ALLOPHONIC IMITATION WITHIN AND ACROSS WORD POSITIONS

A thesis submitted in partial fulfilment of the requirements for the

Degree

of Doctor of Philosophy in Linguistics

in the University of Canterbury and Université d'Aix-Marseille

by Romain Fiasson

University of Canterbury and Université d'Aix-Marseille

2015

To Courtney Havill, a most special person.

Abstract

This dissertation investigates imitation in speech, which is the general tendency shown by a speaker to become more similar to another speaker in the way they speak. Many of us have experienced this while talking to someone who is speaking the same language but with a different accent. Conversing with such a person can affect some characteristics of our speech, so that we come to sound more like them. Imitation in speech has been very extensively studied, especially over recent years. To contribute to this line of research we provide an account of imitation in speech at the allophonic level, that is at the level of the possible phonetic realisations of a phoneme. We are interested in whether imitation of the sound of a given phoneme in a particular word position can influence the other possible realisations of that phoneme in the same word position. We are also interested in determining whether imitation of a speech sound in a particular word position for a given phoneme can affect the realisations of that phoneme in a different word position.

New Zealand English provides of wealth of allophonic variation across word positions for the phoneme /t/. Therefore it is an ideal language to investigate imitation and allophony. Before presenting our experimental designs and our results on imitation however, we verified and further extended the work that has been conducted on the dialect. We analysed large corpora of spoken NZE and found new allophones of /t/. We discuss a fricative realisation in particular.

The fricative realisation in NZE was further examined by means of a palatographic pilot experiment, as well as a perception experiment. Building on our findings from the perception experiment, we investigated imitation in speech towards an artificially created novel fricative allophone in medial position using acoustic and EPG data. For some speakers, the mere exposure to the novel allophone affected the realisation of other allophones in the same word position.

A series of acoustic experiments were then conducted to examine allophonic imitation across word positions. We found that repeated exposure to a given allophone can drive allophonic selection across word positions. We also found that positional transfer can occur, such that exposure to an acoustically manipulated allophone can affect the same allophones in a different word position. Interestingly it can also affect other allophones in a different word position.

Our results are discussed according to hybrid exemplar models and the Direct Realist view. We discuss which theoretical framework best accounts for the results we obtained.

Acknowledgements

I would like to thank Prof. Noël Nguyen and Prof. Jen Hay for their supervision, their time and their unconditional support during the preparation of this dissertation.

I am grateful to have had the opportunity to study at two marvellous institutions: the Laboratoire Parole et Langage, LPL (CNRS et Université d'Aix-Marseille) and the New Zealand Institute of Language Brain and Behaviour, NZILBB (University of Canterbury). I am grateful to Robert Espesser for his patience and his assistance with statistical models, Thierry Legou for creating the palatographic bench, Robert Fromont for his assistance with ONZE Miner and MySQL and to Morgana Mountfort-Davies for her assistance with some of the data collection. I would like to thank all staff and doctoral students at both labs for their assistance and friendship.

This dissertation was supported by a doctoral scholarship from the French Ministry of Research and Higher Education (Ministère de la Recherche et de l'Enseignement Supérieur) and by a doctoral scholarship from the University of Canterbury. The experiments reported in chapter 5 and 6 received financial support from NZILBB and the School of Languages, Cultures and Linguistics at the University of Canterbury.

I would like to thank Jim Scobbie and Allan Wrench who allowed me to use their EPG equipment to conduct experiments at Queen Margaret University. I would like to thank Paul Warren at Victoria University of Wellington and Denis Burnham as well as Cathi Best at the MARCS Institute for welcoming me and giving me a place to work after the Christchurch earthquake in February 2011. I would also like to thank Jonathan Harrington and Anne Cutler for helpful discussions and comments on this dissertation.

The Mobile Unit Recordings were collected by the Mobile Disc Recording Unit of the NZ Broadcasting Service. The Intermediate Corpus was collected by Rosemary Goodyear, Lesley Evans and members of the ONZE team (Origins of New Zealand English). The Canterbury Corpus was collected by members of the NZ English class of the Linguistics Department, University of Canterbury. The work done by members of the ONZE team in preparing the data, making transcripts and obtaining background information is also acknowledged. I would like to thank all the participants who volunteered for the experiments reported in this dissertation.

Many thanks to Cyril and Seb, Andrew McFarlane, Donald Derrick, Jacqui Nokes and Amanda Reid, it has been a pleasure to meet you and to work in your company. And last but not least, many thanks to Gwen, Jean, Matthieu, Rodolphe, Thibaud, Marina, Lynda, Sylvain, Fabrice, Emmanuelle, Callum, Eugene, Courtney and to my family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation investigates imitation in speech, which is the general tendency shown by a speaker to become more similar to another speaker in the way they speak. Many of us have experienced this while talking to someone who is speaking the same language but with a different accent. Conversing with such a person can affect some characteristics of our speech, so that we come to sound more like them. Imitation in speech has been very extensively studied, especially over recent years. To contribute to this line of research we provide an account of imitation in speech at the allophonic level, that is at the level of the possible phonetic realisations of a phoneme. We are interested in whether imitation of the sound of a given phoneme in a particular word position can influence the other possible realisations of that phoneme in the same word position. We are also interested in determining whether imitation of a speech sound in a particular word position for a given phoneme can affect the realisations of that phoneme in a different word position.

1.1 Imitation and phonetic convergence

1.1.1 Imitation

In this section, we will (i) define imitation, (ii) show that imitation is not restricted to the human species, (iii) review recent evidence in neuroscience establishing a neural basis for imitation at least in some species (iv) show that imitation is found at a very early age in humans (v) show that imitation can be mediated by external factors. Our goal here is to give an overview of imitation before addressing imitation in speech more specifically in the next section.

In order to give a definition of imitation we will follow the conclusions reached by Serkhane (2005). They offer an extensive review of literature by cross-referencing work conducted in ethology, phylogenetics and developmental psychology[1]. For them, imitation refers to one or more given surface behaviour(s) largely dependent on hidden psychological mechanisms. One the one hand, imitation corresponds to "a social learning process of another's behaviour [...] by observing that behaviour and reproducing all or parts of it *ie.* the result or the ways of reproducing that behaviour" (p.56, own translation). On the other hand, imitation involves a production-perception link, the coupling of sensory modes or *intermodal mapping*, and representations in memory (p.61). The production-perception link, intermodal mapping and representations in memory will be defined and discussed further below as our introduction progresses.

An interesting aspect of imitation that is worth mentioning is that it is not a human privilege: its use is not restricted to humans. Indeed imitation has been reported in other highly social animal species such as birds (*eg.* Alcock, 1969; Fisher and Hinde, 1949; Epstein, 1984), dolphins (Herman, 2002), monkeys (Kawai, 1965; Ferrari et al., 2006) or apes (Bard and Russell, 1999; Bard, 2007). For example, vocal imitation is well known in birds (Blackmore, 1998): bird songs are acquired progressively by imitating their parents or neighbours. Blackmore (1998) note that "[h]and-raised birds can learn songs from tape recorders and adopted birds sing songs more like their adopted, not biological, parents"[2]. It has been suggested that imitation stems from evolution and natural selection and thus serves as a form of adaptive behaviour (Davis, 1973; Thorpe, 1956). However some the evidence presented in non-human species remains controversial and it has been argued that the observed behaviour of animals could in fact reflect stimulus enhancement[3] or social facilitation[4] amongst other factors (see Serkhane, 2005; Pierce and Cheney, 2008, for discussions).

There seems to be a neural basis for imitation, at least in some species. A subset of neurons, termed 'mirror neurons' , have been found to be active in macaque monkeys when an individual performs an action as well as when this individual observes another performing the same action (Gallese et al., 1996; Rizzolatti and Craighero, 2004). So even by observing an action a monkey's motor cortex is activated. It is argued that the mirror-neuron system plays a fundamental role in action imitation. This has also been taken as evidence for a neural basis for the production-perception link in speech, which will be discussed with respect to gestural theories in the next section.

[1] Defining imitation has been subject of much debate (see Serkhane, 2005) and discussing the ins and outs is outside the scope of this dissertation. We take Serkhane's conclusions as an informed opinion on the subject.

[2] Interestingly parrots and mynahs do not only imitate fellow members of their species but can also imitate human speech.

[3] "Stimulus enhancement is a term that defines the probability that an individual approaches or contacts with an object with which another individual of the species is interacting (Spence, 1937). Because typically animals of the same species interact with identical objects in a rather similar way, "imitative" behaviour is determined by a generic increase of the probability of acting on that object, rather than by "true" imitation [...] of the observed action." (Rizzolatti et al., 2002).

[4] Social Facilitation refers to the effect that impinges on an individual's performance in the presence of others, such that an individual might perform better at simple or rehearsed tasks in front of others.

Imitation starts at a very young age in humans and carries on in adult life. In the widely publicised experiment conducted by Meltzoff et al. (1977) it was found that infants ranging from 12 to 21 days of age imitated manual and facial gestures, such as lip protrusion, mouth opening, tongue protrusion and sequential finger movements. Their results were subject to criticism partly due to the difficulties associated with experimental paradigms in infant imitation research and possible artefacts that could stem from these difficulties (Anisfeld, 1996; Bjorklund, 1987). Their experimental design has since then been refined and Meltzoff and Moore (1999) have argued towards a model of Active Intermodal Mapping. In their experiments infants observe an adult model performing a facial gesture, which is imitated. The matching between facial muscle movements and the visual stimuli is remarkable because it crosses different sensory modes (hence Active Intermodal Mapping): since infants cannot see their own face, they have to monitor their own gestures using proprioceptive feedback that is compared to the visual input and then matched. There is a growing body of evidence supporting the idea that neonates are indeed capable of true imitation (see Pierce and Cheney, 2008, for a discussion). Moreover, mirror neurons, which could be present at early stages of child development, might provide further support to Meltzoff and Moore's model of intermodal mapping. Very interestingly infants are also able to match speech across auditory and visual modes (Kuhl et al., 1982). When presented with faces articulating the vowels /a/ and /i/ with or without congruent visual facial information infants aged between 3 and 4 months were able to detect when the auditory and visual inputs matched. This is in accordance with the view that the visual input carries articulatory information, which adults make use of (*eg.* Schwartz et al., 2004). Khul and Meltzoff also reported that some of the infants' vocalisations bore resemblance to the vowels uttered by the model adult speaker (in terms of vowel quality, duration and intonation contour).

Non-conscious imitative behaviour is an integral part of human interaction. The well-known chameleon effect (Chartrand and Bargh, 1999) illustrates this point. Chartrand and Bargh conducted three experiments. The first one tested whether participants can automatically mimic others' gesture, even those of a stranger (a confederate, whom the participants never met before). The gestures produced by the confederate were face rubbing, foot shaking and smiling. Participants engaging in an interaction with the confederate did spontaneously increase face rubbing, foot shaking and smiling. This suggests that people are willing to imitate a set of different behaviours performed by a stranger. The second experiment reversed the situation: here the confederate were to imitate the movement of the participants. Results found that this mimicry behaviour of the confederate raised the enjoyment of the interaction and it increased the participants' liking of the confederate. This suggests that people like those who are like themselves and it gives support to the idea that mimicry facilitates social bonding. The third experiment investigated whether some participants would display higher levels of automatic mimicry or face rubbing and foot shaking. Participants who are higher on a scale of perspective-taking (the degree to which people naturally take others' perspectives) showed mimicry to a greater extent than participants who scored lower on the scale. The implications of these experiments are that by merely observing an action leads to performing the same action. Moreover, mimicry increases the liking of others and it is mediated by one's empathetic dispositions.

Although imitation seems to occur under various circumstances for a variety of behaviours, it also mediated by a series of factors. For example in a study on interaction, Johnston (2002) found that imitation of another's behaviour is moderated by visible social stigmas. Participants were asked to eat ice cream in the presence of a confederate who was either obese or not in the first experiment and participants were in the presence of a confederate with or without a visible facial birthmark in the second experiment. In both experiments confederates ate either a low amount or a high amount of ice cream. Participants ate the same amount of ice cream as the confederate in all conditions, except for when the confederate was obese and ate a large amount. The implications are that relevant social stigma, such as obesity in relation to food consumption, may prompt divergent behaviours. Thus mimicry is a function not only of social liking, but also of self-presentation.

More complex and less automatic behaviours build up on our very capacity to imitate, which can allow for complex observational learning (Pierce and Cheney, 2008; Gray, 2010). For example, when we are confronted to a new social context, we carefully watch others' behaviours in order to establish what should be an appropriate behaviour. Bandura (1965) conducted an experiment on imitation of aggressive behaviour, in which children watched a short film of adult model hitting, jumping on and verbally insulting a Bobo doll. Each sequence of action was accompanied by a unique verbal response (*eg.* the adult punched the doll in the face and said "Pow, right in the nose, boom boom", or the adult kicked the doll and said "Fly away"). Then the model was either punished, rewarded or left without consequences. After seeing the film, the children were left in a room with toys and the Bobo doll, and were instructed to play with the toys (they were not explicitly told to interact with the doll). Children who saw the model rewarded or left without consequences imitated to a greater extent the aggressive behaviour of the model than those who saw the model being punished. After the experiment an incentive was offered to the all children who could remember the actions and verbal responses performed by the adult model. The first implication of this study is that imitation is facilitated if it is perceived as acceptable. All children were also capable of remembering which verbal responses were associated with the aggressive actions, meaning that they retained the information in memory. In other words, children performed a particular type of imitation – delayed imitation – since they were shown the film first and then left in a room with the doll. Delayed imitation of a complex succession of events involves representations in memory in order to reproduce those events in sequence. Representations in memory have strong implications for the study of imitation in speech, which we will discuss in terms of exemplar theory in the next section and in chapters 5, 6 and 7.

In this section we have presented a definition of imitation, and illustrated some key aspects that will be referred to throughout this dissertation, namely the link between production and perception, intermodal mapping and representations in memory. We have shown that other species than humans have a capacity to imitate, that recent evidence suggest a biological substrate to imitation in some species (mirror neurons), that imitation is present at a very young age in humans and that imitation is an integral part of our behaviour but is not necessarily automatic and can be moderated by external factors such as social liking or social stigma. This introduction on imitation provides the necessary basis to understand and review imitation in speech.

1.1.2 Imitation in speech

Accommodation refers to the general tendency of changing our communicative behaviour in reference to our conversational partners. Amongst the communicative changes we can adopt, one of them has received particular attention, namely convergence. During convergence our communicative behaviour becomes more similar to that of others. Phonetic convergence has to be distinguished from other types of accommodation that may present similar characteristics. For example, in a noisy background it is well established that speech results in increased loudness, pitch, vowel duration or vowel formants (Lane and Tranel, 1971; Van Summers et al., 1988; Junqua, 1993). This is commonly known as the Lombard effect. Thus if two speakers adjust their voice while they are conversing in a noisy setting (*eg.* at a table in a noisy restaurant setting) characteristics of their speech might indeed become more similar but this is not the same as converging towards one another.

Over the recent years, a multitude of studies have focused on imitative behaviour in speech, and on phonetic convergence in particular. Phonetic convergence has been shown to occur in conversational settings (*eg.* Pardo, 2006; Pardo et al., 2011; Aubanel and Nguyen, 2010; Lewandowski, 2012) as well as in controlled laboratory settings (*eg.* Babel, 2010, 2011; Abrego-Collier et al., 2011; Nielsen, 2011; Kim, 2012)[5].

In interaction phonetic convergence has been accounted for by a variety of theories, each of them focusing on specific points. For example, Communication Accommodation Theory (Giles et al., 1991; Giles and Ogay, 2006), henceforth CAT, and the Interactive Alignment Model (Pickering et al., 2004; Pickering and Garrod, 2006), henceforth IAM, attempt to explain changes in communicative behaviour from a different perspective: CAT insisting on the social motivations of those changes and IAM insisting on the cognitive mechanisms underlying them. CAT and AIM presuppose an interaction between speakers, but fewer theories allow for the non-

[5]Note that computer simulations have been used to investigate imitation in speech (Wedel and Van Volkinburg, 2009) or have relied on programmed imitative capacities of agents to study linguistics phenomena such as the emergence of vowels systems (*eg.* Oudeyer, 2006).

presence of reference groups. Bell's theory of audience design (Bell, 1984, 1990) approaches accommodation in very similar way to CAT – namely that accommodation is highly socially motivated – but it allows for communicative changes when only one speaker is present.

However many studies on phonetic convergence have been conducted in very controlled laboratory settings without the need to address social motivations or interactive alignment. Those studies have employed the shadowing paradigm (or modified versions of the shadowing paradigm) which is inherently less ecological than natural interactions. In a shadowing task, a participant hears a pre-recorded model speaker pronouncing a word and he has to repeat it. Imitation occurs very often in this type of experiment[6], but due to the very asocial nature of the task it is difficult to account for imitation using CAT or Bell's theory of audience design. Moreover, speakers are not engaging in a task where they could align their linguistic representations with that of others' interactively. Thus the theory of dialogue proposed by AIM might also be not well suited to account for imitation effects in this type of task.

Studies employing the shadowing paradigm might be remote from speech that occurs in natural interactions but they have had important theoretical implications for (i) how speech sounds might be represented in the mental lexicon and (ii) the study of the link between production and perception. For example, in his highly influential paper, Goldinger (1998) gave further support towards an episodic representation of words in memory using a shadowing task[7], in line with the ongoing development and refinement of exemplar-based models of speech perception and production over the past 20 years. Results obtained using the shadowing paradigm have also fuelled the controversial debate as to whether the link between speech perception is mediated by phonology or not (*eg.* direct realism (Fowler and Galantucci, 2005; Best, 1995) versus the phonological account, see below).

1.2 Theoretical framework

In this section, we discuss the theoretical frameworks that have been used to account for imitation in speech using the shadowing paradigm. We will first give a sketch of the traditional abstractionist perspective. We will compare this approach to the episodic view. Describing both views will enable us to outline what a hybrid approach entails. We will conclude this section by describing gestural theories of speech perception and production, and focus on direct realism in particular. In this dissertation we will attempt to show which theoretical framework is more appropriate to account for our results.

[6]Mitterer and Ernestus (2008) argue that imitation could actually be the default in a shadowing task. See the discussion in section 1.2.2.

[7]See section 1.2.1 below, for an explanation of the shadowing paradigm.

1.2.1 Towards hybrid exemplar models

The traditional abstractionist approach

According to traditional abstractionist accounts speech sounds are phonologically represented as invariant, abstract and discrete categories such as phonemes and features (Chomsky and Halle, 1968). In spite of the high variability in the speech signal, however, listeners are able to extract meaning from the speech signal quite effortlessly. It is thus assumed, from an abstractionist perspective, that listeners must filter out variability in order to map phonetic variation onto abstract phonological categories. It also assumed that listeners do so according to a *normalisation process*, which removes speaker-specific characteristics, also called *indexical properties* (Abercrombie, 1967), such as vocal tract length differences or sociophonetic variation. Mullenix et al. (1989) found that when words are presented in rapid succession, recognition of the words spoken by multiple talkers is slower and less accurate than recognition of the words spoken by the same speaker. They concluded that increased variability in the speech signal increases processing costs because it places a high demand on the normalisation process during word recognition.

The episodic view

A large body of literature has shown that indexical properties are in fact perceptually relevant and are stored in memory (see Nguyen et al., 2009 for a review). In adverse listening conditions for example, word recognition improves with listeners' increasing familiarity with voices (Nygaard et al., 1994; Nygaard and Pisoni, 1998). Also adjusting to different speakers decreases word identification accuracy (Mullenix et al., 1989) and these processing costs seem to reflect the encoding of information about individual voices in memory (Palmeri et al., 1993)[8]. Using the shadowing paradigm Goldinger (1998) showed that the effect for episodic traces occurred in speech production as well. He asked participants to say words out loud after hearing model speakers producing these words. The participants' pronunciation was rated more similar to the model speakers compared to their baseline productions. Imitation was largely influenced by factors such as word frequency and the number of repetitions from the model speakers participants were exposed to. Less frequent words lead to increased imitation and so did an increase in the number of exposure to the words. Goldinger and Azuma (2004) replicated these findings using passive exposure to model speakers and showed that imitation can be found after two weeks for low frequency words that have been presented most. Thus information

[8]Note that speaker variation affects processing when the task performed by the listener is hard, for example in adverse listening conditions (Mullenix et al., 1989) or in lexical decision tasks where non-words resemble words (McLennan et al., 2005). Speaker variation does not affect easy processing, *eg.* in simple lexical decision tasks (Luce and Lyons, 1998) or where non-words are clearly distinguishable from words (McLennan et al., 2005). See Cutler (2012, p.388) for a discussion.

about the model speakers' pronunciation must be stored in long term memory for participants to be able to imitate with such a delay.

Overview of exemplar models based on a purely episodic view

This line of research gave support to the episodic view, which assumes that every experience that we have would leave a unique episodic trace in our memory. The idea dates back to the early 20^{th} century and was originally developed as a theory of memory by Semon. More recently the theory has been adapted for speech perception by Johnson (1997) based on Nosofsky's model (1986) and for word recognition by Goldinger (1998) using the MINERVA 2 model (Hintzman, 1986). The episodic view in its application to speech perception claims that each word that is heard is stored in memory – for a certain amount of time – with detailed acoustic and articulatory characteristics, information on the identity of the speakers and on the communication situation, amongst may other dimensions.

The general approach of a purely episodic model could be described as follows: whenever a speech stimulus is encountered, its trace is stored in memory. Previously stored traces are assigned an activation level and a measure of similarity is established with the most recent one. Greater similarity between the most recent trace and previous traces will increase their activation level. Recognition of the stimulus involves averaging over all activations of previously stored exemplars[9].

This is in sharp contrast with the traditional abstractionist view for which variability in the speech signal must be filtered out in order to map variation onto abstract categories. For example Johnson's exemplar-based model of speech perception (1997) retains speaker-specific variation and thus does not need a speaker normalisation procedure. More generally, speaker normalisation in exemplar-based models is not necessary and this is certainly a strength of the approach.

Overview of hybrid exemplar models

The relevance fine phonetic detail may have been overlooked by some abstractionist models[10] but there is no doubt that some level of abstraction is needed during word recognition. Indeed abstract categories must exist to account for categorical perception (CP), which refers to our tendency to classify an ambiguous sound as belonging to either one perceptual category or another, but never to both at the same time. For example Lisker and Abramson (1970)

[9]Some models extended this approach to speech production (Pierrehumbert, 2001, 2003). Production mirrors perception in that it involves averaging over all activations of previously stored exemplars and selecting the average as a target for speech production.

[10]Nguyen et al. (2009) remark that the relevance of fine phonetic detail is not incompatible with the abstractionist approach. Fine phonetic detail plays a role in Stevens' model of lexical access (Stevens, 2002) or in the TRACE model (McClelland and Elman, 1986). Both models are abstractionists.

showed that manipulating the duration of VOT along a continuum causes an abrupt perceptual shift from voiceless plosives to voiced plosives at one specific location in the continuum (*ie.* it causes listeners to perceive either /p/ or /b/ but not both categories at the same time). The perceptual training literature – in particular findings about the generalisation of perceptual learning – also provides strong evidence that our representations have an abstract component, which will be discussed in detail in the next section.

Some exemplar-based models argue that the storage of exemplars alone cannot explain phonetic variation and that an abstract, phonological component must be taken into account (*eg.* Coleman, 2002; Pierrehumbert, 2006). Pierrehumbert (2006) argues for a *phonological principle* that allows us to generalise phonetic patterns across words that share similar phonological structures. These models are known as hybrid models because they contain both an episodic memory component and an abstract phonological level. A hybrid approach allows to explain the results obtained by Nielsen (2011) for example. She found that speakers extended the VOT of initial /k/ words after being exposed to lengthened VOTs of initial /p/ words. She concluded that an episodic memory component is needed to store the extended /p/ VOTs but that an abstract level is needed so that the sub-phonemic variation can generalise to words with initial /k/ because these sounds are cognitively linked: they belong to the aspirated plosive class.

In summary, exemplar-based models are a departure from traditional abstractionist models. However less extreme models – the hybrid models – do take into account some level of phonological abstraction. In the next section we discuss gestural theories of speech perception and production, and direct realism in particular. Direct realism has been put forward to account for results in imitation in speech, but contrary to exemplar-based models the theory does not involve *stored mental representations of things past* (Wilcox and Katz, 1981).

1.2.2 Gestural theories of speech perception and production

Overview of the theories

The motor theory (Liberman and Mattingly, 1985) and direct realism in its application to speech[11] (*eg.* Fowler, 1986; Fowler and Galantucci, 2005; Best, 1995) focus on the link between speech production and perception. The theories share many features in common, the most important one being that speakers/listeners are able to use and perceive gestures in speech. Here gestures have to be taken in the view of Articulatory Phonology, henceforth AP (Browman and Goldstein, 1986; Browman et al., 1992, and for a more recent overview of AP, see Hall,

[11] The theory of direct realism is a general theory of perception originally developed by Gibson (1979) and its extension to speech by Fowler and colleagues is more recent. For the sake of brevity and since our argumentation is only concerned with speech we will refer to direct realism in its application to speech simply as direct realism.

2010). In AP's view, gestures are actions that create and release constrictions in the vocal tract using a combination of two or more articulators. Most importantly these actions are linguistically relevant and serve as phonological primitives.

The motor theory and direct realism differ in their respective predictions about how listeners perceive these gestures. According to the motor theory, listeners first perceive the acoustic signal and then must reconstruct gestural implementations using an analysis by synthesis. In a direct realist approach, listeners do not need to infer the gestures from the acoustics, rather they directly perceive them in the acoustic signal (or, to be more accurate, they perceive gestures from their sensory experiences, *ie.* auditory, visual or somatosensory experiences). Thus, for the motor theory, gestures are present in the mind while they are already present in the speech events from a direct realist perspective. Also, the motor theory assumes that the motor system is recruited during speech perception, while direct realism remains more agnostic about the subject. From a motor theory perspective, the production perception link is hard-wired in the human brain, which is supported by the possible presence of mirror neurons in humans or early imitation in young infants, as we discussed in the previous section. Further empirical evidence from neuroscience has been put forward by to support that claim: auditory cortical areas which are involved in hearing sounds are activated during silent lip reading (Calvert et al., 1997; MacSweeney et al., 2000), secondary auditory cortical areas are also activated during speech production when speakers are deprived of auditory feedback (Paus et al., 2006), or there is an increase in motor-evoked potentials measured at the tongue when listeners are exposed to utterances that contain lingual consonants compared to utterances that do not (Fadiga et al., 2002).

The theories differ in many aspects but it is outside the scope of this dissertation to present them in detail. What is important here, is that for both gestural theories (i) listeners are able to perceive gestures (whether they achieve that in a distal or in a proximal way depends on the theory), and (ii) there is a tight link between perception and production in speech (although the need to invoke motor control in perception or not also depends on the theory). We will now review findings obtained in shadowing tasks that support the view that speakers/listeners perceive gestures and that there is a tight coupling between perception and action.

The account of direct realism for imitation in shadowing tasks

Some studies have employed the shadowing paradigm in simple and multiple responses tasks. In a simple response task speakers are asked to produce a given syllable regardless of the syllable presented to them. There are two possible scenarios: either the response matches the stimulus (*eg.* the speaker hears /pa/ and has been told in advance to produce /pa/ after hearing the syllable) or the response does not match the stimulus (*eg.* the speaker hears /ta/ and has been told in advance to produce /pa/ after hearing the syllable). In a multiple response task, the speaker is asked to shadow a syllable amongst a limited set of stimuli (*eg.* if the stimulus set contains /pa/, /ta/ and /ka/, the speaker must produce the stimulus that they have just heard). Gestural accounts have predictions in terms of stimulus-response latencies: (i) latencies should be the same in multiple response tasks as they are in a simple response tasks with matching stimulus-response (ii) latencies should be longer in a simple response task when there is a mismatch between stimulus and response compared to when there is a match. Porter and Castellanos (1980) and Fowler et al. (2003) demonstrated that latencies are very short in a multiple response task as well as in simple response task. In the simple response tasks, however, latencies are even shorter when the syllable they are asked to produce matches the syllable that they heard. Fowler et al. (2003) also found that the stops at the onset of the syllables had more aspiration if the model speaker had a longer VOT, thus indicating imitation of phonetic detail. Their interpretation was that a phonological link, rather than a direct link between production and perception, would not be able to explain imitation of sub-phonemic features (in this case VOT). According to them, if speakers first used the acoustic signal and then performed a classification into phonological categories (*ie.* if perception and production were phonologically mediated), then speakers would not copy the model speaker's VOT. Rather they would stick to their own VOT once they aim at producing the speech sound previously identified as belonging to their phonological category.

Honorof et al. (2011) obtained similar findings using a shadowing task. Instead of measuring latencies, they measured tongue and lips movement using electromagnetometry. They found that English speakers were able to reproduce some of the lingual gestures associated with dark [ɫ] when it is presented at the onset of a syllable (the clear [l] variant is expected in this position, while the dark [ɫ] variant is found in coda position in English, see chapter 4). Acoustically, producing a clear [l] or a dark [ɫ] affects the second formant, and so does lip rounding. However speakers did not use lip rounding to produce the dark [ɫ] variant in initial position. Honorof et al. (2011) argued against an acoustic account which predicts the possible use of lip rounding in order to achieve the acoustic target. They concluded that only a gestural approach can account for their finding.

Challenges to the direct realist account

Many other studies have challenged gestural accounts, in particular using the shadowing paradigm in simple response and multiple response tasks. For example, Mitterer and Ernestus (2008) asked Dutch speakers to shadow alveolar or uvular trills and could not replicate some of the results by Fowler et al. (2003). They found that response latencies were not affected when there is a mismatch between stimulus and response in a simple response task, which is not compatible with a gestural account. They also found that speakers did not imitate the trills they were exposed to and produced their habitual articulation instead. Mitterer and Müsseler (2013) investigated accommodation in German using a shadowing task, with two possible realisations of the same phoneme that depend on the dialect spoken by the participants. They replicated previous findings by Mitterer and Ernestus (2008) such that no latency costs were associated with a mismatch between stimulus and response. However they did find imitation, particularly when speakers shadowed non words. They raised that possibility that "the strong tendency to imitate might simply be the default in a shadowing task" (p.15). They added that salience of the stimuli might be the actual cause of an increase in imitation. Finally Mitterer and Ernestus (2008) and Mitterer and Müsseler (2013) argued for a loose link between speech production and perception, and that both systems are phonologically mediated. Similar findings were obtained by Nielsen (2011) who found that long VOTs in English initial voiceless stops were imitated whereas short VOTs were not imitated. It may be that imitation is selective in order to preserve phonological contrast between voiced and voiceless stops.

1.3 Perceptual learning

In a study mentioned above Lisker and Abramson (1970) showed that manipulating the duration of VOT on a continuum causes a perceptual shift from a voiceless plosive category to a voiced plosive category. Other perceptual shifts from one category to another can occur because of compensation for coarticulation (Mann, 1980; Mann and Repp, 1980, 1981), phonotactic knowledge (Massaro and Cohen, 1983; Hallé et al., 1998), or speaking rate (Miller and Liberman, 1979) for example. These short-term and short-range perceptual changes are usually referred to as perceptual adjustment.

Modulation of a phonetic category through experimental training is also possible and is referred to as perceptual learning. For example Norris et al. (2003) presented listeners with words containing an ambiguous sound between [f] and [s]. One group heard the ambiguous sound replacing words ending in [f] but not [s] and one group heard the ambiguous sound replacing words ending in [s] but not [f]. Listeners were then asked to rate whether sounds presented on a [f-s] continuum sounded more like [f] or [s]. Listeners presented with the ambiguous sound in [f] words were more likely to categorise sounds as [f] on the continuum. Conversely listeners presented with the ambiguous sound in [s] words were more likely to categorise sounds as [s] on the continuum. Thus lexically induced perceptual learning has the effect of expanding listeners' phonetic categories so that they include more ambiguous sounding tokens than before.

Perceptual learning has interesting properties. (i) The effect is very robust: McQueen et al. (2012) replicated the findings mentioned above with children aged 6 and age 12. It is also robust as it can be found in a variety of experimental paradigms (McQueen et al., 2006b; Eisner and McQueen, 2006; Clarke-Davidson et al., 2008). (ii) It is durable: it can last at least up to 12 hours after exposure to a deviant stimulus (Eisner and McQueen, 2006). (iii)

It is *thorough* (McQueen et al., 2012) because listeners treat the ambiguous sound as an unambiguous sound after exposure (Sjerps et al., 2010). (iv) Speaker specificity effects arise depending on the phonetic class that the ambiguous sounds belong to, such that perceptual learning generalises across speakers for stops (Kraljic and Samuel, 2006, 2007) but not for fricatives (Eisner and McQueen, 2006; Kraljic and Samuel, 2007). Perceptual learning thus proves to be a robust, long lasting and thorough effect, which also has the ability to generalise. We will now discuss aspects of its generalisation in the next three paragraphs.

Generalisation of perceptual learning can extend to novel words (McQueen et al., 2006a; Maye et al., 2008; Sjerps et al., 2010; Mitterer et al., 2011). For example McQueen et al. (2006a) trained listeners to interpret an ambiguous [f-s] sound as [f] or [s]. In a second phase they tested whether training would influence decisions about minimal pairs that could end in /f/ or in /s/. Training with [f] facilitated the recognition of minimal pairs as ending in [f]. Conversely training with [s] words facilitated the recognition of minimal pairs as ending in [s]. Generalisation to novel words occurred because the items presented in the second phase were not heard in the training phase. The results reported by Nielsen (2011), where VOT lengthens in novel words after training, further extend these findings to speech production.

Kraljic and Samuel (2006) showed that generalisation can occur at the featural level. In their study, participants trained on a [t-d] continuum shifted their perceptual boundaries on a [p-b] continuum. These results also echo the finding in speech production that speakers exposed to longer initial [p] VOTs can lengthen the VOT of initial [k] sounds (Nielsen, 2011).

Perceptual learning can also generalise from one word position to another. Perceptual learning effects have been investigated by training listeners in medial position (Kraljic and Samuel, 2005), in final position (Norris et al., 2003; McQueen et al., 2006b, 2012) or in different positions in the same study (Eisner and McQueen, 2006). Jesse and McQueen (2011) further demonstrated that listeners trained on a phoneme in final position can generalise to the same phoneme in initial position.

1.4 Allophony

This dissertation investigates potential positional transfer effects from one allophone of a single phoneme to another allophone of the same phoneme. In this section we provide a definition of allophony as it is commonly presented in textbooks in phonetics and phonology. We also raise a question about how allophones might be cognitively linked to the phoneme.

Baudoin de Courtenay (1895) was probably amongst the first to offer a definition of the phoneme such that it is an abstract category consisting of a multitude of speech sounds that listeners/speakers feel are all instances of the same category. Speech sounds that can be related to the same abstract category are called allophones. In English for example, the sounds

[t], [tʰ], or [ɾ] can appear word-medially in some varieties, and speakers of these varieties have the intuition that they are instances of the same abstract category /t/. In other words, while these sounds are produced differently – they have different acoustic and articulatory properties – speakers feel that they are all "t-sounds".

Another example of the grouping of concrete realisations of speech sounds into an underlying abstract linguistic category is given by Davenport and Hannahs (1998, p.97) who refer to water as an analogy. Water is a molecule composed of one oxygen atom and two hydrogen atoms (H_2O). The molecule is the underlying representation of water. The molecule can change in terms of its physical aspect according to the context it appears. Below 0°C water is ice, between 0°C and 100°C it appears in its liquid form and above 100°C it becomes vapour. Much like the chemical compound and the forms it can take in specific environments (here at different temperatures), a phoneme can have a multitude of realisations – allophones – in specific environments.

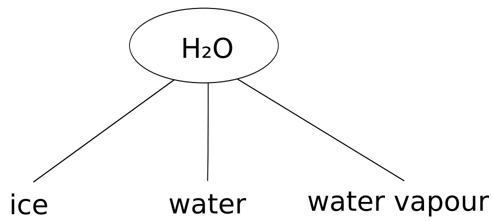

Figure 1.1 – The water analogy (adapted from Davenport and Hannahs, 1998, p.97).

Textbooks in phonetics and phonology teach students to identify allophones using tests such as the *commutation test*[12] for example. If two speech sounds appear to be in *contrastive distribution*[13] using this test then they are said to be allophones of different phonemes. If they appear to be in *complementary distribution*[14] or in *free variation*[15] however they are said to be allophones of a single phoneme. Definitions along with examples of the terms

[12]The commutation test is "[a] procedure for demonstrating the existence of a contrast between two segments. In a given phonological form, a segment of interest (or zero) is replaced by another segment (or zero); if the result is a different phonological form, the segments in question are shown to commute [...] (*ie.* to contrast). For example, English has a form *rap* [ræp]; replacing [r] by [l] yields *lap* [læp], which has a different meaning: hence [r] and [l] commute (contrast) in English. Similarly, *clap* [klæp] is different again showing that [k] commutes with zero. [...]" (Trask, 1996, p.80).

[13]Contrastive distribution is "[t]he relation between two (or more) phones in which either can occur in the same position in identical surroundings to produce different meanings. [...]" (Trask, 1996, p.93).

[14]Complementary distribution is "[t]he relation which holds in a given speech variety between two phones which never occur in the same environment. For example, in the English of England, clear *l* occurs before a vowel while dark *l* never occurs before a vowel; in many varieties of English, [h] only occurs before a stressed vowel, while [ŋ] never occurs before a stressed vowel. In each of these cases, the phones in question are therefore in complementary distribution. Providing additional criteria are satisfied, most notably phonetic similarity, such segments can be assigned to a single phoneme. Clear *l* and dark *l* are accordingly assigned to a single phoneme /l/, while [h] and [ŋ] are not. [...]" (Trask, 1996, p.81).

[15]Free variation is "[...] the phenomenon in which any of two or more phones can appear in the same position without any effect upon meaning. For example, a single speaker of English may at various times pronounce the word *eat* with two or more aspirated [tʰ], unaspirated [t], glottalised [ʔt] or unreleased [t̚]. [...]" (Trask, 1996, p.150).

referred to previously can be found in the footnotes. However it is outside of the scope of the thesis to discuss commutation tests and their limitations in order to identify allophones. This dissertation is concerned with the allophones of the phoneme /t/ in New Zealand English and they have been well identified and extensively studied by scholars on the dialect. An entire section (section 1.6) is dedicated to their inventory, their contexts of occurrence as well as their socio-phonetic patterning.

It is also important to distinguish two types of allophony that may occur within and across word positions. Within word position allophony pertains to phonetic realisations that may be distinct at the same place in a word. For instance, one speaker might pronounce the word *totter* as [tʰɒtə] while another speaker might use a flap instead of a stop in medial position. This is particularly the case for allophones which are in so-called free variation (as mentioned above), where one single speaker may produce [tʰɒtə] or [tʰɒɾə] at different times. Across word positions allophony pertains to phonetic realisations that may be distinct at different places in a word. In [tʰɒɾə] the initial aspirated stop might be phonetically different from a medial tap but they still belong to one single abstract phonemic category.

As far as we know, there has been little research conducted on allophones and their cognitive relatedness to the phoneme. In other words we don't really know whether affecting the phonetic properties of a single allophone can have an impact on other allophones of the same phoneme (*ie.* if they are strongly cognitively linked), or whether the other allophones will remain unaffected (*ie.* if they are loosely cognitively linked). This dissertation will aim at providing insights to the following question: to what degree are different phonetic categories (allophones) cognitively linked to one single abstract category (the phoneme)?

1.5 Research questions addressed by the dissertation

Nielsen (2011) presented speakers with shortened VOTs of initial /p/ and it did not affect their subsequent productions. She argued that producing shorter VOTs was inhibited in order to preserve the contrast between /p/ and /b/. Interestingly training speakers with lengthened VOTs of initial /p/ made them lengthen the VOT of (i) initial /p/ words presented during the exposure phase, (ii) novel initial /p/ words absent from the exposure phase, and (iii) novel initial /k/ words absent from the exposure phase. Nielsen concluded that VOT is a sub-phonemic feature that can be imitated and can generalise to novel words beginning with the same consonant or a different aspirated plosive. A possible interpretation of her results could be as follows: /p/ and /k/ differ by their place of articulation but they belong to one abstract class, namely the class of aspirated plosives. In other words, one might argue that they are cognitively linked at a high order level which is that of the aspirated plosive class. So affecting a member of the class is likely to affect another member.

This is represented in figure 1.2. Gaussian-like curves represent the phonetic distribution of VOT values, with shorter values on the left and higher values on the right. Presenting speakers with shorter initial /p/ VOTs (dotted lines) is not imitated by speakers and overlaps with their typical distribution in production, or *"old" experiences* (black lines). On the other hand presenting speakers with longer VOTs (dashed lines) will shift their distribution to the right. This effect will further spread to the realisation of initial /k/ stops and it happens because /p/ and /k/ might be cognitively linked at an abstract level.

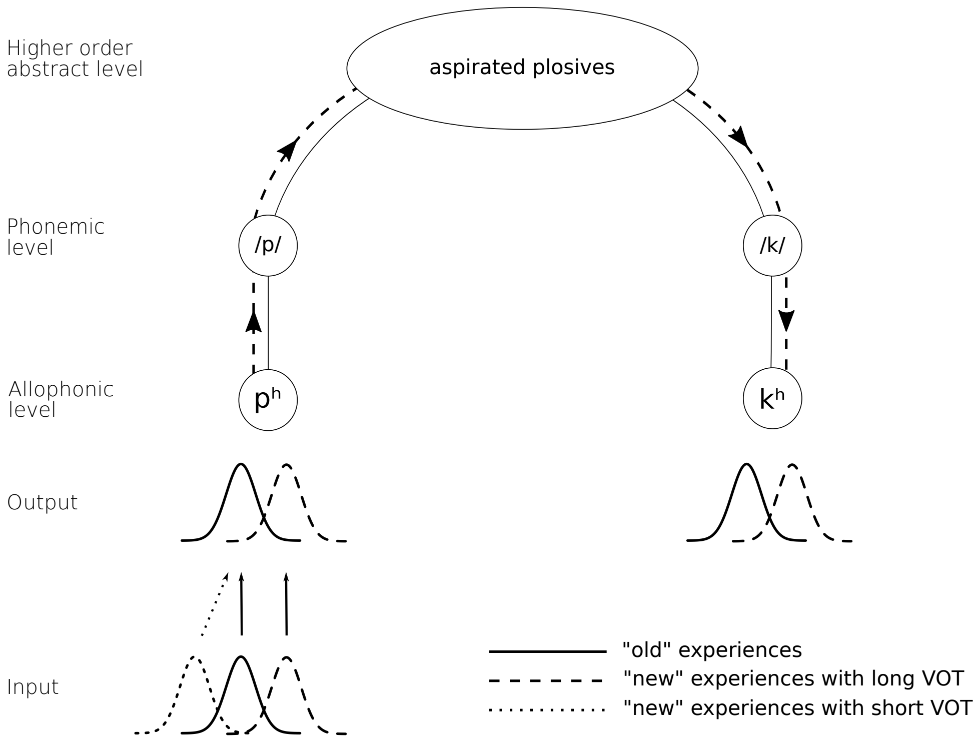

Figure 1.2 – Schematic representation of Nielsen's results.

Nielsen studied imitation effects for different phonemes that are in the same position. However one phoneme can often be realised very differently in a given positional context in English (*eg.* in medial position /t/ can be realised as a stop, a tap or a glottal stop). A phoneme can also be realised very differently in different positional contexts in English (*eg.* /t/ can be realised as a stop in initial position, as a tap in medial position, or it can be unreleased in final position). Thus we will ask the following questions: **(i) how is allophonic variation represented at an abstract level?**, and **(ii) to what degree are different phonetic categories (allophones) cognitively linked to one single abstract category (the phoneme)?**

In this dissertation we will try to shed light on imitation at the allophonic level, which has not been studied as far as we know. To do so, we will investigate imitation effects within and across word positions.

1.5.1 Allophonic imitation within a position in the word

Imitation of fine phonetic detail within the same word position has been well studied (*eg.* imitation of VOT in initial position Fowler et al., 2003; Nielsen, 2011). Yet, one might ask: **to what degree do manipulations of one phoneme extend across variants in the same position?** For example, would lengthening an allophone in a given word position cause other allophones to lengthen in the same position as well?

Goto (1971) reported that native speakers of Japanese who have learned English as adults have difficulty perceiving the acoustic differences between English /l/ and /ɹ/, even if they speak English quite well. Models such as the Perceptual Assimilation Model (Best, 1995) or the Speech Learning Model (Flege, 1995) predict the difficulty of second language learners to discriminate between sounds that do not contrast phonemically in their own language. However less is known about first language speakers who might try to discriminate between two non-contrastive sounds within their own language. But we do know that people vary widely from each other (cite), such that some perform better than others.

Honorof et al. (2011) showed that speakers are also able to produce articulations that are not within their typical realm of production. Shadowing a model speaker who produced a dark [ɫ] – which is usually found in syllable coda position – induced speakers to partially reproduce gestures associated with dark [ɫ] in syllable initial position. This raises the following question: if a speaker can learn to produce an unusual variant, would this ability influence the production of other allophones of the underlying phoneme? And if so, will the pre-existing allophones be produced more like the unusual variant?

Figure 1.3 illustrates this potential effect: if we assume that a novel allophone X can be introduced into the phonetic repertoire of a speaker, then it might cause the distribution of other allophones to shift along some phonetic dimension. It could be the case, however that speakers might indeed learn to produce a novel variant by imitating of novel speech sound, but it might not affect other variants in the same position at all. This is illustrated in figure 1.4.

Also Clark (2013) showed that speakers' production of a variant lead them to produce the variant more in subsequent productions. So would a speaker exposed to a variant in one position be more likely to produce this variant over the others in the same word position? In other words, **would imitation of a variant drive allophonic selection?**

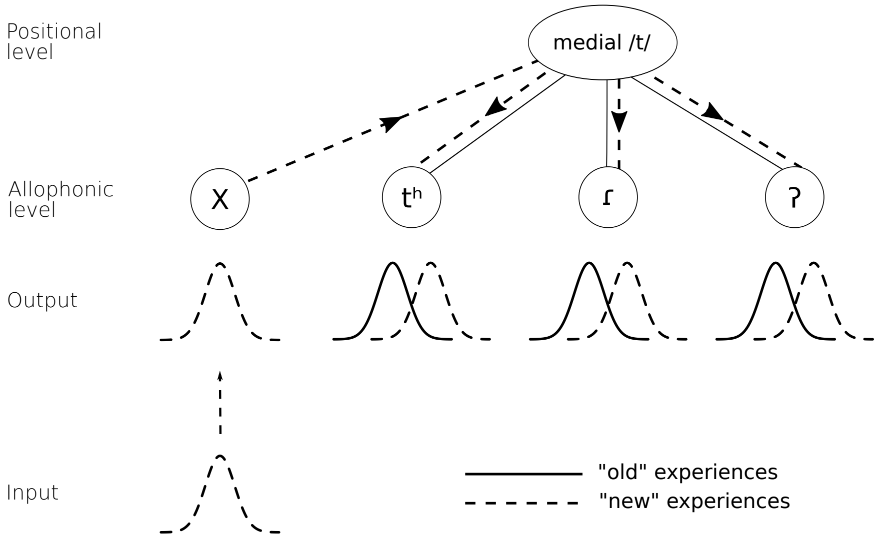

Figure 1.3 – Imitation within position. Here the introduction of a novel allophone X by imitation affects the distribution of other allophones in the same position.

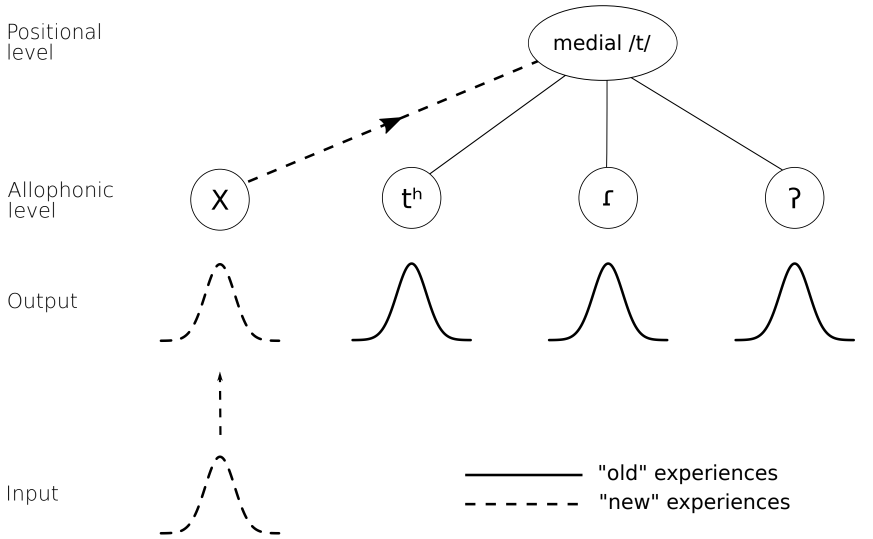

Figure 1.4 – Imitation within position. This time the introduction of a novel allophone X by imitation does not affect the distribution of other allophones in the same position.

1.5.2 Allophonic imitation across word positions

The recent results obtained by Jesse and McQueen (2011) such that training on a phoneme in one position can affect the realisation of the same phoneme in a different position further raise the following questions: **is imitation in speech position sensitive? To what degree do manipulations of one phoneme extend across positions?**

Following Nielsen's results represented in figure 1.2, we could envisage two scenarios: one in which exposure to a deviant variant in one position, let's say a lengthened word initial [t^h], would be imitated and cause the distribution of other allophones to shift in a different word position because they are tightly cognitively linked (figure 1.5). This would replicate Nielsen's finding that imitation of a sub-phonemic feature can spread from one phoneme to another. However this has not been tested in terms of allophonic variation. So we could also envisage the case that lengthening word initial [t^h] can be imitated but might not affect the realisation of other allophones in a different position at all (figure 1.6). This would happen if allophones are not as tightly cognitively linked as previously thought.

New Zealand English provides a wealth of allophonic variation to investigate our research questions. In the next section of our introductory chapter we provide the reader with a literature review of the allophones of the phoneme /t/ that can be found in the dialect.

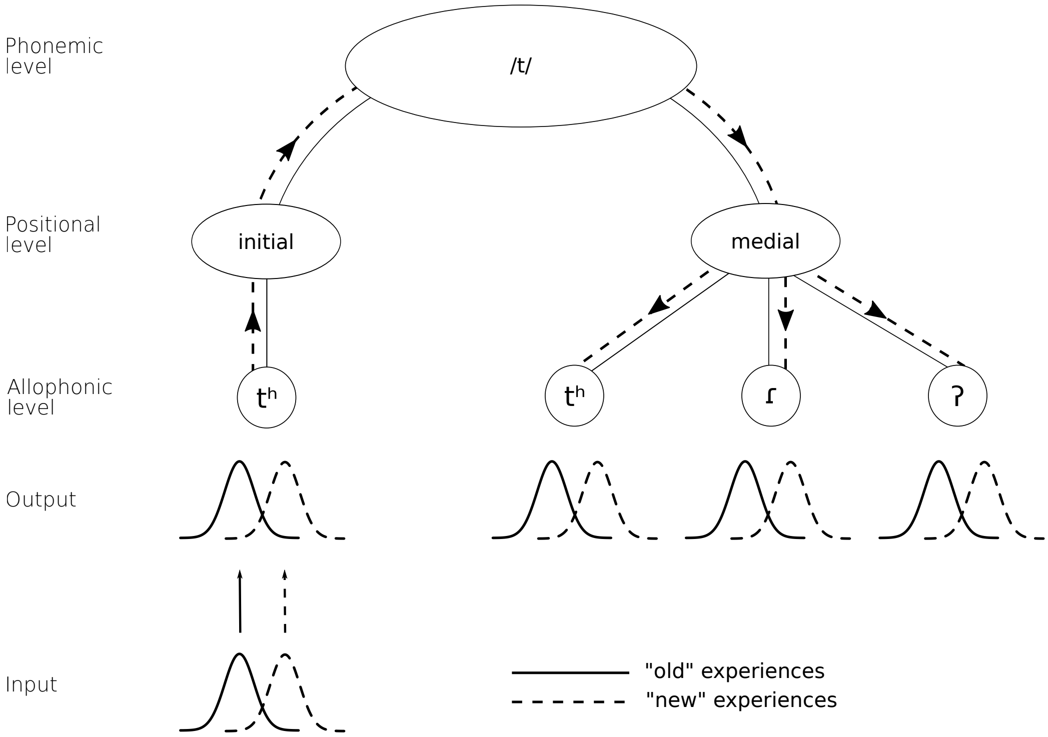

Figure 1.5 – Imitation across word positions. Here the imitation of a phonetic variant in initial position affects the distribution of other allophones in a different position because they are tightly cognitively linked.

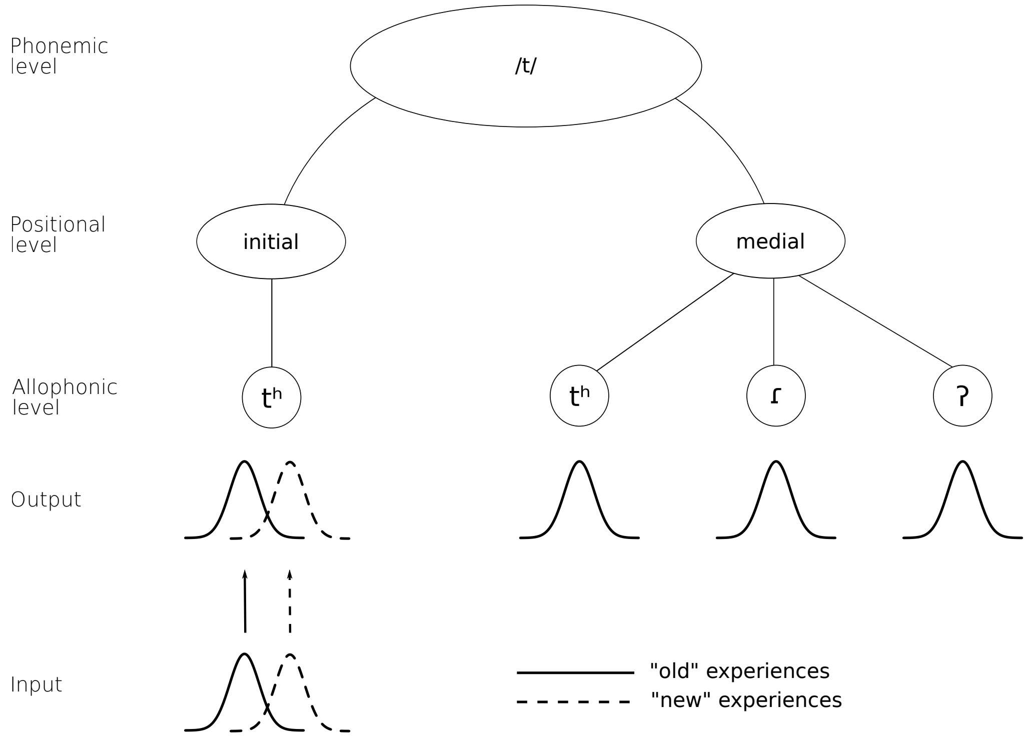

Figure 1.6 – Imitation across word positions. Here the imitation of a phonetic variant in initial position does not affect the distribution of other allophones in a different position because they are more loosely cognitively linked.

1.6 Description of the attested phonetic variants of /t/ in New Zealand English

The pronunciation of NZE has evolved dramatically since English was first spoken in New Zealand and this dialect has proved to be a nest for the study of sound change. For example, a very striking feature of NZE is the front short vowel shift, which resulted in high realisations of the TRAP and DRESS vowel, and a central realization of the KIT vowel. Gordon et al. (2004) traced the start of this push-shift back to very early NZE and showed that speakers born in the 19^{th} century already raised TRAP and DRESS. The KIT vowel started to centralise later in the 20^{th} century (Langstrof, 2006). Other well-known features of NZE are the merging of the NEAR and SQUARE vowels amongst young speakers (Hay et al., 2006) or the merging of ELLEN and ALLEN (Thomas and Hay, 2005).

Consonants have also been studied and the phoneme /t/ has received particular attention from scholars in NZE. In fact /t/ shows a complex patterning of allophonic variation that is both phonologically and socially conditioned. For example studies of word-final /t/ glottalisation demonstrate that the use of this variant has increased in the speech of New Zealanders over the last twenty years (Bayard, 1990; Holmes, 1995a; Docherty et al., 2006).

Age and gender differences have been well established as important factors on sound change in NZE and we will review the different studies on /t/ in this section. We grouped the attested phonetic variants of /t/ into four main categories: (i) voiceless speech sounds involving full alveolar closure, (ii) voiceless speech sounds with partial closure, (iii) voiced speech sounds and (iv) glottal articulations. For each group we will try to give as many details as possible, in terms of the site of occurrence of the allophones as well as in terms of their sociolinguistic patterning[16].

1.6.1 Voiceless speech sounds involving full alveolar closure

In this section we discuss unaspirated [t], aspirated [t^h], affricated [t͡ʃ] and unreleased [t̚].

Sites of occurrence

Unaspirated [t] and affricated [t͡ʃ] are said to occur in all positions: word-initially when they are followed by a stressed or an unstressed vowel (Bauer, 1986; Holmes, 1995b), word-medially (Bell, 1977; Holmes, 1994) and phrase-finally (Docherty et al., 2006).

Aspirated [t^h] is found in initial position either at the beginning of a word or at the beginning of a morpheme. In these contexts, [t^h] wins over other possible phonetic variants: in Holmes' study of initial plosives, "80% of all tokens were aspirated, compared to 12% unaspirated and 8% voiced" (Holmes, 1995b). Bauer (1986, p.15) notes that phrase-final aspirated oral stops can also be in variation with unreleased stops.

Unreleased alveolar [t̚] can be found in final position but in the particular case of NZE they are almost always accompanied by creakiness in the preceding vowel (Docherty et al., 2006) and thus it seems more appropriate to categorise them into glottal articulations.

Sociolinguistic patterning

In initial position Holmes (1995b) found that the unaspirated stop was used more by Māori English (ME) speakers than Pākehā speakers (New Zealand speakers of European descent). She claimed that the variant "may serve as a marker of Maori ethnic identity and can be regarded as a potential feature of ME" (Holmes, 1995b, p.152). She also found a high level of unaspirated [t] amongst rural Pākehā working class males. However the background information on those speakers was not sufficient to tell whether they had more contacts with Māori speakers than Pākehā speakers. She left the question open and concluded that more research is needed.

[16] The acoustic and articulatory characteristics of most sounds won't be discussed into much detail as they have to be found in most varieties of English and they have been extensively described. However we will discuss the case of /t/ when it is realised as a fricative, because its phonetic description poses some problems (see also the discussion in chapter 2 and our articulatory account of the fricative in chapter 3).

In their analysis of phrase-final /t/ in the speech of young New Zealanders Docherty et al. (2006) collapsed the canonical, spirantised and affricated variants under the category released variants. We are treating the spirantised variant separately in this chapter but the authors have showed that the patterning of the released variants is socially conditioned. Females produce the highest rate of released variants and there is an interaction between social class and gender so that professional males produced more released variants than non-professional males.

1.6.2 Voiceless speech sounds involving partial closure

Sites of occurrence

Taylor (1996) identified a variant of /t/ realised as a fricative. The variant that he describes occurs in medial as well as in final position but not in initial position. This was the first mention of a fricated variant in NZE, that Docherty et al. (2006) also call the spirantised variant.

Sociolinguistic patterning

Taylor (1996) looked at the speech of gay men in NZE and found that the fricative realisation of /t/ is a feature of gay men's pronunciation, although its use is not as advanced as it is the in the speech of young straight females. As to straight males, its use was qualified as quite infrequent. Also the fricated variant appeared more often in free conversations that in wordlist readings and Taylor concluded that the fricated /t/ was a prestige form used by both gay men and straight women.

1.6.3 Glottal articulations

/t/ may vary in its degree of glottalisation. Attested variants range from oral stops made with a simultaneous glottal constriction to full-glottal stops (Bayard, 1990). The preceding vowel context may also be affected to some extent by the presence of laryngealisation, even though /t/ does not necessarily happen to be glottalised (Docherty et al., 2006). In other words, /t/ may involve oral closure only while the preceding vowel may retain a certain amount of creakiness. Docherty et al. (2006, p.381) comment on this:

> As already noted, a surprising number of the released tokens also contained a degree of laryngealisation during the preceding vowel, mainly in the second half of the vowel. This is a pattern that has not been reported in other varieties in the absence of a glottal variant, and is somewhat counter-intuitive given the key role of vocal fold abduction in the V-plosive transition in generating the aerodynamic

conditions necessary for the production of the substantial frication associated with spirantisation or affrication [...].

Sites of occurrence

According to Bauer (1986) the preferred sites of occurrence for glottalised/glottal are word-finally and syllable finally before another consonant as well as phrase finally, though very sporadically. More recent studies (*eg.* Holmes 1995a; Docherty et al. 2006) have found that glottalisation in word-final and phrase-final positions has in fact increased dramatically in the speech of New Zealanders over time since Bauer's study (*ie.* in 15 years' time). Also, a common site of occurrence of these variants is when the intervocalic /t/ precedes syllabic nasals, in words such as *fatten* or *button* (Holmes, 1994). Holmes notes that glottal replacement is very rare in other intervocalic contexts and found that they were "all word-final not word-medial as in the stereotypical Cockney [bɪˀə] *bitter* and [bʊˀə] *butter*" (Holmes, 1994, p.213). To finish, word-final glottal stops before vowels are said to "function most frequently as an emphatic device in relation to the following word" (Holmes, 1994, p.220), therefore serving a pragmatic purpose in the speakers' discourse. In that caes they were mainly reported in formal interviews and fewer occurrences were found in conversational speech with friends or family.

Sociolinguistic patterning

Bayard (1990) provided the first evidence of glottalization in NZE and showed than young speakers used more glottal stops than older speakers. In his sample young speakers produced 29% of word final /t/ as glottal stops. In a study conducted a few years later, Holmes (1995c) found an increase in the use of glottal stops for all age groups, with young working class females producing the highest rate (44%). She concluded than females were leading the change. 10 years later Docherty et al. (2006) found that glottalisation occured for 80% of the tokens they analysed in the speech of young New Zealanders. While their findings supported some of Bayard and Holmes results, they questioned the fact that women might still be leading the change as young males produced more unreleased forms than young females.

1.6.4 Voiced speech sounds

In this section we discuss voiced flap or tap [ɾ] and fully-voiced [d]. Studies have usually grouped the voiced flap or tap variants and the fully-voiced stop under the banner of T-voicing (Holmes, 1994, 1995b,c; Bayard, 1999), which is the terminology we will use here.

Sites of occurrence

T-voicing occurs in syllables with a Strong/Weak pattern, in words such as in *letter*, *city* or between words where the following word bears stress, as in *at all* (Holmes, 1994). Some instances have also been found very sporadically in initial position (Holmes, 1995b). T-voicing, however, is unlikely to be found "between an unstressed syllable in word-medial position, *ie.* when /t/ is initial in a stressed syllable (*eg. po'tato*, *ca'tastrophe*)" (Holmes, 1994). Holmes (1995b) also reported that no such variants were found after a pause, before a liquid, or when it as at the beginning of a Māori word.

Sociolinguistic patterning

Bell (1990) showed that the patterning of T-voicing is socially conditioned. He found that the same individual newsreaders used T-voicing more extensively when they spoke on local commercial stations than when they spoke on a radio station with an audience from a higher socio-economic background. Holmes (1994) provided the most extensive socio-linguistic study of T-voicing in NZE to date. Using Varbrul analysis, she looked at age, gender, class and speech style. She found that, overall, T-voicing was more frequent (i) for younger than older speakers, (ii) for males than females, (iii) for working-class speakers than middle class speakers and (iv) in conversations than formal interviews. She also showed that T-voicing is a sound change in progress and that it has entered NZE through the vernacular style of working-class speakers and is increasingly establishing itself in middle class speech, especially in conversational settings. Taylor (1996) came to the same conclusions in terms of the difference between straight males and straight females but added that gay men tended to follow the pattern of straight women, therefore having a less prevalent use of T-voicing in their speech.

1.6.5 Summary

Table 1.1 summarises where the allophones described above have been identified according to their position in the word.

/t/ allophone	initial position	medial position	final position
unaspirated [t]	✓	✓	✓
affricated [t͡ʃ]	✓	✓	✓
aspirated [tʰ]	✓		✓
unreleased [t̚]			✓
fricative		✓	✓
T-voicing	✓	✓	
glottal stop [ʔ]		✓	✓
laryngealisation			✓

Table 1.1 – Summary: /t/ allophones in NZE and their sites of occurrence.

1.7 Summary of the introduction and overview of the dissertation

In this introductory chapter we first discussed imitation at a general level. We then moved on to imitation in speech and justified our choice of studying it in a controlled laboratory setting. We introduced two theoretical frameworks that have been put forward to account for the results observed in the laboratory, namely the exemplar-based account and the gestural account. Our current research investigates imitation and allophony and our aim is to see which of these accounts can best explain our results. Thus we thus ask the following questions:

- how is allophonic variation represented at an abstract level?
- to what degree are different phonetic categories (allophones) cognitively linked to one single abstract category (the phoneme)?

In order to investigate these questions, we study imitation according to two paradigms: one that looks at imitation of the realisation of a phoneme within a given position in the word (*eg.* in medial position), and we ask:

- to what degree do manipulations of one phoneme extend across variants in the same position?
- if a speaker can learn to produce an unusual variant, would this ability influence the production of other allophones of the underlying phoneme?
- if so, will the pre-existing allophones be produced more like the unusual variant?
- would imitation of a variant drive allophonic selection?

The other paradigm that has been chosen is the study of imitation of the realisation of a phoneme across positions in the word. We also ask:

- is imitation in speech position sensitive?
- to what degree do manipulations of one phoneme extend across positions?

New Zealand English provides of wealth of allophonic variation across word positions for the phoneme /t/. Therefore it is an ideal language to investigate imitation and allophony. Before presenting our experimental designs and our results on imitation however, we will verify and further extend the work that has been done on the dialect in the next chapter. We analysed large corpora of spoken NZE. We found new allophones of /t/ and we discuss in this chapter a fricative realisation.

Chapter 3 investigates the fricative realisation in NZE further by means of a palatographic pilot experiment, as well as a perception experiment.

Building on our findings from the perception experiment, we investigate phonetic convergence in NZE towards an artificially created novel allophone in medial position using acoustic and EPG data (chapter 4). We examine its effect on other allophones within the same word position.

In chapter 5 we move on to imitation across word positions in NZE. In chapter 6 we present a follow-up to this experiment. This time we compare Standard French to NZE, because the former does not show much allophonic variation (Tranel, 1987). Different predictions are made for both languages and the implication for how allophonic variation might be represented at an abstract level is discussed.

The conclusion in chapter 7 summarises the findings, discusses what theoretical framework best accounts for them and contributes ideas for future research.

Chapter 2

Verification and extension of documentation work on NZE (corpus analysis)

2.1 Introduction

As we reported in section 1.6 previous studies showed that NZE medial /t/ can be realised as a canonical variant with varying degrees of aspiration, a flap, a fully-voiced variant, a glottal stop and a fricative (Bell, 1977; Holmes, 1994, 1995a; Bayard, 1999; Taylor, 1996). The literature we reviewed show that the patterning of medial /t/ is highly socially conditioned. Holmes (1994) who analysed medial /t/ in conversational settings found that T-voicing entered NZE through the vernacular style of working-class male speakers and increasingly established itself in middle class speech. Furthermore Taylor (1996), who included wordlist data in his sample, found that the voiced variant did not occur at all for females and gay males. He inferred that the prestige variant might be in fact the fricative as it was the most prevalent form in his wordlist data. However the number of participants recruited for the study was low (4 gay males, 3 straight males and 2 straight females). It is unclear whether this pattern would hold true with a larger number of speakers due to individual phonetic variation. Also he analysed 15 tokens per speaker, thus the total number of observations he made was 135. Given his small sample size he was not able to report statistical tests and suggested that the pattern he observed should be taken with caution.

Also, while the auditory identification of T-voicing is relatively straightforward – there is no doubt that the T-voicing analyses done in NZE are very reliable – it is less so for the fricative variant. Auditorily, the fricative variant is not saliently different from an aspirated stop. During the course of this research we have presented tokens containing medial fricatives to trained phoneticians and asked for their judgements. It was not uncommon for listeners to respond that they heard aspirated stops. In her analysis of medial /t/ Holmes (1994) categorised auditorily the realisations as being either voiceless aspirated stops, voiced variants (T-voicing) or glottal stops. Taylor (1996) was the first to discuss a fricative as a possible variant of /t/ in medial position and his analysis was also auditory-based. It is possible that Holmes' identification of /t/ as an aspirated variant was in fact a fricative. So far no acoustic-based categorisation of medial /t/ in NZE has been carried out to our knowledge.

In this chapter, we investigate medial /t/ variants in free conversation as well as in wordlist data. We searched for medial /t/ tokens in archived speech of New Zealanders born as early as the late 19^{th} century. Speakers were stratified according to broad sociolinguistic criteria. We analysed each medial /t/ token acoustically and auditorily. Our goal was to verify claims on the sociolinguistic patterning of medial /t/ in New Zealand as well as to extend previous documentation work on the dialect.

2.2 Method

The University of Canterbury holds recordings that span the history of New Zealand English from the late 19^{th} century until the present. These recordings are part of the ONZE archive (Origins of New Zealand English), which comprises three main corpora : the Mobile Unit, the Intermediate Archive and the Canterbury Corpus (Gordon et al., 2007).

The medial /t/ tokens that were analysed and presented in this chapter were searched within these corpora. All searches were carried out using ONZE Miner (Fromont and Hay, 2008). ONZE Miner is a browser-based interactive database which enables researchers to search across, interact with sound files with time-aligned lexical and phonetic transcriptions, and export the search results into CSV files together with links to the audio files.

We analysed lexical items consisting of two syllables with a strong/weak stress pattern : *city*, *letter*, *fatter*, *scatter*, *better*, *batter* and *Peter*. Those words were chosen because a larger number of speakers produced them in isolation in the Canterbury Corpus as part of a wordlist reading. This enabled us to compare them to the same words produced in free conversations in the other corpora.

2.2.1 Stratification of the speakers

The speakers presented in the corpora below were all stratified according to age and gender. Age is the most self-explanatory sociolinguistic variable used in the ONZE Corpus since it refers to the year of birth of the speakers. Gender refers to the sex of the speakers (*ie.* the state of being male or female). Additionally, the Canterbury Corpus has a binary professional/non-professional status associated with each speaker. The definition of professional status is defined by Gordon et al. (2007) as follows :

> [. . . A]ssigning a social class to speakers in New Zealand immediately raises methodological issues in that New Zealand society does not divide easily into the social classes used in many other English-speaking countries. A social class classification was derived for the Canterbury Corpus that combined information about occupation with information about education level. The score for occupation was arrived at for each speaker using a six-point scale derived from the Elley-Irving codes for census occupations (Elley & Irving, 1985; NZ Ministry of Education, 1990). The lower the number, the higher the social class rating of the occupation. A six-point scale was devised along similar lines to code the speakers' educational attainments (see Gregersen & Pedersen, 1991 for a similar way of assessing social class). These two scores were added together to give a range from 2 (high) to 12 (low). For the Canterbury Corpus, the professional speaker groups have average scores between 4 and 4.5 whereas the non-professional speaker groups have average scores of 8.5 to 9.5, thus indicating that the archive does, indeed, contain speakers representing two very different social groups. While this method for assigning social class is widely used by social scientists in New Zealand, it has a number of shortcomings. The ranking of occupation is made according to level of income but it is well known that some very rich people (such as some farmers) are able to claim very low incomes for taxation purposes. The scale is also out-dated and does not list some occupations popular today (such as counsellors).

2.2.2 Content of the corpora and gathering of the data

The Canterbury Corpus

The data was collected by members of the NZE class of the Linguistics Department at the University of Canterbury (see Gordon et al., 2007). The Canterbury Corpus (CC) includes recordings of New Zealanders born between 1930 and 1985. Speakers were stratified according to age, gender and professional background so that Female Young Professional speakers are labelled as FYP , Old Male Non-professional speakers are labelled as MON and so forth. The wordlist data was taken from this corpus. A search through ONZE Miner provided a total 432 wordlist recordings. 147 recordings were discarded either because of recording quality reasons or because we were not sure of how to categorise certain tokens. 285 wordlist recordings were retained, thus providing 2,023 medial /t/ tokens for analysis. Another search through ONZE miner for the same words in spontaneous speech taken from informal interviews yielded 385 words for analysis.

The Intermediate Archive

The data was collected by Rosemary Goodyear, Lesley Evans and members of the ONZE team (see Gordon et al., 2007). The Intermediate Archive (IA) includes recordings of New Zealanders born between 1890 and 1930. Background information on the speakers included age and gender. A search through ONZE Miner matching the lexical items present in the word lists returned 138 tokens for analysis.

The Mobile Unit corpus

The data was collected by the Mobile Disc recording Unit of the NZ Broadcasting Service (see Gordon et al., 2007). The Mobile Unit corpus (MU) includes recordings of New Zealanders born between 1851 and 1910. Background information on the speakers included age and gender. A search through ONZE Miner matching the lexical items present in the word lists returned 51 tokens for analysis.

2.2.3 Categorisation of phonetic variants

The realisations of intervocalic /t/ were divided into four main groups : canonical articulations, fricated articulations, glottal articulations and taps. Analyses were conducted auditorily and acoustically by looking at spectrograms for each word.

Canonical articulations were defined as having a closure gap (silence after the first vowel) which was followed by a release burst. The release burst could be followed by a certain amount of aperiodic noise[1] preceding voicing in the next vowel. Fricated articulations were defined as having aperiodic noise following the first vowel all the way through the second vowel. Glottal articulations were defined as having a closure gap without any visible burst release between the first and the second vowel. The second vowel could show a certain amount of creaky voice. Contrary to canonical, fricated and glottal articulations, the analysis of which relied solely on waveforms and spectrograms, taps were also analysed auditorily.

Canonical and fricated articulations were frequently pre-aspirated, which has never been reported before in NZE. They were thus classified as being pre-aspirated post-aspirated variants or as being pre-aspirated fricated variants respectively. Pre-aspiration is understood in this study as the brief apparition of aperiodic noise in the 3–10 kHz range in the first vowel offset, similarly reported by Jones and Llamas (2003, 2008)[2]. In the case of pre-aspirated and post-

[1] It is not clear whether the release was followed by aspiration or frication noise. Although most of the tokens could be heard as heavily fricated they were collapsed together with aspirated stops, as the main criterion for grouping articulations in this group was too see a closure followed by a clear burst on waveforms and spectrograms.

[2] Auditorily, pre-aspiration gives the impression of a strong expulsion of air as when breathing out. We

aspirated /t/'s this aperiodic noise ceases completely when closure is achieved. In the case of pre-aspirated and fricated /t/'s there is a drop in spectral energy at the offset of pre-aspiration followed by a rise in energy corresponding to the approximation of closure (*ie.* the frication of /t/). A clear gap between these two energy phases was hardly ever observed in that case. Rather there was a smooth transition from the drop to the rise. A comparison between these articulations is shown on spectrograms in figure 2.1.

Tokens in the three corpora were classified either as canonical articulations, fricated articulations, glottal articulations or taps. The further categorisation of /t/ into pre-aspirated variants was done for the Canterbury Corpus only.

hypothesise that pre-aspiration might be the result of an sudden increase in sub-glottal pressure causing the glottis to spread and maintaining high air velocity for a few milliseconds. These articulatory suppositions will be investigated in further studies.

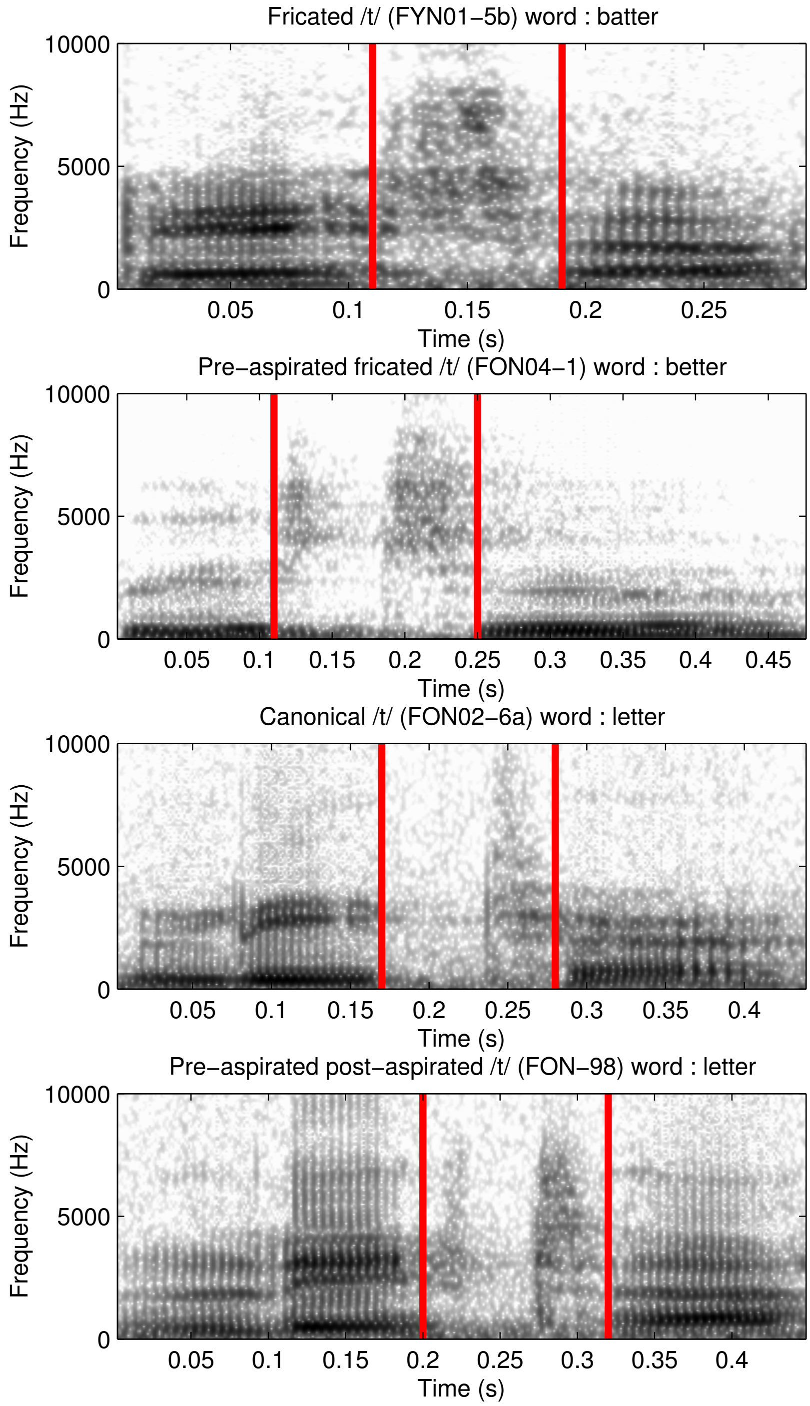

Figure 2.1 – Spectrograms of words produced by four individual speakers in the Canterbury Corpus. Areas of interest are bounded.

2.3 Results

First, the results for the wordlist data within the CC corpus are presented. This corpus is the only one which provides carefully read speech and is treated separately. Secondly, the results for spontaneous speech are presented, including data from the CC corpus, together with data from the MU and IA. Thirdly, a direct comparison between wordlist data and spontaneous speech is established. Finally, the results for pre-aspiration are presented.

2.3.1 Wordlist results within the CC (carefully read speech)

Table 2.1 shows the distribution of the allophones in the wordlist data. The most widespread variants are the fricated ones. They account for 67.33% of the total number of observations. In second position come the alveolar stop variants. They account for 28.87% of the total number of observations. Glottal and tap articulations are very rare in comparison: they account for less than 4% of the total number of observations in this dataset. Interestingly, 43% of all voiceless stops were pre-aspirated and 16% of the fricatives were pre-aspirated as well.

	male		female		
	N	P	N	P	**total**
canonical stop	141	99	69	25	**334**
pre-aspirated post-aspirated stop	57	51	50	92	**250**
fricated /t/	167	247	382	354	**1150**
pre-aspirated fricated /t/	27	70	48	67	**212**
tap	49	16	11	0	**76**
glottal stop	0	0	0	1	**1**

Table 2.1 – Raw counts per allophone present in the wordlist data. Data are separated by gender and professional status. N = non-professional ; P = professional.

In order to analyse the rate of use of the fricated allophones we fitted a logit mixed model (Bates, 2010). The dependent variable was a binary variable coding for either the presence or the absence of the fricated variant. The fricated variant and the pre-aspirated fricated variant were collapsed together. The independent variables were (i) `year of birth` : a continuous variable centred on its mean (1960), (ii) `gender` : a two level factor coding for either male or female, with female selected as the reference level, and (iii) `professional status` : a two level factor coding for either professional or non-professional status, with professional status selected as the reference level. Two random intercepts were added to the model, one that accounts for the variability across the 285 speakers who produced the words in isolation and one that accounts for the variability across the 7 words that were present in the wordlists (*city*, *letter*, *fatter*, *scatter*, *better*, *batter* and *Peter*). A total of 2023 binary measures were submitted to the model. The coefficients for the interaction between `year of birth` and `professional status` as well as `year of birth` and `gender` were non significant ($p = 0.63$ and $p = 0.49$ respectively) and were dropped from the final model.

Table 2.2 shows the output of the logit mixed model. The `year of birth` coefficient is significant and the slope is positive. This means that the probability of observing fricated variants increase over time in the corpus. The `gender` coefficient is negative and significant. As females were selected as the reference level, this means that males produce less fricated

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	2.84032	0.45956	6.180	$p <$.0001
year of birth	0.08877	0.01398	6.352	$p <$.0001
professional status (N)	-0.26280	0.56203	-0.468	0.6401
gender (male)	-1.36131	0.57660	-2.361	0.0182
professional status : gender	-1.93659	0.81044	-2.390	0.0169

Table 2.2 – Output of the model estimating the use of the fricated variant in the CC corpus, wordlist data.

variants than females. There is a significant interaction between `professional status` and `gender` and the slope of the coefficient is negative. Since professional status was selected as the reference level, this means that professional males use significantly more fricatives than non-professional males.

Figure 2.2 makes these effects more clearly visible. For both males and females there is an increase in the use of the fricative variant over time. For any given age, females use the variant more than males. While there is little difference between professional and non-professional females overall, professional males are more advanced in the use of the fricated variant than non-professional males.

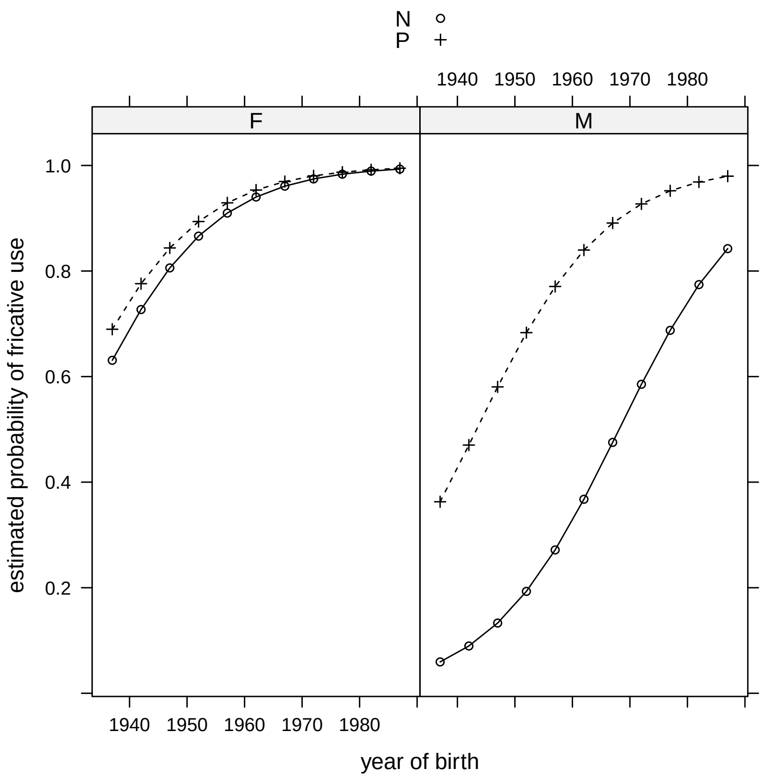

Figure 2.2 – Estimated probability of fricative use obtained from the model coefficients (careful speech). Left panel : females (F). right Panel : males (M). Dashed lines and plus signs : professional speakers (P). Continuous lines and circles : non-professional speakers (N).

We also analysed the probability of observing stops by submitting data to a logit mixed model similar to the one described above. This time the dependant variable was a binary variable coding for either the presence or the absence of the stops. The canonical variant and the pre-aspirated post-aspirated variant were collapsed together. `year of birth`, `gender` and `professional status` were selected as fixed effects and `words` and `speakers` were selected as random effects. Interactions between fixed effects were non-significant and were dropped from the final model.

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	-2.87140	0.33218	-8.644	$p <$.0001
year of birth	-0.09397	0.01411	-6.660	$p <$.0001
gender (male)	1.66470	0.40764	4.084	$p <$.0001

Table 2.3 – Output of the model estimating the use of stops in the CC corpus, wordlist data.

Table 2.3 shows the output of the model of the logit mixed model. The `year of birth` coefficient is significant and the slope is negative. Thus the probability of observing stops decreases over time in the corpus. The `gender` coefficient is positive and significant. As females were selected as the reference level, this shows that males produced more stops than females. These results can be seen on figure 2.3.

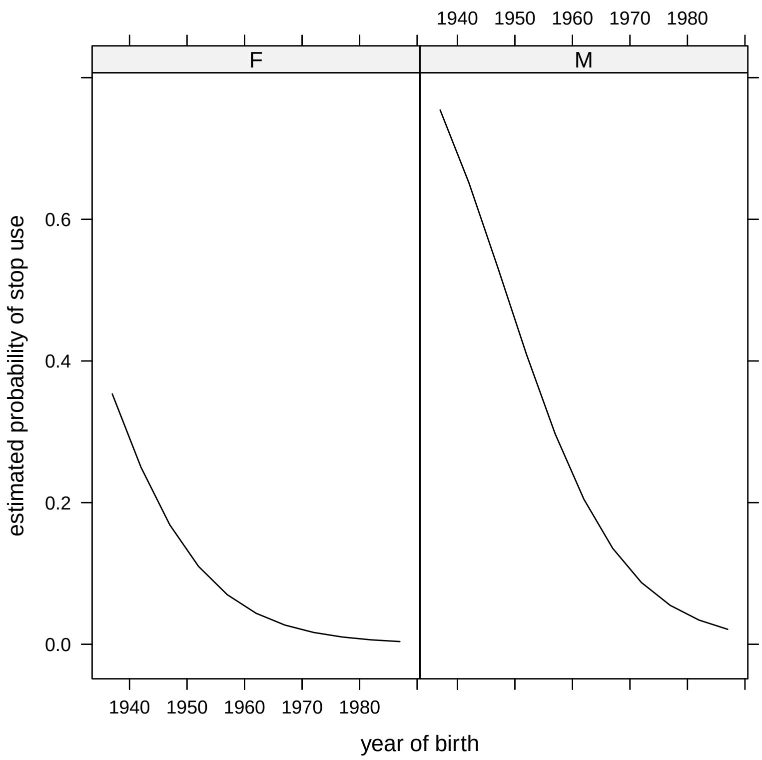

Figure 2.3 – Estimated probability of stop use obtained from the model coefficients (careful speech). Left panel : females (F). Right Panel : males (M).

Finally we focused on /t/ tokens realised as taps. We mentioned previously that 76 tokens out of 2023 were produced as taps, which accounts for less than 4% of the total number of observations in our dataset (see table 2.1). We subset the data and looked at individual patterns across words and speakers for those speakers who did produced taps (27 speakers in total). The results are summarised on table 2.4. Amongst those speakers the majority of them used the variant only once across the list of words (11 males and 4 females). 5 speakers (4 males and one female) used the tap variant consistently. The words that were tapped the most were *city* and *Peter*. In summary, very few tokens in the corpus were tapped and the pattern we observed was a preference for males to tap over females as well as preference for tapping *city* and *Peter* over the other words.

Taken together our results support the conclusion of Taylor (1996) that the fricative is the prestige medial /t/ variant, since it is the prevalent realisation found in careful speech. They also support his result that males used the variant less than females, at least at the time when his study was conducted. He recruited speakers born between 1962 and 1976 and it is clear from figure 2.2 that those male speakers were not as advanced in the use of the fricated variant as they are today. Since the sexual orientation of the speakers was not made available to us, we cannot prove or disprove his claim that gay men were more advanced in the use of the variant than straight men.

	batter	*better*	*city*	*fatter*	*letter*	*Peter*	*scatter*	**total**
fon96-29a	0	0	1	0	0	0	0	**1**
fyn00-20a	0	0	0	0	0	1	0	**1**
fyn02-12a	0	0	0	0	0	1	0	**1**
fyn94-12a	0	0	1	0	0	0	0	**1**
mon95-1a	0	0	0	0	0	1	0	**1**
mon96-10b	0	0	0	0	0	1	0	**1**
mon97-10b	0	0	0	0	0	1	0	**1**
mon99-1b	1	0	0	0	0	0	0	**1**
mop02-2a	0	0	0	0	0	0	1	**1**
mop94-5b	0	0	0	1	0	0	0	**1**
mop95-12c	0	1	0	0	0	0	0	**1**
myn02-19a	0	0	0	0	0	1	0	**1**
myn96-12b	0	0	1	0	0	0	0	**1**
myn96-14	0	0	1	0	0	0	0	**1**
myp08-2	0	0	1	0	0	0	0	**1**
mon95-7b	0	0	1	0	1	0	0	**2**
mon99-16a	0	0	0	0	1	0	1	**2**
myn94-11	0	0	1	0	1	0	1	**3**
myn97-9	0	0	1	1	1	0	0	**3**
mon03-5c	0	0	1	1	1	1	1	**5**
myp99-16b	1	1	1	1	0	1	0	**5**
mon95-4b	1	1	1	1	0	1	1	**6**
fyn95-25b	1	1	1	1	1	1	1	**7**
mon96-22b	1	1	1	1	1	1	1	**7**
myn96-15a	1	1	1	1	1	1	1	**7**
myn98-16b	1	1	1	1	1	1	1	**7**
myp98-8a	1	1	1	1	1	1	1	**7**
total	**8**	**8**	**16**	**10**	**10**	**14**	**10**	**76**

Table 2.4 – Raw counts of taps per speaker and per word.

2.3.2 Results for the CC, IA and MU (spontaneous speech)

In order to allow a comparison with spontaneous speech, we investigated interviews from MU, IA and CC interviews of New Zealanders. There were very few instances of the same words that occurred in the wordlist. To balance our data set, we decided to retain tokens found for the words *city*, *letter* and, *better* and we discarded the other words as only a few instances were found. In total, we found 573 tokens for analysis. The distribution of allophone counts amongst the remaining data set is as shown on table 2.5.

Raw counts from this table were then expressed as a percentage per age group. Thus, to observe the probability prob(c) for canonical stops produced by speakers born between 1850 and 1870, we proceeded as follows : prob(c) $= 8/$total per age group $= 8/5 = 0.5$. This yielded a bar chart tracing the evolution of each allophone, as shown on figure 2.4. Given that there were relatively few glottal stops or pre-aspirated variants in this data set, we focused on canonical, fricated and tap allophones. This is the bottom plot on figure 2.4. Both the bar

year of birth	canonical	pre-aspirated post-aspirated	fricated	pre-aspirated fricated	glottal	tap	**total**
1850-1870	8	0	2	0	0	6	**16**
1870-1890	14	0	4	0	2	10	**30**
1890-1910	15	0	11	0	0	10	**36**
1910-1930	26	0	48	0	0	21	**95**
1930-1950	13	5	41	3	0	31	**93**
1950-1970	1	2	49	2	0	39	**93**
1970-1990	4	0	55	0	1	150	**210**

Table 2.5 – Number of observations per age group for the words *city*, *letter* and, *better*

chart and the bottom plot display the same information but they do it in a different way. On the bottom plot, the size of the dots indicates how many tokens were analysed per age-group (*ie.* their relative size is calculated using the last column – total per age group – from table 2.5). Given that we had very few tokens for analysis for the spontaneous data, this gives an indication as to the reliability of our measures. It is relatively poor for speakers born between 1850 and 1910 but much higher for speakers born after 1910.

We can see that while canonical stops were the most used variants by the oldest speakers, they decrease steadily and are hardly to be found in the speech of young New Zealanders today. Fricatives were already to be found in the speech of the oldest speakers. There is an increase in their use until they reach a plateau amongst speakers born between 1910 and 1950. For those speakers, the fricative is the most widely used variant. From 1950 onwards, there is a decline in its use in favour of the tap, which is the most widespread variant today.

Since the tap variant is the most widely used variant today we fitted a logit mixed model on the CC spontaneous data. The dependent variable was a binary variable coding for either the presence or the absence of taps. The independent variables were (i) `year of birth` : a continuous variable centred on its mean (1960), (ii) `gender` : a two level factor coding for either male or female, with female selected as the reference level, and (iii) `professional status` : a two level factor coding for either professional or non-professional status, with professional status selected as the reference level. Two random intercepts were added to the model, one that accounts for the variability across the 164 speakers who produced the words and one that accounts for the variability across the 3 words that were used (*city*, *letter* and, *better*). A total of 383 binary measures were submitted to the model. Non-significant interactions ($p > 0.2$ using a Likelihood ratio test) were dropped from the final model using a backward stepwise procedure.

Table 2.6 shows the output of the logit mixed model. The `year of birth` coefficient is significant and the slope is positive. This means that the probability of observing taps increases over time in the corpus. The `gender` coefficient is positive and significant. As females were selected as the reference level, this means that males produce more taps than females. The

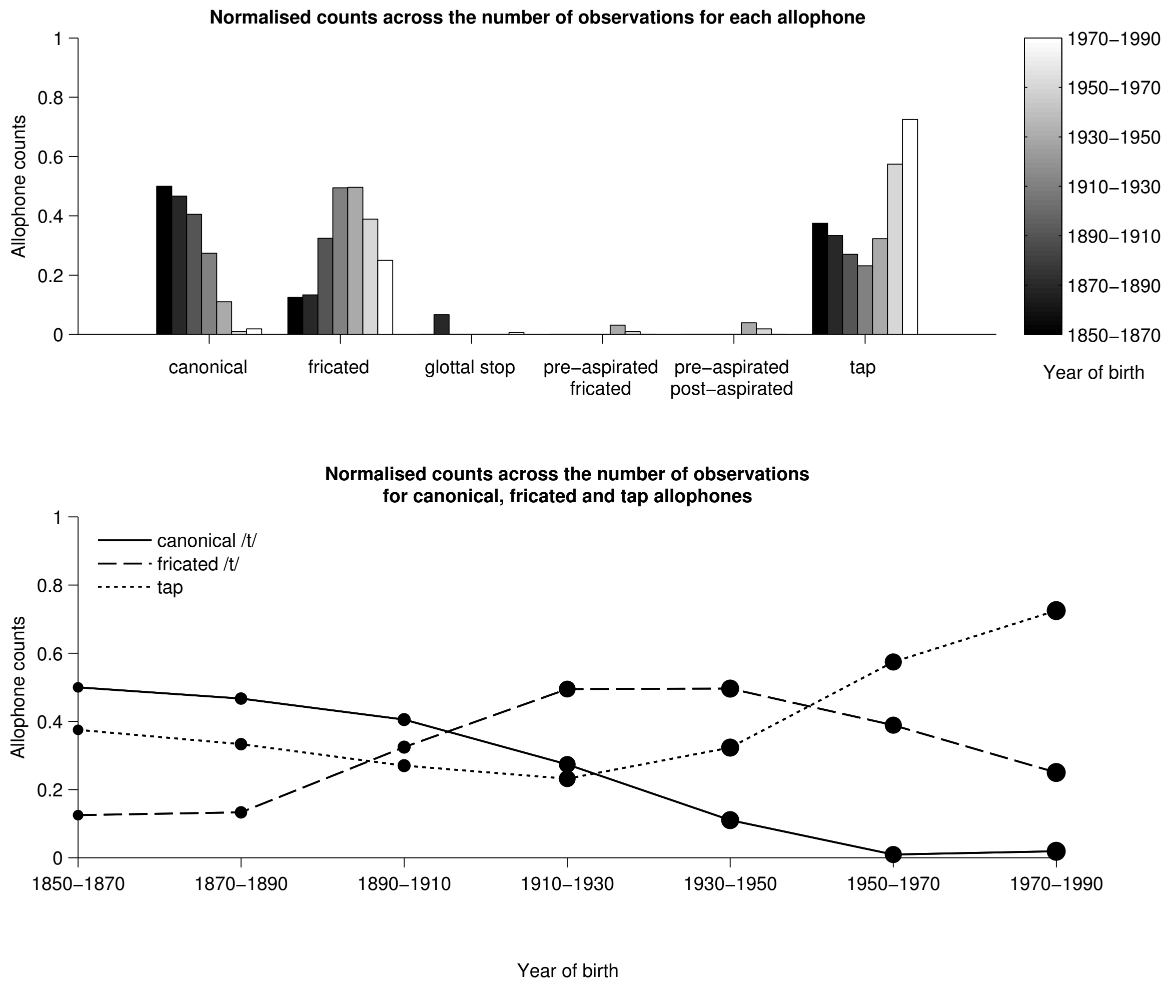

Figure 2.4 – Normalised counts across the number of observations for the words *city*, *letter* and, *better*

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	-0.20957	0.46585	-0.450	0.653
year of birth	0.08817	0.01401	6.295	$p <$.0001
professional status (N)	-1.69448	0.39557	-4.284	$p <$.0001
gender (male)	2.41710	0.40188	6.014	$p <$.0001

Table 2.6 – Output of the model estimating the use of taps in the CC corpus, spontaneous data.

`professional status` coefficient is significant and the slope is positive. Thus professional speakers use less taps than non-professional speakers.

Figure 2.5 makes these effects more clearly visible. For both males and females there is an increase in the use of the tap variant over time. For any given age, males use this variant more than females and, overall, non-professional speakers are more advanced in the use of the tap variant than professional speakers. Our results from spontaneous speech follow the pattern identified by Holmes (1994) whereby T-voicing has entered through the vernacular style of working-class male speakers and has increasingly established itself in middle class speech.

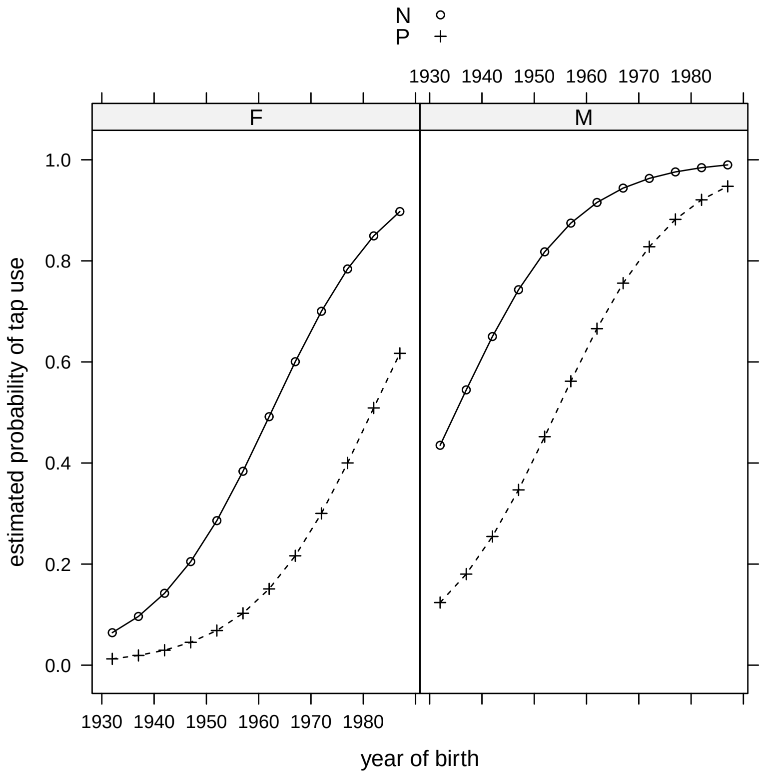

Figure 2.5 – Estimated probability of tap use obtained from the model coefficients (spontaneous speech). Left panel : females (F). Right Panel : males (M). Dashed lines and plus signs : professional speakers (P). Continuous lines and circles : non-professional speakers (N).

We also analysed the probability of observing fricatives by submitting data to a logit mixed model. This time the dependent variable was a binary variable coding for either the presence or the absence of fricatives. The fricated variant and the pre-aspirated fricated variant were collapsed together. `year of birth`, `gender` and `professional status` were selected as fixed effects and `words` and `speakers` were selected as random effects. Interactions between fixed effects were non-significant and were dropped from the final model.

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	1.33717	0.43552	3.070	0.00214
year of birth	-0.05605	0.01260	-4.448	$p <$.0001
professional status (N)	-1.80663	0.37824	-4.776	$p <$.0001
gender (male)	-2.31774	0.38084	-6.086	$p <$.0001

Table 2.7 – Output of the model estimating the use of fricatives in the CC corpus, spontaneous data.

Table 2.7 shows the output of the model of the logit mixed model. The `year of birth` coefficient is significant and the slope is negative. Thus the probability of observing fricatives decreases over time in the corpus. The `professional status` coefficient is negative and significant. Professionals were selected as the reference level, thus non-professionals produced less fricatives than professionals. The `gender` coefficient is negative and significant. As females were selected as the reference level, this shows that males produced less fricated variants than females. These results can be seen on figure 2.6.

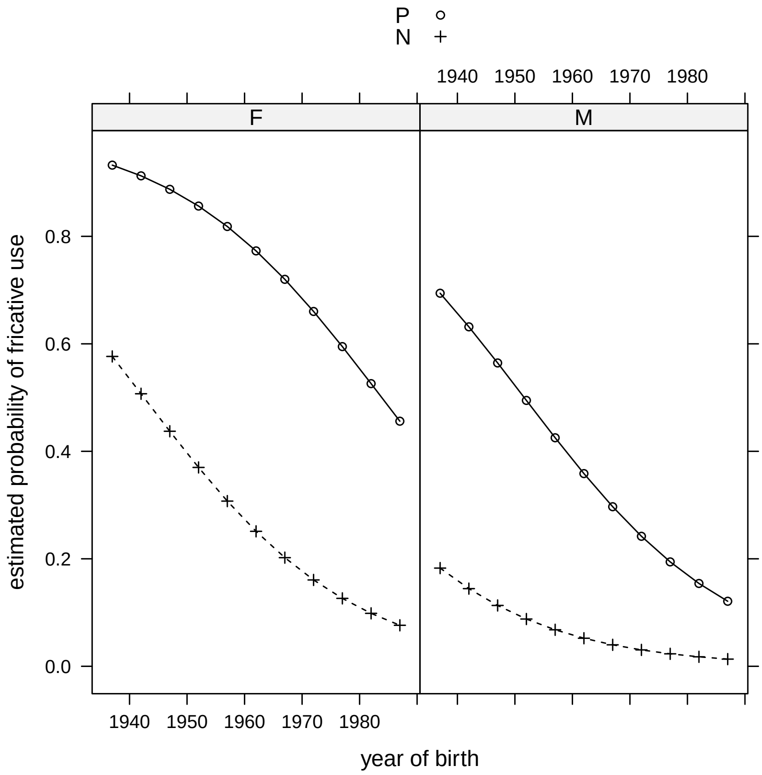

Figure 2.6 – Estimated probability of fricative use obtained from the model coefficients (spontaneous speech). Left panel : females (F). Right Panel : males (M). Dashed lines and plus signs : professional speakers (P). Continuous lines and circles : non-professional speakers (N).

Finally we focused on /t/ tokens realised as taps. We mentioned previously that 76 tokens out of 2023 were produced as taps, which accounts for less than 4% of the total number of observations in our dataset (see table 2.1). We subset the data and looked at individual patterns across words and speakers for those speakers who did produced taps. The results are summarised on table 2.4. 27 speakers produced taps. Amongst those speakers the majority of them used the variant only once across the list of words (11 males and 4 females). 5 speakers (4 males and one female) used the tap variant consistently. The words that were tapped the most were *city* and *Peter*.

2.3.3 Direct comparisons between careful speech and spontaneous speech

Our results have shown that speakers use variants differently in careful speech compared to spontaneous speech (*eg.* the fricated variant is most widespread in wordlist data while the tap is most widespread in conversation). In order to further account for the effect of type of speech on allophonic selection we run models on a subset of data that allows for direct comparison.

We selected the CC corpus and merged wordlist data and spontaneous data. Since we retained 3 words for our previous analysis of spontaneous speech (*city*, *letter* and, *better*) we selected the same words from our wordlist data. Then we retained speakers who were present in both corpora. This yielded a total of 719 tokens for analysis spoken by 141 speakers. We further collapsed the fricated and pre-aspirated variant together as well as the canonical variant and the pre-aspirated post-aspirated variant. One instance of a glottal stop was found and was removed from the dataset. Table 2.8 shows the variants' raw counts in this comparative dataset.

	speech style	
	careful speech	spontaneous speech
stops	130	16
fricatives	280	114
taps	15	164

Table 2.8 – Raw counts of broad variants in the comparative dataset.

Our set of models was obtained by testing separately for the rate of use of broad variants as a function of type of speech using a logit mixed model. This yielded 3 models : one for stops, one for fricatives and one for taps. Each dependent variable was a two level factor coding for the presence or absence of one of the broad variants. The independent variable was `type of speech`, a two level factor coding for either careful speech or spontaneous speech. Careful speech was selected as the reference level. `speakers` and `words` were selected as random effects.

In all 3 models the effect of `type of speech` was significant ($p <$.0001). When predicting the probability of observing fricatives and stops the coefficients were negative ($coef = -1.6680$ and $coef = -3.5253$ respectively). When predicting the probability of observing taps, the coefficient was positive ($coef = 5.3371$). As careful speech was selected as the reference level, the probability of observing fricatives and stops decreases in spontaneous speech while the probability of observing taps increases in spontaneous speech.

2.3.4 Pre-aspiration results within the CC (carefully read speech)

Pre-aspiration could not be clearly identified within the MU and IA data given the quality of the audio recordings. However the CC corpus provided higher quality recordings and pre-aspiration could be clearly seen on spectrograms. Within the CC wordlist data 43% of all voiceless stops were pre-aspirated and 16% of the fricatives were pre-aspirated, as shown on table 2.1. On the other hand, the spontaneous data provided 12 cases of pre-aspiration in total, as seen on table 2.5, which accounts for 3% of the number of observations in spontaneous speech. Thus only the results within the CC wordlist data are presented in this section.

A logit mixed model was fitted on the CC wordlist data. The dependent variable was a binary variable coding for either the presence or the absence of pre-aspiration. The independent variables were (i) `year of birth` : a two level factor coding for either older or younger speakers, as defined by Gordon et al. (2007) , (ii) `gender` : a two level factor coding for either male or female, with male selected as the reference level, and (iii) `professional status` : a two level factor coding for either professional or non-professional status, with non-professional status selected as the reference level. Two random intercepts were added to the model, one that accounts for the variability across the 164 speakers who produced the words and one that accounts for the variability across the 3 words that were used (*city*, *letter* and, *better*). A total of 383 binary measures were submitted to the model. Non-significant interactions ($p > 0.2$ using a Likelihood ratio test) were dropped from the final model using a backward stepwise procedure. Two random intercepts were added to the model, one that accounts for the variability across the 285 speakers who produced the words in isolation and one that accounts for the variability across the 7 words that were present in the wordlists (*city*, *letter*, *fatter*, *scatter*, *better*, *batter* and *Peter*). A total of 2023 binary measures were submitted to the model. Non-significant interactions were dropped from the final model.

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	-2.3958	0.3318	-7.221	$p <.0001$
age (young)	-0.5804	0.4623	-1.256	0.209278
professional status (P)	1.6425	0.4347	3.779	$p <.0001$
professional status (P): age (young)	-1.6223	0.6355	-2.553	$p <.02$

Table 2.9 – Output of the model estimating the use of pre-aspiration in the CC corpus, wordlist data.

Table 2.9 shows the output of the logit mixed model. The `professional status` coefficient is positive and significant. As non-professional status was selected as the reference level, this means that professionals produce more pre-aspirated variants than non-professionals. There is a significant interaction between `professional status` and `age` and the slope of the coefficient is negative. Since non-professional status was selected as the reference level, this means that age has an effect, with older professionals using significantly more pre-aspiration than younger professionals. This interaction also means that the age effect only holds for professionals, which can be more clearly seen on figure 2.7.

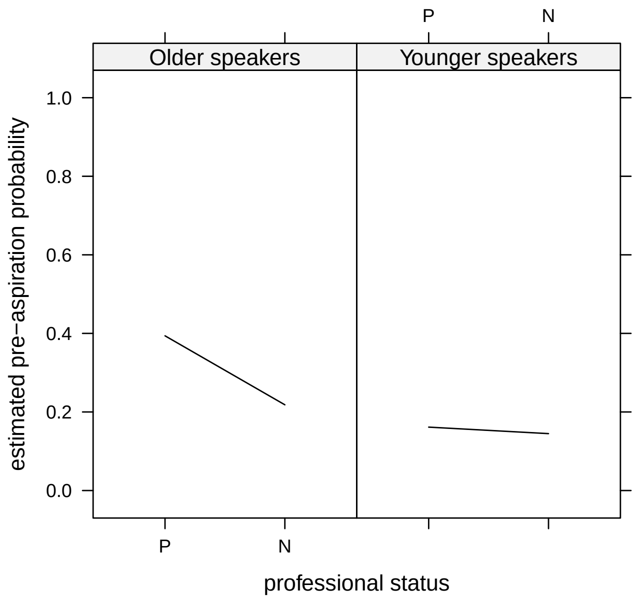

Figure 2.7 – Estimated probability of pre-aspiration obtained from the model coefficients.

2.4 Discussion

Previous studies of medial /t/ have reported the existence of a canonical variant with varying degrees of aspiration, a flap, a fully-voiced variant, a glottal stop and a fricative (Bell, 1977; Holmes, 1994, 1995a; Bayard, 1999; Taylor, 1996). Our study contributed to the literature on NZE by reporting frequent cases of pre-aspiration. Silverman (2003) argues that pre-aspiration is rare both diachronically and synchronically across human languages given its lack of phonetic salience. However recent acoustic evidence shows that pre-aspiration is present is several varieties of English as a non-normative feature (*ie.* a non-obligatory phonetic feature that can variably present or absent in different speakers of a single dialect). Jones and Llamas (2003) reported pre-aspiration of word-final plosives and fricated /t/'s in Middlesbrough English. Gordeeva and Scobbie (2007) identified pre-aspiration of fricatives in Standard Scottish English which may serve to enhance prosodic cues and phonemic voicing contrasts. They did not found a correlation between sociolinguistic variable and rates of pre-aspiration of fricatives. Non-normative pre-aspiration of /t/ has also been found in Northern Welsh dialects (Morris, 2010) and Australian English (Jones and McDougall, 2009). In our wordlist data pre-aspiration occurs in relatively high proportions : it accompanies 43% and 16% of all voiceless stops and fricatives respectively. On the other hand it appears in negligible proportions in spontaneous speech. We might hypothesise that pre-aspiration is a feature of careful speech in NZE. We analysed data using a logit mixed effect similar to the ones reported in this chapter. The dependent variable coded either for the presence or absence of pre-aspiration. The model predicted that younger speakers use little pre-aspiration, regardless of their professional status. Older speakers use more pre-aspiration and even more so when they are professionals. There was no effect of gender, showing that males and females used pre-aspiration in equal proportions. Our results suggest that the use of pre-aspiration is socially conditioned but at this stage it is difficult explain why this would be the case. Further research is needed to investigate this question.

Our results also support the conclusion of Taylor (1996) that the fricative is the prestige medial /t/ variant, since it is the prevalent realisation found in careful speech. Our data shows that, overall, females use the variant more than males. The youngest females achieve a rate of a 100%. Moreover our results show an interaction with gender and professional status so that the youngest male professionals also achieve similar rates. Professional males are more advanced in the use of the fricated variant than non-professional males. This prestige variant has therefore entered NZE through the speech of females and has strongly established itself in the dialect.

Our results on spontaneous speech also support the claim by Holmes (1994) that T-voicing has entered through the vernacular style of working-class male speakers and has increasingly established itself in middle class speech. We found that it was the most widespread variant in the conversational data. We also found evidence in the Mobile Unit corpus that the fricative was already present in NZE and used by speakers born between 1850 and 1870. Clearly, the fricative variant is not a new feature of NZE but apart from the research conducted by Taylor (1996) no mention of the fricative as a possible allophone for /t/ in medial position was made before. However the articulatory nature of the sound is not very well understood. This variant is very well-known to occur in Irish English. Hickey (1984, 234) describes it as "formed by bringing the apex of the tongue close to the alveolar ridge as if for the articulation of /t/ but stopping just before contact". He proposed the symbol [t̮]. (Pandeli et al., 1997) provided EPG data on Irish English and demonstrated that this choice for a phonetic transcription was in fact problematic. They proposed to use the symbol [θ̠] instead. Also it is not guaranteed that the acoustic nature of the fricated variant is the same in Irish English and in NZE. So far we have not proposed a phonetic transcription but we believe it is safer to refer to the variant in plain text as a fricative or as fricated /t/. This is the position we adopt in the rest of the thesis.

Finally we compared careful speech data and spontaneous data directly and found that the probability of observing fricatives and stops decreases while the probability of observing taps increases in spontaneous speech.

In summary we replicated previous results on the distribution of medial /t/ allophones in NZE. We also contributed to the description of the dialect by further identifying pre-aspirated variants and we provided the first statistical results on the sociophonetic patterning of the fricated variant. Further work on medial /t/ in NZE will aim at better understanding the sociophonetic patterning of pre-aspiration that we have observed. Further work is also needed to understand the acoustic, articulatory and perceptual nature of the fricated allophone and we will explore these questions in the next chapters.

It is important to note that this chapter is a means to describe the different variants found in NZE by providing acoustic descriptions as well as socio-linguistic data. Our task here was to provide an exhaustive description of the allophones. However our work on imitation will not focus on socio-linguistic aspects *per se*. Instead, we will focus on acoustic and articulatory aspects of /t/ allophones to study convergence. Section 7.3 will address how socio-linguistic aspects ought to be incorporated into the study of convergence and some of the theoretical questions that may arise.

Chapter 3

Production and perception of lingual fricatives in New Zealand English

3.1 Introduction

In the previous chapter we studied the use of a fricative for /t/ in medial position. Our goal in this chapter is to document the variant further on acoustic, articulatory and perceptual grounds. More specifically we will compare the fricated variant to the fricatives /s/ and /ʃ/ and investigate whether they share some degree of similarity.

Some articulatory data is available on Irish English slit-t (Pandeli et al., 1997) and acoustic data is available on fricated /t/ in Australian English (Jones and McDougall, 2009). In both studies the fricated variant was compared to /s/ and /ʃ/. Pandeli et al. (1997, 68) report typical EPG patterns as shown on figure 3.1 and describe the /t/ variant as following:

> Analysis of the data revealed that contact was concentrated at the sides of the palate across an area stretching from the front to the middle or back of the alveolar ridge. A rather broad fricative channel was evident, whose mean width was estimated to be 10.4 mm. This width was very similar to that estimated for tokens of [ʃ] produced by the same speaker (10.2 mm), but significantly greater than that for tokens of [s] (6.0 mm). [...] Finally, the anterior edge of the area of contact between the tongue and the palate was typically rather flat in shape for slit-t, with no contact at the sides of the palate in front of the major constriction.

They based their conclusions on 36 repetitions for each lingual consonant in 4 different phonetic environments and their data was obtained from one speaker. There might be a great deal of variation that could not be captured in their study (*ie.* intra/inter speaker variation) but their results nevertheless show that the fricated [t] is different from [s] and [ʃ].

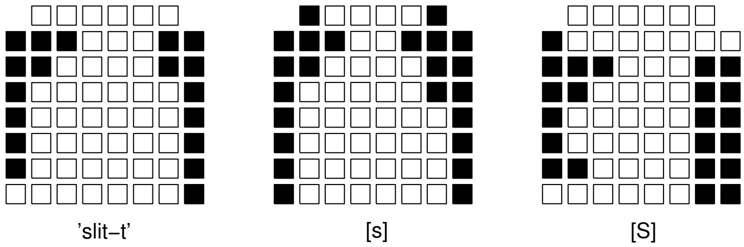

Figure 3.1 – Linguopalatal contact patterns during a single 'target' EPG frame for 'slit-t', [s] and [ʃ] in *met*, *mess* and *mesh* respectively. Adapted from Pandeli et al. (1997).

Jones and McDougall (2009) analysed acoustic data from five female Australian English speakers. They were interested in the durational and spectral properties of the fricatives. Their findings showed a great degree of similarity between /t/ and /ʃ/ in terms of spectral measures, and they argued that duration is the most likely perceptual candidate to distinguish them, with /t/ being significantly shorter than /ʃ/[1].

In this chapter we will first present an acoustic-palatographic pilot experiment conducted on one female NZE speaker. We will try to establish degrees of similarities between the fricated and /s/ and /ʃ/ on articulatory grounds. In the second part of the chapter we will present a perception experiment and we will try to establish whether /t/ is perceptually closer to /s/ than to /ʃ/.

3.2 Palatographic pilot study

3.2.1 Data acquisition

A female NZE speaker was asked to produce in isolation repetitions of [fɒtə], [fɒsə] and [fɒʃə]. These tokens were chosen because they are all non-words with [t], [s] and [ʃ] in intervocalic position, and they involve only one lingual gesture, which is required for palatographic analysis.

The speaker was presented with the non-words in orthographic form (*fotter*, *fosser* and *fosher*) consecutively on a computer screen until 100 repetitions were gathered. In between each word, their tongue was painted with a mixture of charcoal powder and olive oil. They read out the words to a microphone, then placed their mouth onto a mirror attached to a palatographic bench so as to form an 45° angle with the upper jaw. A photo was taken using a camera attached to the other end of the bench and the speaker was asked to rinse their mouth after the photo was taken. We used an AKG C520 head-mounted microphone. The signal was pre-amplified using an external USB-Pre device. Recordings were done using Audacity with a PC. They were saved as WAV files at 44.1 kHz 16 bit. Photos of the palate were taken using a Canon EOS 1000D camera and a 18-55mm kit lens focusing on the mirror with manual

[1]Note that Pandeli et al. (1997) also report shorter durations for /t/ than /s/ and /ʃ/.

exposure and aperture settings to maintain consistency across photographs. Photos were saved as JPEG files. The palatographic bench was designed by the Laboratoire Parole et Langage.

Table 3.1 presents the realisations of the 99 tokens that were collected. Tokens realised as fricatives were selected for analysis. First we present our acoustic results and then we move on to our articulatory results.

speech sounds	realised as a fricative	realised as a stop or an affricate
[t]	27	6
[s]	33	
[ʃ]	33	
Total		**99**

Table 3.1 – Number of realisations

3.2.2 Acoustic results

We measured the duration of the intervocalic consonants on the acoustic signal (the start marker was positioned at the offset of modal voicing for the first vowel and the end marker was positioned at the onset of modal voicing for the second vowel). The duration of the intervocalic consonants is shown on the box and whiskers plots in figure 3.2.

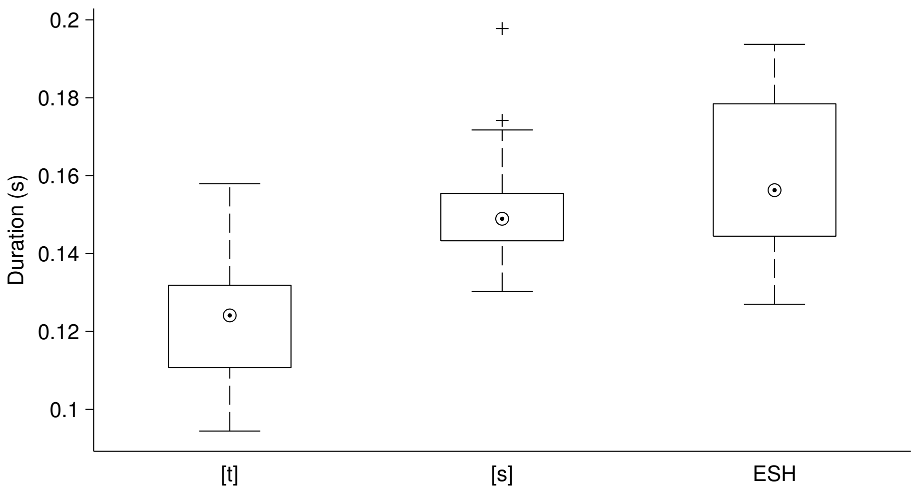

Figure 3.2 – Box and whiskers plots showing the duration of fricated /t/, /s/ and /ʃ/.

We run Wilcoxon non-paired tests for equal medians and found very marked differences in duration: /t/ is shorter than /s/ and /ʃ/ ($p < 0.0001$) while /s/ and /ʃ/ do not differ in duration ($p = 0.058$). We now turn to our palatographic results.

3.2.3 Articulatory results

A total of 90 palotagrams were visually inspected. We analysed tokens realised as fricatives (we discarded the 6 /t/ tokens produced as stops or affricates) and we discarded two tokens of /ʃ/ and one token /s/ where the speaker unintentionally made full tongue-palate contact before we took photographs. The first series of palatograms are shown of figure 3.3. All palatograms can be found in appendix A for reference. We observed the following patterns: /t/ tokens are generally produced with a very large fricative channel compared to /s/ and /ʃ/ tokens. Differences in channel between /s/ and /ʃ/ tokens are less evident. The location of the constriction for /t/ and /s/ tokens is very similar and it is slightly more anterior relative /ʃ/ tokens. Finally the side contacts are relatively thinner for /t/ and /s/ tokens compared to /ʃ/ tokens. Thus it seems that the fricated /t/ differs mainly between /s/ and /ʃ/ tokens in terms of the size of the channel. /t/ seems closer to /s/ with respect to constriction location and side contacts than it is to /ʃ/.

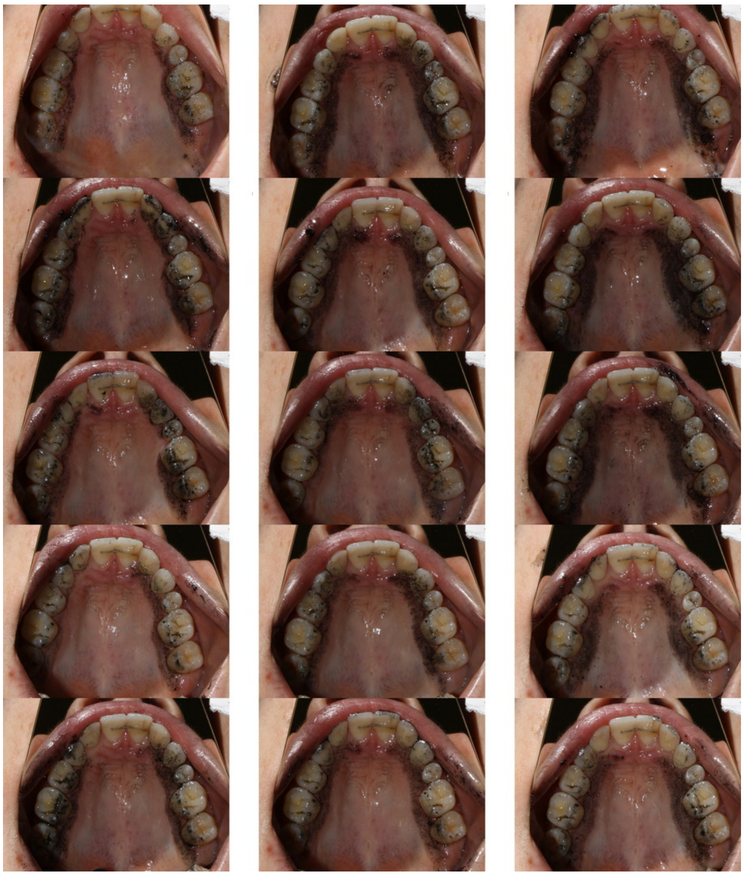

Figure 3.3 – From left to right: palatograms of fricated [t], [s] and [ʃ]. Series 1: tokens 1 to 15.

3.2.4 Short discussion

It is difficult to draw any definitive conclusions about the articulatory make-up of fricated /t/ in NZE since this pilot experiment is limited to one speaker. The gathering of static palatography data is time consuming and recruiting a large number of participants is not an easy task. But this will be made easier in the near future as computational tools for quantitative analyses of palatograms are being currently developed at the Laboratoire Parole et Langage (Legou et al., 2008). An extensive palatographic study presents both theoretical and documentary implications: to further investigate further the mapping between acoustics and articulatory data and to extend documentary work on NZE. Further research could also include a larger set of acoustic measures such as those presented in Jongman et al. (2000) to distinguish places of fricative articulation in American English. They measured spectral peak location, spectral moments (Forrest et al., 1988), locus equations (Sussman and Shore, 1996), normalised RMS amplitude and relative amplitude (Hedrick and Ohde, 1993). Jongman et al. (2000) report that spectral peak location, spectral moments and amplitudinal information prove successful and robust in distinguishing between all places of articulation. Jones and McDougall (2009) used similar measures and concluded that fricated /t/ in Australian English was closer to /ʃ/ than to /s/ acoustically. They also inferred from their acoustic results that /t/ was closer to /ʃ/. Our results show that /t/ is closer to /s/ than to /ʃ/ on articulatory grounds. The articulatory patterns that we observed are similar to the patterns reported by Pandeli et al. (1997) on Irish English slit-t.

According to their acoustic results Jones and McDougall (2009) argued that the main difference between fricated /t/ and /ʃ/ in Australian English is a question of duration. In their view, duration is the most likely candidate to distinguish between these two fricatives. However they did not address directly whether /t/ was perceptually closer to /s/ or to /ʃ/. Our palatographic results show that /t/ is closer to /s/ in NZE in articulatory terms and we will further investigate this finding by means of a perception experiment.

3.3 Perception experiment

3.3.1 Review of literature: acoustic cues to the perception of /s/ and /ʃ/

Numerous perception studies have investigated acoustic cues that play an important role in identifying fricatives in terms of place of articulation, manner of articulation and voicing. In English, and more specifically for /s/ and /ʃ/, two major cues which contribute to their correct identification are the spectral information that they carry (Heinz and Stevens, 1961; Harris, 1958; LaRiviere et al., 1975) and the duration of frication noise (Grimm, 1966; Jongman, 1989;

Kluender and Walsh, 1992). All studies examining the duration of frication noise agree that correct identification scores increase as duration increases. As an example duration can affect the perception of fricatives in terms of manner and place of articulation so that if the duration of naturally produced fricatives in a CV context is manipulated and sufficiently short, listeners perceive a stop consonant (Jongman, 1989). More specifically Grimm (1966) showed that shortening the duration of /s/ and /ʃ/ lead to the perception of /t/. Rise time might also be a cue to correctly identify these fricatives (Cutting and Rosner, 1974; Howell and Rosen, 1983) although Kluender and Walsh (1992) argue that their results were confounded by the duration of frication noise and that rise time alone should not be considered as a primary cue[2]. All the studies mentioned above investigated the perception of /s/ and /ʃ/ in CV sequences using either synthetic speech stimuli or natural utterances but less is known about the perception of these fricatives in CVC sequences. One study of intervocalic consonants in British English was conducted by Smits (2000) using forward and backward gating (*ie.* by presenting the initial part or the final part of an utterance) to examine the temporal distribution of perceptually relevant information for consonant recognition.

3.3.2 Aim of this experiment

From the above mentioned perception studies we have learnt that spectral information and frication duration plays a major role in identifying consonants. Our acoustic data indicate that /t/ and /s/ are spectrally close but /t/ is shorter in duration. Therefore would a shortened /s/ be perceived as /t/ if the main difference between them was merely a question of duration? Also would /t/ be perceptually closer to /s/ than to /ʃ/?

In this experiment we present NZE listeners with natural utterances of CVCV non-words. We truncated parts of the intervocalic consonants by steps of 10ms, yielding a range of durations from long to short. We also included the interdental fricative /θ/ to extend comparisons between all lingual fricatives to be found in NZE.

3.3.3 Methodology

Elicitation of stimuli

Four young Pakeha were recorded in order to produce the stimuli used in the perception experiment. They were two female and two male speakers and they were aged between 18 and 22. At the time of the experiment they were linguistics students at the University of Canterbury and were selected because they fricated a large number of intervocalic-/t/s in a

[2] In a more recent psychophysical study on Japanese, Mitani et al. (2006) provide an account of the interaction between frication duration and rise slope defined as the rate of change in sound pressure

previous pilot acoustic study. The female speakers are referred to as FS1 (for Female Speaker 1) and FS2. The male speakers are referred to as MS1 and MS2.

Our stimuli consist of bi-syllabic non-words. The /t,s,ʃ,θ/ intervocalic consonants are inserted in the following environment: /fɒ_ə/. We used non-words so that listeners would not be influenced by the lexical properties of the stimuli. We used the same phonetic environment as we did in our pilot palatographic study, which was originally designed to avoid the presence of lingual consonants apart from the consonants under study /t,s,ʃ,θ/.

The elicitation of the stimuli was divided into two phases: a training phase and a production phase. The training phase consists in associating non-words to pictures, in order to prevent speakers from focusing on the orthography of the non-words and/or have them focus on the pictures rather than on their pronunciation. The non-words are associated to simple shapes and colours: "fotter" (/fɒtə/), "fosser" (/fɒsə/), "fosher" (/fɒʃə/) and "fother" (/fɒθə/) were associated to a red disc, a green triangle, a blue square, and a black star respectively.

Speakers sat in front of a computer screen and learned to associate non-words to the pictures presented on slides. The 32 first slides include both the written non-words and the corresponding pictures (*ie.* 8 slides per non-word). The next 32 slides include pictures only. During the training phase, all slides are presented in random order. The production phase consists in saying out loud the words associated with the pictures. Speakers are presented with 64 slides with pictures only in random order (*ie.* 16 slides per non-word).

We used a AKG C520 head-mounted microphone. The signal was pre-amplified using an external USB-Pre device. Recordings were acquired onto a PC running Audacity as WAV files at 44.1 kHz 16 bit and were made in a quiet room at the University of Canterbury. The experimenter provided the speakers with an information sheet and also gave instructions orally.

Manipulation of the stimuli

Stimuli for the current perception experiment are obtained according to the following procedure. From the elicited corpus of non-words, we select one /fɒtə/ token (with full frication of /t/), one /fɒsə/ token, one /fɒʃə/ token and one /fɒθə/ token per speaker. We measure the duration of the aperiodic noise for each intervocalic fricative and remove portions of the signal at the middle by increments of 10ms until the duration of the frication noise is equal to 20ms. Portions of the signal are removed between zero crossings. The onset of frication noise is located where the damped pseudo-periodic signal corresponding to /ɒ/ has the smallest amplitude. The offset of frication noise is located where the damped aperiodic noise has the smallest amplitude. Duration of the vowels /ɒ/ and /ə/ is normalised across all tokens for each speaker by identifying the shortest ones and then removing portions of the signal in the stable parts of the longer vowels to match the duration of the short ones. Portions of the signal are removed between zero crossings. The onset of /ɒ/ is located at the onset of glottal vibra-

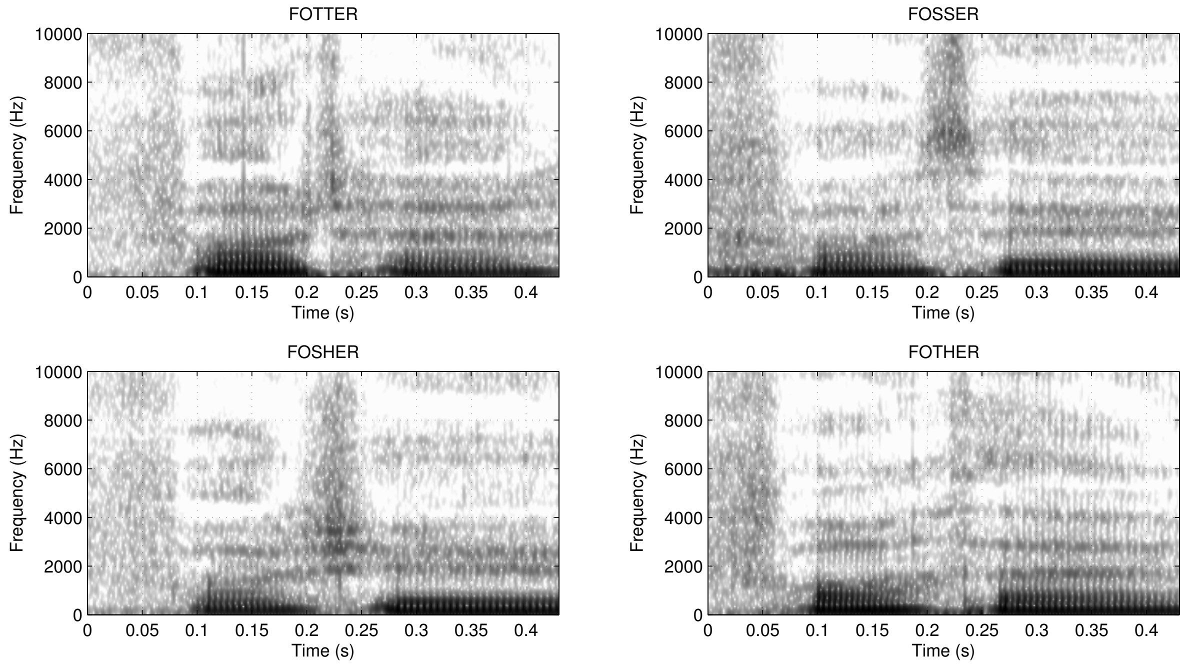

Figure 3.4 – Spectrograms of four test stimuli (DFN=70ms)

tion. The offset of /ɒ/ is located where the damped pseudo-periodic signal had the smallest amplitude. The onset of /ə/ was located where the damped aperiodic noise associated with the intervocalic consonant has the smallest amplitude. The offset of /ə/ is located where the speech signal is not visible anymore on the waveform. As an example, figure 3.4 shows the spectrograms of four stimuli for which the duration of frication noise (DFN hereafter) is set to 70ms.

One advantage of normalising vowels in this way is to preserve formant transitions, which contains cues to the place of articulation of the consonants that they surround. However we acknowledge that altering vowel length might also affect the perception of those consonants to some extent (see discussion).

Description of the forced-choice perception experiment

This perception experiment is a forced-choice identification task, run using a PRAAT script. Listeners sit in front of a computer screen displaying five boxes labelled FOTTER, FOSSER, FOSHER, FOTHER, and ELSE[3]. Participants are presented with one stimulus per trial through headphones and their task is to click on the box that corresponded best to what they heard. Each stimulus is played twice. In total, 195 stimuli were played in random order to each

[3] Participants are given the choice to click on the the ELSE box when they are not sure as to how they should classify a particular stimulus or if they did not hear a stimulus well. We acknowledge that the use of the word ELSE was not the most appropriate choice. However we could not choose the word OTHER because of the medial fricative that word contains.

Group	Listener label	Number of participants
Younger Pakeha Females (18 to 30 years old)	YFL	13
Older Pakeha Females (over 40 years old)	OFL	10
Younger Pakeha Males (18 to 30 years old)	YML	1
Older Pakeha Males (over 40 years old)	OML	4
Total		**28**

Table 3.2 – Constitution and labelling of the listeners

participant.

The experimenter provided the participants with an information sheet and also gave instructions orally. This experiment was reviewed and approved by the University of Canterbury Ethics Committee.

Listeners

There are very well established differences in speech perception due to ageing in terms of hearing loss and cognition so that older listeners perform less accurately and slower than younger listeners (*eg.* Pichora-Fuller and Souza, 2003; Janse, 2009; Adank and Janse, 2010). We recruited 14 "young" listeners aged between 18 and 30 years old and 14 older listeners over 40 years old. All listeners are Pakeha. Table 3.2 presents additional information on our listeners. We investigate a possible interaction between age and perception and recruited an equal number of younger and older speakers for this experiment. Listeners are referred to as YFL1 (Younger Female Listener 1) or OML2 (Older Male Listener), *etc.*

3.3.4 Results

Overview of the data

Percentages of responses for each stimulus as a function of DFN for all listeners are plotted on figure 3.5. Solid thick lines correspond to clicks on the FOSSER box (stimuli perceived as /s/); dashed lines correspond to clicks on the FOSHER box (stimuli perceived as /ʃ/); dotted lines correspond to clicks on the FOTHER box (stimuli perceived as /θ/); dash-dot lines correspond to clicks on the FOTTER box (stimuli perceived as /t/); solid thin lines correspond to clicks

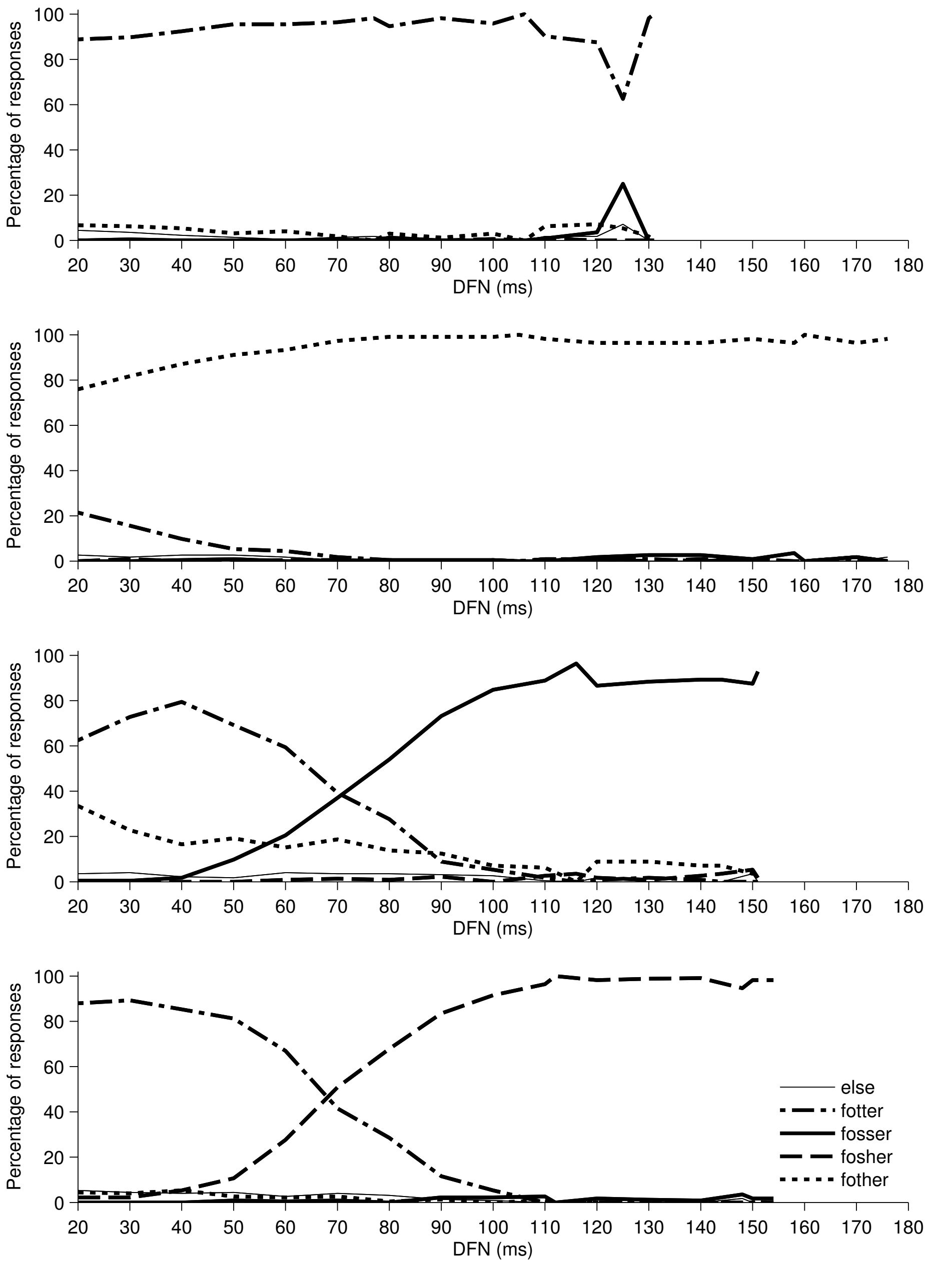

Figure 3.5 – Percentages of responses for /t/ stimuli (first row), /θ/ stimuli (second row), /s/ stimuli (third row) and /ʃ/ stimuli (fourth row)

on the ELSE box. The first and second rows of the figure show that, overall, the perception of /t/ and /θ/ remains quite stable throughout the experiment (*ie.* their perception is not much affected by a reduction in DFN). The first row also shows that /t/ can be mistaken as /s/ between 120ms and 130ms. This provides evidence for the fact that /t/ is closer to /s/ than it is to /ʃ/, which will also be confirmed by subsequent analyses. Also the second row shows that /θ/ can also notably mistaken for /t/ when its duration is less than 30ms. The third and fourth rows show that, as the DFN decreases, /s/ and /ʃ/ are progressively perceived as /t/, or rather, we should say that listeners experienced a percept close to /t/[4]. Finally, when the DFN ranges between 20ms to 70ms, /s/ can be mistaken as /θ/ in a proportion of about 20 percent. Taken together these results show that the perception of /t/ and /θ/ remains relatively stable and that there is a shift towards a percept close to /t/ for /s/ and /ʃ/ as the duration of frication noise decreases. Therefore we now turn to perception of /s/ and /ʃ/ for the remainder of our analyses.

If we look more closely at figure 3.5 we can see that perception shifts for /s/ and /ʃ/ happen at different times: they are located at about 70ms for /s/ and fall in between 60ms and 70ms for /ʃ/. Therefore listeners might have an earlier crossover boundary for /ʃ/ than for /s/. One way of testing this observation was to calculate crossover estimates using linear interpolation. Figure 3.6 exemplifies how we proceeded: for listener YFL7 the crossover between /s/ and /t/ is estimated at 67ms. Note that this method yielded "fictive" values because they not necessarily multiples of ten (whereas DFN increments are multiples of ten). We simply used this method in an exploratory way to identify potential trends in the data.

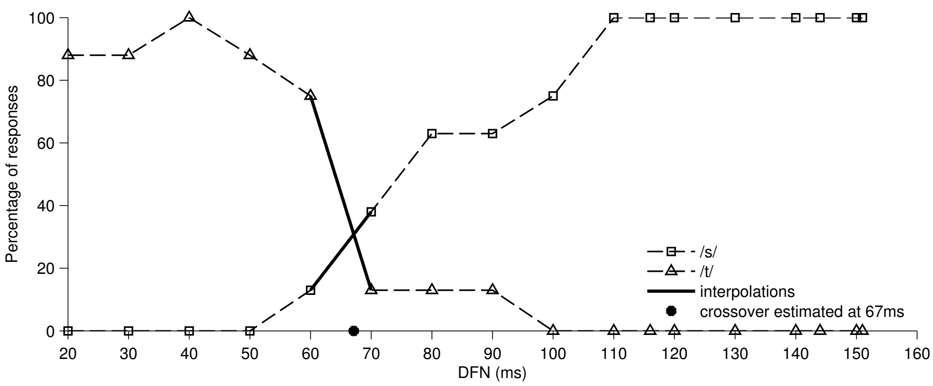

Figure 3.6 – Percentages of responses for /s/ and /ʃ/ as a function of the duration of frication noise for listener YFL7.

We applied this method for all listeners and for both /s/ and /ʃ/. We plotted the distributions of the estimates on a box-and-whisker diagram. Figure 3.7 shows the distributions of the

[4] It would be inexact to state that listeners experienced a true /t/ percept. Apart from durational aspects, the acoustic properties of /s/ and /ʃ/ are still distinct from /t/. As this experiment was a forced-choice categorisation task, listeners selected the FOTTER box, which means that /t/ would have been the closest percept.

estimates, which are significantly different (Wilcoxon non-paired test for equal medians, $p = 0.01$). The distribution of the cross-over estimates indicates that a shorter DFN is required to correctly identify /ʃ/ than /s/.

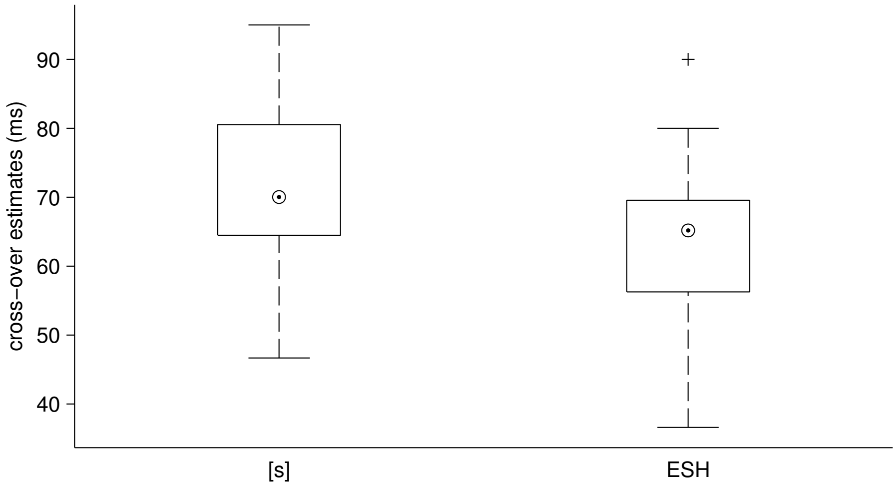

Figure 3.7 – Box-and-whisker diagram showing the crossover estimates for /s/ and /ʃ/ across all listeners

We plotted /s/ and /ʃ/ responses for younger and older listeners on figure 3.8. This plot shows that a smaller amount of frication noise is needed to better identify /ʃ/ than /s/. It also shows that younger listeners identify consonants more accurately and faster than older speakers as the duration frication noise increases. In addition, it is worthwhile noting a dip in older speakers' responses for /s/. When DFN ranges between 110ms and 150ms they achieve between 75% and 85% correct identification, while they achieve 95% correct identification for /ʃ/.

We also plotted the behaviour of the listeners in response to each speaker on figure 3.9 for reference. Younger listeners consistently identify /s/ and /ʃ/ more accurately and faster than older speakers, regardless of the speaker that produced the stimuli.

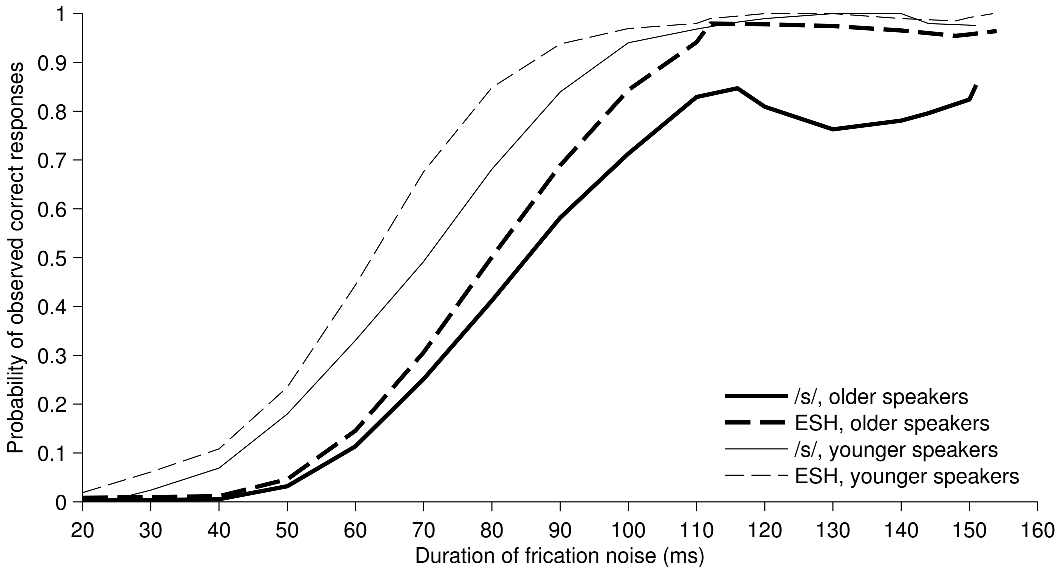

Figure 3.8 – Responses pooled across listeners' age and stimuli type (loess-fitted curves from raw data)

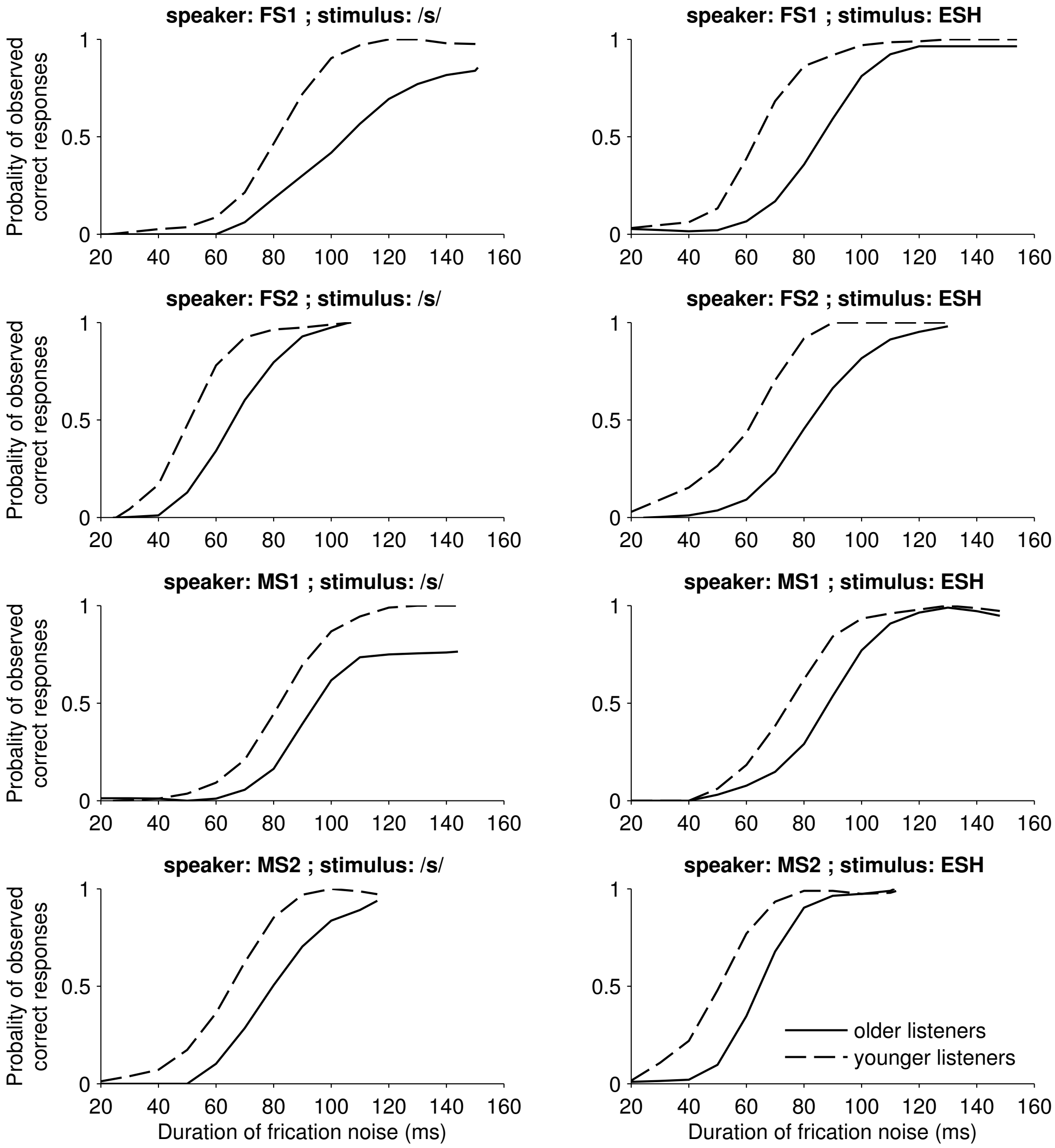

Figure 3.9 – Probability of observed correct responses by listener's age according speaker and stimulus type (loess-fitted curves from raw data)

Modelling of the data

We restricted our analysis to /s/ and /ʃ/ presented as stimuli to the participants. Data were submitted to a logit mixed model and there were 5712 binary responses in total. The dependent variable was a binary effect coded as "correct" or "incorrect" (*ie.* whether there is a match between the stimulus presented and the response given or not). We included 3 independent variables: DFN (a continuous variable centred to its mean: 80ms), listeners' age (a two levels factor where younger listeners were coded as Y and older listeners were coded as O) and the nature of the stimulus (a two levels factor with /s/ *versus* /ʃ/). A random effect structure with a random slope for each unique combination of speaker (4) and listener (28) was used. we obtained to following output:

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	2.4360040	0.2766504	8.805	$<$0.0001
DFN	0.1362303	0.0082704	16.472	$<$0.0001
Age (O)	-2.2916825	0.3737647	-6.131	$<$0.0001
Stimulus /s/	-1.0070268	0.1862224	-5.408	$<$0.0001
DFN by Age (O)	0.0008868	0.0113809	0.078	0.94
DFN by Stimulus (/s/)	-0.0027876	0.0080816	-0.345	0.73
Age by Stimulus (/s/)	0.2050648	0.2387669	0.859	0.39
DFN by Age (O) by Stimulus (/s/)	-0.0318668	0.0102050	-3.123	$<$0.002

Table 3.3 – Model output

The model shows a significant 3 way interaction between age, stimulus type and DFN. This interaction is plotted in figure 3.10. The three significant main effects are clear in this plot. Accuracy of identification was greater at higher DFNs, it was higher for /ʃ/ than for /s/, and younger listeners were more accurate than older listeners. These all interact together, such that the slope for /s/, as heard by older listeners, is steeper than the other slopes. In other words, older listeners make more errors at higher durations for /s/ than younger listeners do.

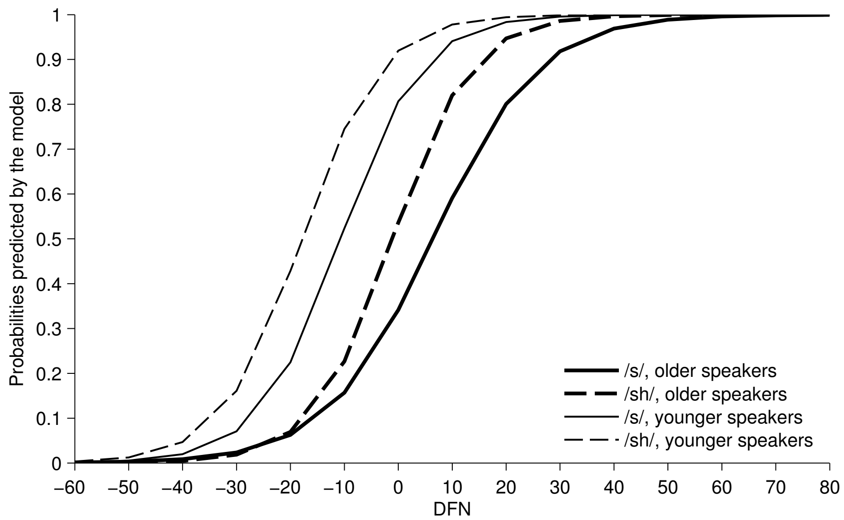

Figure 3.10 – Probabilities predicted by the model as a function of DFN (centred to 80ms)

3.4 Discussion

Our results have shown that correct identification of /s/ and /ʃ/ increases as DFN increases, in accordance with the literature (*eg.* Grimm (1966); Jongman (1989)). In particular, Grimm (1966) showed that the shortening of /s/ and /ʃ/ lead to the perception of /t/. However this is not the case for fricated /t/ and /θ/. In spite of some small variations observed on figure 3.5, their perception remained quite stable throughout our experiment. This is surprising given that Jongman's results in which correct identification of /θ/ decreases by 50% when the duration drops from full frication to 70ms. No previous studies have investigated the effect of duration on the perception of fricated /t/ because its identification is relatively recent (see Jones and Llamas (2003, 2008) descriptive work on this variant in Middlesbrough English and Dublin English). The fact that its perception is resistant to durational manipulations is not entirely surprising though. Shortening the duration of a fricative may result in a change in perceived manner of articulation from fricativeness to plosiveness (Grimm, 1966). But this might bear little relevance to NZE listeners because fricated /t/ and canonical /t/ are allophones. Therefore, regardless of the phonetic implementation, listeners might always have a percept of a /t/.

Our articulatory data indicated that /t/ is closer to /s/ than it is to /ʃ/ but /t/ is shorter in duration than the other two fricatives, which echoes the result of (Pandeli et al., 1997) on Irish English slit-t . Thus we questioned whether a shortened /s/ would be perceived as /t/ if the main difference between them was merely a question of duration. Indeed listeners did experience a percept close to /t/ as DFN decreased but this was also the case for /ʃ/. We also wondered whether /t/ would be perceptually closer to /s/ than to /ʃ/ and we believe that this experiment might provide an answer. Our model showed that for any given score and regardless of the age group, a longer DFN is required for /s/ to obtain the same /ʃ/ score. This may be because of similar and thus competing spectral information about /t/ and /s/ that only a sufficient amount of durational information can resolve. Note they are not spectrally identical, however, because /t/ can be very long without sounding like /s/. Conversely listeners might be able to tell /t/ and /ʃ/ apart more easily because they might not be as spectrally similar, therefore requiring less durational information to make an appropriate judgement. This is also supported by the fact a very long /t/ can be misheard as /s/. We may wonder whether the spectral information contained in the adjacent vowels may also contribute to this effect. Studies which showed that spectral cues contained in /s/ and /ʃ/ are critical for their identification also demonstrated that the information contained in the adjacent vowels might not play any role at all (Heinz and Stevens, 1961; Harris, 1958; LaRiviere et al., 1975). Such findings have been replicated by Whalen (1984), who combined the frication noise of /s/ and /ʃ/ with formant transitions associated with a different fricative and the percept remains that of the noise. Interestingly this slowed reaction times scores relative to those obtained in the condition with matching vocalic information. Whalen concluded that while judgements seem to be based on some cues, all cues taken together affect the perception of the stimuli. Further research is needed to extend the results of our experiment and to test the extent to which vowel information might contribute to spectral and durational cues.

An interesting finding in our experiment was the differences observed between age groups: for any given score, older listeners require a longer DFN than younger listeners to obtain the same score, regardless of whether the stimulus is /s/ or /ʃ/. Presbycusis, which is a deterioration of hearing ability starting from high-frequencies that occurs with ageing (Gates and Mills, 2005), might account for the patterns that we observed, see figure 3.11. A longer duration may be necessary to compensate for hearing loss. Also the slope differential that is typically observed for /s/ responses amongst older speakers may be because the spectrum of /s/ is characterised by higher frequency components relative to /ʃ/. Therefore, categorising /s/ stimuli correctly might be harder for older speakers. This physiological explanation could be tested in further research by simulating a hearing loss using a low-pass filter on younger speakers. If presbycusis is responsible for the difference that we observed between younger and older speakers, then younger speakers with a simulated hearing loss should give similar responses to the older speakers in this experiment.

Jones and McDougall (2009) investigated acoustic properties of fricated /t/ in Australian English. They found that /t/ was closer to /ʃ/ than it is to /s/ and that /t/ was shorter in duration than the other fricatives. However they did not include in their research a perceptual assessment of the /t/–/ʃ/ contrast. The methodology that we presented here could be used in Australian English. If /t/ is perceptually closer to /ʃ/ then we would expect that recognition scores would be higher for /ʃ/ than for /s/ when DFN decreases. In other words, we should observe the reverse tendencies between NZE and Australian English.

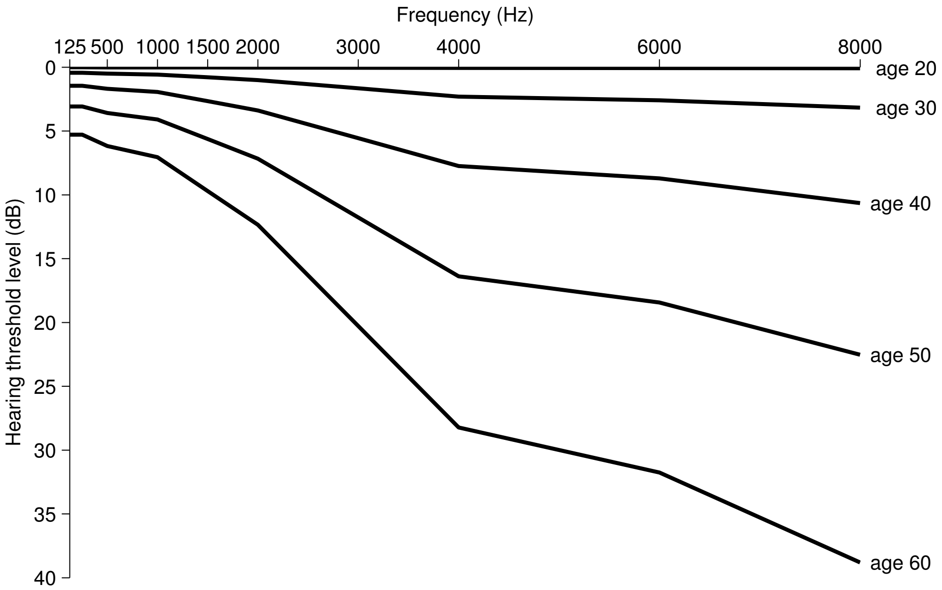

Figure 3.11 – Audiograms showing the effect of presbycusis. These audiograms plot the median values for male speakers according to the ISO standard 7029 (2000). Values are plotted for populations at age 20, 30, 40, 50 and 60.

Chapter 4

Triggering articulatory accommodation in a shadowing task, an EPG study

4.1 Introduction

Honorof et al. (2011) demonstrated that phonetic convergence can occur at the articulatory level using electromagnetometry. They asked an American English model speaker to produce exaggerated distinctions between clear [l] and dark [ɫ] in syllable-initial position of V.CV non-words[1]. Clear [l] is expected in this position while dark [ɫ] is not and is supposed to occur in coda position in American English (Browman and Goldstein, 1995; Scobbie and Pouplier, 2010). Participants were asked to shadow, without being instructed to imitate, the model speaker's speech. They reproduced to varying extents some aspects of the model speaker's production for both clear [l] and dark [ɫ] tokens. The latter qualified as "phonologically atypical" according to the authors since they appeared in an unusual syllable position.

They argue that the speakers reproduced "events that involve unrehearsed motor routines" and that the speakers "recovered the source of the sound without previous experience producing it" (Honorof et al., 2011, 36). However the dark variant is an attested variant within the speech of the speakers. It is certainly atypical in syllable-initial position but the motor routines have been extensively practised in coda position.

Our previous experiment in chapter 3 showed that when the frication noise of intervocalic /s/ or /ʃ/ is sufficiently short, listeners can have a percept closest to /t/. However, despite the fact that listeners have a percept that is close to /t/ the perceptual object is still different from a genuine /t/. Shadowing a shortened /s/ or /ʃ/ presents the advantage of exposing speakers to stimuli that they truly never have experienced before and which is still related

[1] The model speaker produced "(a) [l] with less backing and less lowering of the tongue body into the oropharynx than typical English for his [l] and, conversely,(b) a reduced apical constriction for [ɫ]" (Honorof et al., 2011, 27).

to an existing phoneme category at the allophonic level. Therefore it provides a means to investigate some of our research questions as stated in the introduction to this dissertation: if a speaker can learn to produce an unusual variant, would this ability influence the production of other allophones of the underlying phoneme? And if so, will the pre-existing allophones be produced more like the unusual variant?

Also, our previous perception experiment replicated findings by Grimm (1966) that manipulating the duration of /s/ or /ʃ/ creates a perceptual shift. While we are being careful by saying that listeners have a percept closest to /t/, Grimm argues that listeners experienced a perception shift from fricativeness to plosiveness. If listeners do perceive a plosive, then a shortened /s/ or /ʃ/ should be produced as a stop. However if shadowing a shortened /s/ or /ʃ/ lead to the production of a fricative, then it would mean that speakers are sensitive to the manner of articulation of the shadowed target.

In summary, we will address the following question in this chapter: would shadowing a shortened /s/ or /ʃ/ in medial position lead to the production of a speech sound which is indistinguishable from a shadowed genuine /t/ or would it contain properties (duration, place, and manner of articulation) that would more closely resemble the shadowed target?

To answer our question we set up an experiment using electropalatography (henceforth EPG) combined with simultaneous audio recordings. EPG is an articulatory technique that allows the recording of tongue-palate contacts over the course of time. We measured the Articulatory Centre of Gravity (henceforth ACoG) on the EPG signal, which has been used in numerous studies to locate the place of articulation of lingual consonants on the antero-posterior axis of the palate (*eg.* Hardcastle et al., 1991; Goldstein et al., 1994; Fuchs and Koenig, 2009).

First we present a pilot experiment conducted on one speaker. Then we present a second experiment based on 3 more speakers. Both experiments follow the same methodology essentially. Three CVCV non-words (/fɒtə/, /fɒsə/ and /fɒʃə/) were produced in repetition by each (they were the same non-words as the ones in our previous perception experiment). One repetition of /fɒtə/, /fɒsə/ and /fɒʃə/ was selected from the speakers' base production. The duration of /t/ in /fɒtə/ was not manipulated. Two sets of stimuli were derived from /fɒsə/ and /fɒʃə/: fricatives with their full duration (henceforth called long /s/ and long /ʃ/) and fricatives with a duration set to 50ms (henceforth called short /s/ and short /ʃ/). Speakers were asked to shadow multiple repetitions of the stimuli.

4.2 Pilot experiment

4.2.1 Method

Speaker

One Australian English female speaker (named speaker F1) participated in the experiment. They were the only speaker to have an EPG palate available at the University of Canterbury.

Elicitation of stimuli

Our stimuli consist of bi-syllabic non-words. The /t,s,ʃ,θ/ intervocalic consonants are inserted in the following environment: /fɒ_ə/, as we did in our previous perception experiment (chapter 3).

The non-words `FOTTER` (/fɒtə/), `FOSSER` (/fɒsə/) and `FOSHER` (/fɒʃə/) were prompted on a computer screen using Articulate Assistant (Articulate Instruments Ltd, 2010a). The speaker wore an acrylic palate with 62 electrodes. The EPG signal was acquired at 100Hz and simultaneous audio recordings were made at 44.1kHz using the WinEPG system (Articulate Instruments Ltd, 2010b).

The speaker was asked to repeat the non-words `FOTTER`, `FOSSER`, `FOSSER` in this order 10 times. One item of /fɒtə/, /fɒsə/ and /fɒʃə/ were extracted: for each group of lingual consonants we selected tokens for which the articulatory CoG at the point of maximum contacts was equal or the closest to the median value. We applied the same normalisation method for vowels as in the previous perception experiment. The duration of /ɒ/ and /ə/ was normalised across all tokens by identifying the shortest ones and then removing portions of the signal in the stable parts of the longer vowels to match the duration of the short ones. Portions of the signal were removed between zero crossings. The onset of /ɒ/ was located at the onset of glottal vibration. The offset of /ɒ/ was located where the damped pseudo-periodic signal had the smallest amplitude. The onset of /ə/ was located where the damped aperiodic noise associated with the intervocalic consonant has the smallest amplitude. The offset of /ə/ was located where the speech signal is not visible any more on the waveform.

The duration of /t/ in /fɒtə/ was not manipulated. Two sets of stimuli were derived from /fɒsə/ and /fɒʃə/ : fricatives with their full duration (henceforth called long /s/ and long /ʃ/) and fricatives with a 50ms duration (henceforth called short /s/ and short /ʃ/). As we did in our previous experiment, we measured the duration of the aperiodic noise for each intervocalic fricative and remove portions of the signal at the middle to set their duration to 50ms. Portions of the signal were removed between zero crossings. The onset of frication noise was located where the damped pseudo-periodic signal corresponding to /ɒ/ has the smallest amplitude.

The offset of frication noise was located where the damped aperiodic noise has the smallest amplitude.

Once the stimuli were created, the experimenter asked the participant to perform an informal perception task. This was to ensure that the speaker consistently identify the words `FOSSER` and `FOSSER`the with short /s/ and /ʃ/ as `FOTTER`, which they did. Thus they had a percept closest to /t/.

The experimenter provided the speaker with an information sheet and also gave instructions orally. This experiment was reviewed and approved by the Ethics Committees at the University of Canterbury.

Production experiment

The speaker sat in front of a blank computer screen. They were first presented with the genuine /t/ stimuli 10 times in order to test the EPG equipment. These repetitions were kept in our subsequent analyses. Then they were presented with all the stimuli in random order. Stimuli were presented using headphones and the speaker was asked to repeat the words they heard, without instructing them to imitate. The EPG signal was acquired at 100Hz and simultaneous audio recordings were made at 44.1kHz using the WinEPG system (Articulate Instruments Ltd, 2010b).

Measures

For each lingual consonant that was produced, we measured the Articulatory Centre of Gravity (ACoG) at the point of maximum contacts using EMATOOLS (Nguyen, 2000) for MATLAB. Duration of the consonants was also measured on the acoustic waveform using PRAAT.

4.2.2 Predictions

<u>Prediction 1a:</u> As listeners have a percept closest to /t/ when they are exposed to shortened /s/ and /ʃ/ stimuli we expect our speaker to produce /t/ stops when shadowing short /s/ and /ʃ/ tokens according to Grimm (1966).

<u>Prediction 1b:</u> Alternatively the speaker might also alter their typical articulatory behaviour and converge towards the manner of articulation of the shadowed target. Thus they might produce fricated articulations according to the direct realist account.

<u>Prediction 2:</u> If our speaker is sensitive to the underlying articulatory cues present in the acoustic signal then we expect a change in the ACoG for those shadowed short /s/ and /ʃ/ realised as stops compared to shadowed genuine /t/ stops. More specifically we expect the

ACoG for shadowed short /s/ and /ʃ/ tokens to be lower than genuine /t/ stops since the places of articulation of genuine /s/ and /ʃ/ are more posterior.

Prediction 3: The duration of short /s/ and /ʃ/ is 50ms, which is shorter than the duration of baseline /t/ stops (mean: 0.139s, SD: 0.025s.). Thus we expect our speaker to produce shorter stops or a fricated /t/ when shadowing short /s/ and /ʃ/ stimuli than genuine /t/ stimuli.

4.2.3 Results

Table 4.1 shows the number of repetitions produced by speaker F1 after shadowing long /s/ and /ʃ/ tokens, short /s/ and /ʃ/ tokens and genuine /t/ tokens. Shadowed long /s/ and /ʃ/ tokens were always produced as fricatives. Short /s/ and genuine /t/ tokens were mostly produced as stops, and occasionally as fricatives. Short /ʃ/ tokens were always produced as fricatives. Auditorily those fricatives sounded very much like the fricated variant of /t/ that we discussed in NZE[2]. This articulatory behaviour is not typical of speaker F1: only one shadowed genuine /t/ out 69 was produced as a fricative. Thus the speaker altered their typical articulatory behaviour and converged towards the manner of articulation of the shadowed target for short /ʃ/ tokens and to a smaller extent for short /s/ tokens.

	produced as	
	fricatives	stops
long /s/	70	NA
long /ʃ/	70	NA
short /s/	7	63
short /ʃ/	39	0
genuine /t/	1	68

Table 4.1 – Speaker F1: number of repetitions.

Two-sided Wilcoxon rank sum tests were used to determine whether shadowed short /s/ and /ʃ/ tokens differed significantly from their long counterparts and if they also differed from shadowed genuine /t/ tokens. Tokens were analysed in terms of their Articulatory Centre of Gravity (ACoG) and duration. We compared short tokens produced as stops to genuine /t/ tokens produced as stops, and we compared short tokens produced as fricatives to genuine /t/ tokens produced as fricatives. In order run the tests a minimum of 8 samples per group is needed. Thus, for example, we were not able to compare short /s/ fricatives to genuine /t/ fricatives since only one genuine /t/ fricative was produced.

Figure 4.1 plots the ACoG and the duration for the group of consonants we were able to compare. Short /s/ stops are placed in the middle of the plot and are compared to genuine

[2] These tokens were further presented auditorily to trained phoneticians who classified as /t/-like rather than /s/-like or /ʃ/-like.

/t/ stops and long /s/ fricatives. In terms of ACoG, short /s/ stops are significantly different from genuine /t/ stops and long /s/ fricatives ($p < .01$ and $p < .001$ respectively). The distribution of short /s/ stops lies in between the two. In terms of acoustic duration, short /s/ stops are significantly shorter than genuine /t/ stops and long /s/ fricatives ($p < .001$).

These results show that speaker F1 produced more posterior stops when shadowing short /s/ tokens than genuine /t/ tokens. In other words the speaker diverged from the ACoG of genuine /t/ stops and converged towards the ACoG of long /s/ tokens. In terms of duration, speaker F1 produced shorter stops when shadowing short /s/ tokens than genuine /t/ tokens. Thus the speaker also converged towards the duration of short /s/ tokens. These articulatory and acoustic results provide evidence that a speaker can be sensitive to and converge towards articulations that are not within their typical realm of production.

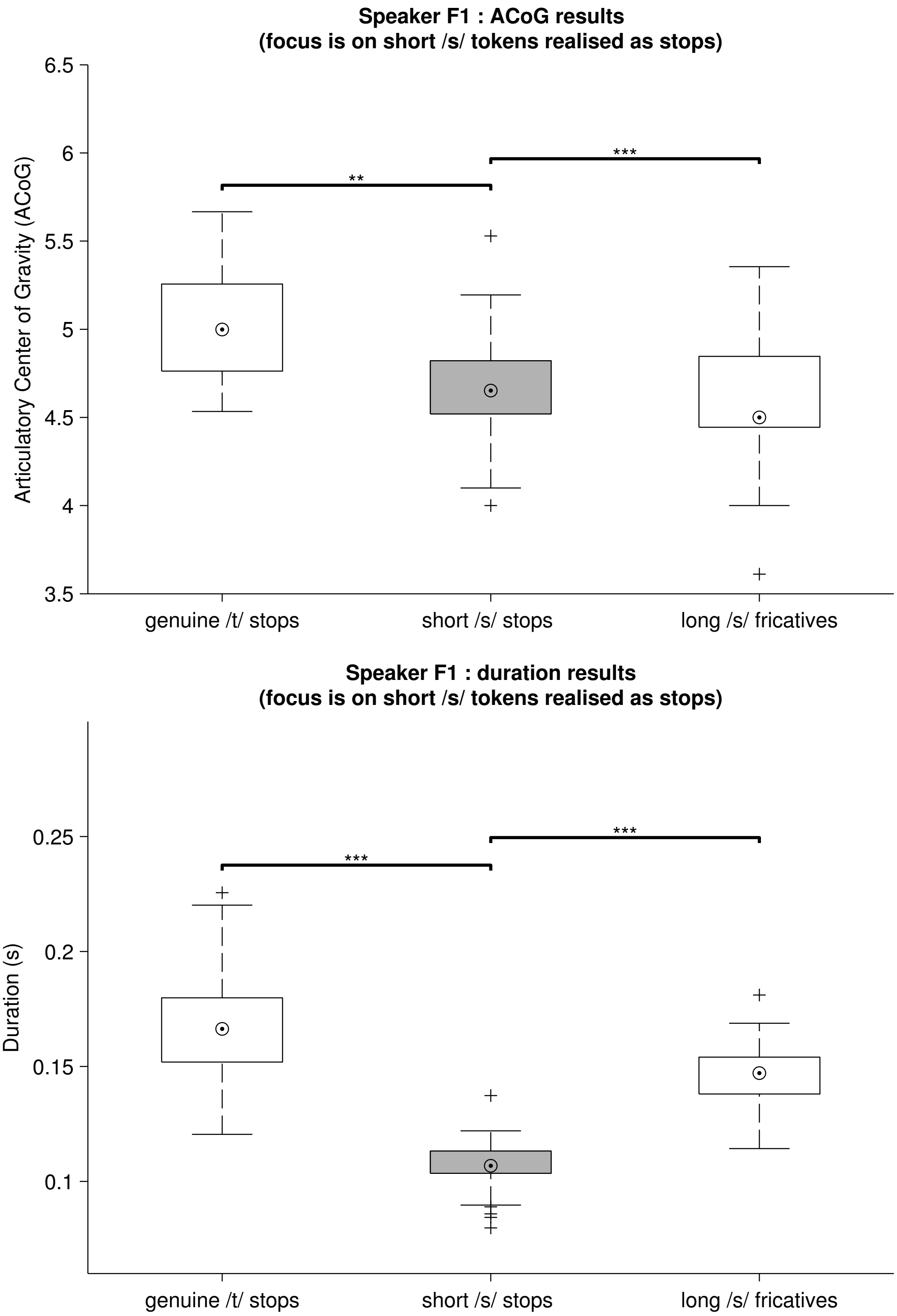

Figure 4.1 – Box and whisker plots focusing on shadowed short /s/ tokens realised as stops (dark box). Top panel: ACoG. Bottom panel: consonant duration. Significance levels were obtained using Wilcoxon rank sum tests and are reported with asterisks: $* = p < .05$, $** = p < .01$ and $*** = p < .001$. n.s. indicates non-significant differences.

4.3 Second experiment

The results of our previous pilot experiment supported our main hypothesis quite conclusively. Thus this led us to investigate it further by recruiting more participants and by gathering more data per participant.

Three speakers participated in this experiment. Essentially the methodological design remained the same except (i) for the number of baseline tokens that were collected (ii) for the order of presentation of the stimuli.

4.3.1 Method

Speakers

One female speaker (F2) and two male speakers (M1 and M2) participated in the experiment. The speakers were international, trained phoneticians who had an EPG palate available[3]. F2 is a Standard Scottish English speaker, M1 is an RP speaker and M2 is a General American speaker.

Stimuli

More baseline tokens were acquired in order to create the stimuli compared to the pilot experiment. Table 4.2 shows how many repetitions were obtained per speaker. The stimuli were created for each speaker exactly in the same way as we did for speaker F1 in our pilot experiment (see section 4.2.1).

	/t/	/s/	/ʃ/
F1 (pilot experiment)	10	9	10
F2	44	44	45
M1	47	47	45
M2	47	47	48

Table 4.2 – Number of repetitions obtained per speaker.

Production experiment

Once the stimuli were created, the participants took part in the production experiment. Speakers sat in front of a blank computer screen. They were presented with the stimuli using headphones were asked to repeat the words they heard (they shadowed their own voice, just as speaker F1 did). Speakers were not instructed to imitate the stimuli. The order of repetition of the stimuli was designed such that every target token occurred at equal probabilities.

The EPG signal was acquired at 100Hz and simultaneous audio recordings were made at 44.1kHz using the WinEPG system (Articulate Instruments Ltd, 2010b) at Queen Margaret

[3]It was not possible to recruit participants who spoke the same variety of English. A limited amount of EPG palates were available and the speakers were from different origins.

University. The experimenter provided the speakers with an information sheet and also gave instructions orally. This experiment was reviewed and approved by both Ethics Committees at the University of Canterbury and at Queen Margaret University.

Measures

For each lingual consonant that was produced, we measured the Articulatory Centre of Gravity (ACoG) at the point of maximum contacts using Articulate Assistant (Articulate Instruments Ltd, 2010a). Duration of the consonants was also measured on the acoustic waveform using PRAAT.

4.3.2 Predictions

Our predictions remains the same as in our pilot experiment.

<u>Prediction 1a:</u> As listeners have a percept closest to /t/ when they are exposed to shortened /s/ and /ʃ/ stimuli we expect our speaker to produce /t/ stops when shadowing short /s/ and /ʃ/ tokens according to Grimm (1966).

<u>Prediction 1b:</u> Alternatively the speaker might also alter their typical articulatory behaviour and converge towards the manner of articulation of the shadowed target. Thus they might produce fricated articulations according to the direct realist account.

<u>Prediction 2:</u> If our speaker is sensitive to the underlying articulatory cues present in the acoustic signal then we expect a change in the ACoG for those shadowed short /s/ and /ʃ/ realised as stops compared to shadowed genuine /t/ stops. More specifically we expect the ACoG for shadowed short /s/ and /ʃ/ tokens to be lower than genuine /t/ stops since the places of articulation of genuine /s/ and /ʃ/ are more posterior.

<u>Prediction 4:</u> The duration of short /s/ and /ʃ/ is 50ms, which is shorter than the duration of baseline /t/ stops (mean: 0.160s, SD: 0.037s)[4]. Thus we expect speakers to produce shorter stops or a fricated /t/ when shadowing short /s/ and /ʃ/ stimuli than genuine /t/ stimuli.

4.3.3 Results

Two-sided Wilcoxon rank sum tests were used to determine whether shadowed short /s/ and /ʃ/ tokens differed significantly from their long counterparts and if they also differed from shadowed genuine /t/ tokens in terms of ACoG and duration. We compared short tokens produced as stops to genuine /t/ tokens produced as stops, and we compared short tokens

[4]The following durations are broke down by speaker. F2 mean: 0.147s, SD: 0.013s. M1 mean: 0.188s, SD: 0.037s. M2 mean: 0.147s, SD: 0.035s.

produced as fricatives to genuine /t/ tokens produced as fricatives. However in order run the tests a minimum number of 8 tokens per population being compared is needed.

Results for speaker F2

Table 4.3 shows the number of repetitions produced by speaker F2 after shadowing long /s/ and /ʃ/ tokens, short /s/ and /ʃ/ tokens and genuine /t/ tokens. Shadowed long /s/ and /ʃ/ tokens were always produced as fricatives. Short /ʃ/ and genuine /t/ tokens were always produced as stops. Short /s/ tokens were almost always produced as stops (one token out of 25 was produced as a fricative). There were enough tokens available to compare shadowed short /s/ and /ʃ/ stops to their long counterparts and to genuine /t/ stops.

	produced as	
	fricatives	stops
long /s/	15	NA
long /ʃ/	15	NA
short /s/	1	24
short /ʃ/	0	23
genuine /t/	0	25

Table 4.3 – Speaker F2: number of repetitions.

Figure 4.2 and 4.3 plots the ACoG and the duration for the group of consonants we were able to compare. Short consonants are placed in the middle of the plot and are compared to genuine /t/ stops and their long fricative counterparts. In terms of ACoG, short /s/ and /ʃ/ stops are not significantly different from genuine /t/ stops but are significantly different from the long fricatives ($p < .001$). In terms of duration the same pattern is observed: short /s/ and /ʃ/ stops are not significantly different from genuine /t/ stops but are significantly different from the long fricatives ($p < .001$).

These results show that speaker F2 produced articulations that were within their typical realm of production when shadowing short /s/ and /ʃ/ tokens. In other words, we did not observe convergence towards the articulatory or acoustic properties of the short fricatives.

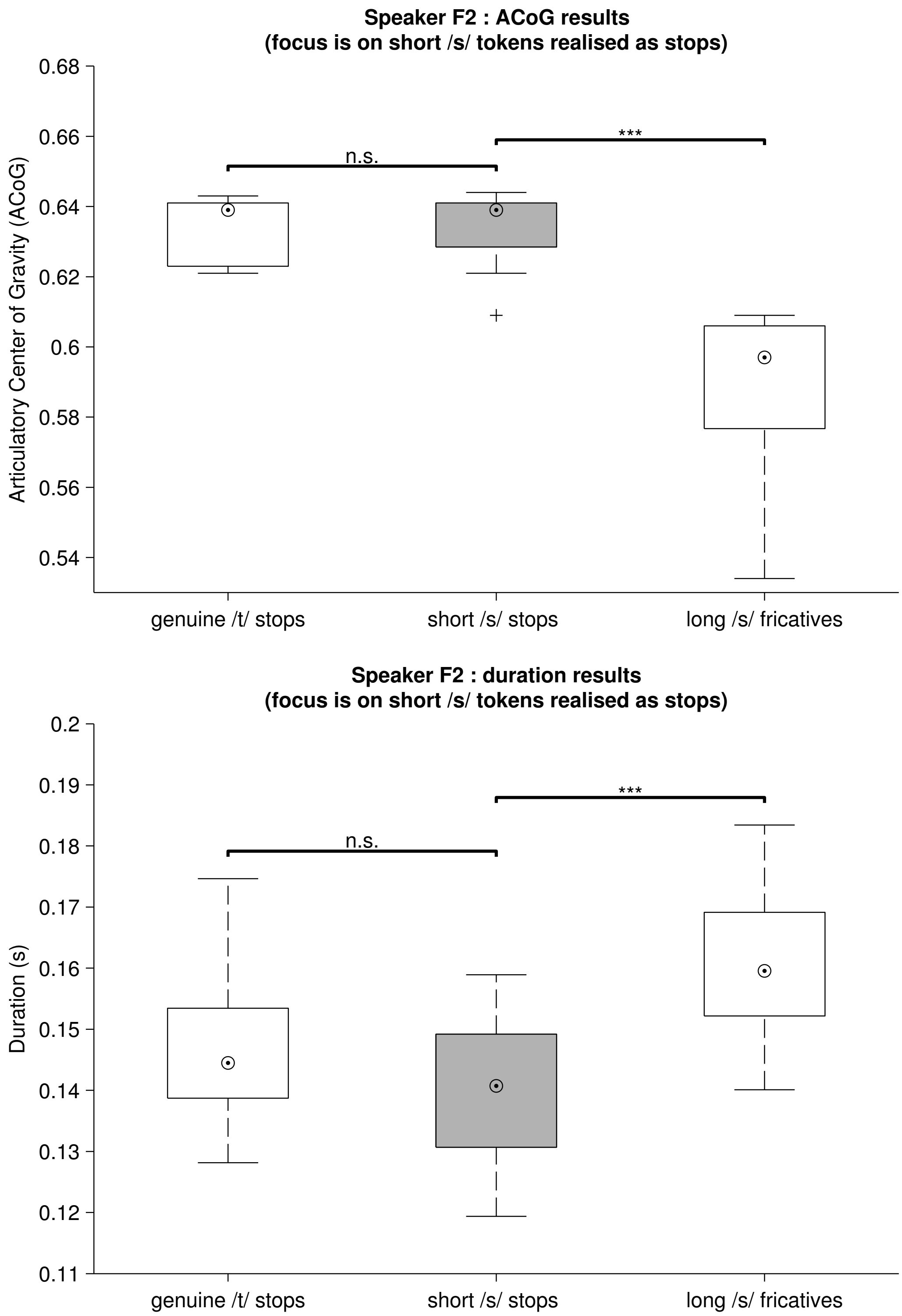

Figure 4.2 – Box and whisker plots focusing on shadowed short /s/ tokens realised as stops (dark box). Top panel: ACoG. Bottom panel: consonant duration. Significance levels were obtained using Wilcoxon rank sum tests and are reported with asterisks: $* = p < .05$, $** = p < .01$ and $*** = p < .001$. n.s. indicates non-significant differences.

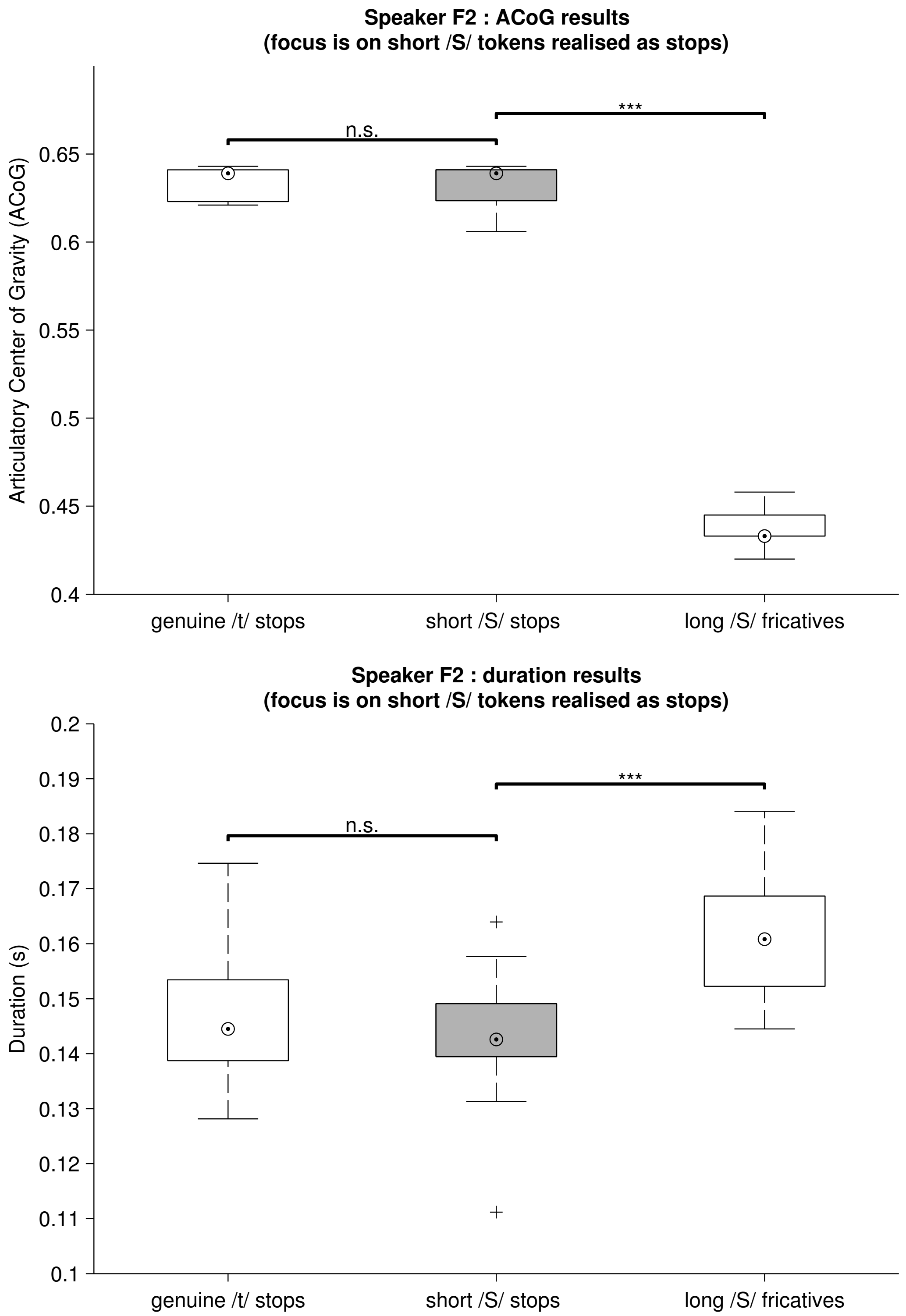

Figure 4.3 – Box and whisker plots focusing on shadowed short /ʃ/ tokens realised as stops (dark box). Top panel: ACoG. Bottom panel: consonant duration. Significance levels were obtained using Wilcoxon rank sum tests and are reported with asterisks: $* = p < .05$, $** = p < .01$ and $*** = p < .001$. n.s. indicates non-significant differences.

Results for speaker M1

Table 4.4 shows the number of repetitions produced by speaker M1 after shadowing long /s/ and /ʃ/ tokens, short /s/ and /ʃ/ tokens and genuine /t/ tokens. Shadowed long /s/ and /ʃ/ tokens were always produced as fricatives. Short /s/, short /ʃ/ and genuine /t/ tokens were mostly produced as stops with occasional fricated realisations.

	produced as	
	fricatives	stops
long /s/	18	NA
long /ʃ/	18	NA
short /s/	1	24
short /ʃ/	4	26
genuine /t/	1	29

Table 4.4 – Speaker M1: number of repetitions.

Figure 4.4 plots the ACoG and the duration of shadowed genuine /t/ stops, shadowed short /s/ stops and shadowed long /s/ fricatives. In terms of ACoG, short /s/ stops are not significantly different from genuine /t/ stops and long fricatives. In terms of duration short /s/ stops are not significantly different from genuine /t/ stops but are significantly different from the long fricatives ($p < .001$).

Figure 4.5 plots the ACoG and the duration of shadowed genuine /t/ stops, shadowed short /ʃ/ stops and shadowed long /ʃ/ fricatives. In terms of ACoG as well as duration short /ʃ/ stops are not significantly different from genuine /t/ stops and are significantly different from the long fricatives ($p < .001$).

These results show that while shadowed short /s/ tokens are not distinguishable from genuine /t/ stops or long alveolar fricatives in terms of ACoG as the speaker produced relatively fronted alveolar fricatives, shadowed short /s/ tokens retained the duration of genuine /t/ stops. As to shadowed short /ʃ/ tokens they are also similar to genuine /t/ stops both in terms of ACoG and duration. Overall this shows that speaker F2 maintained articulations that are within their typical realm of production when shadowing short consonants.

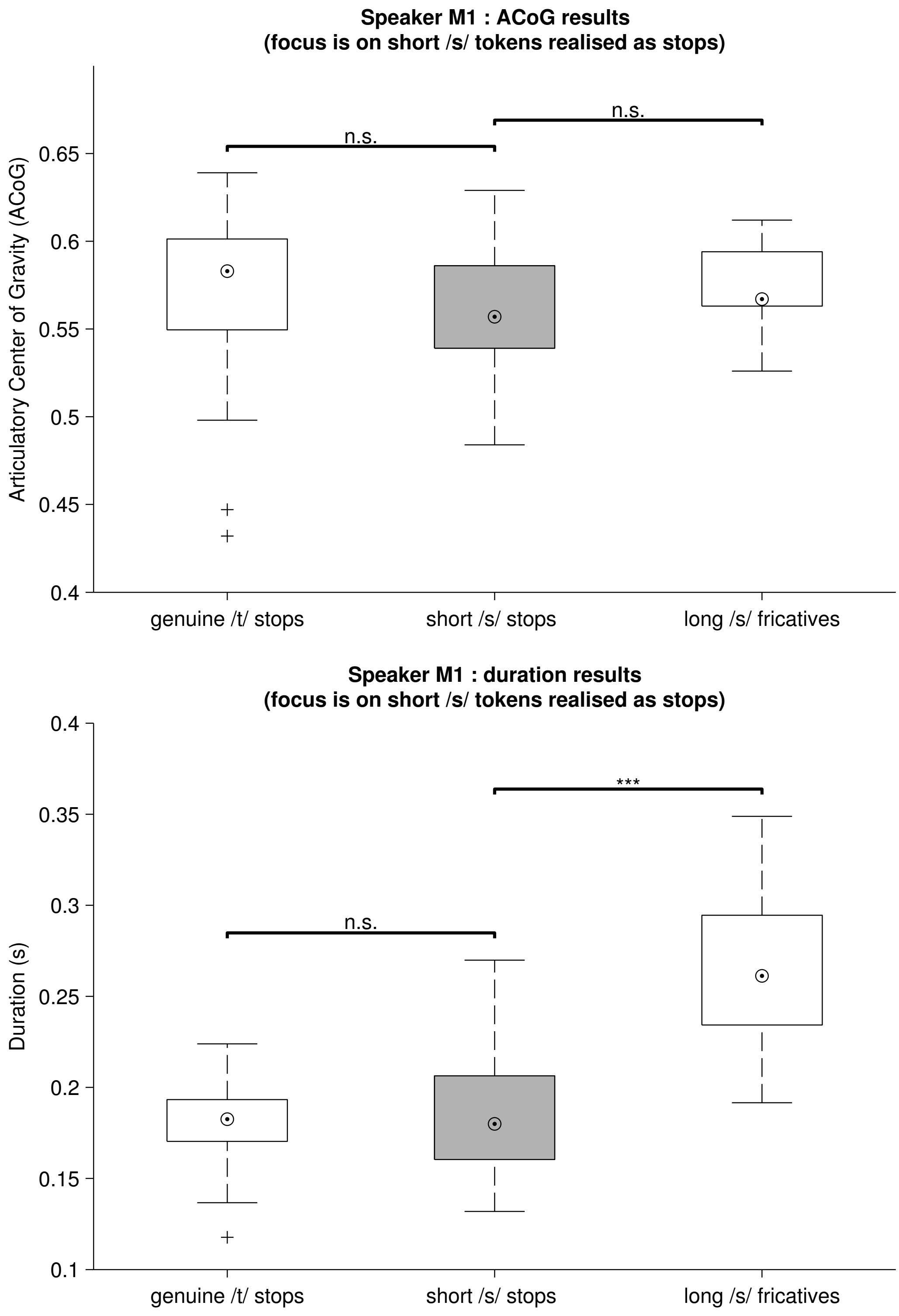

Figure 4.4 – Box and whisker plots focusing on shadowed short /s/ tokens realised as stops (dark box). Top panel: ACoG. Bottom panel: consonant duration. Significance levels were obtained using Wilcoxon rank sum tests and are reported with asterisks: $* = p < .05$, $** = p < .01$ and $*** = p < .001$. n.s. indicates non-significant differences.

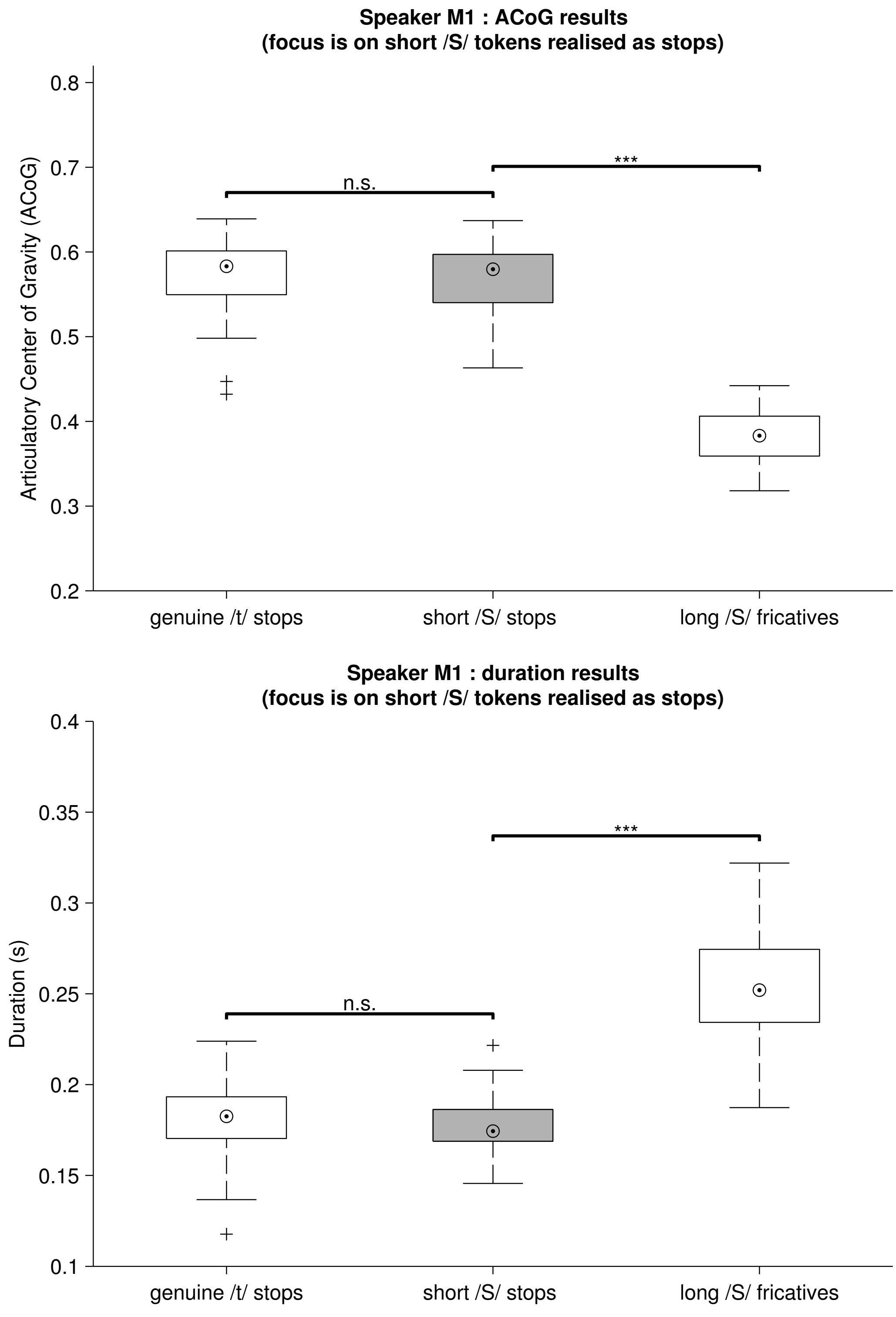

Figure 4.5 – Box and whisker plots focusing on shadowed short /ʃ/ tokens realised as stops (dark box). Top panel: ACoG. Bottom panel: consonant duration. Significance levels were obtained using Wilcoxon rank sum tests and are reported with asterisks: $* = p < .05$, $** = p < .01$ and $*** = p < .001$. n.s. indicates non-significant differences.

Results for speaker M2

Table 4.5 shows the number of repetitions produced by speaker M2 after shadowing long /s/ and /ʃ/ tokens, short /s/ and /ʃ/ tokens and genuine /t/ tokens. Shadowed genuine /t/ tokens were always produced as stops. Shadowed short fricatives were produced mostly as fricatives. As speaker M2 did not produce any fricatives when shadowing genuine /t/ stops, they altered their typical articulatory behaviour and converged towards the manner of articulation of the shadowed target.

	produced as	
	fricatives	stops
long /s/	27	NA
long /ʃ/	27	NA
short /s/	30	14
short /ʃ/	35	9
genuine /t/	0	44

Table 4.5 – Speaker M2: number of repetitions.

Figure 4.6 plots the ACoG and the duration of shadowed genuine /t/ stops, shadowed short /s/ stops and shadowed long /s/ fricatives. In terms of ACoG, short /s/ stops are significantly different from genuine /t/ stops and long /s/ fricatives ($p < .01$ and $p < .001$ respectively). The distribution of short /s/ stops lies in between the two. In terms of acoustic duration, short /s/ stops are significantly shorter than genuine /t/ stops and long /s/ fricatives ($p < .001$).

Figure 4.7 plots the ACoG and the duration of shadowed genuine /t/ stops, shadowed short /ʃ/ stops and shadowed long /ʃ/ fricatives. In terms of ACoG, short /ʃ/ stops are significantly different from genuine /t/ stops and long /ʃ/ fricatives ($p < .001$ respectively). The distribution of short /s/ stops lies in between the two. In terms of acoustic duration, short /s/ stops are significantly shorter than genuine /t/ stops and long /s/ fricatives ($p < .001$).

These results show that speaker M2 produced more posterior stops when shadowing short /s/ and /ʃ/ tokens than genuine /t/ tokens. In other words the speaker diverged from the ACoG of genuine /t/ stops and converged towards the ACoG of the long counterparts. In terms of duration, speaker M2 produced shorter stops when shadowing short tokens than genuine /t/ tokens. Note that short stimuli have a duration of 0.05s and the genuine /t/ stimuli have a duration of 0.14ms. Thus the speaker also converged towards the duration of short /s/ tokens.

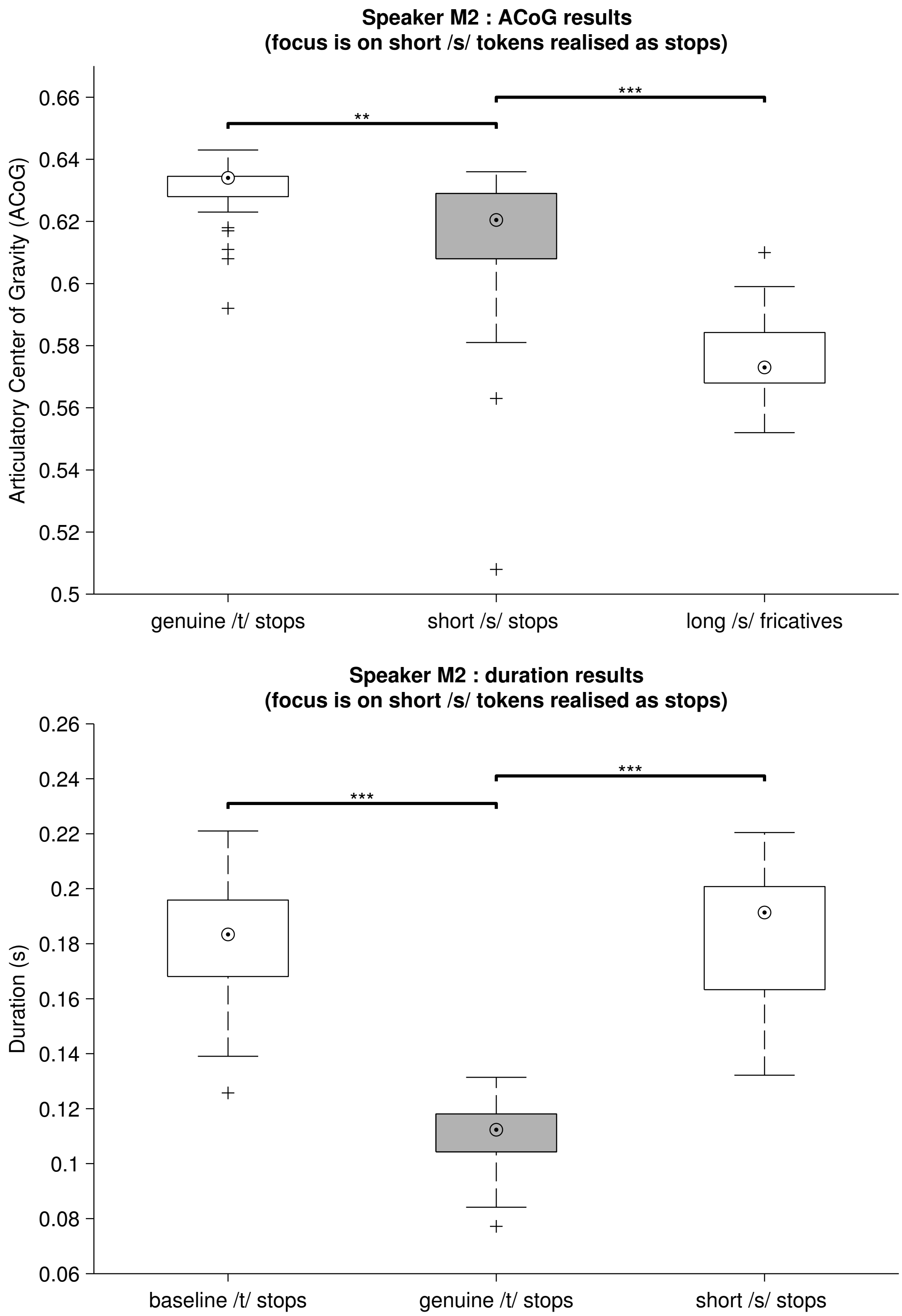

Figure 4.6 – Box and whisker plots focusing on shadowed short /s/ tokens realised as stops (dark box). Top panel: ACoG. Bottom panel: consonant duration. Significance levels were obtained using Wilcoxon rank sum tests and are reported with asterisks: $* = p < .05$, $** = p < .01$ and $*** = p < .001$. n.s. indicates non-significant differences.

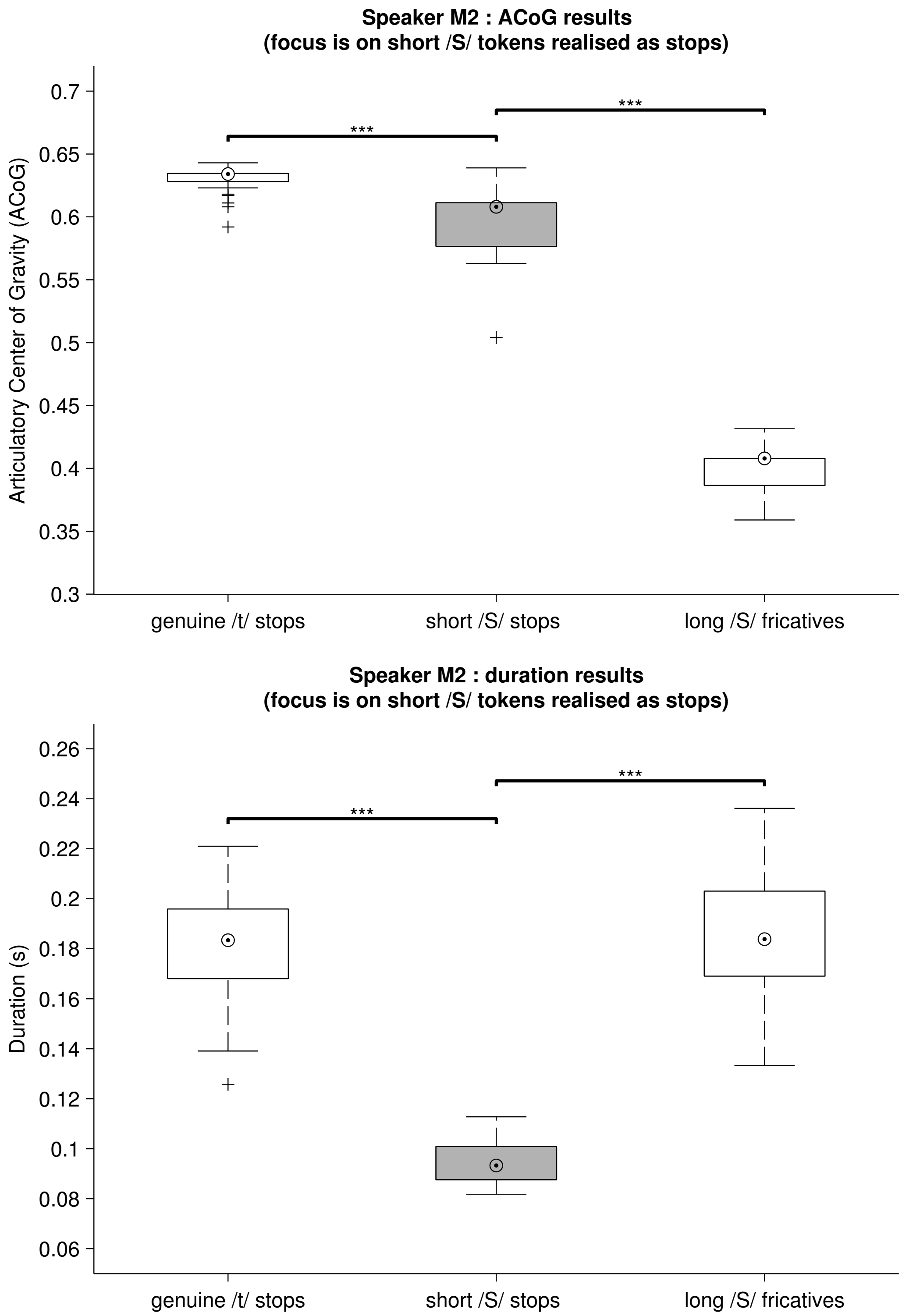

Figure 4.7 – Box and whisker plots focusing on shadowed short /ʃ/ tokens realised as stops (dark box). Top panel: ACoG. Bottom panel: consonant duration. Significance levels were obtained using Wilcoxon rank sum tests and are reported with asterisks: $* = p < .05$, $** = p < .01$ and $*** = p < .001$. n.s. indicates non-significant differences.

4.4 Summary of the results

For speaker F1 shadowing a short /s/ lead to the production of a stop whereas shadowing a short /ʃ/ lead to the production of a fricative. Shadowed short /s/ stops had a significantly lower ACoG and duration compared to shadowed genuine /t/ stops ($p < .01$ and $p < .001$ respectively).

For speaker F2 shadowing short fricatives mostly lead to the production of stops (one token only was realised as a fricative). There were no difference in ACoG and in duration between shadowed short /s/ stops and shadowed genuine /t/ stops. There were also no difference in ACoG and in duration between shadowed short /ʃ/ stops and shadowed genuine /t/ stops.

For speaker M1 shadowing short fricatives mostly lead to the production of stops. There were no difference in ACoG between shadowed short /s/ stops, shadowed genuine /t/ stops or shadowed long fricatives, and there was no difference in duration between shadowed short /s/ stops and shadowed genuine /t/ stops. There were no difference in ACoG and in duration between shadowed short /ʃ/ stops and shadowed genuine /t/ stops.

For speaker M2 shadowing a short fricatives mostly lead to the production of fricatives. Shadowed short /s/ stops had a significantly lower ACoG and duration compared to shadowed genuine /t/ stops ($p < .01$ and $p < .001$ respectively). Shadowed short /ʃ/ stops had significantly lower ACoG and duration compared to shadowed genuine /t/ stops ($p < .001$).

We concluded that both speakers F1 and M2 altered their typical articulatory behaviour as they converged towards the manner of articulation of shadowed targets. We also concluded that these speakers can produce articulations that are not within their typical realm of experience as they produced more posterior and shorter stops when shadowing short fricatives. On the other hand speakers F2 and M1 did not alter their articulatory behaviour when shadowing short fricatives, despite the fact that they are perceived and produced as /t/.

4.5 Discussion

Here we conducted an experiment designed to investigate imitation within medial position. We found that some participants can imitate a medial short fricative that they relate at some abstract level to the /t/ category. Not only can participants imitate the novel variant but it can also disrupt the production of their usual stops, so that they become more phonetically similar to the shadowed target, *ie.* they have a shorter duration and lowered ACoG. One could thus argue that presenting speakers with a novel variant can not only induce imitation but can also affect the phonetic distribution of other allophones of /t/ – at least for stops –. These results provide answers to some the research questions that we asked in our introduction. This also extends findings by Honorof et al. (2011) that speakers (i) can recover articulatory events from the acoustic signal without prior experience to them and (ii) reproduce unrehearsed motor routines.

Two speakers out of four did not imitate the sounds they were exposed to, or did not even reproduce the sounds at least partially. Thus imitation is selective and this has theoretical implications that will be discussed in our final chapter. As we will see these results pose a challenge to the direct realist account that Honorof et al. (2011) used to explain their results.

In the two next chapters we report experiments that are designed to investigate our second research paradigm, namely imitation across word positions. These experiments are also designed to further investigate imitation within a given word position and to extend the findings from this experiment.

Chapter 5

Positional transfer and imitation-driven allophony in New Zealand English

5.1 Introduction

Nielsen (2008, 2011) demonstrated that phonetic imitation can be generalised to new instances of the target phoneme /p/ as well as to the new segment /k/ by presenting English-speaking participants with extended VOT of /p/ sounds. Nielsen concluded that speakers can generalise to new instances which share the same phoneme or the same feature. In her study, she trained participants on initial /p/. Recent findings in speech perception by Jesse and McQueen (2011) further showed that listeners trained on a phoneme in final position can generalise to the same phoneme in initial position. However it is not known whether imitation effects in speech production could generalise to other positions in the word. If participants are trained on a phoneme in initial position, for instance, will the imitation effects extend to other positions on the word?

In this experiment we tested whether shadowing stops with extended closure duration[1] in a given word position lead to an imitation effect, and whether this effect extended to non-shadowed items in the same word position. We also tested whether shadowing stops with extended closure in a given position lead to an imitation effect which extended to a different word position.

Also, our EPG experiment showed that two participants increased the amount of fricatives that they produced – and that we categorised as allophones of /t/ – after shadowing short /s/ and /ʃ/. Those speakers increased the use of fricatives and used fewer stops, the latter

[1]Note that unlike Nielsen, we manipulated stop closure duration instead of VOT. The reason is that stop closure duration can be manipulated in all word positions in NZE, whereas long VOT is a feature that is typically associated with aspirated stops. Aspirated stops are known to aoccur in initial but they do rarely in medial and final positions in NZE, as presented in our review of literature on /t/ allophony in section 1.6.

being their typical variant in baseline. Presenting participants with stops, the duration of which has been extended, could make these stops more noticeable auditorily – especially in medial and final position as stops are in variation with other allophones – and therefore engage participants in producing more stops than other variants. This would also be in line with Clark's finding (2013) that speakers' production of a variant lead them to produce the variant more in subsequent productions. Thus, in this experiment we tested whether participants will produce more medial and final stops than they do compared to their baseline production after exposure to lengthened canonical stops. In other words, we tested whether imitation can drive allophonic selection.

5.2 Methodology

5.2.1 Stimuli

Stimuli words were chosen by querying the CELEX database (Baayen et al., 1993). We selected bisyllabic singular nouns with low frequency, defined as having a score lower than 500 on the "COBUILD" frequency scale[2]. Each word has a Strong/Weak stress pattern. A table showing the selected words can be found in the appendix B. Each word was embedded in the carrier phrase : a/an `WORD` as well. This carrier phrase was chosen because words can be embedded between two vowels, thus making the identification of initial and final stops fairly easy on waveforms and spectrograms. We selected:

- 30 words with initial /t/ used in the pre and post-test conditions. A subset of 20 words was used in the test condition for one group of speakers (group A)
- 30 words with medial /t/ used in the pre and post-test conditions. A subset of 20 words was used in the test condition for one group of speakers (group B)
- 30 words with final /t/ used in the pre and post-test conditions. A subset of 20 words was used in the test condition for one group of speakers (group C)
- 30 words with /d/ used in the pre and post-test conditions only.
- 20 filler words without /d/'s or /t/'s. They were only present in the test condition were designed to distract the speakers' attention from the manipulated stimuli.

[2]The frequency information in the CELEX English database was taken from the COBUILD corpus, which contained 17,900,000 words when it was first published. COBUILD is an acronym for Collins Birmingham University International Language Database.

5.2.2 Data acquisition from the model speaker

A NZE speaker was recruited in France. They were prompted with the stimuli list using the Prorec software (Huckvale, 2003). They were asked to read the prompts and to repeat if they made or error or if they hesitated. A real time spectrogram was running in order to monitor and make sure than the NZE speaker would produce canonical stops. Recordings were acquired using a AKG-C520 head-mounted microphone connected to a Schure X2u XLR to USB signal adapter. The signal was acquired onto a laptop using Audacity and was sampled at 44.1 kHz. The experimenter provided the participant with an information sheet and also gave instructions orally.

5.2.3 Stimuli manipulation

The words previously recorded and containing initial, medial and final /t/ were manipulated to extend the duration of closure. Our intent was to create a very salient and long closure duration that the test participants would be able to notice auditorily and could potentially imitate without being explicitly instructed to. Our manipulation consisted of splicing "silence" at the mid part of consonantal closure at zero crossings. Ambient noise was actually added instead of pure silence as the recordings were not made in a sound-proof booth. Ambient noise was taken from a randomly selected portion of the signal corresponding to a pause made by the model speakers in between two words. 100ms were added to initial /t/ tokens and 80ms were added to medial and final /t/ tokens elicited by the NZE model speaker. After a listening test done amongst trained phoneticians it was concluded that adding 100ms to both medial and final /t/ tokens sounded too artificial in English, while adding 80 ms to initial /t/ did not make the extended closure salient enough.

5.2.4 Data acquisition from the test speakers

28 NZE speakers (23 females and 8 males), all aged between 18 and 30, took part in the experiment. Prior to the experiment, each participant was assigned to a group: A, B or C. In the pre and post-test conditions participants in all groups were prompted with all words using Prorec (see section 6.2.1). In the test condition, participants in group A had to shadow 20 initial /t/ words and 20 filler words, participants in group B had to shadow 20 medial /t/ words and 20 filler words and participants in group C had to shadow 20 final /t/ words and 20 filler words. Participants shadowed the speech of the model speaker. The stimuli were presented binaurally through headphones. Speakers were recorded using a AKG-C520 head-mounted microphone connected to a USB Pre sound card. The signal was acquired onto a laptop using Audacity and was sampled at 44.1 kHz. Recordings were made in a quiet room at the University of Canterbury. Speakers were recorded by a research assistant, a Japanese speaker

of English, hired through a grant generously offered to the author of the thesis by the School of Languages, Cultures and Linguistics at the University of Canterbury. The experimenters provided the participants with an information sheet and also gave instructions orally.

5.2.5 Measures

Durational measures

Four durational measures were computed from the speech of the test speakers. We measured the closure duration of the alveolar stops defined as the corresponding period of silence seen on the waveform. We measured the duration of fricated stops at the offset and onset of surrounding vowels. We measured the length of each carrier phrase to give us an indication of possible speech rate variation between pre-test, test and post-test conditions. Finally we calculated the ratio between the closure duration or the fricative duration and the duration of the whole carrier phrase (CPratio, hereafter).

Labelling of allophones in NZE

As outlined in Experiment 1, NZE has a great deal of allophonic variation. Using the spectrograms, allophones were classified into 4 main categories : stops, fricatives, glottal articulations and miscellaneous. Then they were further subdivided :

Alveolar stops: (see figure 5.1 for sample spectrograms)

- Canonical stops: at the offset of modal voicing, there is a full alveolar closure followed by a burst, aspiration noise and the onset of modal voicing.
- Laryngealised alveolar stops: unlike canonical stops, there is the presence of laryngealisation in the preceding vowel as reported by Docherty et al. (2006) in NZE.
- Pre-aspirated and post-aspirated stops: a spectral leakage can be observed at the offset of modal voicing, which resembles that of a glottal opening letting the air flow freely from the glottis onwards[3].

Fricatives: (see figure 5.2 for sample spectrograms)

[3]Very similar acoustic data coupled with additional EPG data were gathered in a separate study, which are not reported in this thesis. Preliminary results indicate that pre-aspiration of stops takes place at a phonetic level in the way described here. The sound source seems to be generated at the glottis as there does not seem to be any constriction of supra-glottal articulators, which otherwise would create friction noise. Therefore the air seems to flow freely from the glottis onwards. In a separate study conducted on NZE (not reported in this thesis) no significant statistical differences were found between the acoustic intensity taken at the mid-point of pre-aspiration noise and at the mid point of post-aspiration noise.

- Fricated /t/s: (phonologically) voiceless stops realised as fricatives, as described in more details in previous chapters.

- Laryngealised fricated /t/s: (phonologically) voiceless stops realised as fricatives with the presence of laryngealisation in the preceding vowel.

- Pre-aspirated and post-aspirated fricated /t/s: pre-aspiration occurs in the same way as described above for stops and it is followed by a fully fricated release. No burst can be seen on spectrograms. While there might appear to be closure, these speech sounds were included in the fricative category and are discussed in more detail in chapter 2 given their similarity with the pre-aspirated fricated /t/s reported by Jones and Llamas (2003).

Glottal articulations: (see figure 5.3 for sample spectrograms)

- Glottal stops are marked by a period a silence in the speech stream, generally accompanied by laryngealisation on the flanking vowels.

- Creaky voice: there is no marked period of silence but rather a continuous stream of irregular pulses from one vowel to the other.

Miscellaneous:

- Speech sounds classified in the miscellaneous category were taps, flaps, fully-voiced /t/ stops or ambiguous stops that could not be otherwise classified clearly into one of the above-described categories.

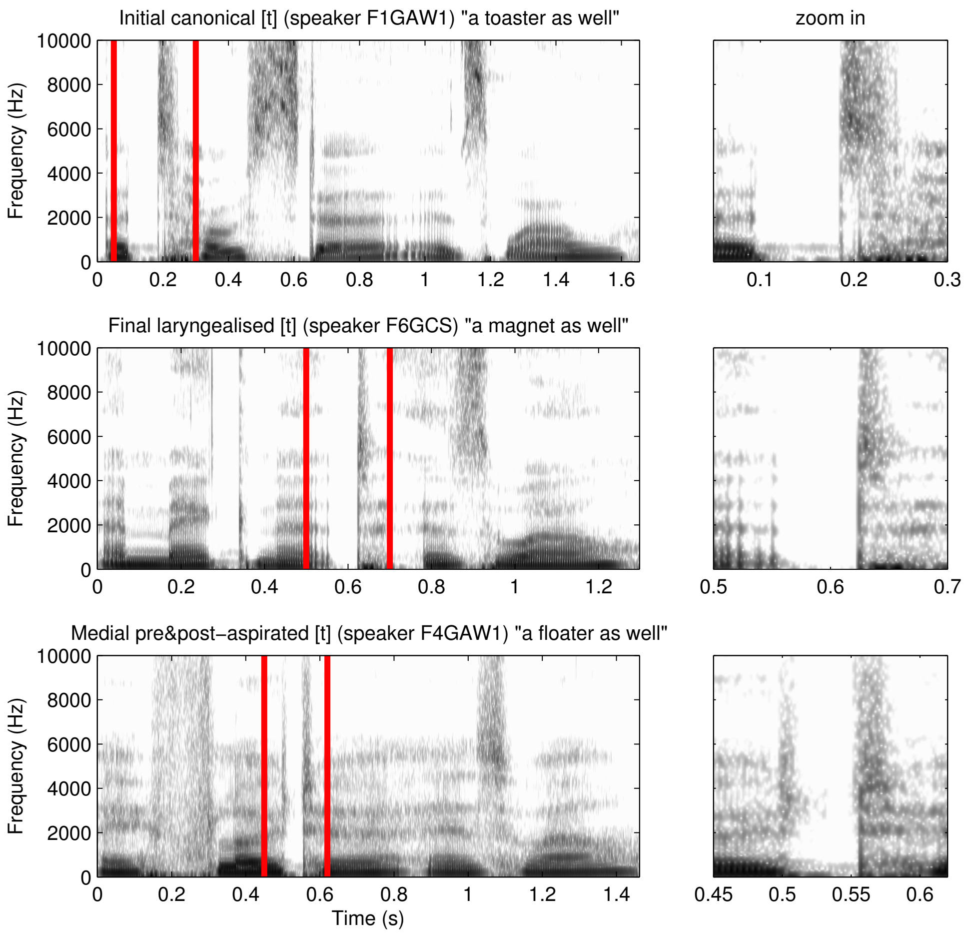

Figure 5.1 – Sample spectrograms of alveolar stops. Right: spectrograms of full phrases. Left: Spectrograms of a segment and portions of neighbouring phonetic context.

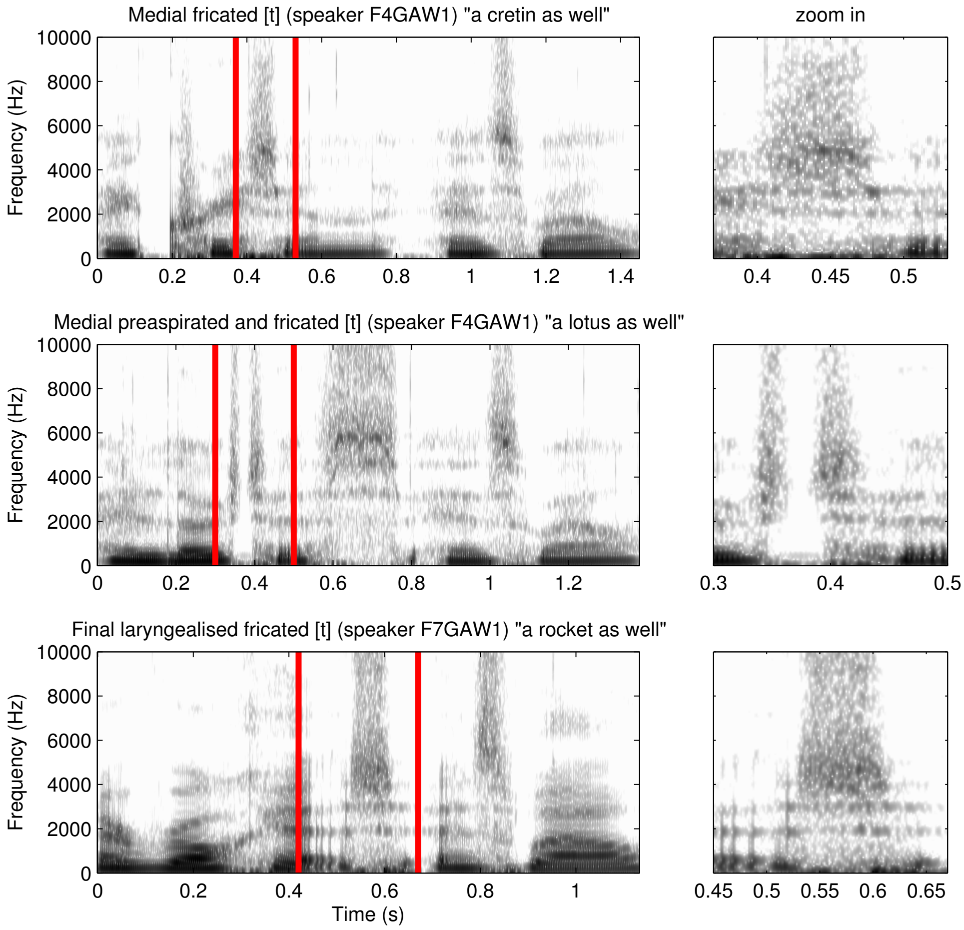

Figure 5.2 – Sample spectrograms of stops realised as fricatives. Right: spectrograms of full phrases. Left: Spectrograms of a segment and portions of neighbouring phonetic context.

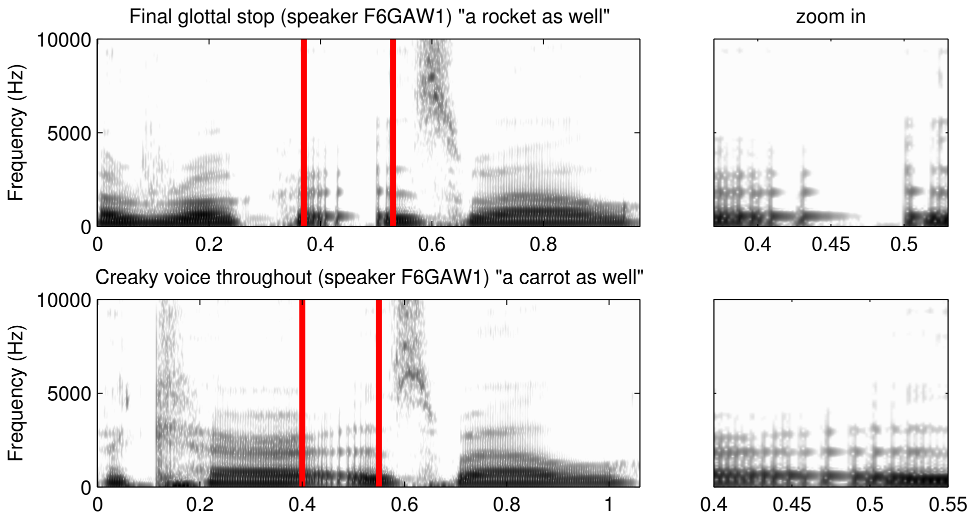

Figure 5.3 – Sample spectrograms of stops realised as glottal articulations. Right: spectrograms of full phrases. Left: Spectrograms of a segment and portions of neighbouring phonetic context.

5.2.6 Predictions

This pilot experiment investigates whether imitation effects can generalise to other positions in the word. If participants are trained on a phoneme in initial position, for instance, will the imitation effects extend to other positions on the word, or to a different phoneme?

Prediction 1: in order to test this question, we must demonstrate that our experimental design is able to generate spontaneous imitation effects. If this is the case, participants will increase the duration and/or CPratio of shadowed /t/ stops in the test condition.

Prediction 2: we also have to demonstrate that imitation effects, if any, should carry-over to /t/ tokens in the same word position. If this is the case, /t/ tokens that were shadowed will increase in terms of duration and/or CPratio in the post-test condition. This will provide evidence for persistent imitation effects across experimental conditions.

Prediction 3: /t/ items that were not-shadowed but in the same word position will increase in terms of duration and/or CPratio in the post-test condition. This will provide evidence for a generalisation to new /t/ items in the same word position.

Prediction 4: provided that predictions 1, 2 and 3 are realised, we can investigate whether participants trained in a given position will increase the duration and/or CPratio in the post-test condition of non-shadowed /t/ items in a different word position. If they do then we can conclude that imitation effects can generalise to other word positions.

Prediction 5: provided that predictions 1, 2 and 3 are realised, we can investigate whether participants trained in a given position will increase the duration and/or CPratio in the post-test condition of (non-shadowed) /d/ items in the same word position. If they do then we can conclude that imitation effects can generalise to /d/ words in the same position.

Prediction 6: provided that all the above predictions are realised, we can investigate whether participants trained in a given position will increase the duration and/or CPratio in the post-test condition of (non-shadowed) /d/ items in a different word position. If they do then we can conclude that imitation effects can generalise to /d/ words in a different position.

Prediction 7: Some participants in our EPG experiment produced speech sounds with acoustic and articulatory characteristics that resembled more closely the stimuli they were exposed to. In this experiment we also investigate whether participants will be influenced by the sounds they are exposed to. Our prediction is that they will produce more medial and final canonical stops than they do compared to their baseline production after exposure to lengthened canonical stops.

5.3 Results

We excluded misproductions and productions when participants hesitated. Statistical outliers in each of the datasets presented below were manually checked. We first present overall results then we move on to the results for each group of speakers.

5.3.1 Overall results

Comparison between the model speaker and the test speakers

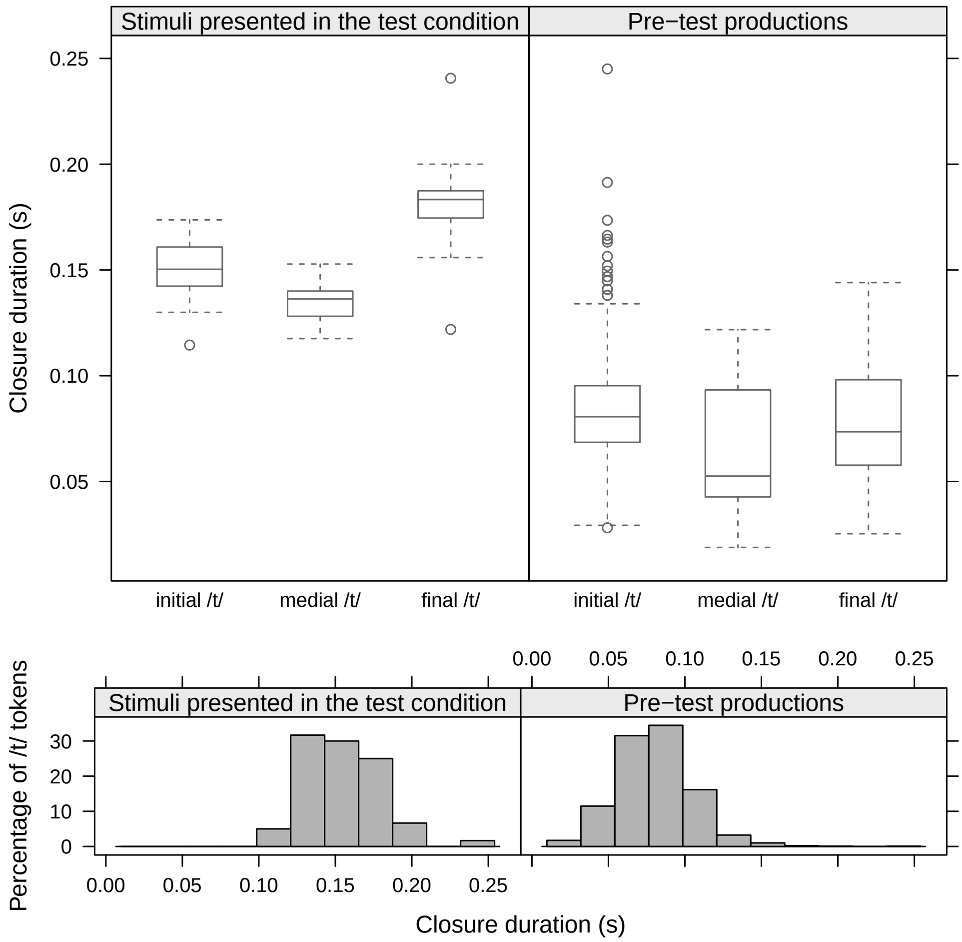

Figure 5.4 – Top: box-and-whisker diagrams plotting the closure duration of /t/ stops presented to participants in the test condition and the closure duration of /t/ stops produced by all participants in the pre-test condition. Bottom : histograms plotting the distribution of /t/ tokens as a function of closure duration.

Figure 5.4 shows that, in all positions, participants' closure duration for /t/ stops during pre-test (*ie.* their baseline production) is shorter than the closure duration of the stimuli presented in the test condition. Wilcoxon rank sum tests for equal medians were carried out for each position to confirm this graphical observation and are all highly significant ($p < .0001$). Outliers in the box-and-whisker diagrams were manually checked. They reflect inter-speaker variability in the production of /t/ and account for a very small proportion of our observations, as shown on the histograms at the bottom of the figure.

Speech rate

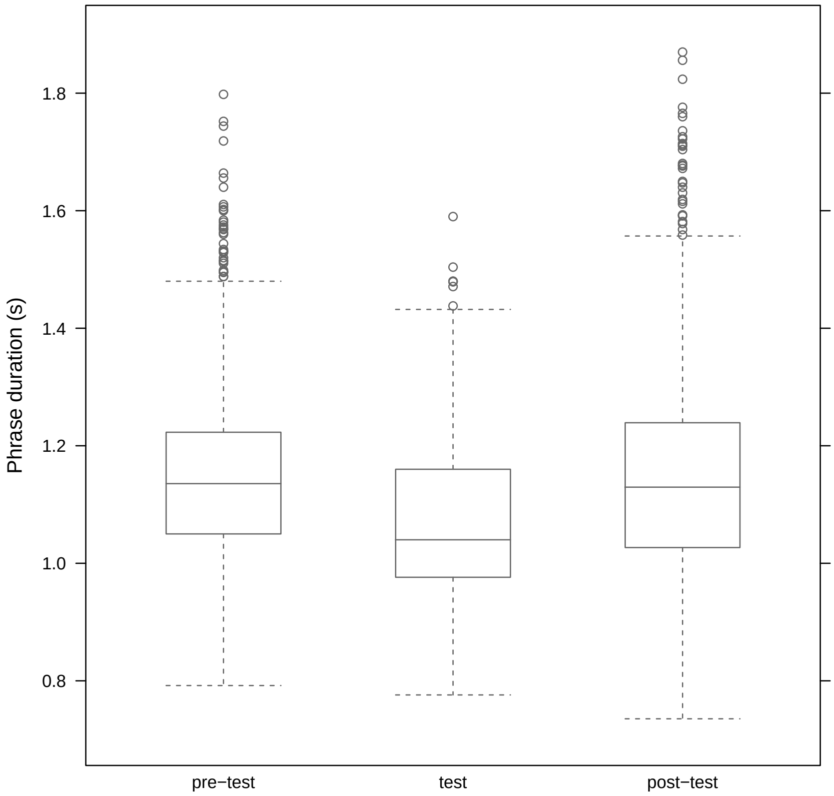

Figure 5.5 – Phrase duration across conditions.

Figure 5.5 shows that phrase duration decreases in the test condition and remains stable in pre and post-test conditions. Data were submitted to a linear mixed-effects model. The dependent variable was phrase duration, a continuous effect. We specified one independent variable as a fixed effect: condition (a 3 level factor including pre-test, test and post-test conditions). Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

Table 5.1 shows that the slope is negative and significantly steeper for the test condition than for the pre-test condition. This result indicates that phrase duration decreases in the test condition and remains stable in pre and post-test conditions. As phrase duration can be used as an indicator of speech rate, participants can be said to speak faster in the test condition.

In this section we presented the results across all groups of participants. In the next sections

	Estimate	$Pr(>\|t\|)$
(Intercept)	1.1450	0.0000
post-test	0.0007	0.8672
test	-0.0703	0.0000

Table 5.1 – Model output. Significant effects are highlighted

we will investigate each group of participants separately in terms of durational imitation. We will also investigate whether imitation can drive allophonic selection.

5.3.2 Results for group A (participants trained on initial /t/)

For the remainder of the analyses, glottal articulations and allophones grouped in the miscellaneous category (as described in section 6.2.5) were discarded because their duration cannot be reliably measured.

Duration of shadowed initial /t/ tokens

This section aims at determining if speakers are able to extend the duration of shadowed initial /t/ tokens in the test condition. If so, it will indicate that spontaneous imitation effects can take place. We also investigate whether tokens that have been shadowed are lengthened in the post-test condition. If so, it will indicate that imitation effects can persist.

We isolated the tokens labelled as shadowed (*ie.* the words containing an initial /t/ shadowed by participants in Group A) and tested whether there was an effect of condition. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we included one independent variable: condition. Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

Table 5.2 shows the output of the model predicting closure duration. In this model, the post-test condition is significant and the slope is negative. There is no significant effect in the test condition. This means that closure duration for shadowed tokens is not affected in the test condition but is shorter in the post-test condition. The fact that there is shortening likely relates to changes in speech rate and word shortening due to repetition. We therefore also consider the CPratio.

Table 5.3 shows the output of the model predicting CPratio. The post-test condition is not significant. The test condition is significant and the slope is positive. This means that CPratio in the test condition increased for shadowed tokens compared to the counterparts in the pre-test condition, and is was not affected in the post-test condition. It is worth noting that the

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0756	0.0000
post-test	-0.0039	0.0015
test	-0.0009	0.4334

Table 5.2 – Model output. DV: closure duration

carrier phrase duration significantly decreased from pre-test condition to test condition, as shown in section 5.3.1, thus the temporal space occupied by consonant closure is even greater when phrases are shorter.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0645	0.0000
post-test	-0.0016	0.1782
test	0.0091	0.0000

Table 5.3 – Model output. DV: CPratio

These results provide evidence for spontaneous imitation effects. However imitation is not to be found in terms of raw closure duration but in terms of the temporal space occupied by closure over the whole length of the carrier phrase. No carry-over effects were found in the post-test condition.

Duration of non-shadowed initial /t/ tokens

This section aims at determining whether initial /t/ tokens that were not shadowed can be affected in the post-test condition. If so, it will indicate that imitation can carry-over effect to non-shadowed tokens in the same position. Given the results reported in the previous section, this is not expected.

We took pre-test and post-test data only and tested for a possible interaction between condition and whether words were shadowed or not. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we included two independent variables: condition and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects.

When predicting closure duration, the interaction coefficient is not significant and was dropped from the model. Table 5.4 shows the output of the model without the interaction. Here, the post-test condition is significant and the slope is negative. There is no significant effect for shadowed tokens.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0782	0.0000
post-test	-0.0043	0.0000
shadowed$_{true}$	-0.0024	0.2508

Table 5.4 – Model output. DV: closure duration

When predicting CPratio, the interaction coefficient is not significant and was dropped from the model. Table 5.5 shows the output of the model without the interaction. The post-test condition is significant and the slope is negative. There is no significant effect for shadowed tokens.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0680	0.0000
post-test	-0.0021	0.0169
shadowed$_{true}$	-0.0032	0.1073

Table 5.5 – Model output. DV: CPratio

The results show a repetition effect: the more often the words are presented to the speakers, the shorter the closure duration and the lower the CPratio. They rule out carry-over imitation effects in initial position.

Duration of medial and final /t/ tokens

This section aims at determining whether speakers trained on initial /t/ are able to extend the duration of /t/ tokens in medial and final position in the post-test condition *ie.* whether positional transfer can occur. It would be surprising, however, to find such effects since no effects were found in the post-test condition for initial tokens.

Data were submitted to a series of linear mixed-effects models. Separate models were run for stop and fricative realisations. Our first series of models included closure duration (for stops) or consonant duration (for fricatives) and CPratio as dependent variables respectively and were run on medial /t/ data. Our second series of models included closure duration (for stops) or consonant duration (for fricatives) and CPratio as dependent variables respectively and were run on final /t/ data. This provided 8 models in total. For each model we specified one independent variable as a fixed effect: condition (a 2 level factor including pre-test and post-test conditions). Speakers and words were specified as random effects.

No significant effects were found for /t/ tokens realised as stops. There was a significant effect found in the dataset investigating medial fricatives, as shown on table 5.6. In the post-test condition, CPratio for /t/ tokens realised as fricatives significantly increases compared to their counterparts in the pre-test condition. However it is difficult to argue that positional transfer from initial position to medial position occurred since no effects were found in the post-test condition for initial tokens, nor for other positions.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0763	0.0000
post-test	0.0037	0.0133

Table 5.6 – Model output. DV: CPratio. Dataset: medial fricatives.

Allophonic variation

This section aims at determining whether participants in group A, who shadowed lengthened canonical /t/ stops in initial position, produced more canonical stops in the experiment. Participants in group A were already at ceiling level since /t/ in initial position is always produced as an aspirated canonical stop in English and participants were trained on that position. Therefore we excluded initial tokens from the analysis.

Data was subset to exclude initial stops and submitted to two logit mixed-effects model, one per position to analyse medial and final realisations. We included one independent variable : condition, a two level factor including pre-test condition and post-test condition. Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

The output of the two models is given in tables 5.7 and 5.8. For both positions participants in Group A significantly decreased the amount of canonical stops they produced.

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	-1.1316	0.3479	-3.253	0.00114
post-test	-0.4414	0.2129	-2.073	0.03814

Table 5.7 – Model output. Data set: medial /t/

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	-1.0370	0.7193	-1.442	0.149
post-test	-1.1155	0.2694	-4.141	<0.0001

Table 5.8 – Model output. Data set: final /t/

5.3.3 Results for group B (participants trained on medial /t/)

Duration of shadowed medial /t/ tokens

This section aims at determining if speakers are able to extend the duration of shadowed medial /t/ tokens in the test condition. If so, it will indicate that spontaneous imitation effects can take place. We also investigate whether tokens that have been shadowed are lengthened in the post-test condition. If so, it will indicate that imitation effects can persist. We investigate /t/ tokens realised as stops well as fricatives.

We isolated the tokens labelled as shadowed (*ie.* the words containing a medial /t/ shadowed by participants in Group B) and tested whether there was an effect of condition. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration (for stops) and consonant duration (for fricatives). In the second model the dependent variable was CPratio. In both models we included one independent variable: condition. Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

Results for /t/ tokens realised as stops:

No significant effects were found when predicting raw closure duration for shadowed medial /t/ tokens realised as stops. Main effects might not have reached significance since there were very few canonical stops in pre-test, test and post-test conditions, as shown on table 5.9.

Pre-test condition	Test condition	Post-test condition
12	36	13

Table 5.9 – Number of observations for shadowed /t/ tokens realised as fricatives

Table 5.10 shows the output of the model predicting CPratio. The post-test condition is not significant. The test condition is significant and the slope is positive. This means that CPratio in the test condition increased for shadowed tokens compared to the counterparts in the pre-test condition.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0567	0.0000
post-test	0.0051	0.4139
test	0.0099	0.0443

Table 5.10 – Model output. DV: CPratio

These results provide evidence for spontaneous imitation effects. However, as for initial position, imitation is not to be found in terms of raw closure duration but it terms of the temporal space occupied by closure over the whole length of the carrier phrase. No carry-over effects were found in the post-test condition. These results have to be taken with care as there were few observations.

Results for /t/ tokens realised as fricatives:

Table 5.11 shows the output of the model predicting raw consonant duration for fricatives. In this model, the post-test condition is not significant. There is a significant effect of test condition and the slope is positive.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0940	0.0000
post-test	-0.0020	0.1411
test	0.0088	0.000

Table 5.11 – Model output. DV: consonant duration

Table 5.12 shows the output of the model predicting CPratio. The post-test condition is not significant. The test condition is significant and the slope is positive. This means that CPratio in the test condition increased for shadowed tokens compared to the counterparts in the pre-test condition. It is worth noting that the carrier phrase duration significantly decreased from pre-test condition to test condition, as shown in section 5.3.1, thus the temporal space occupied by consonant closure is even greater when phrases are shorter.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0836	0.0000
post-test	-0.0014	0.3065
test	0.0120	0.0000

Table 5.12 – Model output. DV: CPratio

These sets of results provide evidence of spontaneous imitation effects in terms of raw consonant duration as well as in terms of the temporal space occupied by the consonant over the whole length of the carrier phrase, but they do not provide evidence of carry-over effects.

Duration of non-shadowed medial /t/ tokens

This section aims at determining whether medial /t/ tokens that were not shadowed can be affected in the post-test condition. If so, it will indicate that imitation can carry-over effect to non-shadowed tokens in the same position. This would be surprising since no carry-over effects were observed for shadowed tokens.

We took pre-test and post-test data only and tested for a possible interaction between condition and whether words were shadowed or not. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration (for stops) and consonant duration (for fricatives). In the second model the dependent variable was CPratio. In both models we included two independent variables: condition and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects.

Results for /t/ tokens realised as stops:

No significant effects were found when predicting raw consonant duration or CPratio. These results rule out any carry-over imitation effects to medial fricatives that were not shadowed. Note that there was also a limited number of observations: 20 observations in the pre-test condition and 21 observations in the post-test condition .

Results for /t/ tokens realised as fricatives:

No significant effects were found when predicting consonant duration and CPratio. These results rule out any carry-over imitation effects for medial fricatives.

Duration of initial and final /t/ tokens

This section aims at determining if speakers are able to extend the duration of /t/ tokens in initial and final positions in the post-test condition *ie.* whether positional transfer can occur. It would be surprising to observe positional transfer since no carry-over effects were observed in medial position.

Data were submitted to a series of linear mixed-effects models. Our first series of models included closure duration and CPratio as dependent variables respectively and were run on initial /t/ data. Our second series of models included closure duration (for stops) or consonant duration (for fricatives) and CPratio as dependent variables respectively and were run on final /t/ data. This provided 6 models in total. For each model we specified one independent variable as a fixed effect: condition (a 2 level factor including pre-test and post-test conditions). Speakers and words were specified as random effects.

There was a significant effect of condition in four of the models. Closure duration and CPratio decreased for initial /t/ in the post-test condition (the output of the models is not reported). Closure duration and CPratio increased in the post-test conditions for final /t/ tokens realised as stops, as shown on tables 5.13 and 5.14.

This may indicate a transfer in duration from medial /t/ stops to final /t/ stops, but it is unclear whether medial stops did lengthen in the post-test condition due to a lack of observations.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0550	0.0000
post-test	0.0129	0.0142

Table 5.13 – Model output. DV: Closure duration. Dataset: final /t/ stops

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0475	0.0000
post-test	0.0088	0.0187

Table 5.14 – Model output. DV: CPratio. Dataset: final /t/ stops

Allophonic variation

This section aims at determining whether participants in group B, who shadowed lengthened canonical /t/ stops in medial position, produced more canonical stops in the experiment.

Data was subset so that initial stops were excluded from the analysis (since initial /t/ is always realised as an aspirated stop in English) and submitted to two logit mixed-effects model, one per position to analyse medial and final realisations.

In medial position we tested whether there was an interaction between condition and whether words were trained on or not. The dependent variable was a binary effect coded as "stop" or "other". we included two independent variables: condition and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects.

In final position we tested whether there was an effect of condition. The dependent variable was a binary effect coded as "stop" or "other". We included one independent variables: condition. Pre-test condition was selected as reference levels. We specified speakers and words as random effects.

In medial position, no significant effects were found. In final position there was an effect of condition, as shown in table 5.15, with significantly fewer stops in the post-test condition that in the pre-test condition.

These results show that participants in group B produced more in the test condition compared to the pre-test condition. They also significantly reduced the use of the canonical variant in the post-test condition in final position.

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	-2.4185	0.6210	-3.895	<0.0001
post-test	-0.7948	0.3429	-2.318	0.0205

Table 5.15 – Model output. Data set: final /t/

5.3.4 Results for group C (participants trained on final /t/)

Duration of shadowed final /t/ tokens

This section aims at determining if speakers are able to extend the duration of shadowed final /t/ tokens in the test condition. If so, it will indicate that spontaneous imitation effects can take place. We also investigate whether tokens that have been shadowed are lengthened in the post-test condition. If so, it will indicate that imitation effects can persist. We investigate /t/ tokens realised as stops well as fricatives.

We isolated the tokens labelled as shadowed (*ie.* the words containing a final /t/ shadowed by participants in Group C) and tested whether there was an effect of condition. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration (for stops) and consonant duration (for fricatives). In the second model the dependent variable was CPratio. In both models we included one independent variable: condition. Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

Results for /t/ tokens realised as stops:

Table 5.16 shows the output of the model predicting closure duration for final /t/ stops. In this model, the post-test condition is not significant. There is a significant effect of test condition and the slope is positive. This means that closure duration in the test condition increased for shadowed tokens compared to the counterparts in the pre-test condition, and is was not affected in the post-test condition.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0729	0.0000
post-test	-0.0032	0.5037
test	0.0220	0.0000

Table 5.16 – Model output. DV: consonant duration

Table 5.17 shows the output of the model predicting closure duration for final /t/ stops. In this model, the post-test condition is not significant. There is a significant effect of test condition and the slope is positive. This means that CPratio in the test condition increased for shadowed tokens compared to the counterparts in the pre-test condition, and is was not affected in the post-test condition. It is worth noting that the carrier phrase duration significantly decreased from pre-test condition to test condition, as shown in section 5.3.1, thus the temporal space occupied by consonant closure is even greater when phrases are shorter.

These results provide evidence for spontaneous imitation effects, in terms of raw closure duration and also in terms of the temporal space occupied by closure over the whole length of

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0633	0.0000
post-test	-0.0045	0.2066
test	0.0142	0.0000

Table 5.17 – Model output. DV: CPratio

the carrier phrase. However, no carry-over effects were found in the post-test condition.

Results for /t/ tokens realised as fricatives:

No significant effects were found when predicting raw closure duration and CPratio for shadowed final /t/ tokens realised as fricatives. Main effects might not have reached significance since there was very few observations in pre-test, test and post-test conditions, as can be seen on table 5.18.

Pre-test condition	Test condition	Post-test condition
19	20	28

Table 5.18 – Number of observations for shadowed /t/ tokens realised as fricatives

Duration of non-shadowed final /t/ tokens

This section aims at determining whether final /t/ tokens that were not shadowed can be affected in the post-test condition. If so, it will indicate that imitation can carry-over effect to non-shadowed tokens in the same position.

We took pre-test and post-test data only and tested for a possible interaction between condition and whether words were shadowed or not. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration (for stops) and consonant duration (for fricatives). In the second model the dependent variable was CPratio. In both models we included two independent variables: condition and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects.

Results for /t/ tokens realised as stops:

No significant effects were found when predicting raw consonant duration and CPratio. These results rule out any carry-over imitation effects to medial stops that were not shadowed.

Results for /t/ tokens realised as fricatives:

No significant effects were found when predicting raw consonant duration and CPratio. These results rule out any carry-over imitation effects to medial fricatives that were not shadowed. Note that was also a limited number of observations : 30 observations in the pre-test condition and 42 observations in the post-test conditions.

Duration of initial and medial /t/ tokens

This section aims at determining if speakers are able to extend the duration of /t/ tokens in initial and medial positions in the post-test condition *ie.* whether positional transfer can occur.

Data were submitted to a series of linear mixed-effects models. Our first series of models included closure duration and CPratio as dependent variables respectively and were run on initial /t/ data. Our second series of models included closure duration (for stops) or consonant duration (for fricatives) and CPratio as dependent variables respectively and were run on medial /t/ data. This provided 6 models in total. For each model we specified one independent variable as a fixed effect: condition (a 2 level factor including pre-test and post-test conditions). Speakers and words were specified as random effects.

There was a significant effect of condition when predicting CPratio, which decreased for initial /t/ tokens realised as stops in the post-test condition, as shown on table 5.19. Taken together these results demonstrate that positional transfer from final position to initial or medial position did not occur.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0756	0.0000
post-test	-0.0026	0.0159

Table 5.19 – Model output. DV: CPratio. Dataset: initial /t/ stops

Allophonic variation

This section aims at determining whether participants in group C, who shadowed lengthened canonical /t/ stops in final position, produced more canonical stops in the experiment.

Data was subset so that initial stops were excluded from the analysis (since initial /t/ is always realised as an aspirated stop in English) and submitted to two logit mixed-effects model, one per position to analyse medial and final realisations.

In medial position we tested whether there was an effect of condition. The dependent variable was a binary effect coded as "stop" or "other". We included one independent variable: condition. Pre-test condition was selected as reference levels. We specified speakers and words as random effects.

In final position we tested whether there was an interaction between condition and whether words were trained on or not. The dependent variable was a binary effect coded as "stop" or "other". we included two independent variables: condition and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects.

In medial position there was an effect of condition, as shown in table 5.20, with significantly more stops in the post-test condition that in the pre-test condition.

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	-1.3463	0.5137	-2.621	0.00878
post-test	0.5912	0.2204	2.683	0.00729

Table 5.20 – Model output. Data set: medial /t/

In final position, the interaction coefficient was not significant and was dropped from the model. There was a effect of condition, as shown in table 5.21, with significantly less stops in the post-test condition than in the pre-test condition. Also, there was no effect of whether the words had been shadowed.

	Estimate	Standard Error	z value	$Pr(>\|z\|)$
(Intercept)	-1.2280	0.5409	-2.270	0.0232
post-test	0.5905	0.2205	2.678	0.0074
shadowed$_{true}$	-0.1776	0.2573	-0.690	0.4900

Table 5.21 – Model output. Data set: final /t/

These results show that participants in group C contributed produced more stops in the test condition. This effect carried-over as they also produced significantly more stops in the post-test condition both in medial and final positions.

5.3.5 Results for /d/ stops

So far we investigated /t/ tokens realised as stops and fricatives. We now turn to the results for /d/ stops realised as stops. We hypothesised that if speakers lengthened their /t/'s in the position they were trained on and if it carried over to /t/'s in the same position in the post-test then speakers could lengthen /d/ stops in the same position. Also, they might lengthen /d/ stops in positions they were not trained on if positional transfer occurred.

Data were submitted to a series linear mixed-effects models per group of speakers. Closure duration and CPratio were specified as dependent variables. We specified one independent variable as a fixed effect: condition (a 2 level factor including pre-test and post-test conditions). Speakers and words were specified as random effects. For each group, we tested whether there was an effect of condition in the position speakers were trained on as well as in other position. This lead to 18 separate models (2 DV's x 3 positions tested x 3 groups of speakers).

There was a significant effect of condition for 5 of the models. Closure duration for initial /d/ stops significantly decreased for participants in Group A, who shadowed initial /t/ tokens (the output of the model is not reported). Closure duration and CPratio for initial /d/ stops significantly decreased for participants in Group B, who shadowed medial /t/ tokens (the output of the models is not reported). Closure duration and CPratio for medial /d/ stops significantly increased for participants in Group B, who shadowed medial /t/ stops, as shown on tables 5.22 and 5.23.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0326	0.0000
post-test	0.0043	0.0153

Table 5.22 – Model output. DV: closure duration. Dataset: medial /d/ stops for participants in group B

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0294	0.0000
post-test	0.0042	0.0062

Table 5.23 – Model output. DV: CPratio. Dataset: medial /d/ stops for participants in group B

5.3.6 Summary of the results

Duration results

Table 6.25 summaries the results in terms of raw duration and CPratio. There were some duration decreases in the post-test condition which we attribute to repetition effects and do not discuss further.

Participants in Group A, who shadowed initial /t/ tokens, significantly increased CPratio for initial /t/ tokens realised as stops in the test condition compared to their counterparts in the pre-test condition. They also significantly increased the CPratio of medial /t/ fricatives in the post-test condition.

Participants in Group B, who shadowed medial /t/ tokens, significantly increased CPratio for medial /t/ tokens realised as stops in the test condition (however this result has to be taken with caution since there were very few observations). They significantly increased closure duration and CPratio for medial /t/ tokens realised as fricatives in the test condition compared to their counterparts in the pre-test condition. They significantly decreased closure duration and CPratio of initial stops in the post-test condition. Finally they significantly increased the closure duration and CPratio of medial /d/ stops in the post-test condition.

Participants in Group C, who shadowed final /t/ tokens, significantly increased closure duration and CPratio for final /t/ tokens realised as stops in the test condition compared to their counterparts in the pre-test condition.

Our results indicate that spontaneous imitation effects can be found in all groups. However no carry-over effects were found in the post-test condition for the shadowed tokens which increased in duration, nor for tokens that were not shadowed and in the same position. Thus, for those participants who increased the duration and/or CPratio of /t/ tokens in a position they were not trained on or increased the duration and CPratio of /d/ stops, it is unlikely that positional transfer or transfer to the /d/ phoneme genuinely occurred.

	Group A		Group B		Group C	
	raw duration	CPratio	raw duration	CPratio	raw duration	CPratio
initial /t/ stops	-	+	NA	NA	NA	NA
medial /t/ stops	NA	NA		+	NA	NA
medial /t/ fricatives	NA	NA	+	+	NA	NA
final /t/ stops	NA	NA	NA	NA	+	+
final /t/ fricatives	NA	NA	NA	NA		

(a) Test condition

	Group A		Group B		Group C	
	raw duration	CPratio	raw duration	CPratio	raw duration	CPratio
initial /t/ stops	-	-	-	-		-
initial /d/ stops	-					
medial /t/ stops						
medial /t/ fricatives		+				
medial /d/ stops			+	+		
final /t/ stops						
final /t/ fricatives						
final /d/ stops						

(b) Post-test condition

Table 5.24 – Summary of each mixed-effect model run per group. A plus sign indicates an significant increase in duration or in CPratio compared to the participants' baseline in the pre-test condition. A minus sign indicates a significant decrease in duration or in CPratio compared to the participants' baseline in the pre-test condition. Empty cells indicate no significant change. NA's indicate that the tests were not available. Table (a) presents the results in the test condition. Table (b) presents the results in the post-test condition.

Allophonic variation results

Participants were presented with stops only during the test condition and, overall, produced significantly more stops in the test condition than in the pre-test condition. We concluded that imitation effects could drive allophonic selection.

Participants in group A, who were trained on initial canonical stops, did not contribute to this effect, as they produced fewer stops in the test condition. However participants in group B and group C both produced more canonical stops in the test condition. Moreover, this effect carried over to the post-test condition for participants in group C who shadowed final /t/ stops.

The carrier phrase duration significantly decreased in the shadowing condition. In such conditions, one might expect speech reduction to take place, causing a decrease in the use of stops and a potential increase towards the use of flaps or taps, in medial position in particular. As participants produced more stops when shadowing than reading, this leads us to think that this is genuine allophonic imitation.

5.4 Discussion

This experiment sought to identify whether imitation effects can generalise to other positions in the word. A prerequisite to answer this question was to demonstrate that our experimental design is able to generate spontaneous imitation effects in the test condition. This prerequisite was met, as participants in all groups extended the duration of stops in the test condition. Also we have to demonstrate that imitation effects, if any, should carry-over to tokens in the same word position. This prerequisite was not met. Our experimental design did not facilitate transfer of durational lengthening in the post-test condition. In this section, we will further discuss why this was the case.

The duration of fricatives was affected when participants in Group B shadowed medial stops. Given that their medial fricatives increased in duration in the test condition, potential carry-over effects might have occurred. However it remains unclear why medial fricatives were not affected in the post-test condition as well. It is also unclear why participants who shadowed initial /t/ stops did not increase the duration of initial /t/ stops in the post-test condition but increased the duration of medial /t/ fricatives in the post-test condition. Given the results we obtain it is difficult to conclude that positional transfer did occur. It is also difficult to conclude that phonemic transfer to /d/ occurred since only medial /d/ stops for participants in group B were lengthened in the post-test condition.

Positional transfer might actually be very restricted (or could even be inhibited) in New Zealand English given the allophonic make-up of the language. NZE has different allophones in initial, medial and final position and the patterning of these allophones is socially conditioned, as outlined in section 1.6. In initial position /t/ is always realised as an aspirated stop whether followed by a stressed or an unstressed vowel (Bauer, 1986); in medial position /t/ can be realised as a canonical variant with varying degrees of aspiration, a flap, a fully-voiced variant, a glottal stop and a fricative (Bell, 1977; Holmes, 1994, 1995a; Bayard, 1999; Taylor, 1996); in final position /t/ can be released, glottalised, spirantised or affricated (Holmes, 1995b; Docherty et al., 2006). Thus we may argue that in NZE there are three distinct phonetic categories, one for each position. For example, for young female NZE speakers who nearly always produce medial /t/ as a fricative and produce initial /t/ as an aspirated stop, these speakers clearly have a separate phonetic distribution for each position. It might be the case that because NZE speakers have three different categories (one for each position), inducing a shift where NZE speakers are sampling in the initial distribution, for example, might not have any effect (or much effect) to their production of medial /t/s and vice versa.

Our further sets of experiment will address this problem, by conducting a cross-linguistic study between French and English. Standard French is known to have unaspirated stops in all word positions. Comparing French and English data will help us shed light on whether positional transfer (if it can occur at all) is actually dependent on the degree of phonetic similarity across positions.

Another motivation for gathering more data and further investigating positional transfer are some of the methodological problems associated with our experimental design. Speakers shadowed words only once and given that speakers produce more fricatives than stops in medial position (as shown by the results of our experiment in chapter 4) this explains why only 36 stops were produced in the shadowing condition for Group B. Also, speakers were presented with 120 tokens in pre-test and post-test conditions compared to 40 in the test condition. This imbalance might have weakened the echo for tokens in the test condition. This could be one of the reasons why duration did not extend to their counterparts in the post-test condition. Furthermore different types of fillers were presented in pre and post-test conditions compared to those in the test condition, which further contributed to the imbalance in our experimental design.

The most robust finding in this experiment was that participants in group B and group C produced significantly more stops in the test condition than in the pre-test condition, and it carried-over to the post-test condition in group C. Both groups have non-canonical realisations of /t/ as their most widely used variants in medial and final position. Since they were only presented with stops in the test condition, our interpretation is that allophonic variation can be driven by imitation, and this is a novel contribution to the literature. Participants in group A were already at ceiling level in the test condition since /t/ in initial position is always produced as an aspirated canonical stop in English and since initial position is also the position where those participants were trained on. This provides an answer to one of the research questions stated in our introduction, of whether imitation of a variant would drive allophonic selection.

Chapter 6

A cross-linguistic study of positional transfer

6.1 Introduction

In the previous chapter we showed that allophonic variation can be driven by imitation since participants were presented with stops only during the test condition and produced significantly more stops in the test condition than in the pre-test condition. However it remains unclear whether carry-over effects to other positions are genuinely taking place.

Positional transfer might not have occurred because of some of the problems associated with our experimental design. We raised the issue that, maybe, positional transfer could actually be very restricted (or could even be inhibited) in New Zealand English given the allophonic make-up of the language. New Zealand English has different allophones in initial, medial and final position and the patterning of these allophones is socially conditioned, as outlined in section 1.6. It might be the case that because NZE speakers have three different categories (one for each position), inducing a shift where NZE speakers are sampling in the initial distribution, for example, might not have any effect (or much effect) to their production of medial /t/s and vice versa.

Standard French, on the other hand, is known to have unaspirated stops in all word positions (Tranel, 1987). Thus /t/ exemplars all overlap in phonetic space, regardless of position and /t/ can be thought of as one large distribution or as one phonetic category. Inducing a shift where SF speakers are sampling from their /t/ category, should shift their production in all positions (*ie.* word-initially and word-medially).

We further investigate the main question asked in the previous chapter by modifying and simplifying our previous experimental design: can imitation effects carry over to other positions in the word? If participants are trained on a phoneme in initial position, for example, will the imitation effects extend to other positions on the word? Comparing French and English data will help us shed light on whether positional transfer (if it can occur at all) is actually dependent on the degree of phonetic similarity across positions.

In this chapter we describe four experiments that were conducted to answer this question, two in French and two in English. In the first experiment in each language, we tested whether shadowing words with extended initial /t/ durations lead to an imitation effect, and whether this effect generalised to non-shadowed words with initial /t/. In the second experiment in each language, we tested whether shadowing words with extended initial /t/ durations lead to an imitation effect which generalised to medial position.

6.2 Methodology

6.2.1 Stimuli

Stimuli words in French were created by querying the Lexique 3 database (New et al., 2001). We selected bisyllabic singular nouns with low frequency, defined as having a score inferior to 1 on the "freqlivres" scale[1]. Each word was embedded in the carrier phrase: le/la/l' `WORD` aussi (the$_{masc.}$/the$_{fem.}$/the$_{preceding\ any\ vowel}$ `WORD` as well). This carrier phrase was chosen because words can be embedded between two vowels, thus making the identification of initial and final stops fairly easy on waveforms and spectrograms.

Stimuli words in English were subset from the previous experiment. They were chosen by querying the CELEX database (Baayen et al., 1993). We selected bisyllabic singular nouns with low frequency, defined as having a score lower than 500 on the "COBUILD" frequency scale[2]. Each word has a Strong/Weak stress pattern. Each word was embedded in the carrier phrase : a/an `WORD` as well. This carrier phrase was chosen because words can be embedded between two vowels, thus making the identification of initial and final stops fairly easy on waveforms and spectrograms.

In the first experiment in French and in English, we used 30 words with initial /t/ in the pre and post-test conditions. A subset of 15 words was used in the test condition. In the second experiment in French and in English, we used 15 words with initial /t/ and 15 words with medial /t/ in the pre and post-test conditions. We used the same 15 words with initial /t/ in the test condition. In both experiments for each language the words presented in the test condition were identical and had /t/ in initial position. A table showing the selected words

[1] This measure corresponds to the frequency of the occurrences of a word in the Lexique book corpus, which contains 14,700,000 words.

[2] The frequency information in the CELEX English database was taken from the COBUILD corpus, which contained 17,900,000 words when it was first published. COBUILD is an acronym for Collins Birmingham University International Language Database.

can be found in appendix C.

6.2.2 Data acquisition from the model speaker

A SF speaker and a NZE speaker were recruited in France. They were prompted with the stimuli list in their native language using the Prorec software (Huckvale, 2003). They were asked to read the prompts and to repeat if they made an error or if they hesitated. A real time spectrogram was running in order to monitor and make sure that the NZE speaker would produce canonical stops. Recordings were acquired using an AKG-C520 head-mounted microphone connected to a Schure X2u XLR to USB signal adapter. The signal was acquired onto a laptop using Audacity and was sampled at 44.1 kHz. We used the same recordings for the NZE speaker as in the previous experiment. The experimenter provided the participant with an information sheet and also gave instructions orally.

6.2.3 Stimuli manipulation

We used the same stimuli in NZE as we did in the previous chapter. In French the words recorded from the model speaker were manipulated to extend the duration of closure. Our intent was to create a very salient and long closure duration that the test participants would be able to notice auditorily and could potentially imitate without being explicitly instructed to. Our manipulation consisted of splicing "silence" at the mid part of consonantal closure at zero crossings. Ambient noise was actually added instead of pure silence as the recordings were not made in a sound-proof booth. Ambient noise was taken from a randomly selected portion of the signal corresponding to a pause made by the model speakers in between two words. 100ms were added to every token elicited by the model speakers in French and in English.

6.2.4 Data acquisition from the test speakers

40 SF speakers and 40 NZE speakers, all aged between 18 and 30, took part in the experiment. 20 SF speakers took part in experiment 1a and the other half in experiment 2a. 20 NZE speakers took part in experiment 1b and the other half in experiment 2b. In all experiments, participants were asked to shadow 15 initial /t/ words 3 times in the test-condition. Stimuli were presented binaurally through headphones. In experiments 1a and 1b, participants were presented with 30 initial /t/ words once in the pre and post-test conditions. In experiments 2a and 2b, participants were presented with 15 initial /t/ words and 15 medial /t/ words once in the pre and post-test conditions. In the pre and post-test conditions participants were prompted with the words using Prorec (Huckvale, 2003).

NZE Speakers were recorded by a research assistant, a New Zealand speaker of English, hired through a grant generously offered to the author of the thesis by the New Zealand Institute of Language, Brain and Behaviour, University of Canterbury. They were recorded using a AKG-C520 head-mounted microphone connected to a USB Pre sound card. The signal was acquired onto a laptop using Audacity and was sampled at 44.1 kHz. Recordings were made in a quiet room at the University of Canterbury. SF speakers were recorded by a research assistant who is a native standard French speaker. They were recorded using a head-mounted microphone connected directly to a laptop. The signal was acquired using Audacity and was sampled at 44.1 kHz. The experimenters provided the participants with an information sheet and also gave instructions orally.

6.2.5 Measures

Durational measures

Four durational measures were computed from the speech of the test speakers. We measured the closure duration of the alveolar stops defined as the corresponding period of silence seen on the waveform. We measured the duration of fricated stops at the offset and onset of surrounding vowels. We measured the length of each carrier phrase to give us an indication of possible speech rate variation between pre-test, test and post-test conditions. Finally we calculated the ratio between the closure duration or the fricative duration and the duration of the whole carrier phrase (CPratio, hereafter).

Labelling of speech sounds

Speech sounds in French were coded as stops while speech sounds in NZE were coded into three broad categories. Canonical stops, laryngealised alveolar stops and pre-aspirated stops (as described in chapter 5) were all coded as stops. Fricated /t/'s, laryngealised fricated /t/'s and pre-aspirated and post-aspirated fricated /t/'s (as described in chapter 5) were all coded as fricatives. All other speech sounds, such as flaps, were coded as NA (Not Available) since only stops and fricatives could be reliably measured in terms of duration.

6.2.6 Predictions

Experiments 1a and 2a:

Our first set of experiments in French and in English (experiments 1a and 2a respectively) investigate whether imitation effects in initial position, if they are to be observed in the test condition, can generalise to new words in the same position, namely in initial position.

Prediction 1: in order to test this question, we must demonstrate that our experimental design is able to generate spontaneous imitation effects. If this is the case, participants will increase the duration and/or CPratio of shadowed /t/ stops in the test condition.

Prediction 2: we also have to demonstrate that imitation effects, if any, should carry-over to /t/ tokens in the same word position. If this is the case, initial /t/ tokens that were previously shadowed will increase in terms of duration and/or CPratio in the post-test condition. This will provide evidence for persistent imitation effects across experimental conditions.

Prediction 3: provided that predictions 1 and 2 realised, we predict that participants trained on initial /t/ words will increase the duration and/or CPratio in the post-test condition of non-shadowed /t/ items in the same word position.

Experiments 1b and 2b:

Our second set of experiments in French and in English (experiments 1b and 2b respectively) investigate whether imitation effects in initial position, if they are to be observed in the test condition, can generalise to new words in a different position, namely in medial position.

Prediction 4: in order to test this question, we must demonstrate that our experimental design is able to generate spontaneous imitation effects. If this is the case, participants will increase the duration and/or CPratio of shadowed /t/ stops in the test condition.

Prediction 5: we also have to demonstrate that imitation effects, if any, should carry-over to initial /t/ tokens. If this is the case, initial /t/ tokens that were previously shadowed will increase in terms of duration and/or CPratio in the post-test condition. This will provide evidence for persistent imitation effects across experimental conditions.

Provided that predictions 4 and 5 are realised, we can investigate whether participants trained in initial will increase the duration and/or CPratio in the post-test condition of non-shadowed /t/ items in a different word position. If they do then we can conclude that imitation effects can generalise to other word positions. Our predictions depend on the languages under study (see section 6.1).

Prediction 6a: In French we expect that participants trained in a given position will increase the duration and/or CPratio in the post-test condition of non-shadowed /t/ items in medial position, since /t/'s in initial and medial position are realised as stops.

Prediction 6b: In English, if we assume that imitation does not depend on the degree of similarity between allophones across positions, then we expect that participants trained on initial stops will increase the duration and/or CPratio in the post-test condition of non-shadowed /t/ items in medial position, whether there are realised as stops or as fricatives.

Prediction 6c: In English, if we assume that imitation does depend on the degree of similarity between allophones across positions, then we expect that participants trained on initial stops will increase the duration and/or CPratio in the post-test condition of non-shadowed /t/ items in medial position and which are realised as stops only.

6.3 Results

We excluded misproductions and productions when participants hesitated. Statistical outliers in each of the datasets presented below were manually checked. We first present overall results then we move on to the results for each group of speakers.

We present the results for the French speakers first, then move on to the results for the New Zealand English speakers.

6.3.1 Experiment 1a: results for the French speakers

This experiment investigates whether imitation effects in initial position, if they are to be observed in the test condition, can generalise to new words in the same position.

Carrier sentence duration

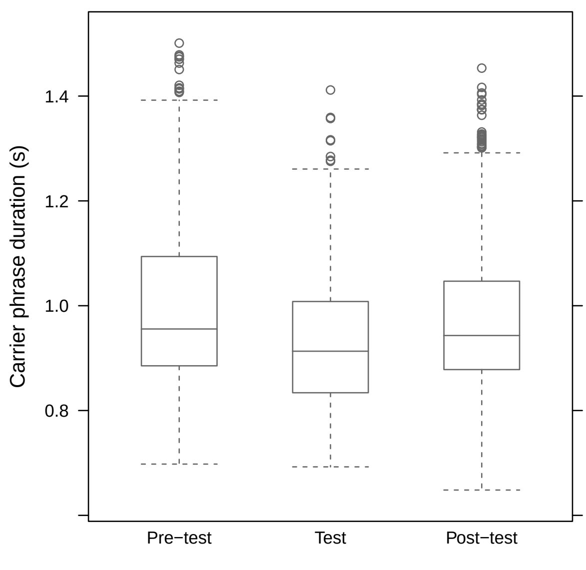

Figure 6.1 – Phrase duration across conditions.

Figure 6.1 shows the duration of the carrier phrases across conditions. Data were submitted to a linear mixed-effects model. The dependent variable was phrase duration, a continuous effect.

We specified one independent variable as a fixed effect: condition (a 3 level factor including pre-test, test and post-test conditions). Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

	Estimate	$Pr(>\|t\|)$
(Intercept)	1.0024	0
post-test	-0.0278	0
test	-0.0278	0

Table 6.1 – Model output

Table 6.1 shows that both slopes are negative and significantly different from the pre-test condition. The slope for the test condition is steeper than the slope for the post-test condition. This result indicates that phrase duration decreases in the post-test condition and even more so in the test condition. In other words, as phrase duration can be used as an indicator of speech rate, participants can be said to speak faster in the test and post-test conditions compared to the pre-test condition.

Duration of shadowed initial /t/ tokens

This section aims at determining if speakers are able to extend the duration of shadowed initial /t/ tokens in the test condition. If so, it will indicate that spontaneous imitation effects can take place. We also investigate whether tokens that have been shadowed are lengthened in the post-test condition. If so, it will indicate that imitation effects can persist.

We isolated the tokens labelled as shadowed (*ie.* the words containing an initial /t/ and that the participants were asked to shadow). Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we tested for an interaction between condition and trial number. Pre-test condition was selected as the reference level. Trial number is a continuous variable numbering the first trial to the last for each condition. We specified speakers and words as random effects. Removed from the models were (i) non-significant interactions and (ii) non-significant main effects which were not present in a significant interaction.

Table 6.2 shows the output of the model predicting closure duration. In this model, main effects did not reach significance. The post-test:trial number and test:trial number interaction coefficients were significant and their slopes are positive. This means that closure duration for shadowed tokens increased both in the test condition and in the post-test condition in the course of the experiment. Note that, although there is significant lengthening throughout the course of the test and post-test condition, the size of this effect is not colossal. An increase of .0003 seconds per token over the course of the shadowing condition, for example, amounts to a total predicted difference in .0135 seconds over the course of the condition. If we omit the interaction from the model, then both test ($p < .0001$) and post-test ($p < .001$) are significantly longer than the pre-test condition.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0866	0.0000
post-test	-0.0006	0.8284
test	0.0008	0.7430
trial number	0.0000	0.7325
post-test:trial number	0.0003	0.0334
test:trial number	0.0003	0.0289

Table 6.2 – Model output. DV: closure duration

Table 6.3 shows the output of the model predicting closure CPratio. Test and post-test conditions were significant and their slopes are positive. The slope for the test condition is steeper than the slope for the post-test condition. This means that CPratio for shadowed tokens increased both in the test condition and in the post-test condition compared to the pre-test condition and that the size of the effect is greater in the test condition. The carrier phrase duration significantly decreased from pre-test condition to test condition, as shown in section 6.3.1, thus the temporal space occupied by consonant closure is even greater when phrases are shorter. There was also a significant effect of trial number and the slope is positive. All things being equal, CPratio for shadowed tokens increased in the course of the experiment.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0841	0
post-test	0.0078	0
test	0.0117	0
trial number	0.0002	0

Table 6.3 – Model output. DV: CPratio

These results provide evidence for spontaneous imitation effects and persistent carry-over imitation effects for the tokens which were shadowed previously in terms of raw closure duration as well as in terms of the temporal space occupied by closure over the whole length of the carrier phrase.

Duration of non-shadowed initial /t/ tokens

This section aims at determining whether initial /t/ tokens that were not shadowed can be affected in the post-test condition. If so, it will indicate that imitation can carry-over effect to non-shadowed tokens in the same position. Given the results reported in the previous section, this is expected. We also test whether, if non-shadowed tokens are affected, whether they are affected to the same degree as tokens which were previously shadowed.

We took pre-test and post-test data only and tested for a possible interaction between condition, trial number and whether words were shadowed or not. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we included 3 independent variables: condition, trial number (a continuous variable numbering the first trial to the last for each condition) and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects. Removed from the models were (i) non-significant interactions and (ii) non-significant main effects which were not present in a significant interaction.

Table 6.4 shows the output of the model predicting closure duration. The interaction coefficient condition:trial number was significant and the slope is positive. This means that the closure duration increased in the post-test condition in the course of the experiment. Note that, although there is significant lengthening throughout the course of the post-test condition, the size of this effect is not colossal. An increase of .0002 seconds per token over the course of post-test condition amounts to a total predicted difference in 6ms over the course of the condition. If we omit the interaction from the model, then post-test is significantly longer than pre-test ($p < .0001$).

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0878	0.000
post-test	0.0020	0.3145
trial number	0.0000	0.6173
post-test:trial number	0.0002	0.0415

Table 6.4 – Model output. DV: closure duration

Table 6.5 shows the output of the model predicting CPratio. The post-test condition was significant and the slope is positive, which means that in this condition /t/ tokens increased in terms of CPratio.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0872	0
post-test	0.0085	0

Table 6.5 – Model output. DV: CPratio

The results appear to show carry-over imitation effects: for initial /t/ tokens there was an increase in terms of raw closure duration as well as in terms of the temporal space occupied by closure over the whole length of the carrier phrase. There is no significant interaction with whether words had been previously shadowed, indicating that shadowed and non-shadowed words behave identically in the post-test condition. As a separate check on this, a model was fit to the post-test condition only, in order to test for an effect of shadowed. No significant effect was found.

6.3.2 Experiment 1b: results for the English speakers

This experiment investigates whether imitation effects in initial position, if they are to be observed in the test condition, can generalise to new words in the same position.

Carrier sentence duration

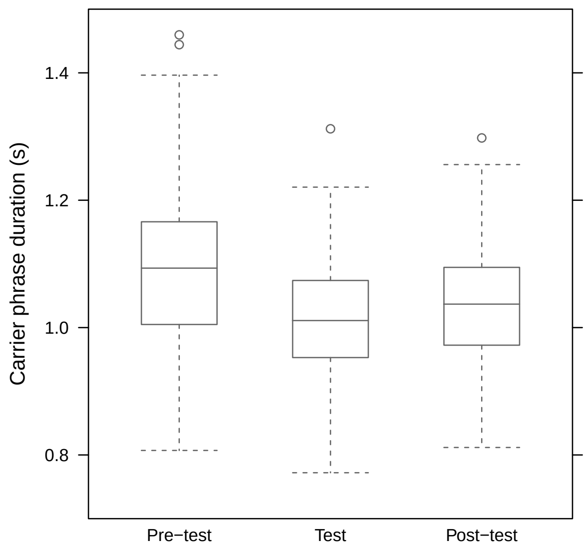

Figure 6.2 – Phrase duration across conditions.

Figure 6.2 shows the duration of the carrier phrases across conditions. Data were submitted to a linear mixed-effects model. The dependent variable was phrase duration, a continuous effect. We specified one independent variable as a fixed effect: condition (a 3 level factor including pre-test, test and post-test conditions). Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

	Estimate	$Pr(>\|t\|)$
(Intercept)	1.0912	0
post-test	-0.0591	0
test	-0.0843	0

Table 6.6 – Model output

Table 6.6 shows that both slopes are negative and significantly different from the pre-test condition. The slope for the test condition is steeper than the slope for the post-test condition. This result indicates that phrase duration decreases in the post-test condition and even more so in the test condition. In other words, as phrase duration can be used as an indicator of speech rate, participants can be said to speak faster in the test and post-test conditions compared to the pre-test condition.

Duration of shadowed initial /t/ tokens

This section aims at determining if speakers are able to extend the duration of shadowed initial /t/ tokens in the test condition. If so, it will indicate that spontaneous imitation effects can take place. We also investigate whether tokens that have been shadowed are lengthened in the post-test condition. If so, it will indicate that imitation effects can persist.

We isolated the tokens labelled as shadowed (*ie.* the words containing an initial /t/ and that the participants were asked to shadow). Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we tested for an interaction between condition and trial number. Pre-test condition was selected as the reference level. Trial number is a continuous variable numbering the first trial to the last for each condition. We specified speakers and words as random effects. Removed from the models were (i) non-significant interactions and (ii) non-significant main effects which were not present in a significant interaction.

Table 6.7 shows the output of the model predicting closure duration. The test condition was significant and the slope is positive. There was no significant effect for the post-test condition. This means that closure duration for shadowed tokens increased only in the test condition compared to the pre-test condition.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0753	0.0000
post-test	0.0000	0.9896
test	0.0064	0.0000

Table 6.7 – Model output. DV: closure duration

Table 6.8 shows the output of the model predicting closure CPratio. In this model, test and post-test conditions are significant and their slopes are positive. The slope for the test condition is steeper than the slope for the post-test condition. This means that CPratio for shadowed tokens increased both in the test condition and in the post-test condition compared to the pre-test condition and that the size of the effect is greater in the test condition. Also, the carrier phrase duration significantly decreased from pre-test condition to test condition, as shown in section 6.3.2, thus the temporal space occupied by consonant closure is even greater when phrases are shorter.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0693	0.0000
post-test	0.0034	0.0003
test	0.0115	0.0000

Table 6.8 – Model output. DV: CPratio

These results provide evidence for spontaneous imitation effects in terms of raw closure duration as well as in terms of the temporal space occupied by closure over the whole length of the carrier phrase (*ie.* CPratio). These results also provide evidence of persistent carry-over imitation effects for those tokens in terms of CPratio.

Duration of non-shadowed initial /t/ tokens

This section aims at determining whether initial /t/ tokens that were not shadowed can be affected in the post-test condition. If so, it will indicate that imitation can carry-over effect to non-shadowed tokens in the same position. Given the results reported in the previous section, this is expected. We also test whether, if non-shadowed tokens are affected, whether they are affected to the same degree as tokens which were previously shadowed.

We took pre-test and post-test data only and tested for a possible interaction between condition, trial number and whether words were shadowed or not. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we included 3 independent variables: condition, trial number (a continuous variable numbering the first trial to the last for each condition) and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects. Removed from the models were (i) non-significant interactions and (ii) non-significant main effects which were not present in a significant interaction.

When predicting closure duration, no significant main effects were found and the model is not reported.

Table 6.9 shows the output of the model predicting CPratio. The post-test condition was significant and the slope is positive, which means that in this condition non-shadowed /t/ tokens increased in terms of CPratio.

The results show carry-over imitation effects: initial /t/ tokens that were not shadowed increased in terms of the temporal space occupied by closure over the whole length of the carrier

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0717	0.0000
post-test	0.0038	0.0000

Table 6.9 – Model output. DV: CPratio

phrase. There is no significant interaction with whether words had been previously shadowed, indicating that shadowed and non-shadowed words behave identically in the post-test condition. As a separate check on this, a model was fit to the post-test condition only, in order to test for an effect of shadowed. No significant effect was found.

6.3.3 Experiment 2a: results for the French speakers

This experiment investigates whether imitation effects in initial position, if they are to be observed in the test condition, can generalise to new words in a different position, namely in medial position.

Carrier sentence duration

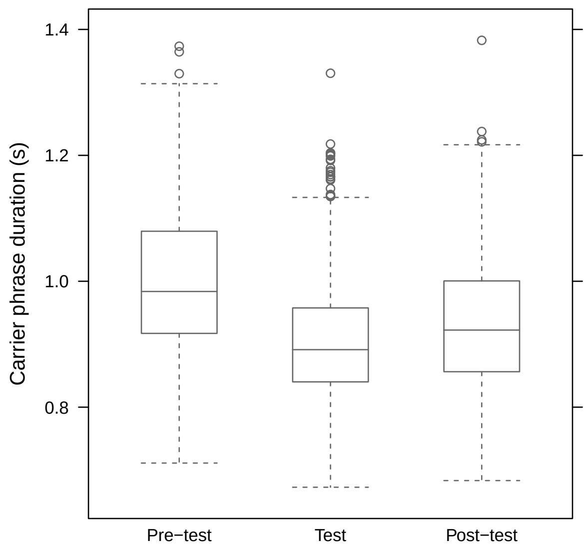

Figure 6.3 – Phrase duration across conditions.

Figure 6.3 shows the duration of the carrier phrases across conditions. Data were submitted to a linear mixed-effects model. The dependent variable was phrase duration, a continuous effect. We specified one independent variable as a fixed effect: condition (a 3 level factor including pre-test, test and post-test conditions). Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

	Estimate	$Pr(>\|t\|)$
(Intercept)	1.0048	0
post-test	-0.0703	0
test	-0.0936	0

Table 6.10 – Model output

Table 6.10 shows that both slopes are negative and significantly different from the pre-test condition. The slope for the test condition is steeper than the slope for the post-test condition. This result indicates that phrase duration decreases in the post-test condition and even more so in the test condition. In other words, as phrase duration can be used as an indicator of speech rate, participants can be said to speak faster in the test and post-test conditions compared to the pre-test condition.

Duration of shadowed initial /t/ tokens

This section aims at determining if speakers are able to extend the duration of shadowed initial /t/ tokens in the test condition. If so, it will indicate that spontaneous imitation effects can take place. We also investigate whether tokens that have been shadowed are lengthened in the post-test condition. If so, it will indicate that imitation effects can persist.

We isolated the tokens labelled as shadowed (*ie.* the words containing an initial /t/ and that the participants were asked to shadow). Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we tested for an interaction between condition and trial number. Pre-test condition was selected as the reference level. Trial number is a continuous variable numbering the first trial to the last for each condition. We specified speakers and words as random effects. Removed from the models were (i) non-significant interactions and (ii) non-significant main effects which were not present in a significant interaction.

Table 6.11 shows the output of the model predicting closure duration. Test and post-test conditions were significant and their slopes are positive. The slope for the test condition is steeper than the slope for the post-test condition. This means that closure duration for shadowed tokens increased both in the test condition and in the post-test condition compared to the pre-test condition and that the size of the effect is greater in the test condition. There was also a significant effect of trial number and the slope is positive. All things being equal, closure duration for shadowed tokens increased in the course of the experiment.

Table 6.12 shows the output of the model predicting closure CPratio. In this model, test and post-test conditions are significant and their slopes are positive. The slope for the test condition is steeper than the slope for the post-test condition. This means that CPratio for

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0698	0.0000
post-test	0.0145	0.0000
test	0.0201	0.0000
trial number	0.0001	0.0056

Table 6.11 – Model output. DV: closure duration

shadowed tokens increased both in the test condition and in the post-test condition compared to the pre-test condition and that the size of the effect is greater in the test condition. There was also a significant effect of trial number and the slope is positive. All things being equal, CPratio for shadowed tokens increased in the course of the experiment. Finally, the carrier phrase duration significantly decreased from pre-test condition to test condition, as shown in section 6.3.3, thus the temporal space occupied by consonant closure is even greater when phrases are shorter.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0692	0
post-test	0.0214	0
test	0.0294	0
trial number	0.0002	0

Table 6.12 – Model output. DV: CPratio

These results provide evidence for spontaneous imitation effects and persistent carry-over imitation effects for the tokens which were shadowed previously in terms of raw closure duration as well as in terms of the temporal space occupied by closure over the whole length of the carrier phrase.

Duration of non-shadowed medial /t/ tokens

This section aims at determining whether initial /t/ tokens that were not shadowed can be affected in the post-test condition. If so, it will indicate that imitation can carry-over effect to non-shadowed tokens in the same position. Given the results reported in the previous section, this is expected. We also test whether, if non-shadowed tokens are affected, whether they are affected to the same degree as tokens which were previously shadowed.

We took pre-test and post-test data only and tested for a possible interaction between condition, trial number and whether words were shadowed or not. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we included 3 independent variables: condition, trial number (a continuous variable numbering the first trial to the last for each condition) and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects. Removed from the models were (i) non-significant interactions and (ii) non-significant main effects which were not present in a significant interaction.

Table 6.13 shows the output of the model predicting closure duration. The post-test condition is significant and the slope is positive. There is a significant effect for tokens that were previously shadowed and the slope is positive. This means that closure duration is longer for initial /t/ tokens than for medial /t/ tokens.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0653	0.000
post-test	0.0135	0.000
shadowed$_{true}$	0.0069	0.0181

Table 6.13 – Model output. DV: closure duration

Table 6.14 shows the output of the model predicting CPratio. The post-test condition is significant and the slope is positive. These results show that all tokens increased in the post-test condition in terms of CPratio. Since the interaction coefficient is significant and the slope is positive, this means that the CPratio for initial tokens that were shadowed increased even more than the medial /t/ tokens (that were not shadowed) in the post-test condition.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0654	0.0000
post-test	0.0183	0.0000
shadowed$_{true}$	0.0065	0.0734
post-test:shadowed$_{true}$	0.0031	0.0455

Table 6.14 – Model output. DV: CPratio

The results show carry-over imitation effects to a different position: medial /t/ tokens that were not shadowed increased in terms of raw closure duration as well as in terms of the temporal space occupied by closure over the whole length of the carrier phrase. Previously shadowed tokens showed an even stronger effect.

6.3.4 Experiment 2b: results for the English speakers

This experiment investigates whether imitation effects in initial position, if they are to be observed in the test condition, can generalise to new words in a different position, namely in medial position.

Speech patterns of the speakers

Table 6.15 presents details about the speakers' speech patterns prior to their exposition to the stimuli. Both the fricated variant of /t/ and canonical stops occur in their speech. As expected from our results on word list data in chapter 2, the fricated variant occurs more often than the canonical variant for most speakers.

speaker	medial /t/ fricatives	medial /t/ stops
participant 1	8	7
participant 2	12	3
participant 3	14	1
participant 4	10	4
participant 5	1	13
participant 6	11	1
participant 7	11	2
participant 8	4	8
participant 9	10	2
participant 10	14	0
participant 11	14	1
participant 12	13	2
participant 13	13	1
participant 14	11	4
participant 15	5	8
participant 16	12	3
participant 17	13	1
participant 18	10	3
participant 19	11	3
participant 20	11	2
total	208	69

Table 6.15 – Raw counts of medial fricatives and stops in the pre-test condition produced in experiment 2b

Carrier sentence duration

Figure 6.4 shows the duration of the carrier phrases across conditions. Data were submitted to a linear mixed-effects model. The dependent variable was phrase duration, a continuous effect. We specified one independent variable as a fixed effect: condition (a 3 level factor including pre-test, test and post-test conditions). Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

Table 6.16 shows that both slopes are negative and significantly different from the pre-test condition. The slope for the test condition is steeper than the slope for the post-test condition. This result indicates that phrase duration decreases in the post-test condition and even more so in the test condition. In other words, as phrase duration can be used as an indicator of speech

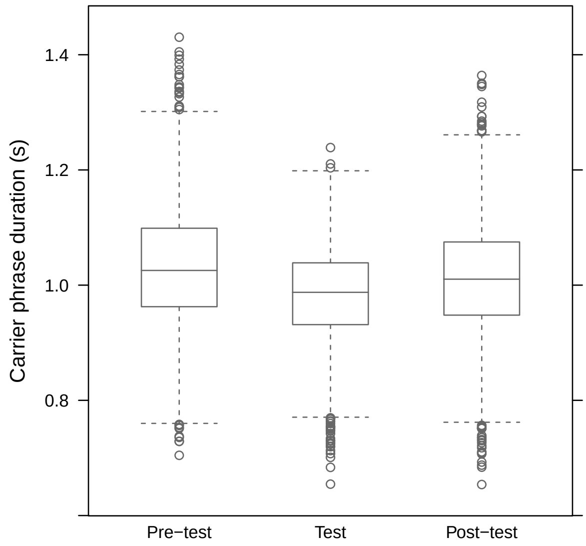

Figure 6.4 – Phrase duration across conditions.

	Estimate	$Pr(>\|t\|)$
(Intercept)	1.0371	0
post-test	-0.0267	0
test	-0.0625	0

Table 6.16 – Model output

rate, participants can be said to speak faster in the test and post-test conditions compared to the pre-test condition.

Duration of shadowed initial /t/ tokens

This section aims at determining if speakers are able to extend the duration of shadowed initial /t/ tokens in the test condition. If so, it will indicate that spontaneous imitation effects can take place. We also investigate whether tokens that have been shadowed are lengthened in the post-test condition. If so, it will indicate that imitation effects can persist.

We isolated the tokens labelled as shadowed (*ie.* the words containing an initial /t/ and that the participants were asked to shadow). Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we tested for an interaction between condition and trial number. Pre-test condition was selected as the reference level. Trial number is a continuous variable numbering the first trial to the last for each condition. We specified speakers and words as random effects. Removed from the models were (i) non-significant interactions and (ii) non-significant main effects which were not present in a significant interaction.

Table 6.17 shows the output of the model predicting closure duration. Test and post-test conditions were significant and their slopes are positive. The slope for the test condition is steeper than the slope for the post-test condition. This means that closure duration for shadowed tokens increased both in the test condition and in the post-test condition compared to the pre-test condition and that the size of the effect is greater in the test condition.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0691	0
post-test	0.0045	0
test	0.0108	0

Table 6.17 – Model output. DV: closure duration

Table 6.18 shows the output of the model predicting closure CPratio. Test and post-test conditions were significant and their slopes are positive. The slope for the test condition is steeper than the slope for the post-test condition. This means that CPratio for shadowed tokens increased both in the test condition and in the post-test condition compared to the pre-test condition and that the size of the effect is greater in the test condition. Also, the carrier phrase duration significantly decreased from pre-test condition to test condition, as shown in section 6.3.4, thus the temporal space occupied by consonant closure is even greater when phrases are shorter.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0666	0
post-test	0.0058	0
test	0.0151	0

Table 6.18 – Model output. DV: CPratio

These results provide evidence for spontaneous imitation effects and persistent carry-over imitation effects for the tokens which were shadowed previously in terms of raw closure duration as well as in terms of the temporal space occupied by closure over the whole length of the carrier phrase.

Duration of non-shadowed medial /t/ tokens

This section aims at determining whether initial /t/ tokens that were not shadowed can be affected in the post-test condition. If so, it will indicate that imitation can carry-over effect to non-shadowed tokens in the same position. Given the results reported in the previous section, this is expected. We also test whether, if non-shadowed tokens are affected, whether they are affected to the same degree as tokens which were previously shadowed.

Results for /t/ tokens realised as stops:

We took pre-test and post-test data only and tested for a possible interaction between condition, trial number and whether words were shadowed or not. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we included 3 independent variables: condition, trial number (a continuous variable numbering the first trial to the last for each condition) and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects. Removed from the models were (i) non-significant interactions and (ii) non-significant main effects which were not present in a significant interaction.

Table 6.19 shows the output of the model predicting closure duration. The post-test condition was significant and the slope is positive. There was a significant effect for tokens that were previously shadowed and the slope is positive. This means that closure duration is longer for initial /t/ tokens than for medial /t/ tokens.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0388	0
post-test	0.0052	0
shadowed$_{true}$	0.0299	0

Table 6.19 – Model output. DV: closure duration

Table 6.20 shows the output of the model predicting CPratio. The post-test condition was significant and the slope is positive. There was a significant effect for tokens that were previously shadowed and the slope is positive. This means that closure duration is longer for initial /t/ tokens than for medial /t/ tokens.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0373	0
post-test	0.0065	0
shadowed$_{true}$	0.0289	0

Table 6.20 – Model output. DV: CPratio

The results show carry-over imitation effects to a different position: medial /t/ tokens that were not shadowed and were realised as stops increased in terms of raw closure duration as well as in terms of the temporal space occupied by closure over the whole length of the carrier phrase.

Results for /t/ tokens realised as fricatives:

We took pre-test and post-test data only and tested for a possible interaction between condition, trial number and whether words were shadowed or not. Data were submitted to two linear mixed-effects models. In the first model the dependent variable was closure duration. In the second model the dependent variable was CPratio. In both models we included 3 independent variables: condition, trial number (a continuous variable numbering the first trial to the last for each condition) and shadowed (a binary variable coded as shadowed$_{true}$ or shadowed$_{false}$). Pre-test condition and shadowed$_{false}$ were selected as reference levels. We specified speakers and words as random effects. Removed from the models were (i) non-significant interactions and (ii) non-significant main effects which were not present in a significant interaction.

Table 6.21 shows the output of the model predicting predicting fricative duration. The post-test condition was significant and the slope is positive. This means that duration of fricatives increased in the post-test condition compared to the pre-test condition.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0818	0.0000
post-test	0.0033	0.0024

Table 6.21 – Model output. DV: fricative duration

Table 6.22 shows the output of the model predicting predicting CPratio for fricatives. The post-test condition was significant and the slope is positive. This means that CPratio for fricatives increased in the post-test condition compared to the pre-test condition.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.0799	0
post-test	0.0055	0

Table 6.22 – Model output. DV: CPratio

The results show carry-over imitation effects to a different position and to a different allophone: medial /t/ tokens that were not shadowed and were realised as fricatives increased in terms of raw duration as well as in terms of the temporal space occupied by the duration of the consonant over the whole length of the carrier phrase.

Duration of tokens other than /t/ tokens

Since our results have shown that imitation affected medial /t/ tokens realised as fricatives, we might question whether imitation could also affect fricatives that are in medial position but are not allophones of /t/. Given that there are very few fricative phonemes in medial position in our experimental design we cannot address this question directly. However if priming participants with lengthened closure duration affected a wide range of consonants, even those allophonically unrelated, then we should expect overall lengthening of consonants in the post-test condition. The words used in our experimental design have /f/, /k/, /n/ and /ʃ/ in initial position, and provide enough tokens to address this question.

Figure 6.5 presents closure duration for stops and consonant duration for nasals and fricatives in the pre-test and post-test conditions. Data were submitted to two linear mixed-effects models to test for an effect of condition for each phoneme. In the first model for each phoneme the dependent variable was closure duration for /k/ and overall duration of /n/, /f/ and /ʃ/. In the second model for each phoneme the dependent variable was CPratio. In both models we included one independent variable: condition. Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

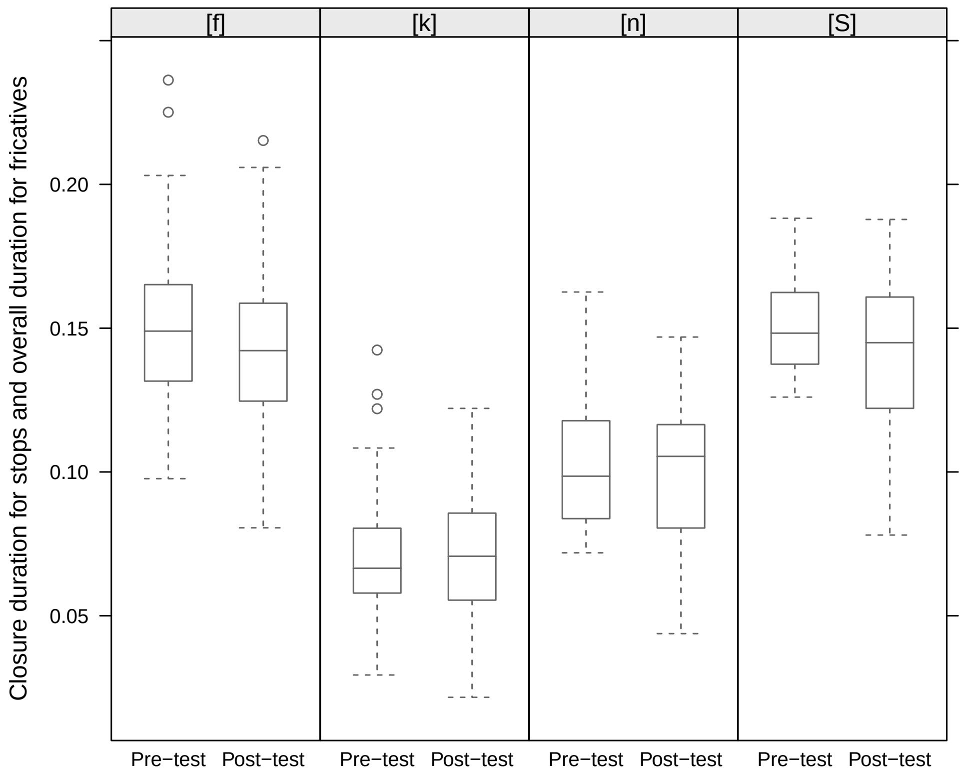

Figure 6.5 – Closure duration for stops and consonant duration for nasals and fricatives.

The models for /k/, /n/ and /ʃ/ did not yield a significant effect of condition. There was a significant effect of condition for /f/ when predicting duration and CPratio, as shown on tables 6.23 and 6.24. The slope was negative which indicates a decrease in duration and CPratio for initial /f/ tokens in the course of the experiment. Thus these results show that priming participants with lengthened closure duration of /t/ tokens did not affect other consonants.

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.1497	0.0000
post-test	-0.0086	0.0001

Table 6.23 – Model output for /f/. DV: duration of the consonant

	Estimate	$Pr(>\|t\|)$
(Intercept)	0.146	0.0000
post-test	-0.006	0.0038

Table 6.24 – Model output for /f/. DV: CPratio

Allophonic variation

In chapter 5 we showed that participants who were presented with stops only during the test condition and, overall, produced significantly more stops in the test condition than in the pre-test condition. We concluded that imitation effects could drive allophonic selection. Participants in group A, who were trained on initial canonical stops, did not contribute to this effect, as they produced fewer stops in the test condition. However participants in group B and group C both produced more canonical stops in the test condition. Moreover, this effect carried over to the post-test condition for participants in group C who shadowed final /t/ stops.

In this section we examine whether participants who shadowed initial /t/ stops, produced more medial stops in the post-test condition. Data was subset so that initial stops were excluded from the analysis (since initial /t/ is always realised as an aspirated stop in English) and submitted to a logit mixed-effects model. Thus we analysed medial realisations only. The dependent variable was a binary effect coded as "stop" or "other". We included one independent variable : condition, a two level factor including pre-test and post-test condition. Pre-test condition was selected as the reference level. We specified speakers and words as random effects.

There was no effect of condition and thus no increase in the use of the canonical variant in the post-test condition compared to the pre-test condition.

6.3.5 Summary of the results

In all experiments phrase duration decreased in the post-test condition and even more so in the test condition. In other words, as phrase duration can be used as an indicator of speech rate, participants can be said to speak faster in the test and post-test conditions compared to the pre-test condition. These results are relevant for the analysis of CPratio because an increase in CPratio would indicate that the temporal space occupied by a consonant becomes greater as phrases are shorter.

We also found that both in French and in English, stop closure duration is longer for initial /t/ than medial /t/ (as reported in tables 6.13 and 6.19).

Table 6.25 summaries the results in terms of raw duration and CPratio. For simplicity the description of our results does not include models in which trial number was tested since overall the effects are the same with or without trial number.

	test condition		post-test condition	
	raw duration	CPratio	raw duration	CPratio
Experiment 1a (French)				
shadowed initial /t/ stops	+	+	+	+
non-shadowed initial /t/ stops	NA	NA	+	+
Experiment 2a (French)				
shadowed initial /t/ stops	+	+	+	+
non-shadowed medial /t/ stops	NA	NA	+	+
Experiment 1b (English)				
shadowed initial /t/ stops	+	+		+
non-shadowed initial /t/ stops	NA	NA		+
Experiment 2b (English)				
shadowed initial /t/ stops	+	+	+	+
non-shadowed medial /t/ stops	NA	NA	+	+
non-shadowed medial /t/ fricatives	NA	NA	+	+

Table 6.25 – Summary of each mixed-effect model run on /t/ tokens. A plus sign indicates an significant increase in duration or in CPratio compared to the participants' baseline in the pre-test condition. Empty cells indicate no significant change. NA's indicate that the tests were not available.

French speakers in experiment 1a significantly increased raw closure duration and CPratio for initial /t/ tokens in the test condition as well as in the post-test condition, which means that spontaneous effects took place and persisted throughout the experiment. They also significantly increased raw closure duration and CPratio for initial /t/ tokens that were not previously shadowed, which means that imitation effects generalised to new /t/ tokens in the same position.

French speakers in experiment 2a significantly increased raw closure duration and CPratio for initial /t/ tokens in the test condition as well as in the post-test condition, which means that spontaneous effects took place and persisted throughout the experiment. They significantly increased raw closure duration and CPratio for medial /t/ tokens that were not previously shadowed, which means that imitation effects generalised to new /t/ tokens in a different position. Also the size of the effect was larger for previously shadowed initial tokens than non-shadowed medial ones in the post-test condition.

New Zealand English speakers in experiment 1b significantly increased raw closure duration and CPratio for initial /t/ tokens in the test condition, which means that spontaneous effects took place. They significantly increased CPratio for initial /t/ tokens in the post-test condition, which means that spontaneous effects persisted throughout the experiment. They also significantly increased the CPratio for initial /t/ tokens that were not previously shadowed, which means that imitation effects generalised to new /t/ tokens in the same position.

New Zealand English speakers in experiment 2b significantly increased raw closure duration and CPratio for initial /t/ tokens in the test condition as well as in the post-test condition, which means that spontaneous effects took place and persisted throughout the experiment. They significantly increased raw closure duration and CPratio for medial /t/ tokens realised as stops that were not previously shadowed. Finally they significantly increased raw duration and CPratio for medial /t/ tokens realised as stops that were not previously shadowed. Thus imitation effects generalised to new /t/ tokens realised as stops as well as fricatives and in a different position.

Given the results obtained in experiment 2b we further investigated whether imitation could also affect a wide range of consonants and tested whether /f/, /k/, /n/ and /ʃ/ tokens in initial position were lengthened. Our results showed that these consonants were not affected by imitation effects.

Finally, we investigated in experiment 2b whether participants would produce more stops in medial position, since participants were exposed to canonical stops in the test condition. Our results showed that it was not the case.

6.4 Discussion

In all experiments, the duration of initial /t/ tokens in the test condition significantly increased compared their counterparts in the pre-test condition. Our experimental design is therefore able to generate spontaneous imitation effects in a robust manner, replicating our findings in the previous chapter and similar findings in the literature on imitation in speech using the shadowing paradigm or modified versions of the paradigm (*eg.* Fowler et al., 2003; Honorof et al., 2011; Nielsen, 2011)

Carry-over effects took place for items that were not shadowed and which were in the same position (experiments 1a and 1b) since their duration significantly increased in the post-test condition. This also relates to Nielsen's finding that speakers can generalise fine-grained phonetic details to new instances which share the same phoneme or the same feature.

Carry-over effects took place for items that were not shadowed and which were not in the same position (experiments 2a and 2b). Interestingly the duration of medial /t/ fricatives in English increased as well as the duration of medial /t/ stops. Also carry-over imitation effects did not affect other consonants in different position.

These sets of results provide a novel contribution to the study of imitation in speech: through imitation, participants trained on a phoneme in initial position with extended duration can extend the duration to other positions in the word. This across-the-board phenomenon adds to the current literature on phonetic imitation in speech and in particular to Nielsen's findings that speakers can generalise to new instances which share the same phoneme or the same feature, in the same word position.

Theoretical implications of are results are discussed in the next chapter.

Chapter 7

General discussion and conclusion

7.1 Summary of the results

In this dissertation we investigated imitation and allophony according to two paradigms. The first one focuses on imitation and allophony within word positions. The second focuses on imitation and allophony across word positions.

7.1.1 Allophonic imitation within a position in the word

In order to investigate this paradigm, we asked: to what degree do manipulations of one phoneme extend across variants in the same position? if a speaker can learn to produce an unusual variant, would this ability influence the production of other allophones of the underlying phoneme? And if so, will the pre-existing allophones be produced more like the unusual variant? Would imitation of a variant drive allophonic selection?

We found that manipulation of one phoneme can extend across variants in the same position. In the experiments reported in chapter 4 speakers shadowed a short fricative that they categorised as /t/ in medial position. In order to converge towards the shadowed target they lowered the articulatory centre of gravity of their stops (as measured on the EPG signal) and they also shortened the duration of their stops. Conversely we reported in chapter 5 that speakers presented with long /t/ stops in medial position lengthened their /t/ fricatives in the same position (they were the participants in group B).

In chapter 4 we induced speakers to produce an usual variant, which verified the claim by Honorof et al. (2011) that speakers can reproduce unrehearsed motor routines in a shadowing task. Also, and as mentioned in the previous paragraph, this affected other allophones such that stops became more similar to the shadowed short fricatives.

Two speakers in chapter 4 produced more fricatives than they typically would in the baseline condition after shadowing a short fricative. Similarly participants in group B and group C in chapter 5 produced more stops than they typically would in the baseline condition after shadowing stops. Moreover the effect persisted in the post-test condition for participants in group C. Note that participants in group B and C were trained on medial and final position respectively. The effect was not found for participants in group A who were trained with stops in initial position because they only used stops in this position *ie.* they were already at ceiling level. Therefore our results show that imitation can drive allophonic selection.

To make these effects clear to the reader we present the results for the participants in group B (from chapter 5) in figure 7.1. Gaussian-like curves represent the duration distribution of speech sounds, with shorter values on the left and higher values on the right. Presenting speakers with long medial stops (dashed lines) is imitated and causes their habitual distribution (black lines) to shift to the right. This effect further spreads to the realisation of medial fricatives (noted F). Taps and glottal articulations are greyed out because their duration was not measured and will we comment on this in the section on future research directions. Note that we also represented in this figure the finding that allophonic selection is driven by exposure to a particular variant, such that stops become more prevalent than other allophones (which is symbolised by a bigger circle).

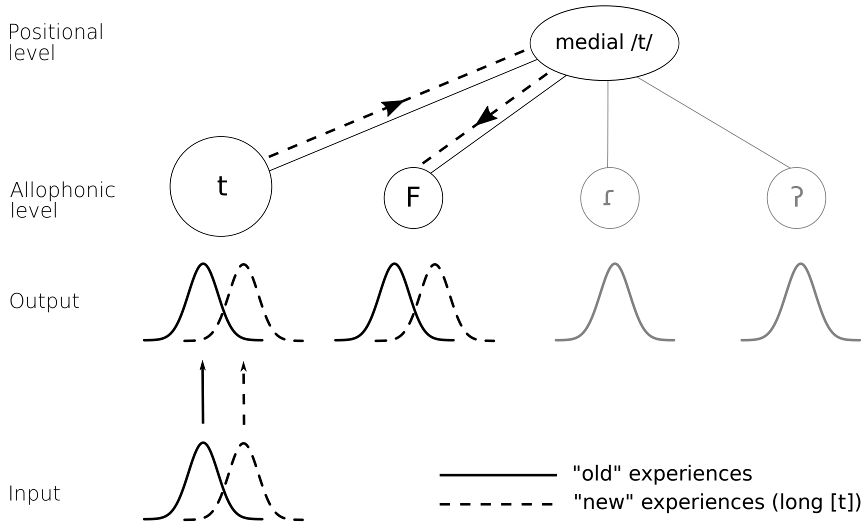

Figure 7.1 – Imitation for a given word position. Here the imitation of a long [t] stop variant in medial position affects the duration of fricatives (noted F) in the same position. It also affects allophonic selection such that stops become more prevalent than other allophones (this is symbolised by a bigger circle). Taps and glottal articulations are greyed out as their duration was not measured.

7.1.2 Allophonic imitation across word positions

In order to investigate this paradigm, we asked: is imitation in speech position sensitive? To what degree do manipulations of one phoneme extend across positions?

Chapter 5 and 6 report experiments aimed at answering these questions. In chapter 5 we found spontaneous imitation effects when participants were trained on a given position. We did not find genuine carry-over effects to different positions. We hypothesised that it could have been due to two possible reasons: (i) our experimental did not facilitate carry-over effects to a different position, or (ii) the allophonic make-up of New Zealand English inhibited carry-over effects.

To tease apart the factors that could have been responsible for our null results, we simplified our experimental design drastically and compared NZE speakers to Standard French speakers in chapter 6. Our reasoning was that French had a greater potential to show positional transfer than NZE because there is not much allophonic variation in SF (/t/ is said to be realised as a stop in all positions). We found that speakers who were trained in initial stops lengthen their medial stops in both languages. Positional transfer can thus occur in both languages. Interestingly, NZE speakers also lengthened their medial fricatives, suggesting that allophones are quite tightly cognitively linked.

The general lengthening observed in experiment 2b for NZE speakers and reported in chapter 6 is represented in figure 7.1. Here a lengthened word initial [t^h] is imitated and causes the distribution of other allophones to shift in medial position. Fricatives and stops are equally affected. Taps and glottal articulations are greyed out because their duration was not measured and will we comment on this in the section on future research directions.

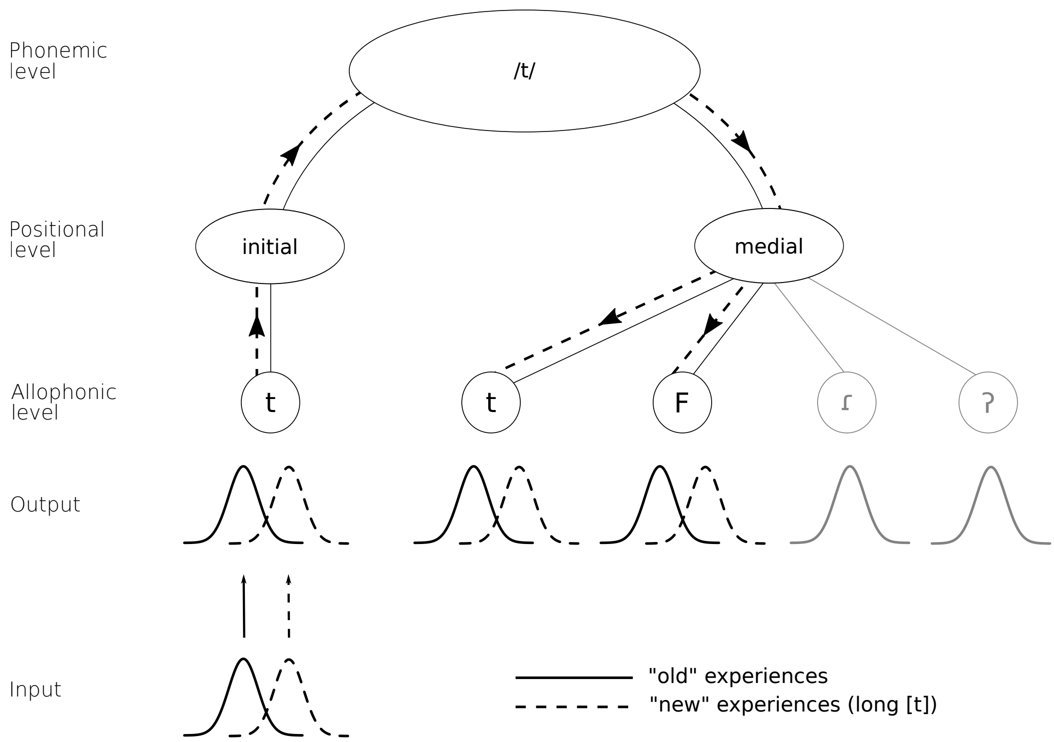

Figure 7.2 – Imitation across word positions. Imitation of a phonetic variant in initial position affects the distribution of other allophones in a different position. Taps and glottal articulations are greyed out as their duration was not measured.

7.1.3 Summary

In this section we list all the research questions stated in our introduction and we briefly state for each one what the answer is that our dissertation provided. Section 7.2 will provide insights into our main research questions, namely: (i) how is allophonic variation represented at an abstract level?, and (ii) to what degree are different phonetic categories (allophones) cognitively linked to one single abstract category (the phoneme)?

- to what degree do manipulations of one phoneme extend across variants in the same position? Shadowing a short medial /t/ fricative has the effect of shortening the duration of medial stops (although we found that the effect was speaker specific in our EPG experiment). Conversely shadowing long /t/ stops in medial position has the effect of lengthening the duration of /t/ fricatives in the same position (chapter 5).
- if a speaker can learn to produce an unusual variant, would this ability influence the production of other allophones of the underlying phoneme? And if so, will the pre-existing allophones be produced more like the unusual variant? Yes it is. Speakers presented with a short /t/ fricative reproduce some of the acoustic and articulatory properties of the usual variant. Moreover /t/ tokens produced as stops in our EPG experiment had a lower ACoG and a shorter duration, which resembles the properties of the novel allophone that they shadowed.
- would imitation of a variant drive allophonic selection? Yes it does. In our EPG experiment, some speakers produced more fricatives than they typically do in the baseline condition when they are exposed to fricatives in the test condition. Similarly speakers produced more stops than they typically do in the baseline condition when they are exposed to stops in the test condition (chapter 5).
- is imitation in speech position sensitive? Yes it is. In chapter 6 we showed that imitation can generalise from /t/ stops in initial position to /t/ stops in medial position.
- to what degree do manipulations of one phoneme extend across positions? Shadowing a long initial /t/ stop has the effect of lengthening the duration of medial stops both in French and in English and it also has the effect of lengthening medial /t/ fricatives in English (chapter 6).

7.1.4 Additional contributions to the literature

While the main focus of this dissertation is imitation in speech, we further documented /t/ allophony in New Zealand English. In chapter 2 we reported a large amount of pre-aspiration for both stop and fricative realisations of /t/ in medial position. This is interesting because pre-aspiration is said to be rare diachronically and synchronically across human languages given its lack of phonetic salience (Silverman, 2003). Furthermore, we found some evidence suggesting that the use of pre-aspiration might be socially conditioned.

We also confirmed the claim by Taylor (1996) that the prestige medial /t/ variant in New Zealand English is a fricative. Analyses of large corpora revealed that the fricative has entered NZE through the speech of females and has strongly established itself in the dialect. In parallel we conducted analyses on T-voicing and confirmed the claim by Holmes (1994) that it has entered NZE through the vernacular style of working-class male speakers and has increasingly established itself in middle class speech.

In chapter 3 we conducted a palatographic pilot experiment to further document the fricative variant. We found similar articulatory patterns to those observed in Irish English by Pandeli et al. (1997). Our data was gathered from one speaker and thus it is difficult to make generalisations. More articulatory data is needed in order to better understand the gestures involved in the production of the consonant.

We also conducted a perception experiment by reducing the duration of /s/ and /ʃ/ in medial position. We replicated findings by Grimm (1966) and Jongman (1989) such that correct identification of /s/ and /ʃ/ decreases as the duration of fricative noise decreases. Listeners had a percept closest to /t/ when the duration of the fricatives was shortened. We further reported that older and younger listeners performed differently in the task, with older listeners having poorer identification scores. This is in accordance with the literature on ageing which shows that older listeners tend to perform less accurately and slower than younger listeners (Pichora-Fuller and Souza, 2003; Janse, 2009; Adank and Janse, 2010).

7.2 Theoretical discussion of the results on imitation

In this section we first discuss the results of the EPG experiment reported in chapter 4. Then we discuss the results on positional transfer from chapters 5 and 6. Results are discussed with regards to direct realism and hybrid exemplar-based models introduced in chapter 1.

7.2.1 Theoretical account of the EPG results

In chapter 4 we asked the following question: would shadowing a shortened /s/ or /ʃ/ lead to the production of a speech sound which is indistinguishable from a shadowed genuine /t/ or would it contain properties that would more closely resemble the shadowed target?

We found imitation for two speakers out of four. Those two speakers converged in terms of duration, place and manner of articulation. Since the sounds do not occur naturally this shows that speakers (i) can recover articulatory events from the acoustic signal without prior experience to them and (ii) reproduce unrehearsed motor routines. This extends findings by Honorof et al. (2011) who presented naturally occurring sounds in an unexpected word position. It also supports their claim that a gestural account can explain the observed results. However the two other speakers did not show any imitation in duration or in place of articulation[1]. These results are not in line with a gestural account.

Moreover, in an informal debriefing session conducted after the experiment, all participants reported hearing "retracted articulations" upon listening to short /s/ or /ʃ/ stimuli. Speakers F1 and M2, who imitated the stimuli, did not report performing any task other than repeating what they heard. Speakers F2 and M1, who did not imitate, reported performing a more complex task involving a stage of phonemic categorisation first and then a stage of production.

It is difficult to account for the discrepancies in our results using a gestural approach to speech perception and production. Gestural theories assume the ability to perceive gestures and a tight coupling between production and perception. As far as gestures are concerned, it seems from our informal interviews with the participants that they could perceive them. However, according to speakers F2 and M1, an intermediate stage of phonological categorisation was involved, which goes against the assumption of a tight production-perception link. Findings by Mitterer and Ernestus (2008), Mitterer and Müsseler (2013) and Nielsen (2011), reported in section 1.2.2, suggest that the link between perception and production is not as tight as direct realism claims. Our results support this conclusion. Moreover the direct realist approach explicitly predicts imitation (Fowler et al., 2003; Honorof et al., 2011) and aside from the doubts we expressed about a tight coupling between perception and production, it is not quite clear how the theory can account for null results.

Fowler and Galantucci (2005) note that, traditionally, the speech production system and the speech perception system have been studied independently, and that few theories have been put forward in order to investigate how one system could inform the other. We could not agree more and we believe that more effort should be put into this theoretical direction. It seems from our results, and from those of Mitterer and colleagues or Nielsen who showed that phonological factors might affect imitation, that the link between production and perception can be loosely coupled. Future research that we would like to conduct will focus on this very link and ask the following questions: how loose or how tight is actually the production-perception link? Under what circumstances can this be affected?

[1] These speakers showed anecdotal imitation of manner of articulation however. But note that, as far as we know, frication of /t/ has not been reported in American English (AmE) of in Standard Scottish English (SSE) *ie.* the dialects spoken by those two speakers who produced sporadic /t/ frication. This does not mean that it is not part of the possible phonetic realisations of /t/ in their dialect. Frication of medial /t/ has been reported for many varieties of British English (*eg.* Buizza, 2010; Pandeli et al., 1997; Watson, 2007) and could possibly occur in SSE. Medial velar voiceless stops can be also lenited and realised as [x] in American English (Lavoie, 2002). Thus /t/ could perhaps be lenited as well in AmE or SSE.

The exemplar approach does not make such a strong prediction of automatic imitation as direct realism does, since it does not assume such a tight coupling between production and perception systems[2]. According to the exemplar account, all speakers stored the non-words along with detailed acoustic and articulatory information as traces in memory. For production purposes these traces are activated and compared to many other activated traces of natural phonetic occurrences of medial /t/ words stored in memory. It is unclear why some speakers would select the most typical traces for production (those speakers who did not imitate) and others would select the most atypical, retracted sounding traces for production (those speakers who did imitate). However, more linguistically-oriented exemplar approaches (eg. Pierrehumbert, 2006) do take into account multiple dimensions that could affect the outcome of our experiment, such as social factors. We suggest that the nature of our task might have biased the behaviour of the participants who did not imitate.

We sound different when we hear a recording of our voice compared to when we speak. Voice recordings played in headphones are perceived through air conduction, whereas hearing as we speak is perceived through air conduction and bone conduction, which in turn amplifies low frequencies (Tonndorf, 1972; Maurer and Landis, 1990). It is also well established that some speakers can report negative opinions about their own voice when it is recorded and played back to them (Holzman et al., 1966; Holzman and Rousey, 1966). Attitudinal factors are known to affect the propensity of imitation in shadowing tasks (*eg.* Babel, 2010, 2011) such that speakers tend to imitate less when they have negative attitudes towards the model speaker. Perhaps those speakers who did not imitate might have been affected negatively by hearing their own voice. However this is only a suggestion; we did not ask the speakers how they felt about hearing their own recorded voice during the informal debriefing session.

In summary, imitation was observed on all levels for half of the participants and no imitation was observed for the other half. A gestural account that assumes a tight production-perception link does not account for this discrepancy since all participants would be expected to imitate to at least some extent. An exemplar approach that allows a mediated production-perception link – here we suggest that it could be mediated by attitudinal factors – can still account for those speakers who refrained from performing retracted gestures that they could otherwise perceive.

7.2.2 Theoretical discussion of the results on positional transfer

In the experiments reported in chapter 5 we did not find genuine positional transfer. In chapter 6 we simplified our experimental design and found genuine positional transfer from

[2]Note that the exemplar approach does assume some coupling between the two systems, since production mirrors perception in that producing speech sounds involves averaging over activations of previously stored exemplars and selecting the average as a target for speech production. The theory however remains agnostic about how tight or how loose the coupling is.

initial position to medial position, in French as well as in English.

These results provide theoretical implications for the study of the interface between phonetics and phonology: in French, the initial and medial variants are very similar as they are realised as stops. In New Zealand English, they are quite different, with the medial /t/ able to be realised as a stop, a tap or a fricative. Medial stops were equally lengthened in both languages and the medial fricatives were also affected in English. Thus speakers' representations of speech sounds seem to be affected at a very abstract level by extended duration. Moreover the extent to which representations can be affected might not necessarily be dependent on a high degree of phonetic similarity between speech sounds.

If only medial /t/ stops lengthened in English, but not medial /t/ fricatives, this would have suggested (i) that the observed effect would have be strongly tied to previously shadowed lexical items and (ii) that both allophones were not tightly cognitively linked. This would have provided support for a more purely episodic representation of speech sounds in the mental lexicon of the speakers.

Our results may provide support to hybrid models such as the one proposed by Pierrehumbert (2006), which include both an episodic memory dimension and an abstract phonological level. An episodic memory dimension may account for the lengthening of medial stops in the same position: phonetic categories previously stored in memory are updated after exposure to a lengthened stimulus, in a given position. However an episodic memory dimension alone cannot account for the lengthening of medial stops in a different position, and cannot account for the lengthening of medial fricatives in the same position. Instead, a more abstract phonological level may mediate similar phonetic categories in different word positions (*eg.* medial stops *versus* final stops), as well as less similar phonetic categories (*eg.* stops *versus* fricatives). A phonological component is thus well-suited to explain across-the-board effects.

These results are also compatible with a direct realist approach, which does not involve *stored mental representations of things past* (Wilcox and Katz, 1981), as opposed to exemplar-based models. From an articulatory point of view, fricatives are quite different from stops, yet they both involve a lingual gesture to achieve closure (or partial closure). Those in favour of a direct realist approach might thus argue that speakers were able to directly perceive, and aimed at reproducing, the duration of the lingual gesture. From such a perspective, maintaining manner of articulation might be less crucial than achieving the desired duration for the target.

7.2.3 Summary and further considerations

Overall, hybrid exemplar-based models are best suited to explain our results. Direct realism is also compatible with most of our findings but fails to capture the null results reported in our EPG experiment.

However, hybrid exemplar models are over-powered in that they allow researchers to ascribe all categorical effects to phonology and gradient effects to episodic and phonetic components. That is, they are compatible with nearly any possible outcome in our experiments. In addition, unconstrained hybrid models risk reinforcing our pre-existing assumptions about the division of phonology as digital and categorical, and phonetics as analogue and gradient. As a result, hybrid models are best when constrained, in particular by other known contributing factors, including but not limited to: language history, environment, and social factors.

The null results reported in our EPG experiment, further add to literature challenging the direct-realist view that perception and production systems are very tightly linked (see section 1.2.2 for a review of literature). It may be more reasonable to assume that some level of phonological representation is needed to mediate the link between perception and production. Further research thus needs to address how tight or our loose the perception-production link may actually be.

7.3 Future research directions

In the future we will continue to investigate imitation in speech and more work remains to be done at the allophonic level. Our results showed that stops and fricatives could be lengthened but we do not know what influence it had on other /t/ allophones. We did not investigate flaps, taps or glottal articulations because measuring their duration accurately on the acoustic signal is difficult. Other experimental designs have to be devised to bypass such measurement issues. Then we could ask the following question: does phonetic similarity play a role in imitation? One could imagine a case where stops, fricatives, taps and flaps would all be lengthened but glottal stops would not. Such an outcome would be interesting because it would mean that imitation occurs at an even lower level than the allophonic level: the level of featural representations *ie.* in this case the level of the lingual gesture[3]

Our experiments were set in controlled laboratory settings using the shadowing paradigm, which may provide a way to study imitation in speech without a direct need to address speakers' social motivations. We favoured simple experimental designs to make the experiments clearer for interpretation but, ultimately, we would like to extend our work to more natural settings, such as conversational interactions.

Chapters 2 and 3 explored the different /t/ variants found in NZE by providing speech production and perception data as well as socio-linguistic data. Acoustic, articulatory and perceptual

[3]What counts as phonetic similarity can be problematic and needs to be clearly defined. Each piece of research on imitation needs to put forward specific measures, or dimensions, of phonetic similarity. In our thesis, we investigated phonetic similarity between productions using specific measures (*eg.* duration, manner of articulation, or articulatory centre of gravity). It is worth noting that a binary contrast, such as similar *versus* dissimilar, might not be very helpful to describe multi-modal distributions of phonetic variation. Rather, a scale of similarity might be more helpful to describe such variation.

descriptions of /t/ allophones were provided as a necessary requirement in order to investigate imitation and allophony in the subsequent chapters of the thesis. However we did not examine socio-linguistic aspects, which ought to be incorporated into future work. Hay and Drager (2007) advocate for probing listeners' uses of socio-phonetic detail in speech perception and understanding "how phonetic variation is produced, performed, and perceived in its social context". For example, speakers of different ages and socio-economic backgrounds might be expected to perceive allophones of a single phoneme differently. Thus we might expect them to perform differently during imitation tasks. Those issues go well beyond the scope of this thesis but ought to be tackled if we want to investigate the connections between socio-linguistics and phonology.

We have shown that speakers, speaking their native language, can converge towards the production of phonetic variants that are not within their typical realm of experience. However, less is known about phonetic convergence among second language learners. We do know that languages differ in their phonemic inventories, and as a result it can be difficult for learners of a second language to perceive phonetic and phonemic differences in their new language. A well-known example is the observation that Japanese learners of English have problems with perceiving English liquid consonants /l/ and /ɹ/. We would like to examine the perception and production of phonetic categories, which might sound alike to second-language learners, through the imitation paradigm in order to uncover patterns of phonetic convergence. Such experiments could provide a contribution to well-established models such as the Perceptual Assimilation Model (Best, 1995) or the Speech Learning Model (Flege, 1995).

A better understanding of the physiological mechanisms involved in pre-aspiration would also be welcome for general phonetic documentation purposes. We gathered preliminary data on pre-aspiration using manometry to evaluate pressure at the level of the pharynx (not reported in this dissertation) and we plan to use aerodynamic measurements. In parallel we will investigate the finding that the use of pre-aspiration might be socially conditioned to further contribute to the literature on linguistic variation and change.

The use of a fricative as a possible realisation of /t/ in New Zealand English also provides some interesting ground for future research. Wagner et al. (2006) presented findings against the claim that formant transitions for some fricatives do not play a role in their identification (Heinz and Stevens, 1961; Harris, 1958; LaRiviere et al., 1975). They concluded that formant transitions might play a role for listeners with spectrally similar fricatives. In this research we showed that the /t/ fricative is similar to /s/ articulatory and perceptually. NZE thus provides a means to investigate whether formant transitions for these fricatives are indeed important for spoken word recognition.

Chapter 8

Résumé de la thèse en français (summary of the thesis in French)

Cette thèse s'intéresse à l'imitation dans la parole, c'est à dire à la tendance pour un locuteur de parler de façon plus similaire à son interlocuteur. Beaucoup d'entre nous font l'expérience de ce phénomène lorsque que nous conversons avec une personne qui possède un accent différent. Certaines caractéristiques de notre propre parole peuvent changer, pour se rapprocher de celle de notre interlocuteur. L'imitation dans la parole a fait l'objet de récentes études. Notre contribution à ce type de recherches est d'étudier l'imitation au niveau allophonique, c'est à dire au niveau des réalisations phonétiques possibles d'un phonème. Nous voulons savoir si l'imitation d'un son phonétique pour un phonème donné, dans une position de mot donnée, peut influencer les autres réalisations de ce phonème, dans la même position de mot. Nous voulons également savoir si l'imitation d'un son phonétique pour un phonème donné, dans une position de mot donnée, peut influencer la réalisation d'autres allophones de ce phonème, dans une position de mot différente.

L'anglais de Nouvelle Zélande possède un inventaire allophonique riche dans chaque position de mot en ce qui concerne le phonème /t/. Ainsi, l'anglais de Nouvelle Zélande est une langue idéale pour étudier le lien entre imitation dans la parole et allophonie. Mais avant de présenter nos protocoles expérimentaux et nos résultats sur l'imitation, nous avons vérifié et étendu le travail qui a déjà été conduit sur cette langue. Nous avons analysé de larges corpus enregistrés d'anglais de Nouvelle Zélande et avons mis en évidence de nouveaux allophones pour le phonème /t/ qui n'avaient pas été identifiés auparavant. Nous avons examiné une réalisation fricative de ce phonème en particulier.

Pour en apprendre plus sur cette fricative, nous avons conduit une expérience pilote en palatographie statique et une expérience en perception de la parole. En s'appuyant sur nos résultats en perception, nous avons étudié l'imitation dans la parole en créant artificiellement une fricative en position médiale et en mesurant les effets de convergence phonétique vers ce nouvel allophone au moyen d'enregistrements acoustiques et électropalatographiques. Pour certains locuteurs, une simple exposition à ce nouvel allophone a eu pour effet d'affecter la réalisation d'autres allophones présents dans la même position de mot.

Une série d'expériences acoustiques a ensuite été conduite pour examiner les phénomènes d'imitation allophonique par delà les positions de mot. Nous avons trouvé qu'une exposition répétée à un allophone donné peut déterminer la sélection allophonique pour les autres positions de mot. Nous avons également trouvé qu'un transfert d'une position vers une autre peut se produire, de telle façon qu'une exposition à un allophone dont les caractéristiques acoustiques ont été manipulées, peut affecter les mêmes allophones dans des positions de mot différentes. Il est intéressant de noter que cela peut affecter d'autres allophones dans une position de mot différente.

Nos résultats ont fait l'objet d'interprétations dans le cadre des modèles à exemplaires hybrides et dans le cadre du réalisme direct. Dans notre discussion générale, nous avons examiné quels cadres théoriques permettent de rendre compte au mieux des résultats que nous avons obtenus dans ce travail de thèse.

8.1 Introduction

8.1.1 Imitation et convergence phonétique

Le terme d'*accommodation* renvoie à la tendance générale qui consiste à changer notre comportement communicatif en référence à nos partenaires conversationnels. Parmi les changements communicatifs qu'il est possible d'adopter, l'un d'entre eux a reçu une attention toute particulière. Il s'agit de la convergence. La convergence implique que nos comportements communicatifs s'appareillent dans une certaine mesure à ceux de nos partenaires en conversation.

Les phénomènes de convergence phonétique ont été beaucoup étudiés récemment. Ils ont été mis en évidence en interaction conversationnelle (*eg.* Pardo, 2006; Pardo et al., 2011; Aubanel and Nguyen, 2010; Lewandowski, 2012) et dans des expériences contrôlées en laboratoires (*eg.* Babel, 2010, 2011; Abrego-Collier et al., 2011; Nielsen, 2011; Kim, 2012)

En ce qui concerne la convergence phonétique en interaction conversationnelle, des théories telles que la *Communication Accommodation Theory*, CAT ci-après, (Giles et al., 1991; Giles and Ogay, 2006) ou bien l'*Interactive Alignment Model*, AIM ci-après, (Pickering et al., 2004; Pickering and Garrod, 2006) ont été utilisées pour en rendre compte. Le modèle CAT met l'accent sur les facteurs sociaux pour expliquer le comportement des locuteurs, tandis que AIM met l'accent sur des mécanismes cognitifs sous-jacents.

Pourtant, beaucoup d'expériences très contrôlées en laboratoire ont été conduites sans faire nécessairement référence aux motivations sociales ou à l'alignement interactif des locuteurs. Le plus souvent, ce genre d'études fait intervenir une tâche dite de *shadowing* dans laquelle on demande à un participant d'écouter un mot pré-enregistré et tel qu'il a été prononcé par un locuteur modèle. Le participant doit ensuite répéter ce mot.

Dans ce type de tâche, il se produit souvent des phénomènes d'imitation, et étant donné la nature asociale du paradigme expérimental (puisqu'il n'y a pas d'interaction entre deux locuteurs à proprement parler), il est assez difficile d'invoquer CAT ou AIM pour rendre compte de l'imitation observée.

Bien que la tâche de shadowing soit assez éloignée des situations naturelles de parole en interaction, les études ayant à recours ce type de paradigme ont permis de mieux comprendre comment les sons phonétiques sont représentés dans le lexique mental et d'en apprendre plus sur le lien entre production et perception. Par exemple, l'étude de Goldinger (1998) a donné du poids en faveur d'une représentation épisodique des mots dans la mémoire à l'aide de ce paradigme. De plus, les résultats obtenus en tâche de shadowing continuent d'alimenter le débat sur le lien entre production et perception, pour savoir si la phonologie sert de médiateur entre ces deux systèmes ou non (voir plus bas, sur le point de vue du réalisme direct (Fowler and Galantucci, 2005; Best, 1995) en opposition à celui de théories phonologiques).

8.1.2 Cadre théorique

Dans cette section, nous présentons les modèles à exemplaires hybrides et le réalisme direct. Nos résultats de thèse seront rattachés à ces cadres théoriques afin de voir lequel permet au mieux d'expliquer les données observées.

L'approche abstractionniste traditionnelle

Dans un cadre abstractionniste traditionnel, les sons phonétiques sont représentés au niveau phonologique comme étants invariants, abstraits et regroupés en catégories discrètes comme les phonèmes ou les traits (Chomsky and Halle, 1968). L'accès à ces catégories abstraites implique un processus de normalisation, de sorte que l'auditeur puisse filtrer la variation au niveau phonétique pour accéder aux catégories phonologiques lors de la reconnaissance des mots. Les propriétés indexicales (âge/sexe du locuteur/accent par exemple), en partie responsables de la variation phonétique, sont alors filtrées.

Le point de vue épisodique

Beaucoup d'études ont montré que ces propriétés indexicales sont en réalité pertinentes au niveau perceptif et sont stockées en mémoire (voir Nguyen et al., 2009 pour une revue détaillée). Le point de vue épisodique veut que chaque expérience que nous faisons laisse une trace dans notre mémoire. Des théories en perception de la parole (Johnson, 1997) ou en reconnaissance des mots ont ainsi vu le jour (Goldinger, 1998). Ces théories veulent que chaque mot qui a été entendu laisse une trace en mémoire (pour un certain laps de temps) avec des caractéristiques acoustiques et articulatoires détaillées, des informations sur l'identité des locuteurs et sur la situation de communication, pour ne citer que certaines dimensions. Ce type de théories est donc en contraste avec l'approche abstractionniste traditionnelle.

Aperçu des modèles à exemplaires hybrides

Bien que la pertinence des détails phonétiques fins ait pu être sous-estimée par les modèles abstractionnistes traditionnels, il ne fait aucun doute qu'un certain niveau d'abstraction soit requis lors de la reconnaissance de mots. En effet, il doit exister des catégories abstraites afin de pouvoir expliquer les phénomènes de perception catégorielle tels que ceux observés par Lisker & Abramson (1970). Certains modèles à exemplaires reconnaissent ainsi qu'une composante phonologique est nécessaire, tout en conservant l'idée que les traces de nos expériences sont stockées en mémoire (*eg.* Coleman, 2002; Pierrehumbert, 2006). Pour cette raison, ils ont reçu le nom de modèles à exemplaires hybrides.

Une approche hybride permet de rendre compte des résultats obtenus par Nielsen (2011) par exemple. Elle a montré que les locuteurs peuvent augmenter la durée d'aspiration de mots avec /k/ à l'initiale après avoir été exposés à des mots avec /p/ à l'initiale et dont la durée d'aspiration avait été augmentée par une manipulation acoustique. Nielsen conclut qu'une composante invoquant la mémoire épisodique est requise pour stocker les durées d'aspiration des sons [p] mais qu'un niveau abstrait est également requis pour que la variation subphonémique soit généralisée à des mots avec /k/ à l'initiale, car ces catégories sont liées d'un point de vue cognitif : elles appartiennent toutes deux à la classe des plosives aspirées.

En résumé, si les premiers modèles à exemplaires s'écartent de l'approche abstractionniste traditionnelle, des modèles moins extrêmes, tels que les modèles hybrides, prennent en compte un niveau d'abstraction phonologique. Dans la section suivante, nous abordons le réalisme direct qui a aussi été mis en avant pour rendre compte des phénomènes d'imitation dans la parole. Contrairement aux modèles à exemplaires, il n'implique pas "des représentations mentales stockées en mémoire pour les événements du passé" (Wilcox & Katz, 1981, notre traduction).

Une explication de l'imitation dans la parole par le réalisme direct

Le réalisme direct dans son application à la parole (*eg.* Fowler, 1986; Fowler & Galantucci, 2005; Best, 1995) s'intéresse au lien entre production et perception. Dans ce cadre théorique, les locuteurs/auditeurs sont capables de percevoir et d'utiliser directement les gestes de la parole à partir de leurs expériences sensorielles (expériences acoustiques, visuelles ou somatosensorielles par exemple). Les gestes sont définis comme des actions qui créent et relâchent des constrictions dans le tractus vocal en utilisant une combinaison de deux articulateurs ou plus. Ces actions sont pertinentes d'un point de vue linguistique et jouent le rôle de primitives phonologiques. Pour le réalisme direct, il existe un lien fort entre perception et production.

Dans une tâche de shadowing, Fowler et al. (2003) ont montré que les occlusives à l'initiale des syllabes ont des durées d'aspiration plus longues si le locuteur modèle produit des durées d'aspiration longues. Pour eux, un lien phonologique entre production et perception ne saurait rendre compte de l'imitation de détails subphonémiques. Selon les auteurs, si les locuteurs utilisaient d'abord le signal acoustique pour ensuite procéder à une tâche de catégorisation phonologique (c'est à dire dans le cas où le lien entre perception et production serait médié par un processus phonologique), alors les locuteurs ne pourraient imiter le degré d'aspiration du locuteur modèle. Au contraire, ils devraient conserver leur degré d'aspiration habituel à partir du moment où ils tentent de produire un son phonétique qui a été préalablement identifié comme appartenant à une catégorie phonologique.

Honorof et al. (2011) ont obtenu des résultats similaires dans une tâche de shadowing. En utilisant un articulomètre, ils ont mesuré les mouvements de la langue et des lèvres. Ils ont montré que les locuteurs sont capables de reproduire certains gestes articulatoires associés avec un [ɫ] sombre volontairement placé en position initiale de syllabe en anglais, alors que cet allophone se trouve habituellement en position de coda. Les locuteurs n'ont pas eu recours a un arrondissement des lèvres pour compenser au niveau acoustique, mais ont bien reproduit les positions linguales auxquelles ils ont été exposés. Les auteurs ont donc conclu que seule une approche gestuelle telle que le réalisme direct peut rendre compte de leurs résultats en imitation.

Toutefois, le réalisme direct a fait l'objet de critiques, en partie par des études utilisant le paradigme du shadowing (voir Mitterer & Ernestus, 2008 et Mitterer & Müsseler, 2013). Selon ces auteurs, il se pourrait que l'imitation soit en fait un biais introduit par le paradigme lui-même. De plus, les résultats de Nielsen (2011) ont montré que les durées d'aspiration longues étaient imitées dans le cas d'occlusives non-voisées à l'initiale de mot en Anglais tandis que les durées d'aspiration courtes ne l'étaient pas. Il se pourrait donc que l'imitation soit sélective afin de préserver le contraste phonologique entre occlusives voisées et non-voisées.

8.1.3 Apprentissage par la perception

Il est possible de moduler une catégorique phonétique par l'intermédiaire d'un entraînement en laboratoire, autrement appelé apprentissage par la perception.

Il a été montré que l'apprentissage par la perception est un phénomène robuste qui se retrouve aussi bien chez les enfants que chez les adultes, et ce dans des paradigmes expérimentaux très variés (McQueen et al., 2012, 2006b; Eisner & McQueen, 2006; Clarke-Davidson et al., 2008). C'est un phénomène durable car il persiste bien longtemps après la phase d'entraînement (Eisner and McQueen, 2006). Enfin c'est un phénomène généralisable d'une catégorie phonétique à une autre (Kraljic & Samuel, 2006, 2007), généralisable à de nouveaux mots (McQueen et al., 2006a; Maye et al., 2008; Sjerps et al., 2010; Mitterer et al., 2011), ou généralisable d'une position de mot à une autre (Jesse & McQueen, 2011).

Ces résultats en terme de généralisation font ainsi écho au travail de Nielsen (2011) en production de la parole où les locuteurs ont augmenté la durée d'aspiration de mots nouveaux et la durée d'aspiration d'une catégorie d'occlusive à une autre après avoir simplement entendu des mots dont l'occlusive à l'initiale fut prononcée avec une durée d'aspiration longue.

8.1.4 Allophonie

Notre travail de thèse porte sur les allophones du phonème /t/ en anglais de Nouvelle Zélande, qui ont été répertoriés par les spécialistes de cette variété. Dans cette section nous voulons simplement attirer l'attention du lecteur sur le fait que s'il existe aujourd'hui de nombreux manuels de phonétique et de phonologie qui permettent de guider les étudiants à identifier les allophones (en ayant recours à des méthodes telles que les tests de commutation par exemple), il existe peu de recherches portant sur les allophones et la façon dont ils sont liés au phonème d'un point de vue cognitif. De façon générale, les manuels expliquent aux étudiants que les locuteurs ont une connaissance passive de leur langue grâce à laquelle ils peuvent établir un lien entre des sons phonétiques parfois très différents et une seule et même catégorie abstraite.

Nous ne savons donc pas si le fait de manipuler les propriétés phonétiques d'un allophone aura un impact sur d'autres allophones du même phonème (*ie.* si ils sont fortement liés d'un point de vue cognitif) ou si les autres allophones resterons inaffectés (*ie.* si ils sont liés de façon ténue d'un point de vue cognitif). Ce travail de thèse a pour objectif d'en apprendre plus sur la question suivante : dans quelle mesure des catégories phonétiques différentes (les allophones) peuvent-elles être liées à une seule et même catégorie abstraite (le phonème)?

8.1.5 Nos questions de recherche

Nielsen (2011) a présenté à des locuteurs des durées d'aspiration courtes dans des mots avec

/p/ à l'initiale et cela n'a pas affecté la réalisation des productions suivantes. Elle soutient que la production de durées d'aspiration courtes a été inhibée afin de préserver le contraste entre /p/ et /b/. En revanche, exposer les locuteurs à des durées d'aspiration longues dans des mots avec /p/ à l'initiale a eu pour effet d'augmenter les durées d'aspiration (i) des mots avec /p/ à l'initiale présentés dans la phase d'entraînement (ii) de nouveaux mots avec /p/ à l'initiale qui n'étaient pas présents dans la phase d'entraînement (iii) de nouveaux mots avec /k/ à l'initiale qui n'étaient pas présents dans la phase d'entraînement. Nielsen conlut que l'aspiration est un trait subphonémique qui peut être imité et être généralisé à de nouveaux mots avec la même consonne à l'initiale ou avec une consonne plosive aspirée différente. Une interprétation possible de ces résultats peut être formulée ainsi : /p/ and /k/ diffèrent en termes de lieu d'articulation mais ils appartiennent à une classe abstraite qui est celle des plosives aspirées. En d'autres mots, on pourrait soutenir que ces consonnes sont liées à un niveau plus haut d'un point de vue cognitif, qui est celui de la classe des plosives aspirées. Ainsi, affecter un membre de cette classe peut avoir pour effet d'affecter un autre membre de la classe.

Nielsen a étudié les effets d'imitation pour des phonèmes différents dans la même position de mot. Cependant un phonème peut souvent être réalisé de façons très différentes dans un contexte de position donné en anglais (*eg.* en position médiale /t/ peut se réaliser en plosive, en tap, ou en coup de glotte). Un phonème peut aussi être réalisé de façons très différentes dans différents contextes de position en anglais (/t/ peut se réaliser en occlusive en position initiale, en coup de glotte en position médiale, ou être non-relâché en position finale). Ainsi nous soulevons les questions suivantes : (i) comment est représentée la variation allophonique à un niveau abstrait? (ii) dans quelle mesure des catégories phonétiques différentes (les allophones) peuvent-elles être liées à une seule et même catégorie abstraite (le phonème)?

Dans notre travail de thèse, nous avons essayé de contribuer à l'étude de l'imitation au niveau allophonique. Pour ce faire, nous avons étudié les effets d'imitation pour une position de mot donnée et par delà les positions de mot.

L'anglais de Nouvelle Zélande possède un inventaire allophonique riche dans chaque position de mot en ce qui concerne le phonème /t/. Ainsi, l'anglais de Nouvelle Zélande est une langue idéale pour étudier le lien entre imitation dans la parole et allophonie.

<u>Imitation allophonique pour une position de mot donnée</u>

L'imitation de détails phonétiques fins au sein d'une même position de mot a été bien étudiée (*eg.* l'imitation de durées d'aspiration en position initiale Fowler et al., 2003; Nielsen, 2011). Pourtant on peut se poser la question suivante : dans quelle mesure manipuler un allophone peut-il affecter d'autres allophones au sein de la même position de mot?

Honorof et al. (2011) ont montré que les locuteurs sont capables de reproduire des articulations qui n'appartiennent pas à leur répertoire gestuel habituel. Les locuteurs ont reproduit certains gestes articulatoires associés avec un [ɫ] sombre volontairement placé en position initiale de syllabe en anglais, alors que cet allophone se trouve habituellement en position de coda. Pour aller plus loin, est-ce que la capacité qu'un locuteur a à produire une variante inhabituelle peut influencer la production d'autres allophones du phonème étudié? Est-ce que l'articulation des allophones pré-existants peut changer au point de devenir plus similaire à cette variante inhabituelle?

De plus, Clark (2013) a montré que lorsque des locuteurs commencent à produire une variante, cela les conduit à utiliser d'avantage cette variante dans les productions suivantes. On peut se demander si un locuteur exposé à une variante dans une position donnée, sera plus enclin à choisir cette variante comparée aux autres allophones possibles dans la même position de mot. En d'autres termes, est-ce que l'imitation peut piloter la sélection allophonique?

Imitation allophonique par delà les positions de mot

Les résultats de Jesse and McQueen (2011) ont montré qu'en entraînant des auditeurs avec un phonème dans une position donnée cela peut affecter la perception de ce phonème dans une position de mot différente. Ainsi, on peut se poser les questions suivantes : est-ce que les effets d'imitation sont sensibles aux positions de mot? Dans quelle mesure manipuler un phonème peut-il se manifester en production par delà les positions de mot?

8.2 Vérification and extension du travail de documentation sur anglais de Nouvelle Zélande (analyses de corpus)

L'anglais de Nouvelle Zélande possède un inventaire allophonique riche dans chaque position de mot en ce qui concerne le phonème /t/. Nous avons analysé des enregistrement organisés sous forme de corpus qui permettent de retracer l'histoire de la langue parlée par les premiers colons jusqu'à nos jours.

Nous avons analysé les réalisations du phonème /t/ en position médiale et avons documenté l'utilisation d'un allophone qui avait peu été étudié auparavant. Il s'agit d'une réalisation en temps que fricative. Il n'est cependant pas étonnant que cet allophone soit passé en grande partie inaperçu puisque la plupart des travaux de corpus précédents ont fait appel à des écoutes auditives et que cet allophone peut facilement se méprendre pour une occlusive. Nous avons entrepris notre travail d'analyse à l'aide d'écoutes auditives et nous avons également visionner les spectrogrammes.

Des analyses statistiques ont montré que cette fricative est désormais la variante la plus utilisée en parole lue. Cet allophone a désormais un statut privilégié au sein de la langue. De plus, nos analyses statistiques ont confirmé les analyses menées par les spécialistes de cette variété d'anglais en parole spontanée : le tap est la variante la plus utilisée par les locuteurs. Pour chaque analyse que nous avons réalisé, nous avons pris en compte des données socio-phonétiques et avons expliqué en détail comment les groupes sociaux (en fonction de l'âge, du sexe et de la catégorie professionnelle des locuteurs) font usage de ces allophones.

Enfin, nous avons également identifié deux nouvelles variantes pré-aspirée de /t/ : une occlusive pré-aspirée et une fricative pré-aspirée.

8.3 Production et perception des fricatives linguales en anglais de Nouvelle Zélande

Pour en apprendre plus sur la variante fricative que nous venons de mentionner, nous avons entrepris un travail pilote en palatographie statique et en acoustique sur un locuteur néo-zélandais. Nous avons comparé cette fricative à deux autres fricatives linguales de la langue : /s/ et /ʃ/.

La réalisation fricative de /t/ a une durée plus courte que celle de /s/ et /ʃ/. Une inspection des spectrogrammes a montré que la distribution des bandes d'énergies est assez similaire entre /t/ et /s/ comparés à /ʃ/. Les durées de /s/ et /ʃ/ n'ont pas atteint de différences significatives. D'un point de vue articulatoire, les lieux d'articulation de /t/ et de /s/ sont similaires et plus antérieures comparés à /ʃ/ et les zones de contacts langue-palais sont peu nombreuses pour /t/ comparé aux autres fricatives.

Ainsi /t/ et /s/ paraissent proches d'un point de vue spectral, mais /t/ a une durée plus courte. Dans ces conditions, nous nous sommes demandé si un /s/ dont la durée serait raccourcie pourrait être perçu comme un /t/ en partant du principe que la différence entre ces deux consonnes est simplement une différence de durée. Dans une expérience en perception, nous avons donc présenté à des auditeurs néozélandais des non-mots avec des consonnes dont la durée a été raccourcie de façon incrémentielle. Cela a donné un panel de consonnes allant de courtes durées à longues durées. Nous avons inclus la fricative /θ/ de manière à établir des comparaisons entre toutes les fricatives linguales de la langue. 14 jeunes locuteurs et 14 locuteurs plus âgés ont pris part à l'expérience dans une tâche d'identification forcée.

Nos résultats ont montré que lorsque les durée de /t/ ou de /θ/ deviennent courtes, les locuteurs perçoivent toujours les fricatives correctement. En revanche lorsque les durées de /s/ et de /ʃ/ deviennent courtes, les auditeurs identifient les consonnes comme étant la consonne /t/. Nous avons également observé des différences entre les auditeurs suivant leur âge en termes de réponses correctes lorsque les durées de /s/ et de /ʃ/ sont manipulées.

8.4 Amorçage d'accommodation articulatoire dans une tâche de shadowing, une étude en palatographie dynamique

Cette expérience adresse deux de nos questions de recherches : est-ce que la capacité qu'un locuteur a à produire une variante inhabituelle peut influencer la production d'autres allophones du phonème étudié? Est-ce que l'articulation des allophones pré-existants peut changer au point de devenir plus similaire à cette variante inhabituelle?

Dans cette expérience, nous avons étudié dans quelle mesure les locuteurs peuvent converger vers différences de lieux d'articulation. Cela fait suite notre expérience en perception. Si les auditeurs méprennent un /s/ ou un /ʃ/ courts pour un /t/ et que leur tâche devient de répéter ces consonnes, vont-ils produire un /t/ dont les propriétés articulatoires sont identiques à leurs /t/ habituels, ou bien vont-ils produire un /t/ dont les propriétés articulatoires s'apparient à la cible qui leur a été présentée?

Quatre locuteurs ont participé à une tâche de shadowing dans laquelle nous avons mesuré la durée de leurs productions par des enregistrements audio et leurs contacts linguo-palataux par des enregistrements en palatographie dynamique. Nous avons présenté les consonnes /t/ (réalisées en tant qu'occlusive), /s/ et /ʃ/ en position médiale. Pour les fricatives /s/ et /ʃ/, nous avons présenté soit des versions longues soit des versions courtes.

Certains participants sont parvenus à imiter les fricatives courtes en position médiale, qu'ils ont relié à un niveau abstrait à la catégorie du phonème /t/. Tous les participants se sont exprimés sur le caractère inhabituelle de ces fricative.

Non seulement les participants peuvent imiter cette nouvelle variante de /t/ mais cela a pour effet de perturber la production de leurs occlusives habituelles, de telle sorte qu'elles sont devenues phonétiquement plus similaires à la cible qui leur est présentée. En effet, elles ont une durée plus courte et un centre de gravité articulatoire plus bas. On peut ainsi soutenir qu'exposer des locuteurs à une nouvelle variante peut provoquer des effets d'imitation mais aussi affecter la distribution des autres allophones de /t/, du moins en ce qui concerne les occlusives. Ces résultats apportent ainsi des réponses à certaines de nos questions de recherche. Ils permettent également d'étendre les résultats d'Honorof et al. (2011) qui ont montré que les locuteurs peuvent recouvrer les événements articulatoires depuis le signal acoustique. De plus ils ont produit plus de fricatives catégorisées comme étant des allophones de /t/ que d'occlusives après avoir été exposés à des /s/ et des /ʃ/ de courtes durées, alors que la variante occlusive est leur variante habituelle.

En revanche deux locuteurs sur quatre n'ont pas imité les sons auxquels ils ont été exposés, et ne sont pas non plus parvenus à reproduire ces sons ne serait-ce que partiellement. Ainsi les effets d'imitation semblent être sélectifs dans le sens où ils sont spécifiques aux locuteurs. Ces résultats posent ainsi un problème pour une interprétation en terme de réalisme direct, cadre théorique sur lequel Honorof et al. (2011) se sont appuyés pour expliquer leurs résultats. Ceci fera l'objet d'une discussion dans la dernière section de ce travail de thèse.

Dans les sections suivantes, nous présentons deux expériences qui s'intéressent à notre second paradigme de recherche : les effets d'imitation par delà les positions de mot. Ces expériences ont également été conduites pour examiner plus en profondeur les effets d'imitation au sein d'une même position de mot et étendre ainsi les résultats de la présente expérience.

8.5 Transfert de position et pilotage allophonique par imitation

Dans cette expérience en shadowing, nous avons testé si des occlusives dont la durée d'occlusion a été augmentée dans une position de mot donnée peuvent conduire à des effets d'imitation, et si ces effets peuvent se propager à des items dans la même position de mot mais qui n'ont pas été présentés dans la phase de test. Nous avons également testé si ces effets peuvent se propager à des items dans une position de mot différente.

Notre expérience précédente a montré que certains participants ont produit plus de fricatives catégorisées comme étant des allophones de /t/ que d'occlusives après avoir été exposés à des /s/ et des /ʃ/ de courtes durées. Si l'on présente aux participants de cette expérience des occlusives dont la durée d'occlusion a été augmentée, on pourrait s'attendre à ce qu'ils produisent plus d'occlusives que d'autres allophones, puisque cela pourrait diriger l'attention des locuteurs sur cette variante inhabituellement longue. Cela ferait également écho aux résultats de Clark (2013). Elle a montré que lorsque des locuteurs commencent à produire une variante, cela les conduit à utiliser cette variante en plus grand nombre dans les productions suivantes. Dans cette expérience, nous avons testé si les effets d'imitation peuvent être capables de piloter la variation allophonique.

28 locuteurs néo-zélandais ont participé à cette expérience. Ils ont été assignés soit au groupe A, dans lequel les mots de test contiennent des occlusives dont la durée d'occlusion a été augmentée en position initiale; soit au groupe B, dans lequel les mots de test contiennent des occlusives dont la durée d'occlusion a été augmentée en position médiale; soit au groupe C, dans lequel les mots de test contiennent des occlusives dont la durée d'occlusion a été augmentée en position finale. Les participants ont d'abord lu une liste de mots dans la phase de pré-test. Puis ils ont répété une partie de ces mots dans la phase de test, où la durée d'occlusion des consonnes a été augmentée. Enfin, dans la phase de post-test, ils ont lu les

même mots que dans la phase de pré-test. Pour chacune des phases, l'ordre de présentation des mots était aléatoire. Les mots étaient tous composés de deux syllabes et étaient de basse fréquence.

Nos résultats ont montrés des effets d'imitation spontanés pendant la phase de test : les participants de chaque groupe ont augmenté la durées de leurs occlusives. Ces résultats apportent une contribution nouvelle à la littérature : par imitation les locuteurs qui ont été entraînés sur une position de mot donnée peuvent changer la façon dont est réalisé un phonème dans une position de mot différente. Aucun effet d'imitation n'a été observé en phase de post-test dans la même position de mots. Dans ces conditions, il n'est pas possible de montrer que des effets d'imitation pourraient se produire dans des positions de mot différentes. En effet, un pré-requis pour démontrer un effet de transfert d'une position de mot à une autre est d'abord de montrer un effet d'imitation au sein de la même position de mot. Nous avons soutenu que la non-imitation pourrait s'expliquer par un biais méthodologique introduit par notre protocole expérimental, ou bien par des propriétés de la structure phonologique de l'anglais. Dans la section suivante nous présentons des expériences similaires mais avec un protocole expérimental simplifié. Nous avons également introduit des locuteurs français, car comme nous l'expliquons en détails dans notre travail de thèse, il serait plus possible de trouver des effets d'imitation en français dans ce type d'expérience, du à des propriétés de la structure phonologique de cette langue.

Le résultat le plus robuste que nous avons obtenu, est que les participants des groupes B et C (entraînés en position médiale et finale respectivement) ont produit significativement plus d'occlusives dans la phase de test. De plus, pour les participants du groupe C, le phénomène s'est également étendu à la phase de post-test. Ces deux groupes ont pour variante principale des consonnes non-occlusives, comme l'a montré la phase de pré-test. Puisqu'ils ont seulement été exposés à des occlusives dans la phase de test nous en avons donc conclu que l'imitation peut piloter la sélection allophonique, ce qui apporte une nouvelle contribution à la littérature. Ces résultats apportent ainsi une réponse à l'une de nos questions de recherche, à savoir que l'imitation peut piloter la sélection allophonique. On peut noter également les participants du groupe A (entraînés en position initiale) avaient déjà atteint un plafond, puisqu'en anglais /t/ est toujours réalisé comme une occlusive en position initiale.

8.6 Une étude sur le transfert de position en français et en anglais

Dans l'expérience précédente nous avons montré que l'imitation peut piloter la variation phonologique. En revanche, nous n'avons pu mettre en évidence des effets de transfert d'une position de mot à une autre. Pour des raisons expliquées en détail dans notre travail de thèse, la non-imitation pourrait s'expliquer par un biais méthodologique introduit par notre protocole expérimental, ou bien par des propriétés de la structure phonologique de l'anglais. Ainsi nous avons choisi d'inclure le français, langue pour laquelle /t/ est le plus souvent réalisé comme une occlusive, quelque soit la position de mot.

Dans cette expérience nous étudions plus en détail la question posée dans la section précédente en simplifiant notre protocole expérimental: est-ce que les effets d'imitation peuvent se propager à d'autres positions de mot?

Nous avons conduit quatre expériences pour répondre à cette question : deux en français et deux en anglais.

- Expérience 1a en anglais et expérience 1b en français : nous avons testé si le fait de répéter des occlusives dont la durée d'occlusion a été augmentée en position initiale peut conduire à des effets d'imitation en phase de test, et si ces effets peuvent être généralisés à des items qui n'ont pas été présentés en phase de test, au sein de la même position. 20 locuteurs du français standard et 20 locuteurs néo-zélandais ont participé aux expériences.
- Expérience 2a en anglais et expérience 2b en français : nous avons testé à nouveau si le fait répéter des occlusives dont la durée d'occlusion a été augmentée en position initiale peut conduire à des effets d'imitation en phase test, et cette fois-ci si ces effets peuvent être généralisés à des items qui n'ont pas été présentés en phase de test et dans une position de mot différente, en position médiale. 20 locuteurs du français standard et 20 locuteurs néo-zélandais ont participé aux expériences. Ces locuteurs n'avaient pas participé aux expériences 1a et 1b.

Les participants ont d'abord lu une liste de mots dans la phase de pré-test. Puis ils ont répété une partie de ces mots dans la phase de test, où la durée d'occlusion des consonnes a été augmentée. Enfin, dans la phase de post-test, ils ont lu les même mots que dans la phase de pré-test. Pour chacune des phases, l'ordre de présentation des mots était aléatoire. Les mots étaient tous composés de deux syllabes et étaient de basse fréquence.

Pour chacune des expériences nous avons trouvé des effets d'imitation spontanés en phase de test : les participants de chaque groupe ont augmenté la durées de leurs occlusives. Pour les participants des expériences 1a et 1b nous avons observé une augmentation de la durée des occlusives en phase de post-test. Les effets d'imitation ont ainsi été généralisés à de nouveaux mots dans la même position. Pour les participants des expériences 2a et 2b nous aussi avons observé une augmentation de la durée des occlusives en phase de post-test. Les effets d'imitation ont ainsi été généralisés à de nouveaux mots dans une position de mot différente. Il est important de noter que pour les locuteurs anglais, cela a également affecté un autre allophone en position médiale : la variante fricative. Ceci suggère que les allophones sont fortement liés d'un point de vue cognitif.

Ces résultats apportent une contribution nouvelle à la littérature : par imitation les locuteurs qui ont été entraîné sur une position de mot donnée peuvent changer la façon dont est prononcé un phonème dans une position de mot différente.

8.7 Conclusion

8.7.1 Réponses à nos questions de recherche

Dans cette section, nous répondons aux questions de recherches que nous avons posées dans notre introduction :

- Dans quelle mesure manipuler un allophone peut-il affecter d'autres allophones au sein de la même position de mot? Dans une tâche de shadowing, répéter une variante fricative courte de /t/ en position médiale a pour effet de réduire la durée des occlusives en position médiale (bien que cet effet ait été spécifique à certains locuteurs dans notre expérience en EPG). Inversement, répéter des occlusives longues en position médiale a pour effet d'augmenter la durée des fricatives de /t/ dans la même position de mot.
- est-ce que la capacité qu'un locuteur a à produire une variante inhabituelle peut influencer la production d'autres allophones du phonème étudié? Est-ce que l'articulation des allophones pré-existants peut changer au point de devenir plus similaire à cette variante inhabituelle? Des locuteurs exposés à une fricative courte créée en laboratoire peuvent reproduire certaines caractéristiques acoustiques et articulatoires de cette variante créé en laboratoire. De plus, les occlusives ont été produites avec un centre de gravité articulatoire plus bas et une durée plus courte lorsque les locuteurs ont été exposés à cet allophone inhabituelle.
- Si un locuteur est exposé à une variante dans une position donnée, sera-t-il plus enclin a choisir cette variante à la place d'autres allophones possibles dans la même position de mot? En d'autres termes, est-ce que l'imitation peut piloter la sélection allophonique? Oui. Dans notre expérience en EPG, certains locuteurs ont produit plus de fricatives en position médiale qu'ils ne le font habituellement lorsqu'ils ont été exposés à une variante fricative dans la même position de mot. De la même façon, les locuteurs ont produits plus d'occlusives en position médiale ou finale qu'ils ne le font habituellement lorsqu'ils ont été exposés à une variante occlusive dans une position respective.
- Est-ce que les effets d'imitation sont sensibles aux positions de mots? Oui. Nous avons montré dans notre dernière expérience que les effets d'imitation rencontrés dans

des occlusives en position initiale peuvent se généraliser à des occlusives en position médiale.

- Dans quelle mesure manipuler un phonème peut-il se manifester en production par delà les positions de mot? Exposer des locuteurs à des occlusives dont la durée a été augmentée en position initiale a pour effet d'augmenter les occlusives en position médiale aussi bien en français qu'en anglais. Cela a aussi eu pour effet d'augmenter la durée des fricatives en anglais.

8.7.2 Résumé de notre discussion théorique

Ici, nous résumons de façon très succincte la discussion présentée dans notre travail de thèse.

Sans entrer dans des détails dans ce résumé, il apparaît que les modèles à exemplaires hybrides puissent expliquer au mieux nos résultats, aussi bien pour notre expérience en EPG que pour nos résultats sur le transfert de position. Le cadre du réalisme direct, bien qu'il permette également d'expliquer une grande partie des résultats, ne permet pas d'expliquer pourquoi les effets d'imitation ne s'appliquent pas à tous les locuteurs dans notre expérience en EPG. Dans notre travail de thèse nous avons mis en question le lien direct qu'établit cette approche entre production et perception. Nos résultats laissent à penser qu'un niveau d'abstraction phonologique entre ces deux systèmes est nécessaire.

8.7.3 Directions de recherches futures

Nous continuerons d'étudier les phénomènes d'imitation au niveau allophonique car notre travail de thèse soulève d'autres questions. Par exemple, les résultats de notre dernière expérience ont montré que la durée des occlusives et des fricatives peut être augmentée, mais nous ne savons pas si cet effet peut se manifester pour d'autres allophones. Nous n'avons pas étudié les flaps, les taps, ou les articulations glottiques, car mesurer leur durée de façon précise à partir du signal acoustique seul est problématique. D'autres protocoles expérimentaux doivent être établis afin de contourner ce problème méthodologique. Nous pourrions alors nous demander si un degré de similarité phonétique est nécessaire entre différents allophones pour jouer un rôle dans l'imitation. Par exemple, imaginons qu'il soit possible d'augmenter la durée des occlusives, des taps, des flaps et des fricatives, mais non la durée des coups de glotte. Un tel résultat pourrait signifier que les effets d'imitation se produisent à un niveau plus bas que le niveau allophonique, c'est à dire au niveau de la représentation des traits phonétiques . Dans notre étude en particulier, les effets d'imitation se sont produit au niveau des gestes linguaux.

Le lien entre production et perception nous intéresse également, et nous continuerons d'étudier dans quelle mesure un niveau d'abstraction phonologique peut articuler ces deux systèmes.

Nous voudrions également mieux comprendre les mécanismes physiologiques impliqués dans le phénomène de pré-aspiration que nous avons identifié, afin d'apporter une contribution à la littérature en phonétique. De plus, nous voudrions savoir dans quelles mesures certains facteurs sociolinguistiques peuvent contribuer à l'utilisation, ou non, de cette caractéristique phonétique très souvent considérée comme rare dans les langues du monde.

Enfin, notre expérience en perception soulève des questions sur le rôle réellement joué par les transitions formantiques pour les fricatives linguales, rôle que nous souhaitons examiner d'avantage.

Appendix A

Palatograms

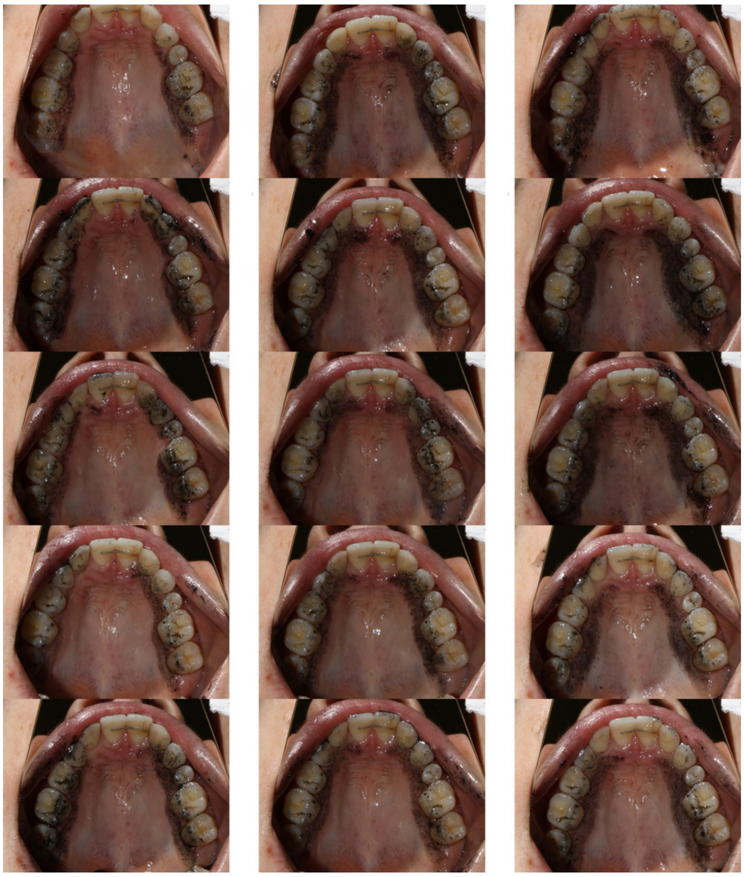

Figure A.1 – From left to right: palatograms of fricated [t], [s] and [ʃ]. Series 1: tokens 1 to 15.

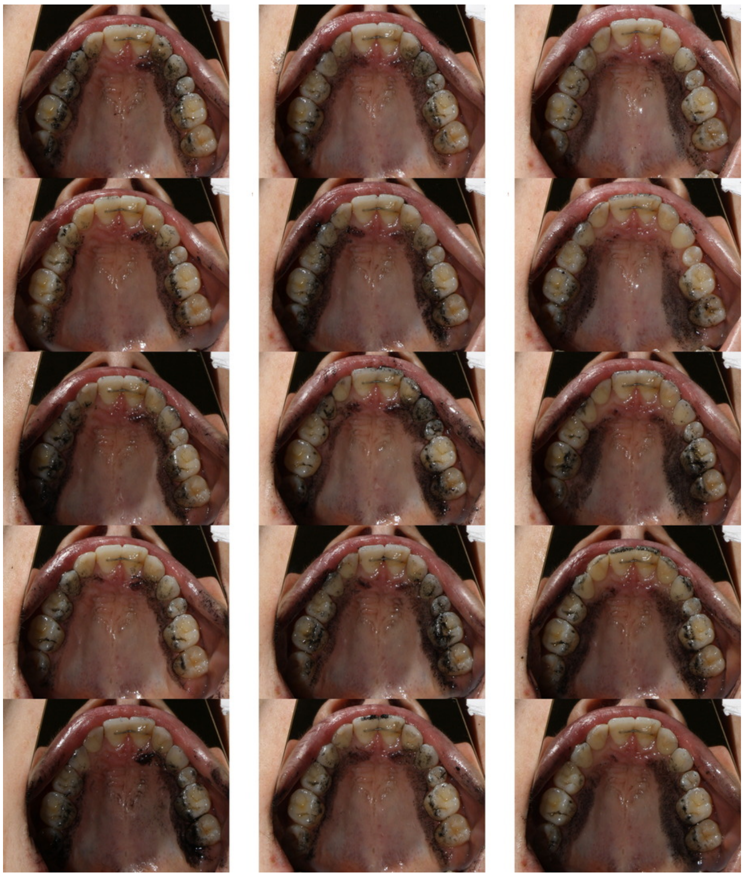

Figure A.2 – From left to right: palatograms of fricated [t], [s] and [ʃ]. Series 2: tokens 16 to 30.

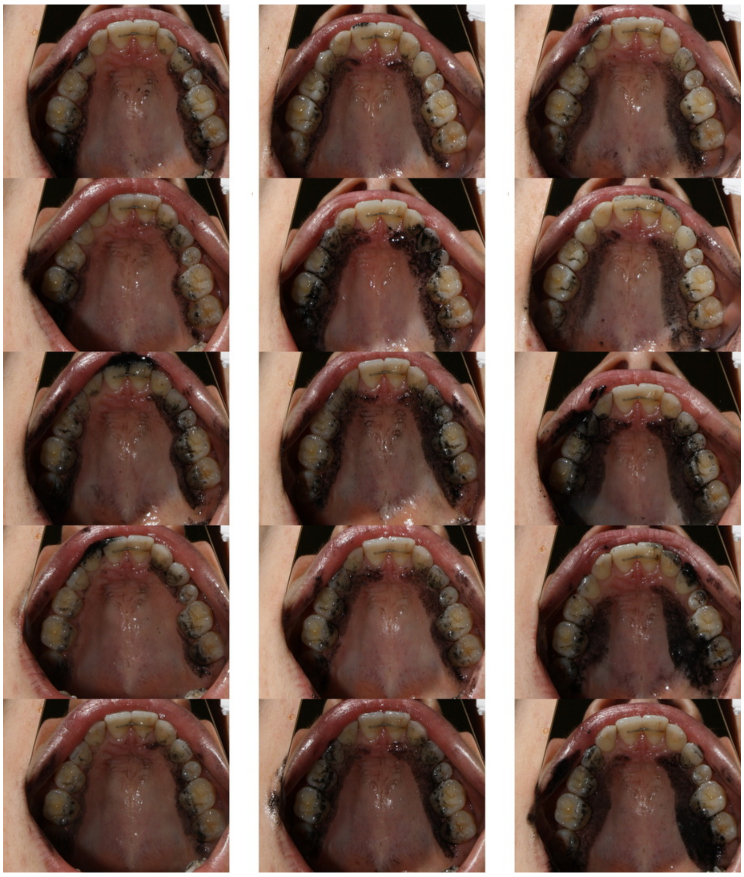

Figure A.3 – From left to right: palatograms of fricated [t], [s] and [ʃ]. Series 3: tokens 31 to 45.

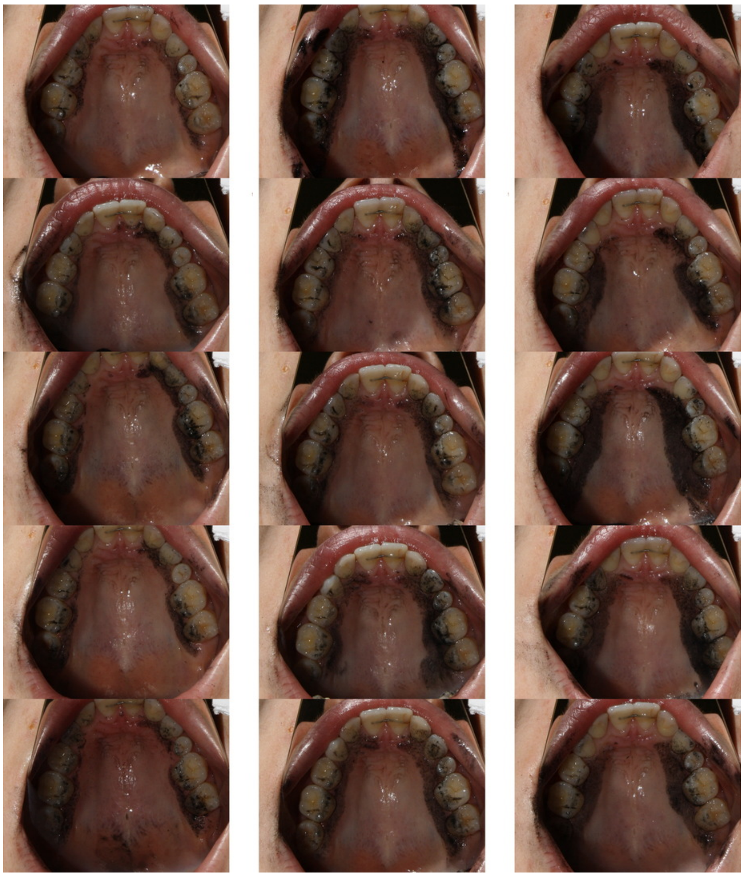

Figure A.4 – From left to right: palatograms of fricated [t], [s] and [ʃ]. Series 4: tokens 46 to 60.

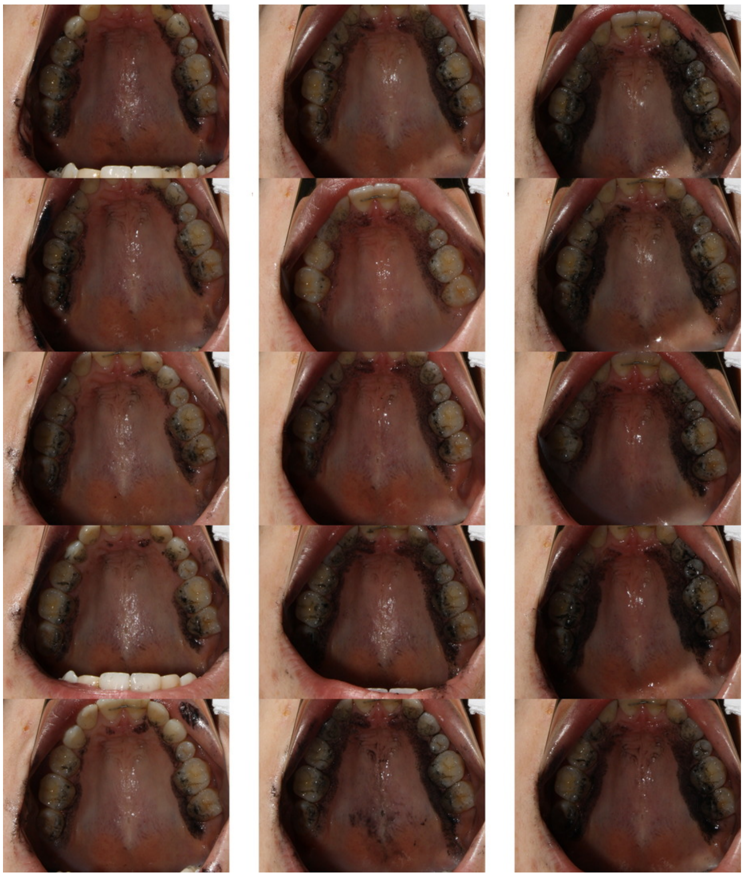

Figure A.5 – From left to right: palatograms of fricated [t], [s] and [ʃ]. Series 5: tokens 61 to 75.

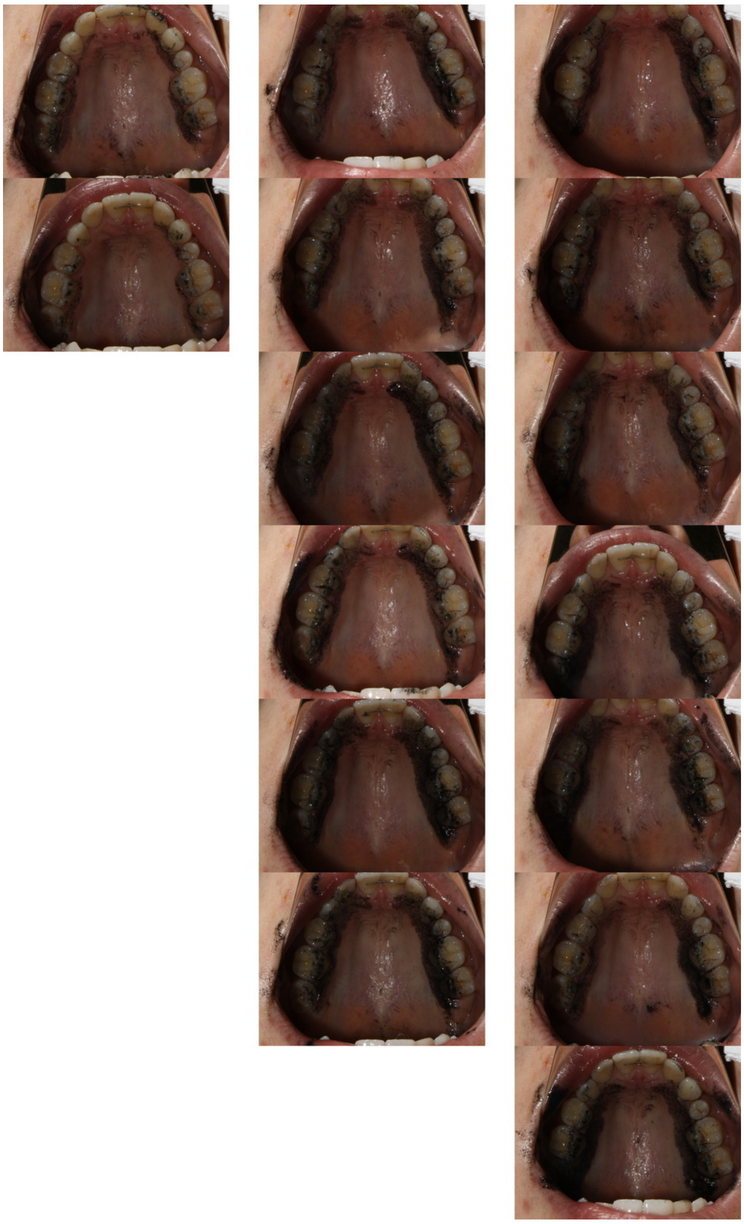

Figure A.6 – From left to right: palatograms of fricated [t], [s] and [ʃ]. Series 6: tokens 75 to 90.

Appendix B

Wordlists used in chapter 5

Initial /t/ tokens	Medial /t/ tokens	Final /t/ tokens	/d/ tokens	Fillers
tailor	coating	basket	daemon	rally
teller	cretin	biscuit	deacon	choir
telly	cutter	bonnet	dealer	snapper
tenner	cutting	bucket	demo	cleaner
tether	fritter	closet	diner	alien
ticker	knitter	concert	dingo	chorus
tiller	native	helmet	disco	kiwi
token	outage	palate	donor	ruby
turner	outing	pirate	duckling	album
turning	shutter	robot	dumper	hammer
taker	crater	bigot	glider	melon
taster	floater	carrot	header	barrow
teaser	flutter	circuit	meadow	cashew
temper	footer	comet	medic	warrior
tempo	freighter	culprit	rider	cargo
tenor	ghetto	goblet	soda	napkin
terrace	grater	hermit	spider	zombie
tester	gutter	hornet	spreader	miner
tiger	lighter	magnet	weirdo	brownie
tiler	lotus	misfit	widow	penguin
timer	motor	mullet	ballad	
tipster	nightie	peanut	blizzard	
tissue	photo	permit	comrade	
toaster	putter	pilot	custard	
toga	quitter	private	hazard	
topaz	rating	rabbit	liquid	
tummy	rotor	rocket	lizard	
turban	sweater	socket	nomad	
turkey	voter	wallet	rapid	
turnip	waiter	yoghurt	shepherd	

Table B.1 – List of words given to participants in group A. Words that participants were trained on are highlighted and were presented in all conditions. /t/ and /d/ tokens that are not highlighted were presented in pre-test and post-test conditions only. Fillers words were presented in the test condition only.

Initial /t/ tokens	Medial /t/ tokens	Final /t/ tokens	/d/ tokens	Fillers
tailor	coating	basket	daemon	rally
teller	cretin	biscuit	deacon	choir
telly	cutter	bonnet	dealer	snapper
tenner	cutting	bucket	demo	cleaner
tether	fritter	closet	diner	alien
ticker	knitter	concert	dingo	chorus
tiller	native	helmet	disco	kiwi
token	outage	palate	donor	ruby
turner	outing	pirate	duckling	album
turning	shutter	robot	dumper	hammer
taker	crater	bigot	glider	melon
taster	floater	carrot	header	barrow
teaser	flutter	circuit	meadow	cashew
temper	footer	comet	medic	warrior
tempo	freighter	culprit	rider	cargo
tenor	ghetto	goblet	soda	napkin
terrace	grater	hermit	spider	zombie
tester	gutter	hornet	spreader	miner
tiger	lighter	magnet	weirdo	brownie
tiler	lotus	misfit	widow	penguin
timer	motor	mullet	ballad	
tipster	nightie	peanut	blizzard	
tissue	photo	permit	comrade	
toaster	putter	pilot	custard	
toga	quitter	private	hazard	
topaz	rating	rabbit	liquid	
tummy	rotor	rocket	lizard	
turban	sweater	socket	nomad	
turkey	voter	wallet	rapid	
turnip	waiter	yoghurt	shepherd	

Table B.2 – List of words given to participants in group B. Words that participants were trained on are highlighted and were presented in all conditions. /t/ and /d/ tokens that are not highlighted were presented in pre-test and post-test conditions only. Fillers words were presented in the test condition only.

Initial /t/ tokens	Medial /t/ tokens	Final /t/ tokens	/d/ tokens	Fillers
tailor	coating	basket	daemon	rally
teller	cretin	biscuit	deacon	choir
telly	cutter	bonnet	dealer	snapper
tenner	cutting	bucket	demo	cleaner
tether	fritter	closet	diner	alien
ticker	knitter	concert	dingo	chorus
tiller	native	helmet	disco	kiwi
token	outage	palate	donor	ruby
turner	outing	pirate	duckling	album
turning	shutter	robot	dumper	hammer
taker	crater	bigot	glider	melon
taster	floater	carrot	header	barrow
teaser	flutter	circuit	meadow	cashew
temper	footer	comet	medic	warrior
tempo	freighter	culprit	rider	cargo
tenor	ghetto	goblet	soda	napkin
terrace	grater	hermit	spider	zombie
tester	gutter	hornet	spreader	miner
tiger	lighter	magnet	weirdo	brownie
tiler	lotus	misfit	widow	penguin
timer	motor	mullet	ballad	
tipster	nightie	peanut	blizzard	
tissue	photo	permit	comrade	
toaster	putter	pilot	custard	
toga	quitter	private	hazard	
topaz	rating	rabbit	liquid	
tummy	rotor	rocket	lizard	
turban	sweater	socket	nomad	
turkey	voter	wallet	rapid	
turnip	waiter	yoghurt	shepherd	

Table B.3 – List of words given to participants in group C. Words that participants were trained on are highlighted and were presented in all conditions. /t/ and /d/ tokens that are not highlighted were presented in pre-test and post-test conditions only. Fillers words were presented in the test condition only.

Appendix C

Wordlists used in chapter 6

Experiment 1a	Experiment 2a	Experiment 1b	Experiment 2b
tailleuse	chatouille	tailor	coating
tigrure	buteur	teller	cretin
taloche	critère	telly	cutter
tannée	crotale	tinker	cutting
texane	potiche	tether	fritter
tanneur	rature	ticker	knitter
tapeur	retouche	tiller	native
taquet	rotule	token	outage
tassement	satire	turner	outing
taulard	métisse	turning	shutter
topaze	mitaine	taker	crater
torpille	mouture	taster	floater
tourbière	fêtard	teaser	flutter
toxine	laitage	temper	footer
tulipe	lutin	tempo	freighter
tireuse	tireuse	tenor	tenor
tisseuse	tisseuse	terrace	terrace
tôlière	tôlière	tester	tester
tombeuse	tombeuse	tiger	tiger
tonsure	tonsure	tiler	tiler
tagueur	tagueur	timer	timer
tamoul	tamoul	tipster	tipster
terrien	terrien	tissue	tissue
tipi	tipi	toaster	toaster
tiret	tiret	toga	toga
tocard	tocard	topaz	topaz
tollé	tollé	tummy	tummy
toucan	toucan	turban	turban
touret	touret	turkey	turkey
tuba	tuba	turnip	turnip

Table C.1 – Wordlists used in the experiments. Words that are not highlighted were presented in pre-test and post-test conditions only. Words that are highlighted were presented in all conditions.

Bibliography

Abercrombie, D. (1967). Elements of general phonetics.

Abrego-Collier, C., Grove, J., Sonderegger, M., and Yu, A. C. (2011). Effects of Speaker Evaluation on Phonetic Convergence. In *Proceedings of the XVth ICPhS, Hong Kong*.

Adank, P. and Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging; Psychology and Aging*, 25(3):736.

Alcock, J. (1969). Observational learning in three species of birds. *Ibis*, 111(3):308–321.

Anisfeld, M. (1996). Only tongue protrusion modeling is matched by neonates. *Developmental Review*, 16(2):149–161.

Articulate Instruments Ltd (2010a). *Articulate Assistant User Guide: Version 1.18*. Articulate Instruments Ltd, Edinburgh, UK.

Articulate Instruments Ltd (2010b). *WinEPG Installation and Users Manual: Revision 1.18*. Articulate Instruments Ltd, Edinburgh, UK.

Aubanel, V. and Nguyen, N. (2010). Automatic recognition of regional phonological variation in conversational interaction. *Speech Communication*, 52(6):577–586.

Baayen, R., Piepenbrock, R., and van H, R. (1993). The CELEX lexical data base on CD-ROM.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(04):437–456.

Babel, M. (2011). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*.

Bandura, A. (1965). Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of personality and social psychology*, 1(6):589.

Bard, K. A. (2007). Neonatal imitation in chimpanzees (pan troglodytes) tested with two paradigms. *Animal Cognition*, 10(2):233–242.

Bard, K. A. and Russell, C. L. (1999). Evolutionary foundations of imitation: Social, cognitive, and developmental aspects of imitative processes in non-human primates. *Imitation in infancy: Progress and prospects of current research*, pages 89–123.

Bates, D. (2010). lme4: Mixed-effects modeling with R. *URL http://lme4. r-forge. r-project. org/book*.

Bauer, L. (1986). Notes on New Zealand English phonetics and phonology. *English World-Wide*, 7(2):225–258.

Bayard, D. (1990). Minder, Mork and Mindy? (-t) glottalisation in New Zealand English. In Bell, A. and Holmes, J., editors, *New Zealand Ways of Speaking English*, pages 149–164. Victoria University Press.

Bayard, D. (1999). Getting into a flap or turning off the tap in Dunedin?: stylistic variation in New Zealand English intervocalic (-t-). *English World-Wide*, 20(1):125–155.

Bell, A. (1977). *The language of radio news in Auckland: a sociolinguistic study of style, audience and subediting variation*. PhD thesis, University of Auckland.

Bell, A. (1984). Language style as audience design. *Language in society*, 13(2):145–204.

Bell, A. (1990). Audience and referee design in New Zealand media language. In Bell, A. and Holmes, J., editors, *New Zealand Ways of Speaking English*, pages 165–194. Victoria University Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience: Issues in cross-language research*.

Bjorklund, D. F. (1987). A note on neonatal imitation. *Developmental Review*, 7(1):86–92.

Blackmore, S. (1998). Imitation and the definition of a meme. *Journal of Memetics-Evolutionary Models of Information Transmission*, 2(11).

Browman, C. P. and Goldstein, L. (1986). Towards an articulatory phonology. *Phonology yearbook*, 3(21):9–252.

Browman, C. P. and Goldstein, L. (1995). Gestural syllable position effects in American English. *Producing speech: Contemporary issues*, pages 19–33.

Browman, C. P., Goldstein, L., et al. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.

Buizza, E. (2010). Frication and affrication of /t/ in RP English. *Proceedings of LangUE 2010*.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., and David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276(5312):593–596.

Chartrand, T. L. and Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893.

Chomsky, N. and Halle, M. (1968). The sound pattern of english.

Clark, L. (2013). *Sociolinguistics in Scotland*, chapter 11: Phonological repetition effects in natural conversation: evidence from TH-fronting in Fife. Palgrave.

Clarke-Davidson, C. M., Luce, P. A., and Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & psychophysics*, 70(4):604–618.

Coleman, J. (2002). Phonetic representations in the mental lexicon. *in Phonetics, phonology, and cognition*.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Cutting, J. and Rosner, B. (1974). Categories and boundaries in speech and music. *Attention, Perception, & Psychophysics*, 16(3):564–570.

Davenport, M. and Hannahs, S. J. (1998). *Introducing phonetics and phonology*. Arnold.

Davis, J. M. (1973). Imitation: A review and critique. *Perspectives in ethology*, 1:43–72.

de Courtenay, J. I. N. B. (1895). *Versuch einer theorie phonetischer alternationen: Ein capital aus der psychophonetik*. KJ Trübner.

Docherty, g., Hay, J., and Walker, A. (2006). Sociophonetic patterning of phrase-final /t/ in New Zealand English. In Warren, P. and Watson, C., editors, *Proceedings of the 11th Australian International Conference on Speech and Technology*, pages 378–383.

Eisner, F. and McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, 119:1950.

Epstein, R. (1984). Spontaneous and deferred imitation in the pigeon. *Behavioural Processes*, 9(4):347–354.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a tms study. *European Journal of Neuroscience*, 15(2):399–402.

Ferrari, P. F., Visalberghi, E., Paukner, A., Fogassi, L., Ruggiero, A., and Suomi, S. J. (2006). Neonatal imitation in rhesus macaques. *PLoS biology*, 4(9):e302.

Fisher, J. and Hinde, R. A. (1949). The opening of milk bottles by birds. *British Birds*, 42(347):57.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, pages 233–277.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, 84(1):115–123.

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., and 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.

Fowler, C. A. and Galantucci, B. (2005). The relation of speech perception and speech production. *The handbook of speech perception*, page 633.

Fromont, R. and Hay, J. (2008). ONZE Miner: The development of a browser-based research tool. *Corpora*, 3(2):173–193.

Fuchs, S. and Koenig, L. L. (2009). Simultaneous measures of electropalatography and intraoral pressure in selected voiceless lingual consonants and consonant sequences of german. *The Journal of the Acoustical Society of America*, 126:1988.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2):593–609.

Gates, G. and Mills, J. (2005). Presbycusis. *The Lancet*, 366(9491):1111–1120.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Lawrence Erlbaum.

Giles, H., Coupland, N., and Coupland, J. (1991). Contexts of accommodation: Developments in applied sociolinguistics. pages 1–68.

Giles, H. and Ogay, T. (2006). Communication accommodation theory. *Explaining communication: Contemporary theories and exemplars*, pages 293–310.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological review*, 105(2):251.

Goldinger, S. D. and Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic bulletin & review*, 11(4):716–722.

Goldstein, P., Ziegler, W., Vogel, M., and Hoole, P. (1994). Combined palatal-lift and epg-feedback therapy in dysarthria: A case study. *Clinical linguistics & phonetics*, 8(3):201–218.

Gordeeva, O. and Scobbie, J. (2007). Non-normative preaspirated voiceless fricatives in Scottish English: Phonetic and phonological characteristics. *QMU Speech Science Research Centre Working Papers*.

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., and Trudgill, P. (2004). *New Zealand English, its Origins and Evolution*. Cambridge University Press, Cambridge.

Gordon, E., Maclagan, M., and Hay, J. (2007). The ONZE corpus. In Beal, J. C., Corrigan, K. P., and Moisl, H. L., editors, *Creating and Digitizing Language Corpora: Diachronic Databases*, volume 2. Palgrave Macmillan.

Goto, H. (1971). Auditory perception by normal japanese adults of the sounds "l" and "r". *Neuropsychologia*, 9(3):317–323.

Gray, P. (2010). *Psychology*. Worth Publishers, sixth edition.

Grimm, W. (1966). Perception of segments of english-spoken consonant-vowel syllables. *The Journal of the Acoustical Society of America*, 40:1454.

Hall, N. (2010). Articulatory phonology. *Language and Linguistics Compass*, 4(9):818–830.

Hallé, P. A., Segui, J., Frauenfelder, U., and Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, 24(2):592–608.

Hardcastle, W. J., Gibbon, F., and Nicolaidis, K. (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics*, 19(3):251–266.

Harris, K. (1958). Cues for the discrimination of american english fricatives in spoken syllables. *Language and Speech*, 1(1):1.

Hay, J. and Drager, K. (2007). Sociophonetics. *Annu. Rev. Anthropol.*, 36:89–103.

Hay, J., Warren, P., and Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34(4):458–484.

Hedrick, M. and Ohde, R. (1993). Effect of relative amplitude of frication on perception of place of articulation. *Journal of the Acoustical Society of America*, 94:2005.

Heinz, J. and Stevens, K. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 33:589–596.

Herman, L. M. (2002). Exploring the cognitive world of the bottlenosed dolphin. *The cognitive animal, ed. M. Bekoff, C. Allen, & GM Burghardt. MIT Press.[BLS]*.

Hickey, R. (1984). Coronal Segments in Irish English. *Journal of Linguistics*, 20:233–51.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological review*, 93(4):411.

Holmes, J. (1994). New Zealand flappers: an analysis of "t" voicing in a sample of New Zealand english. *English World-Wide*, 15(2):195–224.

Holmes, J. (1995a). Glottal stops in New-Zealand English – an analysis of variants of word-final-t. *Linguistics*, 33(3):433–463.

Holmes, J. (1995b). Time for t: initial t in New Zealand English. *Australian Journal of Linguistics*, 15:127–156.

Holmes, J. (1995c). Two for t: flapping and glottal stops in New Zealand English. *Te Reo*, 38:53–72.

Holzman, P. S. and Rousey, C. (1966). The voice as a percept. *Journal of Personality and Social Psychology*, 4(1):79.

Holzman, P. S., Rousey, C., and Snyder, C. (1966). On listening to one's own voice: Effects on psychophysiological responses and free associations. *Journal of Personality and Social Psychology*, 4(4):432.

Honorof, D. N., Weihing, J., and Fowler, C. A. (2011). Articulatory events are imitated under rapid shadowing. *Journal of Phonetics*, 39(1):18–38.

Howell, P. and Rosen, S. (1983). Production and perception of rise time in the voiceless affricate/fricative distinction. *Journal of the Acoustical Society of America*, 73:976.

Huckvale, M. (2003). Prorec (version 1.0).

ISO 7029 (2000). Acoustics – Statistical distribution of hearing thresholds as a function of age.

Janse, E. (2009). Processing of fast speech by elderly listeners. *The Journal of the Acoustical Society of America*, 125:2361.

Jesse, A. and McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic bulletin & review*, 18(5):943–950.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, pages 145–165.

Johnston, L. (2002). Behavioral mimicry and stigmatization. *Social Cognition*, 20(1):18–35.

Jones, M. and Llamas, C. (2003). Fricated pre-aspirated /t/ in Middlesbrough English: An acoustic study. In *Proc. of the 15th International Congress of Phonetic Sciences, Barcelona*, pages 3–9.

Jones, M. and Llamas, C. (2008). Fricated realisations of /t/ in Dublin and Middlesbrough English: an acoustic analysis of plosive frication and surface fricative contrasts. *English Language and Linguistics*, 12(03):419–443.

Jones, M. J. and McDougall, K. (2009). The acoustic character of fricated/t/in Australian English: A comparison with /s/ and /ʃ/. *Journal of the International Phonetic Association*, 39(03):265–289.

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America*, 85:1718.

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108:1252–1263.

Junqua, J.-C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America*, 93:510.

Kawai, M. (1965). Newly-acquired pre-cultural behavior of the natural troop of Japanese monkeys on Koshima Islet. *Primates*, 6(1):1–30.

Kim, M. (2012). *Phonetic Accommodation after Auditory Exposure to Native and Nonnative Speech*. PhD thesis, Northwestern University.

Kluender, K. and Walsh, M. (1992). Amplitude rise time and the perception of the voiceless affricate/fricative distinction. *Attention, Perception, & Psychophysics*, 51(4):328–333.

Kraljic, T. and Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive psychology*, 51(2):141–178.

Kraljic, T. and Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic bulletin & review*, 13(2):262–268.

Kraljic, T. and Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1):1–15.

Kuhl, P. K., Meltzoff, A. N., et al. (1982). The bimodal perception of speech in infancy. *Science*, 218:1138–1144.

Lane, H. and Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech, Language and Hearing Research*, 14(4):677.

Langstrof, C. (2006). Acoustic evidence for a push-chain shift in the intermediate period of New Zealand English. *Language variation and change*, 18(2):141.

LaRiviere, C., Winitz, H., and Herriman, E. (1975). The distribution of perceptual cues in english prevocalic fricatives. *Journal of Speech and Hearing Research*, 18(4):613.

Lavoie, L. M. (2002). Subphonemic and suballophonic consonant variation: The role of the phoneme inventory. *ZAS papers in Linguistics*, 28:39–54.

Legou, T., Marchal, A., Meynadier, Y., and André, C. (2008). 3D Palatography. In *Proc. of the 8th International Seminar on Speech Production, Strasbourg*, pages 369–372.

Lewandowski, N. (2012). *Talent in nonnative phonetic convergence*. PhD thesis, Universität Stuttgart.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36.

Lisker, L. and Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the sixth international congress of phonetic sciences*, volume 563, page 567. Academia Prague.

Luce, P. A. and Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition*, 26(4):708–715.

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., Williams, S. C., Woll, B., and Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: an event-related fmri study. *Neuroreport*, 11(8):1729–1733.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28(5):407–412.

Mann, V. A. and Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]–[[s] distinction. *Perception & Psychophysics*, 28(3):213–228.

Mann, V. A. and Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America*, 69:548.

Massaro, D. W. and Cohen, M. M. (1983). Phonological context in speech perception. *Perception & psychophysics*, 34(4):338–348.

Maurer, D. and Landis, T. (1990). Role of bone conduction in the self-perception of speech. *Folia phoniatrica*.

Maye, J., Aslin, R. N., and Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3):543–562.

McClelland, J. L. and Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, 18(1):1–86.

McLennan, C. T., Luce, P. A., et al. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology-Learning Memory and Cognition*, 31(2):306–321.

McQueen, J. M., Cutler, A., and Norris, D. (2006a). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6):1113–1126.

McQueen, J. M., Norris, D., and Cutler, A. (2006b). The dynamic nature of speech perception. *Language and Speech*, 49(1):101–112.

McQueen, J. M., Tyler, M. D., and Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development*, 8(4):317–339.

Meltzoff, A. N. and Moore, M. K. (1999). Resolving the debate about early imitation.

Meltzoff, A. N., Moore, M. K., et al. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198(4312):75–78.

Miller, J. L. and Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 25(6):457–465.

Mitani, S., Kitama, T., and Sato, Y. (2006). Voiceless affricate/fricative distinction by frication duration and amplitude rise slope. *Journal of the Acoustical Society of America*, 120:1600.

Mitterer, H., Chen, Y., and Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science*, 35(1):184–197.

Mitterer, H. and 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. and Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception–action coupling in speech. *Attention, Perception, & Psychophysics*, pages 1–19.

Morris, J. (2010). Phonetic variation in Northern Wales: preaspiration. In *Proceedings of the Second Summer School of Sociolinguistics*.

Mullenix, J. W., Pisoni, D. B., and Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85(1):365–378.

New, B., Pallier, C., Ferrand, L., and Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année psychologique*, 101(3):447–462.

Nguyen, N. (2000). A matlab toolbox for the analysis of articulatory data in the production of speech. *Behavior Research Methods*, 32(3):464–467.

Nguyen, N., Wauquier, S., and Tuller, B. (2009). The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories. *Approaches to phonological complexity*, pages 5–31.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2):132–142.

Nielsen, K. Y. (2008). *Word-level and feature-level effects in phonetic imitation*. PhD thesis, University of California Los Angeles.

Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, 47(2):204–238.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115(1):39.

Nygaard, L. C. and Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3):355–376.

Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1):42–46.

Oudeyer, P.-Y. (2006). *Self-Organization in the Evolution of Speech*. Oxford University Press.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of experimental psychology. Learning, memory, and cognition*, 19(2):309.

Pandeli, H., Eska, J., Ball, M., and Rahilly, J. (1997). Problems of phonetic transcription: The case of the Hiberno-English slit-t. *Journal of the International Phonetic Association*, 27:65–75.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119:2382.

Pardo, J. S., Gibbons, R., Suppes, A., and Krauss, R. M. (2011). Phonetic convergence in college roommates. *Journal of Phonetics*.

Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., and Evans, A. C. (2006). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience*, 8(11):2236–2246.

Pichora-Fuller, M. K. and Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology*, 42(S2):11–16.

Pickering, M. J. and Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language & Computation*, 4(2):203–228.

Pickering, M. J., Garrod, S., et al. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–189.

Pierce, W. D. and Cheney, C. D. (2008). *Behavior analysis and learning*. Psychology Press, fourth edition.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. *Typological studies in language*, 45:137–158.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, 46(2-3):115–154.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, 34(4):516–530.

Porter, R. J. and Castellanos, F. X. (1980). Speech-production measures of speech perception: Rapid shadowing of VCV syllables. *The Journal of the Acoustical Society of America*, 67:1349.

Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192.

Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V. (2002). From mirror neurons to imitation: Facts and speculations. *The imitative mind: Development, evolution, and brain bases*, 6:247.

Schwartz, J.-L., Berthommier, F., and Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, 93(2):B69–B78.

Scobbie, J. M. and Pouplier, M. (2010). The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics*, 38(2):240–259.

Serkhane, J. E. (2005). *Un bébé androïde vocalisant: Étude et modélisation des mécanismes d'exploration vocale et d'imitation orofaciale dans le développement de la parole*. PhD thesis, Institut National Polytechnique de Grenoble.

Silverman, D. (2003). On the rarity of pre-aspirated stops. *Journal of Linguistics*, 39:575–598.

Sjerps, M. J., McQueen, J. M., et al. (2010). The bounds on flexibility in speech perception. *Journal of experimental psychology. Human perception and performance*, 36(1):195.

Smits, R. (2000). Temporal distribution of information for human consonant recognition in vcv utterances. *Journal of Phonetics*, 28(2):111–135.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, 111:1872.

Sussman, H. and Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Attention, Perception, & Psychophysics*, 58(6):936–946.

Taylor, B. (1996). Gay men, femininity and /t/ in New Zealand English. *Wellington Working Papers in Linguistics*, 8.

Thomas, B. and Hay, J. (2005). A pleasant malady: The Ellen/Allan merger in New Zealand English. *Te Reo*, 48:69.

Thorpe, W. H. (1956). Learning and instinct in animals.

Tonndorf, J. (1972). Foundations of Modern Auditory Theory, volume II, chapter Bone Conduction.

Tranel, B. (1987). *The sounds of French: An introduction*. Cambridge University Press.

Trask, R. (1996). *A Dictionary of Phonetics and Phonology*. Routledge.

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, 84:917.

Wagner, A., Ernestus, M., and Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *The Journal of the Acoustical Society of America*, 120:2267.

Watson, K. (2007). Liverpool English. *Journal of the International Phonetic Association*, 37(3):351–360.

Wedel, A. and Van Volkinburg, H. (2009). Modeling simultaneous convergence and divergence of linguistic features between differently-identifying groups in contact. *Manuscript in preparation*.

Whalen, D. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Attention, Perception, & Psychophysics*, 35(1):49–64.

Wilcox, S. and Katz, S. (1981). A direct realistic alternative to the traditional conception of memory. *Behaviorism*, pages 227–239.