speakers are still able to produce fully voiced variants when they are 'forced' to. We investigate whether all participants are able to produce the whole ranges (at the *individual level*) or whether there are regional differences in this capacity (at the *group level*).

7.2.1 Method

Stimuli

The same phonetic continua of fricatives and stops were used as in the identification tasks (see Section 5.2.1). However, in order to reduce the length and the cognitive load of the experiment, a subset of 25 sounds was selected from the 81-steps original continua, thus every other step along both nine-steps dimensions (i.e., the same subsets as used for the similarity judgment task in Section 5.3). The other sounds were disregarded for this task. The selection of both subsets resulted in a total of 50 sounds, 25 fricative and 25 stop stimuli.

Procedure

Participants were seated in a sound-attenuated booth and listened to the 50 stimuli (25 fricatives and 25 stops) via a headphone Beyerdynamic DT 250 (more details see Section 3.5). They were instructed 'to imitate the speaker as accurately as possible' after each stimulus and to embed their production in a carrier sentence. The carrier sentence was 'ik neem de *target stimulus*' [*I take the target stimulus*], which has also been used in the carrier sentence production task (see Section 4.2). The sentences were recorded through a C420 head-mounted microphone. Unlike previous imitation experiments, there was no need in our case to record participants' baseline production within the imitation task, since this had been done in the frame of the production tasks reported in Chapter 4. Using this carrier sentence, we were both able to measure the duration in stops since the syllables were not imitated in isolation, and to compare the imitated syllables to the baseline syllables from carrier sentence production task where the same carrier sentence was used.

Each participant completed a practice block of twelve trials. After the practice, participants completed the experimental session with all 50 sounds presented in a randomized order. The task was self-paced.

Phonetic measures of variables

All recordings were sampled at 48kHz (24 bits). All realizations of the target variables were segmented and labeled into phonetic segments within `Praat` speech-analysis software package (Boersma & Weenink, 2014). In order to achieve a high degree of consistency and a low error rate, the segmentation was done by hand by the researcher herself, following the same procedure as for the production data (see Section 4.2.2). For fricatives, two phonetic dimensions were measured: VOICING and DURATION. For stops, three phonetic measures were used: VOT, VOICING and DURATION (see discussion in Section 4.2.2).

7.2.2 Results

As for production data, observations greater than four standard deviations from the mean of the measurements were removed (see arguments in Section 4.2.2). In total, 2480 observations were collected for fricatives and 2210 for stops. This section is divided into three main analyses: an analysis of general imitation patterns, an analysis of voicing ranges and an analysis of the best individual imitators.

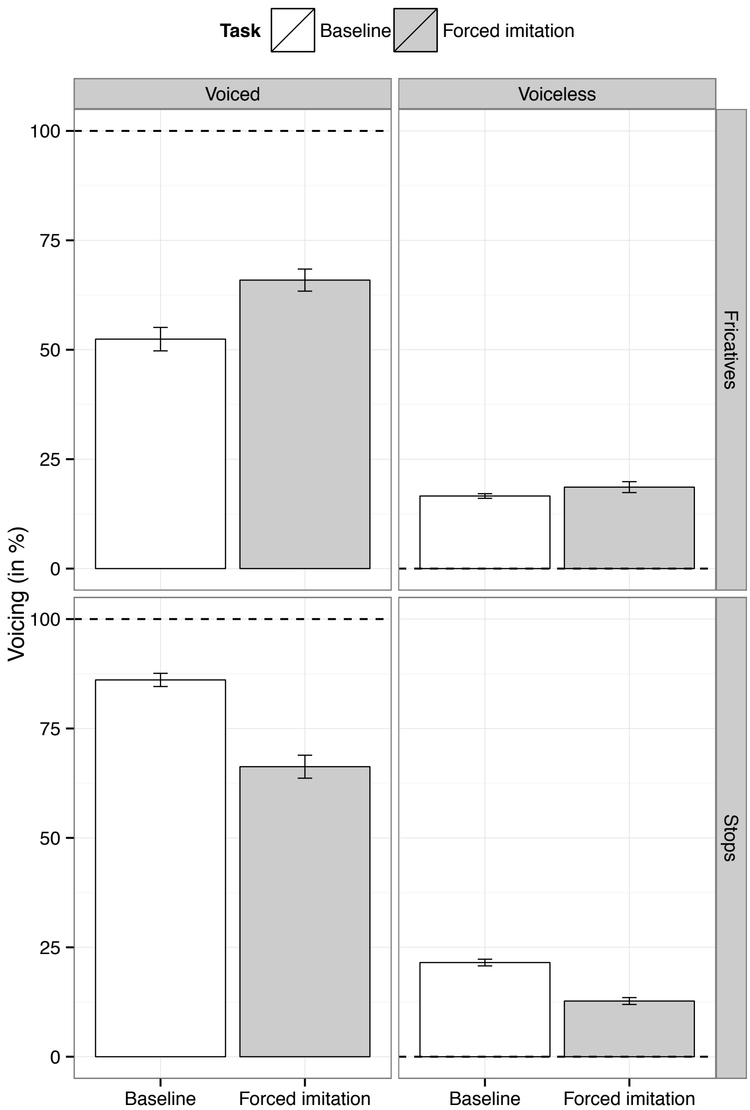
Baseline vs. forced imitation

First, differences in imitation capacities between regions are investigated. In this analysis, only a subset of the data collected in the forced imitation task was selected. In order to analyze voiced fricatives, only the imitation trials triggered by the listening to a fully voiced stimulus (i.e., 100 % voicing) were selected (5 per participant). In the same way, the analysis of voiced fricatives was based on the trials triggered by the listening to a fully voiceless stimulus (i.e., 0 % voicing) (5 per participant). Subsets for voiced and voiceless stops were selected in the same way. Consequently, the remaining 15 imitation trials per participant and per variable were removed for this analysis, leaving 10 trials per participant and per variable available for analysis.

Subsequently, the imitation results of these subsets were compared to the baseline productions obtained from the carrier sentence production task (see Section 4.2). Like imitation productions, baseline productions were divided into voiced and voiceless fricatives, and into voiced and voiceless stops.

The barplot of the results split up by variant and by task is presented in Figure 7.1.

Figure 7.1: Barplot of forced imitation data split up by variant and by task (baseline productions in white and imitation productions in grey). Voiced (left) and voiceless (right) fricatives can be found in the upper panel (N=100) and voiced (left) and voiceless (right) stops in the lower panel (N=100). Error bars represent ±1 standard error. Dotted lines represent the target amount of voicing in the presented stimuli.



In Figure 7.1, the amount of voicing is presented along the y-axes. Voiced (left) and voiceless (right) fricatives can be found in the upper panel and voiced (left) and voiceless (right) stops in the lower panel. Baseline productions are presented in white and imitation production in grey.

These results were analyzed with Mixed Effects Linear regressions (LMM, Baayen et al. (2008)) as implemented in the `lme4` library (Bates et al., 2012) in R (R Core Team, 2014). A model was fitted with voiced and voiceless fricatives data separately, and voiced and voiceless stops separately. In all models, the factor *participant* (N=100), testing for individual differences between participants was added as a random effect, and the factors *region*, *task* (baseline vs. imitation) and the interaction between *region* and *task* were added as fixed effects. The full regression models are presented in Appendix L. A summary of the significant fixed factors is presented in Table 7.1.

Table 7.1: Significance of fixed factors in the mixed models regressions. Asterisks represent significant p values (α =.05).

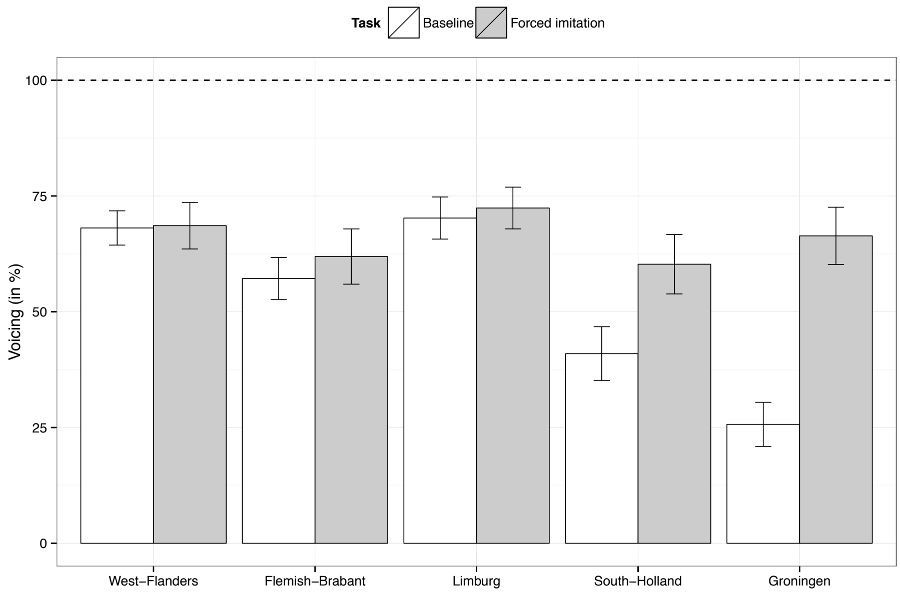
	Fricatives		Stops	
	Voiced	*Voiceless*	*Voiced*	*Voiceless*
Task			*	*
Region	*			*
Task * Region	*			

Note. Degrees of freedom (df) required for statistical significance testing of t values was given by $df = J - m - 1$ (Hox, 2010), where J is the most conservative number of second-level units ($J = 100$ speakers) and m is the total number of explanatory variables in the model ($m = 3$) resulting in $df = 96$.

Fricatives For voiced fricatives (in the upper left part of Figure 7.1), there was a significant effect of region, showing that regions slightly differed in their amount of voicing regardless of the task. Moreover, there was an interaction between task and region. This interaction is plotted in Figure 7.2. In this barplot, the three most conservative regions in the devoicing of fricatives (West-Flanders, Flemish-Brabant and Limburg) showed similar patterns: a high degree of voicing of the voiced fricatives both in the baseline productions (in the carrier sentence reading task) and in the imitation productions. There is thus a ceiling effect in these regions in which participants are able to produce (fully) voiced fricatives in both the baseline and the imitation tasks. The regions of South-Holland and Groningen showed a much lower degree of voicing in the baseline productions. Nevertheless, participants from these regions also achieved a high degree of voicing in the imitation productions. Therefore, it appeared that in all regions they are capable of imitating voiced fricatives in the imitation task, regardless of their baseline productions.

For voiceless fricatives (in the upper right part of Figure 7.1), no effect turned out significant.

Figure 7.2: Barplot of the voiced fricatives imitations split up by region and by task (baseline productions in white and imitation productions in grey) (N=100). Error bars represent ±1 standard error. Dotted lines represent the target amount of voicing in the presented stimuli.



Baseline productions and imitation productions were very similar and did not show any effect of task (mean=16 % vs. 18 % of voicing, respectively). The difference between these results and the results of voiced fricatives again constitutes indirect evidence for the fact that the devoicing of fricatives is a change concerning voiced fricatives, and that voiceless fricatives in contrast show stable patterns (see also Chapters 4 and 5).

Stops For voiced stops (in the lower left part of Figure 7.1), there was a significant effect of task. Imitated productions were more voiceless than baseline productions (mean=86 % vs. 66 % of voicing), showing a divergence towards the fully voiced stimuli targets (represented by the dotted line).

For voiceless stops (in the lower right part of Figure 7.1), there was an effect of region showing that regions slightly differed in their amount of voicing regardless of the task. Moreover, there was a significant effect of task. Imitated productions turned out to be less voiced than baseline productions (mean=21 % vs. 12 % of voicing), showing a convergence towards the fully voiceless stimuli targets (represented by the dotted line).

Ranges of production

This second analysis aims at investigating the range in voicing individual speaker produced in the forced imitation task. In contrast to the previous analysis where subsets of data were selected, the current analysis was run on all data of the forced imitation task.

For each participant, the voicing range between the most voiced and most voiceless realizations was computed. For example, if a participant's most voiced trial yielded 98 % of voicing and his most voiceless trial 8 % of voicing, his voicing range would be 90 %. The individual voicing range therefore varied between 0 % (no range) and 100 % (full range).

The mean results split up by region are presented in Appendix M. The scatterplot of the results are presented in Figure 7.3 for fricatives (N=100) and Figure 7.4 for stops (N=94). Traditional analyses of variance[1] were fitted to the data with the factor *region* as independent variable.

Fricatives Figure 7.3 clearly shows that most participants were able to produce a voicing range in fricatives of 100 % or nearly 100 % in the forced imitation task. In all regions except Limburg, there was however a small proportion of participants who could not achieve maximal voicing. Six participants showed a voicing range around 50 % and four participants showed ranges lower than 25 %.

In conclusion, a great majority of participants was able to produce full ranges when forced. However, around 10 % failed to produce full ranges. Moreover, there was a significant difference in voicing range between regions (ANOVA: df=4;95, F=2.554, p=.044). The post-hoc comparison revealed that participants from South-Holland had significantly lower ranges than participants from Limburg.

Stops Figure 7.4 shows comparable patterns for stops. Most participants were able to produce a voicing range of 100 % or nearly 100 %. In all regions except West-Flanders, there was a small proportion of participants who could not achieve a maximal voicing. This was clearly the case in Flemish-Brabant, South-Holland and Groningen with six participants around 50 % voicing range, and eight participants showing ranges lower than 25 %. There was no significant difference in range between regions.

For both stops and fricatives, it appeared that the few participants who failed to achieve higher ranges in fact failed to produce enough voicing for the voiced variants, but managed to produce voiceless fricatives and stops.

[1]Mixed models were not required for this analysis, since there was only one observation per participant in the aggregated data.

Figure 7.3: Scatterplot of the fricative voicing range split up by region. Each symbol represents a single participant (N=100). The range of voicing in percentage is presented along the y-axis.

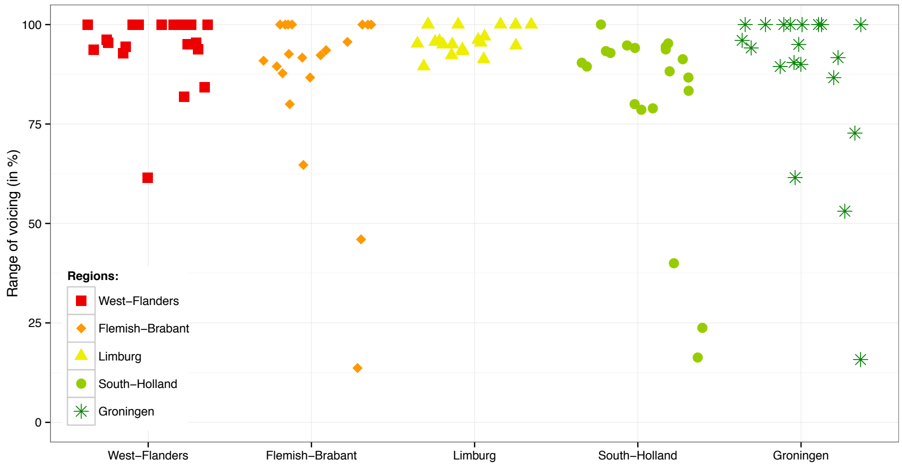
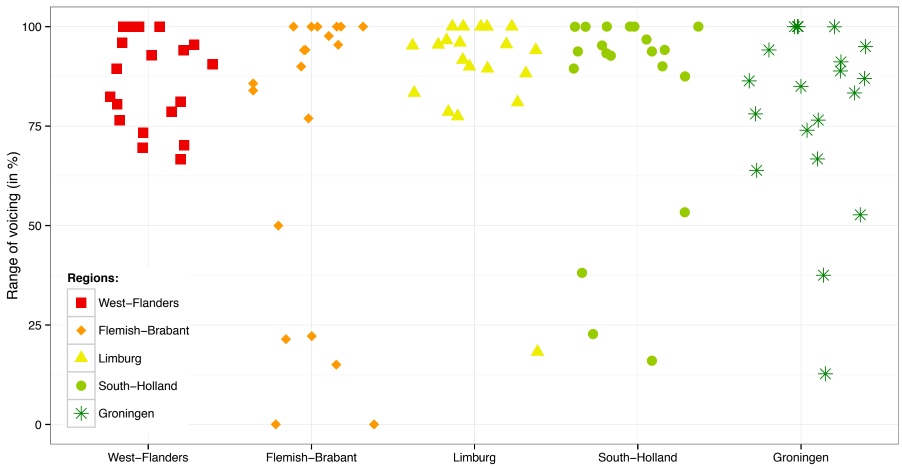


Figure 7.4: Scatterplot of the stop voicing range split up by region. Each symbol represents a single participant (N=94). The range of voicing in percentage is presented along the y-axis.



In order to investigate in more detail who were the participants who showed lower ranges in this imitation task, linear regressions were ran to test relationship between ranges in the imitation task and participant-related factors. Three personality factors (see Appendix C) turned out to significantly predict the range in imitation: *conformism*, *leadership* and *talkativeness*.

The less conformist the participant, the larger his range in imitation (LINEAR REGRESSION t=-2.044, p=.042). The more 'leader-like' the participant, the larger his range (LINEAR REGRESSION t=2.543, p=.012). The more talkative the participant, the larger the range (LINEAR REGRESSION t=1.994, p=.047). Moreover, there appeared to be a relationship between range in imitation and participants' musical abilities. The better the participant's musical abilities, the larger the range (LINEAR REGRESSION t=2.128, p=.034). This finding confirms the results of studies showing that musicians outperform nonmusicians in speech imitation tasks (Christiner & Reiterer, 2013; Pastuszek-Lipińska, 2008). In conclusion, participants with smaller ranges in imitation appeared to be the most conformist, the least 'leader-like', the least talkative and the least musical participants.

Best imitators

In this section, the extent to which individual participants imitated is analyzed quantatively. For each participant, each trial was compared to the target stimuli they were asked to imitate. Subsequently, all trials were aggregated across participant. Two types of imitation score were calculated per participant:

- *The imitation score* (IS) is the aggregated difference between the imitation production and the target production:

$$\text{IS} = \text{voicing}_{imitationproduction} - \text{voicing}_{targetstimulus} \tag{7.1}$$

  A positive IS demonstrates that imitation productions on average were more voiced than the target stimulus, indicating that a participant systematically overshoots the voicing during imitation. A negative IS indicates that imitation productions on average were less voiced than the target stimulus, showing systematic undershoot in voicing during imitation.

- *The absolute imitation score* (AIS) is the absolute value of the aggregated difference between the imitation production and the target production:

$$\text{AIS} = \left|(\text{voicing}_{imitationproduction} - \text{voicing}_{targetstimulus})\right| \tag{7.2}$$

  The AIS ignores the direction of the differences (undershoot vs. overshoot), but concentrates on the relative quality of imitation. It ranges from 0 (best imitation) to 100 (worst imitation).

The results of the imitation scores and of the absolute imitation scores split up by region can be found in Figure 7.5 and Figure 7.6, respectively.

Figure 7.5: Barplot of imitation scores (IS) split up by variable (fricatives on the left and stops on the right) and by region (N=100). Error bars represent ±1 standard error.

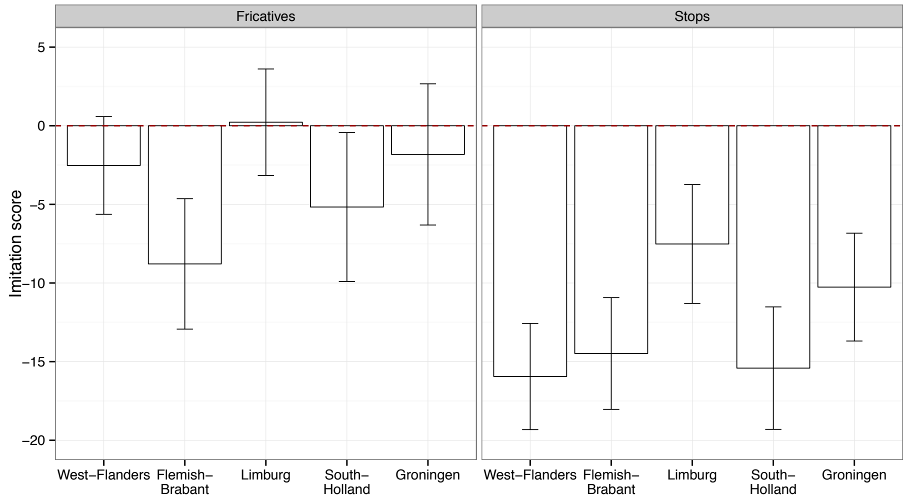
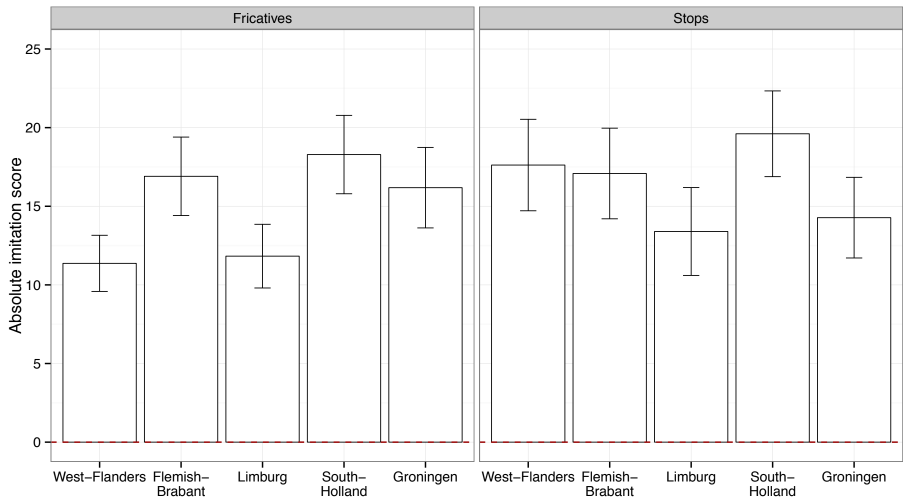


Figure 7.6: Barplot of absolute imitation scores (AIS) split up by variable (fricatives on the left and stops on the right) and by region (N=94). Error bars represent ±1 standard error.



Fricatives Participants' IS in fricatives did not significantly differ from 0, showing that, on average, participants did not tend to more overshoot or undershoot the amount of voicing when imitating. The AIS in fricatives lay on average around 15 and was similar across regions (no significant difference).

Stops Participants' IS in stops were also similar across regions, but did significantly differ from 0, showing a slight tendency for undershoot the amount of voicing in imitated stops. Like for fricatives, the AIS in stops lay on average around 15 and were similar across regions (no significant difference).

Moreover, linear regressions were ran in R (R Core Team, 2014) to test relationship between the AIS and participant-related factors. Two personality factors (see Appendix C) turned out to significantly predict the AIS: *conformism* and *leadership*. The less conformist the participant, the lower the absolute imitation score, thus the better the imitation (Linear Regression t=2.091, p=.038). The more 'leader-like' the participant, the lower the absolute imitation score (Linear Regression t=-1.975, p=.049). Moreover, there appeared to be a relationship between the AIS and participants' musical abilities. The better the participants' musical abilities, the lower the AIS, thus the better the imitation (Linear Regression t=-2.931, p=.004). In conclusion, the best imitators were the least conformist, the most 'leader-like' and the most musical participants.

Finally, there were a moderate correlation between the range in imitation (see previous section) and the AIS (r=-.40, p<.001). The lower the range, the higher the absolute imitation score. Thus, participants who showed the smallest ranges in imitation also were the participants whose imitation was qualitatively worst.

7.2.3 Discussion

The first imitation experiment was a forced imitation task in which participants were instructed to imitate a model talker the best as they could. This experiment was aimed at getting insights in participants' capacities to produce the whole range in voicing in labiodental fricatives and bilabial stops.

The first goal was to investigate, like Dufour and Nguyen (2013), whether participants were effectively able to imitate a degree of voicing when forced to. By comparing baseline productions from the carrier sentence task with imitated productions triggered by the presentation of stimuli, we found that participants imitate voiced fricatives and voiceless stops. No effect was found for voiceless fricatives, since baseline and imitation productions were highly comparable. Voiced stops showed − contrary to expectations − imitation productions with less voicing than baseline productions. It is unclear why participants showed divergence for this specific variant, whereas the task was designed to trigger convergence only.

Interestingly, participants from regions which are quite advanced in fricative devoicing (South-Holland and Groningen) and who showed reduced degree of voicing in their baseline productions turned out to be quite successful at imitating fully-voiced sounds. When explicitly instructed to, participants with high degree of devoicing are still capable of producing voiced variants.

The second goal was to investigate participants' capacities to produce the whole range of voicing in labiodental fricatives and in bilabial stops. In the context of (strong) devoicing, the question was whether individual speakers are still able to motorically produce fully voiced variants when they are 'forced' to, even if they have lost the contrast in their spontaneous speech production. Therefore, it was investigated – at the *individual level* – whether all participants were able to produce the entire voicing range, and – at the *group level* – whether there were regional differences in this capacity. It was shown that the vast majority of participants were able to produce the entire voicing range in both fricatives and stops. However, around one tenth of the participants were not quite successful at producing the entire ranges. They have lost the capacity to motorically produce fully voiced variants.

Two imitation scores were computed to assess the quality of imitation per participant, measuring how well the imitated produced resembled the target. We found that these scores did not differ across regions, showing once again that participants did not differ across regions in their imitation capacities, even though they differed in their baseline productions.

Furthermore, both the individual measures of ranges and imitation quality did correlate with participant-related factors, namely personality factors and musical abilities. The best imitators turned out to be the least conformist, the most 'leader-like', the most talkative and the most musical participants. As explained in Section 2.5.2, the common assumption is that phonetic imitation is a mechanism for a spread of a new variant through a speech community. In this approach, we might expect imitators to be more conformist and less leader-like. However, our findings showed the opposite pattern. These results thus seem possibly more consistent with the *change-by-accommodation model* (see also Section 2.5.2). In view of these interesting results, the link between speech imitation capacities, musical capacities and personality traits deserve more attention in further research.

In conclusion, the forced imitation task succeeded in demonstrating 'how far' participants can go in their production of voiced variants when they are instructed to, and therefore gives insight into motoric production capacities. Whether participants are really willing and ready to use these motoric capacities in social interaction is the question raised in the second imitation task: the spontaneous phonetic imitation.

7.3 Spontaneous phonetic imitation

In order to determine whether participants are likely to use devoiced variables in social interaction, a spontaneous imitation task was conducted. The task was meant to elicit spontaneous imitation, also called 'deliberate imitation' by Delvaux and Soquet (2007), thus imitation patterns which participants are not aware of, and which they produce spontaneously without being instructed to.

As reviewed in the introduction of the current chapter (Section 7.1), this type of phonetic imitation is often investigated in the frame of a shadowing task. In the present study however, an imitation task with a larger amount of social interaction was chosen. The design of the task was largely based on the study by Delvaux and Soquet (2007) who managed to find imitation patterns in vowel realizations between two regiolects of Belgian French.

We investigate the extent to which participants spontaneously imitate reference speaker in a social situation (at the *individual level*) and whether there are regional differences in these patterns (at the *group level*).

7.3.1 Method

Speakers

The stimuli were recorded in a sound-attenuated booth and spoken by two phonetically trained female speakers: reference speaker 1 (26 years old) came from Antwerp (Flanders) and reference speaker 2 (32 years old) from North Brabant (The Netherlands).

Since Yu et al. (2011), Babel (2012) and Babel et al. (2014) have shown that imitation is constrained by evaluative judgments towards the model talker, we carefully chose model talkers with pleasant female voices. Reference speaker 1 came from Flanders and reference speaker 2 came from the Netherlands, so that participants from all regions had at least one reference speaker to whom they could relate, and possibly imitate.

Stimuli

Both reference speakers read 36 target sentences digitally recorded with a sample frequency of 44.1 kHz in a sound-attenuated cabin of the lab of Utrecht University. All target sentences were of the type: 'de (*object*) gaat in de (*container*)' [*the (object) goes into the (container)*]. The objects were 36 easily recognizable objects of the daily life (f.i., a flower, a fork, a banana, a paperclip etc.). There was no particular linguistic or external constraint on those words, expect that they had to refer to objects of a size small enough to get into the containers. The containers were a 'vuilnisbak' [*trash bin*] and a 'boekentas' [*schoolbag*], with initial voiced labiodental fricative and initial voiced bilabial stop respectively. Both were low frequency words, since Goldinger (1998) showed that low frequency words tend to be imitated more than high frequency words. Both containers were three-syllabic words preceded by a schwa (since they have common gender 'de' [*the*]).

For each reference speaker, the best realization of each container word was phonetically manipulated in order to be highly devoiced. The manipulation consisted of replacing segments. The devoiced realizations were obtained by replacing voiced segments by devoiced segments. When manipulating segments, the whole CV syllable was replaced to avoid unnatural coarticulation effects.

Reference speaker 1's realizations of the voiced fricative in 'in de vuilnisbak' contained 36 % of voicing and the voiced stop in 'in de boekentas' contained 36 % of voicing. Reference speaker 2's realizations of the voiced fricative in 'in de vuilnisbak' contained 23 % of voicing and the voiced stop in 'in de boekentas' 29 %. The pitch contours of all model realizations were flattened through the PSOLA pitch manipulation and LPC resynthesis functions of `Praat`.

These manipulated versions of the container complements ('in de vuilnisbak' and 'in de boekentas') were concatenated with the 36 recorded sentences of each reference speaker. In this way, 36 stimuli sentences were obtained for each speaker, each of them containing one of the 36 different objects and the same realization of the devoiced fricative and stop. It is expected that participants gradually converge (or possibly diverge) in their productions from these reference devoiced phonemes.

Procedure

The experiment took the form of a card game played by three players: one was the participant and the two others were the two model talkers called *Anna* and *Lisa*, whose voices were played through headphones Beyerdynamic DT 250 (more details see Section 3.5).

In each trial (see schematic representation of the design in Appendix N), a card representing one of the 36 objects appeared on the screen, together with an arrow pointing towards either the represented schoolbag or the trash bin. When his turn came, each player has to orally formulate the association he saw between the object and the container by pronouncing the model sentence 'de (*object*) gaat in de (*container*)' for the two other players (f.i., 'de aardbei gaat in de vuilnisbak' [*the strawberry goes into the trash bin*].

The instruction given to participants was to perform the task when their turn came (every three trials). The turn of one of the reference players (either Lisa or Anna) was indicated by the coloring of their name on the screen. The corresponding player had to perform the task, and the participant remained silent while the sentence was pronounced. When the circle 'ik' [*me*] was colored, it was the turn of the participant. Participants' sentences were recorded through a C420 head-mounted microphone.

A cover story was used to conceal the real purpose of the experiment and avoid any awareness. Participants were told that the purpose of the experiment was to assess their abilities of memory and attention. They believed that the two other players were other participants in the study. They were instructed that they had to learn the associations between the objects and their container, and were asked in post-test to recall as many of these associations as possible.

Each participant completed a practice block of 18 trials. The test phase consisted of 180 trials in the card game presented in a randomized order

(5 repetitions of 36 different objects). Each player realized 60 trials. In this manner, 30 realizations of voiced labiodental fricatives ('in de vuilnisbak') and 30 realizations of voiced bilabial stops ('in de boekentas') were obtained.

7.3.2 Results

It appeared that the cover story worked very well. Participants were at no point aware of the real purpose of the experiment. On the contrary, they were clearly excited and challenged by the task, since most of them were eager to know their 'memory score' at the end of the experiment.

All recordings were sampled at 48k Hz (24 bits). All realizations of the target variables were segmented and labeled into phonetic segments within `Praat` speech-analysis software package (Boersma & Weenink, 2014). In order to achieve a high degree of consistency and a low error rate, the segmentation was done by hand by the researcher herself, following the same procedure as for the production data (see Section 4.2.2). For fricatives, two phonetic dimensions were measured: VOICING and DURATION. For stops, three phonetic measures were used: VOT, VOICING and DURATION.

Observations greater than four standard deviations from the mean of the measurements were considered as outliers and were removed (see arguments in Section 4.3). In total, 2962 observations were collected for fricatives and 2940 for stops.

This section is divided into two main analyses: an analysis of general imitation patterns compared to the baseline productions and an analysis of gradual convergence or divergence at the individual level.

Baseline vs. spontaneous imitation

Spontaneous imitation results were compared to baseline productions obtained from the carrier sentence production task (see Section 4.2). The barplot of the results split up by variant and by task is presented in Figure 7.7. The mean amount of voicing produced by the reference speakers (29.5 % for fricatives and 32.5 % for stops) is represented by the dotted line. Baseline productions from the carrier sentence tasks are presented in white and imitation production in grey.

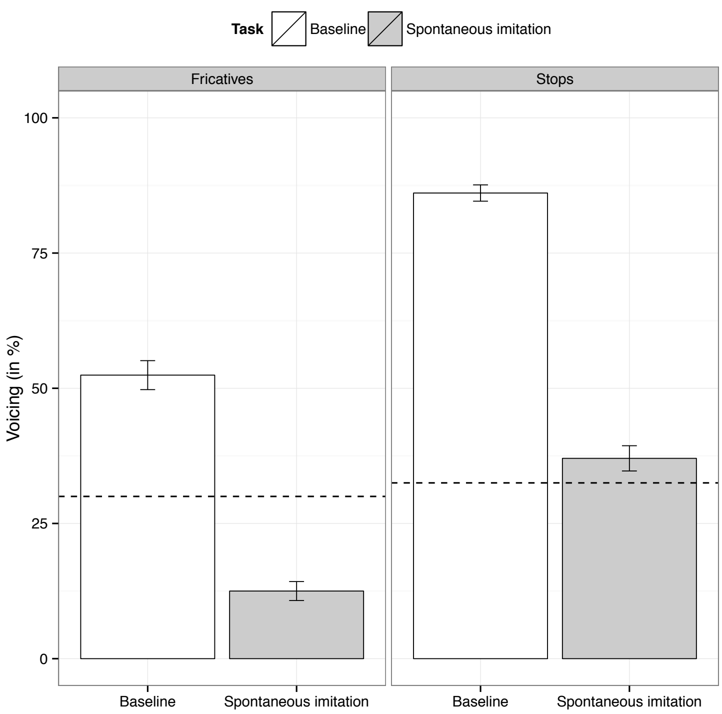
These results were analyzed with Mixed Effects Linear regressions (LMM, Baayen et al. (2008)) as implemented in the `lme4` library (Bates et al., 2012) in R (R Core Team, 2014). A model was fitted with fricative and stop data separately. In both models, the factor *participant* (N=100), testing for individual differences between participants was added as a random effect, and the factors *region*, *task* and the interactions between *region* and *task* were added as fixed effects. The full regression models are presented in Appendix O. A summary of the significant fixed factors is presented in Table 7.2.

Table 7.2: Significance of fixed factors in the mixed models regressions. Asterisks represent significant p values (α =.05).

	Fricatives	Stops
Task	*	*
Region	*	
Task * Region	*	*
In Figure	7.8	7.9

Note. Degrees of freedom (df) required for statistical significance testing of t values was given by $df = J - m - 1$ (Hox, 2010), where J is the most conservative number of second-level units ($J = 100$ speakers) and m is the total number of explanatory variables in the model ($m = 3$) resulting in $df = 96$.

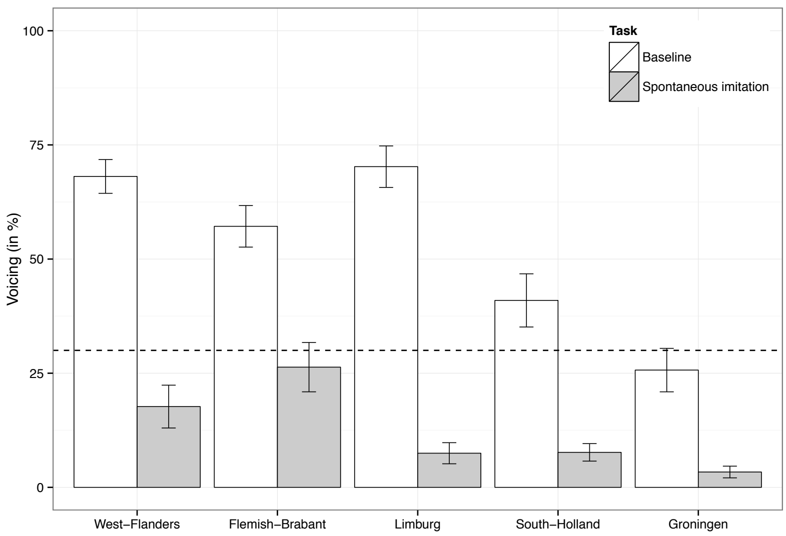
Figure 7.7: Barplot of spontaneous imitation data split up by variant and by task (baseline productions in white and imitation productions in grey). Voiced fricatives can be found in the left panel (n=2962) and voiced stops in the right panel (n=2940). Error bars represent ±1 standard error. Dotted lines represent the target amount of voicing in the presented stimuli.



Fricatives In Figure 7.7 (on the left side), productions from the spontaneous imitation task turned out to be less voiced than baseline productions (12 % vs. 52 % of voicing, respectively), showing an effect of task. Participants showed convergence towards the reference speakers' model (represented by the dotted line) and even more, since most realizations were even more devoiced than those of the reference speakers. There was an effect of region showing that regions slightly differed in their amount of voicing regardless of the task.

Moreover, there was an interaction between task and region. This interaction is plotted in Figure 7.8. In this barplot, we observe − as expected − that in all regions the spontaneous imitation task triggered less voiced productions than the baseline production task, but this difference in smaller in Flemish-Brabant, South-Holland and Groningen than in West-Flanders and Limburg. Interestingly, participants from Groningen who already showed highly devoiced productions in their baseline productions tended to devoice even more, when presented with model speakers who were devoiced themselves.

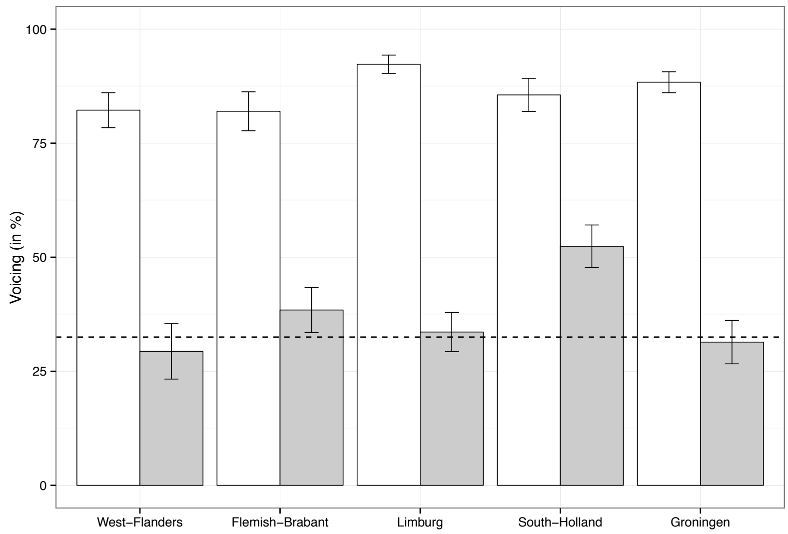
Figure 7.8: Barplot of the fricative spontaneous imitations split up by region and by task (baseline productions in white and imitation productions in grey) (n=2962). Error bars represent ±1 standard error. The dotted line represents the amount of voicing produced by the model speakers.



Stops For stops in Figure 7.7 (on the right side), there was also an effect of task. Productions in the spontaneous imitation task were less voiced than baseline productions (mean=37 % vs. 86 % of voicing). Participants thus spontaneously converged towards the reference speakers (represented by the dotted line in

Figure 7.7). There was no effect of region, but there was an interaction between task and region. This interaction is plotted in Figure 7.9. In this barplot, all regions produced less voiced stop productions in the spontaneous imitation task than in the baseline production task. This difference between baseline and imitation however turned out to be smaller in South-Holland than in the other regions. Participants were quite successful at imitating the voicing of the model speakers with voicing values around 30 %.

Figure 7.9: Barplot of the stop spontaneous imitations split up by region and by task (baseline productions in white and imitation productions in grey). Error bars represent ±1 standard error. The dotted line represents the amount of voicing produced by the model speakers.



Individual convergence vs. divergence

In this section, the extent to which individual participants converge to or diverge from the model speakers is analyzed. Both for fricatives and stops, 30 trials were collected during the experiment. Since it is expected that participants in the course of the experiment gradually converge to (or diverge from) the model speakers, linear regression models were fitted to the realizations of each participant with the factor *order of trials* as independent variable. In each model, it was tested whether the slope significantly differed from 0. In this way, it was determined whether individuals showed patterns of convergence (a negative slope) or divergence (a positive slope) in the course of the experiment.

Figure 7.10: Example of phonetic convergence. Fricative productions of participant #LI20 in the spontaneous imitation task. Each dot represents a realization. Trials are presented along the x-axis and the amount of voicing along the y-axis. The dotted line represents the amount of voicing produced by the model speakers and the full line is the fitted equation line to the production data.

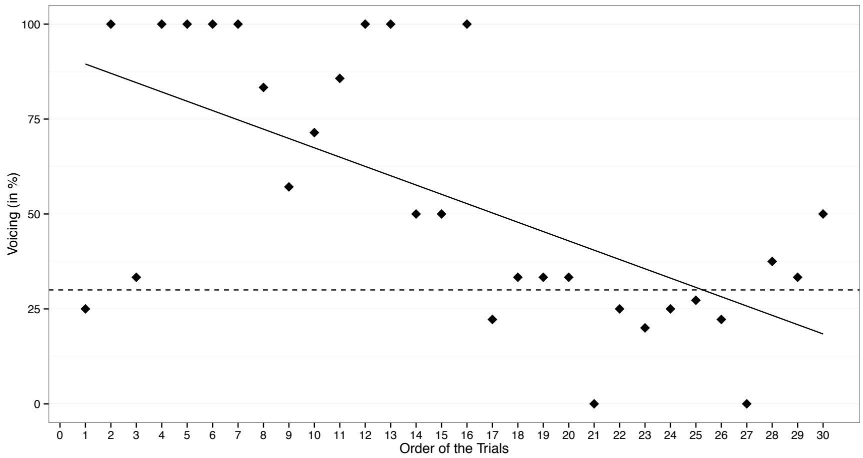
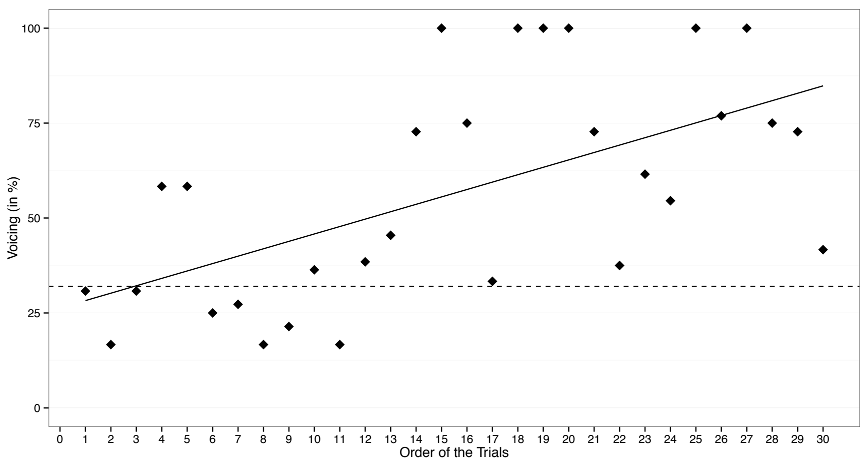


Figure 7.11: Example of phonetic divergence. Stops productions of participant #FB10 in the spontaneous imitation task. Each dot represents a realization. Trials are presented along the x-axis and the amount of voicing along the y-axis. The dotted line represents the amount of voicing produced by the model speakers and the full line is the fitted equation line to the production data.



Figures 7.10 and 7.11 showed examples of individual patterns of convergence and divergence respectively. The order of trials is represented along the x-axis and the amount of voicing in percentage is represented along the y-axis. The dotted line represents the amount of voicing produced by the model speakers and the full line is the fitted equation line to the production data. In Figure 7.10, it is clear that participant #LI20 showed convergence patterns in his fricative realizations, while participant #FB10 in Figure 7.11 showed divergence patterns in her stop realizations.

In Table 7.3, the number of individual participants who showed either convergence, divergence, or no imitation pattern is presented, split up by variable and by region.

Table 7.3: Number of individual showing convergence, divergence or no pattern in spontaneous imitation, split up by variable and by region (N=100).

Fricatives	Convergence	Divergence	No pattern	Total
West-Flanders	2	2	16	20
Flemish-Brabant	1	1	18	20
Limburg	2	0	18	20
South-Holland	2	1	17	20
Groningen	3	0	17	20
Total	10	4	86	100
Stops	Convergence	Divergence	No pattern	Total
West-Flanders	1	1	18	20
Flemish-Brabant	2	2	16	20
Limburg	0	0	20	20
South-Holland	0	1	19	20
Groningen	3	1	16	20
Total	6	5	89	100

For fricatives, a total of 10 participants showed a significantly negative slope converging towards the model speakers, whereas 4 participants showed a significantly positive slope diverging from the models. For stops, 6 participants showed convergence towards the model speakers, and 5 participants showed diverging patterns.

In conclusion, only a rather small proportion of participants showed clear phonetic convergence vs. divergence, when we looked at the gradual imitation patterns.

7.3.3 Discussion

The second imitation experiment was a spontaneous imitation task designed on the model of Delvaux and Soquet (2007), in which each participant was involved in a card game played with two reference speakers. This experiment aimed at gaining insight in participants' willingness and readiness to use devoiced variants of labiodental fricatives and bilabial stops in social situations. We investigated the extent to which participants spontaneously imitate reference speakers (at the *individual level*) and whether there are regional differences in these imitation patterns (at the *group level*).

The first goal was to figure out whether participants imitate a devoiced variant of a model speaker without being instructed to do so. By comparing baseline productions from the carrier sentence task and imitated productions triggered by the players' interaction in the card game, we found that participants imitate devoiced fricatives and devoiced stops. In all regions (even regions which already showed highly devoiced productions), the spontaneous imitation task triggered productions containing less voicing than the baseline productions of the carrier sentence task. This experiment confirms the findings of a range of previous studies on phonetic imitation (e.g., Babel, 2010, 2012; Babel & Bulatov, 2012; Goldinger, 1997, 1998; Goldinger & Azuma, 2004; Namy et al., 2002; Nielsen, 2011; Pardo, 2006; Pardo et al., 2012; Shockley et al., 2004), which showed clear spontaneous imitation patterns.

The second goal was to investigate participants' individual tendencies to converge to or diverge from the reference speakers. It turned out that only a small proportion of participants showed gradual phonetic convergence vs. divergence during the course of the experiment. Possibly, the experiment containing 30 trials for each variable was too short for some participant to yield significant slopes.

The divergence patterns found in some individuals in this study contrast with the convergence effects observed in Nielsen (2007, 2011). The exposure materials in Nielsen (2007, 2011) were words presented isolation, while our materials were embedded in a minimally interactive situation. The difference in divergence patterns might be attributed to the decontextualization of the exposure materials in Nielsen (2007, 2011). As shown by Babel (2012), Babel et al. (2014) and Yu et al. (2011), divergence patterns might be triggered by different evaluative judgments towards the model talkers. Finding large inter-participant variability was however not surprising, since Babel and Bulatov (2012), Nielsen (2011), Miller et al. (2010) and Pardo (2006) also found large individual differences in imitation patterns.

The spontaneous imitation task succeeded in demonstrating that simple exposure to model speakers induces imitation of devoiced fricatives and stops. Hence, there is a tendency to imitate the speech habits of reference speakers, even if the situation of communication is solely minimally interactive. At the *group level*, all regions showed spontaneous imitation, but slightly differed in the extent to which they imitate. At the *individual level*, a small number of participants clearly showed gradual imitation patterns (i.e., increasing phonetic divergence or convergence during the course of the experiment).

7.4 Linking forced and spontaneous imitation

The goal of the forced imitation task was to provide insight into motoric production capacities. By forcing participants to 'imitate' a model speaker as well as they could, their maximal range of production was investigated. By comparing baseline productions with imitated productions, we found that participants are able to imitate voiced fricatives and voiceless stops. Even participants from regions which are quite advanced in fricative devoicing turned out to be quite successful at imitating fully-voiced sounds. With exception of some few individuals, all participants managed to produce the whole range of voicing in fricatives and stops when they were instructed to.

This fact was the prerequisite for the successful conduction of the spontaneous imitation task, since the latter could concentrate on participants' willingness and readiness to devoice in social interaction, regardless of their motoric capacities. The spontaneous imitation task succeeded in demonstrating that simple exposure to model speakers induces imitation of devoiced fricatives and stops. Moreover, a small proportion of participants showed gradual phonetic convergence during the course of the experiment.

In conclusion, both imitation tasks were complementary to the goal of gaining insight in phonetic imitation capacities. In imitation studies, it is often easy to explain why participants do imitate, but more problematic to find an explanation why participants sometimes do not. The absence of imitation might always be attributed to either the lack of linguistic capacity or to the lack of willingness. In this study, this problem was solved by the combination of two imitation tasks (forced vs. spontaneous imitation task).

Still it remains unclear in the case of deficient imitators whether these speakers cannot articulate the sound due to motoric inability or whether they have a poor mental representation of what the speech sound is like. This question can only be answered by confronting imitation, perception and production data of the same individuals. Since this information is available from the other experiments in this study, the link between imitation, perception and production will be investigated in Chapter 8.

CHAPTER 8

Putting the pieces together

This chapter aims at bringing together the results of the experiments reported in the previous chapters (see Chapters 4, 5, 6 and 7). In these chapters, several measures succeeded in grasping the variation that exists in the different aspects of the linguistic competence involved in sound change. These measures will be brought together in order to answer the main research questions of this study as presented in Chapter 2:

> **RQ 2-3: What is the relationship between speech production and speech perception? Does change in perception precede change in production?**
>
> **RQ 1-4-5: What is the role of speech perception, language attitudes and phonetic imitation in sound change?**

Section 8.1 recalls the important measures presented in previous chapters. Section 8.2 discusses the relationship between speech production and speech perception. Section 8.3 investigates the link between the latter two processes and language attitudes and Section 8.4 focuses on the link between production, perception and phonetic imitation capacities. Finally, Section 8.5 presents a model of the role of these different factors in sound change.

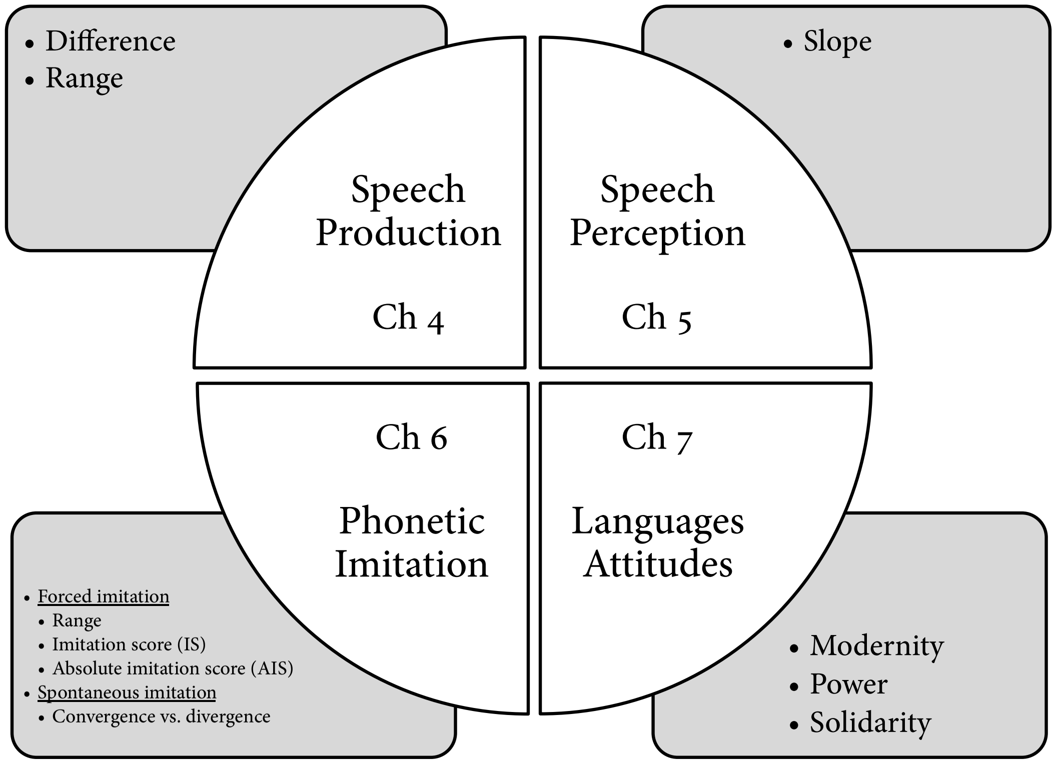


Figure 8.1: Schema of the four main aspects of the linguistic competence involved in sound change and their main measures investigated in Chapters 4, 5, 6 and 7.

8.1 The measures of production, perception, attitudes and imitation

Before investigating how measures of speech perception, production, attitudes and imitation relate to each other, it is important to recall what these measures exactly referred to.

As explained in Chapter 4, the key measure in order to assess the stage of completion of the two sound changes was sound change completion (SC completion) (see Section 4.3.5). This measure ranged from 0 (no change yet) to 100 (fully completed change). For instance, a participant who produces [v] on average with 40 % voicing would get a score of sound change completion of 60 %. In the current chapter, this measure constitutes the reference to assess sound change to which other measures are compared.

Figure 8.1 represents the four main aspects of the linguistic competence involved in sound change which were explored in this study, and − for each of these aspects − the main associated measures. The aspects in this figure are discussed clockwise, starting from the top-left corner.

Firstly, the results of *speech production* were discussed in Chapter 4. Two measures proved to be crucial in order to understand variation in speech production: the difference between the voiced and the voiceless categories and the range of the voiced categories (see Section 4.3.5).

Secondly, the results of *speech perception* were discussed in Chapter 5. The slope of the psychometric curves in the identification task turned out to be the best measure of speech perception patterns involved in sound change (see Section 5.2.2).

Chapter 6 put forth three measures, which succeeded in grasping variation in *language attitudes* towards the changing variables. These measures represented the differences in attitude scores between the voiced and devoiced conditions along the modernity, power and solidarity dimensions (see Section 6.3.2).

In Chapter 7 on *phonetic imitation*, three measures were used to report on forced imitation capacities: the range of forced imitation, the imitation score (IS)[1] and the absolute forced imitation score (AIS) (see Section 7.2.2). Moreover, one measure was used to report on spontaneous imitation capacities: the convergence vs. divergence score (see Section 7.3.2).

8.2 Link between perception and production

In this first section, the focus lies on whether there exists a relationship between variation in speech perception and variation in speech production in the context

[1] In this chapter, the IS is not discussed, since it is intrinsically related with the AIS (i.e., the AIS was computed as the absolute value of the IS) and since its analysis did not provide additional insight (see Section 7.2.2).

of sound change. Moreover, the related research question whether change in speech perception precedes change in speech production is investigated.

8.2.1 Introduction and hypotheses

As explained in Chapter 2, the number of studies on the link between speech perception and speech production is rather limited. Evidence in favor or disfavor of a strong link between the two processes is therefore insufficient and often contradictory. These studies have traditionally tackled the topic by looking at the *group level* (e.g., Harrington et al., 2008; Janson, 1983). They compared production and perception patterns of two or more groups of speakers.

In the current study, we have proceeded in the first steps in the same way as previous studies. In Chapter 5, we compared the perception patterns of participants from five regions from which we know that they show differences in production. The regions where devoicing was less advanced in production turned out to be also the regions where perception was most categorical, indicating a first type of between-group link between production and perception.

In this chapter however, the investigation of evidence in favor of the link between speech perception and speech production is taken further through the analysis of patterns at the *individual level*. Like Harrington et al. (2008), we agree to:

> abandon the notion that there is a uniform relationship between the phonological system and phonetic output for all members of the same speech community. Instead, phonological category boundaries are likely to be specific to different groups of speakers of the same speaking community and above all strongly related to the difference in their own production (Harrington et al., 2008, p.2832).

Only a few studies have directly examined the link between individuals' production and perception. Typically they have looked at the relationship between listeners' judgments on category goodness and their own production of a vowel target (e.g., Evans & Iverson, 2004; Frieda et al., 2000; Johnson et al., 1993; Nakai, 1998; Newman, 1997, 2003). These studies have suggested that adults exhibit a perceptual preference for phonetic tokens that are produced with more extreme features (e.g., higher vowels, longer VOT's) than those they typically produce. This effect has been called the *hyperspace effect*. Besides evidence for the hyperspace effect, Newman (2003) also found significant, though weak, correlations between listeners' perceptual prototypes and their average VOT's for English stop consonants (r=.49, p=.05). Listeners who selected as a perceptual prototype a stop consonant with a longer VOT were likely to show longer VOT's in their production of the same consonants. Others like Grosvald and Corina (2012) also investigated perception and production measures in the same manner, but without finding a statistically significant outcome.

Other studies examining the link between production and perception focused on the specific case of sound change (Beddor & Coetzee, 2014; Harrington et al., 2012; Kleber et al., 2011). Most researchers used different age groups in order to reflect the stages of change. Beddor and Coetzee (2014) investigated whether speakers of Afrikaans who produce more innovative variants also weight this innovative property more heavily in perception. They tested two age groups of Afrikaans speaking women and found that there was a weak correlation (r=.27, p=.01) between production and perception of these variants.

Fridland and Kendall (2012) and Kendall and Fridland (2012) designed a study which resembles most the current study in that they chose to look at sound change by comparing regions instead of age groups. They looked at vocalic changes in the frame of the American English Southern Vowel Shift (SVS) and Northern Cities Shift (NCS). They acoustically analyzed participants' vowel productions and compared them to the same speakers' performance on the vowel identification task. Results showed that both regional affiliation and individual participation in regional shifts in production played a role in perception, indicating a clear relationship between the two processes.

In view of these previous studies, we hypothesize that there exists a relationship between perception and production patterns in the context of sound change and that this relationship is probably of a weak nature. Our analysis seeks evidence for this link both at the *group* and *individual* levels. Moreover, on the basis of studies reviewed in Chapter 2, two hypotheses on the order in which these processes enter sound change were formulated. On the one hand, it is possible that change in perception precedes change in production as proposed for instance by Harrington et al. (2012) and Kleber et al. (2011). On the other hand, change in perception might be slower than change in production and lag behind as shown for instance by Janson (1983).

8.2.2 Evidence for a link between speech perception and speech production

In this section, we present three types of evidence for a link between speech production and speech perception: 1) an analysis *across* individuals, 2) an analysis *within* individuals and 3) an analysis of individuals with reference to the region to which they belong.

Link between perception and production across individuals

In order to obtain insight into the link between perception and production at the group level, we combined the production and perception measures reported in the previous chapters (see Chapter 4 and 5).

The two production measures (i.e., the difference between the voiced and the voiceless categories and the range of the voiced categories) were reported in Section 4.3.5. They turned out to be highly predictable from the timing of sound change. As a recall, these measures are plotted in grey in Figure 8.2 (which is equivalent to Figure 4.8). In this figure, the SC completion score plotted along the x-axis ranged from 0 (no change yet) to 100 (fully completed change).

Figure 8.2: Scatterplot of the of the voiced category and the difference between the voiced and voiceless variants (in voicing) (PRODUCTION) as a function of the sound change completion, split up by variable (N=100, n=200 observations).

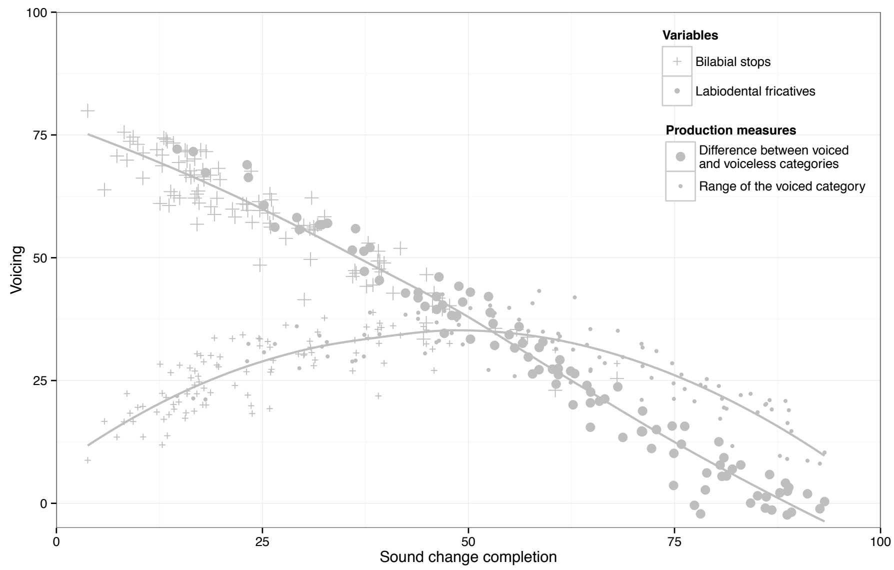
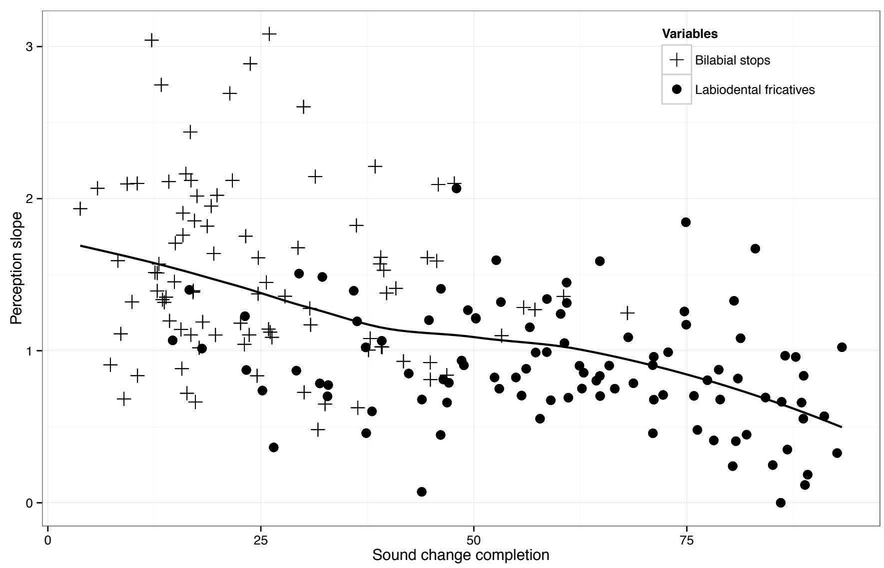


Figure 8.3: Scatterplot of the slope value (PERCEPTION) as a function of the sound change completion, split up by variable (N=100, n=193 observations).



Thus the more to the right, the further in the sound change[2]. Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot. The differences between voiced and voiceless categories are represented by big symbols and the ranges of the voiced category by small symbols. Each symbol stands for a participant. Separate lines are fitted for differences and ranges. As a reminder, the difference measure showed a linear relationship: it decreases in a linear manner as the change is proceeding. The range measure showed a curvilinear relationship with the completion of sound change: it increases in the first phase of the sound change, but crucially decreased in the second phase to form a curve (see Section 4.3.5 for the analysis).

The perception measure (i.e., the slope of categorical perception) was reported in Section 5.2.2. The perception slope was measured as the slope of the logistic regression curves obtained in the identification experiment[3]. The perception slopes are plotted in black in Figure 8.3 against the SC completion. They ranged between 0 (no categorical perception) and 3.083 (very categorical perception)[4]. Slope values greater than two standard deviations from the mean slope were removed (i.e. 6 outliers on a total of 200). These six outliers presented not only an unreasonably high values of perception slope, but also an unusual cut-off point between categories. For both these reasons, they were removed from the analysis. Finally, there was one missing slope value[5], resulting in 193 perception observations. In Figure 8.3, bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot.

In Figure 8.3, it appears that the variation in perception slope in stops is greater than in fricatives. Crucially, it turned out that the slopes in perception show a relationship with the SC completion score. This relationship is a quasi-linear one: perception slopes are high (very categorical perception) at the beginning of the sound change, decrease quite rapidly after the beginning, and keep decreasing until the change is completed. At the end of sound change, it appears however that change in perception slopes is not really reaching 0 (no trace of categorical perception any longer). Some traces of categorical perception are still present at the sound change completion. In conclusion, the more advanced the change, the lower the perception slope value, thus the less categorical the perception.

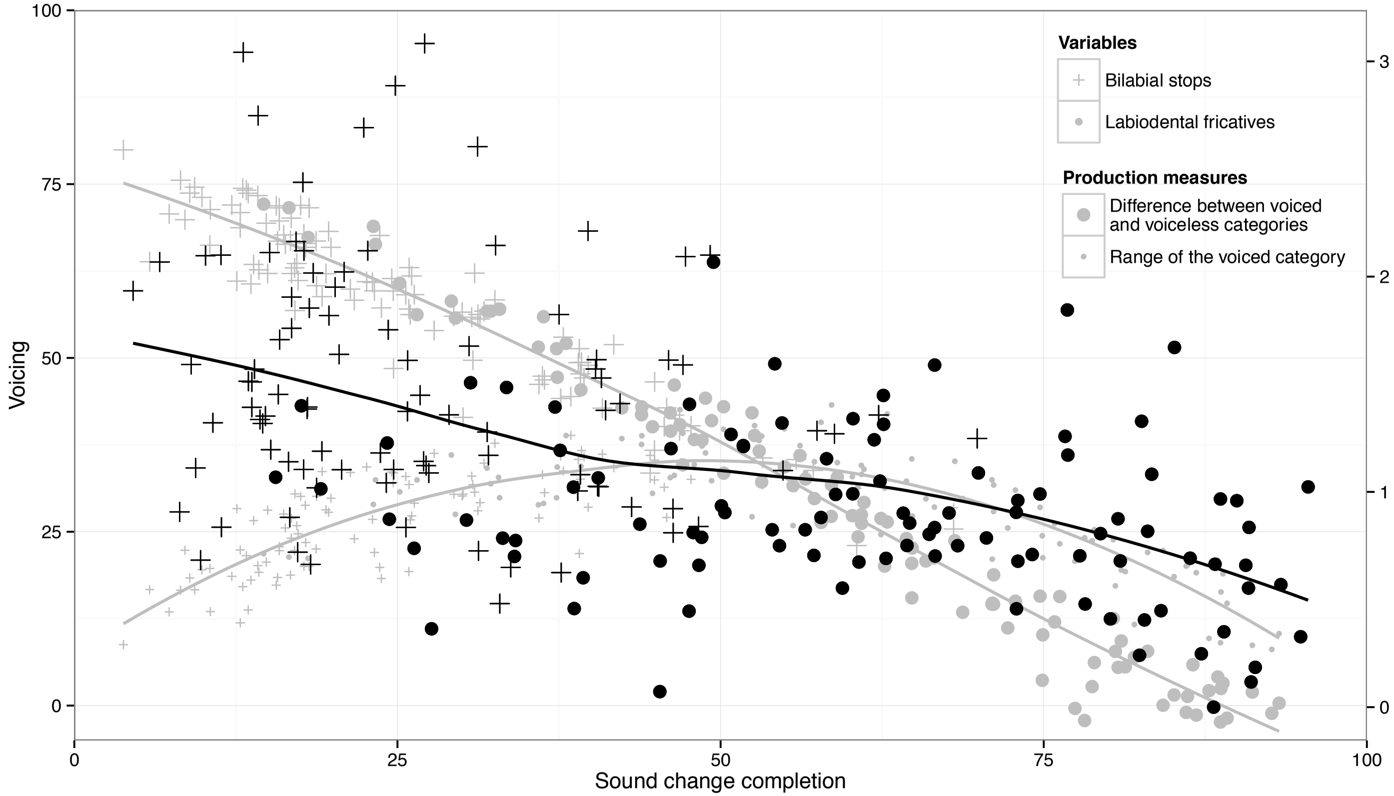
[2]In this chapter, we choose to systematically present the SC completion score along the x-axis, since it has to do with the development of change, representing time. We could also have chosen to represent the SC completion along the y-axis, insisting on the assumption that our measures predict the SC completion. This choice has no statistical consequence.

[3]This way of measuring the slope was chosen above a related measure: the calculation of the slope solely in the centre of the continuum (i.e. the two steps around the cut-off point between categories). Both slope measures (slope of the entire continuum vs. slope around the cut-off point) were highly correlated (r=.831, p>.001), indicating stable results. However, the slope on the entire phonetic continuum appeared to better grasp the existing variation and was finally chosen for this analysis.

[4]Slope values were initially negative (see Section 5.2.2), but are presented here without sign for the ease of visualization.

[5]As explained in Section 5.2.2, the data of the stop identification task were lost for one participant (#WF12).

Figure 8.4: Scatterplot of the range of the voiced category, the difference between the voiced and voiceless variants (both colored in grey and indicated in percentage of voicing on the left y-axis) and the slope in perception (colored in black on the right y-axis) as a function of the sound change completion, split up by variable (n=100, n=200 observations for the range and difference measures and n=193 for the slope measure).



This relationship was tested statistically with the help of a linear regression. The SC completion score was significantly predicted by the perception slope (LINEAR REGRESSION t=-8.279, p<.001)[6], which could account for 26.41 % of the variance in completion scores.

The amount of explained variance was clearly lower than the amount of explained variance by the two production measures (96.62 % and 72.37 %, see Section 4.3.5). However, considering the fact that we are looking at two different processes of speech (perception and production (computed with the SC completion)) and their state in the process of change, explaining a fourth of the variance by the state of perception is not trivial. This relationship constitutes strong evidence for a link between speech production and speech perception.

Subsequently, Figure 8.2 and 8.3 were superimposed in order to obtain Figure 8.4. In this figure, the original Figure 8.2 (showing the relationships between SC completion and the production measures) is presented in grey and the original Figure 8.3 (showing the relationships between SC completion and the perception measures) is presented in black. The SC completion is displayed along the x-axis. The production measures are displayed along the left y-axis and the perception measure along the right y-axis. Each symbol represents one participant (N=100 participants, n=200 observations for the range and difference measures and n=193 for the slope measure). Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot.

Looking at Figure 8.4, the hypothesis that change in perception might be slower than change in production seems to be confirmed. The change in perception seems to slow down half way through the change and not fully reach the end of the process. Hence, the perception slopes remain slightly categorical even when there is no production difference in voicing between the categories any longer. However, the hypothesis that change in perception precedes change in production is not directly confirmed by these results. It appeared that – at least at the beginning of the sound change – that both perception and perception evolve at the same rate (the slope measure and the difference measure decrease parallelly).

Link between perception and production within individuals

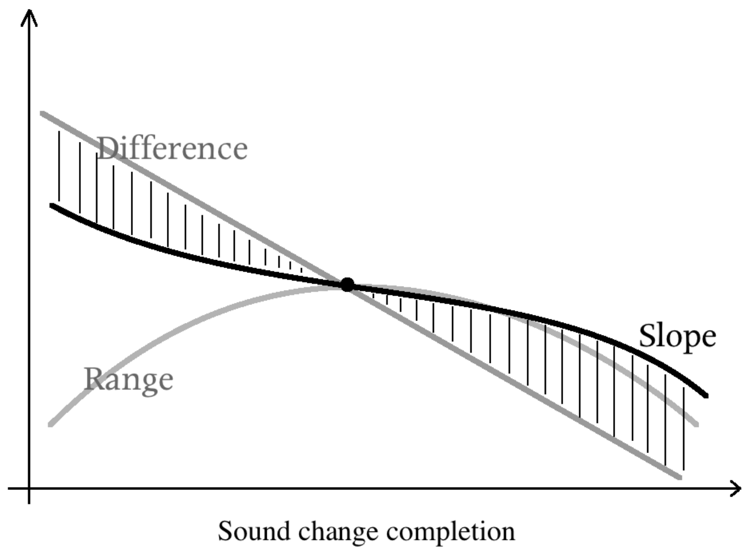
In contrast to the previous analysis *across* individuals, the current analysis focuses on the link between speech production and speech perception *within* individuals.

Figure 8.5 offers a schematization of Figure 8.4. In this figure, the curves of the production measures (difference and range) are presented in grey and the curve of the perception measure (perception slope) is presented in black. The relationship between production and perception *within each individual* is schematized by the black vertical lines.

---

[6] In this chapter, linear regressions are systematically preferred to correlations for the investigation of relationships between dependent variables, since they allow a direction in the relationship (y as a function of x).

Figure 8.5: Schematic representation of the range of the voiced category, the difference between the voiced and voiceless variants and the perception slope as a function of the completion of the sound change.



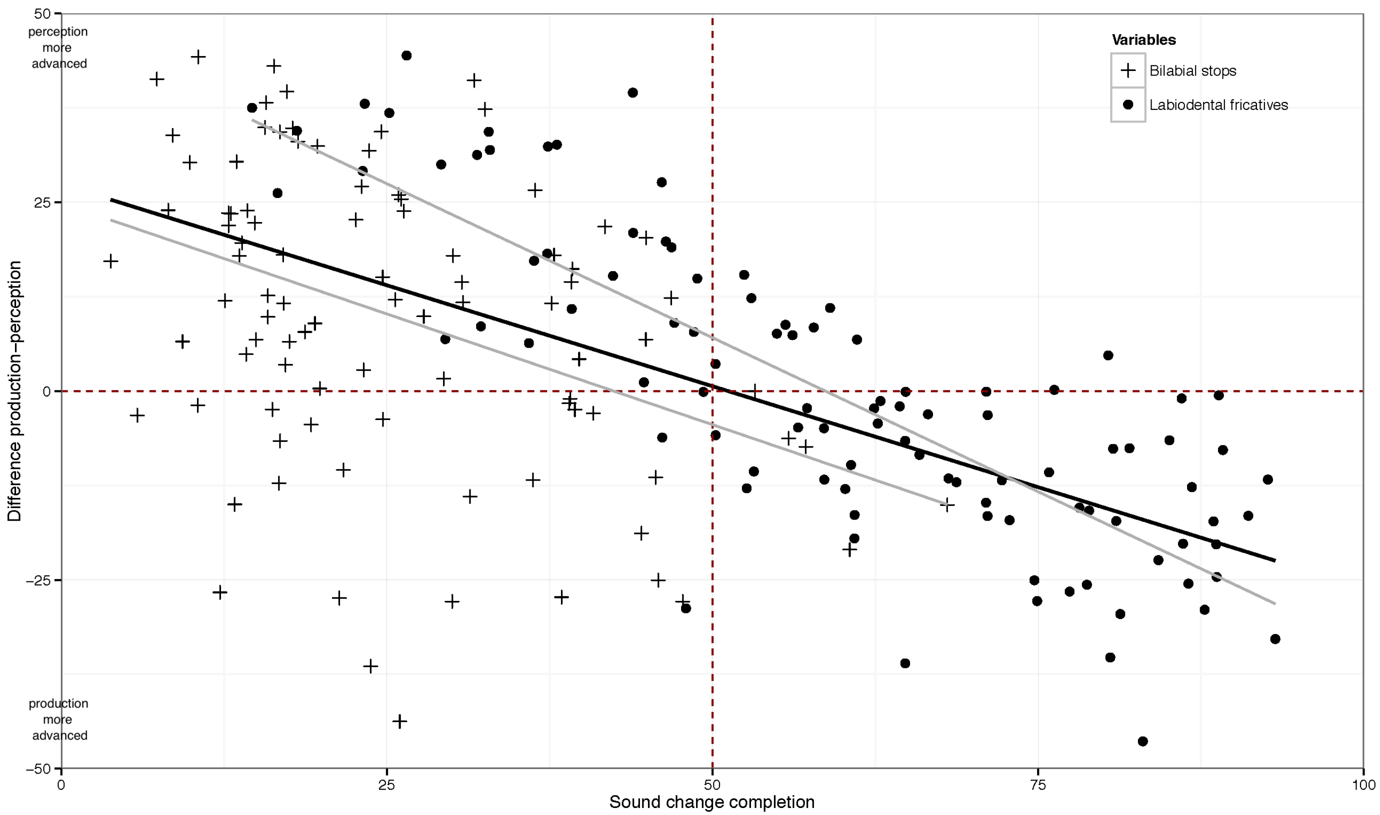
From these schematized individual lines, relation patterns between individual perception and production emerge. In order to obtain more insight into this pattern, a measure of the relationship between production and perception was computed for each individual: the *difference production-perception* (diffPP). This measure is schematized in Figure 8.5 by the vertical lines.

$$\text{diffPP} = \text{difference}_{\text{in voicing between categories}} - (\text{slope}_{\text{perception}} * 32.436) \quad (8.1)$$

First, the perception slopes were multiplied by 32.436, so that participant #WF18 (in the categorization of stops) showing the most categorical slope (3.083) reached a perception score of 100 (maximally categorical). Subsequently, the diffPP was computed as the difference between − on the one hand − the difference in voicing between the voiced and voiceless categories, and on the other hand the perception score. The assumption was made that both production and perception could be measured in the same way, namely ranging from 0 (no more difference in production between the two categories and no categorical perception any longer) to 100 (maximum possible difference in production between the two categories and maximally categorical perception).

As a result, the diffPP constitutes a measure of each participant's state of speech production as compared to the state of his speech perception. If the measure is positive, the individual's production score is higher than his perception score (thus that production is lagging behind in the process of change). If the measure equals 0, the individual's production is equally advanced in the process of change as his perception.

Figure 8.6: Scatterplot of the production-perception difference score (diffPP) as a function of the sound change completion, split up by variable (N=100 participants, n=193 observations). The main fitted line is presented in black and the fitted lines for each separate variable in grey (for fricatives above and for stops below).



If the measure is negative, the individual perception score is higher than the production score (thus that perception is lagging behind in the process of change).

In Figure 8.6, this new measure, diffPP (along the y-axis) is plotted against the sound change completion (along the x-axis). Each symbol represents a participant (n=100 for fricatives, n=93 for stops). Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot. Three lines are fitted to the data: the main fitted line in black and one grey fitted line for each separate variable (one for stops and one for fricatives).

At the beginning of the change, most participants showed a positive diffPP, indicating that their perception is more advanced than their production. A sizeable number of participants however showed – already at the beginning of the change – a slightly negative diffPP with an individual production more advanced than the perception patterns. At around 50 % of completion, the diffPP often equals 0 with individual production and perception approximately at the same stage of advancement. Towards the end of the sound change, the diffPP scores are mostly negative, indicating that production is more advanced than perception. It appeared that the variation in stops is greater than in fricatives, which is consistent with the fact that stop devoicing is incipient.

This relationship was tested statistically with the help of linear regression. The SC completion (of stops and fricatives together) was significantly predicted by the diffPP (Linear Regression t=-10.85, $p<.001$), which could account for 38.13 % of the variance in completion scores. The fact that this relationship is statistically significant is unsurprising, since the diffPP and the SC completion both based their computation on the mean voicing of the voiced category, and are therefore intrinsically related. However, the fact that the correlation is negative and that the equation line crosses the x-axis around 0 are interesting for the timing of change in production and perception. Indeed, if change in production and perception were to happen parallel, participants would have a diffPP around 0 all the way through the change, definitely not the case here. If change in perception would be completed before change in production, the fitted slope would be positive instead of negative. However, the slope is negative. When reaching completion, individual diffPPs lie clearly below 0, indicating that production is reaching complete devoicing, whereas in perception voiced and voiceless categories are still distinguishable. Hence it constitutes evidence in favor of the hypothesis, stating that perception lags behind, yet only towards the end of the sound change.

In conclusion, the score of relationship between individual's production and perception (diffPP) does not directly tell which aspect changes first *within an individual*. Only a longitudinal study of an individual's production-perception link would provide direct evidence. However, when looking at the patterns *across individuals*, it shows evidence for the hypothesis that change in perception precedes change in production and that perception lags behind at the end of the sound change.

Link between perception and production in individuals within their region

This third section on the link between speech perception and speech production focuses on the individual within his own region. We investigated whether individuals who are slightly more advanced in the sound change than the rest of the participants from the same region show specific perception and production patterns.

To obtain further insight into this question, a measure of the relationship between individual devoicing and the devoicing within his region was computed for each participant. The *sound change completion score with respect to the own region* (CompReg) was defined by the following equation:

$$\text{CompReg} = \text{individual SC completion} - \text{mean SC completion of region} \quad (8.2)$$

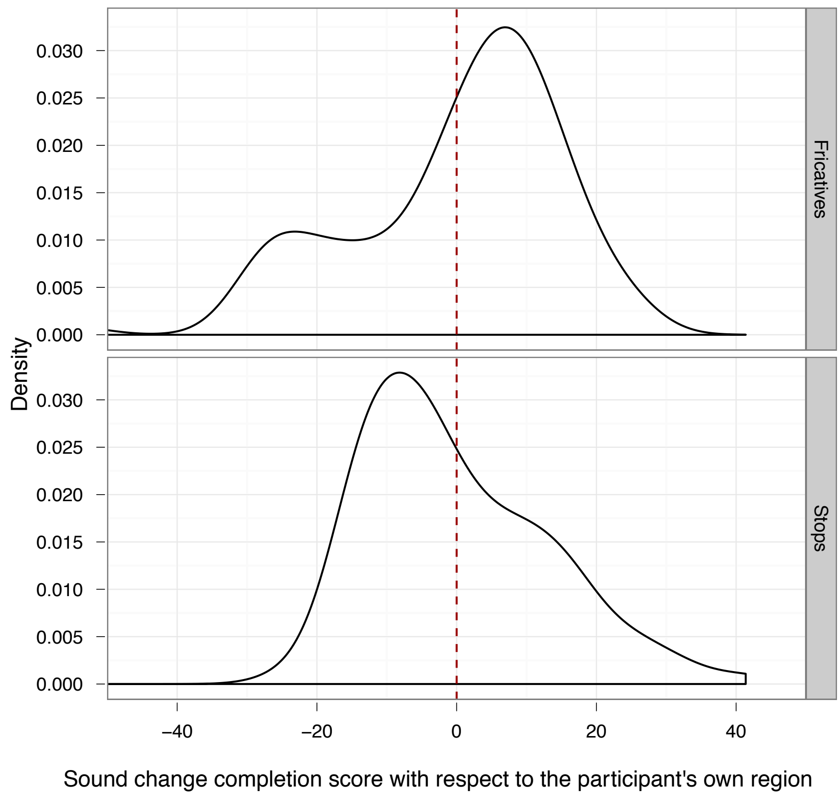
For each participant, the difference was thus computed between his individual score of sound change completion and the mean sound change completion of all individuals from the region to which he belongs. As a result, this measure allows for computation − per participant − his advancement in the sound change as compared to the state of his region. Moreover, the other participants of the same region are representative for the speech environment that an individual hears most.

If the measure is positive, the individual is more advanced in the change than his region. If the measure equals 0, the individual production is equally advanced in the process of change as the other participants from the same region. If the measure is negative, the individual is lagging behind in his own region, since the other participants are on average further in the process of devoicing.

In Figure 8.7, the density plots of the CompReg, split up by variable, are presented. These density plots represent the distributions of the CompReg. In this figure, the mean of the CompReg logically equals 0, but interestingly the distributions are skewed. The distribution for fricatives is skewed to the right (towards the positive scores), indicating that the majority of individuals (N=62) is more advanced than the average sound change completion of their region. In contrast, the distribution for stops is skewed to the left (towards the negative scores), indicating that the majority of individuals (N=59) is less advanced than the average sound change completion of their region.

It appears thus that − for an incipient change like the devoicing of stops − there are only a few individuals who are more advanced and possibly lead the sound change. The other individuals lag behind. In contrast, when a sound change is more advanced like in the devoicing of fricatives, most individuals within a region are further in the process than the mean of the region and there are only some conservative individuals lagging behind in the process.

Figure 8.7: Density plots of the sound change completion score with respect to the own region (CompReg), split up by variable: fricatives (n=100) (in the upper panel) and stops (n=100) (in the lower panel).



In Figure 8.8, the CompReg (along the y-axis) is plotted against the SC completion (along the x-axis). Each symbol represents a participant (n=100 for fricatives, n=100 for stops). Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot.

In this figure, the CompReg for both variables showed a linear relationship with SC completion. This relationship logically follows from the computation of CompReg and SC completion, so that individuals who are conservative within their region are less advanced in the change than individuals who are more progressive within their region. Moreover, the internal relationship between both measures causes the organization of observations of the same region around lines. For fricatives, for instance, the line of observations which crosses as first the x-axis contains participants from West-Flanders and the line which crosses as last the x-axis participants from Groningen. The point in which the fitted lines cross the x-axis gives an indication of the mean stage of completion of each variable: for stops around 25 % of completion and for fricatives around 60 %.

Figure 8.8: Scatterplot of the sound change completion score with respect to the own region (CompReg) as a function of the sound change completion, split up by variable (N=100 participants, n=200 observations).

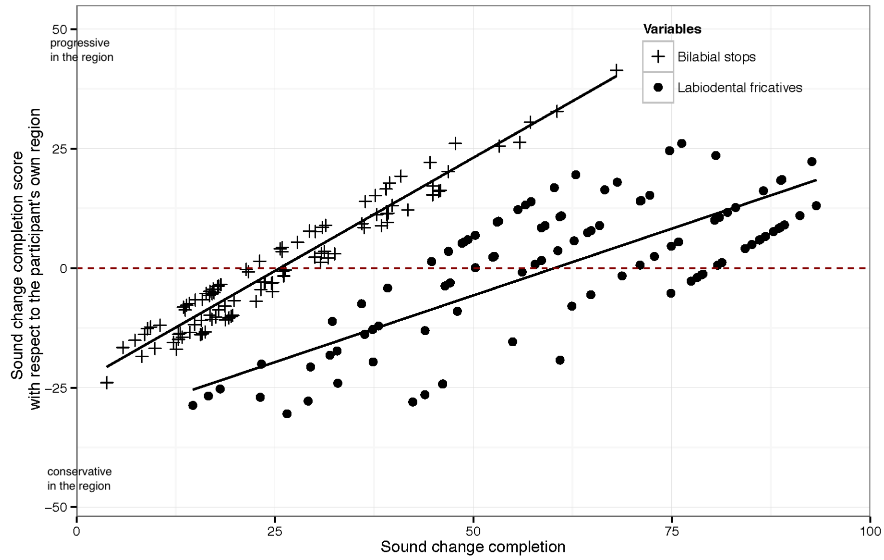
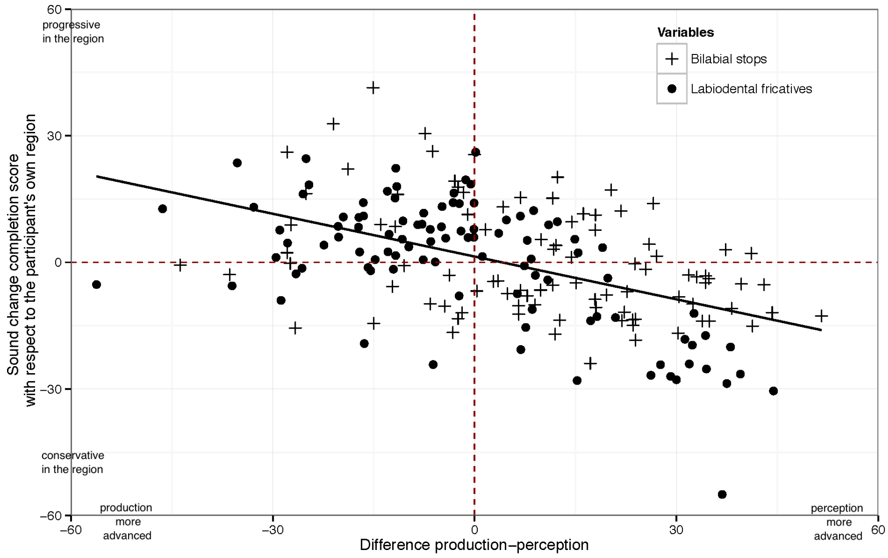


Figure 8.9: Scatterplot of the sound change completion score with respect to the own region (CompReg) as a function of the production-perception difference score (diffPP), split up by variable (N=100 participants, n=193 observations).



In Figure 8.9, the CompReg (along the y-axis) is plotted against the previously computed diffPP (along the x-axis). Each symbol represents a participant (n=100 for fricatives, n=93 for stops). Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot.

In Figure 8.9, there is a negative correlation between the two measures. The most progressive individuals within a region (with a positive CompReg) mostly have a negative diffPP, showing that their production is more advanced than their perception. This is logical in view of Figure 8.8, in which it was shown that the most progressive individuals are also the most advanced in the change, at a stage where perception lags behind. The more conservative the individual is within his region (more negative CompReg), the more positive the diffPP, thus the more advanced their perception in the change as compared to their production.

This relationship was tested statistically with the help of a linear regression. The CompReg was significantly predicted by the diffPP (Linear Regression t=-8.033, p<.001), which could account for 25.25 % of the variance in completion score.

Once again, the relationship between speech perception and speech production is relevant to assess how far an individual is in the process of change, and especially how far he is in comparison to the other individuals of the region to which he belongs. The current analysis also provided indirect evidence in favor of the two hypotheses. First, the negative diffPP scores for the most progressive individuals within a region showed that their production was more advanced than their perception, pointing at the fact that perception is lagging behind at the end of the sound change. Second, the positive diffPP scores for the most conservative individuals within a region showed that their production was more advanced than their perception, confirming that change in perception precedes change in production at the beginning of the sound change.

8.2.3 Discussion and conclusion

We found three types of evidence in favor of the link between speech perception and speech production in the context of sound change: 1) an analysis *across* individuals, 2) an analysis *within* individuals, and 3) an analysis of individuals with reference to the region to which they belong.

The three analyses have provided results which point towards a clear link between speech perception and speech production, both at the *individual level* and at the *group level*. These results are in line with the study of Fridland and Kendall (2012) who also showed − through the analysis of regional patterns of sound change − the relationship between perception and production patterns. In contrast to Fridland and Kendall (2012) however, categorization slopes instead of the categorization cut-off points turned out to be the relevant perception measures in this study. The strength of the relationship between production and perception in these results is also comparable with the results of previous study (e.g., Beddor & Coetzee, 2014; Newman, 2003). The correlation between speech production and speech perception can be said to be relatively weak.

As far as the timing of change is concerned, two hypotheses were formulated. First, it was proposed that change in perception might precede change in production. Second, change in perception might be slower than change in production and lag behind. Important to note is that these hypotheses are not necessary mutually exclusive. In any case, the second hypothesis, the fact that perception lags behind when the sound change is reaching completion, was clearly confirmed by the three types of evidence presented in this section. Determining whether the first hypothesis can also be confirmed was more difficult in this type of cross-sectional data. However, evidence from the analysis at the group level and from the analysis of the individuals within their regions seemed to confirm the fact that change in perception precedes change in production.

Hence, it seems to be the case that – at the beginning of a sound change – perceptual patterns are adjusted first. These perceptual adjustments then necessarily have consequences on speech production patterns. Once the devoicing is launched in production, it has as necessary end point: the full devoicing of the category. However, it was shown that endpoint in perception is not as far as in production: perception lags behind. Participants are still able to hear some categorical contrast in a variable for which they do not make the contrast in production any longer.

The fact that the sound change seems to be triggered by a change in perception first, which is followed a change in production, is not very hard to conceive. As it is through the perceptual system that the ambient forms are perceived, it seems logical that a change happens in that system in the first place. At the same time, it is very plausible that change in perception does not reach an end stage as easily as change in production. Indeed, listeners still maintain some contrast possibly through speech input they receive from outside their region (in which the contrast is still present). Moreover, spelling together with the fact that our society is highly based on literacy and written media obviously contributes to an increased awareness of this specific contrast.

Importantly, the conclusion reached here is based on the analysis of two changes: the devoicing of labiodental fricatives and of bilabial stops. The juxtaposition of these two changes allows us to get insights into the whole process of sound change and points at effects which are either related to the beginning or the end of the process. Even if the two sound changes are very similar types of change (see Section 3.2), it is legitimate to ask whether their intrinsic differences did not bias the results we obtained and consequently the conclusions we reached. As far as speech perception is concerned, it has for example been proposed that the perception of stops is intrinsically more categorical than the perception of fricatives (Healy & Repp, 1982; Liberman et al., 1967). In the literature, evidences are conflicting. Although it is generally accepted that the perception of consonants is more categorical than the perception of vowels (e.g., Fry, Abramson, Eimas, & Liberman, 1962; Kronrod, Coppess, & Feldman, 2012; Pisoni, 1975; Repp, 1984; Schouten & van Hessen, 1992), there has been some disagreement over the degree of categorical perception within consonants.

Repp (1981) showed that fricatives follow patterns similar to the categorical perception found in stop consonants. In contrast, other studies have found that fricative perception is less categorical than stop perception (Healy & Repp, 1982; Liberman et al., 1967). Looking at our data, it is clear that the perception of stops showed a broader variation than the perception of fricatives. It seems that the difference in our data between stop and fricative perception is more related to the amount of variation within slopes than to their mean. Moreover, most conclusions reached in this section would hold even without the stop data. For instance in Figure 8.6, the fitted line for fricatives only (in grey) showed approximately the same slope as the fitted line for all data (in black). Based on these arguments, we conclude that intrinsic differences in the perception of stops and fricatives (if they exist) do not undermine our conclusions.

In conclusion, there is a clear link between production and perception in the context of sound change. Moreover, it was shown that − at the beginning of sound change − changes in speech perception precede change in production. Speech perception appeared to lag behind speech production when the sound change is reaching completion. In the two following sections, it is investigated how the mechanisms of language attitudes and imitation respectively are related to these processes.

8.3 Link between perception, production and language attitudes

In this section, the mechanism of evaluation of linguistic variables through language attitudes is brought into relationship with change in speech perception and speech production. The focus lies on the question to what extent language attitudes play a role in sound change.

8.3.1 Introduction and hypotheses

Due to the difficulty to collect production, perception and attitude data in the context of the same sound change, studies on the link between these processes are very rare. Hay, Warren, and Drager (2006), however, succeeded in investigating the specific role of attitudes in speech production and speech perception in the context of sound change.

Hay, Warren, and Drager (2006) addressed the issue of listener- and speaker-specific characteristics in speech processing. They investigated the expanding near-square merger in New Zealand English and found that the perception of words involved in this sound change was dependent both on the listeners' own production of the merger and on social attributes associated with the speaker. Providing evidence in favor of exemplar theories, they concluded that both social information and the participant's own production played a role in his speech perception.

Despite the lack of evidence around these themes, we proposed two hypotheses related to the role of language attitudes in the context of sound change. First, it is hypothesized that language attitudes play a role in speech production patterns undergoing sound change. It is predicted that the more devoicing in speech production, the more positive the attitudes towards this change. Second, it is hypothesized that language attitudes also play a role in speech perception. We predict that the better the sound is perceived (more categorical), the more language attitudes are associated with it.

8.3.2 Evidence for a link between perception, production and attitudes

In this section, evidence related to the role of language attitudes in the context of sound change is presented. First, the relationship between languages attitudes towards the two variables and the sound change completion is investigated. Second, the relationship between languages attitudes and perception is discussed.

A link between production and attitudes

This analysis investigates the relationship between languages attitudes and speech production. Therefore, the correlations between three measures of language attitudes (i.e., the differences in factor scores between the voiced and devoiced conditions along the modernity, power and solidarity dimensions (see Section 6.3.2)) and the SC completion score were tested. These correlations and their significancy are reported in Table 8.1.

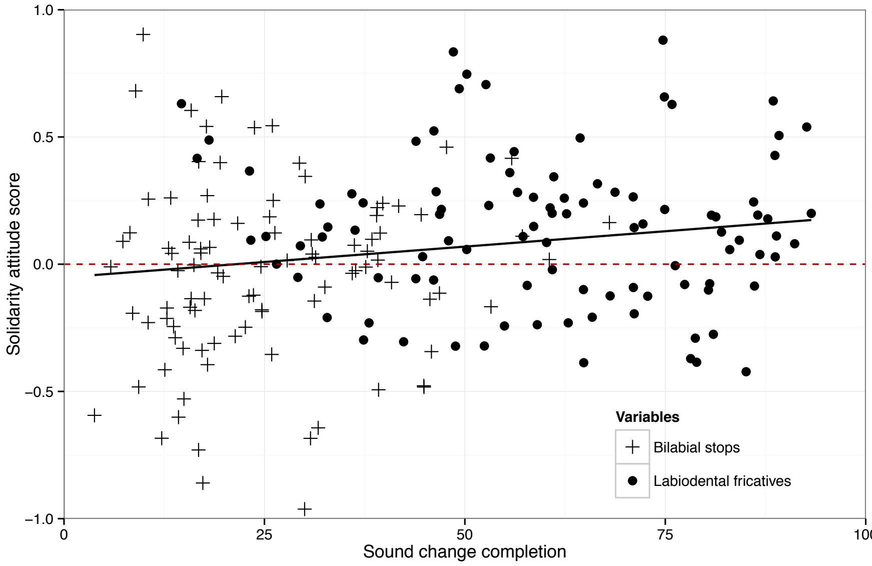
Table 8.1: Correlations between the three measures of language attitudes (along the modernity, power and solidarity dimensions) and the sound change completion score.

	r	p value	
Modernity	-.03	.689	
Power	.07	.318	
Solidarity	.17	.016	*

As seen in Table 8.1, only the correlation between the solidarity measure and the sound change completion was significant. This correlation is plotted in Figure 8.10. In this figure, the SC completion score is plotted along the x-axis, and the solidarity attitude score along the y-axis. Each symbol represents a participant (N=100 participants, n=200 observations). Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot.

In this figure, the attitude scores for stops show more variation than the scores for fricatives. There is a positive correlation between the attitudes towards the variables along the solidarity dimension and the SC completion. At the beginning of the sound change, speakers have slightly negative attitudes towards the variables along the solidarity dimension, indicating that none of the conditions (voiced vs. devoiced) was preferred. The more the sound change is progressing, the more positive the attitude scores. This means that the devoiced condition was gradually more and more preferred above the voiced condition along the solidarity dimension scales, as the change is proceeding.

Figure 8.10: Scatterplot of the solidarity attitude score as a function of the completion of the sound change, split up by variable (N=100 participants, n=200 observations).



This relationship was tested statistically with the help of a linear regression. The SC completion was significantly predicted by the solidarity attitude score (Linear Regression t=2.42, p=.016), which could account for 2.87 % of the variance in completion scores.

This analysis provides evidence for a relationship between language attitudes and speech production, which confirms our first hypothesis. The more devoicing of a variable, the more positive devoicing is rated along the solidarity dimension. This relationship, however, turned out to be weak.

A link between perception and attitudes

In our second step, we investigated the relationship between perception and languages attitudes. Crucially, we found no direct relationship between attitudes measures and perception measures (Linear Regression all t's<1.5).

Therefore, we need to reject our second hypothesis. In our data, language attitudes were not directly shown to play a role in speech perception.

8.3.3 Discussion and conclusion

Two hypotheses were formulated around the link between speech perception, speech production and language attitudes.

First, the hypothesis that language attitudes play a role in speech production in the context of sound change was confirmed. The results showed that the more devoicing, the more positive devoicing was rated along the solidarity dimension. Second, the hypothesis that language attitudes play a role in speech perception was not confirmed, since no relationship was found between the attitude measures and the perception measure.

The significant correlation between the attitude measure and the sound change completion turned out to be quite weak with only a rather small amount of variance explained. The evidence was shown for the attitude dimension *solidarity* only. The attitude scores on the two other dimensions did not correlate with measures of speech production nor perception.

It is important to note that the link found between attitudes and production does not imply causal relationship. It was shown that the more devoiced the speech production, the more positive the attitudes towards this change. This does not imply that the devoicing in speech production induces more positive attitudes, nor that positive attitudes cause change in production.

In conclusion, we showed that language attitudes towards a variable undergoing change do not seem to be present from the beginning of the sound change onwards. The positive evaluation of a change along the solidarity dimension seems to arise when the system of speech production is already undergoing the change. Language attitudes, which were shown to develop during the process, seem to be a reinforcing mechanism of sound change.

8.4 Link between perception, production and phonetic imitation

In this third section, the mechanism of phonetic imitation is linked with the change in speech perception and speech production. The focus lies on the main research question to what extent imitation capacities play a role in sound change.

8.4.1 Introduction and hypotheses

In her study on spontaneous phonetic imitation, Babel (2012) raised the question of individual differences in imitation capacities. She asked why some participants tended to imitate more than others and suggested that this fact could be linked to individual's differences in baseline production. Unfortunately, she did not collect information on participants' regional background and baseline productions (Babel, 2012, p.186). In contrast, the design of our study allows for such a comparison between participants' imitative capacities and their speech production and perception patterns.

The investigation of imitative capacities in Chapter 7 resulted in three measures. Two measures succeeded in grasping the variation within the forced imitation task: the range in forced imitation and the absolute forced imitation score (AIS) (see Section 7.2.2). These measures gave insights in the range participants are motorically able to produce and the quality of the imitation, respectively. Another measure came from the spontaneous imitation task: the convergence vs. divergence score (see Section 7.3.2). This task was meant to determine participants' willingness and readiness to imitate in a social interactive situation. Together, these three measures were representative for individual forced and spontaneous imitative capacities.

Two hypotheses are formulated about the link between imitation capacities and the state of an individual's speech perception and production in the context of sound change.

The first hypothesis about the role of imitation in sound change is the most straightforward. It is hypothesized that imitation is a reinforcing mechanism of sound change. Imitation is conceived as a capacity which allows speakers to reproduce sounds of their surroundings and, in this way, helps to spread a sound change. From this hypothesis follows that good imitators (i.e., the individuals who have developed this capacity best) are further in the sound change. Even if this hypothesis seems the most straightforward, it is not unproblematic. Indeed, it is possible that this imitation capacity which helps to reproduce and spread new variants is used at the same time to maintain older variants. If an individual is especially good at reproducing the sounds of the environment, he might do this both for new and older forms he hears in the input, and therefore just be good at copying the current stage of input he processes.

Therefore, we felt the need to propose a second opposite hypothesis: bad imitators are further in sound change. If we assume that production and perception are two separate systems (see *dual representation* (Garrett & Johnson, 2013) in Section 2.3.2), imitation can logically be conceived as the linking mechanism between these two systems. Hence, a good imitator is an individual with both a good perception and a strong link between his perception and production system. He will perceive the detailed acoustic structure of a segment and will subsequently successfully translate the details of this segment into a motoric production. In contrast, a bad imitator is less successful in this process, and this may occur for two possible reasons. The first possible explanation is that a bad imitator fails to hear the details of the target segment due to a weakened perception and consequently can't reproduce it correctly. In this case, a bad imitation is only due to the fact that perception is already engaged in the change. The second possible explanation for a bad imitation is a weak link between production and perception. In this case, a bad imitator fails to translate the details of the target segment (which he managed to perceive well) into production. In other words, the bad imitation − regardless of the quality of the perception system − follows from a weak link between the production and perception systems.

8.4.2 Evidence for a link between perception, production and imitation

In this section, we present evidence testing the two hypotheses cited above. It is tested whether good imitators are the most progressive participants in the sound change, or whether − on the contrary − bad imitators are the most progressive participants in the sound change.

Good imitators are progressive in change

In order to test the first hypothesis stating that good imitators are progressive in change, the measures of quality of imitation (the absolute forced imitation score (AIS) and the convergence vs. divergence score) were correlated with the measures of sound change completion (the general SC completion and the CompReg computed in Section 8.2).

Firstly, there was a significant relationship between the AIS and the sound change completion. The correlation between these two measures is plotted in Figure 8.11. The SC completion is plotted along the x-axis, and the AIS along the y-axis. Each symbol represents a participant (N=100 participants, n=200 observations). Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot.

In Figure 8.11, there is a positive correlation between the two measures. At the beginning of the sound change, most participants have an imitation score around 0, indicating that they succeeded in imitating fairly well. The more advanced the sound change, the higher the AIS. The more advanced a participant in the change, the worse he gets at the forced imitation task.

This relationship was tested statistically with the help of a linear regression. The AIS was significantly predicted by the SC completion score (Linear Regression t=2.583, p=.011), which could account for 3.26 % of the variance in imitation scores.

Secondly, we examine the advancement of participants in the change, but within their own region through the sound change completion score with respect to the own region (CompReg) (computed in Section 8.2). There was a significant relationship between the AIS and the CompReg. This relationship is plotted in Figure 8.12. The CompReg is plotted along the x-axis, and the AIS along the y-axis. Each symbol represents a participant (N=100 participants, n=200 observations). Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot.

Figure 8.11: Scatterplot of the absolute forced imitation score (AIS) as a function of the completion of the sound change, split up by variable (N=100 participants, n=200 observations).

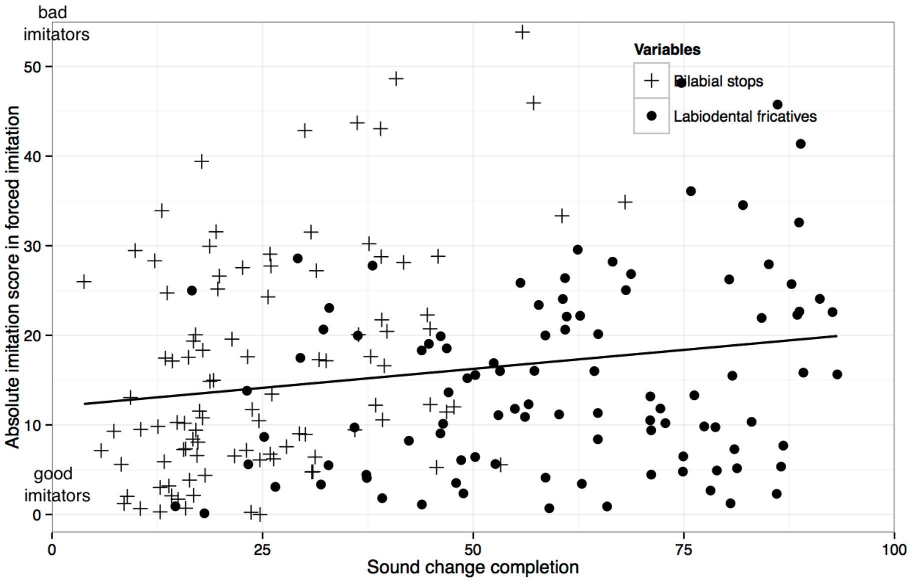
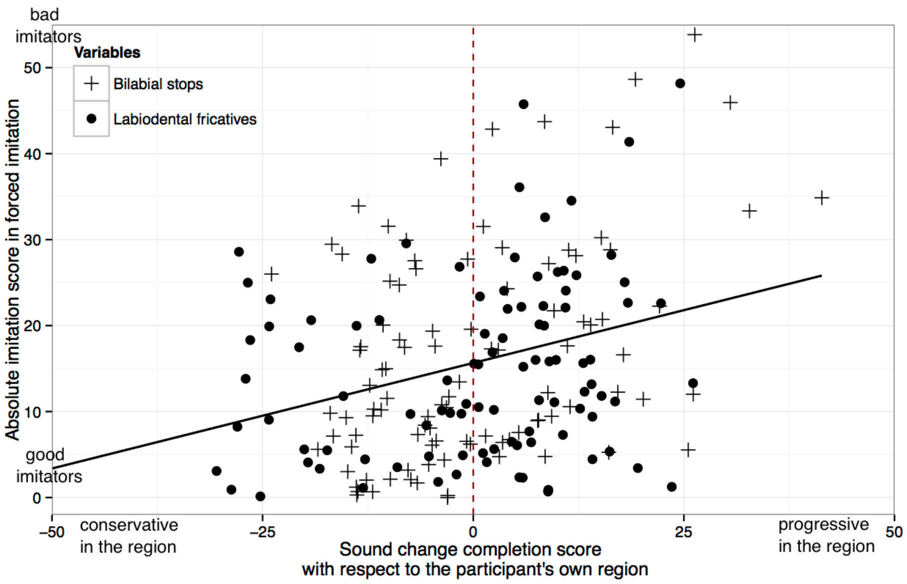


Figure 8.12: Scatterplot of the absolute forced imitation score (AIS) as a function of the sound change completion score with respect to the own region (CompReg), split up by variable (N=100 participants, n=200 observations).



In Figure 8.12, there is a positive correlation between the two measures. The more progressive the individual within his region (more positive completion scores), the higher the AIS, thus the worse his imitation capacities. The existing correlation seems to be partly triggered by a group of progressive individuals within their region (with a highly positive CompReg). These individuals clearly showed higher AIS scores, meaning that they are bad imitators.

This relationship was tested statistically with the help of a linear regression. The AIS was significantly predicted by the CompReg (LINEAR REGRESSION t=4.469, p<.001), which could account for 9.16 % of the variance in imitation scores. This model remained significant when the worst imitators were considered as outliers and were removed from the data.

In conclusion, we presented two pieces of evidence for the link between sound change completion and imitation. The pieces of evidence forced us to reject our first hypothesis stating that good imitators are progressive in change. In both analyses, the opposite pattern was shown: the best imitators were individuals who were at the beginning of the sound change and who were more conservative within their region, whereas bad imitators were further in the sound change and more progressive within their region. This confirms the second hypothesis.

The evidence used to reject the first hypothesis came from the analysis of the AIS, the measure of quality of the forced imitation. The convergence vs. divergence score, the measure issued from the spontaneous imitation task, did not show any relationship with the sound change completion.

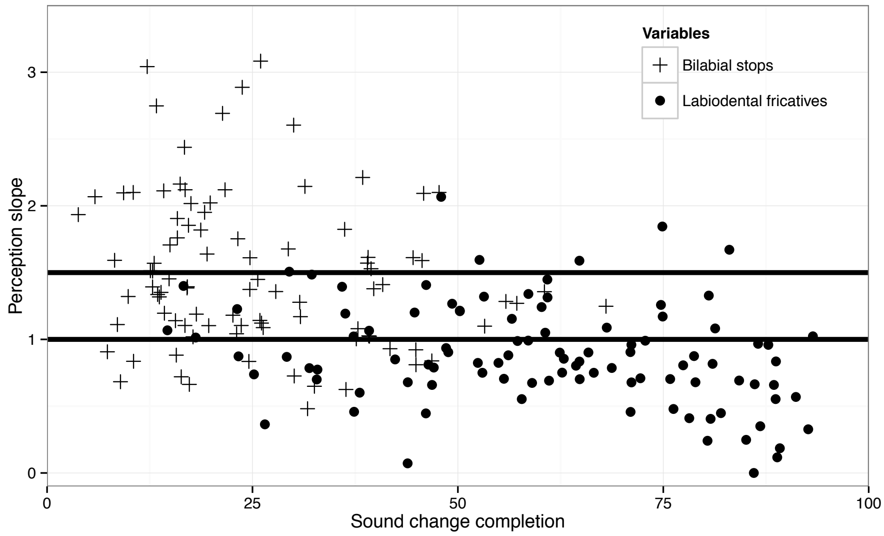
Bad imitators are progressive in change

In the previous analysis, we found evidence in favor of our second hypothesis: bad imitators are the most progressive participants in the sound change. In this section, we verify the possible underlying explanations for this fact. As explained in the introduction of this section, a bad imitation may have two different causes: a bad perception solely or a weak link production-perception.

Crucially, the state of perception system is intrinsically linked to sound change, as established in Section 8.2.2. The further in the change, the worse the perception (see Figure 8.3). Consequently, in order to isolate a possible role of the link between production and perception in the change, we need to isolate the cases where the quality of the imitation only depends on the strength of the link between production and perception. A subset of participants with comparable average perception for fricatives and stops was therefore selected. This selection is presented in Figure 8.13.

Participants with a slope in perception higher than 1 and lower than 1.5 were selected, resulting in a subset of 64 observations. Importantly, the subset was selected so that it consisted of a group of participants that showed no relationship between the perception/production and the sound change completion. In this way, if the analysis still shows a relationship between imitation capacities and sound change completion, this relationship can only be attributed to the strength of the link between production and perception.

Figure 8.13: Selection of the subset with comparable average perception (n=193 observations, 64 selected within the black lines).



Subsequently, the same analyses as in the previous section were run on this subset to test whether it is still true for these 64 participants that bad imitators are more advanced in sound change.

Firstly, it was tested whether there was a significant relationship between the AIS and the sound change completion. The correlation between these two measures is plotted in Figure 8.14. The figure represents the same relationship as Figure 8.11, but with the relevant subset of participants only (n=64). The SC completion is plotted along the x-axis, and the AIS along the y-axis. Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot. This relationship did not turn out significant (LINEAR REGRESSION t=1.649, p=.104). The relationship present for the analysis of all data was lost when looking at the subset of data only.

Secondly, we tested the relationship between the AIS and the advancement of participants in the change within their own region. The relationship between the AIS and the CompReg is plotted in Figure 8.15. In this figure, the CompReg is plotted along the x-axis, and the AIS along the y-axis. The figure represents the same relationship as Figure 8.12, but with the relevant subset of participants only (n=64 observations). Bilabial stops are represented by a plus sign and labiodental fricatives by a filled dot. The correlation between the two measures turned out significant. The more progressive the individual within his region (more positive completion scores), the higher the AIS, thus the worse his imitation capacities.

Figure 8.14: Scatterplot of the absolute forced imitation score (AIS) as a function of the completion of the sound change, split up by variable (n=64 observations).

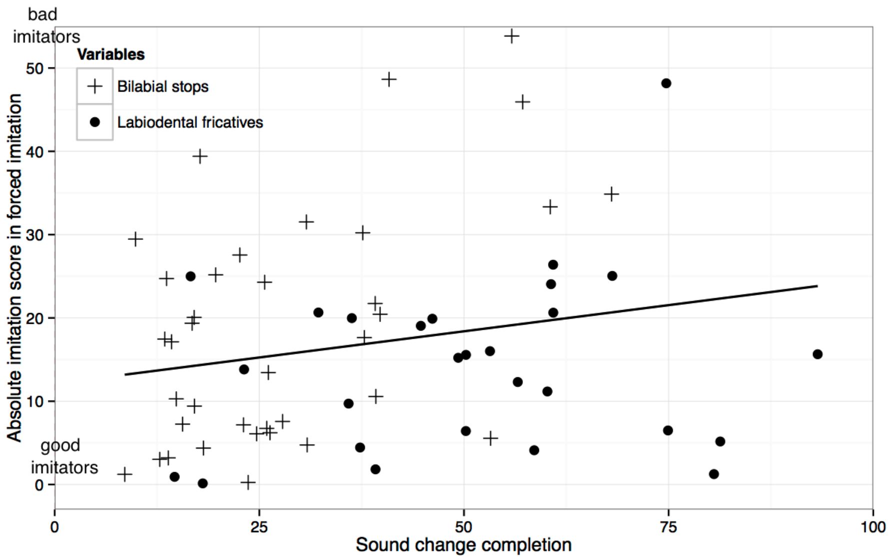
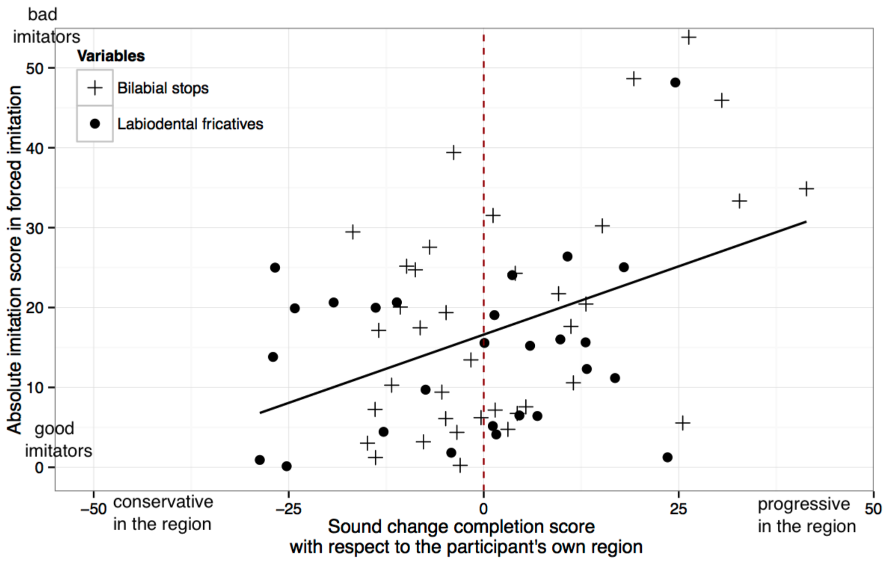


Figure 8.15: Scatterplot of the absolute forced imitation score (AIS) as a function of the sound change completion score with respect to the own region (CompReg), split up by variable (n=64 observations).



This relationship was tested statistically with the help of a linear regression. The AIS was significantly predicted by the CompReg (Linear Regression t=3.524, p<.001), which could account for 16.68 % of the variance in imitation scores. The relationship was maintained when looking at the subset of data only.

This evidence shows that − even if we control for perception − it is still true that bad imitators are further in the sound change. Being a bad imitator is thus more than just having a weakened perception. The strength of the link between production and perception is playing a crucial role in sound change.

8.4.3 Discussion and conclusion

In this section, we found different pieces of evidence for the link between speech perception, speech production and imitation. Specifically, two hypotheses about the link between imitation capacities and the state of individual speech perception and production in the context of sound change were tested.

The first hypothesis stated that good imitators are further in the sound change. This analysis of the data forced us to reject this first hypothesis, since the opposite pattern was shown: good imitators were more conservative in the sound change. Two types of correlation between production measures and imitation measures showed that the best imitators were individuals who are at the beginning of the sound change and who are more conservative within their region.

Furthermore, two underlying explanations were proposed for the fact that bad imitators are further in the change. The first possible explanation was that a bad imitation is due to a weakened perception (as perception is already engaged in the change). The second possible explanation for a bad imitation was a weak link between production and perception, regardless of the quality of the perception system. This second explanation was confirmed, since − for a subset of participants who were equally good at perceiving the sounds − it was still true that bad imitators are more progressive in the change than good imitators. Bad imitations are not just an artefact of a weakened perception, but are caused by the weakness of the cognitive link between production and perception. Consequently, the strength of this link appears to play an important role in sound change.

Even if we have seen that the quality of the link between production and perception plays an important role in sound change, it is still unclear why participants would differ with this respect. It might be the reflection of a priori individual differences (i.e. some individuals being better than others at imitating regardless of the target sound) or it might also be the case that the production-perception link is weakened in some individuals due to their involvement in the sound change. Further evidence in which general imitation capacities and imitation capacities related to the sound change in question are torn apart is required to answer this question.

Moreover, we speculate that the fact that bad imitators are further in the change is related to the internal nature of the sound changes we investigated. Indeed, it seems plausible that individual differences in the link between perception and production come into play when changes are internal to the language system and thus not induced by contact with another system. In external language change (i.e. contact-induced change) the reverse process might be the case: good imitators are further in the change. Further evidence is needed to verify this hypothesis.

All analyses presented in this section involved imitation measures issued from the forced imitation experiment. The measure related to the spontaneous imitation experiment did not show any relation to measures of perception and production.

In conclusion, it was shown that phonetic imitation capacities are related to individual states of production, as suggested by Babel (2012). It was demonstrated that the best imitators are found at the beginning of the change and have conservative production patterns. The further advanced in the change, the worse the imitation. Moreover, it was shown that bad imitations were caused by the weakness of the cognitive link between production and perception.

8.5 A model of sound change

The principal aim of this final section is to present a model for the role of the four different aspects (speech perception, speech production, language attitudes and phonetic imitation) in sound change. In the previous sections, these different aspects were assessed separately in order to determine their exact role in the context of sound change. In the current section, a more general perspective is taken by putting all aspects in one model.

The model aims to predict the sound change completion. The reasoning behind this model is the assessment of the extent to which an individual production can be predicted on the basis of his perception, language attitudes and phonetic imitation capacities. All measures recapitulated in Figure 8.1 (except the production measures which are logically intrinsically correlated with the sound change completion in order to avoid colinearity) were added in a STEPWISE LINEAR REGRESSION. The results of the linear regression model are presented in Table 8.2.

Table 8.2: Estimated parameters of the stepwise linear regression predicting the sound change completion.

Fixed effects	Estimates	S.E.	t value	p value	
Intercept	63.212	3.898	16.218	<.001	*
Perception slope	-22.310	2.585	-8.633	<.001	*
Solidarity attitude score	9.633	4.332	2.223	.027	*
Absolute imitation score (AIS)	0.384	0.128	2.990	.003	*

Three unrelated factors could significantly predict sound change completion: the slope in perception, the solidarity attitude score, and the AIS. The other measures either did not manage to predict variance in sound change completion or were correlated with the three cited measures and consequently did not succeed in explaining extra variance. Perception was a significant factor in the model (t=-8.633, p<.001). It showed that − if the individual perception is getting worse (less categorical) − the participant is more advanced in the change (a strong relationship described in Section 8.2). Secondly, the solidarity attitude score also turned out to be a significant predictor in the model (t=2.223, p=.027). In this case, the more positive the individual towards the changing variable, the more advanced he is in the sound change (see also Section 8.3). Finally, the AIS was the third significant factor (t=2.990, p=.003). As a recall, the lower the AIS, the better the imitation. It was shown that the worse the imitation, the more advanced the individual in sound change (see also Section 8.4). Interactions between these three predictors were removed since they did not improve the model.

Together these three measures can account for 31.97 % of the variance in sound change completion. Hence, when measuring categorical perception, attitudes towards the variable and forced imitation capacities of an individual, his advancement in the sound change (in production) can be predicted to a fair extent thanks to this model.

It is important to note that language attitudes along the other two dimensions (power and modernity) did not turn out to predict sound change completion significantly. Attitudes along the solidarity dimension appeared the most important. Similarly, the measure of divergence-convergence related to the spontaneous imitation experiment did not show any relation to the sound change completion. Consequently, it appeared that the motoric capacities, more than the willingness and readiness to imitate, play a role in sound change.

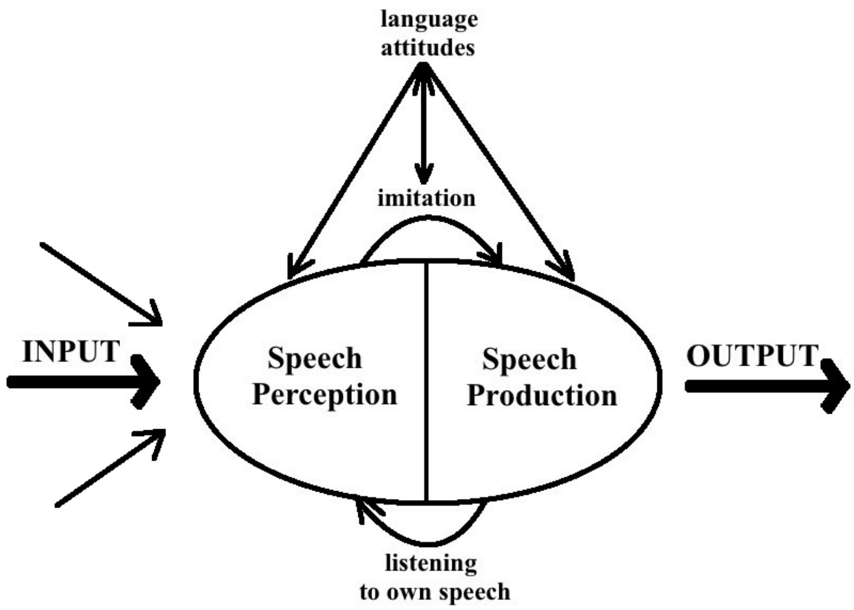
Based on the insights from this model, a schematic representation of the factors playing a role in sound change at the *individual level* is proposed. This schema is presented in Figure 8.16.

In this figure, the individual is represented by the central oval. Like Garrett and Johnson (2013), we adopt a *dual representation model* (see Section 2.3) in which perception and production are based on different sets of exemplars. Within the individual, production and perception are therefore represented as separate systems. Furthermore, the speech input processed by the perception system is much larger than the speech output an individual is able to produce, since listeners understand variants they cannot produce themselves.

Even through production and perception do not consist of exactly the same representational spaces, both systems are linked. This link is implemented by two mechanisms. The first mechanism relates speech perception to speech production through imitation processes. As seen in the data, individuals use the ability to translate the details of a segment they hear into motoric alternations of this segment in production, and this mechanism was shown to play a significant role in sound change, independently from the state of change in perception.

Indeed, sound change appeared to be led by bad imitators, because they have a weak link between production and perception.

Figure 8.16: Schematic representation of the factors playing a role in sound change at the individual level.



The second mechanism links speech production to speech perception and was already proposed in the literature. As proposed for example by Harrington et al. (2008), Kendall and Fridland (2012), Sumner and Samuel (2009) and Newman (2003), listeners do process speech through the filter of their own production. Since people are likely to have heard their own productions more than those of any other single individual, their productions are likely to highly influence their perceptual patterns. Our data also confirms the fact that the speech norms used by an individual himself constitutes a crucial part of his auditory experience.

On top of these relations, we showed that language attitudes play a significant role in sound change. Language attitudes significantly contribute to strengthen the process of devoicing, showing a clear relationship between production of the variables and its evaluation.

Concerning the relationship between language attitudes and imitation capacities, this study did not succeed in providing additional information. Previous studies showed that attitudes play an important role in the mechanism of imitation. For example, Pardo (2006) insisted on the effect that the social value of certain exemplars has on the likelihood that they will be imitated.

Reversely, imitation also positively affects language attitude. Adank, Stewart, Connell, and Wood (2013) for instance, recently showed that imitating a speaker had a positive effect on the perceived social attractiveness of that speaker. As explained in Chapter 2, this important multiplex role of language attitudes in sound change has not yet been implemented well in exemplar-based theories, and require attention in further studies.

In conclusion, we have shown that four factors turned out to play a significant role in sound change. As indicated by the global linear regression model, the perception system, the mechanism of imitation linking the production and perception systems and language attitudes all influence the process of sound change which expresses itself in speech production. These insights – together with the insights provided by previous studies – allow us to provide a model for sound change at the *individual level*.

Once it is clear how it works at the individual level and what the role of the different factors is, sound change might more easily be conceived as the iterative process in which the output (the production) changes incrementally every time a speaker (as described in the model) speaks to a listener (i.e., an other individual with a comparable system).

CHAPTER 9

Conclusion

The actuation problem in the study of sound change is the question as to why changes take place in a particular language at a given time. This dissertation aimed at contributing to potential solutions for this problem and gaining a better understanding of the phenomenon of sound change.

As explained in Chapter 1, the approach of the current study to sound change crucially differed from previous research in a number of ways. Firstly, our research aimed to step away from the notions of innovation and spread in sound change and the understanding of sound change as a two-step process associated with them. The process of actuation was reconsidered as an iterative process at the *individual speaker level*, in which minimal target changes incrementally accumulate every time a speaker speaks to a listener. Hence, sound change was redefined as a purely synchronic process, residing in the variability inherent to the individual systems of speech production and speech perception.

Secondly, this research aimed at reconciling the fields of sociolinguistics, phonetics and psycholinguistics by making use of their respective strengths and compensate for their weaknesses. The two-levels design of the study incorporated the tradition of cross-sectional studies in sociolinguistics. Experimental methods and measures from psycholinguistics, social psychology and phonetics were adopted in the design. At the same time, it succeeded in giving weight to individual speakers and their linguistic histories. The study put the focus both on intra- and inter-speaker variability.

Thirdly, the study concentrated on two sound changes in progress in the Dutch language. These sound changes were similar in many ways, but crucially differed in their degrees of completion. The first was a sound change in an advanced stage (i.e., the devoicing of labiodental fricatives), and the second was a sound change which is still incipient (i.e., the devoicing of bilabial stops). The juxtaposition of these two changes allowed to observe the iterative process of actuation at each step of completion.

The implementation of these research aims led to the development of a comprehensive experimental design in which four crucial aspects of the linguistic competence involved in sound change were tested on the same pool of participants. These four aspects of the linguistic competence were the processes of *speech perception* and *speech production*, and the mechanisms by which linguistic variables are evaluated and imitated, *language attitudes* and *phonetic imitation* respectively. A *multi-experimental one-shot cross-sectional study* was designed so that the speech production, speech perception, imitation capacities and attitudes patterns of individual speakers within a language community could be linked together.

The study was articulated around two participant levels: the *group* and the *individual*.

- The *group level* consisted of five regions within the Dutch-speaking area of Belgium and the Netherlands. These different geographical regions were chosen to represent different stages of sound change. At the group level, comparisons were made between these regions.
- The *individual level* was represented within each region by a sample of twenty participants. For the selection of these participants, the factors age and level of education were kept constant. At the individual level, comparisons were made between participants within the same region and across regions.

The study was implemented in the framework of recently developed exemplar-based theories. In this theoretical framework, an individual's linguistic knowledge consists of a detailed record of all of the exemplars of the words and the phonemes he has already heard. Moreover, exemplar representations continue to be updated throughout the lifespan of an individual. Given that no two speakers are ever exposed to the same language input throughout their life course, perceptual patterns are likely to vary across individuals. In this manner, exemplar-based theory succeed in modeling both individual variation and the influence of the ambient language on the individual, so that life-span change can easily be conceptualized.

Furthermore, an exemplar-based model can account for the gradual character of change, so that sound change can be conceived as an iterative process. It anchors social indexation together with linguistic representation. Additionally, the perceptual activation of a set of exemplars may lead to a shift in production, showing that the process of imitation is also inherent to the model.

In view of its theoretical benefits, this study was grounded within exemplar-based theory. Five main research questions related to the role of different aspects of the linguistic competence in sound change were formulated.

In this concluding chapter, the most important results of the previous chapters are summarized, resulting in an integrative account of sound change in Section 9.1. In Section 9.2, a range of limitations in the current study are presented. Potential steps for future research are outlined in Section 9.3.

9.1 Summary of the results

This section summarizes the main findings of this dissertation and integrates these findings into possible answers for our five main research questions.

In *Chapter 4*, an extensive production experiment was conducted, aiming at determining the stages of sound change of the two target variables in each region and for each individual. As predicted from previous production studies, differences were found between regions (at the *group level*) in the amount of devoicing in fricatives. In all regions, starting with the Netherlandic regions, the contrast in initial labiodental fricatives is fading. This sound change appeared to be even further advanced than reported a decade ago by Kissine et al. (2003, 2005). Stops, on the contrary, did not directly show the expected regional differences. Patterns of stop devoicing only appeared when looking at the *individual level*. In several regions, a minority of individuals showed clear patterns of stop devoicing, mostly along the voicing dimension.

The juxtaposition of the production patterns of both variables gave us precise insights into the relationships between differences between categories and range of these categories on the one hand, and the development of sound change on the other hand. These insights allowed us to redefine the mechanism of category merging in production. Before the change, the situation is stable with two separated categories of comparable range. At the change onset, the merging category (in our case the voiced category) increases its range, while the difference between categories decreases in a linear manner. In the second phase of the change however, the range of the merging category starts to decrease, whereas the difference between categories keeps shrinking. At the change completion, both categories are merged (no or very small difference between categories) and have a comparably small range.

RQ 1: What is the role of speech perception in sound change?

In *Chapter 5*, the perception of labiodental fricatives and bilabial stops was investigated through two perception experiments: an identification task and a similarity judgment task. The identification task was aimed at gaining quantitative insights in the regional and individual perceptual patterns. The similarity judgment task aimed at studying phonetic perception in a more qualitative manner without requiring that listeners identify sounds. In these experiments, we examined whether differences in the perception of labiodental fricatives and bilabial stops could be found between regions within the Dutch language area (at the *group level*) and between individuals within these regions (at the *individual level*). We succeeded in showing regional differences in the perception of phonemes undergoing sound change (i.e., fricatives). The regions where devoicing was most advanced in production turned out to be also the regions where perception was the least categorical. Listeners presented differences in perception, which were thus in line with speech production patterns (reported in Chapter 4). Therefore, the regional differences shown in the perception of fricatives established the first link between speech production and speech perception. As a result, variation appears to be involved in sound change is not limited to production, but is also a perceptual phenomenon.

RQ 2: What is the link between variation in speech perception and variation in speech production?

The investigation of evidence in favor of the link between speech perception and speech production concentrated on the group level in *Chapter 5*, and was taken further through the analysis of patterns at the individual level in *Chapter 8*. The analyses provided results which pointed towards a clear link between speech perception and speech production. For both variables, it was shown that the more devoicing in an individual's production, the least categorical his perception. Even though production and perception are thought to be represented by two different systems (*dual representation*), these systems are linked with each other, and this link is crucial for sound change.

RQ 3: Does change in speech perception precede change in speech production?

In *Chapter 8*, the order in which the systems of perception and production evolve in sound change was under investigation. The analyses seemed to reveal that – at the beginning of a sound change – perceptual patterns are adjusted first. These perceptual adjustments seemed to have consequences on speech production patterns. The language user first incorporates the new form in his perception and only later, if at all, he starts to use the form in production. Once the devoicing is launched in production, its end point is the full devoicing of the category. However, it was shown that the change in perception is not as far proceeded as in production: at the end of the sound change, perception 'lags behind'. Participants are still able to hear some categorical contrast in a variable for which they no longer contrast in their own production.

In conclusion, it appeared that changes in speech perception precede change in production. However, speech perception appears to lag behind speech production when the sound change is reaching completion.

RQ 4: What is the role of language attitudes in sound change?

In *Chapter 6*, it was investigated to what extent listeners make judgments about the indexical properties of a sound change in progress. The goals were to determine to what extent a changing sound triggers prestige across the different groups of listeners (at the *group level*) and among individual listeners (at the *individual level*). For the devoicing of fricatives, there was clear evidence that the devoicing of labiodental fricatives is considered as the prestigious variant. At the *group level*, some regional differences were revealed. The regions with the least fricative devoicing showed more positive attitudes towards devoicing than regions where the devoicing is more advanced. It seemed that participants coming from regions where the change is not so advanced were more inclined to associate the new form with positive attitudes. In that sense, having positive attitudes towards a variable seems to be a sign for potential change towards this variable. In contrast, more evidence is needed to claim that the devoicing of stops is associated with positive attitudes. Since the devoicing of stops is an incipient change, it is possible that the change is not yet far enough, so that participants would have encoded a social index together with the phonetic information.

In *Chapter 8*, attitude, production, and perception results were brought into relationship. At the *individual level*, the hypothesis that language attitudes play a role in speech production in the context of sound change was confirmed. The results showed that the more devoicing of a variable, the more positive devoicing is rated (along the solidarity dimension). Second, the hypothesis that language attitudes play a role in speech perception was not directly confirmed.

In conclusion, we showed that language attitudes towards a variable undergoing change does not seem to be present from the beginning of the sound change onwards. The positive evaluation of a change seems to arise when the system of speech production is already undergoing the change. Language attitudes thus develop during the process of sound change and seem to be a reinforcing mechanism. This result is in line with previous sociolinguistic research which traditionally investigates the role of language attitudes in the process of spread of the change (and not so much in the innovation of a sound change since it is too early to observe attitudes at that stage) and which views the development of language attitudes as a phenomenon related to the group level.

RQ 5: What is the role of phonetic imitation in sound change?

In *Chapter 7*, it was investigated through two imitation experiments (a forced and a spontaneous imitation task) to what extent individuals imitate changing sounds (at the *individual level*) and whether this ability differs across regions (at the *group level*). The goal of the forced imitation task was to give insights into motoric production capacities. By forcing participants to imitate a model speaker as well as they could, their maximal range of production was investigated. With exception of a few individuals, all participants managed to produce the whole range of voicing in fricatives and stops when they were instructed to. Even participants from regions where productions were devoiced turned out to be successful at imitating fully voiced sounds. The spontaneous imitation task concentrated on participants' willingness and readiness to devoice in social interaction and succeeded in demonstrating that simple exposure to model speakers induces imitation of devoiced fricatives and stops. At the *group level*, all regions showed imitation, but slightly differed in the amount of spontaneous imitation. At the *individual level*, a small number of participants clearly showed increasing imitation patterns during the course of the imitation.

In *Chapter 8*, the individual imitation results were examined in relationship with insights on the link between speech perception and speech production. It was shown that bad imitators were more progressive in the change than good imitators. Correlations between production measures and imitation measures showed that the best imitators were individuals who are at the beginning of the sound change and who were more conservative within their region. Individual differences in imitation capacities were attributed to differences in the strength of the link between the perception and productions systems, regardless of the quality of the perception system.

In conclusion, phonetic imitation capacities were shown to be related to individual states of production, as suggested by Babel (2012). It was demonstrated that the best imitators are at the beginning of the change and show conservative production patterns. The further advanced in the change, the worse the imitation capacities. Bad imitators seemed to lead the change, because of their weak link between perception and production.

9.2 Limitations of the study

This section offers some thoughts on possible limitations of this study. First, some possible limitations related to the methodology of the study are presented. Second, the question is raised whether the devoicing of bilabial stops is a real sound change.

9.2.1 Methodological limitations

Firstly, some methodological limitations are related to the use of laboratory settings for the study of a phenomenon that is social and interactional in nature. It is important to remember that the artificiality of the setting in which the experiments were conducted may produce unnatural behaviors or experimental artifacts (Foulkes, Scobbie, & Watt, 2010). Laboratory materials may also lack in naturalness. The design of the current study attempted to circumvent these potential problems by eliciting both read and spontaneous speech and by using phonetically manipulated speech instead of synthesized stimuli. The task in our design, which was the most sensitive to these methodological concerns, was the spontaneous imitation task. Many parameters in this task were difficult to control for in laboratory settings (e.g., participant's evaluation of the interactional degree of the task, their attitudes towards the two reference voices, their readiness to imitate in a minimally interactive situation, etc.). The design of this task would definitely be improved if it could be run in a more real-life situation.

Secondly, as usual in this type of research, our participants were selected among the university student population of Belgium and the Netherlands. University students make often convenient and readily available participants. In our design, the advantage of using a student population was related to the sociolinguistic requirements of homogeneity of age and educational level. Moreover, the student population is more acquainted with the situation of experimental testing in a lab. Hence, their comprehension of the task instructions is easier. However, our impression was that it is more common for students to participate in a paid experiment in the Netherlands than in Belgium. The extent to which this fact might have influenced our results is unclear, but should be taken into account in follow-up studies.

9.2.2 The devoicing of stops as an incipient sound change

It is necessary to raise the question whether our second variable in the study, the devoicing of bilabial stops, is a real sound change. As explained in previous chapters, the devoicing of bilabial stops was chosen to represent an incipient change in the Dutch language. Basing our hypotheses on previous studies (e.g., Smits & van Alphen, 2004; Ziliak & Van de Velde, in progress), we expected to find regional variation in the realizations of stops. In contrast, only individual patterns of devoicing could be found.

The question is therefore raised whether the devoicing of stops is really a sound change in a very incipient form for which we did not manage (yet) to find patterns of regional variation or whether it is merely stable variation (not leading to sound change). The problem in this case is that − because the sound change is incipient (if we assume that it is a sound change) − it is situated in the first phase of the S-shape typically representative for the course of sound change. In this first phase, patterns are often subtle and cannot be associated with certitude to change or stable variation. Moreover, our study was designed to test for one single source of variation: regional variation. Hence, we have no information regarding possible age-related or socially-relation variation, which might have provided additional evidence for change in bilabial stops.

Nevertheless, two arguments plead for the consideration of the devoicing of stops as an incipient sound change. First, we saw in Chapter 4 that the devoicing of stops did not show any particular pattern of stylistic variation. Therefore, it could be concluded that there is no stable pattern of style shifting. Second, it was shown in Chapter 5 that stops also present a great variability in perception. Since the language user first incorporates a new form in his perception, the fact that participants' perception of stops showed great variation seems more compatible with a variable being in an incipient phase of change than with stable variation patterns. As explained in Chapter 8, conclusions on the link between the four different aspects of the linguistic competence still held when we considered the fricative results only.

In conclusion, the paradox of the current study was to find a sound change incipient enough to reveal the nascent patterns of change that other studies have failed to observe. However, there is a risk at that stage of confounding sound change with stable patterns of variation. Future studies will help to confirm that bilabial stops are indeed undergoing devoicing in Dutch.

9.3 Future research

This study was innovative in a number of aspects (e.g., focus on regional variation instead of age variation, one-shot design, sound changes in different degrees of completion, etc.). The conclusions reached here may motivate researchers to adopt or expand these aspects in further studies.

An first line of expansion is related to phonetic dimensions. We started the analyses with two phonetic dimensions, but rapidly selected one single dimension, voicing, as it turned out to be the most important cue for the concerned variables. Further studies may however profit from insight about the multi-dimensional aspect of phonetic change. This would help to investigate whether there are new dimensions replacing the older ones helping the listener/speaker to maintain the contrast in perception and in production.

The two variables undergoing sound change chosen for this study were very similar (i.e., both were word-initial consonantal mergers undergoing a process of devoicing). Their similarity was an advantage for the experimental design and the interpretation of the results. We however feel the need to extend our analyses to other types of sound changes, which could be analyzed in a similar way in follow-up studies.

A crucial step would be to test our conclusions and the predictions that follow from them in a longitudinal study. In the frame of a doctoral dissertation, looking at the developments of individuals' speech production and perception over time is not feasible, because of the timing of change. However, a larger-scaled research project could examine the evolution of these aspects longitudinally for a series of individuals and their response to changes in the environment speech.

Related to this, examining the link between speech production and speech perception from the acquisitional perspective could contribute to additional insights and reinforce the theoretical anchorage. Studies on sociolinguistic variation and change in children are rare. Such a line of research would bring together insights on processing, acquisition and variation.

This study was grounded within exemplar-based theories. Insights from these models allowed for reconsidering the process of change as intrinsically iterative and interpreting the results accordingly. Our research, however, raised important questions on how some crucial aspects in sound change are encoded cognitively. For instance, it is still unclear how socio-indexical information is represented cognitively and at which level it is stored and activated. How can the awareness and the salience of a linguistic variable be accounted for in exemplar theory? How do we model the influence of speech perception on speech production in sound change? Future research on these topics will help to elaborate the existing models of exemplar theory.

In addition to theoretical matters, the results of this dissertation also question the future of laboratory linguistics. In this dissertation, we have seen that the methodology of experimental testing in the lab suffered from some limitations (e.g., for the study of spontaneous imitation) in that it was not perfectly tailored to the interactional and dynamic nature of the research object. In our view, the study of the production/perception interface and its role in sound change could therefore take two different directions. Either we lay the focus on more dynamic experimental settings with possible disadvantages for the quality of phonetic measurements. Or we place greater emphasis on articulatory measurements with the help of newly developed tools in order to improve our phonetic measures before adding the social and interaction aspect to the research.

9.4 Conclusion

This dissertation reported on a series of experiments aimed at gaining a better understanding of the phenomenon of sound change and its actuation. The results of the experiments provided insights into the role of different aspects of the linguistic competence in sound change.

Following Garrett and Johnson (2013), we adopted the dual representation model in which speech perception and speech production are based on different sets of exemplars. Within each individual language user, the perceptual space is separated from the production space. Since listeners understand variants they cannot produce themselves, the perceptual space is larger than the production space.

Even through production and perception form separate systems, both systems are linked by at least two mechanisms. First, individuals use their ability to translate the details of a segment they hear into production through the mechanism of imitation. Second, listeners process speech through the filter of their own production. Positive language attitudes associated with specific linguistic features reinforce these mechanisms.

Considering the model of individual functioning, our research is pleading for the conception of sound change as the iterative process by which an individual's output (the production) changes incrementally every time he speaks to a listener. The actuation of change is directly linked to individual differences in the quality of the perception and production systems and differences in the strength of the link between these systems. The discrepancy that exists between the systems of perception and production is at the core of the process of sound change. We propose that actuation happens *within the individual* every time his perception and production try to align with each other. Future study will help to define how the directionality of sound change is influenced by external factors (external in the sense that they do not proceed at the individual level), like articulatory, distributional and functional constraints on the existing language system.

This dissertation has contributed to the study of sound change and the actuation problem through a sociophonetic investigation of the role of four aspects of the linguistic competence engaged in sound change. It has extended our understanding of processes taking place at the individual and at the group level and managed to bridge the gap between these levels.

References

Abbs, J. H. (1986). *Invariance and variability in speech production: A distinction between linguistic intent and its neuromotor implementation.* Lawrence Erlbaum Associates.

Adank, P., Stewart, A. J., Connell, L., & Wood, J. (2013). Accent imitation positively affects language attitudes. *Frontiers in Psychology*, *4*.

Ainsworth, W., & Paliwal, K. (1984). Correlation between the production and perception of the English glides/w, r, l, j/. *Journal of Phonetics*, *12*(3), 237–243.

Ali, A., Van der Spiegel, J., & Mueller, P. (2001a). Acoustic-phonetic features for the automatic classification of fricatives. *Journal of the Acoustical Society of America*, *109*(5), 2217–2235.

Ali, A., Van der Spiegel, J., & Mueller, P. (2001b). Acoustic-phonetic features for the automatic classification of stop consonants. *IEEE Transactions on Speech and Audio Processing*, *9*(8), 833–841.

Allen, H. W., & Herron, C. (2003). A mixed-methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals*, *36*(3), 370–385.

Allport, G. W. (1935). *Attitudes.* Clark University Press.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Babel, M. (2009). Selective vowel imitation in spontaneous phonetic accommodation. *UC Berkeley Phonology Lab Annual Report (2009)*, 163–194.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(04), 437–456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189.

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, *55* (2), 231–248.

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, *5* (1), 123–150.

Bailey, P. J., & Haggard, M. P. (1973). Perception and production: Some correlations on voicing of an initial stop. *Language and Speech*, *16* (3), 189–195.

Baker, A. (2008). Addressing the actuation problem with quantitative models of sound change. *University of Pennsylvania Working Papers in Linguistics*, *14* (1), 3.

Barr, D. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, *4* .

Barr, D., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68* (3), 255–278.

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes.*

Bauer, L. (1985). Tracing phonetic change in the received pronunciation of British English. *Journal of Phonetics*, *13* (1), 61–81.

Beddor, P., & Coetzee, A. (2014). Sound change propagation: the relation between perception and production in individual language. *Presentation at the workshop Sound Change in Interacting Human Systems. Berkeley.*

Bell, A. (1984). Language style as audience design. *Language in Society*, *13* (02), 145–204.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some relationships between speech production and perception. *Phonetica*, *36* (6), 373–383.

Blaauw, E. (1995). *On the perceptual classification of spontaneous and read speech.* University Utrecht.

Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer.* Retrieved from `http://www.praat.org/`

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. *Contrast in phonology theory, perception, acquisition*, 145–172.

Bosker, H. R. (2014). *The processing and evaluation of fluency in native and non-native speech.* Utrecht University.

Boves, T. (1992). *Speech accommodation in co-operative and competitive conversations.* Radboud Universiteit Nijmegen.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, *75* (2), B13–B25.

Brown, R., & Gilman, A. (1960). *The pronouns of power and solidarity* (S. T, Ed.). MIT Press.

Campbell-Kibler, K. (2007). Accent, (ing), and the social logic of listener perceptions. *American Speech*, *82* (1), 32–64.

Casserly, E. D. (2010). Perceptual similarity across multiple sociolinguistic variables. In *IULC Working Papers Online* (Vol. 10).

Cassier, L., & Van de Craen, P. (1986). Vijftig jaar evolutie van het Nederlands. In J. Creten, G. Geerts, & K. Jaspaert (Eds.), *Momentopnamen van de sociolinguïstiek in België en Nederland* (pp. 59–73). Acco.

Chistovich, L., Fant, G., de Serpa-Leitao, A., & Tjernlund, P. (1966). Mimicking of synthetic vowels. *Quarterly Progress and Status Report, Speech Transmission Lab, Royal Institute of Technology, Stockholm*, *2*, 1–18.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*(2), 121–157.

Christiner, M., & Reiterer, S. M. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, *4*.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 521–533.

Clopper, C. G., Conrey, B., & Pisoni, D. B. (2005). Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology*, *24*(2), 182–206.

Clopper, C. G., & Pisoni, D. B. (2004). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change*, *16*(1), 31–48.

Clopper, C. G., & Pisoni, D. B. (2008). Perception of dialect variation. *The Handbook of Speech Perception*, 313.

Cohen, A., Ebeling, C. L., Fokkema, K., & Holk, A. G. F. (1961). *Fonologie van het Nederlands en het Fries.* Martinus Nijhoff.

Corvalan, C. S. (1984). The social profile of a syntactic-semantic variable: Three verb forms in Old Castile. *Hispania*, 594–601.

Coupland, A., Peter, W., & Nikolas, G. (1999). Dialect recognition. *Handbook of Perceptual Dialectology*, *1*, 345.

Croft, W. (2010). The origins of grammaticalization in the verbalization of experience. *Linguistics*, *48*(1), 1–48.

Dancy, C., & Reidy, J. (2007). *Statistics without maths for psychology.* Pearson Education.

Debrock, M. (1977). An acoustic correlate of the force of articulation. *Journal of Phonetics*, *5*, 61–80.

Delvaux, V., Demolin, D., & Soquet, A. (2004). Sound change as a result of mimetic interactions between speakers: implementation of imitation in the laboratory. In *9th Conference on Laboratory Phonology* (pp. 24–26).

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*(2-3), 145–173.

Díaz-Campos, M., & Killam, J. (2012). Assessing language attitudes through a matched-guise experiment: The case of consonantal deletion in Venezuelan Spanish. *Hispania*, *95*(1), 83–102.

Dijksterhuis, A., & Bargh, J. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology*, *33*, 1–40.

Di Stefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1–11.

Docherty, G. J. (2007). Speech in its natural habitat: accounting for social factors in phonetic variability. *Laboratory Phonology*, *9*, 1–36.

Donaldson, B. (1983). *Dutch. A linguistic history of Holland and Belgium.* Martinus Nijhoff.

Drager, K. (2005). From bad to bed: the relationship between perceived age and vowel perception in New Zealand English. *Te Reo*, *48*, 55–68.

Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass*, *4*(7), 473–480.

Drager, K. (2011). Speaker age and vowel perception. *Language and Speech*, 1-23.

Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology*, *4*.

Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, *25*(5), 582–602.

Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern british english sentences. *Journal of the Acoustical Society of America*, *115*(1), 352–361.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*, *121*(6), 3814–3826.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological review*, *116*(4), 752–782.

Ferguson, S., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America*, *135*(6), 3570–3584.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438.

Foulkes, P., Docherty, G., & Jones, M. (2010). Analysing stops. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics* (p. 58-71). Routledge.

Foulkes, P., Scobbie, J., & Watt, D. (2010). Sociophonetics. *The Handbook of Phonetic Sciences, Second Edition*, 703–754.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*(1), 3–28.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, *49*(3), 396–413.

Fox, R. A. (1982). Individual variation in the perception of vowels: implications for a perception-production link. *Phonetica*, *39*(1), 1–22.

Fox, R. A. (1983). Perceptual structure of monophthongs and diphthongs in english. *Language and Speech*, *26*(1), 21–60.

Fox, R. A. (1985). Auditory contrast and speaker quality variation in vowel perception. *Journal of the Acoustical Society of America*, *77*(4), 1552–1559.

Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America*, *97*(4), 2540–2551.

Fridland, V., Bartlett, K., & Kreuz, R. (2005). Making sense of variation: Pleasantness and education ratings of southern vowel variants. *American Speech*, *80*(4), 366–387.

Fridland, V., & Kendall, T. (2012). Exploring the relationship between production and perception in the mid front vowels of US English. *Lingua*, *122*(7), 779–793.

Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (2000). Adults' perception and production of the English vowel /i/. *Journal of Speech, Language, and Hearing Research*, *43*(1), 129–143.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171–189.

Gandour, J., & Harshman, R. (1977). Cross language differences in tone perception. *Journal of the Acoustical Society of America*, *61*(S1), 1–33.

Garrett, A. (2005). Attitude measurements. In U. Ammon, N. Dittmar, K. Mattheier, & P. Trudgill (Eds.), *Sociolinguistics. An International Handbook of the Science of Language and Society* (Vol. 2, pp. 1251–1260). De Gruyter.

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. *Origins of sound change: Approaches to phonologization*, 51–97.

Gentilucci, M., & Bernardis, P. (2007). Imitation during phoneme production. *Neuropsychologia*, *45*(3), 608–615.

Giles, H., Coupland, N., & Coupland, J. (1991). *Contexts of accommodation: Developments in applied sociolinguistics* (Vol. 1). Cambridge University Press.

Giles, H., & Smith, P. (1979). *Accommodation theory: Optimal levels of convergence* (Vol. Language and Social Psychology; H. Giles, S. Clair, & R. N., Eds.). Baltimore: Basil 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.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*(5), 1166–1183.

Goldinger, S. D. (1997). *Speech perception and production in an episodic lexicon* (K. Johnson & J. Mullennix, Eds.). Academic Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, *105*(2), 251.

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review*, *11*(4), 716–722.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. *Phonetics and phonology in language comprehension and production: Differences and similarities*, 159–207.

Goossens, J. (1974). *Historische phonologie des niederländisches*. Niemeyer.

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, *32*, 141–174.

Gregory, S. W. (1990). Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior*, *14*(4), 237–251.

Grondelaers, S., & Van Hout, R. (2010). Do speech evaluation scales in a speaker evaluation experiment trigger conscious or unconscious attitudes? *University of Pennsylvania Working Papers in Linguistics*, *16*(2), 93–102.

Grondelaers, S., & van Hout, R. (2013). Do we need corpus or speaker evaluation evidence for ongoing syntactic change? In *Presentation at ExApp Copenhague.*

Grondelaers, S., Van Hout, R., & Steegs, M. (2010). Evaluating regional accent variation in Standard Dutch. *Journal of Language and Social Psychology*, *29*(1), 101–116.

Grosvald, M. (2009). *Long-distance coarticulation: A production and perception study of English and American Sign Language*. University of California Davis.

Grosvald, M., & Corina, D. (2012). *The production and perception of subphonemic vowel contrasts and the role of the listener in sound change* (M. Solé & D. Recasens, Eds.). John Benjamins Publishing.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*(2), 1111–1121.

Gussenhoven, C. (1999). Illustrations of the IPA: Dutch. In *Handbook of the International Phonetic Association* (p. 74-77). Cambridge University Press.

Gussenhoven, C., & Bremmer, R. H. (1983). Voiced fricatives in Dutch: sources and present-day usage. *North-Western European Language Evolution*, *2*, 55–71.

Hamann, S., & Sennema, A. (2005). Acoustic differences between German and Dutch labiodentals. In *ZAS papers in linguistics* (Vol. 42, pp. 33–41).

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, *34*(4), 439–457.

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. *Laboratory Phonology*, *9*, 125–144.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, *123*(5), 2825–2835.

Harrington, J., Kleber, F., & Reubold, U. (2012). The production and perception of coarticulation in two types of sound changes in progress. *Speech Planning and Dynamics*, 39–62.

Harrington, J., Palethorpe, S., & Watson, C. (2005). *Deepening or lessening the divide between diphthongs? An analysis of the Queen's annual Christmas broadcasts* (W. Hardcastle & B. J. Mackenzie, Eds.). Routledge.

Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English? *Nature*, *408*(6815), 927–928.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*(3), 373–405.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4), 865–892.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3), 351–379.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484.

Healy, A. F., & Repp, B. H. (1982). Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(1), 68–80.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, *119*(5), 3059–3071.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). Codes and their vicissitudes. *Behavioral and Brain Sciences*, *24*(05), 910–926.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge University Press.

Hox, J. (2010). *Multilevel analyses: Techniques and applications*. Routledge.

Hruschka, D., Christiansen, M., Blythe, R., Croft, W., Heggarty, P., Mufwene, S., et al. (2009). Building social cognitive models of language change. *Trends in Cognitive Sciences*, *13*(11), 464–469.

Huang, T. (2004). *Language-specificity in auditory perception of Chinese tones.* Ohio State University.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*(12), 942–951.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*(1), 553–562.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/and /l/. *Journal of the Acoustical Society of America*, *99*(2), 1130–1140.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), 47–57.

Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434–446.

Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialogue* (Vol. 8). Academic Press New York.

Janson, T. (1979). Vowel duration, vowel quality, and perceptual compensation. *Journal of Phonetics*, *7*(2), 93–103.

Janson, T. (1983). Sound change in perception and production. *Language*, 18–34.

Janssens, G., & Marynissen, A. (2005). *Het Nederlands vroeger en nu.* Acco.

Jassem, W. (1979). Classification of fricative spectra using statistical discriminant functions. In B. Lindblom & S. Öhman (Eds.), (pp. 77–91). New York: Academic Press.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, 145–165.

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*(1), 127–136.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 505–528.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384.

Kager, R., van der Feest, S., Fikkert, P., Kerkhoff, A., & Zamuner, T. (2007). Representations of [voice]: Evidence from acquisition. In J. van de Weijer & E. van der Torre (Eds.), *Voicing in dutch: (de)voicing - phonology, phonetics, and psycholinguistics* (pp. 41–80). John Benjamins.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286-319.

Kendall, T., & Fridland, V. (2012). Variation in perception and production of mid front vowels in the US Southern Vowel Shift. *Journal of Phonetics*, *40*(2), 289–306.
Keuleers, E., Brysbaert, M., & New, B. (2010). Subtlex-nl: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650.
Kewley-Port, D., & Atal, B. S. (1989). Perceptual differences between vowels located in a limited phonetic space. *Journal of the Acoustical Society of America*, *85*(4), 1726–1740.
Kissine, M., Van de Velde, H., & van Hout, R. (2003). An acoustic study of standard Dutch /v/, /f/, /z/ and /s/. *Linguistics in the Netherlands*, *20*(1), 93–104.
Kissine, M., Van de Velde, H., & van Hout, R. (2005). *Acoustic contributions to sociolinguistics: Devoicing of /v/ and /z/ in Dutch* (Vol. 10) (No. 2).
Kleber, F., Harrington, J., & Reubold, U. (2011). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, 1–23.
Knops, U., & van Hout, R. (1988). *Language attitudes in the Dutch language area: An introduction.* Walter de Gruyter.
Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81.
Kristiansen, T. (2009). The macro-level social meanings of late-modern Danish accents. *Acta Linguistica Hafniensia*, *41*(1), 167–192.
Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, *1*(3), 199–244.
Kronrod, Y., Coppess, E., & Feldman, N. H. (2012). A unified model of categorical effects in consonant and vowel perception. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 629–634).
Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107.
Labov, W. (1966). *The social stratification of English in New York City.* Washington: Center for Applied Linguistics.
Labov, W. (1981). Resolving the neogrammarian controversy. *Language*, 267–308.
Labov, W. (1982). Building on empirical foundations. *Perspectives on Historical Linguistics*, *24*, 17.
Labov, W. (1994). *Principles of language change: Internal factors* (Vol. 1). Blackwell.
Labov, W. (2001). *Principles of linguistic change: Social factors* (Vol. 2).
Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress* (Vol. 1). US Regional Survey.

Lacerda, F. (1995). The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 2, pp. 140–147).

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, *29*(1), 98–104.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, *60*(1), 44.

Latour, B., Grondelaers, S., & van Hout, R. (2012). De schoonheid van taal. hoe wezenlijk is het oordeel mooi in taalattitudes? *Taal en Tongval*, *64*(2), 243–261.

Lev-Ari, S., & Peperkamp, S. (2014). An experimental study of the role of social factors in language change: The case of loanword adaptations. *Laboratory Phonology*, *5*(3), 379–401.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431-461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403–439). Springer.

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

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th international congress of phonetic sciences* (Vol. 563, pp. 563–567).

Livijn, P. (2002). Distribution of dental and retroflex l-sounds across some Swedish dialects. In *Proceedings from fonetik 2002, the XVth Swedish Phonetics Conference* (pp. 29–31).

Macmillan, N., & Creelman, C. (1991). *Detection theory: A user's guide.*

Malderez, I. (1995). *Contribution à la synchronie dynamique du français contemporain: le cas des voyelles orales arrondies.* University of Paris 7.

Mayo, C., Clark, R. A., & King, S. (2005). Multidimensional scaling of listener responses to synthetic speech. In *International speech communication association.*

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 539.

Mees, I., & Collins, B. (1982). A phonetic description of the consonant system of standard Dutch. *Journal of the International Phonetic Association*, *12*, 2-12.

Mees, I., & Collins, B. (1999). Cardiff: a real-time study of glottalisation. *Urban voices: Accent studies in the British Isles*, 185–202.

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, *325*(5938), 284–288.

Meyerhoff, M. (2011). *Introducing sociolinguistics*. Taylor & Francis.

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics*, *72*(6), 1614–1625.

Milroy, J. (1992). *Linguistic variation and change: on the historical sociolinguistics of English*. Blackwell.

Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, *21*(02), 339–384.

Mitterer, H. (2009). *Research stuff.* Retrieved 17.05.2015, from `http://www.holgermitterer.eu/research.html`

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173.

Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception–action coupling in speech. *Attention, Perception, & Psychophysics*, *75*(3), 557–575.

Nahkola, K., & Saanilahti, M. (2004). Mapping language changes in real time: A panel study on Finnish. *Language Variation and Change*, *16*(02), 75–92.

Nakai, S. (1998). The effect of vowel prototypicality and extremity on discrimination sensitivity. *Journal of the Acoustical Society America*, *103*, 2983.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, *21*(4), 422–432.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790.

Nettle, D. (1999). Using social impact theory to simulate language change. *Lingua*, *108*(2), 95–117.

Newman, R. (1997). *Individual differences and the link between speech perception and speech production*. State University of New York.

Newman, R. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *Journal of the Acoustical Society of America*, *113*(5), 2850–2860.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62–85.

Nielsen, K. (2007). Implicit phonetic imitation is constrained by phonemic contrast. In *Proceedings of the 16th International Congress of Phonetic Sciences, Dudweiler, Germany* (pp. 1961–1964).

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Ohala, J. J. (1981). The listener as a source of sound change. *Papers from the Parasession on Language and Behavior*, 178–203.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In *The Production of Speech* (pp. 189–216). Springer.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. *Language change: Contributions to the study of its causes*, 173–198.

Ohala, J. J. (1993a). The phonetics of sound change. *Historical linguistics: Problems and perspectives*, 237–278.

Ohala, J. J. (1993b). Sound change as nature's speech perception experiment. *Speech Communication*, *13*(1), 155–161.

Paliwal, K., Lindsay, D., & Ainsworth, W. (1983). Correlation between production and perception of English vowels. *Journal of Phonetics*, 77–83.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, *119*(4), 2382–2393.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190–197.

Pastuszek-Lipińska, B. (2008). Musicians outperform nonmusicians in speech imitation. In *Computer music modeling and retrieval. sense of sounds* (pp. 56–73). Springer.

Paul, H. (1880). *Prinzipien der sprachgeschichte*. Tübingen: Niemeyer.

Pear, T. H. (1931). *Voice and personality*. Chapman and Hall.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America*, *116*(4), 2338–2344.

Perkell, J. S., & Klatt, D. H. (2014). *Invariance and variability in speech processes*. Psychology Press.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169–190.

Pierrehumbert, J. (2001). *Lenition and contrast* (Vol. 45). John Benjamins Publishing.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory phonology*, *7*, 101–139.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2-3), 115–154.

Pisoni, D. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, *3*(1), 7–18.

Pisoni, D. (1997). Some thoughts on 'normalization' in speech perception. In K. Johnson & J. Mullennix (Eds.), (pp. 9–32). Academic Press.

Plichta, B., & Preston, D. (2005). The/ay/'s have it: The perception of /ay/ as a North-South stereotype in United States English. *Acta Linguistica Hafniensia*, *37*(1), 107–130.

Pols, L. C., Van der Kamp, L. T., & Plomp, R. (1969). Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America*, *46*(2B), 458–467.

Preston, D. (1989). *Perceptual dialectology: Nonlinguists' views of areal linguistics* (Vol. 7). Walter de Gruyter.

Preston, D. (2013). Language with an attitude. *The Handbook of Language Variation and Change*, *80*, 157.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, *18*(1), 10–30.

Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*(1), 103–121.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425.

R Core Team. (2014). *R: A language and environment for statistical computing.* Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rakerd, B. (1984). Vowels in consonantal context are perceived more linguistically than are isolated vowels: Evidence from an individual differences scaling study. *Perception & Psychophysics*, *35*(2), 123–136.

Rakerd, B., & Verbrugge, R. R. (1985). Linguistic and acoustic correlates of the perceptual structure found in an individual differences scaling study of vowels. *Journal of the Acoustical Society of America*, *77*(1), 296–301.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & psychophysics*, *30*(3), 217–227.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language: Advances in basic research and practice*, *10*, 243–335.

Rietveld, T., & Van Heuven, V. (2009). *Algemene fonetiek.* Coutinho.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

Rojczyk, A. (2011). Perception of English voice onset time continuum by Polish learners. In J. Arabski & A. Wojtaszek (Eds.), *The acquisition of L2 phonology* (pp. 37–58). Bristol, Buffalo, Toronto: Multilingual Matters.

Romaine, S. (1988). *Historical sociolinguistics: Problems and methodology.* (K. J. Mattheier, Ed.). Berlin: Walter de Gruyter.

Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, *14*(3), 337–353.

Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, 560–588.

Schouten, M., & van Hessen, A. (1992). Modeling phoneme perception. I: categorical perception. *Journal of the Acoustical Society of America*, *92*(4), 1841–1855.

Sebregts, K. (2015). *The sociophonetics and phonology of Dutch /r/.* LOT University Utrecht.

Shadle, C. H. (2012). The aerodynamics of speech. *The Handbook of Phonetic Sciences*, *79*, 39.

Shepard, R. N. (1972). Psychological representation of speech sounds. In E. David & P. Denes (Eds.), *Human communication: A unified view.* (pp. 67–113). McGraw-Hill New York.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422–429.

Sievers, E. (1901). *Grundzüge der Phonetik.* Breitkopf & Härtel.

Simon, E. (2009). Acquiring a new second language contrast: an analysis of the English laryngeal system of native speakers of Dutch. *Second Language Research*, *25*(3), 377–408.

Singh, S., & Woods, D. R. (1971). Perceptual structure of 12 American English vowels. *Journal of the Acoustical Society of America*, *49*, 1861–1866.

Sinnott, J. M., Brown, C. H., Malik, W. T., & Kressley, R. A. (1997). A multidimensional scaling analysis of vowel discrimination in humans and monkeys. *Perception & Psychophysics*, *59*(8), 1214–1224.

Slis, I. (1985). *The voiced-voiceless distinction and assimilation of voice in Dutch.* Radboud Universiteit Nijmegen.

Slis, I. (1986). Assimiliation of voice as a function of stress, word boundaries, and sex of speaker and listener. *Journal of Phonetics*, *14*(311–326).

Slis, I., & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction I and II. *Language and Speech*, *12*, 80-102.

Slis, I., & van Heugten, M. (1989). Voiced-voiceless distinction in dutch fricatives. In H. Bennis & A. van Kemenade (Eds.), *Linguistics in the Netherlands* (pp. 123–132).

Smits, R., & van Alphen, P. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics*, *32*(4), 455–491.

Sonderegger, M., & Adviser-Goldsmith, J. A. (2012). *Phonetic and phonological dynamics on reality television.* University of Chicago.

Stevens, K. N. (2000). *Acoustic phonetics.* MIT press.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, *18*(1), 86–100.

Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural language processing and speech technology. results of the 3rd KOVENS conference* (pp. 14–26). Berlin: Mouton de Gruyter.

Strange, W. (1995). *Speech perception and linguistic experience: Issues in cross-language research.* York Press.

Stuart-Smith, J. (2007). A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. *Proceedings of ICPhS, Saarbrücken.*

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*(4), 487–501.

Syrdal, A. K. (1996). Acoustic variability in spontaneous conversational speech of american english talkers. In *Proceedings of the Fourth International Conference on Spoken Language* (Vol. 1, pp. 438–441).

Tagliamonte, S. A. (2011). *Variationist sociolinguistics: change, observation, interpretation.* John Wiley & Sons.

Terbeek, D. (1977). *A cross-language multidimensional scaling study of vowel perception* (No. 37). University of California.

Thomas, E. R. (2000). Spectral differences in /ai/ offsets conditioned by voicing of the following consonant. *Journal of Phonetics*, *28*(1), 1–25.

Thomas, E. R. (2002). Sociophonetic applications of speech perception experiments. *American Speech*, *77*(2), 115–147.

Tops, E. (2009). *Variatie en verandering van de /r/.* VUB Press.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, *1*(2), 179–195.

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

Trudgill, P. (1988). Norwich revisited: recent linguistic changes in an English urban dialect. *English World-Wide*, *9*(1), 33–49.

Trudgill, P. (2008). Colonial dialect contact in the history of european languages: On the irrelevance of identity to new-dialect formation. *Language in Society*, *37*(02), 241–254.

van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology*, *18*(1), 31–48.

Vandekerckhove, R. (2009). Dialect loss and dialect vitality in Flanders. *International Journal of the Sociology of Language*, 73=-97.

van den Berg, R. (1988). *The perception of voicing in Dutch two-obstruent sequences.* Radboud Universiteit Nijmegen.

van der Harst, S. (2011). *The vowel space paradox. A sociophonetic study on dutch.* LOT University Utrecht.

van der Wal, M., & van Bree, C. (1992). *Geschiedenis van het nederlands.* Spectrum.

Van de Velde, H. (1996). *Variatie en verandering in het gesproken standaardnederlands (1935–1993).* University of Nijmegen.

Van de Velde, H., Gerritsen, M., & van Hout, R. (1996). The devoicing of fricatives in standard dutch: A real-time study based on radio recordings. *Language Variation and Change*, *8*(02), 149–175.

Van de Velde, H., Kissine, M., Tops, E., van der Harst, S., & van Hout, R. (2012). Will Dutch become Flemish? Autonomous developments in Belgian Dutch. *Multilingua. Journal of Cross-Cultural and Interlanguage Communication*, *29*(3/4), 385–416.

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 Reenen, P. (1994). Driemaal /r/ in de Nederlandse dialecten. *Taal en Tongval*, 54–72.

van Son, R., & Pols, L. (1996). An acoustic profile of consonant reduction. In *Proceedings of the Fourth International Conference on Spoken Language* (Vol. 3, pp. 1529–1532).

Veenker, T. (2012). *The ZEP experiment control application* (Tech. Rep.). Utrecht Institute of Linguistics OTS, Utrecht University.

Weinreich, U., Labov, W., & Herzog, M. (1968). *Empirical foundations for a theory of language change.* University of Texas Press.

Yaeger-Dror, M. (1994). Phonetic evidence for sound change in Quebec French. *Phonological Structure and Phonetic Form*, 267–293.

Yu, A. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. *Origins of sound change: Approaches to phonologization*, 201–227.

Yu, A., Abrego-Collier, C., Baglini, R., Grano, T., Martinovic, M., Otte, C., et al. (2011). Speaker attitude and sexual orientation affect phonetic imitation. *University of Pennsylvania Working Papers in Linguistics*, *17*(1), 26.

Ziliak, Z., & Van de Velde, H. (2008). *Stop variation in Dutch.* Presentation at New Ways of Analyzing Variation 37 Houston TX.

Ziliak, Z., & Van de Velde, H. (in progress). Stop variation in dutch.

Zonneveld, W. (1983). Lexical and phonological properties of Dutch voicing assimilation. In M. van den Broecke, V. van Heuven, & W. Zonneveld (Eds.), (Vol. 13, pp. 297–312). Dordrecht: Foris.

Appendices

Appendix A

Table 1: List of the participants with their ID number, sex, age and origin, and split up by region (N=100).

Region	ID number	Sex	Age	Origin
West-Flanders	WF01	M	21	Izegem
	WF02	W	24	Geluwe
	WF03	M	25	Izegem
	WF04	W	25	Kortrijk
	WF05	W	22	Gits
	WF06	M	22	Diksmuide
	WF07	W	21	Roeselare
	WF08	W	20	Roeselare
	WF09	W	20	Roeselare
	WF10	W	25	Koolskamp
	WF11	M	24	Kortrijk
	WF12	M	23	Veldegem
	WF13	M	21	Zwevegem
	WF14	W	25	Ieper
	WF15	W	22	Ingelmunster
	WF16	M	22	Oostrozebeke
	WF17	M	22	Otegem
	WF18	W	21	Watou
	WF19	M	22	Deerlijk
	WF20	M	22	Zwevegem

Region	ID number	Sex	Age	Origin
Flemish-Brabant	FB01	W	22	Overijse
	FB02	W	21	Werchter
	FB03	M	21	Overijse
	FB04	W	26	Kessel-Lo
	FB05	M	22	Binkom
	FB06	W	18	Humbeek
	FB07	M	23	Leefdaal
	FB08	M	23	Holsbeek
	FB09	W	21	Tienen
	FB10	W	19	Goetsenhaven
	FB11	W	19	Huldenberg
	FB12	W	22	Leuven
	FB13	M	22	Leuven
	FB14	M	20	Sint-Agatha-Rode
	FB15	W	20	Landen
	FB16	W	24	Kampenhout
	FB17	M	25	Leefdaal
	FB18	M	20	Rotselaar
	FB19	M	20	Begijnendijk
	FB20	M	20	Lubbeek

Region	ID number	Sex	Age	Origin
Limburg	LI01	M	18	Sittard
	LI02	W	23	Mechelen
	LI03	W	22	Well
	LI04	W	20	Geijsteren
	LI05	W	22	Simpelveld
	LI06	M	23	Heerlen
	LI07	W	25	Landgraaf
	LI08	W	21	Heerlen
	LI09	W	22	Beek
	LI10	M	21	Blitterswijck
	LI11	W	21	Maastricht
	LI12	M	21	Bunde
	LI13	M	21	Thorn
	LI14	M	20	Blitterswijck
	LI15	M	20	Maasbree
	LI16	M	22	Borgharen
	LI17	W	20	Wijlre
	LI18	W	20	Venlo
	LI19	M	21	Sittard
	LI20	M	22	Venlo

Region	ID number	Sex	Age	Origin
South-Holland	SH01	W	27	Zoeterwoude
	SH02	M	22	Zoetermeer
	SH03	W	21	Leiden
	SH04	W	28	Voorhout
	SH05	M	28	Leiden
	SH06	M	23	Leiderdorp
	SH07	W	18	Alphen aan den Rijn
	SH08	M	22	Oegstgeest
	SH09	M	21	Leiden
	SH10	M	21	Leiden
	SH11	W	21	Gouda
	SH12	W	19	Den Hoorn
	SH13	M	21	Katwijk aan den Rijn
	SH14	W	19	Haastrecht
	SH15	M	19	Leiderdorp
	SH16	W	22	Wateringen
	SH17	M	21	Leiden
	SH18	W	27	Delft
	SH19	W	20	Zoeterwoude
	SH20	M	20	Delft

Region	ID number	Sex	Age	Origin
Groningen	GR01	W	21	Slochteren
	GR02	W	25	Roden
	GR03	W	22	Borgercompagnie
	GR04	W	27	Assen
	GR05	M	28	Winsum
	GR06	W	20	Groningen
	GR07	M	27	Vries
	GR08	W	20	Groningen
	GR09	W	24	Zuidbroek
	GR10	M	23	Winsum
	GR11	W	21	Groningen
	GR12	M	27	Thesinge
	GR13	M	23	Assen
	GR14	M	21	Harkstede
	GR15	M	24	Zuidhorn
	GR16	M	23	Groningen
	GR17	M	21	Groningen
	GR18	W	24	Roden
	GR19	M	20	Groningen
	GR20	W	20	Roden

Appendix B

Intake questionnaire for Dutch participants

Met deze enquête verzamelen we achtergrondinformatie. Dit is bij het analyseren van de resultaten zeer belangrijk. De enquête bestaat uit 17 vragen die te maken hebben met je herkomst en taalachtergrond. Niet alle vragen zijn voor jou van toepassing. In zo'n geval laat je de vraag open. Vanzelfsprekend worden alle gegevens vertrouwelijk behandeld en anoniem verwerkt. De gegevens worden enkel gebruikt in het kader van dit onderzoek. Alvast bedankt voor je medewerking!

Deelnemerscode:
Datum :

Algemene gegevens

1. Ik ben een:
X man
X vrouw

2. Geboortejaar:

3. Ik heb gehoorproblemen:
X Nee
X Ja. Zo ja, welk(e)?

4. Mijn huidige beroep/studie is:

5. Ik ben:
X rechtshandig
X linkshandig

6. Ik ben dyslectisch:
X Nee
X Licht
X Matig
X Sterk

Taalachtergrond

1. Mijn moedertaal is/zijn:
X Nederlands
X Frans

X Engels
X Duits
X Andere:

2. Ik spreek dialect:
X Nee
X Ja. Welk dialect?
X In welke situaties/met wie?

3. Als ik Standaardnederlands spreek, denk IKZELF dat ik helemaal kan wegstoppen uit welke streek/regio ik vandaan kom.
X Nee, ik kan het niet helemaal wegstoppen.
X Ja, ik kan het wel wegstoppen.

4. Als ik Standaardnederlands spreek, kunnen ANDERE MENSEN horen uit welke streek/regio ik vandaan kom.
X Nee
X Ja. Welke regio/welk accent?

Afkomst en taalcontact

1. Als mensen in Nederland vragen waar ik vandaan kom, dan antwoord ik:

2. Mijn moeder komt oorspronkelijk uit: (Geef een stad, dorp, regio en/of land aan.)

3. Als mijn moeder met mij spreekt, gebruikt ze:
X Standaardnederlands
X Standaardnederlands met een accent, welk?
X Een dialect, welk?
X Een andere taal, welke?

4. Mijn vader komt oorspronkelijk uit: (Geef een stad, dorp, regio en/of land aan.)

5. Als mijn vader met mij spreekt, gebruikt hij:
X Standaardnederlands
X Standaardnederlands met een accent, welk?
X Een dialect, welk?
X Een andere taal, welke?

6. Waar heb je tot nu toe gewoond? Wees zo nauwkeurig mogelijk (Periode, Plaats, Provincie, Land).

Exit questionnaire for Dutch participants

Na afloop van alle experimenten willen we je nog een paar vragen stellen. Alle gegevens worden vertrouwelijk behandeld en anoniem verwerkt. De gegevens worden enkel gebruikt in het kader van dit onderzoek. Hartelijk dank voor je deelname!

Deelnemerscode:
Datum :

Taalcontact

1. Ik heb in mijn vriendenkring en/of familie mensen die UIT EEN ANDERE REGIO VAN NEDERLAND afkomstig zijn:
X Nee (dat betekent dat je alleen maar omgaan met mensen uit je eigen regio)
X Ja, Waar komen ze vandaan? (Je kunt uiteraard meerdere regio's aanvinken)
X Noord-Brabant
X Gelderland
X Drenthe
X Limburg
X Flevoland
X Overijssel
X Friesland
X Noord-Holland
X Groningen
X Zeeland
X Zuid-Holland

2. Ik heb in mijn vriendenkring en/of familie mensen die uit VLAANDEREN afkomstig zijn:
X Nee
X Ja, waar komen ze vandaan?

3. Hoe vaak hoor je Vlaams/Belgisch Nederlands? bv. op tv, in films, door kennissen, enz.
X Dagelijks
X Een paar keer per week
X Een keer per week
X Een keer per maand
X Zelden

4. Er zijn mensen die – zodra ze met iemand uit een andere regio spreken – zich heel gemakkelijk aanpassen aan het accent van hun gesprekspartner. Doe jij dat ook?

X Ja, ik pas mijn accent snel aan aan dat van andere mensen
X Ik pas soms mijn accent aan
X Nee, ik pas mijn accent niet aan aan het accent van anderen

Muziek

1. Ik heb een instrument bespeeld/ik bespeel een instrument:
X Nee
X Ja, Welk(e) instrument(en)? Hoe lang al?

2. Ik heb in een groep, een koor of solo gezongen/ik zing in een groep, een koor of solo:
X Nee
X Ja, Welk(e) instrument(en)? Hoe lang al?

3. Ik vind van mezelf dat ik muzikaal aangelegd ben:
X Ja
X Ja, redelijk
X Eerder niet
X Absoluut niet

Talen

1. Ik spreek/begrijp een andere taal/talen dan Nederlands:
Taal 1:
Ik begrijp het: heel slecht X X X X X heel goed
Ik spreek het: heel slecht X X X X X heel goed
Ik ben ermee begonnen toen ik ... was

Taal 2:
Ik begrijp het: heel slecht X X X X X heel goed
Ik spreek het: heel slecht X X X X X heel goed
Ik ben ermee begonnen toen ik ... was

Taal 3:
Ik begrijp het: heel slecht X X X X X heel goed
Ik spreek het: heel slecht X X X X X heel goed
Ik ben ermee begonnen toen ik ... was

Taal 4:
Ik begrijp het: heel slecht X X X X X heel goed
Ik spreek het: heel slecht X X X X X heel goed
Ik ben ermee begonnen toen ik ... was

Taal 5:
Ik begrijp het: heel slecht X X X X X heel goed
Ik spreek het: heel slecht X X X X X heel goed
Ik ben ermee begonnen toen ik ... was

2. Ik gebruik een andere taal(en) dan Nederlands minstens 1 keer per week:
X Geen
X Engels
X Duits
X Frans
X Andere:

3. Ik vind van mezelf dat ik goed ben in vreemde talen:
X Ja
X Ja, redelijk
X Niet echt
X Absoluut niet

Persoonlijkheid

Tenslotte willen we je nog enkele stellingen voorleggen om je persoonlijkheidsprofiel te bepalen. Het gaat hier om je eigen mening, niet die van anderen. Op de volgende pagina kan je je mening over de stellingen geven. Kruis 1 aan als je het helemaal niet eens bent met de stelling. Kruis 5 aan als je het helemaal eens bent met de stelling.

Ik volg niet graag bevelen	XXXXX
Ik ben een betrouwbaar persoon	XXXXX
Ik kan gemakkelijk mensen overhalen tot bijna alles	XXXXX
Al mijn relaties met andere zijn gebaseerd op vriendelijkheid	XXXXX
Ik haat vastgelegde sociale gedragsregels	XXXXX
Wat kleren betreft volg ik graag de mode	XXXXX
Ik ben gevoelig voor de problemen van anderen	XXXXX
Ik breng mensen gemakkelijk aan het lachen	XXXXX
Ik spreek graag in het openbaar	XXXXX
Ik ben niet zo goed in mensen vermaken	XXXXX
Ik ben koppig	XXXXX
Ik probeer niet te veel af te wijken van mijn groepsgenoten	XXXXX
Ik doe liever niets dat mij in gevaar brengt	XXXXX
Ik schiet goed op met de meeste mensen	XXXXX
In een ideale wereld zou ik willen leven zonder enige vorm van bestuur (bv. zonder regering, baas op het werk, enz.)	XXXXX
Ik besteed veel tijd aan praten met mensen	XXXXX
Ik vind het leuk om veel verschillende mensen te ontmoeten als ik naar een feest ga	XXXXX
Ik doe liever wat andere mensen willen dan ze te moeten overtuigen van mijn mening	XXXXX

Ik doe graag dingen op mijn eigen manier	XXXXX
Ik ben actief in openbare zaken die belangrijk zijn voor onze samenleving	XXXXX
Ik waardeer ideen van anderen, zelfs als ze anders zijn dan de mijne	XXXXX
Ik ben minder spraakzaam dan mensen rondom mij	XXXXX
Ik heb geen moeite om mensen iets op te leggen	XXXXX
Ik ben snel verlegen	XXXXX
Ik kan door mijn gedrag gemakkelijk de aandacht trekken van grote groepen	XXXXX

Contact

Het kan zijn dat de analyse van de experimenten extra vragen bij ons oproept. Mogen we dan contact opnemen voor extra informatie en/of vervolgonderzoek. Indien wel, geef hier je emailadres aan:
Wil je op de hoogte gehouden worden van de resultaten?
X Ja
X Nee

Dank je wel!

Appendix C

List of the external factors

1. Factors related to production/perception

- *Dialect proficiency*
  Binary factor assessing dialect proficiency (yes, I do speak a dialect or no, I don't speak a dialect).
- *Regional accent*
  Self-reported presence of a regional accent in the participants' speech (yes, I do have a regional accent or no, I don't have any regional accent).
- *Mobility*
  Factor assessing the mobility of participants based on 1) whether they had lived for a significant period in a different region of the country, 2) whether they have at least one parent coming from a different region, and 3) whether they have friends or family coming from other regions. The factor ranges from 0 (not mobile) to 3 (very mobile).
- *Contact with the other country within the Dutch language area*
  Factor assessing participants self-estimated frequency of contact with the variety of Dutch spoken in the other country (Flanders or The Netherlands). It is ranging from 1 (seldom) to 5 (everyday).

2. Factors related to imitation capacities

- *Imitation capacity*
  Factor assessing participants' self-reported capability to adapt to someone else's accent (imitator, occasional imitator or no imitator).
- *Musical abilities*
  Factor assessing participants' ability to play music instruments. Singing is treated as an instrument. The score is ranging from 0 (the participant does not play any instrument) to 4 (the participant plays multiple instruments since a long time).
- *Estimated musical abilities*
  Factor assessing participants' self-reported musical abilities ranging from 0 (very poor musical abilities) to 3 (very good musical abilities).
- *Foreign languages proficiency*
  Factor assessing participants' language proficiency, calculated as the sum of the self-reported proficiency on a 5-point scale for the production and comprehension of the different foreign languages. It is ranging from 0

(no proficiency in any foreign language) to 50 (very broad capacities in different foreign languages).

- *Estimated foreign languages proficiency*
  Factor assessing participants' self-reported language capacity ranging from 0 (very poor foreign language capacities) to 3 (very good foreign language capacities).
- *Foreign languages use*
  Factor assessing participants' use of foreign languages on a regular basis (at least once a week), ranging form 0 (uses no foreign language) to 3 (regular use of three foreign languages).
- *Personality factors*:
  - *Conformism*
    Participants' self-reported evaluation on the personality trait 'conformism', calculated as the sum of the scores on the five related questions and ranging from 0 (very nonconformist) to 20 (very conformist).
  - *Sociability*
    Participants' self-reported evaluation on the personality trait 'sociability', calculated as the sum of the scores on the five related questions and ranging from 0 (very antisocial) to 20 (very social).
  - *Talkativeness*
    Participants' self-reported evaluation on the personality trait 'talkativeness', calculated as the sum of the scores on the five related questions and ranging from 0 (very nontalkative) to 20 (very talkative).
  - *Egocentrism*
    Participants' self-reported evaluation on the personality trait 'egocentrism', calculated as the sum of the scores on the five related questions and ranging from 0 (very nonegocentric) to 20 (very egocentric).
  - *Conformism*
    Participants' self-reported evaluation on the personality trait 'leadership', calculated as the sum of the scores on the five related questions and ranging from 0 (bad leader) to 20 (good leader).

Appendix D

1. Test words in the word reading task (n=78)

Table 2: Test words from the word reading task, split up by variant.

	Fricatives			Stops		
Following phon.	/v/	/f/		/b/	/p/	
/ɑ/	vallen	fazant		bad	pad	*
/a/	varen	falen		baden	paden	*
/ɛ/	vet	fet	*	bed	pet	*
/e/	veter	feest		beer	peer	*
/ə/	vertrek	–		benoemen	–	
/ø/	veulen	feut		beurt	peuter	
/i/	vier	fier	*	bier	piek	
/ɛi/	veilig	feit		bijlage	pijl	
/ɪ/	vis	film		bitter	pil	
/l/	vlot	fles		bleek	pleiten	
/l/	vlag	fluiten		blik	plagen	
/o/	vogel	fooi		boden	poot	
/u/	voet	foerier		boel	poes	
/ɔ/	vos	folder		bot	politie	
/ɔu/	vouwen	fout		bouwen	pauze	
/r/	vreet	friet		braden	preken	
/r/	vrezen	fruit		brug	proef	
/œʸ/	vuist	fuif		buik	puinhoop	
/ʏ/	vullen	fut		bus	put	
/y/	vuur	fuut		buur	puur	*

Note. * The asterisk indicates that these words form a minimal pair.
Note. The words *fazant, politie, benoemen* and *vertrek* were removed post-experimentally, because their word stress did not fall on the first syllable whilst Slis and van Heugten (1989) have shown that stress might influence duration and voicing of fricatives.

2. Words in the word reading task taken from the spontaneous imitation task (n=43)

aardbei *strawberry*, appel *apple*, banaan *banana*, blad *leaf*, bloem *flower*, boek *book*, bril (2x) *glasses*, cadeau *present*, diamant *diamond*, enveloppe *envelope*, golfbal *golf ball*, hoed *hat*, ijs (2x) *ice*, kaart *map*, kleerhanger *coat hanger*, konijn *rabbit*, kous (2x) *stocking*, lamp *lamp*, laptop *laptop*, lepel *spoon*, paddestoel (2x) *mushroom*, paperclip *clip*, paraplu *umbrella*, peer *pear*, potlood (2x) *pencil*, schaar *scissors*, schelp (2x) *shell*, slak *snail*, sleutel *key*, spuit *syringe*, vork *fork*, wereldbol *globe*, wortel *carrot*.

3. Fillers in the word reading task (n=39)

arm *poor*, dapper *brave*, deur *door*, ding *thing*, garage *garage*, golf *golf*, gras *grass*, groen *green*, hek *fence*, hooi *hay*, hopen *hope*, hout *wood*, huis *house*, hut *hut*, keel *throat*, kerk *church*, maten *sizes*, mes *knife*, mijn *mine*, militair *soldier*, minister *minister*, naakt *naked*, nat *wet*, nek *neck*, neus *nose*, olifant *elephant*, pianist *pianist*, saus *sauce*, schijn *appearance*, soep *soup*, taken *task*, trui *sweater*, tuin *garden*, twijfel *doubt*, wet *law*, wolk *cloud*, zout *salt*, zuivel *dairy*, zwoel *muggy*.

Appendix E

Table 3: Test words and fillers from the carrier sentence reading task, split up by variant.

	Test words				Fillers			
	Fricatives		Stops					
Final consonants	/v/	/f/	/b/	/p/	/w/	/d/	/m/	/k/
/g/	vieg	fieg	bieg	pieg	wieg	dieg	mieg	kieg
/f/	vief	fief	bief	pief	wief	dief	mief	kief
/k/	viek	fiek	biek	piek	wiek	diek	miek	kiek
/l/	viel	fiel	biel	piel	wiel	diel	miel	kiel
/m/	viem	fiem	biem	piem	wiem	diem	miem	kiem
/n/	vien	fien	bien	pien	wien	dien	mien	kien
/p/	viep	fiep	biep	piep	wiep	diep	miep	kiep
/s/	vies	fies	bies	pies	wies	dies	mies	kies
/t/	viet	fiet	biet	piet	wiet	diet	miet	kiet

Note. We are aware of the fact that some of these words are real words in Dutch. However, the analysis presented in Chapter 4 did not show any difference when these words are taken away.

Appendix F

Table 4: Test words from the sentence reading task, split up by variant (n=56).

	Fricatives			Stops		
Foll. phon.	/v/	/f/		/b/	/p/	
/ɑ/	valt	fakkel		bad	pad	*
/a/	varkenvlees	fabel		baard	paard	*
/ɛ/	vechten	ferme		benzine	penseel	
/e/	vee	fee	*	beer	peer	*
/ɛi/	vijf	feiten		bijlage	pijnboompit	
/ɪ/	vis	fin		binnen	pink	
/i/	vier	fiets		biet	piet	*
/l/	vlek	flessen		blikje	plek	
/ɔ/	vordering	formele		bondgenoot	pompoen	
/u/	voeren	foelie		boekentas	poeder	
/o/	vorig	folie		bonen	Polen	
/r/	vriest	fries		brief	priester	
/œʸ/	vuilnisbak	fuif		buik	puist	
/y/	vuur	fuut		buurt	puur	

Note. * The asterisk indicates that these words form a minimal pair.

- Mijn nummer is 0 7 4 5.
  My number is 0 7 4 5.
- De goede fee heeft Doornroosje geholpen.
  The well-meaning fairy helped Sleeping Beauty.
- In de fabel was er een priester met een baard.
  There was a bearded priest in the fairytale.
- Een feest heet in Vlaanderen een fuif.
  A party is called a fuif in Flanders.
- Hij stuurde de formele brief als bijlage.
  He sent the formal letter as an attachment.
- Het oude vee staat in de groene weide.
  The old cattle is grazing in the green meadow.
- Het meisje nam een peer mee in haar boekentas.
  The girl took a pear in her schoolbag.
- Hij voegde de pijnboompitten, de foelie en de bonen aan de soep toe.
  He added the pine nuts, the mace, and the beans to the soup.
- Jelle is een Fries en Seppe is een Fin.
  Jelle comes from Friesland and Seppe from Finland.
- In het reservaat hebben ze een beer en een fuut gezien.
  They saw a bear and a great crested grebe in the reserve.

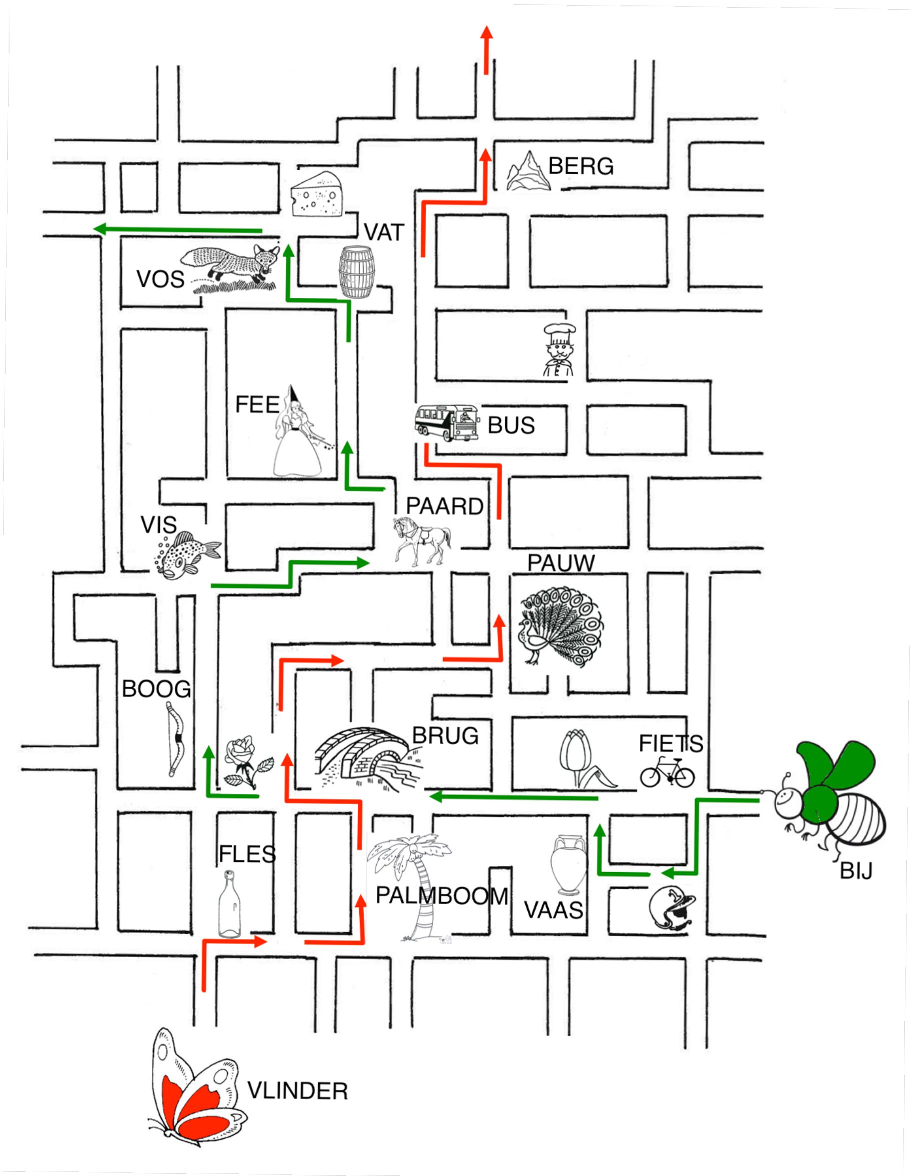
- In Polen maken ze soep van pompoen of rode biet.
  In Poland they prepare soup with pumpkin and beetroot.
- Ze valt van haar fiets en breekt haar pink.
  She felt from her bicycle and broke her little finger.
- Hij lust geen varkensvlees.
  He does not like pork.
- Ze gebruikt dit zachte penseel voor het aanbrengen van poeder.
  She uses this soft brush for powdering.
- Bart gooide een blikje in de vuilnisbak.
  Bart threw a can in the trash.
- Hij werd wakker met een dikke puist op zijn neus.
  He woke up with a big pimple on his nose.
- Dit is de beste plek in de buurt.
  This is the best place in the neighborhood.
- Jan en Piet kwamen aan op hun paard.
  Jan and Piet arrived on their horse.
- Vijf jaar na de feiten is er nog geen vordering.
  After five years, there is still no progress.
- De kinderen blijven liever binnen spelen omdat het bijna vriest.
  The children prefer playing inside, because it is freezing.
- Schotland was toen een bondgenoot van Frankrijk.
  At that time Scotland was an ally of France.
- Volgens het parket zouden twee flessen benzine vuur gevat hebben.
  According to the public prosecutor, two bottles of petrol caught fire.
- Ze heeft een grote vlek op haar buik.
  She has a big stain on her belly.
- Alice nam een pad door het grote bos.
  Alice took a path through the forest.
- Esther vindt een bad veel praktischer dan een douche.
  Esther finds that a bath is far more practical than a shower.
- Vorig jaar gaf ze de fakkel door aan Caroline.
  Last year she handed on the torch to Caroline.
- Partijen die ferme uitspraken doen, trekken veel kiezers.
  Parties which use harsh rhetoric are very popular.
- In Syrië voeren ze oorlog en vechten ze om territorium.
  In Syria they go to war and fight for territory.
- Hij legt daarna de vis op een stukje folie.
  He then puts the fish on a piece of foil.
- Het gaat hier puur om de resultaten.
  It is all about the results.

Appendix G

Figure 1: Drawing used for the semi-spontaneous task: Part 1.

Note. The names of the objects were not present on the original test stimuli, but were added here to indicate the most frequently used target words.

Figure 2: Drawing used for the semi-spontaneous task: Part 2.



Note. The names of the objects were not present on the original test stimuli, but were added here to indicate the most frequently used target words.

Appendix H

Table 5: Speech production results for fricatives in all speech styles measured for duration (in ms) and voicing (in %) split up by variant.

			Duration in ms		Voicing in %	
	n	n/part.	/v/	/f/	/v/	/f/
Word reading	3686	37	134 (34)	155 (37)	23 (31)	2 (4)
Carrier sentence r.	1793	18	119 (31)	160 (29)	52 (32)	16 (8)
Sentence reading	2790	28	99 (31)	135 (25)	47 (35)	17 (8)
Semi-spontaneous sp.	1157	20	120 (41)	153 (40)	44 (35)	14 (11)
Spontaneous speech	835	10	106 (32)	131 (30)	52 (36)	29 (26)
Total and Mean	10261	113	115.6	146.8	43.6	15.6

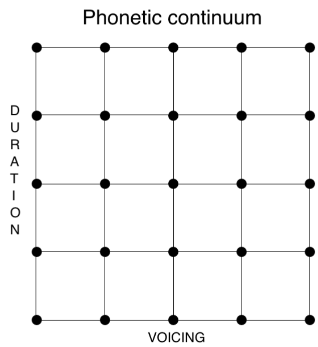
Table 6: Speech production results for stops in all speech styles measured for duration (in ms), VOT (in ms) and voicing (in %) split up by variant.

			Duration in ms		VOT in ms		Voicing in %	
	n	n/part.	/b/	/p/	/b/	/p/	/b/	/p/
Word reading	3700	37	102 (38)	–	-66 (56)	19 (8)	63 (36)	5 (5)
Carrier sentence r.	1790	18	123 (24)	157 (26)	-105 (32)	21 (12)	86 (20)	21 (9)
Sentence reading	2772	28	79 (26)	113 (27)	-61 (28)	18(9)	80 (26)	17 (11)
Semi-spontaneous sp.	1656	20	105 (46)	139 (47)	-78 (45)	18 (9)	74 (28)	16 (12)
Spontaneous speech	886	10	86 (33)	106 (43)	-66 (32)	19 (10)	81 (23)	44 (32)
Total and Mean	10804	113	102.3	136.3	-77.5	19	75.8	14.9

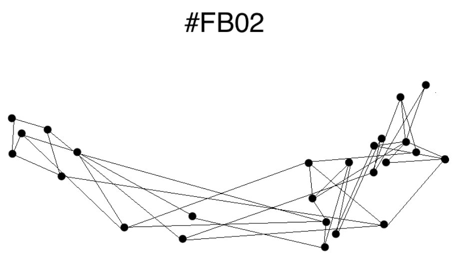
Note. Numbers between round brackets indicate standard deviations.

Appendix I

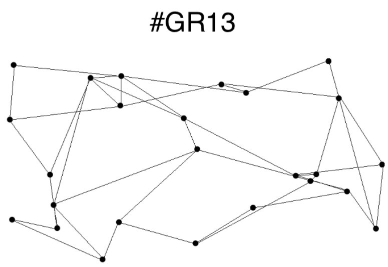
Figure 3: The reference phonetic continuum (a) and two bi-dimensional MDS solutions of the fricative data for two individual participants : (b) #FB02 which exemplifies the use of the voicing dimension, but no use of duration and (c) #GR13 which exemplifies the use of both dimensions. The voicing dimension is presented horizontally (the more to the right, the more voicing) and the duration dimension vertically (the more to the bottom, the longer the duration).



(a) Phonetic continuum



(b) Voicing & no duration



(c) Voicing & duration

Appendix J

Table 7: List of the speakers in the matched guise attitude task, their country (N=Netherlands vs. F=Flanders), sex, region, origin and age (N=28).

	Use	Country	Sex	Region	Origin	Age
1	Example	N	Woman	North	Almelo	23
2	Example	V	Woman	Flanders	Antwerpen	23
3	(v)	N	Woman	North	Leeuwarden	22
4	(v)	N	Woman	Middle	Den Haag	26
5	(v)	N	Woman	South	Maastricht	26
6	(v)	V	Woman	Flanders	Antwerpen	21
7	(v)	N	Man	North	Drenthe	26
8	(v)	N	Man	Middle	Delft	30
9	(v)	N	Man	South	Goes	34
10	(v)	V	Man	Flanders	Westmalle	32
11	(b)	N	Woman	North	Zwolle	25
12	(b)	N	Woman	Middle	Amsterdam	26
13	(b)	N	Woman	South	Vugt	24
14	(b)	V	Woman	Flanders	Kalmthout	26
15	(b)	N	Man	North	Assen	25
16	(b)	N	Man	Middle	Utrecht	25
17	(b)	N	Man	South	Nijmegen	24
18	(b)	V	Man	Flanders	Gent	24
19	Filler	N	Man	South	Utrecht	24
20	Filler	N	Woman	Middle	Zoetemeer	26
21	Filler	N	Woman	Middle	Apeldoorn	24
22	Filler	V	Woman	Flanders	Assenede	25
23	Filler	N	Man	North	Veluwe	33
24	Filler	N	Woman	Middle	Woerden	21
25	Filler	N	Woman	South	Breda	25
26	Filler	V	Woman	Flanders	Brugge	36
27	Filler	N	Woman	Middle	Amersfoort	27
28	Filler	N	Man	North	Arnhem	24

Appendix K

Results of the linear mixed-model regressions on the attitude data.

Table 8: Estimated parameters of mixed-effects modelling on attitudes towards fricatives for the first component (*Modernity*).

Random effects	Variance	S.D.			
Participant (Intercept)	0.093	0.305			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	-0.014	0.091	-0.151	.880	
Isvoiced	-0.080	0.084	-0.952	.344	
Flemish-Brabant	0.097	0.128	0.760	.449	
Limburg	0.062	0.128	0.487	.627	
South-Holland	-0.080	0.128	-0.621	.536	
Groningen	0.148	0.128	1.154	.251	
Isvoiced * Flemish-Brabant	-0.015	0.119	-0.124	.902	
Isvoiced * Limburg	-0.109	0.119	-0.911	.365	
Isvoiced * South-Holland	0.253	0.119	2.117	.037	*
Isvoiced * Groningen	-0.047	0.119	-0.398	.692	

Table 9: Estimated parameters of mixed-effects modelling on attitudes towards fricatives for the second component (*Power*).

Random effects	Variance	S.D.			
Participant (Intercept)	0.120	0.347			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	0.079	0.095	0.834	.406	
Isvoiced	-0.091	0.077	-1.185	.239	
Flemish-Brabant	-0.190	0.134	-1.414	.161	
Limburg	-0.158	0.134	-1.176	.243	
South-Holland	-0.042	0.134	-0.316	.753	
Groningen	0.050	0.134	0.376	.708	
Isvoiced * Flemish-Brabant	0.223	0.109	2.045	.044	*
Isvoiced * Limburg	-0.027	0.109	-0.251	.802	
Isvoiced * South-Holland	0.085	0.109	0.779	.438	
Isvoiced * Groningen	0.065	0.109	0.592	.555	

Table 10: Estimated parameters of mixed-effects modelling on attitudes towards fricatives for the third component (*Solidarity*).

Random effects	Variance	S.D.			
Participant (Intercept)	0.078	0.279			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	-0.007	0.078	-0.085	.932	
Isvoiced	-0.235	0.065	-3.595	<.001	*
Flemish-Brabant	-0.096	0.110	-0.875	.384	
Limburg	0.117	0.110	1.068	.288	
South-Holland	0.235	0.110	2.140	.035	*
Groningen	0.122	0.110	1.113	.269	
Isvoiced * Flemish-Brabant	0.041	0.092	0.445	.657	
Isvoiced * Limburg	0.183	0.092	1.980	.051	(*)
Isvoiced * South-Holland	0.120	0.092	1.298	.198	
Isvoiced * Groningen	0.140	0.092	1.515	.133	

Table 11: Estimated parameters of mixed-effects modelling on attitudes towards stops for the first component (*Modernity*).

Random effects	Variance	S.D.		
Participant (Intercept)	0.090	0.300		
Fixed effects	Estimates	S.E.	t value	p value
Intercept	-0.014	0.092	-0.147	.883
Isvoiced	-0.085	0.090	-0.941	.349
Flemish-Brabant	0.051	0.131	0.387	.700
Limburg	0.002	0.131	0.016	.987
South-Holland	0.063	0.131	0.483	.630
Groningen	0.143	0.131	1.095	.276
Isvoiced * Flemish-Brabant	0.113	0.127	0.889	.376
Isvoiced * Limburg	0.014	0.127	0.113	.910
Isvoiced * South-Holland	-0.048	0.127	-0.375	.709
Isvoiced * Groningen	-0.038	0.127	-0.300	.765

Table 12: Estimated parameters of mixed-effects modelling on attitudes towards stops for the second component (*Power*).

Random effects	Variance	S.D.			
Participant (Intercept)	0.126	0.355			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	-0.076	0.094	-0.810	.420	
Isvoiced	0.120	0.072	1.672	.098	
Flemish-Brabant	-0.048	0.133	-0.356	.723	
Limburg	0.153	0.133	1.143	.256	
South-Holland	0.086	0.133	0.645	.521	
Groningen	0.1133	0.133	0.849	.398	
Isvoiced * Flemish-Brabant	-0.119	0.102	-1.172	.244	
Isvoiced * Limburg	-0.055	0.102	-0.540	.591	
Isvoiced * South-Holland	-0.212	0.102	-2.082	.044	*
Isvoiced * Groningen	-0.060	0.102	-0.587	.559	

Table 13: Estimated parameters of mixed-effects modelling on attitudes towards stops for the third component (*Solidarity*).

Random effects	Variance	S.D.		
Participant (Intercept)	0.112	0.335		
Fixed effects	Estimates	S.E.	t value	p value
Intercept	-0.113	0.096	-1.178	.242
Isvoiced	-0.102	0.084	-1.220	.226
Flemish-Brabant	-0.072	0.135	-0.532	.596
Limburg	0.141	0.135	1.046	.298
South-Holland	0.202	0.135	1.490	.140
Groningen	0.206	0.135	1.526	.130
Isvoiced * Flemish-Brabant	0.231	0.119	1.947	.055
Isvoiced * Limburg	0.135	0.119	1.134	.260
Isvoiced * South-Holland	0.192	0.119	1.620	.109
Isvoiced * Groningen	0.126	0.119	1.057	.293

Note. Degrees of freedom (df) required for statistical significance testing of t values was given by $df = J - m - 1$ (Hox, 2010), where J is the most conservative number of second-level units ($J = 100$ speakers) and m is the total number of explanatory variables in the model ($m = 3$) resulting in $df = 96$.

Note. * Asterisks represent significant p values (α =.05).

Appendix L

Results of the linear mixed-model regressions on the forced imitation data.

Table 14: Estimated parameters of mixed-effects modelling on forced imitation data for voiced fricatives.

Random effects	Variance	S.D.			
Participant (Intercept)	219.18	14.805			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	68.093	5.216	13.055	<.001	*
Forced imitation	0.497	5.700	0.087	.931	
Flemish-Brabant	-10.926	7.376	-1.481	.142	
Limburg	2.143	7.376	0.290	.773	
South-Holland	-27.148	7.371	-3.680	<.001	*
Groningen	-42.410	7.376	-5.750	<.001	*
Forc. imit. * Flemish-Brabant	4.254	8.061	0.528	.599	
Forc. imit. * Limburg	1.668	8.061	0.207	.837	
Forc. imit. * South-Holland	18.819	8.061	2.334	.022	*
Forc. imit. * Groningen	40.209	8.061	4.988	<.001	*

Table 15: Estimated parameters of mixed-effects modelling on forced imitation data for voiceless fricatives.

Random effects	Variance	S.D.			
Participant (Intercept)	0.002	<0.001			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	15.608	2.106	7.410	<.001	*
Forced imitation	3.254	2.979	1.092	.277	
Flemish-Brabant	-1.085	2.979	-0.364	.717	
Limburg	3.816	2.979	1.281	.203	
South-Holland	0.142	2.979	0.048	.962	
Groningen	1.996	2.979	0.670	.505	
Forc. imit. * Flemish-Brabant	-3.717	4.213	-0.882	.380	
Forc. imit. * Limburg	-5.739	4.213	-1.362	.177	
Forc. imit. * South-Holland	0.928	4.213	0.220	.826	
Forc. imit. * Groningen	2.387	4.213	0.566	.826	

Table 16: Estimated parameters of mixed-effects modelling on forced imitation data for voiced stops

Random effects	Variance	S.D.			
Participant (Intercept)	183.040	13.529			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	82.240	4.650	17.686	<.001	*
Forced imitation	-14.110	4.994	-2.825	.006	*
Flemish-Brabant	-0.240	6.576	-0.036	.971	
Limburg	10.060	6.576	1.530	.129	
South-Holland	3.338	6.576	0.508	.613	
Groningen	6.128	6.576	0.932	.354	
Forc. imit. * Flemish-Brabant	-5.086	7.470	-0.681	.498	
Forc. imit. * Limburg	-3.453	7.129	-0.484	.630	
Forc. imit. * South-Holland	-9.435	7.063	-1.336	.185	
Forc. imit. * Groningen	-12.419	7.063	-1.758	.082	

Table 17: Estimated parameters of mixed-effects modelling on forced imitation data for voiceless stops.

Random effects	Variance	S.D.			
Participant (Intercept)	25.130	5.013			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	20.433	1.684	12.132	<.001	*
Forced imitation	-8.651	1.778	-4.866	<.001	*
Flemish-Brabant	-1.644	2.382	-0.690	.492	
Limburg	6.026	2.382	2.530	.013	*
South-Holland	0.220	2.382	0.092	.926	
Groningen	0.825	2.382	0.346	.730	
Forc. imit. * Flemish-Brabant	1.363	2.661	0.512	.610	
Forc. imit. * Limburg	-4.317	2.538	-1.701	.092	
Forc. imit. * South-Holland	1.041	2.514	0.414	.680	
Forc. imit. * Groningen	0.563	2.514	0.224	.823	

Note. Degrees of freedom (df) required for statistical significance testing of t values was given by $df = J - m - 1$ (Hox, 2010), where J is the most conservative number of second-level units ($J = 100$ speakers) and m is the total number of explanatory variables in the model ($m = 3$) resulting in $df = 96$.

Note. * Asterisks represent significant p values (α =.05).

Appendix M

Voicing ranges in the forced imitation task.

Table 18: Mean values and standard deviations of voicing ranges (in %) in the forced imitation task, split up by region.

Regions	Fricatives		Stops	
	Mean	SD	Mean	SD
West-Flanders	94.216	9.289	86.850	11.616
Flemish-Brabant	86.247	21.766	71.329	37.485
Limburg	96.010	3.180	88.544	18.076
South-Holland	80.276	24.126	82.839	26.908
Groningen	86.827	21.362	78.644	22.807
Total	88.714	18.491	81.641	25.214

.

Appendix N

Schematic representation of the design of the spontaneous imitation task.

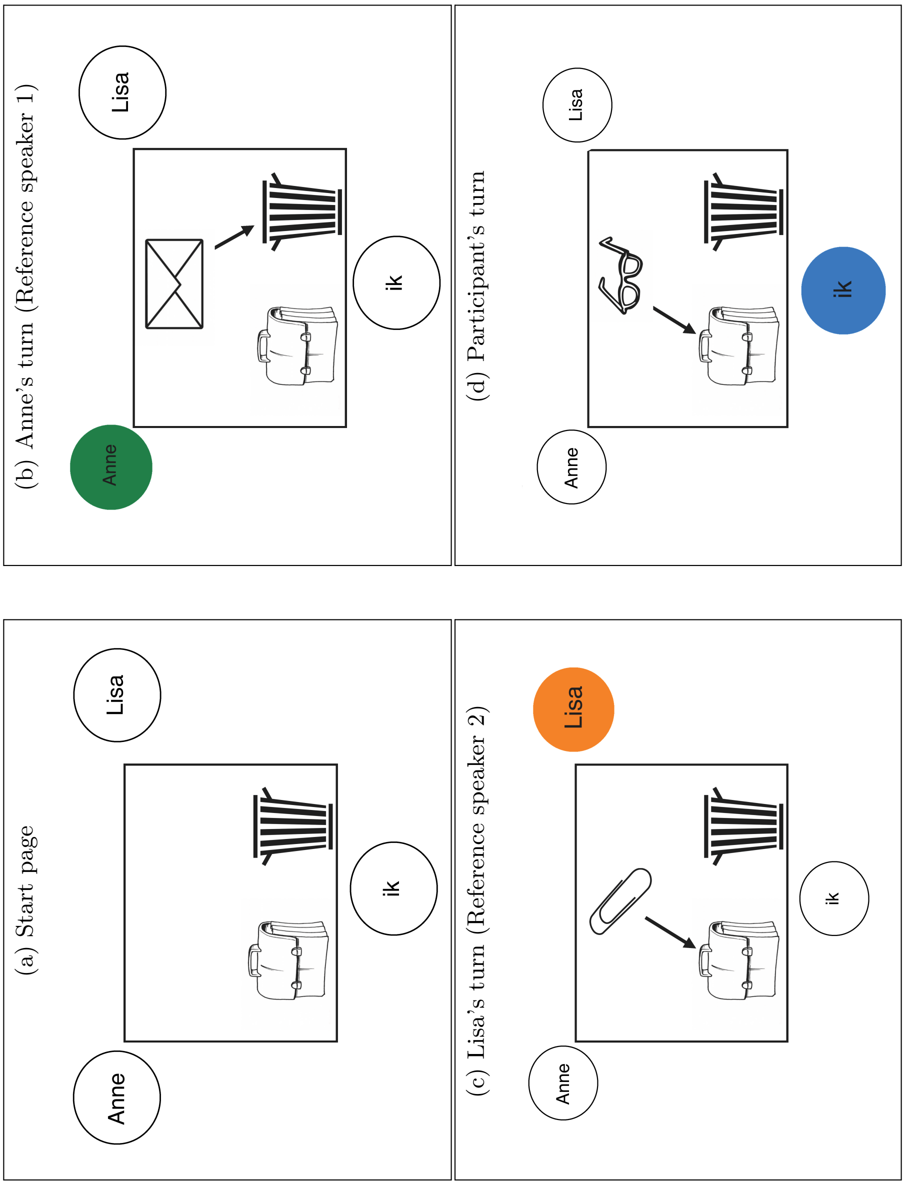


Figure 4: Schematic representation of the design of the spontaneous imitation task in four steps.

Appendix O

Results of the linear mixed-model regressions on the spontaneous imitation data.

Table 19: Estimated parameters of mixed-effects modelling on spontaneous imitation data for fricatives.

Random effects	Variance	S.D.			
Participant (Intercept)	71.675	8.466			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	68.093	4.164	16.352	<.001	*
Spont. imitation	-50.403	5.246	-9.609	<.001	*
Flemish-Brabant	-10.926	5.889	-1.855	.067	
Limburg	2.143	5.889	0.364	.717	
South-Holland	-27.148	5.889	-4.610	<.001	*
Groningen	-42.410	5.889	-7.202	<.001	*
Spont. imit. * Flemish-B.	19.562	7.418	2.637	.010	*
Spont. imit. * Limburg	-12.356	7.418	-1.666	.099	
Spont. imit. * South-Holland	17.122	7.418	2.308	.023	*
Spont. imit. * Groningen	28.083	7.418	3.786	<.001	*

Table 20: Estimated parameters of mixed-effects modelling on spontaneous imitation data for stops.

Random effects	Variance	S.D.			
Participant (Intercept)	110.09	10.492			
Fixed effects	Estimates	S.E.	t value	p value	
Intercept	82.240	4.234	19.422	<.001	*
Spont. imitation	-52.873	4.985	-10.606	<.001	*
Flemish-Brabant	-0.240	5.988	-0.040	.968	
Limburg	10.060	5.988	1.680	.096	
South-Holland	3.338	5.988	0.557	.579	
Groningen	6.128	5.988	1.023	.309	
Spont. imit. * Flemish-B.	9.306	7.050	1.320	.190	
Spont. imit. * Limburg	-5.819	7.050	-0.825	.412	
Spont. imit. * South-Holland	19.698	7.050	2.794	.006	*
Spont. imit. * Groningen	-4.100	7.050	-0.582	.562	

Note. Degrees of freedom (df) required for statistical significance testing of t values was given by $df = J - m - 1$ (Hox, 2010), where J is the most conservative number of second-level units ($J = 100$ speakers) and m is the total number of explanatory variables in the model ($m = 3$) resulting in $df = 96$.

Note. * Asterisks represent significant p values (α =.05).

Samenvatting

Introductie en methodologie

Het thema van deze dissertatie is het zogenaamde *actuation problem* (Weinreich et al., 1968) dat de vraag stelt waarom klankverandering in een bepaalde taalvariëteit op een bepaald moment plaatsvindt. Deze studie is opgezet om oplossingen voor dit probleem aan te dragen en om ons begrip van het fenomeen van klankverandering te verdiepen.

Onze benadering van klankverandering verschilt in drie opzichten van vorige studies. Ten eerste is er afstand genomen van de concepten innovatie en verspreiding en van de opvatting dat klankverandering een tweedelig proces is. We beschouwen de *actuation* als een iteratief proces op het individueel sprekersniveau, waarbij minimale veranderingen bij elke uiting accumuleren. Op deze manier wordt klankverandering gezien als een zuiver synchroon proces dat gebaseerd is op de variatie binnen individuele productie- en perceptiesystemen.

Ten tweede worden in deze studie de linguïstische subdomeinen sociolinguïstiek, fonetiek en psycholinguïstiek bij elkaar gebracht. De opzet van deze studie is sociolinguïstisch van aard. Het cross-sectioneel design wordt gecombineerd met methodes uit de psycholinguïstiek, de sociale psychologie en de fonetiek. Naast de analyse op groepsniveau besteden we ook aandacht aan de individuele sprekers en hun linguïstische geschiedenis.

Ten derde richten we ons op twee variabelen die in het Nederlands taalgebied aan klankverandering onderhevig zijn. Deze twee veranderingen lijken in vele opzichten op elkaar, maar verschillen in de mate van vordering. De eerste klankverandering is de *verstemlozing van initiële labiodentale fricatieven* (/v/-/f/). Deze verandering is relatief gevorderd. De tweede verandering is de *verstemlozing van initiële bilabiale plosieven* (/b/-/p/). Deze verandering is beginnend. De combinatie van deze twee veranderingen maakt het mogelijk om het iteratief proces van *actuation* bij elke fase van de vordering te bestuderen.

In onze studie worden vier aspecten van de linguïstische competentie betrokken bij klankverandering getest zijn op dezelfde groep proefpersonen. Deze vier aspecten zijn de processen van *spraakperceptie* en *spraakproductie* en de mechanismen waarbij linguïstische variabelen geëvalueerd en geïmiteerd worden: *taalattitudes* en *fonetische imitatie*.

De data zijn op twee niveaus geanalyseerd: het *groepsniveau* en het *individuele niveau*.

- Het *groepsniveau* bestond uit vijf regio's binnen het Nederlands taalgebied. Deze geografische regio's zijn gekozen om verschillende fases van klankverandering te vertegenwoordigen. Er waren drie regio's in Nederland: Groningen, Zuid-Holland en Nederlands Limburg, en twee regio's in Vlaanderen: West-Vlaanderen en Vlaams-Brabant. Binnen elke regio werden twintig proefpersonen geselecteerd. Bij de keuze van deze proefpersonen werden de factoren leeftijd (tussen 18 en 28 jaar) en opleidingsniveau (hoger onderwijs) constant gehouden. Op het groepsniveau werd er tussen regio's vergeleken.
- Op het *individuele niveau* werd er tussen proefpersonen uit dezelfde regio of uit verschillende regio's vergeleken.

De studie werd ontwikkeld binnen het kader van de *exemplar-based* theorie. In dit theoretische kader bestaan individuele linguïstische representaties uit een gedetailleerde representatie van alle voorbeelden van de woorden en fonemen die individuen ooit gehoord hebben. Bovendien wordt de individuele linguïstische representatie constant bijgewerkt onder invloed van de nieuwe voorbeelden die elk individu hoort. Aangezien geen twee individuen precies dezelfde spraakinput krijgen, is het heel waarschijnlijk dat individuen in hun perceptiepatronen verschillen. Op die manier kan de *examplar-based* theorie individuele variatie en de invloed van de omgevingstaal op het individu modelleren.

Daarnaast kunnen *exemplar-based* modellen het graduele karakter van klankverandering verklaren waardoor klankverandering als iteratief proces beschouwd kan worden. *Exemplar* theorie biedt een benadering waarin sociale informatie samen met linguïstische informatie opgeslagen wordt. Verder kan de perceptuele activering van een reeks voorbeelden tot een verschuiving in productie leiden, wat laat zien dat het imitatiemechanisme inherent aan het model is. Omwille van zijn theoretische voordelen werd deze studie verankerd binnen *exemplar-based* theorie.

Vijf onderzoeksvragen gerelateerd aan de rol van de verschillende aspecten van de linguïstische competentie in klankverandering werden geformuleerd.

Resultaten

De resultaten van de studie worden hier bondig gerapporteerd in de vorm van antwoorden op de vijf onderzoeksvragen (OV).

In *Hoofdstuk 4* werd er een grootschalig productie-experiment uitgevoerd om de status van de verandering van de twee variabelen in elke regio en voor elk individu te bepalen.

Zoals in vorige productiestudies bleken er regionale verschillen te zijn in de mate van fricatiefverstemlozing. In alle regio's en vooral in de Nederlandse regio's verdwijnt het contrast tussen initiële /f/ en /v/. Deze verandering blijkt verder gevorderd dan een tiental jaar geleden gerapporteerd door Kissine et al. (2003, 2005). Bij plosieven daarentegen bleken de verwachte regionale verschillen niet aanwezig te zijn. Verstemlozing bij plosieven blijkt enkel op *individueel niveau*. In verschillende regio's vertoont een klein deel van de sprekers duidelijke patronen van verstemlozing, vooral op de dimensie stemhebbendheid. Omdat *Voice Onset Time* geen betrouwbare meting bleek te zijn, hebben we de relatieve mate van stemhebbendheid bij plosieven gemeten.

Door beide klankverandering samen te nemen, werd inzicht verkregen in de relatie tussen verschillen tussen categorieën (/v/ en /f/, /b/ en /p/), het bereik van deze categorieën en de vordering van klankverandering. Deze inzichten hebben bijgedragen tot een nauwkeurige definitie van het mechanisme waarbij twee categorieën in productie samensmelten (*merger*). Vóór het begin van klankverandering is er een stabiele situatie met twee afzonderlijke categorieën van vergelijkbaar bereik. Aan het begin van klankverandering stijgt het bereik van de stemhebbende categorie (/v/ of /b/), terwijl het verschil tussen beide categorieën geleidelijk aan kleiner wordt. In de tweede fase van de verandering wordt het bereik van de stemhebbende categorie echter kleiner en het verschil tussen de categoriën blijft verkleinen. Bij de voltooiing van de verandering zijn beide categorieën samengesmolten (geen of heel klein verschil tussen /v/ en /f/, en /b/ en /p/ in productie) en alle klanken hebben een vergelijkbaar klein bereik.

OV 1: Wat is de role van spraakperceptie in klankverandering?

In *Hoofdstuk 5* werd de perceptie van labiodentale fricatieven en bilabiale plosieven onderzocht in twee perceptie-experimenten: een identificatietaak en een gelijkenisbeoordelingstaak. Aan de hand van de identificatietaak werd er kwantitatief inzicht verkregen in de regionale en individuele perceptiepatronen. Met de gelijkenisbeoordelingstaak kunnen we spraakperceptie op een kwalitatieve manier zonder dat de luisteraars klanken hoeven te identificeren. In die experimenten werd er gekeken naar regionale verschillen (op *groepsniveau*) en verschillen binnen regio's (op *individueel niveau*). Er bleken regionale verschillen te zijn in de perceptie van labiodentale fricatieven. De regio's waar de verstemlozing het sterkst in productie was, bleken ook de regio's te zijn waar de perceptie van het contrast het minst categorisch was. Luisteraars verschilden in hun perceptie en deze verschillen correleren met de staat van hun productie (zie *Hoofdstuk 4*). Hierdoor werd er een eerste type relatie vastgesteld tussen de perceptie en de productie van de variabelen.

Conclusie: Het bleek dat de variatie bij klankverandering niet enkel gekoppeld is aan spraakproductie, maar ook een perceptie-gerelateerd fenomeen is.

OV 2: Wat is de relatie tussen variatie in spraakperceptie en variatie in spraakproductie?

De analyse van evidentie voor het bestaan van een relatie tussen spraakperceptie en spraakproductie focuste op het groepsniveau in *Hoofdstuk 5*, en werd verfijnd in een analyse op individueel niveau in *Hoofdstuk 8*. Resultaten laten een duidelijke relatie zijn tussen perceptie en productie. Voor beide variabelen werd aangetoond dat hoe meer verstemlozing in een individuele productie, hoe minder categorisch zijn perceptie.

Conclusie: Hoewel productie en perceptie als twee afzonderlijke representatiesystemen beschouwd worden, zijn deze systemen gelinkt, en die link is cruciaal voor klankverandering.

OV 3: Gaat verandering in perceptie vooraf aan verandering in productie?

In *Hoofdstuk 8* werd de volgorde geanalyseerd waarin de productie- en perceptiesystemen zich tijdens klankverandering ontwikkelen. De analyses tonen aan dat − aan het begin van klankverandering − perceptiepatronen eerst veranderen. De eerste veranderingen in perceptie hebben gevolgen voor productiepatronen. Een spreker voegt dus eerst de nieuwe vorm in zijn perceptie toe, en pas later zal hij deze vorm beginnen produceren. Zodra de verstemlozing in productie gestart is, gaat dit proces zich volledig voltooien (tot de volledige verstemlozing van de stemhebbende categorie (/v/ en /b/)). Tegelijk werd er aangetoond dat de veranderingen in perceptie niet even ver raken: perceptie blijft achter bij de voltooiing van de verandering. Luisteraars zijn nog in staat om een categorisch contrast te horen waarbij ze geen contrast meer produceren.

Conclusie: Er werd aangetoond dat verandering in perceptie voorafloopt aan verandering in productie. Perceptie lijkt echter achter te blijven, als de verandering zich voltooit.

OV 4: Wat is de rol van taalattitudes in klankverandering?

In *Hoofdstuk 6* werd onderzocht hoe luisteraars de veranderende klank en zijn kenmerken beoordelen. Het doel was om te bepalen of luisteraars prestige toekennen aan de nieuwe variant en of er verschillen (op *groepsniveau* of *individueel niveau*) in attitudes bestaan. Voor fricatieven was er evidentie dat de verstemloosde variant als de prestigevariant beschouwd wordt. Op *groepsniveau* waren er regionale verschillen. Regio's met het minst verstemlozing toonden positievere attitudes ten opzichte van verstemlozing dan regio's waarin de verstemlozing meer gevorderd is. Het lijkt dat proefpersonen die uit regio's komen waar de verandering nog niet gevorderd is, meer geneigd zijn om de nieuwe vorm met positieve attitudes te associëren. Op die manier lijken positieve attitudes ten opzichte van een variabele een teken te zijn voor mogelijke verandering. Daarentegen hebben we extra evidentie nodig om te beweren dat de verstemlozing van plosieven met positieve attitudes geassocieerd wordt. Gezien de verstemlozing van plosieven een beginnend verandering is, is het mogelijk dat de verandering niet ver genoeg is omdat luisteraars nog geen sociale beoordeling samen de fonetische informatie hebben opgeslagen.

In *Hoofdstuk 8* zijn attitude-, productie- en perceptieresultaten met elkaar gekoppeld. Op *individueel niveau* werd de hypothese dat taalattitudes een rol spelen in productie bij klankverandering, bevestigd. Hoe meer verstemlozing van een variabele, hoe positiever verstemlozing wordt beoordeeld (op de solidariteitsdimensie). Maar de hypothese dat taalattitudes een rol spelen in perceptie werd niet direct bevestigd.

Conclusie: We hebben aangetoond dat attitudes ten opzichte van een veranderende variabele niet vanaf het begin van de verandering aanwezig zijn. De positieve evaluatie van de klankverandering lijkt zich pas te ontwikkelen als het productiesysteem al aan verandering onderhevig is. Taalattitudes ontstaan dus tijdens het proces van klankverandering en versterken dit proces. Dit resultaat stemt overeen met vorig sociolinguïstisch onderzoek waar de rol van taalattitudes traditioneel op het verspreidingsproces van taalverandering onderzocht wordt (en niet zo zeer op de innovatie van een verandering, gezien het te vroeg is om attitudes te observeren in dat stadium) en waar de ontwikkeling van attitudes als een groepsgerelateerd fenomeen beschouwd wordt.

OV 5: Wat is de rol van fonetische imitatie in klankverandering?

In *Hoofdstuk 7* werd in twee imitatie-experimenten (een gedwongen en een spontane imitatietaak) getest in welke mate individuen veranderende klanken kunnen imiteren (op *individueel niveau*) en of deze capaciteit tussen regio's verschilt (op *groepsniveau*). Het doel van de gedwongen imitatietaak was om inzicht te krijgen in de motorische productiecapaciteiten. Door proefpersonen ertoe te dwingen een modelspreker zo goed mogelijk te imiteren, werd hun maximale productiebereik gemeten. Met uitzondering van een beperkt aantal individuen konden alle proefpersonen het hele bereik van stemhebbendheid in fricatieven en plosieven produceren. Zelfs in regio's waar producties al verstemloosd waren, waren de sprekers in staat om stemhebbende klanken imiteren. De spontane imitatietaak onderzocht de bereidwilligheid van proefpersonen om in sociale interactie te verstemlozen. Deze taak heeft aangetoond dat eenvoudige blootstelling aan modelsprekers de imitatie van verstemloosde fricatieven en plosieven veroorzaakt. Op *groepsniveau* werd spontane imitatie vastgesteld in alle regio's, maar regio's verschilden significant in de mate van imitatie. Op *individueel niveau* vertonen een klein aantal proefpersonen stijgende imitatiepatronen in de loop van het experiment.

In *Hoofdstuk 8* werden individuele imitatieresultaten geanalyseerd in relatie met de link perceptie en productie. Uit de resultaten blijkt dat beste imitatoren progressieven zijn in de verandering. Correlaties tussen productie- en imitatiemetingen toonden dat de beste imitatoren individuen waren die in de beginfase van de verandering zaten en die binnen hun regio conservatieve sprekers waren. Individuele verschillen in imitatiecapaciteiten werden toegewijt aan verschillen in de sterkte van de link tussen productie- en perceptiesysteem, ongeacht de kwaliteit van hun perceptiesysteem.

Conclusie: Er werd aangetoond dat imitatiecapaciteiten gerelateerd zijn aan individuele productiepatronen. De beste imitatoren bleken aan het begin van klankverandering te zitten en conservatieve productiepatronen te vertonen. Hoe meer gevorderd in de verandering, hoe slechter de imitatiecapaciteiten. Slechte imitatoren lijken verandering te leiden omwille van hun zwakke link tussen spraakproductie en spraakperceptie.

Conclusie

In een reeks experimenten hebben we geprobeerd betere grip te krijgen op het fenomeen van klankverandering en de manier waarop het op gang komt (*actuation*). De resultaten hebben inzicht gegeven in de rol van verschillende aspecten van de linguïstische competentie in klankverandering: spraakperceptie, spraakproductie, fonetische imitatie en taalattitudes.

We gebruiken het *dual representation model* (Garrett & Johnson, 2013) waarin spraakperceptie en -productie op verschillende representatiesystemen gebaseerd zijn. Bij elke individuele taalgebruiker is het perceptuele systeem gescheiden van het productiesysteem. Gezien luisteraars varianten waarnemen die ze zelf niet kunnen produceren, is het perceptuele systeem breder en meeromvattend dan het productiesysteem. Hoewel productie en perceptie als afzonderlijke systemen beschouwd worden, zijn beide systemen aan elkaar gelinkt. Sprekers gebruiken hun capaciteit om de details van een klank die ze horen, te vertalen in hun eigen productie via het imitatiemechanisme. Daarnaast nemen luisteraars spraak waar via de filter van hun eigen productie. Positieve attitudes geassocieerd met specifieke linguïstische kenmerken versterken dit proces.

Onze resultaten pleiten voor de opvatting van klankverandering als een iteratief proces waarbij individuele productiesystemen gradueel veranderen in elke uiting. Hoe deze verandering op gang komt, is gerelateerd aan individuele verschillen in de kwaliteit van de productie- en perceptiesysteem en aan individuele verschillen in de sterkte van de link tussen deze systemen. De discrepantie tussen het productie- en het perceptiesysteem vormt de kern van het proces van klankverandering. We stellen voor dat klankverandering *binnen het individu* op gang komt elke keer het productie- en perceptiesysteem aan elkaar gekoppeld worden. Vervolgonderzoek zal helpen bepalen hoe de richting van klankverandering door externe factoren beïnvloed wordt (extern omdat ze niet binnen het individu plaatsvinden), zoals articulatorische, distributionele en functionele factoren.

Deze dissertatie heeft bijgedragen tot de studie van klankverandering en tot de oplossing van het *actuation problem* met een sociofonetisch onderzoek van de rol van vier aspecten van de linguïstische competentie in klankverandering. Het heeft onze begrip van deze processen op groeps- en individueel niveau versterkt.

Résumé

Introduction et méthodologie

Le problème du déclenchement du changement phonétique (*actuation problem*) concerne la question du pourquoi un changement phonétique se déclenche, à un moment particulier, dans une variété langagière particulière. Cette thèse doctorale a pour objectif de contribuer à une solution à ce problème et de mieux comprendre le phénonème de changement phonétique.

L'approche scientifique du changement phonétique adoptée dans cette étude se distingue des recherches antécédentes de trois manières.

Premièrement, les notions d'innovation et de propagation du changement phonétique sont mises de coté, ainsi que la conceptualisation du changement phonétique en tant que procédé en deux étapes. Le déclenchement du changement (*actuation*) est donc considéré, dans nos travaux, comme un processus itératif au *niveau individuel*, au sein duquel tout changement minimal s'accumule chaque fois qu'un locuteur s'adresse à un auditeur. Par conséquent, le changement phonétique est redéfini comme processus purement synchronique qui réside dans la variation existante à l'intérieur des systèmes individuels de perception et de production.

Deuxièmement, cette étude cherche à rapprocher les domaines de sociolinguistique, de phonétique et de psycholinguistique en alliant leurs avantages respectifs. Le protocole expérimental de cette étude s'inscrit dans la tradition des études sociolinguistiques cross-sectionnelles. Les méthodes expérimentales sont inspirées des domaines de psycholinguistique et de phonétique. Une attention particulière est portée aux locuteurs individuels et à leur histoire linguistique. Notre étude se concentre sur la variation inter- et intra-locuteurs.

Troisièmement, cette étude traite de deux changements linguistiques en cours dans la langue néerlandaise standard actuelle. Ces changements sont comparables sur plusieurs points. Néanmoins, il est important de noter qu'ils se distinguent par leurs degrés d'achèvement. Le premier changement (le dévoisement des fricatives labio-dentales) constitue un changement attesté comme avancé; le second (le dévoisement des plosives bilabiales) est un changement naissant. La juxtaposition de ces deux changements permet alors l'observation du processus itératif d'actuation à chaque étape d'avancement.

La mise en oeuvre de ces objectifs de recherche conduit au développement d'un dispositif expérimental de grande ampleur où quatre aspects centraux de la compétence linguistique impliquée dans le changement phonétique sont testés sur une même population. Ces quatre aspects de la compétence linguistique sont les systèmes de perception et de production, et les mécanismes grâce auxquels les variables linguistiques sont évaluées et imitées, les attitudes langagières et l'imitation phonétique. Dès lors, une grande étude cross-sectionnelle a été développée de manière à mettre en relation les systèmes de production, perception, imitation et attitudes des locuteurs individuels d'une communauté linguistique.

Cette étude s'articule autour de deux niveaux de participants: le niveau *groupal* et le niveau *individuel*.

- Le niveau *groupal* est constitué de cinq régions situées sur le territoire néerlandophone de Belgique et des Pays-Bas. Les trois régions des Pays-Bas sont la région de Groningue, de la Hollande-Méridionale et du Limbourg, et les deux régions de Flandre (Belgique néerlandophone) sont le Brabant-Flamand et la Flandre occidentale. Ces différentes régions géographiques sont sélectionnées pour représenter les phases de changement phonétique. On compare, au niveau groupal, les régions entre elles. Au sein de chaque région, une population de vingt participants est sélectionnée. Les facteurs âge (de 18 à 28 ans) et niveau d'éducation (études supérieures) sont maintenus constants.
- Les comparaisons s'effectuent, au niveau *individuel*, entre participants d'une même région et entre les participants de différentes régions.

Cette étude s'inscrit dans le cadre de la théorie d'instances (*exemplar-based theory*) récemment développée. Dans ce cadre théorique, les représentations linguistiques individuelles consistent en une série détaillée de toutes les instances, tous les exemples des mots et des phonèmes que les individus ont entendus jusqu'à présent. De plus, cette représentation des instances continue à être mise à jour au cours de la vie des individus et au gré des nouvelles instances qu'ils perçoivent. Considérant le fait que chaque individu est confronté à des instances différentes au cours de sa vie, il est plus que probable que les systèmes de perception diffèrent d'un individu à un autre. La théorie des instances permet donc de modéliser la variation individuelle et l'influence de la langue ambiante sur le locuteur individuel.

Par ailleurs, le modèle des instances donne les moyens d'expliquer le caractère graduel du changement phonétique de façon à ce qu'il soit conçu comme un processus itératif. La théorie des instances est la première à permettre que l'information sociale soit enregistrée en relation avec l'information linguistique. De plus, l'activation perceptuelle de certaines instances a la capacité d'influencer directement la production de ces mêmes instances, incluant directement la capacité d'imitation au modèle.

Cinq questions de recherche centrales, chacune liée à un aspect particulier de la compétence linguistique impliquée dans le changement phonétique, ont été formulées.

Les résultats

Les résultats de notre étude sont décrits succinctement sous forme de réponse aux cinq questions de recherche (QR) principales.

Le *Chapitre 4* présente les résultats du test de production dont l'objectif est de déterminer le stade de changement phonétique pour les deux variables (le dévoisement des fricatives et des plosives), et ce pour chaque individu. Comme le décrivent les études antérieures, on constate de claires différences régionales (différences groupales) dans le degré de dévoisement des fricatives. Dans chaque région, à commencer par les régions des Pays-Bas, le contraste entre fricatives sourdes et sonores se perd. On observe que ce changement phonétique est encore plus avancé que lors de sa description par Kissine et al. (2003, 2005).

Les plosives, au contraire, n'ont pas révélé les différences régionales attendues. Les traces de dévoisement des plosives n'apparaissent que lors de l'analyse au niveau individuel. Dans plusieurs régions, une minorité de participants présentent des signes clairs de dévoisement, surtout au niveau de la dimension phonétique de sonorité.

La juxtaposition des résultats productionnels pour les deux variables nous permet de comprendre les relations entre la différence entre catégories (/f/ et /v/, /b/ et /p/), l'étendue de ces catégories et le développement du changement phonétique. Ces résultats nous permettent de redéfinir le mécanisme de fusion de catégorie en production. Avant le déclenchement du changement, on observe une situation stable avec deux catégories phonétiques distinctes et une étendue comparable. Au déclenchement du changement, la catégorie qui va fusionner (/v/ ou /b/) accroît son étendue, tandis que la différence entre les catégories diminue de manière constante. Dans la phase de changement suivante pourtant, l'étendue de la catégorie sonore diminue, alors que la différence entre catégories continue à diminuer. A l'achèvement du changement, les catégories ont fusionné et leur étendue est minimale.

QR 1: Quel est le rôle de la perception dans le changement phonétique?

Dans le *Chapitre 5*, deux expériences centrées sur la perception ont pour but l'analyse des processus perceptuels des fricatives labio-dentales et des plosives bilabiales, à savoir une expérience d'identification phonétique et une expérience de jugement de similarité. Le test d'identification nous permet d'étudier de manière quantitative le système de perception aux niveaux régional et individuel. Le test de jugement de similarité, en revanche, vise à analyser la perception de manière plus qualitative sans que les auditeurs ne doivent identifier les réalisations phonétiques. Grâce à ces expériences, il est possible de déterminer si la perception des fricatives labio-dentales et des plosives bilabiales présente des différences régionales dans la communauté néerlandophone (au niveau *groupal*) et des différences entre individus dans chacune de ses régions (au niveau *individuel*).

L'analyse des deux tests a revélé des différences régionales dans la perception des phonèmes soumis au changement. Les régions où le dévoisement est très avancé s'avèrent également être les régions où la perception est la moins catégorielle. Les auditeurs présentent des différences perceptuelles qui correspondent au système productionnel (présenté dans le *Chapitre 4*). Ces résultats contribuent à l'établissement d'un premier lien entre perception et production. En conclusion, il apparaît clairement que la variation qui est essentielle au changement phonétique ne se limite pas aux systèmes de production, mais qu'elle est aussi présente au sein des systèmes de perception.

QR 2: Quel est le lien entre la variation au sein des systèmes de perception et la variation au sein des systèmes de production?

L'analyse du lien entre les systèmes de perception et de production s'est concentrée au niveau groupal dans le *Chapitre 5*. Le *Chapitre 8* poursuit l'analyse de ce lien au niveau individuel. Les résultats de cette analyse individuelle démontrent un lien évident entre perception et production. Pour chacune des variables, il est prouvé que plus un individu a tendance à dévoiser sa production, moins sa perception sera catégorielle. Même s'il est nécessaire de postuler deux systèmes de représentation différents pour la production et la perception (*dual representation*), ces systèmes sont clairement liés l'un à l'autre, et ce lien est central dans le processus de changement phonétique.

QR 3: Les changements perceptuels précèdent-ils les changements productionnels?

Dans le *Chapitre 8*, l'ordre dans lequel les systèmes de production et de perception évoluent est inspecté. Il ressort des analyses qu'au début du changement, les systèmes perceptuels sont ajustés en premier lieu. Ces ajustements perceptuels semblent alors avoir des conséquences directes sur le système de production. Le locuteur incorpore d'abord la nouvelle variante à son système de perception, et ce n'est que dans une seconde étape qu'il commence à utiliser cette forme en production. Dès lors, une fois que le dévoisement est lancé en production, l'étape finale est nécessairement le dévoisement total de la catégorie. En revanche, l'étape finale du changement pour la perception n'est pas équivalente à celle pour la production: à la fin du changement, le système de perception est à la traîne. Les participants sont encore capables de percevoir le contraste catégoriel d'une variable dans une certaine mesure, alors qu'ils ont totalement perdu la capacité de produire ce même contraste.

En conclusion, il est démontré que les changements perceptuels précèdent les changements productionnels. Néanmoins, il apparaît que, vers la fin du processus, le système de perception est en retard.

QR 4: Quel est le rôle des attitudes dans le changement phonétique?

Dans le *Chapitre 6*, il est question des jugements que les auditeurs associent à une variable soumise au changement phonétique. L'objectif est de déterminer dans quelle mesure une nouvelle forme est évaluée comme prestigieuse par certains groupes d'auditeurs (au niveau *groupal*) et par certains auditeurs individuels (au niveau *individuel*). En ce qui concerne le dévoisement des fricatives, il apparaît clairement que les fricatives labio-dentales dévoisées sont considérées comme prestigieuses. Au niveau *groupal*, on constate des différences régionales. Les régions qui comportent le moins de dévoisement montrent des attitudes plus positives envers ce dévoisement que les régions où le dévoisement est plus avancé. Il semble donc que les participants issus des régions où le changement n'est pas très avancé sont plus susceptibles d'associer des attitudes positives à la nouvelle variante. De cette manière, on observe que l'évaluation positive d'une nouvelle forme est un signal pour un changement phonétique qui est encore potentiel. Contrairement aux fricatives, le dévoisement des plosives ne semble pas directement associé à des attitudes positives. Vu que le dévoisement des plosives est encore dans une phase très primaire, il est possible que les participants n'aient pas encore encodé d'évaluation sociale pour cette nouvelle variante.

Dans le *Chapitre 8*, les résultats d'attitude, de production et de perception sont mis en relation. Au niveau *individuel*, l'hypothèse proposant que les attitudes jouent un rôle dans la production dans le contexte du changement phonétique est confirmée. Les résultats montrent que plus une variable est dévoisée, plus ce dévoisement est évalué positivement (selon la dimension de solidarité). En revanche, l'hypothèse proposant que les attitudes jouent un rôle dans la perception n'a pas été directement confirmée.

En conclusion, les attitudes envers une variable soumise au changement phonétique ne semblent pas être présentes dès le début du processus. L'évaluation positive d'une nouvelle variante se forme au fur et à mesure que le changement progresse et contribue à renforcer le développement en cours. Ce résultat correspond aux analyses sociolinguistiques traditionnelles qui examinent le rôle des attitudes dans le processus de propagation du changement (plutôt que dans l'innovation de ce changement, puisqu'il est trop tôt, à ce stade, pour observer des attitudes) et correspondent, par ailleurs, à l'idée que le développement des attitudes est intrinsèquement un phénomène lié au niveau groupal.

QR 5: Quel est le rôle de l'imitation dans le changement phonétique?

Le *Chapitre 7* présente deux expériences d'imitation (un test d'imitation volontaire et un test d'imitation spontanée) qui visent à déterminer dans quelle mesure les individus imitent les variables soumises au changement (au niveau *individuel*) et dans quelle mesure cette capacité diffère entre régions (au niveau *groupal*). Le but du test d'imitation volontaire est d'analyser la capacité de production motorique. En obligeant les participants à imiter un locuteur modèle le mieux qu'ils peuvent, l'étendue maximale de leur production peut être mesurée. A l'exception de quelques individus, tous les participants se sont montrés capables de produire toute l'étendue de sonorité des fricatives et des plosives. Même les participants issus des régions où les productions sont largement dévoisées parviennent à imiter les consonnes sonores. Le test d'imitation spontanée met l'accent sur la disposition des participants à imiter au sein d'interactions sociales et permet de démontrer que la simple exposition à un locuteur modèle peut conduire un individu à imiter le dévoisement des fricatives et des plosives. Au niveau *groupal*, l'imitation spontanée est observée dans toutes les régions, celles-ci diffèrent néanmoins dans l'intensité de l'imitation. Au niveau *individuel*, un nombre limité de participants montrent une imitation croissante au cours de l'expérience.

Dans le *Chapitre 8*, les résultats individuels d'imitation sont examinés en relation avec les données sur le lien entre perception et production. Il ressort que les moins bons imitateurs sont plus loin dans le changement que les meilleurs imitateurs. Les corrélations entre les mesures de production et d'imitation montrent que les meilleurs imitateurs sont au début du changement phonétique et sont les plus conservateurs au sein de leur région. Les différences individuelles de capacité d'imitation sont attribuées aux différences d'intensité du lien entre les systèmes de perception et de production, quelle que soit la qualité du système de perception.

En conclusion, il est démontré que les capacités d'imitation phonétique sont étroitement liées aux stades d'avancement des productions individuelles. Ainsi, les meilleurs imitateurs ont amorcé le changement phonétique et présentent une production conservatrice. Plus le changement progresse, plus les capacités d'imitation sont réduites. Les moins bons imitateurs semblent dès lors mener le changement à cause de leur lien affaibli entre les systèmes de production et de perception.

Conclusion

Cette thèse présente une série d'expériences qui visent à mieux comprendre le phénomène de changement linguistique et son déclenchement. Les résultats fournissent des explications sur le rôle de différents aspects de la compétence linguistique dans le changement phonétique, c'est-à-dire les systèmes de perception et de production, la capacité d'imitation et les attitudes langagières.

Nous adoptons le modèle de la double repésentation (*dual representation*) où les systèmes de production et de perception sont basés sur des ensembles d'instances différents Garrett and Johnson (2013). En d'autres mots, pour chaque locuteur, le système de perception est séparé du système de production. Etant donné le fait que les auditeurs sont capables de percevoir des variantes qu'ils ne peuvent produire, le système de perception est nécessairement plus large que le système de production.

Malgré que le fait que les systèmes soient séparés, ils sont liés l'un à l'autre grâce à deux mécanismes. En premier lieu, les locuteurs utilisent leur capacité à transformer les détails des phonèmes qu'ils perçoivent en production par le biais du mécanisme d'imitation. En second lieu, les auditeurs décodent la parole par le biais de leur propre production qui fonctionne comme un filtre. Les attitudes langagières positives associées à certaines propriétés linguistiques viennent renforcer ces deux phénomènes.

Si l'on considère le modèle individuel de fonctionnement, ces recherches plaident pour la conceptualisation du changement phonétique en tant que processus itératif lors duquel la production individuelle change graduellement chaque fois qu'un locuteur s'adresse à un auditeur. Le déclenchement du changement est directement lié aux différences individuelles dans la qualité des systèmes de production et de perception et aux différences de solidité du lien qui unit ces systèmes. La divergence qui existe au sein de ces systèmes est au centre du processus de changement. Nous proposons, dès lors, qu'il y a 'déclenchement' (*actuation*) chaque fois qu'au niveau individuel, les systèmes de production et de perception tentent de s'aligner l'un sur l'autre. Des travaux futurs permettront sans doute de définir comment la directionalité du changement est influencée par des facteurs externes (externes dans le sens où ils ne se situent pas au niveau individuel), comme les contraintes articulatoires, distributionnelles et fonctionnelles sur les systèmes linguistiques existants.

Cette thèse contribue à l'étude du changement phonétique et son déclenchement grâce à l'exploration du rôle de quatre aspects de la compétence linguistique engagée dans le processus de changement. Elle permet d'étendre notre compréhension des procédés aux niveaux individuel et groupal en analysant, non seulement ces deux niveaux, mais aussi leur relation.

Curriculum Vitae

Anne-France Pinget was born on July 30^{th} 1987 in Charleroi (Belgium). In 2005, she obtained her Gymnasium diploma from the Collège Saint-Joseph in Chimay (Belgium). In the same year, she started her bachelor in Language and Literature (Dutch and German) at the University of Namur (Belgium). She participated in the Erasmus-programme and spent one study semester at the Free University Berlin (Germany). She obtained her BA in 2008. In 2009, she received her MA in Language and Literature (Dutch and German) from the University of Ghent (Belgium). She specialized in Linguistics and obtained her ResMA in 2011 at the Utrecht University (The Netherlands). Between 2011 and 2015, she worked as a PhD at the Utrecht Institute of Linguistics OTS (The Netherlands). During her PhD, she spent a semester in Bologna (Italy) at the Collegio dei Fiamminghi.