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

Voeten, C.C.

Citation

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

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

CHAPTER 3

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

This chapter has been submitted.

Abstract

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

3.1 Introduction

3.1.1 Investigating the adoption of on-going sound change

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

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

3.1.2 Sound change as second-dialect acquisition

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

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

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

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

3.1.3 The present study

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

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

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

3.2 Experiment 1: rhyme decision

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

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

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

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

3.2.1 Method

Participants

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

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

Stimuli

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

---

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

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

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

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

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

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

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

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

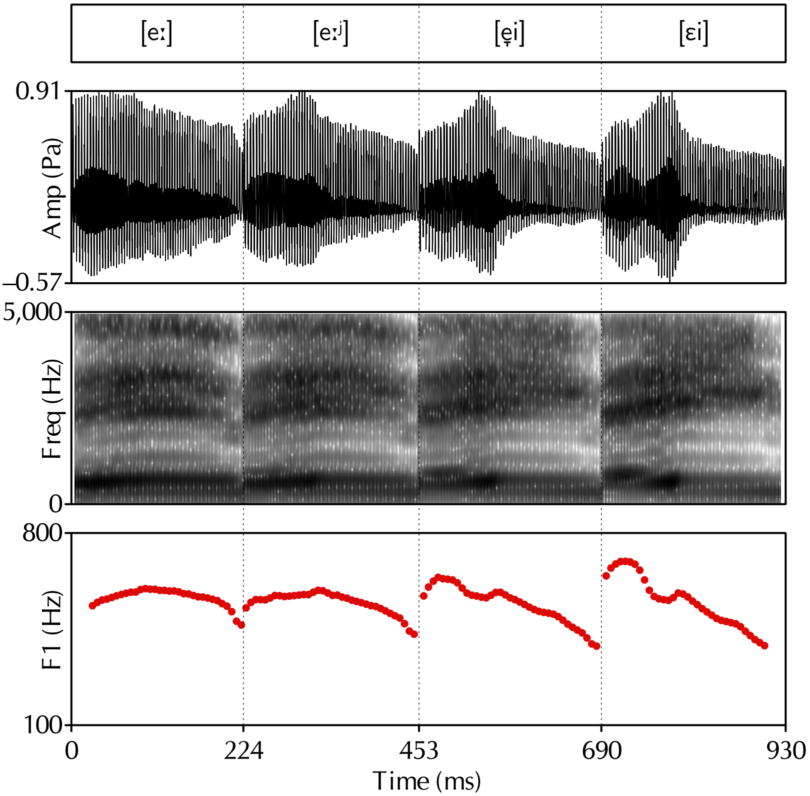

Figure 3.1: Example waveforms, spectrograms, and F1 trajectories for the [eː∼ɛi] contrast. The four tokens shown in this figure correspond to the stimuli containing 20%, 40%, 60%, and 80% morphing, respectively. Note the lowering of the nucleus and the increase in the diphthong's slope over the four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

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

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

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

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

Procedure

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

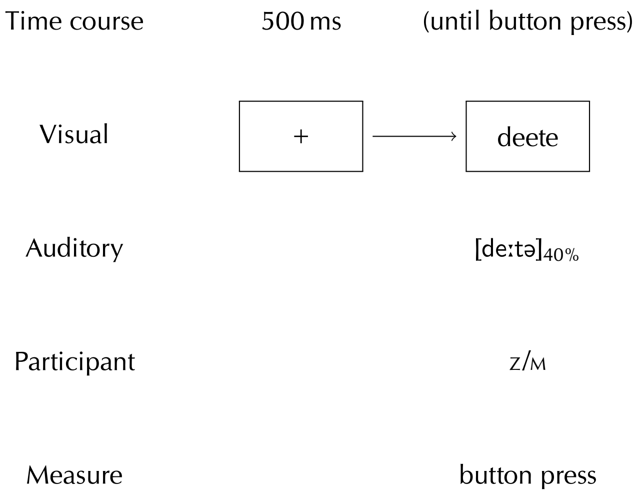

Figure 3.2: Example trial for the rhyme-decision task. At a later trial, the same auditory stimulus will be presented, but the visual target ⟨deete⟩ (/deːtə/) will be replaced by ⟨dijte⟩ (/dɛitə/).

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

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

Data analysis

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

3.2.2 Results

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

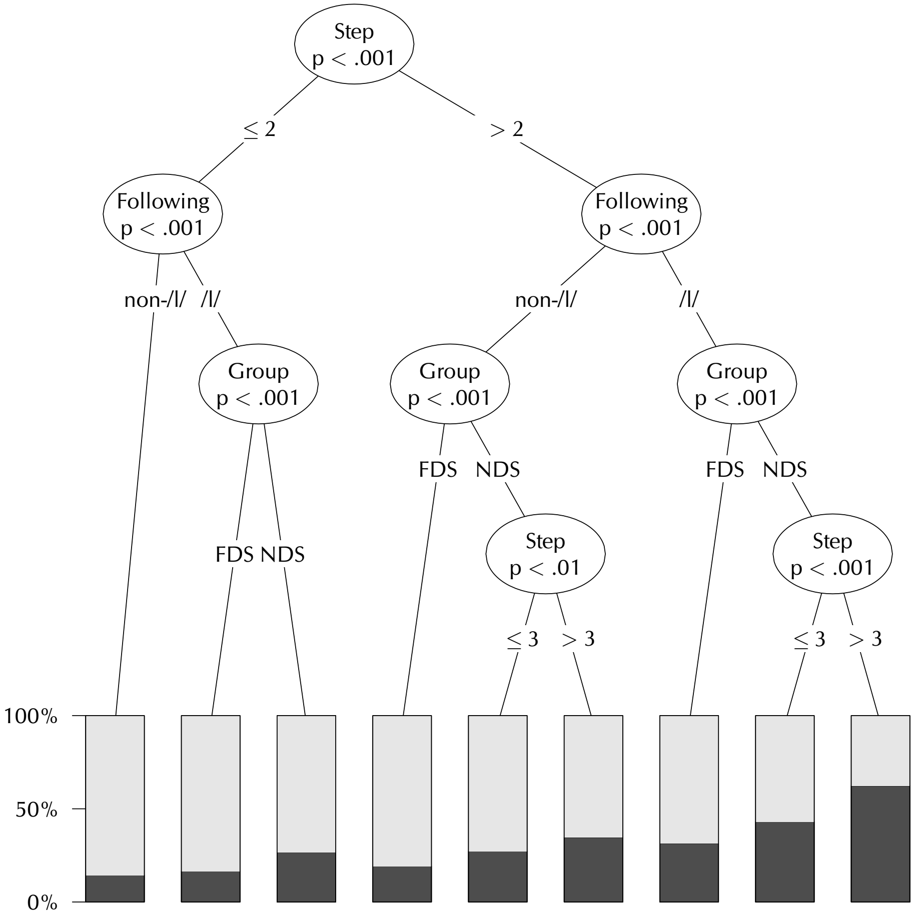

Figure 3.3: Logistic mixed-effects regression tree for the [eː∼ɛi] continuum (20 participants, 68 items). The target variable is the probability of indicating an [ɛi] percept.

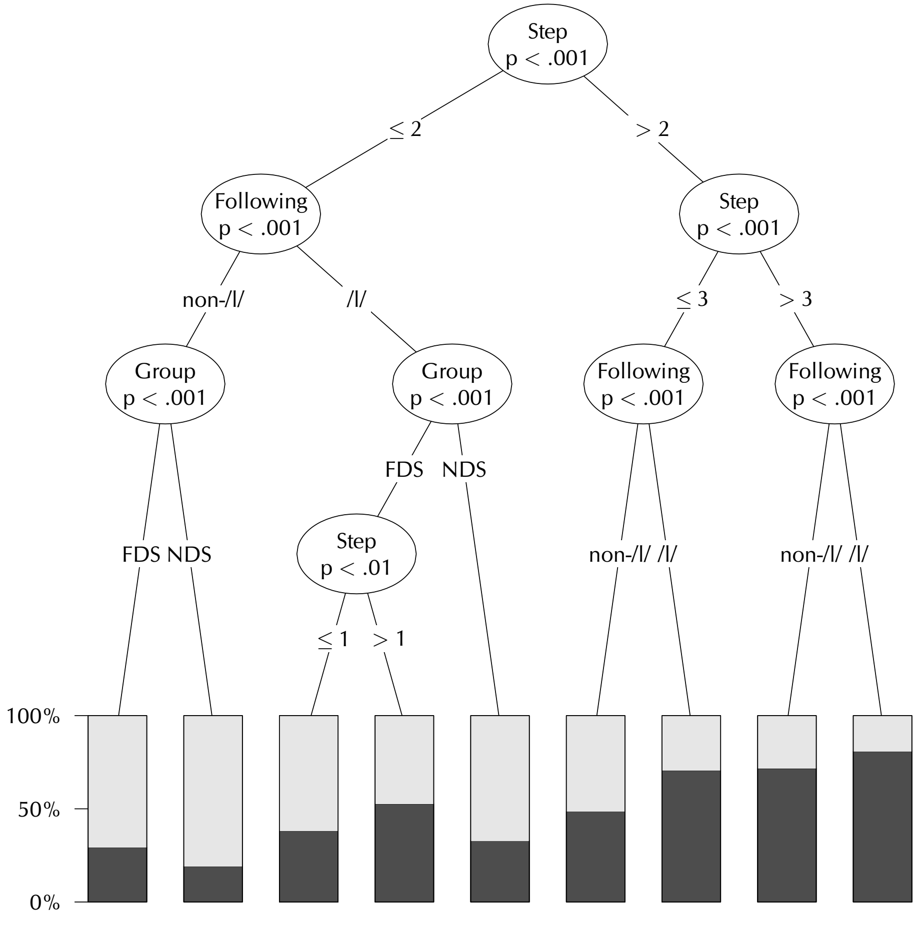

Figure 3.4: Logistic mixed-effects regression tree for the [oː∼ɑu] continuum (20 participants, 68 items). The target variable is the probability of indicating an [ɑu] percept.

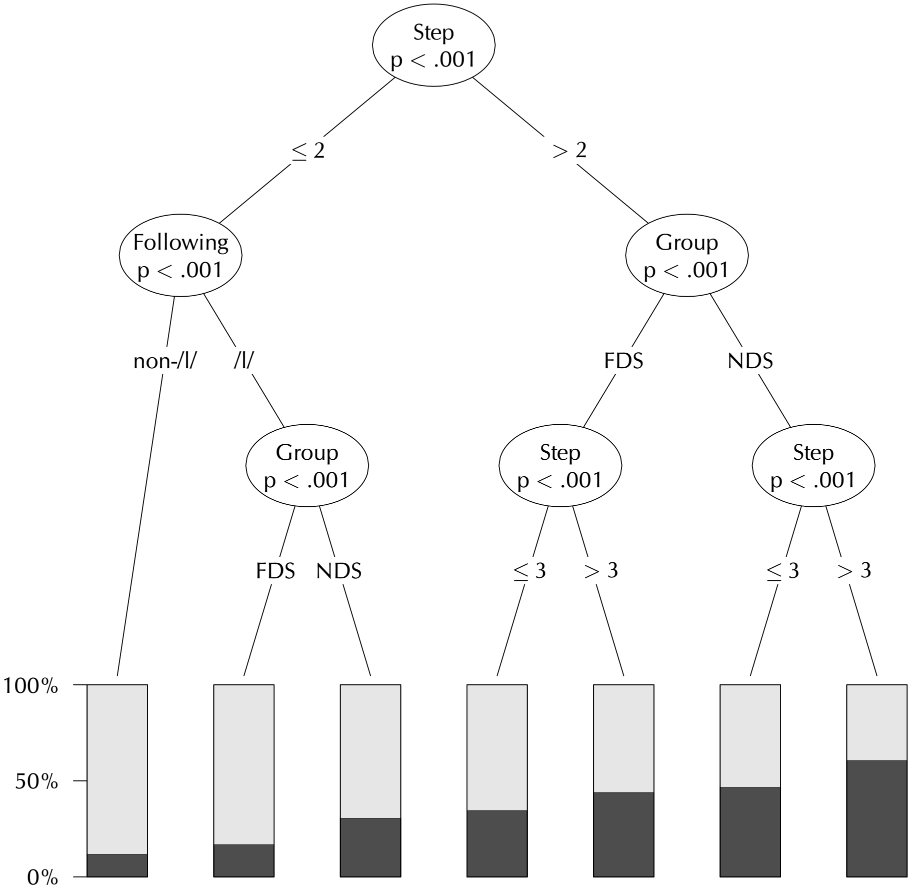

Figure 3.5: Logistic mixed-effects regression tree for the [ɛ∼ɛi] continuum (20 participants, 68 items). The target variable is the probability of indicating an [ɛi] percept.

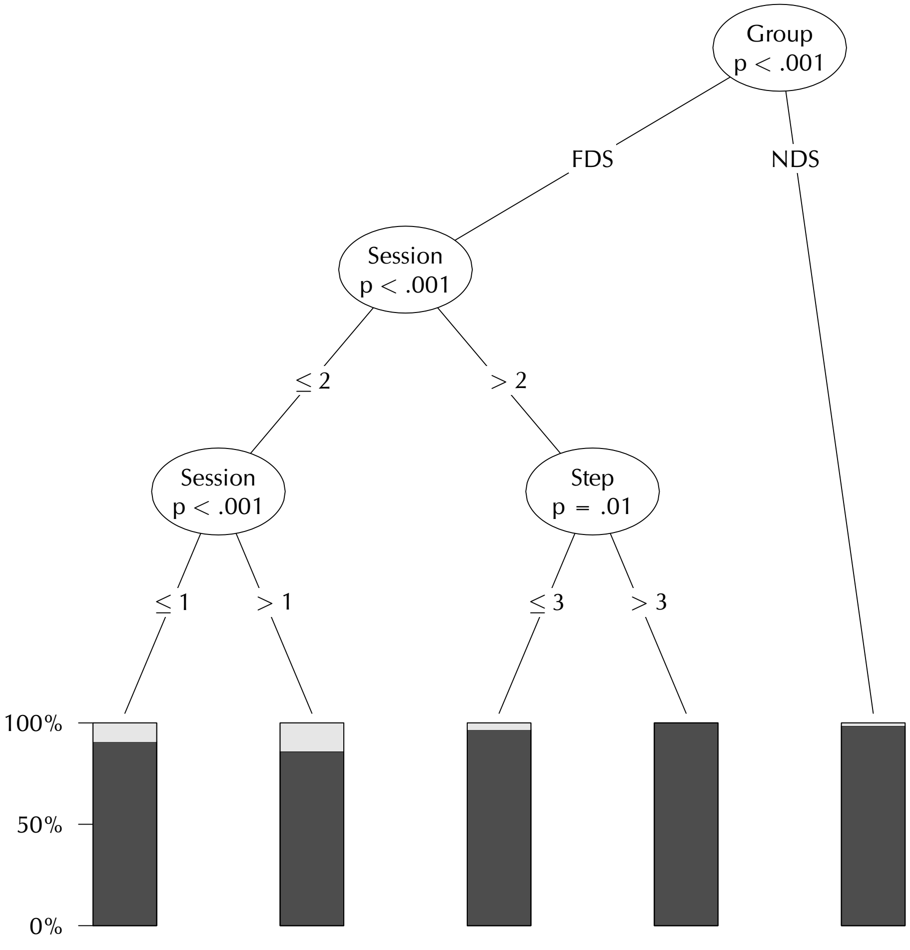

Figure 3.6: Logistic mixed-effects regression tree for the [ʀ∼ɹ] continuum (20 participants, 34 items). The target variable is the probability of indicating an [ɹ] percept.

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

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

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

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

3.2.3 Discussion

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

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

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

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

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

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

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

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

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

3.3 Experiment 2: word production

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

3.3.1 Method

Participants

The participants were the same as in Experiment 1.

Stimuli

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

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

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

---

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

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

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

Procedure

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

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

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

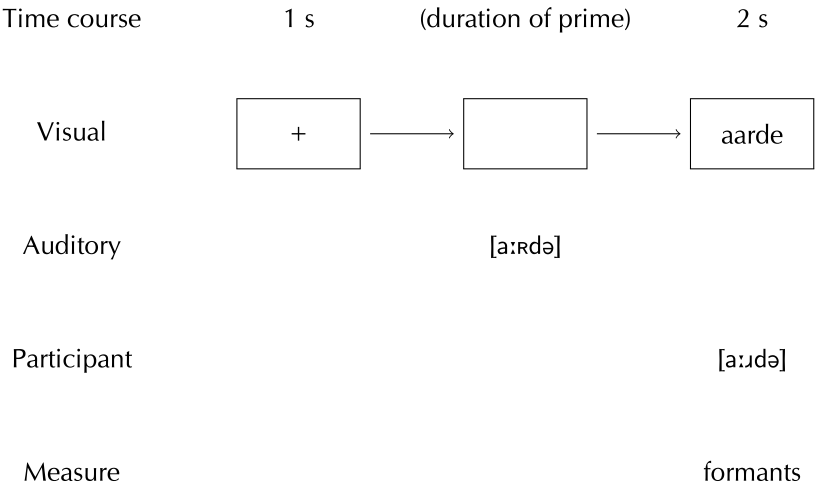

Figure 3.7: Example trial for the production task.

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

Data analysis

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

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

---

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

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

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

3.3.2 Results

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

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

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

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

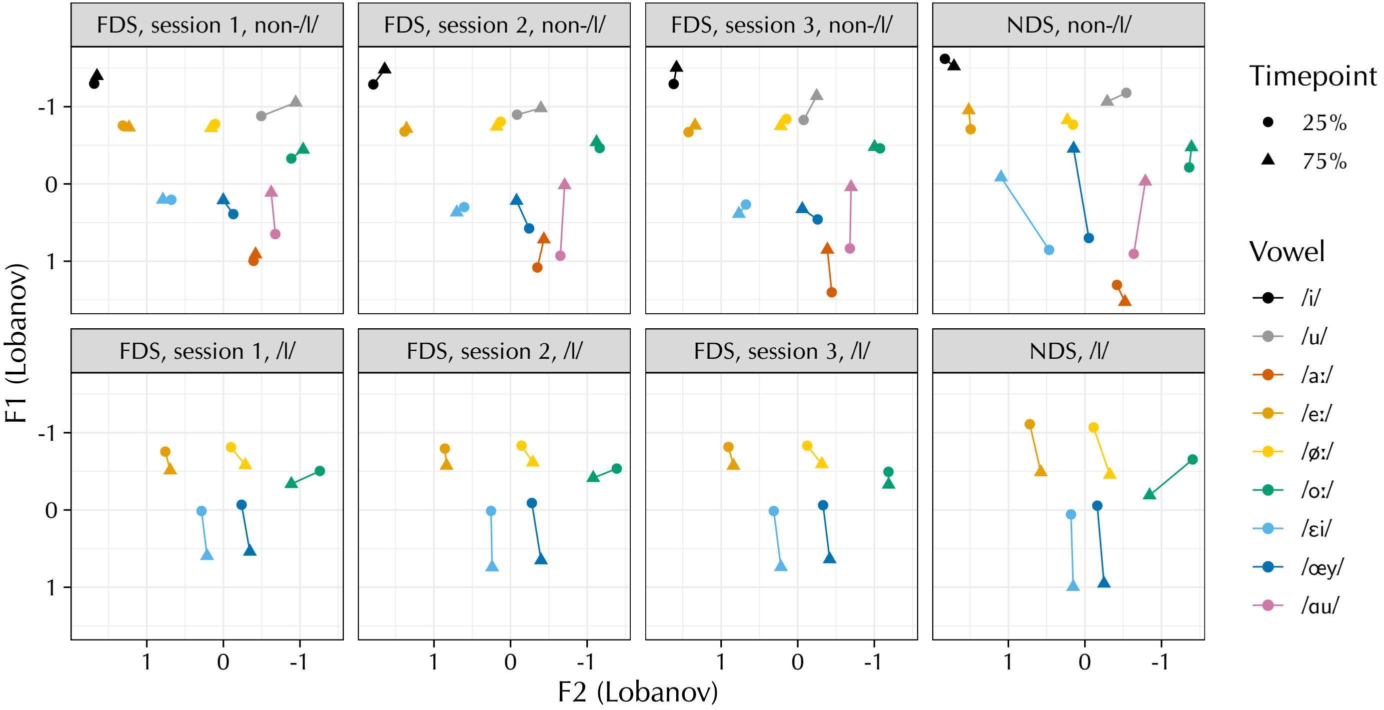

Figure 3.8: Vowel-space plot of the raw F1/F2 data (20 participants, 550 items). For reasons of space, the data for the NDS controls have been collapsed over the three sessions, as has the effect of prime realization. Observe how the NDS have upgliding realizations of the vowels under investigation in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

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

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

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

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

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

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

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

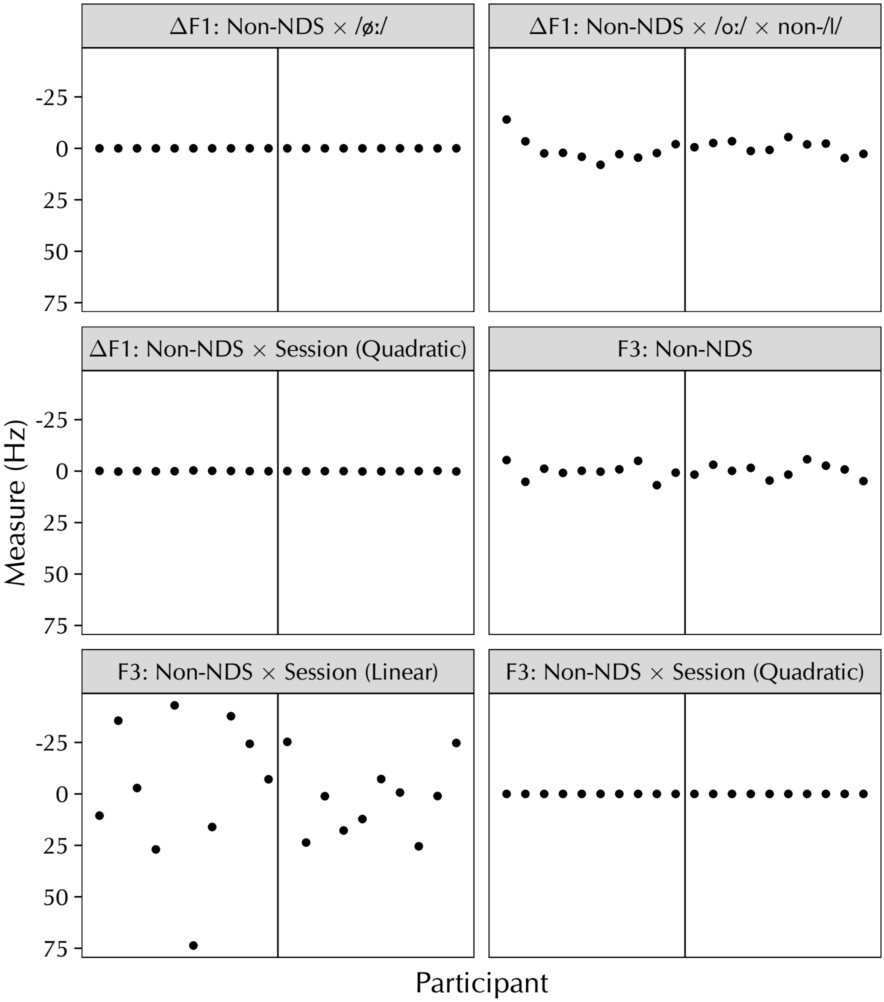

Figure 3.9: The estimated by-participants random effects (also known as "BLUPs", for "best linear unbiased predictors") for the random slopes involving the factor "Prime realization" in the supplementary models. Random-effect vectors that were estimated with zero variance (i.e. all by-participants random-effect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

3.3.3 Discussion

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

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

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

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

3.4 General discussion and conclusion

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

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

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

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

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

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

Acknowledgments

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

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