Articulatory Evidence for Interactivity in Speech Production

Corey T. McMillan, M.Sc.

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

to

School of Philosophy, Psychology and Language Sciences
University of Edinburgh

2008

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Corey T. McMillan

Abstract

Traditionally, psychologists and linguists have assumed that phonological speech errors result from the substitution of well-formed segments. However, there is growing evidence from acoustic and articulatory analyses of these errors which suggests activation from competing phonological representations can cascade to articulation. This thesis assumes a cascading model, and investigates further constraints for psycholinguistic models of speech production. Two major questions are addressed: whether such a cascading model should include feedback; and whether phonological representations are still required if articulation is not well-formed. In order to investigate these questions a new method is introduced for the analysis of articulatory data, and its application for analysing EPG and ultrasound recordings is demonstrated.

A speech error elicitation experiment is presented in which acoustic and electropalatography (EPG) signals were recorded. A transcription analysis of both data sets tentatively supports a feedback account for the lexical bias effect. Crucially, however, the EPG data in conjunction with a perceptual experiment highlight that categorising speech errors is problematic for a cascaded view of production. Therefore, the new analysis technique is used for a reanalysis of the EPG data. This allows us to abandon a view in which each utterance is an error or not. We demonstrate that articulation is more similar to a competing phonological representation when the competitor yields a real word. This pattern firmly establishes evidence for feedback in speech production.

Two additional experiments investigate whether phonological representations, in addition to lower-level representations (e.g., features), are required to account for ill-formed speech. In two tongue-twister experiments we demonstrate with both EPG and ultrasound, that articulation is most variable when there is one competing feature, but not when there are two competing features. This pattern is best accounted for in a feedback framework in which feature representations feedback to reinforce phonological representations.

Analysing articulation using a technique which does not require the categorisation of responses allows us to investigate the consequences of cascading. It demonstrates that a cascading model of speech production requires feedback between levels of representation and that phonemes should still be represented even if articulation is malformed.

Acknowledgements

Despite spending the past few years studying how speakers transform ideas into articulation, I still cannot articulate how grateful I am for the love, humour and enormous support that Sara Park McMillan has provided throughout my studies.

Special thanks are extended especially to the McMillan and Park families for their generosity, love and support. Also, I would like to give special thanks to the Beavercreek family for getting me out of bed each day, whether I wanted to or not.

Sincere thanks are extended to Martin Corley; his enthusiasm, skillful programming, knowledge and support have been undeniably instrumental in my academic growth.

Additionally, I would like to thank Robin Lickley for always generously sharing his phonetic knowledge and giving valuable guidance throughout.

The University of Edinburgh and Queen Margaret University have provided a rich environment to develop the ideas presented in this thesis. While there are too many people to individually thank, a few deserve formal recognition: The members of the Edinburgh Disfluency Group, especially Phil Collard, Lucy MacGregor, Suzy Moat and Michael Schnadt who have provided substantial support from discussing ideas to making cups of tea. Steve Cowen, Jim Scobbie and Alan Wrench who made the articulatory recordings possible. Robin Hellier for volunteering countless hours of time to segment and transcribe audio files. Jim Hurford for encouraging me to study at Edinburgh University. George Montgomery for his thoughtful conversations and for taking me on as a research assistant. Scott McDonald for his conversations about statistics. Heartfelt thanks also go to Jimmy Duncan, Katie Keltie, Steven McGauley, Debbie Moodie and the late George Tait. Also, grateful thanks are extended to several anonymous volunteers who dedicated their time to participate in experiments.

Thanks also to participants and speakers at AMLaP and CUNY conferences for their valuable feedback and discussion of the ideas presented in this thesis.

I would like to also thank Jack Brougham, Sandy Burns, the DeBusks, the Dickins Family, James Dodson, Kes Evoy, Jasmine Falconer, the Hams, Brody Hunt, Martin Jones, Stevie Kearney, Nigel Kennedy, Jane Manson, Graham Rushworth, Katherine Shearer, the Sykewoods, Jon Todman, and Elly Veitch whose friendship, banter, curiosity, and good taste have always kept me in good spirits over the course of my studies.

Lastly, I would like to thank Mark Wheeler, Murray Grossman and Robin Clark who graciously took me into their laboratories, inspired me to pursue cognitive psychological studies and have provided me with invaluable support, advice and friendship throughout my subsequent studies.

Funding for this work was provided by a NRSA Predoctoral Fellowship from the National Institute of Health (DC07282) and a University of Edinburgh College of Humanities and Social Sciences Scholarship. Travel funding to attend conferences was provided by an Experimental Psychology Society Grindley Grant and the School of Philosophy, Psychology, and Language Sciences.

Additional thanks go to the Scottish weather for ensuring that outdoor pursuits and the threat of sunburn never distracted me from my studies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Speaking is a complex process involving several levels of planning. A speaker must form a concept of an intended message, retrieve the necessary representations to generate the message, and then articulate the intended message. There is widespread agreement about these levels of planning, however, researchers disagree about how levels interact (Dell, 1986; Levelt, Roelofs, & Meyer, 1999). A primary source of evidence used to discriminate between models of production are speech errors. A long tradition of research has focused on categorising speech errors into different types such as: lexical exchanges (*Bonnie Prince Billy* → "Prince Bonnie Billy"), morphological exchanges (*Beastie Boys* → "Beasties Boy"), and phonological exchanges (*Johnny Cash* →"Connie Jash"). In this thesis we focus on phonological speech errors and how they can inform us about the interaction between levels of planning.

Phonological speech errors have long been considered well-formed substitutions of one phonological representation by another phonological representation (Boomer & Laver, 1968; Fromkin, 1971; Nooteboom, 1969; Wells, 1951). This has been motivated by the observation that errors can be best described as phonological rather than featural (e.g., Fromkin, 1971), they obey positional constraints (e.g., Nooteboom, 1969), and adhere to phonotactic constraints of the language (e.g., Boomer & Laver, 1968). However, this evidence has not been without exceptions. More recently, an emerging discipline that focuses on the articulatory and acoustic characteristics of speech errors has challenged the traditional categories of speech errors (Pouplier & Hardcastle, 2005). Detailed investigations have demonstrated that speech errors can be *non-canonical*. The articulation of a /k/, which would normally only include tongue-dorsum raising, may include both tongue tip and tongue-dorsum raising (Frisch, 2007; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; Pouplier, 2003, 2007; Stearns, 2006). Similarly, the voice onset time (VOT)

of an unintended /k/ may be shorter in duration than a canonical /k/, but longer in duration that a canonical /g/ (Goldrick & Blumstein, 2006, see also Frisch & Wright, 2002). As a result, articulatory and acoustic evidence has been used to challenge the traditional view of well-formed errors.

Two different classes of psycholinguistic models can account for non-canonical errors. On the one hand, there are staged models which posit that the planning of an utterance must be completed prior to articulation (e.g., Dell, 1986; Levelt et al., 1999). According to staged models, non-canonical errors result from an error in articulatory implementation. On the other hand, cascading models propose that the activation of partially activated representations can flow to articulation (e.g., Goldrick & Blumstein, 2006). A cascading account attributes non-canonical speech errors to the simultaneous articulation of partially activated phonological representations.

In this thesis we argue that a cascading model of production provides a parsimonious account for speech errors. Instead of assigning malformed responses to categories of 'errors', we investigate the extent to which speech production may reflect the continuous activation of phonological representations. To achieve this novel approach we present an analysis method, the *Delta method*, which quantifies articulatory variability from electropalatography (EPG) and midsagittal tongue ultrasound recordings. Specifically, we investigate how articulation is influenced by competition during planning by comparing articulation recorded during error-elicitation tasks relative to articulation recorded during conditions which should not promote 'errorful' speech.

Using the Delta method, we investigate the consequences of assuming a cascading model of production. We first investigate whether a feedback flow of information is required to account for the lexical bias effect: the finding that speech errors are more likely to result in real words than predicted by chance (e.g. Dell & Reich, 1981). Researchers have attributed the lexical bias effect to feedback (Dell, 1986; Humphreys, 2002), self-monitoring (Baars, Motley, & MacKay, 1975), or both (Hartsuiker, Corley, & Martensen, 2005; Nooteboom & Quené, in press). However, these accounts have been based on speech error investigations that required responses to be categorised as "errorful" or "correct" and, therefore, assumed a staged model of production. Since the aim of this thesis is to investigate speech production in a cascading model, which does not discriminate between categories of responses, we reevaluate the evidence for the lexical bias effect.

We also investigate whether lower-level representations (e.g., features) are required in a cascading model of production. Staged accounts of non-canonical speech errors have attributed errors to the misselection of a lower-level representation (e.g., articulatory gesture; Goldstein et al., 2007; Mowrey & MacKay, 1990). However, if partially activated phonological representations can cascade to articulation, lower-level representations may not be required in models of production. To investigate the role of lower-level representations we reevaluate the evidence for the phonological similarity effect: the finding that speech errors are more likely to occur if phonological representations are similar as opposed to dissimilar (e.g., Shattuck-Hufnagel & Klatt, 1979). We also investigate whether feedback is also required if lower-level representations are incorporated into a cascading model.

We conclude that measuring articulation, without using categorisation, allows us to investigate of the consequences of assuming a cascading model of speech production. Our experimental findings support a cascading model of speech production which includes feedback between phonological to lexical representations and feedback between feature and phonological representations.

Chapter 2

A Cascading Approach to Speech Error Investigations

2.1 Chapter Overview

The focus of this chapter is on how speech errors can be used to investigate the extent to which different levels of processing in speech production influence one another. In Section 2.3 of this chapter we present a review of speech error research and focus on phonological speech errors. Traditionally phonological speech errors have been viewed as well-formed segment exchanges and this view has motivated much of current psycholinguistic theory on speech production (e.g., Fromkin, 1971; Garrett, 1975). However, an increasing amount of evidence has challenged this traditional view (e.g., Pouplier & Hardcastle, 2005). We discriminate between two classes of models that can account for the challenging evidence. Specifically, in Section 2.4 a distinction is drawn between staged models of production, which require the selection of a phonological representation prior to articulation, and cascading models which posit that articulation can include partial activation of phonological representations. Ultimately, a functional model of production is proposed that allows phonological representations to cascade to articulation. Lastly, in Section 2.5 we present a discussion on how different levels of speech production interact with one another with a focus on feedback and feedforward interactivity. In the final section we outline the structure of this thesis which will investigate the consequences of cascading activation on speech production.

2.2 Introduction

Turning an intended message into articulated speech is a complex process. Anyone who has tried to utter a phrase like *fresh fried flesh of fowl* will have little

doubt that the resulting articulation often differs substantially from the intended phrase. Several factors of speech planning may contribute to this difficulty. There is widespread agreement that speech planning involves several subprocesses including: the conceptualisation of an intended message, the selection of lexical, syntactic, morphological and phonological representations, and the movement of the articulators in the best way to communicate the intended message (e.g., Levelt, 1989). However, researchers disagree about how information flows from one level to another (e.g., Dell, 1986; Rapp & Goldrick, 2000; Levelt, 1989; Levelt et al., 1999).

This dissertation examines the extent to which levels of planning an utterance influence one another before and during articulation. A primary source of evidence used to constrain models of production is the occurrence of speech errors. Investigations on the phonological aspects of speech errors have demonstrated that errors such as *cream of chicken soup* → "cheam of cricken soup" are not random noise and therefore can inform researchers about the nature of speech planning.

2.3 Speech Errors

Speech errors occur regularly in spontaneous speech. Estimates of the frequency of slips of the tongue involving a sound error range from 31 (Garnham, Shillcock, Brown, Mill, & Cutler, 1981) to 160 (Shallice & Butterworth, 1977) occurrences out of every 100,000 words spoken. While these rates may appear low at first glance, a comparison of the lowest estimate with spoken word frequencies suggests speakers are likely to produce an error as often as they say "war" or "London". Therefore, the regular occurrence of errors has prompted linguists (e.g., Fromkin, 1971) and psychologists (e.g., Garrett, 1975) to investigate them in more detail. Perhaps the most established finding is that speech errors adhere to a strict pattern: planning units are typically only substituted by other planning units of the same type. Planning units can include words (e.g., *Bonnie Prince Billy* → "Prince Bonnie Billy"[1]), syllables (e.g., *Captain Beefheart* → "Beeftain Capheart"), or morphemes (e.g., *Beastie Boys* →"Beasties Boy"), but the focus of this dissertation will be on phonological errors (e.g., *Johnny Cash* →"Connie Jash").

[1]We use *italic* font to represent an intended utterance, a rightward arrow (→) to represent how the utterance was realised, and quotation marks to represent the spoken utterance.

2.3.1 Traditional View

A long tradition of psycholinguistic research has viewed phonological speech errors as the product of well-formed substitutions of one phoneme by another phoneme (Fromkin, 1971; Garrett, 1979; Meringer & Mayer, 1895; Shattuck-Hufnagel & Klatt, 1979). These errors can include full exchanges (e.g., *Johnny Cash* →"Connie Jash"), perseverations (e.g., *Johnny Cash* →"Jonnie Jash"), and anticipations (e.g., *Johnny Cash* →"Connie Cash"). For the purposes of this thesis, phonological errors will be referred to as *substitutions* independent of whether the error was a full exchange or partial exchange unless otherwise specified.

There have been several motivations for the traditional assumption of well-formed phoneme substitutions. Several researchers have argued that errors are best described as segmental phoneme substitutions as opposed to feature substitutions. Fromkin (1971) reported a variety of errors from her self-recorded corpus of speech errors, including:

1. (a) *spell* → "smell" (p. 35)
   (b) *reveal* → "refeal" (p. 36)
   (c) *pussy cat* → "cussy pat" (p. 33)
   (d) *play the victor* → "flay the pictor" (p. 41)

Fromkin argued that regardless of the possibility that some errors may involve a distinctive feature substitution, the phoneme is the primary segment substituted in such errors. For example, in (1a) the intended production of /p/ becomes nasal to yield an /m/. In (1b) the intended /v/ is devoiced to result in /f/. While these errors could be specified phonetically, for example */p/* → "/k/" in 1c could be described as *<anterior–,high–,back–>* → "<anterior–,high+,back+>", the errors can be more simply described phonologically as *bilabial consonant* → "velar consonant". In other words, it is impossible to discriminate between whether a single feature or whole phoneme was substituted. According to this view, adopting a phonemic substitution standpoint sufficiently captures the difference between the intended and uttered speech signal without over-specification of distinctive features. In later research Fromkin (1973) argues that distinctive features play a role in substitution errors but that their interaction does not exclude segmental substitution.

Another motivation for a segmental view of speech errors is that segment substitutions are far more frequent than feature-level errors (Shattuck-Hufnagel, 1982; Shattuck-Hufnagel & Klatt, 1979). For example, Shattuck-Hufnagel and Klatt

(1979) identified 70 speech errors in the MIT corpus for which a single feature substitution was possible (defined as phonemes that would yield a legal outcome in English, unlike voiceless /l/) and observed that only 3 of the errors involved a single feature substitution. They used this evidence to argue that "features are not independent movable entities at the level where most substitution and exchange errors are made" (Shattuck-Hufnagel & Klatt, 1979, p. 50). Similarly, Meyer (1992) reported that less than 5% of errors could be accounted for by a single feature substitution.

In addition to the evidence for well-formed phoneme substitutions in speech errors, it has been demonstrated that speech errors also obey positional constraints. For example, onsets are substituted by other onsets, vowels by other vowels, and final consonants by other final consonants (Boomer & Laver, 1968; Fromkin, 1968, 1971; MacKay, 1970; Nooteboom, 1969). This evidence has generally been used to motivate frame based accounts of speech production in which each phoneme of a phonological word is filled with an appropriate phoneme for each position (Dell, 1986, 1988; Hartley & Houghton, 1996; MacKay, 1987; Roelofs, 1996; Shattuck-Hufnagel, 1979), though frame constraints may not be required (Dell, Juliano, & Govindjee, 1993; Vousden, Brown, & Harley, 2000). Although frame-based accounts suggest that substitutions occur at all positions within a word, in practice most corpus and experimental research on speech errors has focused on word onsets. In this thesis we follow this practice and do not address the issue of errors that occur beyond the initial onset position.

Another source of evidence used to support segmental errors is that speech errors typically obey phonotactic constraints (Boomer & Laver, 1968; Fromkin, 1971; Wells, 1951). Wells' (1951) First Law of Speech Errors stated that errors are nearly always phonotactically legal. In a qualitative assessment of a corpus of speech errors, Fromkin (1971) demonstrated that errors do not result in phonemes which are disallowed in a given language. One convincing example she gave was in (1d) above in which /v/ was substituted by /p/. Fromkin explained that one possible interpretation of this error is that a full exchange, which would have resulted in "vlay the pictor", would have been phonotactically illegal, so the speaker therefore devoiced the /vl/ to produce /fl/ which is a legal consonant cluster. More recently, it has been demonstrated that speakers' errors will obey phonotactic constraints after only two days of implicit training on new phonotactic rules (Warker & Dell, 2006).

Taken together, the fact that errors can be more simply specified as phonological (as opposed to featural) errors, the very low rates of feature-level errors and the evidence that errors obey phonotactic constraints suggests that speech errors result from well-formed phonological segment substitutions. However, researchers have typically reported exceptions: a phonetic description of errors may be too specific but can capture the units that have been substituted (Fromkin, 1971), feature errors, despite a very low occurrence, do exist (Shattuck-Hufnagel & Klatt, 1979), and Wells' (1951, p. 86) First Law states that errors "practically always" adhere to phonotactic constraints. In the following section the errors that are exceptions to the traditional view will be discussed in more detail.

2.3.2 Challenges to the Traditional View of Speech Errors

In 1980, Laver investigated vowel quality during the repetition of CVC pairs such as *peep–pip*, *pup–poop*, *parp–peep*. He observed that for each speaker 1–4% of responses contained an error in vowel production. For example, during repetitions of *parp–peep* speakers occasionally responded "pipe". These findings suggest that speakers may have produced a blend of /ɑ/ and /i/. However, Laver (1980, p. 26) reasoned "it is improbable to think of the brain as sending out simultaneous but contradictory neuromuscular commands to the same muscles".

Since the work of Laver there has been a history of reports of anomalous speech errors (Laver, 1980; Butterworth & Whittaker, 1980) and, more recently, an emerging field of articulatory and acoustic investigations of speech errors (Pouplier & Hardcastle, 2005) challenging the traditional view of speech errors. Together these investigations have provided evidence that errors may not be best described as substitutions of one canonical phoneme by another phoneme (Frisch, 2007; Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein et al., 2007; Laver, 1980; Mowrey & MacKay, 1990; Pouplier, 2003, 2007; Stearns, 2006). In this section a review of the literature in challenge of the traditional view will be presented. Henceforth, we refer to errors outside the traditional view of speech errors as *non-canonical*, and those within the traditional view will be referred to as *canonical*.

In the first systematic investigations of errors produced during tongue-twisters, Butterworth and Whittaker (1980) observed that approximately 27% of errors produced reflected cluster errors, such as *bat gat* → "gbat gat". Moreover, they specified the errors were genuine clusters because they did not contain a vowel between /g/ and b such as /gəbæt/. They reasoned that traditional production models could account

for these errors, but due to the nature of the tongue-twister task, only if they were interpreted as errors of articulatory implementation. Together, the reports of Laver (1980) and Butterworth and Whittaker (1980) suggest errors may not necessarily result from whole phoneme substitutions. Rather errors may include components of both the target phoneme (e.g., /b/ in *bat gat*) and an unintended phoneme (e.g., /g/) in the recorded utterances. However, these reports were only based on perceptual transcriptions of errors and therefore do not provide direct articulatory or acoustic evidence for such non-canonical errors.

More recently, systematic investigations of articulatory properties of errors have demonstrated evidence similar to Laver's (1980) and Butterworth and Whittaker's (1980) reports. In an electromyographic (EMG) investigation, Mowrey and MacKay (1990) observed transversus/verticalis muscle activity normally associated with /l/ production, during the production of "bay" in repetitions of *Bob flew by Bligh Bay*. Similarly, they also observed anticipatory transversus/verticalis muscle activity during the production of "fried" in the tongue-twister *fresh fried flesh of fowl*. They argued that these segmental "blends" were the most common error type observed in their investigation. However, their analysis was restricted to a qualitative visual inspection of EMG amplitude rather than a quantitative analysis (Pouplier & Hardcastle, 2005).

Quantitative investigations of the articulation of speech errors have also provided evidence for non-canonical phoneme substitutions (Frisch, 2007; Goldstein et al., 2007; Pouplier, 2003, 2007; Stearns, 2006). In an investigation using electromagnetic midsagittal articulometry (EMMA; sometimes referred to as electromagnetic articulometry: EMA) articulation was measured while speakers uttered monosyllabic words with alternating consonants such as *cop top*. Goldstein et al. (2007, see also Pouplier, 2003) observed errorful tongue tip raising in addition to the normal tongue-dorsum raising during /k/ production. Similarly, both errorful tongue-dorsum and non-errorful tongue-tip raising was observed for /t/ production. They interpreted these results as "gestural intrusions" in which two competing articulatory gestures are simultaneously articulated. In another EMMA investigation gestural intrusions were also observed in a non-repetition-based speech error elicitation task – the SLIP task (Pouplier, 2007). An ultrasound investigation of the mid-sagittal contour of the tongue also revealed similar results: tongue-dorsum raising during the articulation of intended alveolar consonants in a repetition task (Stearns, 2006, see also Frisch, 2007).

Non-canonical phoneme substitutions have also been observed in acoustic analyses of speech errors (Frisch & Wright, 2002; Goldrick & Blumstein, 2006). In an analysis of percent voicing in /s/–/z/ errors, Frisch and Wright (2002) observed abnormal proportions of periodicity in the duration of onset consonants during a tongue-twister task. This analysis further revealed a continuum of percent voicing for both /s/ and /z/ productions ranging from 0-100% voicing. In another acoustic tongue-twister investigation, Goldrick and Blumstein (2006) demonstrated that the VOT of an errorfully produced stop consonant (e.g., /g/ in *keff* → "geff") differs from both a canonical /g/ and canonical /k/.

Taken together, detailed articulatory and acoustic investigations suggest the traditional view of well-formed substitutions in speech errors should be questioned. A primary motivation for the traditional view was that, if non-canonical errors exist, they are too rare to be considered. However, instrumental investigations have provided evidence against this assumption. In fact, while the typical speech error rate across speech error elicitation experiments is 1–8% (Nooteboom & Quené, 2007, based on a review of SLIP experiments), the observed rate of non-canonical errors in instrumental experiments is much higher: 28% (Pouplier, 2007), 32% (Mowrey & MacKay, 1990), and 36% (Goldstein et al., 2007). Crucially, if the traditional view of errors is rejected because of recent instrumental observations, models of speech production must be constrained to account for non-canonical errors.

2.4 Model Constraints for Non-Canonical Speech Errors

Most psycholinguistic models of speech production posit that planning occurs over three stages (Dell, 1986; Levelt, 1989; Levelt et al., 1999; Meyer, 1990). First, conceptualisation must take place. This is the stage of retrieving the words representing the concept one wants to convey. Second, after the words are retrieved, the phonological representation required to make the sounds of those words must be retrieved. Finally, those sounds must be articulated by moving the articulators in the appropriate way to communicate the intended sounds. Investigations of phonological speech errors are primarily concerned with the latter two stages: phonological encoding and articulation. In this section a distinction will be made between staged models, which propose that a representation must be selected prior to articulation, and cascading models, which allow the activation of partially activated representations to flow to articulation.

All stage-based models of speech production posit that the articulation stage does not begin until the phonological encoding stage is complete. Phonological encoding is considered complete when at least one entire phonological word has been planned (Dell, 1986; Levelt, 1989; Meyer, 1990). Selection in staged models is "winner takes all", whereby at a given deadline the most highly activated phonological representation is selected for each position (e.g., onset, nucleus, coda). Once these representations are selected they are then passed on to a later stage to be articulated. Stage-based production models are therefore based on the presumption that articulatory programs are retrieved following phonological encoding. While there has been a debate over whether these programs constitute stored syllables (Crompton, 1982; Levelt, 1989; Levelt & Wheeldon, 1994; Meyer, Roelofs, & Levelt, 2003; Roelofs, 1997a) or alternative units (Dell, 1986; Dell, Burger, & Svec, 1997; MacKay, 1987; Stemberger, 1985a; Vousden et al., 2000), there is still agreement that some form of retrieval must take place.

In contrast to stage-based models, cascading models generally propose that activation can flow across levels of processing (McClelland, 1979). Cascading models of production propose that, instead of "winner takes all" selection, responses can reflect partial activation of competing representations (Rapp & Goldrick, 2000; Goldrick & Blumstein, 2006). Several investigations on cascading activation have focused on a slightly different question concerned with incremental processing: whether articulation can be initiated prior to the completion of planning (Damian, 2003; Damian & Dumay, 2007; Kawamoto, Kello, Jones, & Bame, 1998; Kawamoto, Kello, Higareda, & Vu, 1999; Kello, Plaut, & MacWhinney, 2000; Kello & Plaut, 2000). However, the relevant focus for accounting for instrumental speech error data is the extent to which partial activation of phonological representations can cascade to articulation.

In Section 2.3.2 we presented several sources of evidence for non-canonical speech errors. For example, articulatory investigations of speech errors have demonstrated evidence for tongue tip and tongue-dorsum raising during the production of an intended alveolar during the production of phrases such as *top cop* (Goldstein et al., 2007; Pouplier, 2003, 2007). Acoustic investigations have revealed that the VOT for errorful productions differs from the VOT of the canonical intended onset and the canonical unintended onset (Goldrick & Blumstein, 2006, see also Frisch & Wright, 2002 for acoustic evidence). The occurrences of non-canonical errors can be accounted for in a staged and cascading framework. In the following section we present the different accounts.

2.4.1 Staged and Cascading Accounts for Speech Errors

Staged and cascading models of speech production account for non-canonical speech errors in different ways. According to staged models of production canonical speech errors are the result of a misselection of a phonological representation (e.g., Dell, 1986; Levelt, 1989; Levelt et al., 1999). More specifically, only one phonological representation can be (mis)selected: either the intended phonological representation or some other unintended phonological representation. This selection process occurs only during phonological encoding and the process is complete before the initiation of articulation. In a later section (Section 2.5) we discuss the details of how phonological representations become activated.

However, staged models can not account for non-canonical errors in the same way as canonical errors. Since only one phonological representation can be selected, it is not possible for a non-canonical error, which contains properties of two different phonemes, to be attributed to phonological encoding. Therefore, the only way this set of models can account for non-canonical errors is if these errors are attributed to articulatory implementation. In fact, Levelt et al. (1999) are proponents of this argument and explicitly state that Mowrey and MacKay's (1990) observation of inappropriate articulatory muscle movements can not rule out that non-canonical errors result from a late motor execution stage.

Attributing non-canonical errors to articulatory implementation presents a paradox for speech error research. Traditionally, (canonical) errors have been attributed to planning at the level of phonological encoding. In fact, experimental speech error investigations on tongue-twisters have provided evidence that errors are the result of planning and can not be attributed to articulatory implementation (Dell & Repka, 1992; Wilshire, 1999). Similarly, researchers have demonstrated that the SLIP task (described in Section 2.5.1), a laboratory speech error elicitation paradigm, successfully elicits planning errors rather than errors from another source (Dell, 1986; Hartsuiker et al., 2005, Experiment 1b). Therefore, it is not clear why some errors should be attributed to execution, while others attributed to planning. Especially given that the errors under investigation have been elicited using the same methods: tongue-twisters and the SLIP task.

An alternative account for non-canonical errors comes from cascading models of production. Cascading models of production allow articulation to begin prior to completion of phonological encoding. From this account non-canonical errors can be attributed to the cascading of partially activated representations that are competing

during phonological encoding. Articulation is initiated before a single phonological representation is selected and therefore the movement of the articulators exhibits properties of the partially activated representations. A cascading model can also account for canonical errors: only one representation, specifically the unintended one, cascades to articulation and therefore articulation only contains properties of the incorrect phoneme. Therefore a cascading account does not discriminate between canonical and non-canonical errors. Instead, different patterns of articulation simply reflect the extent to which phonological representations are partially activated.

A cascading account of non-canonical errors has several advantages over a staged account. Primarily, a cascading model can accommodate a "simultaneous activation" account of non-canonical speech errors. The repeated observation that non-canonical errors include properties of both intended and competing phonemes, rather than random articulatory or acoustic properties (Frisch, 2007; Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein et al., 2007; Mowrey & MacKay, 1990; Pouplier, 2003, 2007; Stearns, 2006), suggests that planning is involved in the production of these errors. Another advantage of a cascading account is it's ability to parsimoniously accommodate the possibility of canonical and non-canonical errors occuring along a continuum. In other words, there are no reasons to assume that canonical errors are a categorically different result of phonological encoding compared to non-canonical errors which, according to a staged account, would result from articulatory implementation. Indeed, most instrumental investigations have demonstrated non-canonical errors occur along a continuum (Frisch & Wright, 2002; Frisch, 2007; Goldstein et al., 2007; Pouplier, 2007).

Finally, a cascading account can also accommodate higher-level influences on non-canonical error production. For example, there is some evidence that lexical status increases the likelihood of a non-canonical error (e.g., Frisch & Wright, 2002; Goldrick & Blumstein, 2006, see Section 2.5 for further discussion). There are no reasons to assume that articulatory implementation, as proposed by a staged account, would differ for nonwords compared to real words.

Taken together, a cascading model of speech production can best accommodate non-canonical speech errors. Perhaps the strongest evidence for cascading of phonological representations to articulation comes from an acoustic investigation of speech errors. Goldrick and Blumstein (2006) measured the VOT of responses in a tongue-twister task. They observed the VOT of an errorfully produced /g/ differed from both a canonical /k/ and a canonical /g/. They interpreted this observation in the

context of a cascading model: the produced phoneme contained traces of both the target (e.g., /k/=.3) and the competitor (e.g., /g/=.7) and therefore the VOT duration was some length between those typical of a canonical /k/ and /g/. Goldrick and Blumstein (2006) explicitly argued that their observation of errorful VOTs, which were a duration between a canonical voiced and canonical voiceless phoneme, reflected partial activation of competing representations. Throughout this thesis, we will assume a cascading model of speech production that is consistent with Goldrick and Blumstein's (2006) interpretation. In the following section we detail the assumptions of the model and the consequences for investigating non-canonical speech errors.

2.4.2 A Functional Model

Throughout this thesis we will assume a cascading model of production. This model allows partially activated phonological representations to cascade to articulation. A consequence of assuming a cascading model is that it requires a new approach to the investigation of speech errors. As stated in the previous section, a cascading model can not differentiate between a canonical error and a non-canonical error. Both types of responses simply reflect differing degrees of partial activation of phonological representations.

In fact, to anticipate findings reported in this thesis, in Chapter 3 a speech error experiment is presented in which responses were categorised as "errors" or "correct" using transcription of auditory and articulatory records. We conclude from this investigation that categorising responses as errors is problematic if a cascading model is assumed. In other words using a discrete "error or not" methodology can not capture the extent to which there is partial activation of phonological representations in a response.

In order to sustain the proposed cascading model, the remainder of the thesis focuses on an *a priori* definition of speech errors. Rather than categorising experimental responses, a method is developed (see Chapter 4) to compare all articulations of responses in error-invoking conditions relative to articulations in non-error-invoking conditions. Using this approach to investigating speech errors grants the ability to measure the extent to which competition during phonological encoding influences articulation.

In order to refine a cascading model it is necessary to investigate how phonological representations become activated during phonological encoding. As stated previously, some articulations may include partial activation of phonological representations. In the following section we discuss how activation spreads across representations during phonological encoding. We focus on two aspects of information flow. First, we consider the influence of lexical representations on phonological representations. Then we discuss whether lower-level representations such as features are required.

In an aside, some researchers have proposed the notion of a mental syllabary (Crompton, 1982; Levelt, 1989; Levelt & Wheeldon, 1994; Meyer et al., 2003; Roelofs, 1997a), which we will not consider further. In general, syllabary-based models propose a stage in production that syllabifies the intended phonemes through the retrieval of syllabic units. For the purposes of the research presented throughout this thesis all theoretical arguments will take a neutral stance on syllabification. According to the proposed cascading account two partially activated phonemes can be simultaneously articulated. This does not rule out the possibility that these two phonemes activate two corresponding syllables which are both simulutaneously articulated.

2.5 Interactivity in Models of Production

Thus far, we have considered the flow of information from phonological encoding to articulation. We have discriminated between a staged and cascading flow of information and have proposed a functional model that allows partially activated representations to cascade to articulation. However, we have not discussed how representations become activated in order to flow to articulation. Therefore, it is important to explain how information flows during phonological encoding to activate the phonological representations that are articulated. In this section we discuss how different types of information flow influence the activation of representations. We first discriminate between different types of information flow and then present a detailed review of information flow between lexical and phonological representations and between phonological and feature representations.

Discrete models of phonological encoding propose the most basic flow of information from lexical to phonological representations. According to discrete models, a single lexical representation is selected and only the selected lexical representation will influence the activation of phonological representations (Levelt, 1989).

For example, in a seminal study by Levelt et al. (1991) participants were presented with pictures of objects (e.g., SHEEP) which had to be named. Crucially, between picture presentation and naming, participants were presented with a real word or nonword distractor and had to make a lexical decision choice. Distractors were manipulated in three ways: semantically related to the target (e.g., GOAT), phonologically related to the target (e.g., SHEET), or phonologically related to a semantic competitor (e.g., GOAL). The critical finding was that no priming occurred for distractors that were phonologically related to the semantic competitor. This was interpreted as evidence that competing representations at the lexical level (such as GOAT) do not activate their corresponding phonological representations.

In contrast to discrete models, *interactive* models propose that partially activated representations at a higher level (e.g., lexical) can influence activation of representations at a lower level (e.g., phonological). For example, two lexical representations can be partially activated (e.g., SHEEP, GOAT) and their activation can cascade to partially activate phonological representations (e.g., /ʃ/, /i/, /p/, /g/, /o/, /t/). However, the activation of SHEEP would be stronger than the activation of GOAT because GOAT only partially represents the intended concept to be spoken. In an experimental priming investigation R. R. Peterson and Savoy (1998) demonstrated that phonologically-related distractors of items with dominant (e.g., COUCH) and secondary (e.g., SOFA) names yielded priming. This finding is contrary to the results discussed above from Levelt et al.'s (1991) investigation. R. R. Peterson and Savoy (1998) interpreted their observation of priming as evidence for an interactive flow of information: activation from competing lexical representations flows to activate their corresponding phonological representations.

In the account provided by R. R. Peterson and Savoy (1998) a competitor phonological representation was activated and in turn facilitated lexical decision. More generally, once interactivity is incorporated into a model of production, the interactive flow of information serves to increase the activation of competitors. A *competitor* can be defined as a non-target representation that becomes activated during planning. To understand how competitors become activated in more detail consider the production of *dead cab* in Dell's (1986) spreading activation model. In this model, the lexical representation of the first target word[2] (*dead*) receives an arbitrary quantity of activation (60 units) and the representation of the second target word (*cab*) receives half the amount of activation (30 units). Since *cab* has partial

[2]Dell (1986) discusses activation in terms of morphemes, however since this example and all discussion throughout this thesis will be restricted to monosyllabic words, I will use "word" rather than "morpheme".

activation, it becomes a competitor. Activation of the target and competitor representations then spreads automatically to neighbouring representations (e.g., for *dead*: the lexical representation LIFE and phonological representations /d/, /ɛ/, /d/). These neighbouring representations, which previously were not activated, then become both activated and competitors. There is therefore an increase in competitor activation and a decrease in target activation. Since an interactive flow of information increases the activation of competitor representations, interactivity increases the likelihood of a target or competitor being selected for articulation.

In the example above, an interactive feedforward flow of information from lexical representations to phonological representations increased competitor activation. However, it is logically possible for phonological representations to also influence lexical representations. In other words, a *feedback* flow of information can also increase competitor activation. A critical component of Dell's (1986) model is that activation of phonological representations can feedback to lexical representations. In the example above, the activation of the lexical competitor LIFE will in turn activate the competitor phonemes /l/, /i/, /f/. Then, for example, the activation of /l/ (from LIFE) and the activation of /d/ and /ɛ/ (from DEAD) will feedback to the lexical level and activate the lexical representation LEAD. Therefore, the feedback flow of information further increases the number of competing representations.

In summary, there are three different ways that information can flow to yield activation of competing phonological representations. Information can flow in a discrete manner so lexical representation activation does not yield competitor phonological representation activation. Alternatively, information can flow in an interactive manner. An interactive flow of information may be feedforward to yield phonological competitor activation or an interactive flow of information may be feedbackward to yield phonological competitor and lexical competitor activation.

Throughout this thesis we investigate whether a feedback flow of information is required for models of speech production. Feedback can occur between different levels of speech production. For example, there can be phonological to lexical feedback (as above). We will discuss this in Section 2.5.1. There also can be feedback between subphonemic representations, such as features, to phonological representations. We discuss the latter in Section 2.5.2.

2.5.1 Phonological to Lexical Feedback: Lexical Bias Evidence

A primary source of evidence used to discriminate between feedforward and feedback models of production are phonological speech errors. Garrett (1976) argued that speech errors occur independent of lexical processing. According to this account substitutions are as likely to result in real words (e.g., *built to spill* →"spilt to bill") as nonwords (e.g., *loose fur* → "foose lur"). In two corpus analyses Garrett (1976) found that only 38–41% of substitutions yielded real words. Similarly, both Fromkin (1973) and Fry (1977) reported substitutions tended to result in nonwords. These analyses which were used as evidence to argue for lexically-independent phonological processing.

In another corpus analysis Dell and Reich (1981) provided evidence for a *lexical bias effect* in which real words were more likely to occur than would be predicted by chance. This analysis differed from the Garrett (1976) analysis because the proportion of real words was compared to a chance estimate rather than to the proportion of nonwords. Including a chance estimate is important because it can account for other factors that may increase the likelihood of an error, such as the size of the phonological neighbourhood of words involved in errors. For example, an onset substitution for the word *cop* has more chances to result in a real word (e.g., "mop", "pop", "top",...) than a word such as *cup*. Also, using a different chance calculation, Nooteboom (2005b) demonstrated a lexical bias in Dutch. However, using a similar chance estimate to Dell and O'Seaghdha (1991), del Viso, Igoa, and Garcia-Albea (1991) and Pérez, Santiago, Palma, and O'Seaghdha (2007) found no evidence for a lexical bias in Spanish speech errors (but see Hartsuiker, Antón-Méndez, Roelstraete, & Costa, 2006, who demonstrated a Spanish lexical bias in an error-elicitation task).

Corpus investigations have provided evidence of a lexical bias effect which suggests that lexical and phonological processing may have an influence on one another. However, the results of corpus analyses have been inconsistent. One source of inconsistency, suggested by Nooteboom (2005b), is that the calculation of chance can be difficult. Another potential source of inconsistency is the fact that speech error corpora are based on transcribed responses of everyday observations by investigators. Transcription biases for speech errors have long been reported in the literature (Ferber, 1991), which has even prompted researchers to argue against the use of speech error data all together (Meyer, 1992). For example, a large proportion of

speech errors may not be perceived and the source of the error can be difficult to determine (Ferber, 1991, see also Chapter 3 for a detailed discussion on transcription limitations).

An alternative method for investigating speech errors is through experimentally eliciting speech errors in a laboratory (Baars, 1992; Baars & MacKay, 1978). The basic premise of speech error elicitation paradigms is that errors are elicited by somehow creating competition between alternative speech plans (Baars, 1992). Experimental speech error investigations have two primary advantages over corpora analyses. The first and primary advantage is that experimental materials can be controlled to allow valid comparisons across conditions. Therefore, hypotheses about speech errors can be directly tested deeming it no longer necessary to estimate chance error rates. Second, responses can be recorded so that repeated listening can reduce the chances of a transcription bias such as the misperception of an error (Ferber, 1991).

One of the most often reported methods for eliciting speech errors is the SLIP task (originally reported by Baars et al., 1975; Baars & Motley, 1976). In this task each trial consists of silently reading several sequentially presented biasing items (e.g., *keet fime*) and then, on occasion, speaking out loud a target with reversed onsets (e.g., *feep kive*). The aim of the task is to create substitution errors (e.g., "keep five") through repeated priming. To investigate lexical bias, targets can be designed so that errors would result in real words (e.g., *feep kive* → "keep five") or nonwords (e.g., *feeb kise* → "keeb fise"). In the original SLIP investigation Baars et al. (1975, Experiment 1) observed 42 onset errors, of which 32 resulted in real words and 10 nonwords. This result established experimental evidence for a lexical bias.

Baars et al. (1975) interpreted their observation of a lexical bias as evidence for some form of editing of the speech signal. According to this account, the speaker monitors their speech plan using inner speech. This is accomplished by using a comprehension-based mechanism that parses the speech plan after phonological encoding, but before articulation (Levelt, 1989; Levelt et al., 1999; Roelofs, 1992, 2003; Roelofs & Hagoort, 2002). Because the comprehension system is unable to detect that real words are violations, real words are more likely to be spoken than nonwords. However, Experiment 1 of Baars et al.'s (1975) investigation was limited because it could not be determined whether the unbalanced proportion of real word to nonword errors was due to editing the production of nonwords or increasing the production of real words.

In order to discriminate between these editing accounts Baars et al. (1975) manipulated the context in which targets were presented in an additional experiment. Target word pairs were embedded into context lists which consisted either entirely of nonwords, or of a mixture of words and nonwords. In the mixed context, Baars et al. observed more phoneme exchanges resulting in real words than nonwords, replicating the lexical bias effect. In the nonword context, however, the exchange levels did not differ for real word and nonword outcomes. They reasoned that in conditions where real words were never encountered (nonlexical context), there was no need for a self-monitor to make use of a lexicality criterion (e.g., *is this a word?*) to monitor and edit the speech plan.

In a more recent SLIP investigation that manipulated word outcome and context in a manner similar to Baars et al. (1975, Experiment 2), a different pattern of errors was observed. Humphreys (2002, Experiment 4) observed a lexical bias effect in both the nonlexical (26 nonword, 43 real word exchanges) and mixed contexts (24 nonword, 40 real word exchanges). She interpreted her experimental results of a context-independent lexical bias effect as support for an interactive model of production which incorporates feedback from phonological to lexical representations (e.g., Dell, 1986).

Interactive models of production propose the activation of a representation automatically spreads to activate other neighbouring representations (Dell, 1986; Dell & O'Seaghdha, 1991, 1992). For example, the activation of RAT spreads to activate a neighbouring lexical representation (e.g., CAT) which in turn activate their corresponding phonological representations (e.g., /k/, /r/, /a/, /t/). According to this account the lexical bias effect results from feedback from phonological to lexical representations during which lexical representations are reinforced. Since by definition nonwords do not have lexical representations there is no activation to reinforce lexical representation, thereby reducing the chances of a nonword being produced.

A further prediction of Dell's (1986) feedback model is that the lexical bias effect will increase with longer durations of activation. This is because early in the production process competitors are activated almost as strongly as the target. With time (and therefore spreading activation) a real word target will become much more active through reinforcement, while a nonword competitor becomes less active through lack of reinforcement. In a SLIP task investigation by Dell (1986) it was demonstrated that there was no lexical bias for a 500ms deadline and the lexical bias effect increased across 700ms and 1000ms deadlines. This finding confirmed the prediction

of the model and established additional experimental support for feedback between phonological and lexical representation.

More recently, Hartsuiker et al. (2005) reported two further SLIP experiments with outcome and context factors. They observed a lexical bias effect in the mixed context (Experiment 1: 8 nonword, 20 real word; Experiment 2: 12 nonword, 41 real word), as was the case for Baars et al. (1975). However, in their experiments rather than fewer nonlexical outcomes, there were, in fact, more real word outcomes in the mixed context relative to the other three conditions. Hartsuiker et al. attributed the differences between their results and previous studies to experimental design differences. First, Baars et al. did not statistically test whether there was an interaction effect. Second, Baars et al. did not use counterbalanced target lists, rendering a comparison across contexts invalid. Third, Humphreys (2002), who showed a context-independent lexical bias, used a stringent time criterion such that participants were required to respond quickly and slow responses were discounted. Fast responses may have encouraged participants to rely less on self-monitoring (see Hartsuiker, 2006, for further discussion).

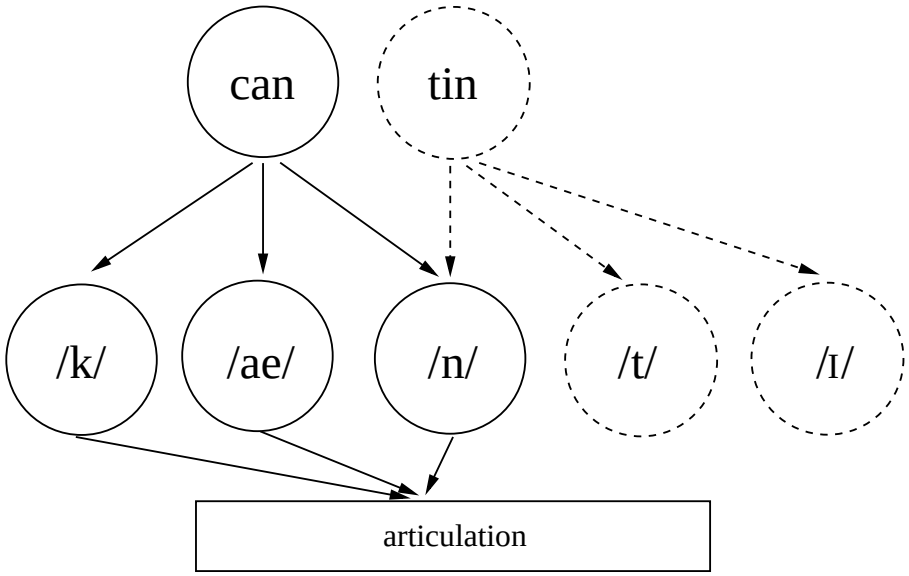
Having accounted for the differences between studies, Hartsuiker et al. (2005) suggest that their findings support a view in which both feedback *and* monitoring play a role. According to this account the lexical bias is initially caused by feedback, but self-monitoring in the nonlexical context increases the likelihood that errors resulting in real words are filtered. Thus the lexical bias is only manifest in the mixed context. On this account, the monitor is functionally adaptive: In the nonlexical context, it is possible to determine that anything resulting in a real word is an error. In the mixed context, on the other hand, participants encounter both real words and nonwords, and the intention of uttering one or the other provides no evidence as to whether it is likely to be an error. Hence the monitor has no functional role to play in the mixed context. Similarly, Nooteboom and Quené (in press) have recently proposed a feedback with monitoring model to account for their distribution of SLIP task data.

Collectively, experimental investigations of speech errors have provided clear evidence of a lexical bias effect, but contextual influences have been a topic of debate. The current speech error evidence appears to implicate feedback (Hartsuiker et al., 2005; Humphreys, 2002; Nooteboom & Quené, in press), however the role that a monitor may play is less clear (Hartsuiker et al., 2005; Nooteboom & Quené, in press). All of the theoretical accounts of the lexical bias effect discussed in this section assumed a staged model of speech production: a phonological representation

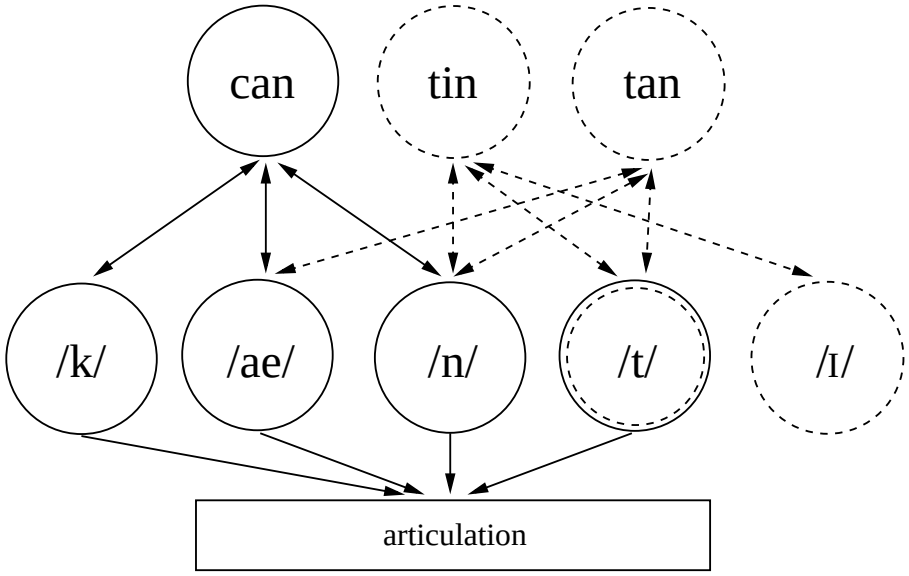
is more likely to be misselected when the misselection yields a real word. However, if a cascading model is assumed it is not clear whether these accounts of the lexical bias hold.

In particular, there is some evidence that non-canonical errors are influenced by lexical status, though these findings have been mixed. In an acoustic analysis of tongue-twister data, Frisch and Wright (2002) demonstrated that /s/ → /z/ errors were more likely for real word competitors, but that /z/ → /s/ were less clear. For the latter there was a clear lexical bias for canonical substitutions, but non-canonical rates of percent voicing were more likely for nonword competitors. It is possible that this asymmetric pattern can be attributed to frequency effects: the /s/ → /z/ real word competitors (zit, zip, zap, zoo) were lower frequency words compared to the /z/ → /s/ real word competitors (sit, sip, sap, sue). It has been well-established in the speech error literature that low-frequency words are more prone to errors than high-frequency words (Dell, 1988, 1990; Stemberger, 1984; Stemberger & McWhinney, 1986). In a different post-hoc analysis on the VOT duration of substitution errors, Goldrick and Blumstein (2006) demonstrated that the VOT of an errorfully produced target was more similar in duration to the VOT of the competitor phoneme if the competitor phoneme yielded a real word. Lastly, a SLIP investigation in which articulation was measured with EMMA did not reveal a positive or negative lexical bias, but the error rates used in the analysis were very low to warrant the statistical analysis (Pouplier, 2003). One consideration about these mixed findings is that all analyses of lexical influences on non-canonical errors have been *post-hoc*; experiments were not specifically designed to investigate lexicality. Additionally, none of the non-canonical speech error investigations manipulated context.

Given the fact that the theoretical accounts of the lexical bias effect have assumed staged planning, and there is some evidence that non-canonical errors are influenced by lexical processing, it is necessary to re-evaluate lexical bias accounts in a cascading model of production. As previously stated, stage-based models of production must attribute non-canonical errors to articulatory implementation and therefore are not able to account for lexical planning influences on the rate of non-canonical speech errors. There is no reason why articulatory implementation should be affected by lexical status. A cascading account, on the other hand, can easily account for a lexical bias effect if the model also includes feedback from phonological to lexical representations.



(A)



(B)

Figure 2.1: Lexical bias effect predictions of a feedforward and feedback cascading model; solid lines represent strong activation, dotted lines represent weak activation, and a dotted line within a solid line represents initially weak activation that was reinforced and became strong activation. Only strong activation can cascade to articulation. (A) represents a feedforward model in which only strong activation of the phonological representations /k/, /æ/ and /n/ cascade to articulation; (B) represents a feedback model in which the activation of the phonological representations /k/, /t/, /æ/ and /n/ cascade to articulation because /t/ is reinforced by a feedback flow of information.

To illustrate a cascading account of the lexical bias effect, two different models of production are presented in Figure 2.1. The rules of the models are simple: there can be strong activation (solid lines) or weak activation (dotted lines). Activation that is initially weak can be reinforced and become strong (dotted line within a solid line). Only strong activation can cascade to articulation.

In both models the target response is *can* from the phrase *tin can*. The lexical representation for CAN has strong activation and the lexical representation for TIN has weak activation. Once the lexical representations become activated, their corresponding phonological representations also become activated. Since TIN only has weak activation the phonological representations (/t/, /ɪ/) also only receive weak activation. In the feedforward model (A) only the strongly activated phonological representations (/k/, /ae/, /n/) cascade to articulation. However, in the feedback model (B) activation from phonological representations feeds back to activate the lexical representation TAN. Then the corresponding phonological representations are activated. Because /t/ already had weak activation from TIN, its activation level becomes stronger as it is being reinforced by TAN. As a result, the strongly activated phonological representations (/k/, /ae/, /n/, and /t/) cascade to articulation. Therefore, if there is cascading activation from phonological encoding to articulation, it appears that only a model that includes feedback can account for the lexical bias effect.

In summary, several accounts of the lexical bias effect have been proposed based on speech error evidence (Baars et al., 1975; Dell, 1986; Levelt, 1989; Levelt et al., 1999; Hartsuiker et al., 2005; Humphreys, 2002). However, these accounts have all been based on a staged model of production, which is problematic given the evidence for non-canonical speech errors. A cascading model of production, which allows the activation of phonological representations to flow to articulation, can account for the occurences of non-canonical errors and can also account for the lexical bias effect if there is feedback from phonological to lexical representations.

In order to investigate the feedback predictions of a cascading model, we report an experiment in Chapters 3 and 5 designed to investigate lexical competitor and context influences on articulation. We discuss our approach in more detail in Section 2.6. However, before continuing there is one additional component of a cascading model that must be considered. Some models of production have proposed that in addition to lexical and phonological representations, feature representations are

required (e.g. Dell, 1986; Dell et al., 1993). In the following section we present a discussion of whether features are required in a cascading model of production, and if so, whether feedback is required between feature and phonological representations.

2.5.2 Featural to Phonological Feedback

So far our discussion of a cascading model of production has been based on the assumption that phonological representations can cascade to articulation. According to this model non-canonical speech errors can be attributed to the simultaneous articulation of two partially activated phonological representations. However, it is possible that a model of production including representations at a level lower than the phonological level can account for non-canonical speech errors. Goldstein et al. (2007) and Mowrey and MacKay (1990) have specifically argued ocurrences of non-canonical speech errors implicate a role for lower-level planning units. Goldstein et al. (2007) interpreted their observation of non-canonical speech errors as evidence for gestural unit planning. According to this account two phonological representations can not be selected, but two competing gestural representations can be selected for articulation. Mowrey and MacKay (1990) also interpreted their non-canonical speech error observation as evidence for lower-level planning units. In their EMG investigation they observed inappropriate muscle activation that is normally associated with /l/ production during repetitions of *bay* in the phrase *Bob flew by Bligh Bay*. They argued that the abnormal muscle activity reflected planning at a subfeatural level of representation. A subfeatural representation rather than featural representation was proposed because muscle activity was only recorded at one point on the tongue. They therefore could not conservatively infer whether an entire featural representation was misactivated in the error.

It is important to note that the nature of the proposed lower-level representations differs between Goldstein et al.'s (2007) and Mowrey and MacKay's (1990) accounts for non-canonical errors. Goldstein et al.'s (2007) account is based on a gestural model of speech motor control while Mowrey and MacKay's (1990) account is based on features. Feature-based theories posit that phonological representations include a set of phonological features that specify articulatory or acoustic symbolic units and do not include any temporal specifications (e.g., Chomsky & Halle, 1968). On the other hand, gestural theories propose an alternative to phonological representations called gestural constellations which contain primitive action units (gestures) specifying spatio-temporal characteristics (e.g. Browman & Goldstein, 1989). The distinction between these types of representations is beyond the scope

of the research questions presented throughout this thesis. Since features have been discussed more widely in the psychological literature, we will discuss lower-level representations using feature terminology. This specification does not rule out that lower level representations can be described as having a different nature, but merely provides a language for a discussion on lower level representations.

Models of speech production differ with regards to whether feature representations are required during phonological encoding. On one hand, there are models which propose phonological representations are the lowest level units during phonological encoding. For example, in the Weaver++ model of production, phonological encoding is completed when a phonological representation has been selected. In this model, finer-grained representations for the selected representations are not activated until a later stage of planning called phonetic encoding (Levelt et al., 1999). On the other hand, other models of production propose that feature representations are required during phonological encoding (Dell, 1986; Dell et al., 1993). For example, Dell's (1986) spreading activation model, like Levelt et al.'s (1999) model, involves the selection of a phonological representation at the end of phonological encoding. However, unlike Levelt et al.'s (1999) model, the selection of phonological representations can be influenced by feature representations.

A primary source of evidence for features in models of production comes from the so-called phonological similarity effect: phonological speech errors are more likely to occur if the interacting phonological representations are similar as opposed to if they are dissimilar (Butterworth & Whittaker, 1980; Dell & Reich, 1981; del Viso et al., 1991; Kupin, 1982; Levitt & Healy, 1985; Nooteboom, 2005a, 2005b; Shattuck-Hufnagel & Klatt, 1979; Shattuck-Hufnagel, 1986; Stemberger, 1982; Vousden et al., 2000; Wilshire, 1999). The similarity between phonological representations increases with the more features they have in common. Evidence for the phonological similarity effect has been established in both corpora analyses (Dell & Reich, 1981; del Viso et al., 1991; Stemberger, 1982; Shattuck-Hufnagel & Klatt, 1979; Shattuck-Hufnagel, 1986) and experimental investigations (Butterworth & Whittaker, 1980; Kupin, 1982; Levitt & Healy, 1985; Vousden et al., 2000; Wilshire, 1999) of speech errors.

An important consequence of incorporating features into a model of production is the logical possiblility to have feedback from feature representations to phonological representations. In fact, incorporating feedback between featural and phonological representations provides a straightforward account for the phonological similarity effect (Dell, 1986; Stemberger, 1982, 1985a). According to this account activation

from competing phonological representations (e.g., /d/ and /g/) flows to activate the corresponding feature representations (e.g., <voice+, alveolar+>, <voice+, velar+>). The feature activation then feeds back to reinforce the phonological representations. When competing phonological representations share more features in common, they will receive a greater amount of reinforcement. A representation receiving additional activation via reinforcement is more likely to be selected for articulation. Therefore, a similar phonological representation is more likely to be produced in an error.

Ultimately, there is some evidence that suggests features are required for models of production. First, articulatory investigations suggest lower-level representations may be involved in non-canonical speech errors (Goldstein et al., 2007; Mowrey & MacKay, 1990). Second, evidence for phonological similarity suggests competing phonological representations with features in common are more likely to yield an error. However, these sources of evidence for feature representations have been based on staged models of production.

Throughout this thesis we assume that speech production is a cascading process. In a cascading model, non-canonical speech errors may be accounted for by either phonological or featural representations cascading to articulation. For example, a non-canonical articulation which includes tongue tip raising and tongue-dorsum raising could result from the partial activation of /t/ and /k/ or the partial activation of the associated features: <alveolar+, velar+, voice–>. It is therefore important to evaluate whether features are required in a cascading framework. Furthermore, if features are required it is necessary to evaluate whether feedback between feature and phonological representations is also required. In Chapters 6 and 7 we present a more detailed discussion on the role of features in models of production. We also present two articulatory and acoustic investigations designed to investigate the role of features in a cascading model of production.

2.6 Structure of this Thesis

This thesis investigates the implications of cascading from phonological encoding to articulation for models of speech production. Throughout this thesis we assume that partially activated representations can cascade to articulation. Two major questions are addressed: first, we investigate whether feedback is required in a cascading model of production; second, we investigate whether feature representations are required.

In order to investigate whether feedback is required we first focus on the lexical bias effect. In Chapter 3 we present an auditory and articulatory transcription analysis of speech errors. Speech errors were elicited using a word order competition (WOC) task; an error elicitation paradigm designed to investigate the influences of context and lexical status on error production. Articulation was recorded using electropalatography (EPG), a technology that allows the direct measurement of tongue-to-palate contact over time. We also present a perceptual experiment that investigates whether non-canonical speech errors can be auditorily perceived. We conclude from these investigations that many more errors are produced than can be detected using transcription and, importantly, that categorising responses as errors has limitations.

In order to investigate articulation without using categorisation, we developed a new method for EPG data analysis which is reported in Chapter 4. This technique, the Delta method, is a relative measure of articulatory variability that does not require responses to be categorised. From Chapter 5 onwards we use an *a priori* definition of "error". Instead of categorising responses, we investigate the variability in articulation of responses during error invoking conditions (e.g., an error-elicitation paradigm and tongue-twisters) relative to comparable responses recorded during non-error invoking conditions. Analysing the variability of responses in this way grants the ability to investigate cascading processing without having to categorise responses as having one phonological representation or another.

In Chapter 5 we present a reanalysis of the articulations recorded for the experiment in Chapter 3. For this analysis, we use the Delta method to compare the articulations recorded in the WOC task relative to articulation recorded in a comparable non-error invoking task. The primary result of this analysis is that patterns of articulation are more similar to the competing phonological representation when that representation yields a real word. We argue that this pattern strongly implicates feedback in a cascading model of production.

In the remaining two experimental chapters, we investigate whether features are required in a cascading model of production. In Chapter 6 we present an acoustic and EPG analysis of articulation recorded during tongue-twisters. The tongue-twisters were designed to investigate phonological similarity influences on variability of articulation. We demonstrate with EPG and measurements of VOT that articulation is most variable when there is competition between phonological representations that differ by one feature compared to phonological representations that differ by two features. We argue that this pattern can be best accounted for by a cascading

model of production that incorporates feature representations and includes feedback between feature and phonological representations.

In Chapter 7 we extend the research presented in earlier chapters by using ultrasound; an alternative articulatory imaging technique. We first adapt the Delta method for ultrasound analysis and then present a replication of the phonological similarity results from Chapter 6. We additionally extend the previous findings, which were limited to stop consonants, to include fricatives. The results from this chapter are largely consistent with Chapter 7, establishing further evidence for the role of features in a cascading model, feeback between feature and phonological representations and validating the Delta method as a useful method for articulatory analysis.

Finally, in Chapter 8 we conclude that analysing articulation using a technique that does not require categorisation allows the consequences of a cascading model to be investigated. We argue for a cascading model of production that includes feedback between phonological and lexical representations and feedback between feature and phonological representations. We discuss the broader implications of these findings and potential directions for future research.

Chapter 3

Transcription Investigation of the Lexical Bias Effect[1]

3.1 Chapter Overview

This chapter presents a study designed to investigate the extent to which lexical and phonological representations interact. Experiment 1 is a Word Order Competition (WOC) task designed to investigate the nature of the lexical bias effect. The results of the experiment provide evidence for a feedback account of the lexical bias effect. However, a comparison between an auditory and an articulatory transcription of speakers' responses reveals that many more errors are produced than are transcribed. Experiment 2 addresses the perceptual consequences of such errors and demonstrates that they can be perceived by listeners. The limitations of transcription are discussed.

3.2 Introduction

There is an ongoing debate in the speech production literature about whether feedback is required in models of speech production. While there are feedforward models which only allow a unidirectional flow of information from lexical to phonological representations (Levelt, 1989; Levelt et al., 1999; Roelofs, 1992, 1993, 1996, 1997b), there are also feedback models of production allowing for a bidirectional flow of information between lexical and phonological representations (Dell, 1986; Dell & O'Seaghdha, 1991, 1992; Dell et al., 1997; Harley, 1993). These two classes of models make different predictions about aspects of speech production. One primary

[1]Portions of the data reported in Experiment 1 are included in McMillan, Corley, and Lickley (in press)

source of evidence used to discriminate between the models is phonological speech errors.

Phonological speech error investigations have established firm evidence for a lexical bias effect (Baars et al., 1975; Dell, 1986; Dell & Reich, 1981; Humphreys, 2002; Hartsuiker et al., 2005; Nooteboom, 2005a, 2005b; Nooteboom & Quené, in press, but see del Viso et al., 1991 and Perez et al., 2007 and see also Section 2.5.1 for a detailed discussion). Dell and Reich (1981) demonstrated in a speech error corpus analysis that real word outcome errors were more likely to occur than would be predicted by chance. Similarly, Nooteboom (2005b) also demonstrated evidence for a lexical bias in a corpus analysis of Dutch speech errors. Experimental investigations of speech errors, which allow materials to be controlled across conditions, have also provided substantial evidence for a lexical bias (Baars et al., 1975; Dell, 1986; Humphreys, 2002; Hartsuiker et al., 2005; Nooteboom & Quené, in press). For example, using the SLIP task, Baars et al. (1975) provided evidence that exchange errors occurred more often for real word outcome errors than nonword outcome errors.

Feedforward and feedback models of production offer different accounts for the lexical bias effect. Feedforward accounts attribute the occurrences of more real word outcomes than nonword outcomes to self-monitoring (Baars et al., 1975; Levelt, 1989; Levelt et al., 1999; Nooteboom, 2005a, 2005b). After a speech plan has been generated, but before it is articulated, it is checked using a comprehension-based monitor. The monitor parses the speech plan just as it would normal speech and, if it detects a violation, passes the message back to the production system to be edited. The monitor can make use of a lexicality criterion (*Is this a word?*: Baars et al., 1975; Levelt, 1989; Levelt et al., 1999) or compare the generated speech plan to the intended output (*Is this what I intended to say?*: Nooteboom, 2005a, 2005b). In both cases, a real word error is less likely to be detected as a violation than a nonword error.

In contrast, feedback accounts attribute the lexical bias to an automatic process (Dell, 1986). According to this account, activation of lexical representations spreads to activate phonological representations. The phonological representation activation can then feed back and reinforce the lexical representations. Since, by definition, real words have lexical representations and nonwords do not have representations, real words can be reinforced by feedback while nonwords can not. As a result, real word outcome errors are more likely to be produced.

In an effort to discriminate between these accounts, three very similar SLIP task experiments investigated the effect of context on the lexical bias effect (Baars et al., 1975; Hartsuiker et al., 2005; Humphreys, 2002). In each study, the methodology was broadly similar. Target word pairs were embedded into context lists consisting of either entirely nonwords (nonlexical context), or a mixture of words and nonwords (mixed context). In each trial, participants silently read several sequentially presented biasing items (e.g., *keet fime*). On occasion they were cued to repeat aloud a target with reversed onsets (e.g., *feep kive*). Targets were designed so that exchange errors could result in real words (e.g., *feep kive* → "keep five") or nonwords (e.g., *feeb kise* → "keeb fise"). A feedforward model predicts that context will influence the lexical bias effect due to differences in monitoring across contexts, while a feedback account predicts a context-independent lexical bias effect since the reinforcement of lexical representations is an automatic process.

Each of these comparable investigations yielded a different pattern of results. In an investigation by Baars et al. (1975, Experiment 2) a lexical bias effect was observed in the mixed context but not in the nonlexical context. The pattern of errors included fewer nonword outcomes in the mixed context relative to the other conditions (though the interaction was not tested statistically). This pattern was interpreted as evidence for self-monitoring in a feedforward model. According to this account, in the mixed context nonword errors were edited by a monitor using a lexicality criterion. However, in the nonlexical context, since only nonwords were encountered, Baars et al. (1975) reasoned speakers did not monitor their speech plan.

In a different SLIP task investigation, Hartsuiker et al. (2005) also only observed a lexical bias effect in the mixed context. However, in their investigation there were more real word outcome errors in the mixed context relative to the other conditions. They proposed that the lexical bias effect resulted from feedback that reinforced the activation of lexical representations. As a result, real words were produced more often than nonwords in the mixed context. They further argued that monitoring could account for the pattern observed in the nonlexical context. According to this account, the monitor is functional and edits the speech plan for inappropriate content (see also Hartsuiker, 2006). Since only nonwords should be produced in the nonlexical context, the monitor edited the production of real words.

Lastly, Humphreys (2002, Experiment 4) observed a context independent lexical bias effect: real word outcome errors were greater than nonword outcome errors in both contexts. This pattern is consistent with a feedback-only account: lexical

representations are reinforced by a feedback flow of information from phonological representations (Dell, 1986). Since lexical representations for real words can be reinforced, and there are no representations for nonwords to be reinforced, the production system produced a higher proportion of real word outcome errors.

Taken together, three previous investigations have yielded inconsistent findings for context influences on the lexical bias effect. This has prompted researchers to argue for models of speech production that include self-monitoring (Baars et al., 1975), feedback (Humphreys, 2002), or both (Hartsuiker et al., 2005). However, almost all previous SLIP task investigations (Baars et al., 1975; Daneman, 1991; Dell, 1986; Hartsuiker et al., 2005; Humphreys, 2002; Motley, Baars, & Camden, 1981, 1983; Nooteboom, 2005a; Nooteboom & Quené, in press, 2007) have assumed a staged account of production. According to this account, any speech errors produced are considered canonical substitutions which result from the misselection of a competing phonological representation.

However, several instrumental speech error investigations have demonstrated that articulation can include acoustic or articulatory properties of both competitor and target phonological representations (Frisch, 2007; Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein et al., 2007; Laver, 1980; Mowrey & MacKay, 1990; Pouplier, 2003, 2007, see Section 2.3.2 for a detailed discussion). In a recent investigation, articulation was recorded using electromagnetic midsagittal articulometry (EMMA) during the SLIP task (Pouplier, 2007, see also Pouplier, 2003). In each trial participants were presented with three priming pairs. The first pair (e.g., *nap flip*) always contained the same rime as the target (e.g., *tap kip*) and the second two priming pairs contained the same onsets as the target, but in reverse order (e.g., *case tick* and *can tim*). Pouplier (2007) categorised responses as "correct", an "intrusion error" (non-target gesture articulation), and a "reduction error" (reduced target gesture articulation). She observed that 28% of responses included either an articulatory intrusion (17%) or reduction (11%). This establishes clear evidence that the SLIP task can give rise to articulatory movements which do not correspond to canonical phonemes. However, Pouplier's (2007) study did not include an analysis of lexical outcome due to a low number of errors.

The evidence for non-canonical speech errors observed in the Pouplier's (2007) study is consistent with both staged and cascaded models of production. A staged-based model account would attribute the observed errors to difficulties with articulatory implementation. Since staged models only allow one phonological representation (for each frame) to be selected during phonological encoding, the simultaneous

articulation of an additional phoneme can not result from a planning error during phonological encoding. On the other hand, a cascaded model of production can also account for the non-canonical errors. According to this account activation of competing phonological representations can spread to articulation. Therefore, it is possible that the errors resulted from planning during phonological encoding, not from articulatory implementation. However, there is no direct evidence to support a cascading account since higher-level influences on planning were not tested.

Given the possibility that Pouplier's (2007) observation of non-canonical speech errors may reflect the cascading of partially activated phonological representations, it is important to reevaluate the existing evidence for context and lexicality influences on speech errors. In this chapter we present a speech error elicitation experiment designed to investigate context and lexical status on speech errors. But before describing the experiment, it is important to review two potential methodological limitations of the previous inconsistent SLIP task investigations.

The first potential limitation is that the previous investigations all relied on auditory transcription. The high rate of non-canonical errors observed by Pouplier's (2007) raises the possibility that many more errors were produced in the previous inconsistent investigations (Baars et al., 1975; Hartsuiker et al., 2005; Humphreys, 2002) than were detected during transcription. Problems with auditory transcription of speech errors have long been highlighted in the literature (Cutler, 1982; Ferber, 1991; Fromkin, 1980; Stemberger, 1985b; Tent & Clark, 1980; MacKay, 1980). Ferber (1991) argued that transcription suffers from two major problems: a perceptual bias (some errors are harder to perceive than others) and transcription error. The latter was largely a criticism of on-line transcription, however transcription errors can be reduced through repeated listening to audio recordings (Stemberger, 1985b). Nonetheless, Ferber's (1991) analysis of transcriber accuracy revealed that none of the errors in her radio corpora were more likely to be correctly (or incorrectly) transcribed than other errors. If anything, her data suggested if there was any bias at all it was due to a transcriber bias, but not an error type bias. However, her analysis only focused on transcription differences between phonological, grammatical, and lexical errors.

The most relevant criticisms of auditory transcription are those concerned with the detection of non-canonical substitutions or double articulations (Buckingham & Yule, 1987; Laver, 1980; Mowrey & MacKay, 1990; Pouplier & Goldstein, 2005). For example, Laver (1980) suggested that non-canonical errors are reported as being very rare due to a perceptual artefact: listeners "edit" the speech signal to hear

canonical phonemes (Boomer & Laver, 1968). This argument is consistent with the categorical perception literature, which has demonstrated listeners are equally likely to categorise between-boundary phonemes as belonging to one category or another and also that reaction times are longer when making decisions about between-boundary phonemes (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, 1997; Pisoni & Tash, 1974). Similarly, phonemic restoration effects have been reported in which listeners do not hear extraneous sounds like coughing (e.g., Warren, 1970). Therefore it is possible that any non-canonical error produced in the SLIP task may go undetected.

The second possible limitation of previous investigations is concerned with the nature of the SLIP task. The SLIP task method relies on repeated priming of phonological onsets. For example, a typical trial would include the following materials: *kalve taith*, *whack foul*, *kims tigs*, *kilm tidge*, and *tiss kint* (materials from Hartsuiker et al., 2005). The basic premise behind the task is that repeated presentation of biasing items (e.g., "/k/... /t/...") will in some way compete with the occasional presentation of the target (e.g., "/t/... /k/...") and that any resulting error will reflect a speech planning process (Baars, 1992). However, Baars and Motley (1976) reported that speakers sometimes sub-vocalised the biasing items which speakers were instructed to read silently to themselves. Therefore, the errors produced in the SLIP task may result from repeating articulatory movements. This potential confound of the SLIP task would be especially pertinent if articulatory measurements were also recorded during the task because it would not be possible to discriminate between a staged (motor) account and a cascaded (planning) account for errors. In fact, recent experimental investigations on the articulation of speech errors, including Pouplier's (2007) study discussed in the previous section, have argued that the demands of articulating alternating consonants may give rise to non-canonical speech errors (for example, abnormal tongue-tip raising during velar productions Goldstein et al., 2007; Pouplier, 2003, 2007).

An additional challenge for methods based on repeated priming is that in a cascading model of speech production, which is assumed throughout this thesis, it is not possible to discriminate between partial activation of a currently competing representation and residual activation of a previously competing representation. For example, an articulation that includes a blend of a /k/ and /t/ could reflect partial activation from the current experimental item (*keam turve* onsets) or activation of the current /k/ in *keam* and residual activation of the /t/ in the prime. In fact, priming investigations are based on the premise that residual activation of a prime

will in some way influence the processing of the subsequent target (Bock, 1996). Therefore, in order to investigate the influence of local competition on production, the potential confound of residual activation must be eliminated.

Lastly, and of more practical importance, a disadvantage of the SLIP task is its length. First, the materials required to generate all of the priming materials in addition to the experimental items are limited. Therefore, it is difficult to accurately control filler items across conditions. This is especially important when investigating the effects of context in a block design. Second, and most importantly, the demands put on the participant could potentially be reduced with a shorter experiment. Experiment 1 of this chapter includes an articulatory investigation that requires participants to wear an artificial implement in their mouths not only during the experiment, but also for at least 20 minutes beforehand to allow for accommodation. An experiment that allows context and lexical outcome to be manipulated, but is shorter, will reduce any unnecessary discomfort.

3.3 Experiment 1

In order to investigate the extent to which lexical and phonological levels of representation interact, a WOC task (Baars & Motley, 1976) was used for Experiment 1. In this task participants see pairs of words or nonwords, which disappear and are followed by an arrow on the screen. Participants are told to repeat the words they have just seen aloud, in the direction of the arrow (i.e., for *tup golve* followed by a right arrow, the correct response is "tup golve"; followed by a left arrow, "golve tup"). Items followed by left arrows may give rise to onset exchanges rather than full exchanges of words (participants may say "gup tolve" when the prompt following *tup golve* points left). The use of a directional arrow cue, rather than repeated priming, eliminates potential influences of repeated priming on target production.

The design of the WOC task included both lexical outcome and context manipulations similar to previous SLIP task investigations. We use the term *competitor*, to refer to substitutions that might result in nonwords (e.g., *golve tup* → "gup tolve"), or in real words (e.g., *tum gop* → "gum top"). We also manipulated the *context* in which target pairs were embedded: nonlexical (all nonwords) vs. mixed context (nonwords and real words). Each participant saw a full set of target word pairs counterbalanced across two (blocked) context conditions, resulting in a fully within-participants design.

Experiment 1 innovates from previous comparable SLIP investigations (e.g., Baars et al., 1975; Hartsuiker et al., 2005; Humphreys, 2002) because articulation was recorded for a subset of speakers in the WOC task. Measuring articulation with electropalatography (EPG) allows the measurement of tongue-to-palate contact across time. An investigation of the articulatory record allows for a more detailed analysis of the elicited responses. In particular, if a cascading model of production is accepted in which target and competitor phonemes can simultaneously affect production, an auditory transcription may not be able to capture these responses. For comparison we report both auditory and articulatory transcription results.

3.3.1 Method

Participants

Forty-eight native speakers of English from the University of Edinburgh and Queen Margaret University research community participated in the experiment. Eight of these participants were additionally recorded using electropalatography (EPG). We excluded one of the speakers recorded with EPG from all analyses due to a difficulty with responding to stimuli at the experimental presentation rate. All participants were treated in accordance with the University of Edinburgh and Queen Margaret University ethical guidelines.

Materials

Competitor Pairs The targets consisted of 96 CVC(C) nonwords, 24 each with the onsets /k/, /g/, /t/, and /d/. Each target ended in a bilabial or labio-dental phoneme (/p/, /b/, /f/, /v/, /m/), sometimes preceded by a liquid (/l/). The targets were designed to achieve firm tongue contact with the EPG palate at word onset while minimising the amount of tongue contact at word offset. Due to restrictions in the targets and in the generation of competitor pairs (see below), only 92 unique nonwords could be used in the experiment. Four of these were repeated to complete the design.

Each velar onset target (/k/, /g/) was paired with an alveolar onset target (/t/, /d/) to generate 48 *competitor pairs*, yielding 12 pairs of each onset combination (/d/–/g/, /t/–/g/, /d/–/k/, /t/–/k/). Targets were paired on the basis of their *competitors*—that is, the type of outcome that would be observed if participants were to exchange the onsets of the words in the pair. Half of the pairs had real word competitors, such that an exchange would result in two phonologically well-formed

words; half had nonword competitors, where the result of an exchange would be two nonwords. For each real-word competitor pair there was a phonologically similar nonword pair. For example, the real word competitor pair *keam turve* was matched to the nonword competitor pair *keeb turp*. The 24 competitor pairs in each category (real-word and nonword) included six pairs of each of the onset combinations. For presentation in the experiment, half of each set of six pairs was reversed so that participants were equally likely to encounter pairs with onsets /t/–/g/ or /g/–/t/. Refer to Appendix A for a complete list of competitor pairs.

Foil Pairs In addition to the competitor pairs, 48 nonword foil pairs were generated. The foils consisted of 12 pairs of each combination of /s/–/m/, /s/–/n/, /r/–/m/, and /r/–/n/. The foil pairs were generated to obscure the matched alveolar-velar pattern of the competitor pairs and therefore did not contain matched alveolar and velar onsets. All foil pairs had nonword competitors so that any onset substitution error could only yield nonword outcomes.

Contexts Competitor pairs were embedded in either *nonlexical* contexts (participants never saw real words), or *mixed* contexts (participants saw a total of 62% nonwords and 38% real words, including competitor, foil, and context pairs). The two contexts were created independently. The nonlexical context consisted of 150 pairs of nonwords. The mixed context consisted of 75 pairs of real words and 75 pairs of nonwords. None of the context items contained any of the onsets (/k/, /g/, /t/, /d/) used in the competitor pair items.

Experimental Lists Four experimental lists were generated to present the competitor, foil, and context pairs in a fully counterbalanced design. This was accomplished by creating two lists of competitor pairs, each consisting of 12 real-word competitor pairs and 12 nonword competitor pairs and organised such that if a real-word competitor pair appeared in one list, the corresponding nonword competitor pair appeared in the other list. A further 24 foil pairs was added to each list and then each list was combined once with each context: nonlexical and mixed. This yielded four context-list combinations comprising 198 pairs each (24 competitor pairs, 24 foil pairs, 150 context pairs). Competitor pairs and foil pairs were randomly distributed in the context lists and each competitor and foil pair was preceded by 2-4 context pairs. Finally, two experimental treatments were created, each consisting of a nonlexical context list and a mixed context list such that each participant saw every competitor pair over the two lists.

Apparatus

The experiment took place in a sound-treated recording studio at either University of Edinburgh or Queen Margaret University.

Acoustic–Only Recording The acoustic signal of participants responses were recorded on to a DAT recorder with a Sony ECM-TS125 condenser microphone and digitally converted into .wav files with a 22,050Hz sampling rate. Stimuli were presented on a 15" LCD monitor using a desktop computer and E-Prime software (Schneider, Eschman, & Zuccolotto, 2002). Audio materials were played over stereo headphones.

Acoustic & EPG Recording Prior to testing, each participant was fitted with a custom electropalatography (EPG) palate (manufactured by Incidental, Newbury, UK or Grove Orthodontics, Norfolk, UK) moulded to fit a dental cast from an impression of the hard palate. The EPG palate is made of acrylic and contains 62 embedded silver contacts on the lingual surface of the artificial palate, organised in 7 rows of 8 contacts and 1 row of 6 contacts (see Figure 4.1 for a a sample illustration of an EPG record). EPG data was recorded at a rate of 100Hz using the WinEPG system (Articulate Instruments Ltd, Edinburgh, UK), which connected the palate to a multiplexer unit that transferred the data to an EPG3 scanner and then to the serial port of a desktop computer. Acoustic recordings of participants' responses were recorded at 22,050Hz using an Audio Technica ATM10a microphone. A desktop computer, to which the microphone and WinEPG system were attached, was used to record participants' responses with Articulate Assistant (Wrench, 2003) software. A laptop computer was used to control stimulus presentation, using E-Prime (Schneider et al., 2002): participants saw word pairs on a 15" LCD monitor and auditory signals were played over stereo headphones.

Procedure

Once participants were seated, they were given instructions to repeat aloud, as quickly as possible and in the direction of an arrow which appeared, the word pairs that had appeared on the screen. For example, a correct response to *perch house* followed by a right arrow would be "perch house"; when followed by a left arrow, it would be "house perch". Participants were told to respond as quickly as possible to each pair.

All competitor pairs were followed by a left arrow. Foil pairs were also followed by left arrows, to prevent participants noticing the alveolar-velar pattern of competitor

pairs. All other items were followed by right arrows, creating a 3:1 ratio of right to left arrows throughout the experiment.

Each word-pair appeared on-screen for 1000ms, after which it was immediately replaced with a (right or left pointing) arrow. The arrow remained on the screen for up to 1000ms. If a response onset triggered the E-Prime voice cue before the 1000ms deadline, the arrow disappeared. To encourage rapid responses, if a response was not initiated within the 1000ms arrow presentation, a loud buzzer was played together with a red flash on the monitor. The next item pair appeared on the screen 400ms after the response initiation or buzzer warning.

The experiment started with a 10 item practise session containing 2 mock competitor items, followed by a break to allow for questions and feedback. The experiment was then presented in two blocks with a one minute break between runs. Participants listened to brown noise throughout the task at a volume as loud as comfortable. The total duration of the experiment was approximately 15 minutes.

Data Treatment

Following the experiment, we analysed participants' responses in two ways. The first analysis was based on the perceptual transcription of responses from all 47 speakers included in the experiment. Each target response was assigned to a category, discussed in more detail in *Auditory Transcription Method*. The second analysis was based on the articulation data recorded from seven speakers and the method is discussed in more detail in *Articulation Method*. For both analyses, each nonword within every competitor pair was treated independently: For example, within the competitor pair *gope doof*, *gope* and *doof* were coded independently. This is different from many studies using phonological exchange elicitation paradigms, which report outcomes analysed by pairs of targets (e.g., Baars et al., 1975; Hartsuiker et al., 2005). In part, we adopted this approach because the WOC task encourages the confusion of left and right hand targets, making it difficult to distinguish between anticipation (*gope doof* → "doof dope") and perseveration (*gope doof* → "goof gope") errors. It also allowed us to isolate onsets for the articulatory analyses. Also, for both analyses we focus exclusively on target onsets. We use the term *competitor substitution* to refer to slips of the tongue that involve the substitution of the competitor's onset for the target onset.

Auditory Transcription Method Each competitor item (e.g., *gope* in the competitor pair *gope doof* was coded as Correct (*gope* → "gope"), a → "dope"), or as

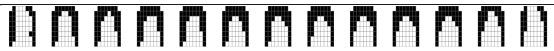
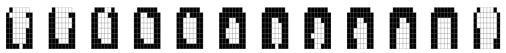
Other. The "Other" category included all responses that did not fit into the "competitor substitution" or "correct" categories. This included cases in which the responses were pronounced in the presented order as opposed to the cued reverse order, which could not be distinguished from a full word substitution or combination of competitor and rime substitutions. Where participants produced more than two (non-)words in response to a competitor pair, the first two items constituting a response resembling the intended competitor pair was selected as the response.

All of the responses from five speakers (10.6% of the total data) were cross-transcribed by another transcriber to test for inter-rater reliability. There was 96% agreement on categorising the onset of each competitor item. The transcribers came to a consensus on the transcriptions of the other 4% of items.

Articulatory Transcription Method Articulatory records were created by identifying the cue to speak (or acoustic offset of the previous word) and the onset of the spoken vowel for each item. The EPG contact data was then extracted between these two time points. Each articulatory record was visually trimmed to include the first palate before full closure through to the first palate after the final full closure release before vowel onset. This trimming method was conducted blind to the intended articulatory pattern to ensure that a velar closure was as equally likely to be identified as an alveolar closure. Full closure was defined as any lateral continuous path across the EPG palate. Responses in which velar closure did not yield a continuous path across the palate were reexamined. If closure was achieved across all but the middle two posterior contacts at any point during the articulation, this was treated as full closure and the record was trimmed accordingly. Data points that did not have this degree of velar closure, or did not achieve full closure at other positions, were excluded from the articulation analysis because the start and end points could not be reliably identified. This included 10.8% of the competitor pair items. No more than 14.5% of the data was excluded for any given participant and the excluded items were evenly distributed across cells in the design matrix.

The articulatory analysis did not take into account contact that may have occurred during rime production for two reasons: first, all competitor pairs were designed to minimise tongue-to-palate contact after onset production in an effort to reduce contact between nonwords within each competitor pair; second, while some contact may have been recorded as the result of higher vowels or liquids, the identification of onsets based on the acoustic offset of the previous word minimises the chance of

Table 3.1: Sample EPG recordings of intended alveolar articulations and their categorical codes from Experiment 1: (a) and (b) contain closure in the alveolar region of the palate, (c) contains both alveolar and velar closure, and (d) contains clear velar closure

	Token	Categorical Code
a		correct
b		correct
c		other
d		comp. substitution

mistaking rime closure of the first nonword with onset closure of the second nonword of each competitor pair.

These records corresponded to target onsets, and were assigned to the same categories used for the auditory transcription. In this case, the assignments were based on the presence or absence of the relevant articulatory closure patterns for each onset. Tables 3.1 and 3.2 show examples of articulatory contact patterns for each category. In rows (a) and (b), we see a correct onset in response to a target alveolar (e.g. *doof*), featuring alveolar closure (in the anterior [top] region of the palate). Row (c) shows a double articulation coded as "other". In each case the articulation contains both alveolar and velar closure. Lastly, row (d) shows a competitor substitution: closure occurred in the competitor region of the palate.

Table 3.2: Sample of EPG recordings of intended velar articulations and their categorical codes from Experiment 1: (a) and (b) contain closure in the velar region of the palate, (c) contains both velar and alveolar closure, and (d) contains clear alveolar closure

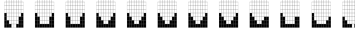
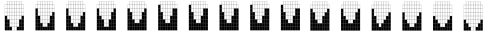
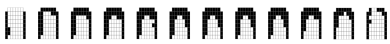
	Token	Categorical Code
a		correct
b		correct
c		other
d		comp. substitution

Table 3.3: Total numbers of competitor substitutions by each experimental condition for the auditory and articulatory transcription analyses of Experiment 1.

		Context	
Analysis		Nonlexical	Mixed
Auditory Transcription (n=47)			
	Nonword competitor	6	8
	Real word competitor	16	21
Articulatory Transcription (n=7)			
	Nonword competitor	2	2
	Real word competitor	8	7

3.3.2 Auditory Transcription Results

A total of 4512 targets were presented to the 47 speakers. Speakers did not respond to 73 target items. Of the 4439 responses recorded, 3003 were coded as correct (67.7%), 51 as competitor substitutions (1.1%), and 1385 as other errors (31.2%). Table 3.3 shows the numbers of competitor substitutions by condition. ANOVA statistics are reported including: by-participant (F1), by-item (F2), and minF′ statistics with Competitor (real word, nonword) and Context (mixed, nonlexical) as within-participant and within-item factors and Group (behavioural, EPG) as a between-participant but within-item factor.

The analysis revealed a significant effect of Competitor, showing that competitor substitutions occurred more frequently when the target had a real word competitor (1.6%) compared to a non-word competitor (0.6%; 95%CI ± 0.01). We did not observe significant effects of Context or Group, or any significant interactions. See Table 3.4 for a summary of ANOVA statistics.

Table 3.4: ANOVA statistics for auditory transcription analysis from Experiment 1

Source of variance	By participants			By items		minF′	
	df	F1	MSe	df	F2	df	minF′
Competitor	1,45	8.53**	.0006	1,94	6.26*	1,133	3.61^{m}
Context	1,45	<1	.0008	1,94	<1	1,122	<1
Group	1,45	2.00	.0009	1,94	1.91	1,124	<1
Competitor × Context	1,45	<1	.0006	1,94	<1	1,122	<1
Competitor × Group	1,45	2.29	.0006	1,94	1.67	1,133	<1
Context × Group	1,45	<1	.0008	1,94	2.63	1,80	<1
3 Way Interaction	1,45	1.03	.0006	1,94	1.81	1,96	<1

[m]marginal effect, p = .06; * p <.05; ** p<.01

Table 3.5: ANOVA statistics for articulatory transcription analysis from Experiment 1

Source of variance	By participants			By items		MinF'	
	df	F1 value	MSe	df	F2 value	df	minF′ value
Competitor	1,6	8.68*	.0001	1,47	4.13*	1,37	2.80
Context	1,6	<1	.002	1,47	3.87^{m}	1,9	<1
Competitor × Context	1,6	<1	.002	1,47	<1	1,21	<1

*p <.05

3.3.3 Auditory Transcription Discussion

When the speech produced in the WOC paradigm is transcribed and analysed in a similar way to previous SLIP studies, there is evidence for a lexical bias, with more competitor substitutions for real word competitors than for nonword competitors. This establishes that the effect of the WOC paradigm on speech production is similar to that of the SLIP task and adds to the growing body of experimental support for a lexical bias in the substitution of word onsets (e.g., Baars et al., 1975; Hartsuiker et al., 2005; Humphreys, 2002). In the WOC experiment, context has no effect, consistent with Humphreys' (2002) SLIP findings, which have been taken as support for a feedback account of the lexical bias effect (Dell, 1986). According to such an account, lexical bias results from the reinforcement of lexical representations from a back flow of information from phonological representations. This reinforcement is an automatic process and occurs independent of context.

However, this analysis was restricted to items that could auditorily be identified as competitor substitutions. The following analysis is based on the articulatory transcription of elicited responses.

3.3.4 Articulatory Transcription Results

Out of 632 recorded responses, 502 responses were coded as correct articulations (79.4%), 19 as competitor substitutions (3.0%), and 111 as other articulations (17.6%). Table 3.3 shows the numbers of competitor substitutions by condition. ANOVA statistics were calculated for by-participant (F1), by-item (F2), and minF′ analyses of the competitor substitution error distribution. Refer to Table 3.5 for a summary of ANOVA statistics. The analysis revealed a significant effect of competitor, with significantly more real word competitor substitutions (5.0%) relative to nonword competitor substitutions (1.3%; 95%CI ±0.01).

3.3.5 Experiment 1 Discussion

Both the auditory and articulatory transcription analyses of competitor substitution errors provide further experimental evidence of the lexical bias effect, with more competitor substitution errors for real word competitors compared to nonword competitors. Moreover, the competitor substitution error pattern in the articulatory transcription mirrors the pattern in the acoustic transcription. This suggests that competitor substitution error patterns can successfully be captured using articulatory transcription. More importantly, it adds to the growing body of experimental support for the lexical bias effect (Baars et al., 1975; Hartsuiker et al., 2005; Humphreys, 2002). The specific pattern of the lexical bias effect for competitor substitution errors, with no context effect, is consistent with an interactive feedback account of the lexical bias effect (Dell, 1986; Humphreys, 2002). This account posits that the lexical bias effect results from the reinforcement of lexical representations from a backward flow of information from phonological representations. This reinforcement is an automatic process, and occurs independent of context.

However, the analyses of competitor substitution were limited to only a small percentage of the data: 1.1% in the auditory transcription and 3.0% in the articulatory transcription. Error rates in other comparable investigations have also been low: In total, 8.2% of responses were categorised as full exchanges by Baars et al. (1975, Experiment 2); Humphreys (2002, Experiment 4) reports 2.7%; the figure for Hartsuiker et al. (2005) is 2.3%. The categorisation of responses results in a high proportion (over 98%) of mispronunciations being excluded from further analysis. With such low numbers of occurrences and empty cells, the data may be susceptible to noise (see also Nooteboom & Quené, 2007).

A comparison of competitor substitutions from the auditory and articulation transcriptions for the seven speakers who were recorded with EPG revealed 97% agreement between transcriptions. The discrepancy arose from one item which was identified auditorily as a competitor substitution, despite the articulatory record not showing evidence of a competitor substitution. This suggests that auditory transcription can successfully capture competitor substitutions. On the other hand, "other" error rates greatly differed across transcription types: 17.6% in auditory transcription compared to 31.2% in the articulatory transcription.

To investigate the nature of 'other' responses a *post hoc* inspection of the articulatory records was performed. In particular, EPG records containing 'double articulations' were identified. These included closure at both alveolar (defined as a continuous

seal of contact in the anterior half of the palate) and velar (posterior palate) palatal regions. See rows (c) in Tables 3.1 and 3.2 for example double articulations. This inspection revealed that 79 (13.2%) of the responses recorded with EPG contained a double articulation. However, in a post-hoc auditory transcription of the speakers recorded with EPG only 14 (2.2%) of the responses could be identified as a double articulation (defined as having two detectable closures, such as /tkim/ or /t.kim/). If double articulations represent cascading errors, then this discrepancy between 14 and 79 identifiable double articulations suggests that auditory transcription misses over 80% of the errors that occur.

An inspection of the distribution of double articulations revealed that these errors do not appear to be affected by the lexical status of competitors: out of 79 double articulations, 38 errors occurred for nonword competitors and 41 errors for real word competitors. However, context does appear to have an influence: 48 occurred in the nonlexical context and 31 in the mixed context. One possible account for the context effect is that double articulations provide evidence for some form of repair. This is consistent with the view that speech is internally monitored and can be rapidly revised when errors are detected (Baars et al., 1975; Levelt, 1989). If this were the case, then the most frequent pattern of double articulation observed should reflect a successful repair, ending with a closure pattern matching the intended onset. Further, the time between the first and second closures should feasibly represent the stopping of one motor plan and its replacement with another.

A detailed examination of the recorded double onsets revealed that 92.4% were of the form *competitor closure* → *intended closure*, compatible with a repair, as opposed to *intended closure* → *competitor closure*. To investigate the time-course, we calculated the time between the onset of the first closure and the onset of the second closure. It has been suggested that 180–200ms is required to initiate a repair (Laver, 1980; Levelt, 1983; Logan & Cowan, 1984; Hartsuiker & Kolk, 2001). Our analysis revealed a mean inter-contact duration of 254ms (SD 104ms), with 82.2% of responses having inter-contact durations of more than 180ms. Taking the patterns and durations of the double onsets into account, it is plausible to suggest that at least a subset of instances reflect rapid repairs.

Pouplier (2007) performed a similar analysis of double articulations recorded with EMMA during a SLIP task. In her analysis she subtracted the time of the competitor peak height from the time of the target peak height. This yielded a inter-target duration metric that also reflected the direction of closure: a positive value reflected a *competitor closure* → *intended closure* pattern and a negative value a *intended*

closure → *competitor closure* pattern. This analysis revealed a range of inter-target durations from -180 to 274ms with a large majority having a duration of 100ms or longer. Therefore a large proportion of these inter-contact durations appear to be too fast for a rapid repair, though some of the positive values are consistent with repairs of a minimum duration of 180ms. Also, Pouplier (2007) observed more positive compared to negative values which is consistent with the observation of more EPG *competitor closure* → *intended closure* articulations.

Taken together, the distributions of competitor substitutions in both the auditory and articulation analysis provide evidence for feedback in production. However, the high rate of "other" errors, and in particular those that constitute double articulations, may have additional implications. First, some double articulations may provide evidence for rapid repair and therefore implicate a role for monitoring in speech production. Second, the occurrence of double articulations raises a concern over the accuracy of auditory transcription. Primarily, the concern raised is with regard to what constitutes an "error".

One could argue that the occurrence of double articulations reflects an error in speech production regardless of whether they can be auditorily perceived. In other words, the general observation of some additional motor movement not normally associated with a motor goal suggests that some factor has influenced planning. One could also argue that double articulations are not "errors" if they can not be perceived auditorily. This is because the mapping of articulatory movements to acoustic signal can be ambiguous (Atal, Chang, Mathews, & Tukey, 1978). We align with the former argument and view double articulations as evidence for a planning process. However, in order to establish that double articulations can also be interpreted as errors for those that subscribe to the "must be perceivable" camp we conducted a perceptual study to determine if listeners are sensitive to double articulations.

3.4 Experiment 2: Perception of Non-Canonical Errors

This experiment was designed to investigate if double articulations can be perceived auditorily. A classic finding in the categorical perception literature is that when listeners are presented with a voice-onset time continuum (e.g. /d/↔/t/) they have more difficulty identifying phonemes in the middle of two boundaries compared to phonemes at boundary end points (Liberman et al., 1967; Liberman, 1997). It is also the case that reaction times are consistently slower when participants have to

identify mid-boundary phonemes (Pisoni & Tash, 1974). Therefore, the form of this experiment is a basic categorisation task. Participants were presented with recordings of double articulations from Experiment 1 and recordings of matched normal articulations and asked to categorise the responses as one onset or another. If participants have more difficulty categorising double articulations compared to normal articulations the evidence would suggest that listeners are indeed perceptually sensitive to double articulations.

3.4.1 Methods

Participants

Twenty native English speakers (8 males, 12 females) from the University of Edinburgh participated in the experiment for pay. None of the participants reported a language or hearing impairment. All participants were treated in accordance with the University of Edinburgh and Queen Margaret University ethical guidelines.

Materials

The experimental items consisted of 80 recordings of nonword productions from speakers recorded with EPG. Half of the items were identified as double articulations from the EPG recording of Experiment 1. The other half were identified as normal articulations from the EPG recordings for a control task (details of the latter reported in Section 5.3.1; these recordings were of the same nonword stimulus items but recorded in a non-error invoking condition). Double articulations and normal articulations were matched across item (e.g. the same nonword) and speaker. Recordings were selected from five speakers who had the highest quality recording, with a minimum of 10 items per speaker. All of the double onset items were of the form competitor closure → intended closure. All of the normal onsets only contained the intended closure. Half of the items were intended alveolar onsets (/d/, /t/) and half were intended velar onsets (/g/, /k/). Within the set of alveolar onsets half were voiced (/d/) and half were voiceless (/t/). However, due to limited materials velar onset items were not balanced for voicing [7 voiced (/g/), 13 voiceless (/k/)].

Each item was extracted using Praat (Boersma & Weenink, 2006) from 100ms prior to the acoustic release of the onset consonant to the end of the coda. We then spliced 300ms of silence to the beginning of each experimental item. Materials

were controlled for release duration to reduce the chances that any potential perceptual differences between double and normals onsets could be attributed to one being longer than another. Release duration was measured in Praat in milliseconds from the release onset to the vowel onset. A two way ANOVA with Articulation Type (normal onset,double articulation) and Place of Articulation (alveolar,velar) as within-item factors revealed no significant differences in release duration across conditions: Articulation Type [$F(1,19)<1$], Place of Articulation [$F(1,19)<1$], and Articulation Type $\times$ Place of Articulation interaction [$F(1,19)=1.14$,ns].

Presentation Lists

Each of the 80 stimulus items were repeated across four presentation lists. Presentation lists were equally split into pairs so that each of the eight lists of 40 stimulus items contained an equal number of each stimulus type (ten velar normal onsets, ten velar double articulations, ten alveolar normal onsets and ten alveolar double articulations). The distribution of stimulus items in each list of the pairs was controlled so that one of the lists contained the normal articulation of a stimulus item and the other list contained the matched double articulation of the same stimulus item.

Apparatus

The experiment took place in a sound attenuated room at the University of Edinburgh. A computer was used to control stimulus presentation, using E-Prime (Schneider et al., 2002): participants saw onset choices on a 17 inch M2-CRT monitor and auditory materials were played using Technics RP-F200 stereo headphones.

Procedure

Once participants were seated, they were instructed to choose, as quickly as possible, the onset of each experimental item [t or d] or [k or g]. Choices were counterbalanced so that half of the participants responded with the left button for [t or d] and the right button for [k or g] and half in the reverse order. Choices remained on the screen throughout the duration of the experiment. Participants did not have a response deadline and the next item was played 1000ms after each response. The experiment started with a practise session, consisting of ten items (half alveolar, half velar and two items from each speaker). After the practise session participants were given time to ask questions before the experiment started. In the experiment, each pair of presentation lists was presented twice in random order with a short

Table 3.6: Mean accuracy and response times (with standard errors) for perceptual judgements of normal onsets and double onsets from Experiment 2.

		Articulation Type	
Measure	Place of Articulation	Normal Onset	Double Onset
Accuracy (% correct):			
	Alveolar	95.5 (1.5)	93.7 (2.0)
	Velar	95.4 (1.6)	91.9 (2.7)
Reaction times (ms):			
	Alveolar	969 (20)	1008 (22)
	Velar	992 (20)	1034 (26)

break in between list pairs. Participant were presented with a total of 640 items (40 items per list × 8 lists × 2 list presentations). The entire experiment lasted approximately 25 minutes.

3.4.2 Experiment 2 Results

ANOVA statistics were calculated for by-participant (F1), by-item (F2), and minF′ analyses with Articulation Type (normal onset, double onset) and Place of Articulation (alveolar, velar) as within-participant variables. See Table 3.6 for means and Table 3.7 for a summary of ANOVA statistics. An analysis of categorisation accuracy revealed an Articulation Type main effect that was significant by-participants, but not by-items. Participants were less accurate at identifying double onsets (92.8%) relative to normal onsets (95.5%; 95%CI ± 1.58). Neither the Place of Articulation main effect or the Articulation Type × Place of Articulation interaction for accuracy was significant.

Given the lack of a by-items effect for articulation type, a response time analysis was also performed on correctly categorised onsets. This analysis revealed a main effect of Articulation Type, significant both by-participants and by-items. The mean response time for double onsets (mean=1021ms) was significantly slower than the mean response time for normal onsets (mean=980ms; 95%CI ± 10.40). There was also a significant by-participants main effect for Place of Articulation. The mean response time for velar onsets (mean=1013) was significantly longer than for alveolar onsets (mean=989; 95%CI ± 12.56). The Articulation Type × Place of Articulation interaction was not significant.

3.4.3 Experiment 2 Discussion

Experiment 2 demonstrates that listeners were less accurate and slower at identifying the correct onset consonant for double articulations compared to normal articulations. This suggests listeners are sensitive to double onsets, though, as demonstrated in the Experiment 1 transcriptions, they may not be able to consciously perceive them. This findings has implications for both speech error transcription methods and defining what constitutes an error. The implications for transcription are discussed in more detail in the General Discussion (Section 3.5).

Some speech error investigations have reported an asymmetry in error production in which less frequent phonemes are more likely to be substituted with more frequent phonemes or vice versa (Stemberger, 1991a, 1991b). However, it is possible that this asymmetry reflects perceptual biases in transcription (Pouplier & Goldstein, 2005). For example, the transcription of a double articulation may be affected by the frequency of the phoneme being transcribed. In an experiment designed to investigate perceptual biases Pouplier and Goldstein (2005) observed that listeners had more difficulty perceiving /t/ compared to /k/ double articulation errors. Their finding suggests that asymmetries in speech error production may be accounted for by perceptual biases in transcription rather than by some production-based mechanism. However, Pouplier and Goldstein (2005) did not observe a bias in the perception of /s/ and /ʃ/ errors. In the present experiment, the only place of articulation effect was observed for reaction time, which was only significant by-participants, and was in the opposite direction of Pouplier and Goldstein's (2005)

Table 3.7: ANOVA statistics for the accuracy and response time analyses of perceptual judgements from Experiment 2

Source of variance	By participants			By items		minF′	
	df	F1	MSe	df	F2	df	minF′
Accuracy Analysis:							
Articulation Type	1,19	6.14*	22.71	1,76	2.66	1,89	1.86
Place of Articulation	1,19	1.29	13.22	1,76	<1	1,71	<1
Type × Place	1,19	1.42	9.66	1,76	<1	1,74	<1
Reaction Time Analysis:							
Articulation Type	1,19	32.81****	988.30	1,76	7.23**	1,95	5.92*
Place of Articulation	1,19	8.52**	1440.22	1,76	<1	1,90	<1
Type × Place	1,19	<1	398.25	1,76	<1	1,74	<1

**** p <.0001; ** p <.01; * p <.05

results. This finding therefore does not appear to support a perceptual account of speech error asymmetries.

3.5 General Discussion

Two experiments were reported in this chapter. Experiment 1 provides additional experimental support for the lexical bias effect. Both an auditory and an articulatory transcription analysis demonstrated that competitor substitutions are more likely to occur if the outcomes yield real words than if they yield nonwords. These results are discussed in more detail below. However, the observation of double articulations using EPG suggests that auditory transcription may be problematic for speech error research. Experiment 2 highlights this problem by demonstrating that listeners are sensitive to double articulations despite not being able to transcribe them. The implications of the transcription of speech errors are also discussed.

3.5.1 A Feedback Account

Previous SLIP task investigations of the lexical bias effect have given rise to inconsistent results (Baars et al., 1975; Hartsuiker et al., 2005; Humphreys, 2002). Both Baars et al. (1975) and Hartsuiker et al. (2005) observed patterns of lexical bias that differed across contexts, which they argued could only be accounted for by a production model incorporating some form of self-monitoring. Specifically, Baars et al. (1975) observed less exchanges that resulted in nonwords in the mixed context which, they argued, supported the editing of the production of nonwords. Hartsuiker et al. (2005) on the other hand observed less exchanges for real word outcomes in the nonlexical context, which they claimed supported an adaptive monitor that edits out the production of real words since they are inappropriate in a nonlexical context. In contrast to both Baars et al. and Hartsuiker et al., Humphreys (2002) observed a context-independent lexical bias effect, which she argued supports a feedback account of the lexical bias effect because feedback is an automatic process that occurs independently of context.

The results from Experiment 1 provide additional evidence for a context-independent lexical bias effect for competitor substitutions. Both the auditory and articulatory transcription analyses confirm this pattern. This finding replicates Humphreys's (2002) Experiment 4 and provides evidence for feedback in speech production, consistent with models such as Dell's (1986). In particular, phonological representations of the competitor onsets are reinforced by the flow of information from

phonological representations to lexical representations. Since it is a consequence of the architecture of the model, the flow of information is automatic and therefore context-independent. However, this flow of information is not possible for nonword competitors since information can not feedback to a nonword representation because, by definition, they do not exist.

However, the support for a feedback account from Experiment 1 must be interpreted with caution. The analysis was restricted to competitor substitutions which only constitute 1.1%-3.0% of the total responses. Similarly, analyses from previous comparable investigations are restricted to a small subset of data. While a comparison of the auditory and articulatory transcription analysis revealed 97% agreement across transcription methods for competitor substitutions, a discrepancy in double articulations for the 'other' errors suggests that transcription may be problematic.

As noted in Section 3.3.5, one interpretation of the high rate of double articulations (13.2% of responses) is that they reflect self-monitoring. Double articulations may be some form of a rapid repair from detection of unintended articulation of one phoneme to articulation of the intended phoneme. A repair-based double articulation may be accounted for by a staged model of production in which one phoneme is selected and articulated followed by the selection and articulation of another phoneme. However, an inspection of the timing of double articulations revealed that some are too fast to be accounted for by a repair. A different interpretation is that they result from cascading activation in a model of production. In particular overlapping (or nearly overlapping) articulation of both a competitor and an intended phoneme can be interpreted as the outcome of partial activation of both phonemes.

Importantly, the WOC task appears to yield similar results to the traditional SLIP task for eliciting speech errors. The primary benefit of the WOC task over the SLIP task is that investigations assuming cascading activation are not confounded by the possibility that repeated priming may influence articulation. For example, double articulation of a competitor and intended phoneme are likely to reflect activation associated with the target pair since those phonemes will not have been produced for at least three preceding items. This is not the case for SLIP task investigations in which primed stimulus items may retain some activation during presentation of the target pairs.

A disadvantage of the WOC task is that transcription can be more difficult than for the SLIP task, due to the directional cuing that elicits errors. The transcription

analyses were conservatively restricted to transcribing responses as 'correct', 'competitor substitution' and 'other'. Analysis was restricted to these categories because in a response such as *gope doof* → "gope doof" (following a left arrow, which should have been spoken as "doof gope") it was impossible to determine whether the arrow had been ignored or a full lexical exchange had occurred. Similarly, a response such as *gope doof* → "goof dope" could be interpreted as a rime exchange or a word reversal with an onset exchange.

3.5.2 Transcription of Speech Errors

In Experiment 1 of this chapter it was demonstrated that a high proportion of the double articulations observed in the articulatory transcription were not detected in the auditory transcription. However, Experiment 2 established that listeners are sensitive to double articulations because accuracy and especially reaction times for identifying the onset were affected. A likely source of these conflicting findings is that transcriptions systems are segmental in nature (see, for example, Frisch & Wright, 2002; Mowrey & MacKay, 1990; Pouplier & Goldstein, 2005). While double articulations may contain some acoustic distortion detectable in categorical perception task it is difficult to transcribe a response that sounds like a /t/ but differs slightly from a typical /t/. Moreover, Frisch and Wright (2002) highlighted that most speech error investigations typically rely on a forced-choice (e.g., *Is this an error or not?*) transcription method.

The use of acoustic and articulatory instruments to investigate speech errors has largely been motivated by the inability to transcribe non-canonical errors (Frisch & Wright, 2002; Goldstein et al., 2007; Pouplier, 2003, 2007). Indeed, Experiment 1 was, in part, designed to investigate the discrepancy between errors observed with auditory and articulatory transcription. However, while instrumental investigations may be less susceptible to the limitations of not hearing 'other' errors, they are still faced with the difficulty associated with assigning responses to categories. For example, if a specific category for 'double articulations' was created to account for the articulatory data in Experiment 1, there would still be difficulty in deciding which responses belonged to this category. The double articulation in Table 3.1(c) contains velar and alveolar closures that are clearly occurring simultaneously, while the articulation in Table 3.2(c) also contains velar and alveolar closures but these may be sequential rather than simultaneous. Likewise, some double articulations are of the form *competitor closure* → *intended closure*, while others are *intended closure* → *competitor closure*. The difficulty of creating a response category is not

only limited to double articulations. It is not even clear what constitutes a 'correct' response: The alveolar articulation in Table 3.1(b) is clearly more midpalatal than the articulation in Table 3.1(a), but it is also clearly not a velar articulation.

The difficulty with categorising responses is not limited to capturing the detail observed in articulatory investigations. Even across auditory transcription investigations the categorisation of errors varies. For example, Hartsuiker et al. (2005) considered abortions such as *pote vark* → "vote p..." as full exchanges because the onset of the second (non)word would have been an exchange if the response had been completed. Humphreys (2002) on the other hand used an additional category for aborted errors. Also Hartsuiker et al. (2005) and Humphreys (2002) only categorised responses that included a correct rime as substitutions (or exchanges) while Baars et al. (1975) categorised responses as substitutions provided they contained the relevant onset and medial vowel. Therefore a comparison of different results from these three experiments may not be informative. Other investigations have included analyses of additional categories and have based theoretical claims on the pattern of these errors (Nooteboom, 2005a; Nooteboom & Quené, in press). For example, Nooteboom and Quené (in press) analysed "interrupted spoonerisms" (e.g., *barn door* →"d...barn door") and argued that only a feedback with self-monitoring model such as Hartsuiker et al.'s (2005) model could account for their distribution.

Taken together, the primary limitation of both auditory and articulatory transcription is the practise of assigning responses to categories. Categorising responses makes specific assumptions about what components of output should be observed in a response, such as what the acoustic signal should sound like or what articulatory movements should be observed. A cascading model of production does not require the production of categorical output: an articulation may include partial activations of both alveolar and velar contact. In other words, a cascading model specifically predicts that spoken responses may be noisy. Therefore, any categorisation-based method for investigating speech seems likely to fail.

In order to investigate articulation in a cascading model of production a non-categorical technique for evaluating articulation is required. In Chapter 4 a new technique, the Delta method, for analysing articulation data is presented. The Delta method quantifies the articulatory signal and rates how (dis-)similar an articulation is relative to another articulation. By quantifying articulation in this way it is possible to abandon categorising responses as 'errors' or 'correct' and instead

to measure general variability in articulation. In Chapter 5 a reanalysis of the articulation data recorded for Experiment 1 is presented and the influences of lexical competitors and context on articulation are discussed.

3.6 Chapter Summary

The WOC task is a speech error elicitation technique that yields results similar to the SLIP task. Experiment 1 provides support for a feedback account of the lexical bias. However, the results must be interpreted cautiously due to the limitations associated with transcription. It is therefore necessary to develop a means of analysing articulatory data without making assumptions about what constitutes a category of response. In Chapter 4, a new technique for analysing articulation data is presented that allows categorisation of responses to be abandoned. In Chapter 5 the data reported in this chapter is reanalysed without relying on transcription.

Chapter 4

Measuring Articulation

4.1 Chapter Overview

Articulatory investigations typically rely on the categorisation of responses and require spatial and temporal assumptions about articulation. In this chapter we introduce electropalatography (EPG), a technique that provides a direct measure of tongue-to-palate contact across time. We also present a review of several analysis methods used for EPG and argue that they all have some limitations. In the remainder of the chapter we present the Delta method: an analysis method we developed for quantifying variability in articulation that does not suffer from the limitations of previous methods. We conclude that the Delta method is a useful approach for our research which investigates the interface between speech planning and articulation.

4.2 Introduction

Articulatory and acoustic investigations have challenged the traditional view of speech errors by demonstrating speech errors do not necessarily result in well-formed outcomes (Frisch, 2007; Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein et al., 2007; Laver, 1980; Mowrey & MacKay, 1990; Pouplier, 2003, 2007; Stearns, 2006, also see Section 2.3.2 for a detailed discussion). However, previous articulatory and acoustic investigations of speech errors have relied on categorising responses. This reliance on categorisation is problematic because it assumes that speech production is a staged process in which whole representations are selected for articulation. In order to investigate speech production in a cascading model of production, an articulatory analysis method which does not rely on categorisation is required.

In this chapter we present a method for quantifying patterns of articulation recorded with electropalatography (EPG). First, we present a review of EPG and how the articulatory imaging technique can be used to measure patterns of tongue-to-palate contact over time. We then present a review of several methods which have been used to analyse EPG data and argue that all of these methods require some form of categorisation and are therefore inadequate. In the remainder of the chapter we present a detailed description and demonstration of the Delta method. The Delta method is an articulatory analysis technique that does not require categorisation.

4.3 Electropalatography (EPG)

EPG is an articulatory imaging technique that measures tongue-to-palate contact over time. Speakers are fitted with a custom manufactured artificial palate formed to fit their hard palate using a dental mould. The artificial palate contains a set of embedded switches that record whether the tongue made contact to the palate. Tongue-to-palate contact data is sampled at 60-200Hz at each switch point on the artificial palate. This information is transferred from a set of wires extending from the palate, out the corners of the speaker's mouth and connecting to a multiplexer unit that processes the data. The contact data can then be analysed in its raw form or converted into a graphical representation (Hardcastle, Gibbon, & Nicolaidis, 1991; Byrd, Flemming, Mueller, & Tan, 1995). Refer to Figure 4.1 for a graphical representation of an EPG record based on the WinEPG system (Wrench, 2003, see also Fletcher, 1989 and Michi, Suzuki, Yamashita & Imai, 1986 for specifications about other palatography systems). The top of the image corresponds with the alveolar (anterior) portion of the palate and the bottom of the image corresponds with the velar (posterior) region of the palate. The contact points are arranged in seven rows of eight contacts and one row of six contacts. The articulatory record in the figure contains one epoch consisting of four frames collected over time.

EPG has predominantly been used for clinical applications, such as providing visual feedback about articulation during speech therapy (see Hardcastle & Gibbon, 1997, for a review), but we will focus on research applications of EPG. The use of EPG to measure articulation during experiments allows the direct observation of how articulation unfolds over time. Unlike other articulatory imaging techniques such as electromagnetic midsagittal articulometry (EMMA) or ultrasound of the tongue, EPG provides detailed information about articulatory movement outside of a single sagittal or coronal plane of the mouth. Therefore a primary benefit is

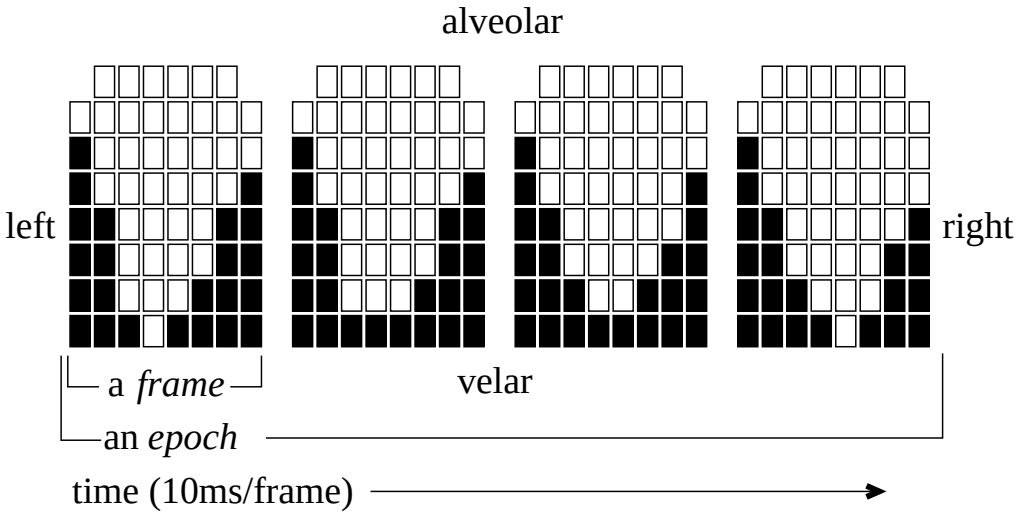


Figure 4.1: A graphical representation of an EPG record based on the WinEPG System: the top of the image represents the alveolar region of the palate, the bottom of the image represents the velar region of the palate, and each small square represents a contact point on the palate arranged in seven rows of eight contacts and one row of seven contacts. A black square represents contact and a white square represents no contact. An *epoch* of data contains several *frames* collected over time.

that articulatory patterns can be observed across the lateral (dorsal) surface of the palate.

In addition to the recording of articulatory data, the acoustic signal can be recorded simultaneously while individuals speak with an artificial palate. Despite the potential risk of measuring acoustic patterns during articulation with an oral implement, it has been demonstrated that speech intelligibility does not differ between speaking with or without an artificial palate (Searl, Evitts, & Davis, 2006; Flege, 1986). This is especially true when participants wear the palate for an initial adaptation phase of approximately 30 minutes (Searl et al., 2006) and when speakers are very experienced in wearing a palate (McLeod, 2006).

Due to technical reasons of fitting an artificial palate the distribution of switches does not extend to the velum and is restricted to the hard palate. A potential consequence of this restriction is it may be difficult to record velar closures because closure may occur posterior to the palate. For example, it has been reported that an average of 19% of velar articulations do not yield observable full closure on the palate (Liker & Gibbon, 2007). However, evidence for this has been inconsistent. In Liker and Gibbon's (2007) evaluation of velar closure characteristics strong individual differences were observed such that incomplete closure across seven speakers ranged from 4–41%. Also closure patterns were affected by the preceding and following

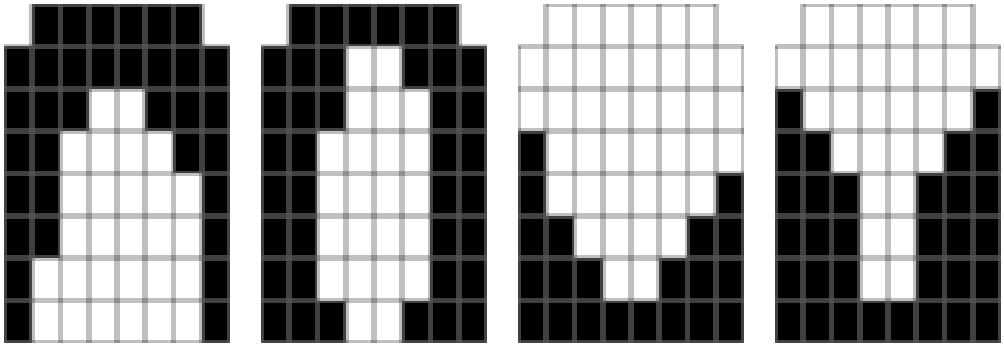


Figure 4.2: Examples of different closure patterns observed with EPG: (a) typical alveolar closure, (b) less typical alveolar closure, (c) typical velar closure, and (d) less typical velar closure.

vowel context of the spoken items. In another investigation, Byrd (1996) observed full velar closure for every /g/ recorded in 50 repetitions of *say bag gab again*. Given these inconsistent findings the experimenter should be aware of potential individual differences during the analysis of EPG closure patterns.

Once the experimenter has collected EPG data, the articulatory patterns can be evaluated using a variety of methods. The most basic method is to qualitatively identify exemplars of different types of articulatory patterns (Barry, 1985; Marchal, 1988; Nolan, 1992). Consider, for example, the different patterns of articulation in Figure 4.2. The palates in (a) and (c) can be interpreted as examples of typical alveolar and velar closure patterns observed in EPG recordings. However, the articulations in (b) and (d) look similar to the typical examples but differ in some respects: (b) has closure in the posterior region of the palate in a way that (a) does not; (d) has a greater magnitude of contact which is more anterior than the contact in (c). Similarly, in Chapter 3 we reported patterns of closure including both alveolar and velar contact (see Tables 3.1 and 3.2 for examples). Simply categorising those response as 'double articulations' does not capture different characteristics of the articulations. For example, there were a range of inter-contact durations and some closures were velar followed by alveolar and others vice versa. Lastly, reports on pathological speech have also highlighted difficulty in categorising mid-palatal closures as either velar or alveolar (Friel, 1998; Hewlett, Gibbon, & Cohen-McKenzie, 1998). Taken together, the categorisation of EPG records is a limited approach because any defined category may be based on an arbitrary boundary and it is not clear how such a boundary should be defined.

An ideal EPG analysis method would not require responses to be categorised. Also, an ideal method would satisfy three additional constraints. First, the analysis of EPG should be quantitative to allow statistical hypothesis testing. Simply identifying exemplar articulations may provide an idea about the types of articulations observed in a specific experimental condition, but it is important to be able to make generalisations about how articulation is affected by one experimental condition compared to another. Second, any EPG analysis method should take into account individual differences in articulation. Since articulatory patterns vary from speaker to speaker in, for example, velar closure, it is important that an analysis method account for individual differences. Third, since EPG allows the measurement of articulatory patterns over time, an ideal analysis method would account for both spatial and temporal characteristics of articulation.

Several quantitative measurements for analysing EPG records have been proposed in the literature (see Byrd et al., 1995; Hardcastle et al., 1991, for detailed reviews). The most basic quantitative measurement is a region index, which simply involves calculating the amount of tongue contact made in a particular region of the palate (Byrd et al., 1995). Regions that are often reported include velar (e.g., rows 1-2), mid-palatal (e.g., rows 3-5), alveolar (e.g., rows 6-8), and total palate (e.g., rows 1-8) contact. Once the amount of contact in a relevant region has been measured, the region index is typically reported as percent contact in that region. However, a regional index value does not necessarily provide information about where contact occurred within the region being measured. For example, a mid-palatal regional index of 33% could represent full closure across the fourth or sixth row of the palate. Moreover, it is a form of categorisation: defining *a priori* regions for analysis assumes that it is more important to group together, for example, rows 3 through 5 than rows 4 through 6. Lastly, a region index can only account for spatial and not temporal characteristics of articulation.

Alternative spatial measurements have been proposed that represent the position of contact relative to other positions on the palate (Byrd et al., 1995). For example, one commonly reported measure is the centre of gravity (COG) index. COG is a weighted mean of contact that reflects how anterior contact occurs on the palate ($\text{COG} = (row_1 \times 1) + (row_2 \times 2) + (row_3 \times 3) \ldots + (row_8 \times 8)$; where row 1 is the most posterior row of switches). A higher value represents contact closer to the alveolar ridge while a lower value represents contact closer to the velum. A similar index can be calculated to reflect laterality ($\text{Laterality} = (column_1 \times 1) + (column_2 \times 2) + (column_3 \times 3) \ldots + (column_8 \times 8)$; where column 1 is the far

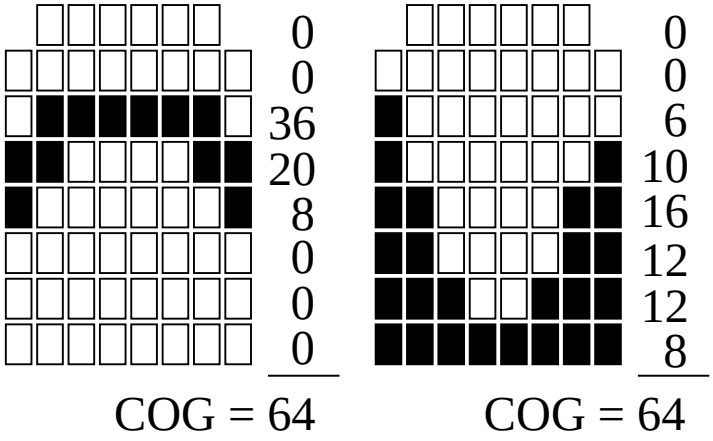


Figure 4.3: Sample centre-of-gravity (COG) values for two frames of EPG data. Despite having clearly different patterns of articulation, the COG value is equal to 64 for both frames, which illustrates a limitation for using COG for EPG analysis

left column of switches). The use of these alternative spatial indices can provide more detail about articulatory patterns than a regional index because they are not restricted to predefined regions on the palate. However, because they are based on a (weighted) mean calculation, they do not provide a full description of the specific articulatory pattern observed. Consider, for example, the EPG palates displayed in Figure 4.3. Both closure patterns yield a COG value of 64, but the palates contain clearly different articulatory closure patterns.

A further potential difficulty with spatial indices is they are often only calculated for one frame of the EPG record and therefore do not account for temporal characteristics of articulatory patterns. A strong benefit of EPG over other articulatory imaging techniques such as X-Ray or MRI is that articulatory patterns can be recorded over time. Therefore, an ideal quantitative measurement of EPG contact patterns would account for both the spatial and temporal characteristics of articulation. If spatial indices are used, it is possible to calculate an index such as COG over time. Then one could demonstrate that, for example, articulation becomes more anterior over the time course of a production.

Using spatial indices like COG also has limitations. They can only be used to account for changes within one stimulus production, which may have little value for making generalisations about articulation. Alternatively a mean COG could be calculated across several repetitions of a stimulus to demonstrate that articulation becomes more anterior over the course of production in one condition compared to another. However, creating a mean articulation can be difficult, though possible (see Section 4.4.3), since repetitions of the stimulus items often vary in length.

A serious limitation of using spatial indices is physical differences in palate size and shape across speakers means these measures do not correspond with absolute physiological landmarks (Byrd et al., 1995). As a result spatial indices cannot account for individual differences in articulatory closure patterns. An approach used by Byrd et al. to compensate for this limitation is to create speaker-specific regions of interest. To achieve this, the experimenter records the EPG signal for a series of control utterances and then performs a comparison between the region index value for experimental utterances relative to the control utterances. Using this relative comparative approach rather than an absolute measure is useful given the constraints of both not having absolute landmarks on the palate and individual differences in closure patterns. However, the method proposed by Byrd et al. (1995) is restricted to a predefined portion of the palate and is therefore limited by spatial assumptions about where tongue contact is likely to occur on the palate.

Taken together, EPG is valuable research tool that allows spatial and temporal characteristics of articulation to be measured during speech production. A variety of methods for analysing EPG data have been proposed in the literature, but they all have some limitations. In particular, none of the methods reviewed above satisfy all the ideal characteristics of an analysis method: qualitative methods require categorisation and do not allow generalisations about patterns of articulation; regional indices require spatial assumptions and can not account for temporal characteristics; spatial indices can yield the same value despite clearly different patterns of articulation; and the relative measure proposed by Byrd et al. (1995) requires spatial assumptions.

In the following section we present a new analysis algorithm, the Delta method. The Delta method is a relative measure of tongue-to-palate contact variability that satisfies all of the ideal characteristics of an analysis method. A relative measure of articulation has a clear advantage over absolute measures. Instead of describing where tongue contact occurred on the palate, a relative measure can simply quantify whether closure was different for one articulation compared to another articulation. This approach, therefore, does not require assumptions about where and when closure should have been. It also can account for individual differences in articulation by making relative comparisons within each speaker. In addition to the benefits of using a relative measure, the Delta method yields a single value (Δ) representing spatial and temporal variability of one articulation compared to another articulation. The Δ-value can then be used for standard statistical tests to make generalisations about articulation.

4.4 Delta Method for Quantifying EPG Data

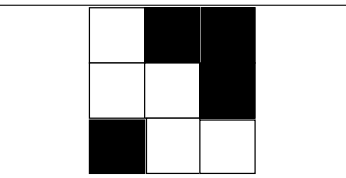
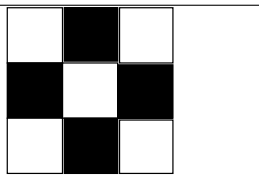
In this section, we describe a method we developed to quantify articulatory patterns recorded with EPG. This method, the *Delta method*, is based on relative rather than absolute measurements of differences in articulation. It is based on the comparison of variability between two articulatory records. Variability in the Delta method is calculated by measuring the *mean Euclidean distance* between two articulatory records which yields a single value, Δ (or "delta"). Since Δ is a difference (or distance) score, a low value can be interpreted as greater similarity and a high value can be interpreted as lesser similarity. The articulatory records used for comparisons can consist of a single frame of data, an epoch of data which consists of several frames, or an epoch of data that consists of different numbers of frames (e.g., 10 frames compared to 12 frames).

In order to show how the Delta method works, we demonstrate three types of calculations: two single frame comparisons (Section 4.4.1), two epoch comparisons of several frames (Section 4.4.2), and two epoch comparisons of several frames where the number of frames differs (Section 4.4.3). For each demonstration we first illustrate the Delta method calculation using a simplified 3×3 grid. For an example refer to Figure 4.1.a. The grid represents arbitrary data which allows us to visually demonstrate the method. Each of the nine points in the grid is a value of one, representing contact, or zero, representing no contact. After illustrating the method with the arbitrary grid data, we extend each sample comparison to include "realistic" 62 contact EPG data, which we refer to as *mock-EPG* data. The mock-EPG data was generated to provide clear examples of how Δ varies for different patterns of articulation. Each point in the mock-EPG data takes the form of real EPG data with binary information (e.g, 0=no contact, 1=contact) for each contact switch being measured.

4.4.1 Single frame comparisons of EPG data

The first demonstration of the Delta method compares two single frames of data. To calculate a Δ value for this comparison each frame of data must first be converted into a vector. Each vector contains the binary contact information for the frame being measured. A vector is generated by transforming the data in a 3×3 grid into a nine element array, beginning with the southwestern point and moving from left to right across each row of the grid.

Table 4.1: A demonstration of a Δ calculation for comparing single frames of grid data. The images in (a) are a representation of the raw data and the items (b) are the corresponding vectors. The vector is created by transforming the 3×3 grid into a nine element array, beginning with the southwestern corner and moving from left to right across each row. Δ is equal to the mean Euclidean distance between the two vectors in (b).

	Frame_A	Frame_B	Euclidean Distance
(a)			
(b)	(1 0 0 0 0 1 0 1 1)	(0 1 0 1 0 1 0 1 0)	2.00
			$\Delta = 2.00$

For single frame comparisons, Δ is equal to the Euclidean distance between the vectors. Refer to Table 4.1 for an illustration of this process: (a) provides a graphical representation of the raw data frames to be compared, (b) represents the vector that was generated, and Δ is equal to the Euclidean distance between the vector for Frame_A and Frame_B.

To demonstrate how Δ varies for EPG data we calculated a Δ value for different frames of mock-EPG data relative to a reference velar frame and a reference alveolar frame. Since Δ is a relative measure of differences in articulation a comparator is always required. Table 4.2.a and 4.2.b displays the reference frames used for these comparisons. The remainder of the table contains different frames of mock-EPG data and the Δ values for each frame relative to the references. There are a few interesting patterns to note. The velar articulation are similar to the velar reference while the less similar velar articulation has a higher Δ value. Likewise, the alveolar articulations are similar to the alveolar reference and the less similar alveolar articulation has a higher Δ value. Additionally, velar articulations have higher Δ values when compared to the alveolar reference and alveolar articulations have higher Δ values when compared to the velar reference. Lastly, the final articulation, which contains both velar and alveolar contact, has comparatively high Δ values relative to both references. This pattern demonstrates that the Delta method successfully yields lower Δ values for similar articulations and higher Δ values for dissimilar articulations.

Table 4.2: Δ values for the comparison of single frames of mock-EPG data. All Δ values are calculated relative to a velar reference frame (a) and an alveolar reference frame (b). The comparisons in the table demonstrate that the Delta method yields low values for similar articulation comparisons and high values for dissimilar articulation comparisons.

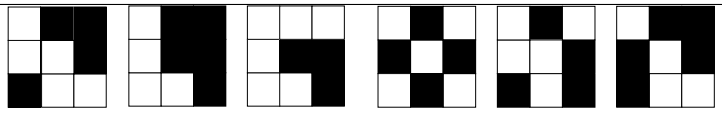
			Reference	
	Articulation Type	Mock-EPG Frame	Velar	Alveolar
(a)	reference velar		–	–
(b)	reference alveolar		–	–
(c)	velar		0.10	0.61
(d)	velar		0.20	0.58
(e)	alveolar		0.61	0.10
(f)	alveolar		0.54	0.41
(g)	alveolar and velar		0.41	0.46

However, these results are only based on the comparison of single frames of data. In the following two sections we demonstrate how the Delta method can be used to compare full epochs of data and can therefore account for temporal aspects of articulation.

4.4.2 Equal length epoch comparisons of EPG data

In this section we demonstrate how to calculate Δ for a comparison of epochs. Epochs contain several frames of data acquired over time. The procedure for calculating Δ is similar to the single frame calculations. First, each frame within the epochs must be converted into a vector. See, for example, Table 4.3 in which each frame of the raw data (a) is converted into vectors (b, c, d). Once the vectors are generated, Euclidean distance must be calculated between each pair of vectors. For example, Euclidean distance is calculated for the difference between the first vector

Table 4.3: A demonstration of a Δ calculation for two equal length epochs of grid data: (a) two epochs of raw data which each have three frames, (b) the vectors corresponding to the first frame of the epochs, (c) the vectors corresponding to the second frame of the epochs, (d) the vectors corresponding to the third frame of the epochs. Δ is equal to the mean Euclidean distance of each pairwise vector comparison.

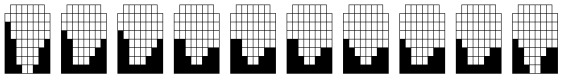
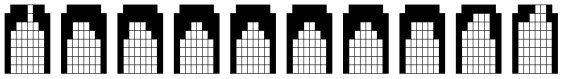
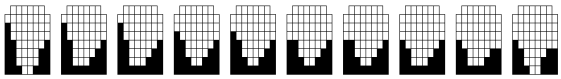
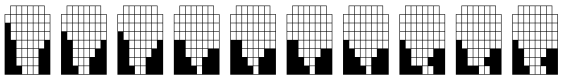
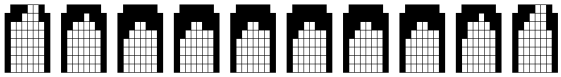
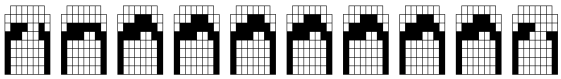
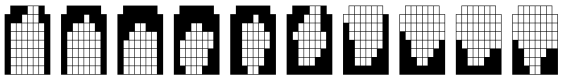
	Epoch_A	Epoch_B	Euclidean Distance
(a)			
(b)	(1 0 0 0 0 1 0 1 1)	(0 1 0 1 0 1 0 1 0)	2.00
(c)	(0 0 1 0 1 1 0 1 1)	(1 0 1 0 0 1 0 1 0)	+ 1.73
(d)	(0 0 1 0 1 1 0 0 0)	(1 0 0 1 0 1 0 1 1)	+ 2.45
			6.18 / 3
			$\Delta = 2.06$

of Epoch_A and the first vector of Epoch_B. The final Δ value is equal to the mean Euclidean distance for all of the pairwise vector comparisons.

To illustrate how Δ differs for EPG analyses we calculated Δ for a set of mock-EPG epochs relative to a reference velar epoch and a reference alveolar epoch. We always define the starting point of each epoch as the palate before full closure and the ending point of each epoch as the palate after the closure release. Full closure includes any continuous seal of contact across the lateral axis of the palate. An exception to this is when full velar closure is not recorded because it occurred posterior to the palate. For the purposes of this demonstration we extended velar cases using a two-contact rule: we select the frame before two contacts are open, the frame after more than two contacts are open and the intermediate frames (see Table 4.4.d).

The results of the Δ calculations are displayed in Table 4.4. It is clear from the calculations that comparisons of similar articulations yield lower Δ values and comparisons of dissimilar articulation yield higher Δ values. For example, the velar articulations have low Δ values relative to the velar reference and one of the alveolar articulations has a low Δ value relative to the alveolar reference. Other articulations have high Δ values when they are dissimilar to the reference. There are high Δ values for the comparison of the alveolar articulation with posterior contact (lower on the palate) relative to the alveolar reference and for velar articulations compared

Table 4.4: Δ values for the comparisons of equal length epochs of mock-EPG data. All the Δ values are calculated relative to the velar reference epoch (a) and the alveolar reference epoch (b). The values demonstrate that the Delta method yields low values for similar articulation comparisons and high values for dissimilar articulation comparisons.

	Articulation Type	Mock-EPG Epoch	Reference	
			Velar	Alveolar
(a)	reference velar		–	–
(b)	reference alveolar		–	–
(c)	velar		0.82	5.66
(d)	velar		1.66	5.61
(e)	alveolar		5.66	1.48
(f)	alveolar		5.02	4.30
(g)	alveolar and velar		3.54	3.80

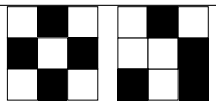
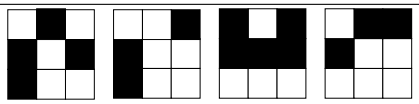
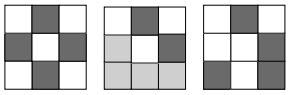
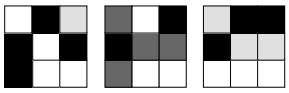
to the alveolar reference. Lastly, the articulation that contains both alveolar and velar contact has a relatively high Δ value when compared to either the alveolar or the velar references. This pattern of results is consistent with the results from the previous section: low Δ values for similar articulations and high Δ values for dissimilar articulations. This suggests that comparing articulations with the Delta method can capture relative differences between different patterns of articulation.

Note that so far we have only compared epochs of equal length. This can potentially be limiting since articulations are likely to vary in duration. It is therefore required that the Delta method be extended to include comparisons of epochs of different lengths. In the following section we demonstrate how this can be accomplished.

4.4.3 Unequal length epoch comparisons of EPG data

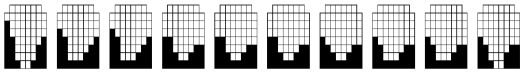
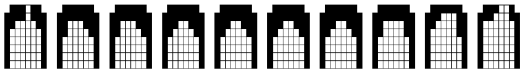
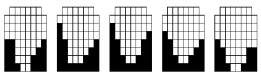
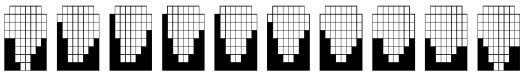
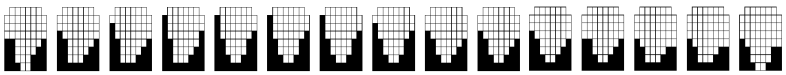
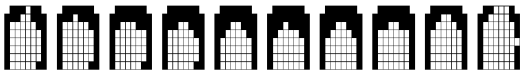
To use the Delta method to compare epochs of different lengths, the epochs must first be standardised to an equal length. This is accomplished by stretching or

Table 4.5: A demonstration of a Δ calculation for the comparison of unequal length epochs of grid data: (a) raw epoch data in which Epoch_A is two frames and Epoch_B is four frames; (b) epoch data that has been standardised to a length of three frames using an averaging algorithm; (c) the vectors corresponding to the first frame of standardised data; (d) the vectors corresponding to the second frame of standardised data; and (e) the vectors corresponding to the third frame of standardised data. Δ is equal to the mean Euclidean distance of all the pairwise vector comparisons.

	Epoch_A	Epoch_B	Euclidean Distance
(a)			
(b)			
(c)	 (0 .66 0 .66 0 .66 0 .66 0)	 (1.33 0 0 1.33 0 1 0 1 .33)	1.73
(d)	 (.33 .33 .33 .33 0 .66 0 .66 0)	 (.66 0 0 1.33 .66 .66 .66 0 1.33)	+ 2.10
(e)	 (.66 0 .66 0 0 .66 0 .66 0)	 (0 0 0 1.33 .33 .33 .33 1 1.33)	+ 1.76
			5.59 / 3
			$\Delta = 1.86$

shrinking the number of raw data frames with an averaging algorithm. The averaging algorithm equally distributes contact across the standardised epoch using contact values that begin with 0 (no contact) and range upwards so that a higher contact value represents more continuous contact across frames. Table 4.5.a contains two epochs of grid data that differ in length: Epoch_A has two frames and Epoch_B has four frames. To compare these epochs the number of frames must be standardised, which we arbitrarily set to three frames. To standardise Epoch_A we stretched the raw epoch to three frames using the following formula: $\text{frame1}(\frac{2}{3}) + (\text{frame1}(\frac{1}{3}) + \text{frame2}(\frac{1}{3})) + \text{frame3}(\frac{2}{3})$. To standardise Epoch_B to shrink to three frames we used a similar formula: $(\text{frame1} + \text{frame2}(\frac{1}{3})) + (\text{frame2}(\frac{2}{3}) + \text{frame3}(\frac{2}{3})) + (\text{frame3}(\frac{1}{3}) + \text{frame4})$. In the figure, darker coloured squares represents a higher contact value. Once the standardised epochs are generated they must then be converted into vectors. Then Δ is calculated using the same procedure of the previous section since the epochs being compared are now the same length: Δ is equal to the mean Euclidean distance of the pairwise vector comparisons.

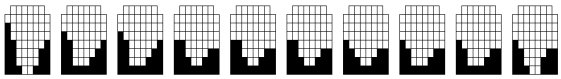
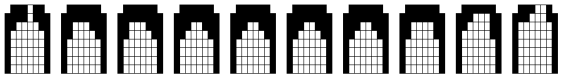
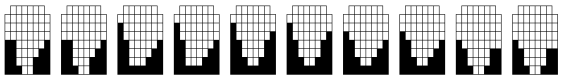
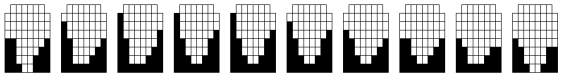
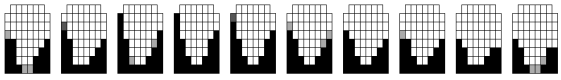
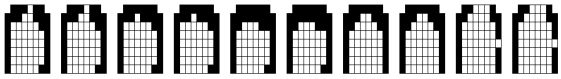
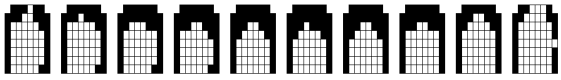
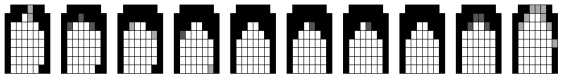
Table 4.6: Raw mock-EPG epochs of different lengths which were used for the demonstration of the Delta method

	Articulation Type	Raw Mock-EPG Epochs
(a)	velar reference	
(b)	alveolar reference	
(c)	velar short	
(d)	velar medium	
(e)	velar long	
(f)	alveolar short	
(g)	alveolar medium	
(h)	alveolar long	

To illustrate how Δ values vary for different length epochs we generated another set of mock-EPG data. This set includes velar and alveolar articulations that are short (five frames), medium (ten frames), and long (15 frames) in length and a velar and alveolar reference that have a medium length. These mock-EPG epochs are displayed in Table 4.6. We then standardised all of the epochs (excluding the references) to have an equal length of ten frames. This was accomplished using the averaging algorithm. The standardised epochs are displayed in Table 4.7.

We calculated a Δ value for each standardised epoch relative to the velar and alveolar reference. The results are presented in Table 4.7. An inspection of the Δ values for the standardised epochs reveals that all of the velar articulations have lower Δ values compared to the velar reference and all of the alveolar articulations have lower Δ values compared to the alveolar reference. This provides additional evidence that the Delta method yields lower Δ values for comparisons of similar articulations and higher Δ values for comparisons of dissimilar articulations. Importantly, the Δ values are also sensitive to the length of the articulation. The articulations which, in their raw form were equal in length to the references, yielded the lowest Δ values. This suggests that they are more similar to the reference than the shorter and

Table 4.7: Δ values for the comparisons of standardised epochs of mock-EPG data. All of the values were calculated relative to the velar and alveolar references. The values demonstrate that the Delta method yields values that are sensitive to similarity in both spatial and temporal differences.

	Articulation Type	EPG Frames	References	
			Velar	Alveolar
(a)	velar reference		–	–
(b)	alveolar reference		–	–
(c)	short velar		2.47	5.03
(d)	medium velar		1.61	5.63
(e)	long velar		3.16	6.99
(f)	short alveolar		4.14	3.31
(g)	medium alveolar		5.73	1.35
(h)	long alveolar		7.74	3.12

longer articulations. This pattern establishes evidence that the Delta method can account for temporal variability in addition to spatial variability.

4.4.4 Real EPG Demonstration

In the preceding three sections we demonstrated how the Delta method is calculated. This demonstration also illustrated that spatial and temporal variability is captured by the Delta method. However, these demonstrations were based on mock data to simplify the procedure for illustration purposes. In this section we present a demonstration of the Delta method using real EPG data collected from one speaker.

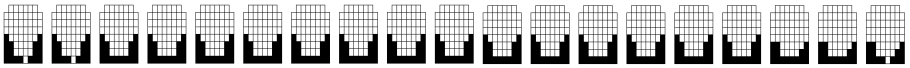
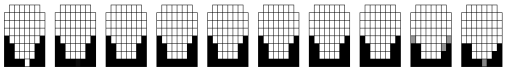
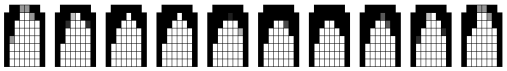
Prior to testing, the speaker was fitted with a custom electropalatography (EPG) palate (manufactured by Incidental, Newbury, UK or Grove Orthodontics, Norfolk, UK). The palate was moulded to fit a dental cast from an impression of the speaker's hard palate. The EPG palate is made of acrylic and contains 62 embedded silver

switches on the lingual surface of the artificial palate, organised in seven rows of eight contacts and for the anterior most part one row of six contacts. EPG data was recorded at a rate of 100Hz using the WinEPG system (Articulate Instruments Ltd, Edinburgh, UK), which connected the palate to a multiplexer unit that transferred the data to an EPG3 scanner and then to the serial port of a desktop computer. An acoustic recording of the speaker's responses was recorded at 22,050Hz using an Audio Technica ATM10a microphone. A desktop computer, to which the microphone and WinEPG system were attached, was used to record the speaker's responses with Articulate Assistant (Wrench, 2003) software. Stimuli were presented using the display prompt of Articulate Assistant on a 15" LCD monitor. Auditory signals of a metronome beat were played at a rate of 100 syllables/minute through monaural headphones worn in the speaker's preferred ear. The speaker was instructed to repeat *kom* sixteen times and then to repeat *tom* sixteen times at a rate of one word per metronome beat.

After recording, the EPG and audio data were exported from Articulate Assistant for analysis. The articulatory records of interest (all /k/s and /t/s) were created by identifying the offset of the previous word and the onset of the vowel of each target word from the acoustic signal in Praat (Boersma & Weenink, 2006). The EPG data was then visually inspected and trimmed to only include the palate before full closure through to the palate following the closure release. Full closure was defined as any continuous path across the lateral axis of the palate. Once the articulatory records were created, they were standardised to an equal length of ten frames. This was accomplished by using the averaging algorithm described in Section 4.4.3. Samples of the raw and standardised data for a /k/ and /t/ production are displayed in Table 4.8.

To demonstrate the range of Δ values calculated for the full set of EPG data, we calculated Δ for each articulatory record relative to every other articulatory record and generated a (dis)similarity matrix containing all of the values. We then used a multidimensional scaling algorithm (Cailliez, 1983; Cox & Cox, 1994) to visualise the results. Multidimensional scaling takes a set of similarity values (e.g., Δ values) and returns a set of points on a scatter plot so the distance between the points of the plot are approximately equal to similarity values between the points. The results from the multidimensional scaling analysis are presented in Figure 4.4. An inspection of this plot reveals a clear cluster of /k/ articulations and another cluster of /t/ articulations. This provides clear evidence that the Delta method can successfully capture relative differences between velar and alveolar articulations.

Table 4.8: Sample EPG records of a /k/ and /t/ used for the multidimensional scaling analysis. The first set of records represents the raw epochs of data and the second set represents the same data after it has been standardised to a length of ten frames

Onset Phoneme	EPG Record
Raw Data:	
/k/	
/t/	
Standardised Data:	
/k/	
/t/	

This finding is important because it establishes evidence that the Delta method can successfully capture differences between experimentally elicited EPG records.

4.5 Conclusions

In this chapter we introduced the use of EPG to measure articulation during speech production. This articulatory imaging technique allows the measurement of spatial and temporal characteristics of tongue-to-palate contact and the simultaneous collection of acoustic data. The ability to measure articulation during speech production allows researchers to directly investigate the interface of speech planning to articulation.

Several methods for analysing EPG data have been proposed in the literature, however, they are inadequate for our purposes. Categorising responses implies that representations are selected that fit into the assigned categories. If partially activated phonological representations can cascade to articulation, then articulation can not be categorised. According to a cascaded model, articulation can reflect continuous levels of activation and therefore can not be categorised. A quantitative measure of articulation is required to demonstrate that articulation in one condition differs from articulation in another condition.

Existing quantitative methods for analysing EPG suffer from some limitations. Most methods for quantifying EPG data are based on predefined spatial indices.

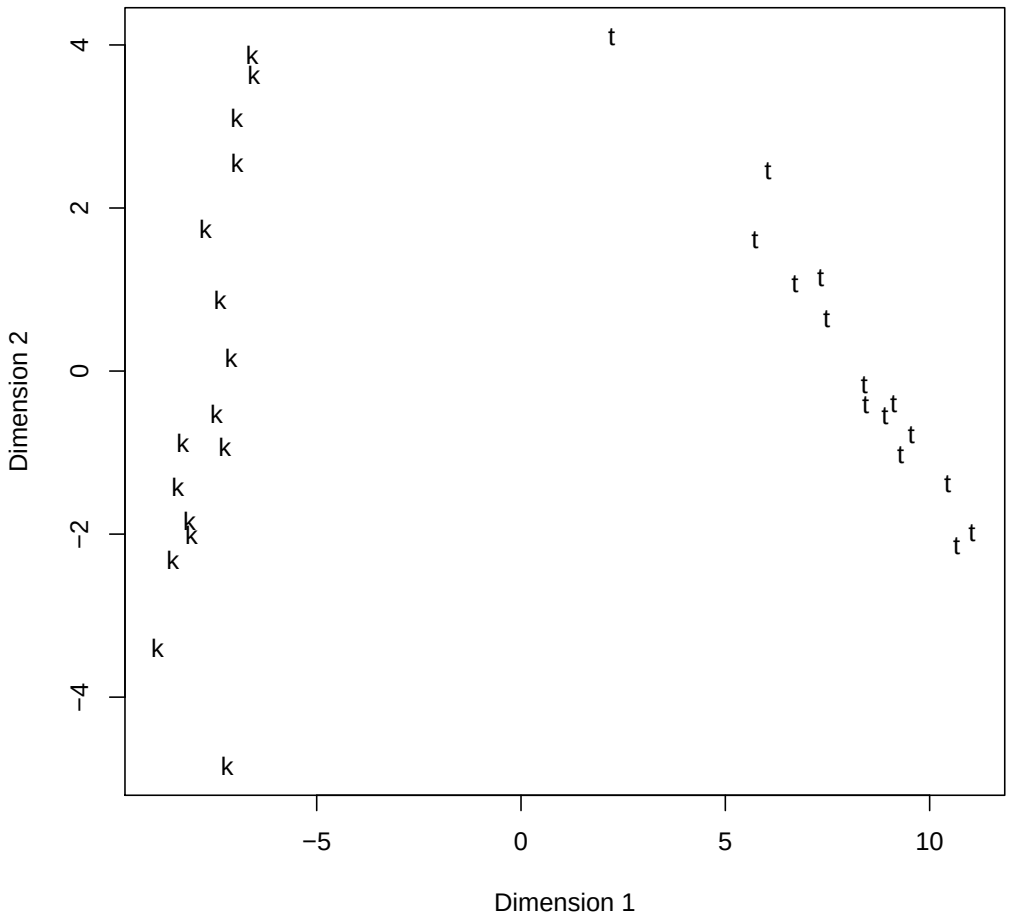


Figure 4.4: A multidimensional scaling plot for the comparisons of 16 /t/ articulations and 16 /k/ articulations recorded with EPG. Δ was calculated for each articulation relative to every other articulation. The plot reveals a clear cluster of /t/s and a clear cluster of /k/s which demonstrates that the Delta method can capture the differences between alveolar and velar articulations

These require specific assumptions about the space in which articulation will occur. Relatedly, the spatial regions on which the indices are based may not correspond with absolute landmarks in the oral cavity, which limits the conclusions that can be drawn about spatial characteristics of articulation and individual differences between speakers. Finally, spatial indices often do not account for temporal characteristics of articulatory patterns. Since articulation is a serial processes that occurs over time this limitation results in the exclusion of an important aspect of articulation.

To account for the inadequacies of previously used analysis methods we developed the Delta method for quantifying articulatory patterns. This method generates a value (Δ) that represents how (dis)similar one articulation is compared to a reference articulation. A lower value indicates that two articulations are similar and a higher value indicates that two articulations are dissimilar. The values produced by the Delta method are sensitive to variability in both spatial and temporal characteristics of articulation

The most important feature of the Delta method is that it is based on relative rather than absolute comparisons. Using a relative measure of articulation can accommodate for the limitations of previous EPG analysis methods. Specifically, a relative measure does not require responses to be categorised or assumptions about the spatial and temporal characteristics of articulation. By using different references (e.g., a velar reference and an alveolar reference) for the Δ comparisons we can then interpret whether an articulation was more velar-like or more alveolar-like based on whichever reference comparison yielded the lowest (or most similar) value. Moreover, references can be created for individual speakers to account for individual differences.

Since the Delta method is sensitive to spatial and temporal characteristics of articulation and does not require categorisation, it is a useful technique for investigating the research questions throughout this thesis. The articulatory analyses presented in later chapters were not designed to investigate the fine phonetic or physiological details of how different tongue positions yield different acoustic sounds, but rather our research question is whether articulation in one condition differs in some way from articulation in a comparable condition. Since Δ is a ratio variable it can be used for further analysis using standard quantitative methods. One limitation of EPG is that a custom artificial palate must be manufactured for each speaker. A practical consequence of this precision is that it is very costly to manufacture

palates for each speaker. As a result EPG investigations typically have a low number of speakers. Therefore, the ability to use standard statistical analyses with by-participant and by-item factors allows us to directly test whether articulation differs across experimental conditions.

In the remainder of this thesis we use the Delta method to investigate whether feedback is required in models of production that allow partial representations to cascade to articulation. We report two experiments that use the Delta method for analysing EPG data. In Chapter 7 the Delta method is extended for the analysis of ultrasound data. Importantly, the results of an ultrasound experiment reported in Chapter 7 replicate the results of an EPG experiment reported in Chapter 6. Among other factors, this replication establishes that the Delta method can be used for different modalities of articulatory imaging.

4.6 Chapter Summary

In this chapter we presented an overview of EPG and the methods that have been used to analyse articulatory records. All of the analysis methods reviewed had some limitations that rendered the methods inadequate for our research. In contrast, the Delta method which we developed has advantages over previous methods. Importantly, it does not require categorisation or spatio-temporal assumptions about articulation. Therefore, we argue that the Delta method is a useful approach for investigating the interface between speech planning and articulation in the remainder of this thesis.

Chapter 5

The Influences of Lexical & Contextual Competition on Articulatory Variability[1]

5.1 Chapter Overview

This chapter begins with a review of the methods that have been used to identify non-canonical speech errors from articulatory and acoustic records. However, like transcription-based investigations, these techniques have relied on categorising responses which can not be used to investigate speech in a cascaded model of production. We present a reanalysis of the EPG data from the word order competition (WOC) task using the Delta method, which does not require assigning responses to categories. The results of this analysis demonstrate that onsets are articulated more similarly to the competitor onset when they have real word competitors than when a competitor substitution would result in a nonword. An account that incorporates feedback between phonological and lexical representations in a cascading model of speech production is proposed.

5.2 Introduction

The observation of a high rate of 'other' errors, in particular double articulations, in Experiment 1 (Chapter 3) is consistent with previous acoustic and articulatory investigations of speech errors (Frisch, 2007; Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein et al., 2007; Laver, 1980; Mowrey & MacKay, 1990; Pouplier, 2003, 2007; Stearns, 2006). Together, these investigations provide experimental evidence for the existence of non-canonical errors. The most straightforward account for these errors is within a cascading model of production: articulation reflects simultaneous partial activation of both competitor and target phonological

[1]Portions of the data reported in this chapter are included in McMillan et al. (in press)

representations. Importantly, in a cascading activation model there is no *a priori* reason to assume that errorful and non-errorful articulations will differ qualitatively from one another. Instead the *a priori* assumption should be that all articulatory responses reflect the extent to which there is competition between representations. In this chapter, a review of how instrumental errors have been categorised will be presented and the limitations of this approach will be discussed. A new approach for investigating slips of the tongue will be presented together with a reanalysis of the articulation data from the WOC task (Chapter 3).

5.2.1 Categorical Limitations

An oft-cited motivation for instrumental investigations of speech errors is that they are not vulnerable to limitations associated with auditory transcription (Frisch & Wright, 2002; Goldstein et al., 2007; Pouplier, 2007; Pouplier & Hardcastle, 2005). However, all of these investigations operate by assigning responses to categories. Previous instrumental investigations of speech errors have relied on either qualitative assessments of errors (Laver, 1980; Mowrey & MacKay, 1990) or an arbitrary criterion for identifying errors (Goldstein et al., 2007; Frisch, 2007). These approaches, while less vulnerable to the difficulties of auditory transcription, are still confined to the categorisation of responses as 'correct' or 'errors'.

One method used for the quantitative categorisation of articulatory recordings is identifying those articulations that substantially differ from a mean articulation as "errorful" (Frisch, 2007; Goldstein et al., 2007). In an EMMA analysis of speech errors Goldstein et al. (2007) aligned the signal of each relevant transducer (e.g., glued to tongue-dorsum and tongue blade) for all targets (e.g., *cop top*). From these overlaid signals, which represent the degree of transducer raising over time, a mean signal was calculated. Errors were identified in two ways: any deviation greater than two standard deviations from the mean intended signal (e.g., /k/ relative to the mean-/k/) was labelled as a "full" error and any deviation greater than two standard deviations from both the mean intended and mean competitor signals (e.g., /k/ relative to the mean-/k/ and the mean-/t/) was labelled as a "partial" error. Similarly, Frisch (2007) and Stearns (2006) use a two standard deviation criterion to identify errors in the midsagittal plane of ultrasound recordings.

Using a two standard deviation criterion is problematic for two primary reasons. First, a standard deviation criterion in any analysis restricts the number of relevant observations to a small subset of the data. If a distribution of tongue-tip raising

was normally distributed then a two standard deviation cut-off would only allow 5% of items to be categorised as errors. As has been previously argued, limiting analyses to a small proportion of observations may add noise to any statistical analysis (Nooteboom & Quené, 2007).

Second, and more importantly, a two standard deviation cut-off is an arbitrary criterion. There is no *a priori* reason to assume that any variation in the articulatory signal greater than an arbitrary criterion reflects a qualitatively different process than variability that is close to the mean. In using such a criterion one is making an assumption about what constitutes an error despite the lack of any theoretical or experimental evidence that highly deviant articulation reflects a qualitatively different process from less deviant articulation.

Another criterion that has been used to categorise errors is based on auditory perception. Pouplier (2007) argued that articulatory measurements are superior to auditory transcription because of transcription limitations, but excluded target responses including any auditorily perceivable errors beyond the initial consonant. This approach makes the assumption that onsets may differ in some way from normal variation if their corresponding rime is perceivably different. If auditory perception is not suitable for categorising onsets then it also should not be suitable for categorising rimes. If analyses are to account for normal variation then there is no reason to exclude data unless it can not be analysed for a particular reason, such as technical limitations.

Two instrumental investigations have presented descriptive arguments based on the distributions of all recorded responses, rather than limiting analyses to the distributions of 'errors'. Frisch and Wright (2002) observed a continuum of percent-voicing in an analysis of /s/-/z/ tongue-twister productions, which they interpreted as support for two types of errors: "categorical" errors which contained percent voicing equivalent to a canonical production of the competing phoneme and "gradient" errors which contained non-canonical percent voicing. Goldstein et al. (2007) also proposed a distinction between "partial" and "full" errors. However, in a similar analysis on SLIP task productions of /t/-/k/ Pouplier (2007) suggested that she observed a (mostly) normal, rather than bimodal, distribution of tongue-tip and tongue-dorsum raising. This was interpreted as lack of evidence for categorically different errors. However, these analyses also have methodological limitations. Frisch and Wright's (2002) distribution was based on assigning responses to categories of voicing (e.g., 0%, 0–5%, 5–10%, 10–30%...100%) and Pouplier's (2007) distribution was not evaluated quantitatively.

In summary, instrumental investigations of speech errors have been limited by the implicit categorisation of responses. Categorising responses assumes a staged model of production in which each articulatory category must reflect the (mis)selection of a phonological representation. However, articulatory responses can vary continuously. In Chapter 3, a qualitative evaluation of articulatory patterns recorded with EPG revealed that it is difficult to assign responses to categories. For example, it is not clear where to create a boundary between different kinds of 'double articulations'. Even categorising articulations as 'correct' is difficult because any boundary between categories is arbitrary. Rather than arbitrarily defining a response as belonging to one category or another, the quantification of articulation during error elicitation tasks allows the variability between articulations in different experimental conditions to be measured.

In this chapter we investigate the influences of lexical competition and context on articulatory variability. This investigation, which does not rely on categorising responses, is important since previous instrumental investigations relying on categorisation yielded mixed findings. In an acoustic investigation Goldrick and Blumstein (2006) auditorily identified substitution 'errors' and measured the VOT duration of those responses. A *post-hoc* analysis demonstrated that the VOT of an errorfully produced target was more similar in duration to the VOT of a correctly produced competitor phoneme if the competitor phoneme yielded a real word. In another acoustic investigation, it was demonstrated that /s/ → /z/ errors were more likely for real word competitors, but the findings for /z/ → /s/ were less clear. For the latter there was a clear lexical bias for canonical substitutions, but 'non-canonical' rates of percentage voicing were more frequent for nonword competitors (Frisch & Wright, 2002), although this may reflect the fact that low frequency words tend to be more error-prone (Dell, 1988, 1990; Stemberger, 1984; Stemberger & McWhinney, 1986). Finally, a *post-hoc* articulation analysis of SLIP task responses did not reveal a positive or negative lexical bias though the frequency of errors was too low to warrant statistical analysis (Pouplier, 2003, see also Pouplier, 2007).

We also investigate influences of phonological neighbourhood size on articulatory variation. A prediction of feedback models is that the size of a target's phonological neighbourhood will influence the likelihood of a speech error (Dell & Gordon, 2003; Vitevitch & Sommers, 2003; Vitevitch, 2002). For example, Vitevitch (2002) demonstrated in an error elicitation task and tongue-twister experiment that phonological errors were more likely for targets with sparse neighbourhoods

compared to those with dense neighbourhoods. Their finding suggests that phonological neighbours facilitate production through reinforcement of phonological and lexical representations. The activation of the target phonological representation is stronger for dense neighbours and therefore an error is less likely. However, the activation of the competitor phonological representation is stronger for sparse neighbours making a substitution error more likely. Additional evidence for facilitatory neighbourhood effects include picture naming studies in healthy (Vitevitch & Sommers, 2003; Roach, Schwartz, Martin, Grewal, & Brecher, 1996) and aphasic adults (Gordon, 2002), and in the occurrence of tip-of-the-tongue phenomenon (Harley & Brown, 1998; Vitevitch & Sommers, 2003). These accounts, however, have been based on the assumption that production is staged and that a single phonological representation is selected for articulation.

A cascading model of production that includes feedback also makes a prediction about the influence of phonological neighbourhood size on articulatory variability. If competing phonological representations have several lexical representations to feed back to, the target phonological representation will receive more reinforcement since activation will be distributed to several different phonological competitors. However, if there are only one or two lexical representations to feed back to, both the target and the competitor will be reinforced since activation will be less distributed. Therefore, articulation should be more variable for smaller phonological neighbourhoods since the articulatory signal may include traces of the target and competitor phonological representations.

In order to investigate articulation in a cascading model of production a new approach is required that does not rely on categorisation. In this chapter we present a reanalysis of the EPG data from Experiment 1 (Chapter 3) using the Delta method (described in detail in Chapter 4). Using the Delta method allows articulatory patterns to be analysed without pre-assigning participants' responses to categories. Reanalysing the EPG recordings in this way abandons the concept of categories of speech errors. Instead, an *a priori* assumption is made that all articulations under conditions designed to promote speech errors may have been affected in some way. Therefore, the EPG recordings from the WOC task are compared to those collected during a control task, which can be defined *a priori* as reference articulations because they were obtained under conditions that should not promote tongue slips above normal variation. Reanalysing the WOC data, which includes a context (whether the participant sees any real words) and competitor (whether a full phonemic substitution would result in a real word or not) manipulation, allows the

effects of higher-level speech planning processes on the articulation of individual phonemes to be investigated.

5.3 Reanalysis of WOC Data

5.3.1 Articulation Method

WOC Task

The EPG records from seven speakers collected during the WOC task reported in Chapter 3 were used for this analysis. In the task participants were presented with nonword target pairs (e.g., *keam turve*) and cued with a leftward arrow to produce the nonwords in reverse order (e.g., "turve keam"). Targets were manipulated in a counterbalanced design so that a competitor substitution would result in a real word (e.g., *keam (turve)* → "team") or a nonword (e.g., *keeb (turp)* → "teeb"). Targets were also embedded in either a nonlexical (all nonwords) or a mixed (nonwords and real words) context.

Control Task

The seven participants who were recorded with EPG in the WOC task (see Chapter 3) were also recorded in a control task. We used the same stimulus items as the WOC Task, but only presented the original target pairs, each preceded by one filler pair. The procedure was identical except that all target pairs were followed by a right arrow, cueing the participants to speak the target and filler items in the presented order. For example, in the WOC Task participants were presented with *gope doof* and cued with a left arrow to respond "doof gope", but in this task they were cued with a right arrow to respond "gope doof". The control recording task resulted in the presentation of 672 nonword target onsets in circumstances which were not designed to elicit errors.

EPG Processing

Tongue-to-palate contact variability was calculated for all EPG records that were suitable for analysis from the WOC and Control tasks. Tokens were identified as suitable provided they included full closure across any lateral continuous path of the EPG palate and could be trimmed to include the palate before full closure to the palate after closure release. Targets in which velar closure did not yield a continuous

path across the palate were reexamined. If closure was achieved across all but the middle two posterior contacts at any point during the articulation, this was treated as full closure. EPG records that did not include this degree of velar closure, or full closure at other positions, were excluded from the articulation analysis because the start and end points could not be reliably identified. After excluding items without full closure the analysis included a total of 564 WOC task tokens and 609 Control task tokens out of 672 possible responses for each task. No more than 14.5% of the data was excluded for any given participant and the excluded items were evenly distributed across cells in the design matrix.

To calculate tongue-to-palate contact variability, a reference articulation was created for each speaker for each relevant place of articulation from the control task EPG records. This was accomplished by standardising the epochs of all of the EPG recordings which contained full closure (see Section 4.4) so that each record contained ten frames. The standardised tokens were then averaged together to create a mean velar and mean alveolar reference articulation for each speaker.

For each of the relevant EPG recordings from the WOC Task (i.e., all recordings which included full closure at any point) we then calculated difference scores from that speaker's reference articulations. This was accomplished using the Delta method (see Section 4.4), in which each EPG record is treated as a series of 62-dimensional vectors over time. To quantify variability each WOC record was standardised to include ten epochs. The Euclidean distance between each record and the corresponding reference articulation was then calculated.

A difference calculation was conducted twice for each EPG recording from the WOC task: once relative to the intended articulation reference and once relative to the competitor articulation reference. The results of the calculations were measures of deviance in contact from the intended onsets (the higher the score, the less like a 'typical' intended onset a particular recorded articulation was) and of deviance in contact from the competitor onsets (the lower the score, the more like a 'typical' competitor onset). Table 5.1 give examples of EPG recordings and the derived deviance scores relative to an alveolar and velar reference.

5.3.2 Reanalysis Results

To investigate the influences of lexical competition and context on articulation two analyses of the tongue-contact variability data were conducted. The first used the difference scores calculated relative to the intended onset references; the second

Table 5.1: Sample EPG recordings of intended alveolar and velar articulations from Experiment 1 with their corresponding difference scores (Δ) relative to an alveolar and velar reference: (a) a typical alveolar; (b) a less typical alveolar indicated by a higher difference score relative to the alveolar reference; (c, d) 'double articulations' which contain either overlapping (c) or non-overlapping (d) alveolar and velar contact; (e) less typical velar indicated by a higher difference score relative to the velar reference; and (f) a typical velar articulation.

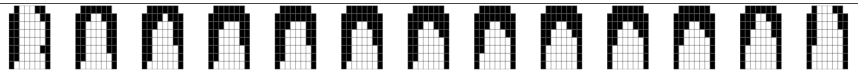
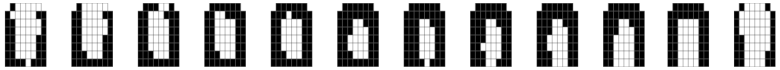
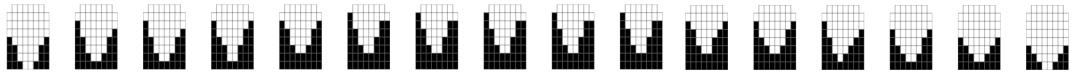
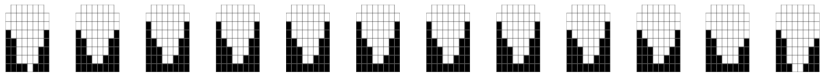
	Token	Reference	
		Alveolar	Velar
a		1.38	5.13
b		3.95	4.59
c		3.15	4.00
d		3.62	3.89
e		4.31	2.56
f		5.17	1.04

Table 5.2: Mean difference scores (Δ; standard deviation in brackets) for the reanalysis of the EPG data recorded in Experiment 1. Higher values indicate greater tongue contact variability.

Context	Nonlexical	Mixed
Variability from intended onset		
Nonword competitor	2.07 (.20)	1.89 (.17)
Real word competitor	2.12 (.24)	1.98 (.29)
Variability from competitor onset		
Nonword competitor	4.52 (.29)	4.58 (.31)
Real word competitor	4.40 (.35)	4.43 (.27)

used difference scores relative to the competitor onset references. For both analyses by-participant (F1), by-item (F2), and minF' ANOVA statistics with Competitor (real word, nonword) and Context (nonlexical, mixed) as within-subjects factors are reported. See Table 5.2 for means and standard deviations of variability across conditions.

The first analysis includes tongue contact variability calculated relative to the reference articulation of the intended target onset. There was a marginal effect of Context; articulation tended to be more similar to the intended target in the mixed context than in the nonlexical context (difference scores of 1.93 vs. 2.09; 95% CI ± 0.06). See Table 5.3 for a summary of ANOVA statistics.

The second analysis was based on the tongue contact variability of each target calculated relative to the reference articulation of the competitor onset. This analysis showed a significant effect of Competitor (marginal by-items): Articulation was more similar to the competitor onsets when targets had real word competitors than when they did not (difference scores of 4.41 vs. 4.55; 95% CI ± 0.08). See Table 5.4 for a summary of ANOVA statistics.

Table 5.3: ANOVA statistics for the analysis of tongue-contact variability calculated relative to the intended target.

Source of variance	By participants			By items		minF'	
	df	F1	MSe	df	F2	df	minF'
Competitor	1,6	2.21	.002	1,47	<1	1,38	<1
Context	1,6	4.81^{m}	.004	1,47	3.39^{m}	1,28	1.99
Competitor × Context	1,6	<1	.002	1,47	<1	1,21	<1

^{m}marginal effect, p<.075; * p<.05

Table 5.4: ANOVA statistics for the analysis of tongue-contact variability calculated relative to the competitor.

Source of variance	By participants			By items		minF'	
	df	F1	MSe	df	F2	df	minF'
Competitor	1,6	15.71**	.0008	1,47	3.58^{m}	1,50	2.92
Context	1,6	<1	.003	1,47	1.06	1,20	<1
Competitor × Context	1,6	<1	.002	1,47	<1	1,21	<1

^{m}marginal effect, p=.065; ** p<.01

Two additional analyses were conducted to test the feedback prediction for phonological neighbourhood size. To focus on the role of competing onsets, phonological onset neighbourhood size of each target was calculated by counting the number of onset substitutions that yield a real word in the CELEX English database (*CELEX English database - Release E25 [On-line]*, 1993). For example, the target "dolf" has a phonological onset neighbourhood size of 1 because GOLF is the only real word that can result from a substitution. A Pearson's product-moment correlation of phonological onset neighbourhood size and contact variability from the intended target revealed a significant negative correlation [r(597)=–.12, p<.005] in which articulation was more variable from the target as the phonological neighbourhood size decreased. A correlation of phonological onset neighbourhood size and contact variability towards the competitor target was also significant [r(597)=–.10;p<.05] in the same direction.

5.3.3 Reanalysis Discussion

The analysis of articulation variability confirms the pattern of results reported from the transcription analysis in Chapter 3; speech is more affected in cases where there is a real word competitor but this effect is not sensitive to the nature (mixed, or fully nonlexical) of the context. However, this analysis extends the earlier findings in three important ways. First, it provides evidence that lexical competition affects articulation. Where target onsets have real word competitors (e.g., *gome* could result in "dome"), the articulation is more similar to the competing real word onset (/d/) than in cases where there is no real word competitor (*gofe* → "dofe").

Second, a comparison of the analyses using target and competitor onset references establishes clearly that the differences in articulation must be attributed to the influence of competitor onsets: although *gome* → "dome" results in a "more /d/-like" onset than *gofe* → "dofe", *gome* → "dome" does not result in a "less /g/-like"

onset than for the equivalent nonword onset competitor. In other words, onsets are attracted *toward* a real word competitor, rather than repelled *away* from the target onset. This extends evidence showing that VOT can be affected by lexical status (Goldrick & Blumstein, 2006) because VOT is a continuous measurement. Therefore a "less /g/-like" /g/ will tend to be "more /k/-like" along that dimension, regardless of whether the difference is due to the influence of a /k/ (a similar argument can be made for degree of voicing, as in Frisch & Wright, 2002).

Finally, to our knowledge, this is the first demonstration of the effects of context on articulation. Participants produced articulations which varied more from the intended target onsets when the context consisted entirely of nonwords than when it was mixed. Importantly, this effect also highlights a distinction between variability from an intended articulation that differs from variability towards a competitor articulation.

The correlation analyses of phonological neighbourhood and articulatory variation revealed that articulation is more variable the smaller the neighbourhood size. This pattern was found when variability was calculated relative to the target and the competitor onset references. We return to this point in the following section.

5.4 General Discussion

Several theorists have proposed that activation in the speech production system can cascade from phonological encoding to articulation (e.g., Goldrick & Blumstein, 2006; Kello et al., 2000). According to such a view, competition, such as that caused by the WOC paradigm, can result in partial activation of not only the intended but also the competitor onset phonemes. To the extent that the competitor onset is active, it will influence the eventual articulation of a target.

Such a view is clearly compatible with our findings. In cases where there is competition, articulation of the onset becomes more similar to a 'canonical' articulation of the competitor phoneme. Moreover, the clear influence of the competitor is modulated by lexical status; the effect is greatest when a substitution of the competitor onset would result in a real word, whether the context contains real words or not. Importantly, we can rule out the possibility that competition from the lexical level simply adds noise to articulation, because dissimilarity from the target phoneme is not affected by lexical status. Our analysis demonstrates that, for example, variation in /k/ articulation may be more /g/-like or less /k/-like, but not necessarily

both. This distinction is not possible in any investigation that only measures VOT since a difference between /k/ and /g/ occurs on a continuum so that any less /k/-like production will be more /g/-like.

5.4.1 Lexical Effects on Articulation

The observation of lexical influences on articulation is an important finding given mixed previous findings (Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Pouplier, 2003), which have relied on different categorical methods. The most straightforward way of accounting for the overall lexical bias in articulation found in the present study is to incorporate feedback between phonological and lexical representations into the model. According to this account, the activation of a competing phonological representation and the other relevant activated phonological representations would feed back to activate a lexical representation. The activation of the lexical representation would reinforce the activation of the competitor phonological representation, which in turn, because it is sufficiently activated, would cascade to articulation. The resulting articulation would then contain properties of the competing phonological representation. However, in cases in which there were no real word competitors and therefore no lexical representations to feed back to, the competing phonological representation would not be reinforced and would remain relatively inactive.

The two correlation analyses also revealed a pattern that can be best accounted for by feedback. According to a feedback account, competing phonological representations which have fewer lexical representations to feed back to will receive more reinforcement than competing phonological representations which have several lexical representations to feed back to. Reinforcement increases the activation of a competitor and therefore increases articulatory variability since articulation may contain both the target and competitor phonological representations. Both correlation analyses confirmed this predicted pattern: articulatory variability increased as the phonological neighbourhood size decreased.

The correlation analyses additionally raise an interesting direction for future articulatory research. For example, one could evaluate the influence of syllable frequency on articulatory variability. Such an analysis could provide a novel approach for testing whether the syllable is a functional production unit (e.g., Levelt & Wheeldon, 1994) or not (e.g., Schiller, Costa, & Colomé, 2002). However, such an analysis is not possible with the current experimental materials: only 17 out of 96 items are

included as syllables in the CELEX database (*CELEX English database - Release E25 [On-line]*, 1993).

5.4.2 Context Effects on Articulation

In contrast to the effects of lexical status, context appears to affect the general noisiness of articulation. Articulation is more variable from the target phoneme in the nonlexical context than in cases where participants encounter both nonwords and real words. It is not immediately clear how these effects might be accounted for within a model incorporating cascading and feedback between levels. In Chapter 3, an inspection of EPG records revealed that 13.2% of responses contained clear evidence of both alveolar and velar contact during the production of a word onset. These were among the records which resulted in relatively high variability scores (see Table 5.1(c) and (d) for examples). Furthermore, in Chapter 3 an inspection of the distribution of double articulations revealed that the majority of these errors were in the nonlexical context and that the timing and pattern of some of the errors was compatible with a repair-based account.

A cautious suggestion can therefore be made that a repair is more likely where the context is entirely nonlexical (regardless of the lexical status of the competitor). It is interesting to note that the repairs here are not "covert", according to Levelt's (1983) definition, in the sense that they are detectable (using appropriate instrumentation). But neither are they straightforwardly "overt"; many are not detectable by listeners. Similarly, the distinction between "internal" and "external" monitoring appears unclear. Criteria such as the "0ms repair" (Blackmer & Mitton, 1991) appear less tenable when the time between articulations can be measured precisely. But whether internal or external, Hartsuiker (2006) has emphasised that a putative monitor must be functional, in the sense that the repairs it instigates serve the present communicative purpose. In previous work, considerations concerning the functionality of the monitor have centred around the interaction between context (what sort of words should I be producing?) and lexical status (what sort of word am I about to produce? e.g., Baars et al., 1975; Hartsuiker et al., 2005). Because the context effect in the present study is not affected by the lexical status of the competitor, it is hard to discern the operational criteria of a self-monitor, even one which has access to the intended speech plan (cf. Nooteboom, 2005a, 2005b).

5.4.3 Conclusions

In summary, our findings strongly implicate feedback from the phonological to the lexical levels, given a model of speech production which includes cascading activation between levels. There is also some evidence that participants are repairing slips of the tongue, and that these repairs are affected by context. A possible cause of such repairs is a self-monitor, although the details of its operation remain unclear. It is important to note that the potential presence of a monitor in production does not preclude feedback. Hartsuiker et al. (2005) and Nooteboom and Quené (in press) have proposed models of production that incorporate both feedback and self-monitoring. In such a model it may be the case that monitoring is manifested through the detection of aberrant activation within a cascading framework (e.g., Vigliocco & Hartsuiker, 2002).

One implication for models of production that needs to be addressed is the nature of representations that are competing. The cascading and feedback account of the present data was discussed in terms of competing phonological representations: variability results from partial activation of both a target and competitor phoneme. However, non-canonical articulations of this form could also be accounted for through activation of additional sub-phonemic units such as features. It is not possible to discriminate between these accounts in the present analysis. In Chapters 6 and 7 two further articulation investigations are reported to discriminate between phoneme and feature-based accounts of non-canonical articulation. To anticipate the findings, the results of both chapters demonstrate that variability can only be accounted for in a model that allows partial activation of competing phonemes.

5.5 Chapter Summary

Previous instrumental investigations of the acoustic and articulatory properties of speech errors have relied on categorising responses as 'errors'. However, in a cascading model of production articulation must be investigated in a manner that does not rely on categorisation. In an analysis of articulatory variability we demonstrated that responses are articulated more similarly to a competing phonological representation when that representation yields a real word. This finding is consistent with a cascading model of production that has feedback between phonological and lexical representations. We also observed context influences on articulation that

may implicate a role for monitoring in models of speech production, but the details are not clear.

Chapter 6

The Role of Features in a Cascaded Model I: EPG & VOT Evidence

6.1 Chapter Overview

The work presented in this thesis so far has been based on the assumption that phonological representations can cascade to articulation during speech planning. In this chapter we refine this assumption by investigating the role of phonological features in models of speech production. The focus is on whether feature representations are also required in models of production and, if so, whether there is feedback from there to phonological representations. We report an articulatory EPG and acoustic analysis of a tongue-twister experiment designed to investigate these issues.

6.2 Introduction

So far in this thesis we have assumed that phonological representations can cascade to articulation. A consequence of this cascading processes is that articulation may contain both alveolar and velar contact when alveolar and velar phonological representations compete during planning. We demonstrated in Chapter 3 that a high proportion of articulations contain both alveolar and velar tongue contact. In Chapter 5 we performed a quantitative analysis of articulatory variability which demonstrated that when competition yields a real word, tongue contact is more similar to the competing phonological representation than the target phonological representation. These results are consistent with a model of production with feedback between phonological and lexical representations and allows cascading from competing phonological representations to articulation. However, it is possible that lower level representations are required to account for the simultaneous articulation

of competing phonological representations. In other words, simultaneous articulation could result from the cascading of phonological representations or from the cascading of lower level representations such as features.

Some researchers have interpreted the occurrence of non-canonical errors as evidence for lower level representations in planning (Goldstein et al., 2007; Mowrey & MacKay, 1990; Pouplier, 2003, 2007). In an electromagnetic midsagittal articulometry (EMMA) investigation Goldstein et al. (2007, see also Pouplier, 2003) observed a high rate of articulations that contained simultaneous tongue-tip and tongue-dorsum raising during production. Their account for this observation is based on a gestural model of production. According to the gestural model, speech is planned by creating a gestural constellation which comprises several smaller units called gestures. Gestures are dynamic units that specify spatio-temporal motor commands for articulation (see Browman & Goldstein, 1989, for a detailed description of the model). In the model, two gestures can be simultaneously activated and therefore articulated due to the spatio-temporal dynamics of the system. However, two segmental units (e.g., gestural constellations) can not be coactivated. Critically, in order to account for non-canonical errors in this framework, lower level representations must be present in the model.

Mowrey and MacKay (1990) also argued for a lower level of representations. In an electromyography (EMG) investigation they observed abnormal orbicularis oris muscle activity that could likely be associated with an errorful addition of the labial+ feature in /s/ production. However, because EMG measurement was restricted to only one muscular point, they argue that their observation of abnormal muscle activity cannot be interpreted as "full" feature activation (such as labial+). Instead they interpret their observation as evidence for an even lower level of sub-featural representations (one muscle movement associated with labial+). Whether levels of representation extend as low as sub-featural activation remains an open question, but importantly it presupposes the existence of a featural level of representation.

While the investigations of Goldstein et al. (2007, see also Pouplier, 2003) and Mowrey and MacKay (1990) suggest that lower level representations may be required, their interpretations are based on the assumption that speech planning is staged. More specifically, gestures are included in the account of Goldstein et al. (2007) because phonological representations (or in their terminology "gestural constellations") can not be simultaneously activated. However, in this thesis we have

assumed a cascaded framework that allows the simultaneous activation and articulation of competing representations. This raises an important question: *If phonological representations can cascade to articulation, are lower level representations required?*

In this chapter and in Chapter 7 we investigate whether lower level representations should be included in a cascading model of speech production. We focus on whether *features* are required and we do not discriminate between feature and gesture representations. For the purposes of addressing our research question, the distinction is not important since both types of representations are at a lower level than phonological representations. We chose a feature approach because this is the lower level unit that has been more widely discussed in the psychological literature. Future research may address this distinction.

In the remainder of this chapter we present evidence that suggests features are required in models of production. Primary evidence in support of a role for features comes from investigations on the phonological similarity effect: similar phonological representations are more likely to interact than dissimilar phonological representations. An important consequence of incorporating features is the possibility of feedback interaction between feature and phonological representations. Hence, two mutually dependent issues are under question: are features required, and if so, is feedback required between representations? We report an articulatory and acoustic study designed to investigate these questions. We conclude that not only are features required, but there must also be feedback between featural and phonological representations.

6.2.1 The Role of Features in Models of Production

A primary source of evidence for the role of features in speech production is the phonological similarity effect: similar phonological representations are more likely to interact with one another than dissimilar phonological representations. The phonological similarity effect has been demonstrated in a variety of cognitive tasks, including working memory tasks (Baddeley, 1966) and picture naming (e.g., Bock, 1986), but our discussion will focus on speech errors. Phonological speech errors are more likely to occur if the target and competitor are phonologically similar compared to dissimilar (Dell & Reich, 1981; del Viso et al., 1991; Butterworth & Whittaker, 1980; Kupin, 1982; Levitt & Healy, 1985; MacKay, 1970; MacKay, 1980; Nooteboom, 2005a, 2005b; Shattuck-Hufnagel, 1986; Shattuck-Hufnagel & Klatt,

1979; Stemberger, 1982, 1985a; Vousden et al., 2000; Wilshire, 1999). Phonological similarity can be defined in several ways (see Frisch, 1996, for a discussion), but the most common way of defining phonological similarity is by counting the number of features that two phonemes share or do not share. By this definition, /k/ and /g/ only differ by one feature (place of articulation: alveolar *vs.* velar) and are therefore more similar than /k/ and /d/ which differ by two features (place of articulation as above; voicing: voice+ *vs.* voice–).

Several corpus analyses have established evidence for the phonological similarity effect (Dell & Reich, 1981; del Viso et al., 1991; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1982). In a corpus analysis of self-recorded speech errors Stemberger (1982) observed a high rate of single feature substitutions in contextual errors (e.g., *big pocket* → "pig pocket") and non-contextual errors (e.g., *at least* → "at weast"). Similarly, Shattuck-Hufnagel and Klatt (1979) carried out a confusion matrix analysis of the 1977 MIT Corpus and observed that errors often only differed by one distinctive feature. Phonological similarity effects have also been demonstrated in languages other than English. In an analysis of a German speech error corpus, MacKay (1970) demonstrated clear evidence for a phonological similarity effect on speech errors; over 55% of substituted phonemes differed by one distinctive feature, while less than 5% differed by four distinctive features. Similarly, a phonological similarity effect has been reported for Spanish speech errors (del Viso et al., 1991).

Effects of phonological similarity have also been demonstrated experimentally in tongue-twister tasks (Butterworth & Whittaker, 1980; Kupin, 1982; Levitt & Healy, 1985; Wilshire, 1999). In one tongue-twister experiment by Levitt and Healy (1985) it was demonstrated that stimulus items that differing by only one feature yielded more substitution errors than items differing by more than one feature. However, tongue-twister errors are often recorded during very rapid repetitions of stimuli (e.g., 180-210 syllables/minute: Kupin, 1982). This has raised criticism over whether tongue-twisters elicit errors of phonological encoding or errors of articulatory execution (cf. Laver, 1980). Importantly, Wilshire (1999) provided additional evidence for phonological similarity effects in a tongue-twister experiment with a rate of 100 syllables/minute, which is much slower than estimates of spontaneous speech rates. She observed that tongue-twisters with similar onsets (e.g., *cape turf tip calf*) yielded a higher error rate than those with less similar onsets (e.g., *cough dot deaf kit*). This finding rules out the suggestion that phonological similarity is only a factor in rapid repetition.

In order for models of speech production to account for the phonological similarity effect, they must incorporate some way for similar phonological representations to interact with one another more than dissimilar phonological representations. The most straightforward way to achieve this is to include a lower level of representation for features (Dell, 1986; Stemberger, 1982, 1985a). Consider, for example, Dell's (1986) model of production. According to this model there are both phonological and featural representations, but phonological encoding is completed when a phonological representation achieves sufficient activation. In Dell's (1986) description of the model he identifies a "featural paradox": the units that 'slip' in speech errors are phonological units (based on the traditional view of speech errors, see Section 2.3 for details), but features are still required to account for phonological similarity effects. According to Dell (1986, p. 294), the phonological similarity effect provides evidence that "features seem to be exerting their influence as units, but this is not revealed by the features themselves slipping ... clearly, a model of phonological encoding must assign roles to the various linguistic units so as to account for the [featural paradox]".

A consequence of including features and having phonological representation output is that there must be feedback for the lower level activation to influence higher level interaction. For example, activation from competing phonological representations (e.g., /k/ and /t/) cascades to activate their corresponding features (e.g., <voice–, velar+>, <voice–, alveolar+>), and then featural activation feeds back to reinforce the activation of phonological representations. Activation of phonological competitors is more reinforced when there is a greater proportion of features in common (as above) compared to when there is a lesser proportion of features in common (e.g., /k/ and /d/ would activate <voice–, velar+>, <voice+, alveolar+>).

Taken together, there is a wealth of evidence for the phonological similarity effect in speech errors. This evidence has been instantiated by both corpus-based and experimental analyses of speech error distributions. The most straightforward account for phonological similarity is to incorporate features into models of production and to have feedback between the feature and phonological representations. However, the feedback account is based on a staged account of production. This is because responses have been categorised as "errors" and it has been assumed that planning is based on "winner takes all" selection by which either a phonological representation is selected or not selected.

In contrast, throughout this thesis we have proposed that partially activated representations can cascade to articulation. If representations can cascade and responses

are not categorised as errors it is important to evaluate how a cascading model of production can account for phonological similarity. As stated earlier, there are two mutually dependent issues under question: are features required and if so, is there feedback between featural and phonological representation?

In order to investigate phonological similarity we present a tongue-twister experiment designed to investigate phonological similarity in a cascading model. We measure articulatory (using EPG) and acoustic variability (using VOT) and compare responses in a control condition (e.g., repeating *kef kef kef kef*) relative to responses in conditions of differing phonological similarity (e.g., *kef tef tef kef vs. kef def def kef*).

To illustrate how a cascading model of production may account for phonological similarity we present four different models in Figure 6.1. In each model the target phonological representation is /t/ and it is competing with another phonological representation. The models on the left of the figure represent competition between phonological representations that differ by one feature (/t/, /k/) and the models on the right represent competition of phonological representations that differ by two features (/t/, /g/). The rules of the models are simple: they are all interactive above the phonological level (see evidence in Chapters 3 and 5), there can be strong activation (solid lines) or weak activation (dotted lines), activation that is initially weak can be reinforced and become strong (dotted line within a solid line), and only strong activation can cascade to articulation. The cascading of two competing representations to articulation increases variability.

In Figure 6.1(A) the model only includes phonological representations which feedforward to articulation. This model clearly cannot account for phonological similarity because there is no reason why /k/ would be more likely to cascade to articulation than /g/. The model presented in Figure 6.1(B) includes featural representations. These are the units that cascade to articulation. This model is also unable to account for phonological similarity because there is no activation that makes features <velar+, voice–>, activated by /k/, more likely to cascade than the features <velar+, voice+> activated by /g/. Since models A and B can not account for phonological similarity, we would predict that variability of articulation would not differ across conditions in which phonological representations differ by one or two features.

The models presented in Figure 6.1(C) and (D) both have feedback from feature representations to phonological representations. Both of these models can account

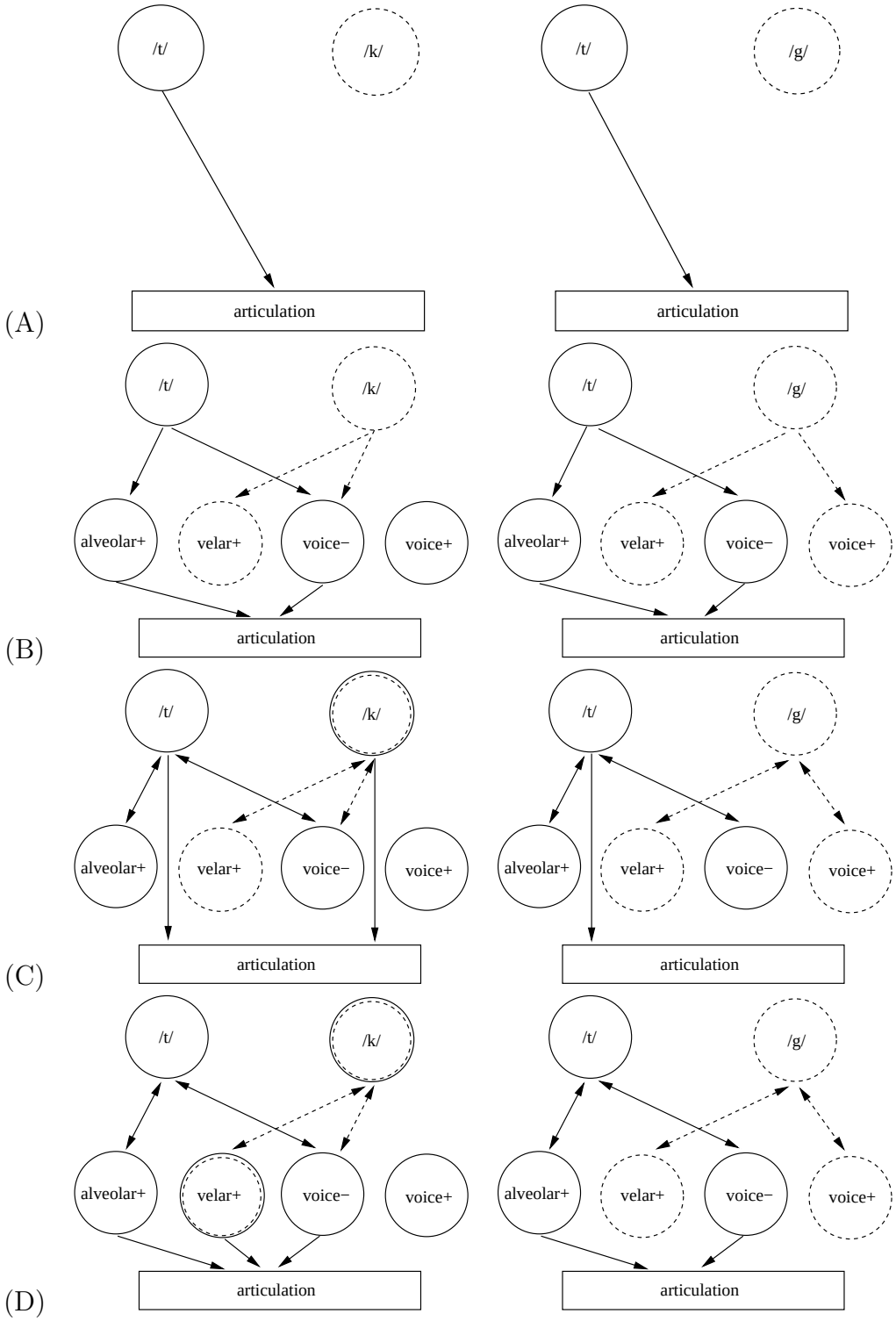


Figure 6.1: Different models to account for phonological similarity effects in articulation; solid lines represent strong activation, dotted lines represent weak activation, and a dotted line within a solid line represents initially weak activation that was reinforced and became strong activation. Only strong activation can cascade to articulation. (A) a feedforward model with phonological representation action units without feature representations; (B) a feedforward model with feature unit output; (C) a feedback model with phonological representation action units and feature representations; (D) a feedback model with feature unit output. Only models (C) and (D) can account for phonological similarity.

for phonological similarity. For example, in (C) activation from /k/ flows to activate the <voice–> representation. Because <voice–> is also receives activation from /t/ it feeds back to reinforce the activation of /k/. Due to reinforcement, both /t/ and /k/ cascade to articulation. However, in the case of /t/-/g/ competition none of the feature representations receive enough activation to feedback and reinforce /g/ activation. Therefore, only /t/ cascades to articulation. The model in (D) makes the same predictions except, because the output is at the feature level of representation, activation must flow from /k/ to <voice–>, feed back to reinforce /k/, and then feed forward again to reinforce the <velar+> representation. Since competitors are more likely to cascade to articulation when similar phonological representations are competing we predict that articulation should be more variable for similar phonological representations than dissimilar phonological representations.

These example models suggest that feature level representations are required to account for phonological similarity and that there must be feedback between the feature and phonological representations. Interestingly, models C and D make the same predictions independent of whether the output unit of the model to articulation is a phonological representation or a feature representation. We return to this point in the General Discussion.

6.3 Experiment 3: EPG Tongue-Twisters

Tongue-twister tasks are often used to elicit speech errors in the laboratory (Butterworth & Whittaker, 1980; Dell & Repka, 1992; Kupin, 1982; Levitt & Healy, 1985; MacKay, 1982; Schwartz, Saffran, Bloch, & Dell, 1994; Shattuck-Hufnagel, 1983, 1987, 1992; Wilshire, 1999). However, this methodology has been criticised for yielding errors that result from factors outwith speech planning; primarily researchers have argued that errors may be due to articulatory implementation or short-term memory difficulty (e.g. Laver, 1980). In a seminal tongue-twister study Wilshire (1999) established that the errors produced during tongue-twisters can be attributed to phonological encoding. First, errors were elicited during a slower than spontaneous speaking rate, which demonstrates that errors can not be solely attributed to speaking at rapid rates. Second, errors were elicited while participants read stimulus materials, rather than reciting them from memory, which weakens the contention that errors have their origin in short-term memory limitations. This is not to say that speaking rate (Kupin, 1982; MacKay, 1982) and short-term memory (McCutchen, Bell, France, & Perfetti, 1991; Saito & Baddeley, 2004) do not influence speech error production, but rather that tongue-twister errors cannot be

attributed solely to these factors. Finally, Wilshire demonstrated that the errors produced in tongue-twister repetitions have the same properties and adhere to the same rules as errors produced in spontaneous speech.

Following Wilshire (1999), we use a tongue-twister design in the present experiment. Given the limitations of articulatory data collection, tongue-twister tasks have a clear advantage over other methodologies; a high number of observations can be recorded in a short period of time. In a similar recording time we collect 1024 observations per participant in the present experiment, compared to 96 observations per participant in the WOC task (Chapters 3 and 5).

This experiment innovates from previous investigations by including analyses of both articulation and acoustic variability of the same recordings. All previous investigations that have focused on non-canonical errors have only reported one dependent measure based either on articulation (Goldstein et al., 2007; Mowrey & MacKay, 1990; Pouplier, 2003, 2007; Frisch, 2007; Stearns, 2006) or acoustics (Frisch & Wright, 2002; Goldrick & Blumstein, 2006). In this experiment, articulation is measured using electropalatography (EPG), which records tongue-to-palate contact over time. Our analysis of the acoustic signal is based on voice onset time (VOT), a robust measure of voicing for initial-onset stop consonants (Lisker & Abramson, 1964).

This investigation also innovates by including materials that differ in phonological similarity. Previous investigations have typically only included stimulus items that differ by a single feature. Because these investigations have only focused on one factor, either articulation or acoustics, the stimulus items used have only differed by that factor. For example, acoustic studies have only included materials that differ in voicing (e.g., /k/-/g/) and articulatory studies have only included materials that differ in place or manner (e.g., /k/-/t/). Therefore most stimulus items from previous investigations have been highly phonologically similar.

Importantly, by measuring both articulation and the acoustic signal we can independently investigate how each feature (e.g., place of articulation and voice) is affected by phonological similarity. For example, we test how variability in articulation is affected by a competing place of articulation representation (e.g., *kef tef tef kef*) compared to both a competing place of articulation and a voicing representation (e.g., *kef def def kef*). Likewise, we investigate how variability in VOT is affected by a competing voice representation (e.g., *kef gef gef kef*) compared to when both voice and place of articulation are competing (e.g., *kef def def kef*). An observation

of a phonological similarity effect for each dependent measure will provide strong evidence for feedback between featural and phonological representations.

6.3.1 Method

Participants

Seven native speakers of English from the Queen Margaret University research community participated in the experiment. All participants were experienced in speaking while wearing an EPG artificial palate and reported no speech or hearing impairments. All participants were treated in accordance with the Queen Margaret University and University of Edinburgh ethical guidelines. Two speakers were excluded from the analysis due to missing data as a result of technical failure during recording.

Materials

The tongue-twisters were designed to include Place of Articulation and/or Voicing contrasts. The targets were designed to ensure that onset consonants would yield firm tongue contact with the EPG artificial palate at word onset. Vowels and coda consonants were selected to minimise the amount of EPG contact following each onset. A set of tongue-twisters was generated which contained alternating combinations of two onsets (/k/, /g/, /t/, /d/). Each pair of onset combinations was also assigned a vowel (/ɪ/, /e/, /a/). Vowels were selected to control for forward transitional probabilities across conditions. This resulted in 15 ABBA sequences (e.g., /k.../ /t.../ /t.../ /k.../) which were inverted to create an additional 15 BAAB sequences (e.g., /t.../ /k.../ /k.../ /t.../). Each of these 30 sequences was then paired with an /f/ and a /v/ final consonant to yield 60 tongue-twisters such as /gef/ /kef/ /kef/ /gef/. The labio-dental final consonants were selected to balance the occurrence of a voice± coda and to make acoustic segmentation between each syllable easier. An additional four control tongue-twisters were generated to contain a non-alternating repetition of each onset together with a randomly assigned vowel and final consonant (e.g., /kef/ /kef/ /kef/ /kef/). Refer to Appendix B for a list of all 64 tongue-twisters.

Apparatus

The experiment took place in a sound-treated recording studio at Queen Margaret University. Prior to testing, each participant was fitted with a custom elec-

tropalatography (EPG) palate (manufactured by Incidental, Newbury, UK or Grove Orthodontics, Norfolk, UK) moulded to fit a dental cast from an impression of the hard palate. The EPG palate is made of acrylic and contains 62 embedded silver contacts on the lingual surface of the hard palate, organised in eight rows of eight contacts (except only six contacts in the most anterior row). EPG data was recorded at rate of 100Hz using the WinEPG system (Articulate Instruments Ltd, Edinburgh, UK), which connected the palate to a multiplexer unit that transferred the data to an EPG3 scanner and then to the serial port of a desktop computer. The acoustic signal of participants' responses were recorded on to one auditory channel at 22,050Hz using an Audio Technica ATM10a microphone. A desktop computer, to which the microphone and WinEPG system were attached, was used to record participants' responses with Articulate Assistant (Wrench, 2003) software. Stimuli were presented on a 15" LCD monitor using the prompt function of Articulate Assistant.

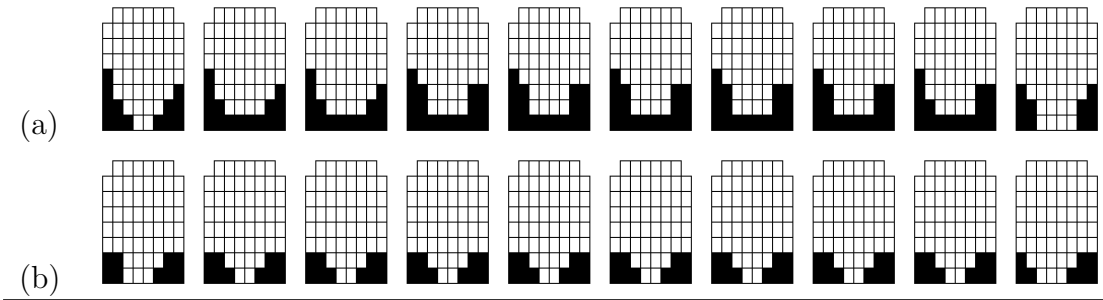
To control speaking rate, participants were presented with an auditory beat at a rate of 150 beats per minute using metronome software on a laptop computer. The metronome signal was fed to a set of mono headphones (worn on participants' preferred ear) and to a direct audio line into the computer running Articulate Assistant. The latter was recorded onto a second auditory channel at a 22,050Hz sampling rate.

Procedure

Once participants were seated, they were instructed to read out loud each experimental item once. This was done because stimulus items were presented orthographically and we wanted to make sure that participants used the correct vowel. Participants were given feedback about their pronunciation and, if necessary, asked to repeat each item until it was pronounced correctly.

After the practise session, each tongue-twister was presented individually on the screen and participants were instructed to repeat each phrase four times. Participants were additionally instructed to speak at a rate of one item per metronome beat. Following each sequence the experimenter advanced to the next sequence with a short pause (approx 3s). Participants were allowed to take a longer break by notifying the experimenter. The first four tongue-twisters were the control sequences (e.g., *kef kef kef kef*). These were followed by the 60 experimental sequences, presented in random order.

Table 6.1: A sample of EPG recordings that have been trimmed to only include full closure: (a) a velar articulatory record trimmed to only include the palate before full closure through to the palate after the full closure release; (b) a velar articulatory record which does not contain visible full closure and therefore was trimmed to include the palate before maximal closure through to the palate after maximal closure.



Data Treatment

Following the experiment, we performed measurements on both the EPG and acoustic recordings. We treat each word initial onset as an item for all measurements. The only items excluded were those items not collected due to technical failure of the recording equipment (83 items out of 5120 possible responses).

EPG Data Each recorded onset was identified in Praat (Boersma & Weenink, 2006) using the acoustic signal. The key time points identified were the offset of the previous item (or for the first item, a time point 150ms prior to the onset release) and the onset of the vowel of the target item. The EPG record for this duration was extracted for preliminary inspection. Each articulatory record was then trimmed to include the first palate before full closure through to the first palate after the full closure release. Full closure was defined by any continuous path across the lateral axis of the palate. In some cases velar closure did not include a continuous path across the posterior row of contacts. These items were trimmed to include the palate before the maximal closure to the palate following the maximal closure. See Table 6.3.1 for two sample closures: (a) a trimmed velar item with full closure; (b) a trimmed velar item without full closure.

Once the EPG records were extracted and trimmed they were analysed using the Delta method described in Chapter 4. To calculate tongue-to-palate contact variability all EPG recordings were standardised so that each record contained ten frames. Then a mean reference articulation was created for each onset phoneme (/k/, /g/, /t/, /d/) for each speaker. The tokens used to create the references

were from the control tongue-twister sequences (e.g., *kef kef kef kef*) which did not contain any competing phonemes. We then calculated Delta difference scores for each recorded token compared to the relevant mean reference (refer to Chapter 4 for details on the Delta calculation). The higher the distance score, the greater the variability of the target from the mean reference.

Acoustic Data The VOT for each target item was measured from the acoustic signal with Praat (Boersma & Weenink, 2006). This was defined as the duration (in milliseconds) between the acoustic burst of the onset to the onset of the periodicity associated with the following vowel. A mean reference VOT was created for each speaker by calculating the average VOT for each target onset (/d/,/g/,/k/,/t/ from the control tongue-twisters. Since we are interested in the amount of variability resulting from competition, we calculated a VOT difference score for each item. The difference scores were calculated as the absolute value of the target VOT minus the relevant mean reference VOT.

6.3.2 Experiment 3: Results

To investigate the influence of phonological similarity on production we separately analysed the EPG and VOT data using Generalised Linear Mixed-Effects models, with the lme4 (Bates & Sarkar, 2007) and languageR (Baayen, in press) packages in R. Both analyses include every recorded observation and include Place (change, no change) and Voice (change, no change) as fixed effect factors and item and participant as random effect factors. Each tongue-twister sequence was treated as an independent experimental item. The fixed effect factors were centred to reduce multi-collinearity in an unbalanced design (Landsheera, van den Wittenboerb, & Maassena, 2006). The model includes the interaction of Place × Voice because this was the primary contrast of theoretical interest.

The *t*-values for each contrast are reported along with probabilities based on 10,000 Markov Chain Monte Carlo (MCMC) samples. Using an MCMC probability estimate has been suggested because it can be difficult to calculate accurate degrees of freedom corresponding to each *t*-value for the intercepts (Bates & Sarkar, 2007). All reported mean values are based on MCMC mean estimates. Confidence intervals were calculated for each intercept, calculated in the 2 × 2 design by modelling each factor separately. The confidence intervals correspond to the effect size within each of the respective models.

An important distinction between Linear Mixed-Effects models and the more traditional ANOVA should be noted. In an ANOVA analysis the mean and variance within each experimental condition are compared relative to one another. However, a Linear Mixed-Effects model is based on whether the variance contributed by one condition is greater than the variance contributed by another condition. For example, in the current experiment a main effect for Place (change, no change) indicates that a change in place of articulation has an influence on the dependent variable (e.g., articulatory variability). Likewise, a significant Place × Voice interaction indicates that a change in both place of articulation and voicing significantly influences the dependent variable. For analyses of the present experiment, a significant positive effect for one factor (e.g., Place) plus a significant *negative* effect for the combination of factors (Place × Voice) can be interpreted as the equivalent of a traditional ANOVA interaction in which a change in Place causes greater articulatory variability than changes in Place and Voice combined.

Articulation Analysis (EPG)

The articulation analysis included 5037 EPG observations for five speakers and 64 items. A mixed effects analysis revealed a significant main effect for Place Change [t=7.06, p<0.0001] in which articulation was more varied when place of articulation changed (2.28) compared to no change (1.77; 95%CI ±0.15). We also observed a main effect for Voice Change [t=2.29, p<0.02] in which articulation was more variable when there was a change in voice (1.91) compared to no change (1.77; 95%CI ±0.25). Lastly, there was a significant Place Change × Voice Change interaction [t=5.03; p<0.0001]: Articulation was more variable when only place changed (2.28) compared to when both voice and place changed (1.65; 95%CI ±0.31). Refer to Figure 6.2 for MCMC mean estimates.

Acoustic Analysis (VOT)

The acoustic analysis included 5037 VOT observations for five speakers and 64 items. A mixed effects analysis revealed a significant main effect of Voice Change [t=2.18, p<0.05] in which VOT variability was greater when there was a change in voice (11.99ms) compared to no change (10.84ms; 95%CI ±1.04ms). There was also a significant Voice Change × Place Change interaction [t=2.02, p<0.05]: variability in VOT was greatest when only voice changed (11.99ms) compared to when both voice and place of articulation changed (9.35ms; 95%CI ±2.59ms). The main effect

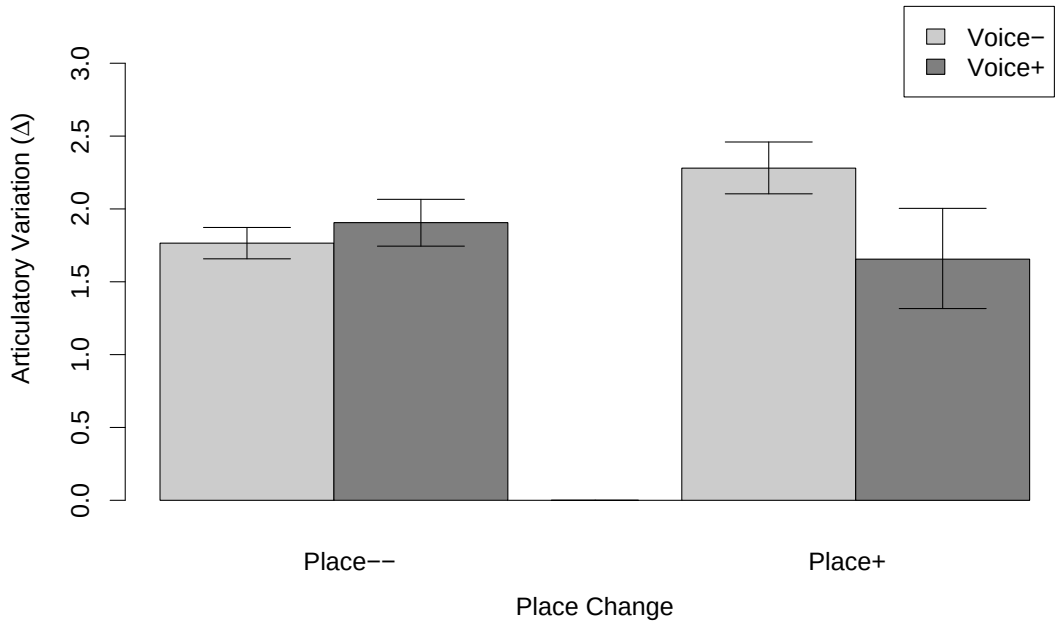


Figure 6.2: MCMC mean estimates for articulatory variation (Δ) in the EPG analysis of tongue-twisters from Experiment 3 with 95% confidence intervals.

for Place Change was not significant [t<1]. Refer to Figure 6.3 for MCMC mean estimates.

6.3.3 Experiment 3: Discussion

The independent analyses of articulation and the acoustic signal establish experimental evidence for phonological similarity. For the articulation analysis of EPG recordings, we compared the articulation of phonemes spoken when they were competing with similar representations (e.g., /k/–/t/) or dissimilar representations (e.g., /k/–/d/) relative to when they were spoken without competing representations (e.g., /k/–/k/). This analysis revealed that articulation was more variable when only place of articulation was competing (e.g., /k/–/t/) compared to when place of articulation and voice were competing (e.g., /k/–/d/). This finding establishes evidence that articulation is more variable when phonologically similar representations are competing.

In a similar analysis we compared the VOT of phonemes spoken in contexts with competing representations relative to the same phonemes when they were spoken

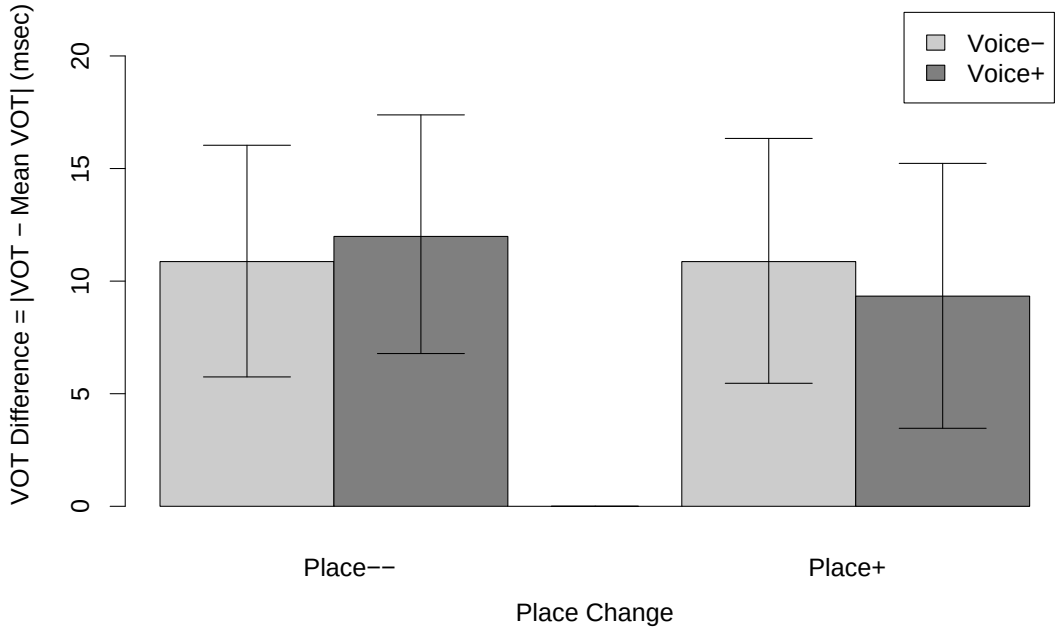


Figure 6.3: MCMC mean estimates for VOT variation (msec) in the acoustic analysis of tongue-twisters from Experiment 3 with 95% confidence intervals.

without competing representations. This acoustic analysis revealed that there was more variability in VOT when there was competition for voicing (e.g., /k/–/g/) compared to when there was competition for both voicing and place of articulation e.g., /k/–/d/). This result provides additional evidence for phonological similarity: VOT is most variable when phonemes are competing with similar phonological representations.

These findings are important for two reasons. First, we established evidence for phonological similarity without categorising responses. Second, we provided evidence for phonological similarity using two different dependent measures of speech: variability in articulation and VOT.

6.4 General Discussion

There are two sources of evidence in the literature that implicate a role for features in speech production. First, speech error investigations have demonstrated evidence for the phonological similarity effect: substitution errors are more likely

when similar phonological representations, rather than dissimilar representations, are competing (Dell & Reich, 1981; del Viso et al., 1991; Butterworth & Whittaker, 1980; Kupin, 1982; Levitt & Healy, 1985; MacKay, 1970; MacKay, 1980; Nooteboom, 2005a, 2005b; Shattuck-Hufnagel, 1986; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1982, 1985a; Vousden et al., 2000; Wilshire, 1999). Second, articulatory investigations of speech errors have suggested lower level representations may be required to account for the simultaneous articulation of two competing representations (Goldstein et al., 2007; Mowrey & MacKay, 1990). However, both sources of evidence have relied on categorising responses as errors. A consequence of categorising responses is that a staged planning process must be assumed.

Throughout this thesis we have assumed that, unlike staged processing, partially activated representations can cascade to articulation. Allowing partially activated representations to cascade raises an issue of whether feature representations are still required. Importantly, the results of the present study implicate a role for features in a cascading model of production. Without categorising responses we demonstrated evidence for phonological similarity. In analyses of the articulatory and acoustic records we demonstrated that variability is greatest when phonological representations differ by only one relevant feature. Importantly this evidence was established from two independent measures of variability.

The most straightforward way to account for phonological similarity is to incorporate feedback from featural to phonological levels of representation (Dell, 1986; Stemberger, 1982, 1985a). According to this account activation from phonological representations flows to activate featural representations, which in turn feeds back to reinforce the phonological representation. Competing phonological representations will receive more reinforcement if they share feature representations and less reinforcement if they do not share feature representations. Reinforcement increases the activation level of competitors which leads to the cascading of strongly activated representations to articulation.

It is difficult for a model without feedback to account for phonological similarity effects. More generally, it is not clear how a lower level of representations can influence a higher level of representation without a bottom-up flow of information. One potential way in which this influence could be accomplished is through a monitoring loop. Monitoring mechanisms such as the Perceptual Loop Theory propose that speakers monitor their speech plans using the comprehension system (Levelt, 1989; Levelt et al., 1999). This is accomplished because the speech plan is represented in phonological units that can be parsed by the comprehension system. For example,

monitoring accounts of the lexical bias effect propose that the perceptual system can detect nonwords and edit the speech plan if necessary. However, this type of monitor cannot account for phonological similarity: an error such as *meat puppets* → "peat muppets" yields real word outcomes so the perceptual system would not detect an error unless it had access to the intended utterance. To account for phonological similarity in a monitoring-based model Nooteboom (2005a, 2005b) proposed that speakers do have access to the intended utterance. According to this account the phonological similarity effect results from speakers being less likely to detect an error in a speech plan that sounded similar to the intended plan compared to a speech plan that sounded different. However, it is not clear how such a monitor would be implemented. How can a correct intended speech plan be generated for comparison and, if it is possible, why is an incorrect plan generated? Therefore, a monitoring account of phonological similarity seems implausible.

While feature representations and feedback are required to account for phonological similarity, it is not clear if the phonological or featural representations cascade to articulation. Goldstein et al. (2007) and Mowrey and MacKay (1990) posited that lower level representations are the units of planning that flow to articulation. However, their accounts did not allow partial activation of competing representations. In the cascading models presented in Figure 6.1, the same pattern of results are predicted independent of whether featural or phonological representations cascade as long as feature representations are included in the model and there is feedback between representations. Future research is required to investigate the nature of the cascading unit that flows to articulation.

A source of caution from the present investigation is that analyses were restricted to stop consonants differing, at most, by two features. There is some evidence that some features are qualitatively different from other features (Shattuck-Hufnagel & Klatt, 1979). Shattuck-Hufnagel and Klatt (1979) demonstrated in an analysis of a confusion matrix of speech errors that error patterns are asymmetric: /s/ → /š/ (68 occurrences) errors are more frequent than /š/ → /z/ (33 occurrences) errors in the 1977 MIT speech error corpus. Similarly, much of the work by Stemberger and colleagues (1991a, 1991b; 1986; 1991) has argued for asymmetries in speech production. Asymmetries in speech errors are most often attributed to differences in the (under)specification of certain features (e.g., coronal is an under-specified feature) or differences in individual feature properties such as frequency (Stemberger, 1991a, 1991b; Stemberger & Treiman, 1986; Stemberger & Stoel-Gammon, 1991).

Our results do not address asymmetric patterns because our analysis were counterbalanced to include, for example, all /k/–/t/ and /t/–/k/ productions within the same cell. Future research is also required to determine if phonological similarity influences on variability extend to additional types of feature representations. To address whether different types of features equally influence variability, we report an additional phonological similarity experiment in Chapter 7. The experiment investigates phonological similarity in articulation across place of articulation (velar and alveolar) and manner (plosive and fricative) features.

In summary, the study reported in this chapter establishes a role for features in cascading models of speech production. Models must also allow feedback between feature and phonological representations. Future work is required to determine whether the phonological or featural representations cascade to articulation and to determine whether all features exert the same influence on phonological similarity.

6.5 Chapter Summary

This chapter provided an investigation on whether features are required in models of production that have cascading activation. We presented an EPG and acoustic analysis of phonological similarity influences on speech production. The analyses revealed that articulatory and acoustic variability was greatest when only one feature was competing as opposed to when two features were competing. This pattern is consistent with a cascading model of speech production that includes features and allows feedback between feature and phonological representations.

Chapter 7

The Role of Features in a Cascaded Model II: Ultrasound & VOT Evidence

7.1 Chapter Overview

This chapter extends the articulation investigations of previous chapters by using a different articulatory imaging technique. In this chapter we use ultrasound to measure tongue movement during articulation. We present an overview of ultrasound imaging and argue that previous analysis methods are limited. We then present a demonstration of how the Delta method can be used for ultrasound analysis. Finally we present an experiment designed to replicate the phonological similarity evidence reported in Chapter 6. We conclude that features are required for models of production, there must be feedback between featural and phonological representations, and that the Delta method is a useful tool for articulatory analyses.

7.2 Introduction

The articulatory investigations in previous chapters have suggested that feedback should be incorporated into a cascading model of speech production. In Chapter 5 we demonstrated that tongue-to-palate contact is more similar to a competing phonological representation when that representation yields a real word. This finding can best be accounted for by feedback from phonological representations to lexical representations. In Chapter 6 we observed that tongue contact variability is greatest when phonological representations have similar competitors compared to when they have dissimilar competitors or no competitors at all. A feedback account that includes a flow of information from feature representations to phonological representations provides the most straightforward account of the observed articulatory variability.

All of the articulation evidence presented so far has been based on electropalatographic (EPG) recordings of tongue-to-palate contact. One limitation of using EPG is that articulatory measurements are restricted to tongue movements that yield palatal contact. EPG does not allow the recording of partial movements towards an articulatory goal or any measurement of the shape of the tongue during articulation. Since measuring articulation during the production of speech errors is a young methodology, it is important to consider the methodological implications of choosing one articulatory imaging method over another (see Frisch, 2007, for a discussion).

In this chapter we extend our research to include articulatory analyses from a different articulatory imaging technique. Ultrasound imaging of the tongue provides a record of the full midsagittal contour of the tongue over time. The benefit of ultrasound over EPG is partial movement toward the palate can be observed and information about tongue shape can be recorded. This benefit is particularly important since previous investigations of speech errors have interpreted some articulatory movements as evidence for partial errors (Goldstein et al., 2007; Pouplier, 2003, 2007). Partial errors have been defined as tongue movements that contain properties of a competing representation, but differ from the canonical articulation of the competing and target representations. For example, in a study using electromagnetic midsagittal articulometry (EMMA) Goldstein et al. (2007) observed velar articulations with a tongue-tip height that was higher than typical velar articulation, but lower than typical alveolar articulations.

To investigate articulation measured with ultrasound we present an extension of the Delta method developed in Chapter 4. The Delta method was originally developed for EPG analysis, but can easily be adapted for ultrasound analysis. We first present a review of ultrasound imaging techniques and discuss the limitations of some commonly used analysis techniques. We then present a demonstration of how the Delta method is used for ultrasound and discuss how it benefits over other analysis methods.

Later in the chapter (Section 7.5) we present an experiment designed to further investigate the role of features in a cascading model of production. In Experiment 3 we observed, using EPG, that articulation was more variable when only place of articulation was competing in a tongue-twister compared to when both place of articulation and voicing features were competing. We also observed in Experiment 3 that acoustic variability in VOT is influenced by phonological similarity: VOT was more variable when only voicing was competing compared to when both

voicing and place of articulation were competing. Both of these findings provide experimental evidence for features and evidence for feedback between featural and phonological representations. An observation of a phonological similarity effect that replicates Experiment 3 using ultrasound will not only have theoretical implications for feedback between featural and phonological representations, but will also have an important methodological implication. A replication will demonstrate that the Delta method is a technique that can be used across different articulatory imaging methods.

7.3 Ultrasound for Articulation Analysis

Ultrasound, unlike EPG, is a medical imaging technique used for a variety of therapeutic and diagnostic functions. The most popular use of ultrasound is for foetal imaging, but it is also a useful method for imaging the tissue of most internal organs including the tongue. In this thesis we focus on the imaging of the midsagittal spline of the tongue, though ultrasound can also be used for coronal (Slud, Stone, Smith, & Goldstein, 2002) and three-dimensional tongue measurements (Lundberg & Stone, 1999; Stone & Lundberg, 1996). Henceforth, we use the general term *ultrasound* to refer to midsagittal tongue imaging.

The basic principle behind ultrasound is that a series of mechanical vibrations are converted into sound waves and are projected into an object (e.g., the mouth). The waves travel until they reach a barrier (e.g., the surface of the tongue) and then bounce back towards the original source of the sound waves. A barrier results from any change in density within the object being measured. In the case of oral ultrasound the sound wave encounters the tongue, which is surrounded by air or the bone of the hard palate, and because there is a change in density the wave bounces back to the source. Once the duration for each wave to travel to a barrier and back is calculated, it can be converted into a distance measure and represented in a visual display (see Hedrick, Hykes, & Starchman, 1995; Stone, 2005, for a detailed discussion on general ultrasound physics and oral ultrasound physics, respectively).

Figure 7.1 provides a sample midsagittal ultrasound image of the tongue. The upper surface of the tongue is represented by a bright whitish-coloured band across the middle of the image, with the tongue root on the far left and the tongue tip on the far right. Because the sound wave is reflected at the tongue's surface no other features above the tongue in the oral cavity, such as the palate, are displayed.

Figure 7.1: A single frame of an ultrasound recording of the midsagittal contour of the tongue. The upper surface of the tongue is represented by a whitish-coloured band that spans the width of the image. The tongue root is on the left and the tongue-tip on the right of the image.

A black shadow is present on the right side of the image because the sound wave reaches the mandible and hyoid bones before reaching the tongue (Stone, 2005).

During ultrasound imaging, the probe which emits the sound waves must be stabilised relative to the participant's head position. This can be accomplished in a variety of ways, from using dental chairs, specialised helmets, or a specialised head and transducer support system (see Stone, 2005). For example, the technique for head stabilisation used in the laboratory at Queen Margaret University makes use of a custom-designed helmet that holds the transducer in a vertical position centred beneath the chin. The use of this helmet ensures that all recordings collected during a session are comparable because all images are relative to the stationary probe position.

The primary advantage of ultrasound over other recording techniques is the entire midsagittal contour of the tongue can be measured continuously. EPG does not provide information about tongue shape; electromagnetic midsagittal articulometry (EMMA) only records predefined points on the articulators where transducers

are affixed. However, analysing the curve of the tongue remains a challenge for ultrasound research. In the following section we discuss different approaches that have been used for ultrasound analysis and discuss some of the limitations associated with them.

7.3.1 Ultrasound Analysis Methods

The first step in nearly all ultrasound analysis methods is to identify the contour of the tongue. This can be accomplished using a variety of methods. One method involves hand tracing the contour of the tongue for each video frame. Stone and colleagues (1983; 1988) have demonstrated that both inter-tracer and intra-tracer reliability are very high and tracing accuracy can have an error as low as a one pixel difference. These findings suggest that hand tracing the tongue contour is a reliable measure. A strong drawback of hand tracing is that it can be very time consuming. More recently algorithms have been developed for automatic tongue edge detection (Li, Kambhamettu, & Stone, 2005; Unser & Stone, 1992). These algorithms identify the brightest pixels in the ultrasound image and draw a spline over these pixels. However, the tongue may not always be the brightest feature in an ultrasound image due to noise and potential artefacts. Therefore, edge detection algorithms in practise are only semi-automatic and researchers are required to inspect and hand edit the detected splines.

Once the contours of the tongue have been identified there are a variety of ways in which the contours can be analysed (see Stone, 2005, for a review). The most basic analysis method is to calculate the height of the tongue contour at predefined points. This is accomplished by using a Cartesian coordinate system. Points of interest are identified on the x-axis of the ultrasound image and then the height of the tongue is calculated on the y-axis. This approach for analysing ultrasound data is limited for two reasons. First, the length of the tongue surface can vary across phoneme repetitions. Therefore, a height measurement from the same point on the x-axis may not always correspond with the same position on the tongue. This is especially a concern for the measurement of tongue-tip height since this is where the tongue lengthens the most during articulation. Second, an analysis that only accounts for height differences does not capture any detail about the overall tongue shape. Since the ability to measure tongue shape is the primary benefit of ultrasound, this approach is potentially self-defeating.

An analysis method that does measure the full contour of the tongue uses curve fitting or polynomial functions (Morrish, Stone, Stonies, Kurtz, & Shawker, 1984; Morrish, Stone, Shawker, & Sonies, 1985; Stone, 2005). Each contour of the tongue can be defined by an equation that represents the distributions of points on the tongue. Depending on the complexity of the tongue contour, the order of a polynomial equation can be as low as one, representing the tongue slope. Additional details may be required such as the degree of curvature or number of bends in the contour. Studies on curve-fitting for midsagittal tongue ultrasound have demonstrated that lower-order functions are sensitive to the length of the tongue (see Stone, 2005, for evidence of this from unpublished data). Given that tongue length changes during articulation the finding that curve fitting is sensitive to length is potentially problematic. On the other hand, higher-order polynomial functions yield better fits of the data, but can be much more difficult to interpret physiologically (Stone, 2005).

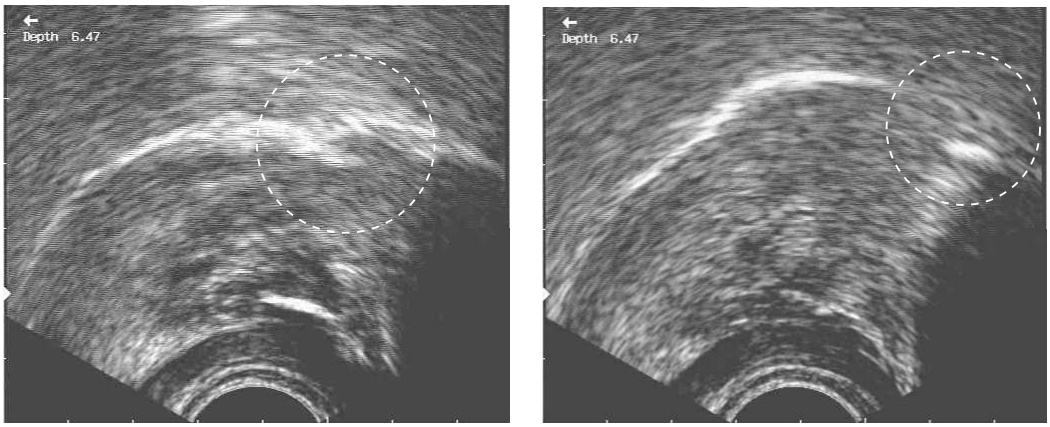
Another method for analysing tongue-contours is based on a principal components analysis (PCA) algorithm (Harshman, Ladefoged, & Goldstein, 1977; Hoole, 1999; Jackson, 1988; Maeda, 1990; Slud et al., 2002; Stone, 2005). PCA is a general statistical procedure that reduces the dimensionality of high-dimensional data to the principal components (or factors) that account for the most variance in the data. To use PCA for ultrasound analysis, a series of arbitrary reference points are marked on the ultrasound image and then the distance between each reference point and the tongue contour is calculated. PCA can then be used to identify which reference point distances account for the most variability in articulation. For example, one PCA-based investigation demonstrated that two references in a midsagittal plane could account for approximately 90% of the variance between tongue shapes for English and Icelandic vowels (Jackson, 1988). One potential difficulty with using PCA is that the results will depend on which arbitrary points have been defined and measuring the distance from the arbitrary points to the tongue contour can be challenging.

The last method we discuss for analysing tongue contours is based on similarity comparisons of different articulations (Davidson, 2004). This approach involves averaging the contours of the tongue from several repetitions of the same stimulus item. Once a mean contour is generated the difference between curve positions can be calculated. For example, if at point X the height of one contour is 100mm and at point X the height of another contour is 120mm, the difference would be equal to 20mm. Davidson (2004) proposed three different metrics to calculate a similarity value. First, a vector that represents point-to-point differences can be defined using

L_n norms. However, since L_n norms are generated by summing the distance of each point the final value is dependent on the number of measured points. Provided an equal number of points are measured throughout each comparison this should not be problematic. A second metric involves calculating the area between the tongue contours, but it is problematic given that tongue length differs across articulations. The final metric proposed by Davidson (2004) uses a root-mean-square (RMS) calculation; a statistical measure of magnitude which yields a low value if similar items are compared and a high value if dissimilar items are compared. Davidson's (2004) procedure involves calculating an RMS for each stimulus item repetition relative to every other stimulus item repetition. A sign-test is used to determine if the RMS values for one condition are significantly smaller than the RMS values for another condition. If the RMS values are significantly smaller for one condition then it suggest that that articulations in that condition are more similar.

All of the ultrasound analysis methods discussed so far require tongue contour tracing. However, there are technical limitations to tongue tracing methods. One limitation is ultrasound images may contain visual artefacts (see Stone, 2005, for a discussion on artefacts). For example, when the ultrasound waves are emitted from the transducer they are distributed in a fan-like shape. If the tongue is on the same angle as each beam in the fan a double edge artefact may be observed. This occurs because the time for the two beams to reach the tongue and bounce back to the transducer are equal, and are therefore interpreted as equidistant from the probe. A double edge artefact can present challenges for tracing the tongue contour because it is difficult to infer the accurate position of the tongue. See for example, the ultrasound image in Figure 7.2(a). In this frame of ultrasound there is a double edge which makes it difficult to determine where the contour of the tongue should be traced.

A further challenge for contour tracing is that it may not always be possible to identify the full tongue surface in ultrasound recordings. This is especially prevalent when investigating tongue-tip movement. The tongue-tip can be partially obscured by a black shadow formed from an obstruction of the jaw bones (Stone, 2005). Additionally, there is air beneath the tongue tip that hinders the ability of the ultrasound signal to reach the tongue surface and bounce back to the source. As a result the tongue tip tends to disappear from the ultrasound image when it is extended vertically (Stone, 1990). An ultrasound image in Figure 7.2(b) provides an example of when the tongue-tip is not visible during recording: the area within the overlaid circle does not contain a traceable contour.



(a) (b)

Figure 7.2: Two sample ultrasound frames which contain visual artefacts: (a) has a double edge artefact in which the contour of the tongue appears to split inside the overloaded circle; (b) displays an example of how the tongue-tip can disappear during ultrasound imaging.

A limitation for analysis methods that rely on analysing tongue contours is that all measurements are absolute. Using absolute measurements makes assumptions about that spatial properties of tongue movement during articulation. The use of the Cartesian coordinate system to measure tongue height at arbitrary positions does not account for the full contour of the tongue. The use of methods that do include the full contour of the tongue require measurements to be taken from arbitrary points on the ultrasound image. Identifying arbitrary points requires spatial assumptions that the distance between points will differ between two articulations. And arbitrary points are truly arbitrary because absolute physiological landmarks can not be detected using ultrasound (Stone, 1990, 2005). The only potential exception are the curve-fitting algorithms, but due to visual artefacts it is not clear how the equations should be modified to account for double edges and disappearing tongue-tips.

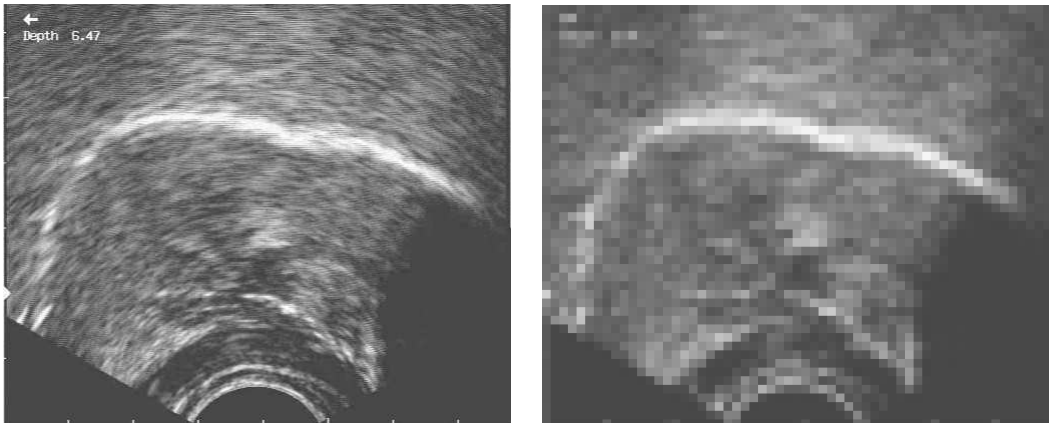
In summary, there are a variety of methods for extracting tongue contours and for analysing the differences between them. However, there are several disadvantages or limitations of using tongue tracing methods for ultrasound analysis. The first, more practical limitation, is the time required for tracing all of the tongue contours. With a minimum image acquisition rate of 25 frames per second this means that a considerable amount of time is required to perform tongue tracing. For example, if each contour could be traced in 5 seconds (which is a rapid estimate) it would take over two hours to trace just a minute of data. While edge detection algorithms can

reduce the amount of time required they still require the user to inspect and hand edit all of the identified tongue contours. A second limitation is visual artefacts can make tongue contour tracing difficult. As a result, data often has to be discarded or the experimenter must infer where the tongue contour should be visible. Lastly, any analysis that relies on tongue tracing requires assumptions about spatial properties of the tongue.

In the following section we present an alternative method for analysing ultrasound data. We extend the Delta method developed for EPG (refer to Chapter 4 for a detailed description) to be applicable to ultrasound. The Delta method is a relative measurement which returns a value representing how similar two articulations are to one another. Rather than identifying the tongue contour in each ultrasound image, we use a novel approach that calculates differences between articulations based on the greyscale values of the entire ultrasound image. Each pixel of an ultrasound image can be represented as a greyscale value ranging from 0 (black) to 255 (white). Using these values, we create vectors that can be compared across articulations in a similar manner to the way we treat EPG contact data as a series of vectors. This approach removes the time-consuming and tedious task of tracing contours. Importantly, the approach can also deal with the visual artefacts that pose challenges for contour tracing. Provided the same artefacts occur across a given speaker and circumstance, the greyscale values will be consistent. For example, if a speakers' articulation of /t/ results in some loss of tongue-tip contour then all repetitions of /t/ will look similar. Additionally, while the contour of the tongue-tip may not be discernible, the image becomes brighter in the region that the contour would have occupied. Therefore, the greyscale value of that region will be higher relative to other regions. A higher value will indicate tongue movement in a given region. One potential objection to using the entire ultrasound image is that there is a considerable amount of noise in ultrasound imaging. However, noise is random by definition so it should not affect comparisons across different images. The following section includes three demonstrations of the Delta method for quantifying ultrasound data.

7.4 Ultrasound Demonstration of the Delta Method

The Delta method for calculating variability in ultrasound recordings of articulation is very similar to the EPG method. In fact, the analysis method is identical to EPG except that the ultrasound data is represented differently. First, since ultrasound does not contain discrete landmark points like the electrodes on an EPG palate we



(a) (b)

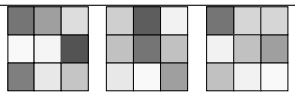
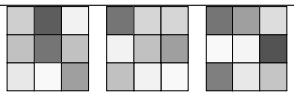
Figure 7.3: A sample of two ultrasound images: (a) a frame of raw ultrasound data; (b) the same frame of ultrasound data after the number of pixels have been reduced for analysis. Each pixel in (b) corresponds to a set of 144 (12 × 12) pixels from (a).

treat each pixel of the video image in a similar way to a contact point. Each pixel has a value ranging from 0 (black) to 255 (white). Since the tongue surface appears as a distinct whitish coloured band, pixels in the portion of the image that shows the tongue will have higher values than those in other portions of the image.

The ultrasound data that we report in this thesis was recorded using the equipment from the Queen Margaret University ultrasound laboratory (a full description of the equipment is presented later in this section). The raw video includes a header window at the top of the image that reports information such as the time and date of the recording. For all analyses we crop the video images to a 507 × 418 pixel window which only includes the oral cavity that was imaged. We refer to the cropped image as *raw* data.

A consequence of measuring each pixel is that ultrasound data embodies much more data than EPG: 507 × 418 pixels compared to the 62 contact points for EPG data. To reduce the number of pixel values, and therefore reduce some noise in the data, we take an average pixel value for every 12 × 12 grid of pixels. This grid size was chosen because an inspection of ultrasound data revealed that the tongue surface is generally 12 pixels in height. An example of a raw video frame of ultrasound data and the same frame of data after the number of pixels have been reduced is presented in Figure 7.3. In both of the images the spline of the tongue is clearly visible as a whitish coloured band that spans the width of the image, with tongue-root on the left and tongue-tip on the right

Table 7.1: A demonstration of a Δ calculation for two epochs of grid data generated to represent ultrasound data: (a) raw epoch data; (b) vectors corresponding to the first frame of the epochs; (c) vectors corresponding to the second frame of the epochs; and (d) vectors corresponding to the third frame of the epochs. Δ is equal to the mean Euclidean distance of the pairwise vector comparisons.

	Epoch_A	Epoch_B	Euclidean Distance
(a)			
(b)	 (127 232 197 249 246 83 117 159 221)	 (232 249 159 193 117 193 207 93 242)	239.86
(c)	 (232 249 159 193 117 193 207 93 242)	 (193 242 249 242 193 159 117 214 214)	206.22
(d)	 (193 242 249 242 193 159 117 214 214)	 (127 232 197 249 246 83 117 159 221)	+ 137.36
			583.44 / 3
			Δ = 194.48

In order to demonstrate the Delta method for ultrasound we generated two epochs of grid data (see Section 4.4 for a description) to simplify the calculation. Table 7.1.a displays the two epochs of raw grid data to be compared and rows b-d are the vectors corresponding to each frame of the epochs. The vectors are generated by starting in the southwestern most corner of each frame and moving from left to right across each row of the grid. The values in rows b-d are the greyscale values used for the calculation. Similar to the EPG calculations, Δ is equal to the mean Euclidean distance of each array comparison. The only difference between this calculation and the calculation used for EPG data in Section 4.4.2 is that values range from 0–255, rather than 0–1. Note that the Δ value (194.48) from this calculation is much higher than the Δ value for the EPG calculation (2.06). This is not due to these epochs being much more dissimilar than the epochs used for the EPG calculation, but is simply a consequence of the 0–255 scale of values used for the calculation. Therefore, Δ values generated from the Delta method should only be directly compared to a relevant set of data since EPG and ultrasound calculations yield values on different scales.

To demonstrate how the Delta method works for real ultrasound data we recorded pilot data for one speaker. Ultrasound data was collected using a Concept M6 Digital Ultrasonic Diagnostic Imaging System (Dynamic Imaging: Livingston, UK)

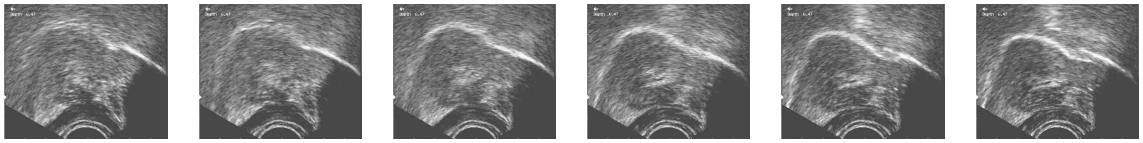
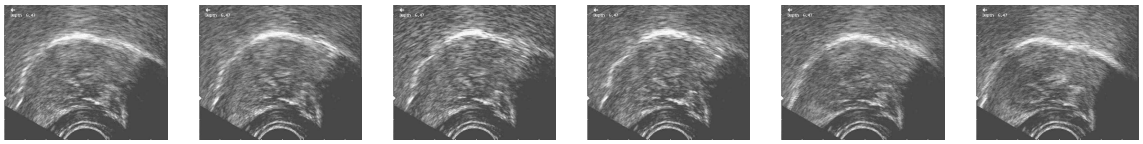
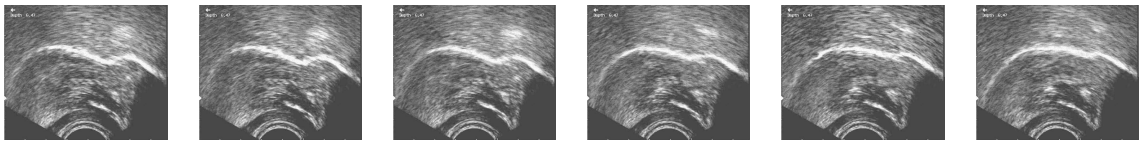
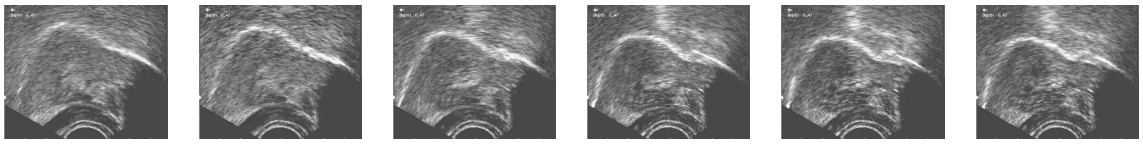
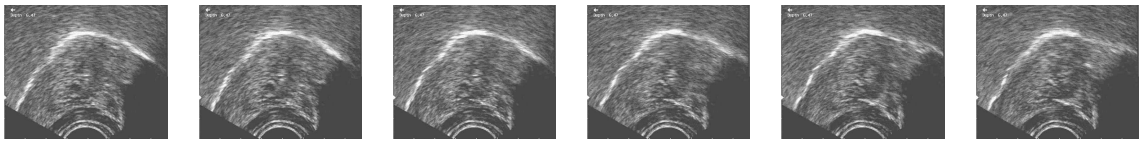
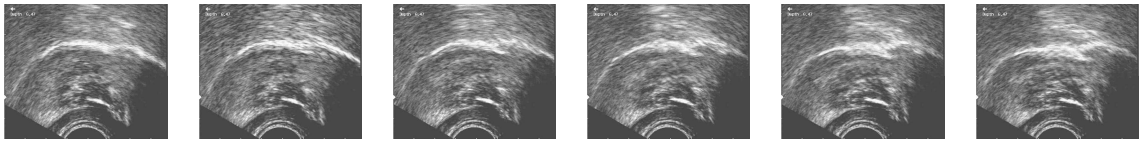
together with an endocavity transducer probe (Model 65EC10EA; Mindray: Shenzhen, China). The probe was secured at an approximately 90° angle beneath the chin with a custom manufactured lightweight helmet (Articulate Instruments Ltd.; Edinburgh, UK). Ultrasound images were acquired with a 6.5MHz image frequency, 120° image field sector, and a 25Hz acquisition rate. The axial resolution, when measured in water, was 0.5mm with a penetration depth of 95mm. Acoustic recordings of participants' responses were recorded at 22,050Hz using an Audiotechnica ATM10a microphone. The acoustic and ultrasound data were synchronised using Articulate Assistant Advanced software (Articulate Instruments Ltd.: Edinburgh, UK). Auditory signals of a metronome beat were played at a rate of 100 syllables/minute through monaural headphones worn in the speaker's preferred ear. The speaker was instructed to repeat *kom*, *tom*, and *som* sixteen times each at a rate of one word per metronome beat.

After recording, the entire video file for each stimulus item was exported from Articulate Assistant in AVI format using a MPEG-4 (mp42) Video Codec. Each frame of the video file was then converted to a PNG still frame using Mplayer (http://www.mplayerhq.hu) software. To identify the articulatory records of interest the onset of the acoustic release for each initial consonant was identified using Praat software(Boersma & Weenink, 2006). Using the time point of the acoustic release, each articulatory record was defined as 0.3 seconds preceding the release to 0.3 seconds following the release.

This method of identifying articulatory records differs slightly from the EPG analysis. For the ultrasound data we always use equal length epochs (though it is possible to use unequal epochs) because it is difficult in ultrasound data to identify a discrete starting point and discrete ending point of an articulation, whereas in EPG you can more easily define a start and end point based on the closure pattern. In pilot analyses we tested the Delta method with all variations of epoch lengths ranging from 0.3 before release to 0.3 seconds after the release (e.g., 0.3s pre-release to 0.1s post-release; 0.2s pre-release to 0.2s post-release) and found that the Δ values were similar for all epoch lengths. Therefore we chose the longest epoch (0.3s pre-release to 0.3s post-release; 15 frames) to allow the most variance to be captured.

We calculated Δ values for three randomly selected ultrasound records of a /k/, /t/, and /s/ relative to three randomly selected references of the same phonemes. The middle six frames (5-10) of each recording are presented in Table 7.2 along with the corresponding Δ values relative to each reference. It is clear from the Δ values in the table that the /k/ articulation is most similar to the /k/ reference, the /t/

Table 7.2: Results from a comparison of six different ultrasound recordings using the Delta method. For each articulation (d, e, f) a Δ value was calculated relative to each of the reference articulations (a, b, c). The Δ values demonstrate that /k/ is most similar to the /k/-reference, /t/ is most similar to the /t/-reference, and /s/ is most similar to the /s/-reference.

		Frame						Reference		
		5	6	7	8	9	10	/k/	/t/	/s/
(a)	/k/ ref.							–	–	–
(b)	/t/ ref.							–	–	–
(c)	/s/ ref.							–	–	–
(d)	/k/							764.40	894.55	1013.07
(e)	/t/							1141.75	773.15	910.22
(f)	/s/							1088.82	979.93	820.74

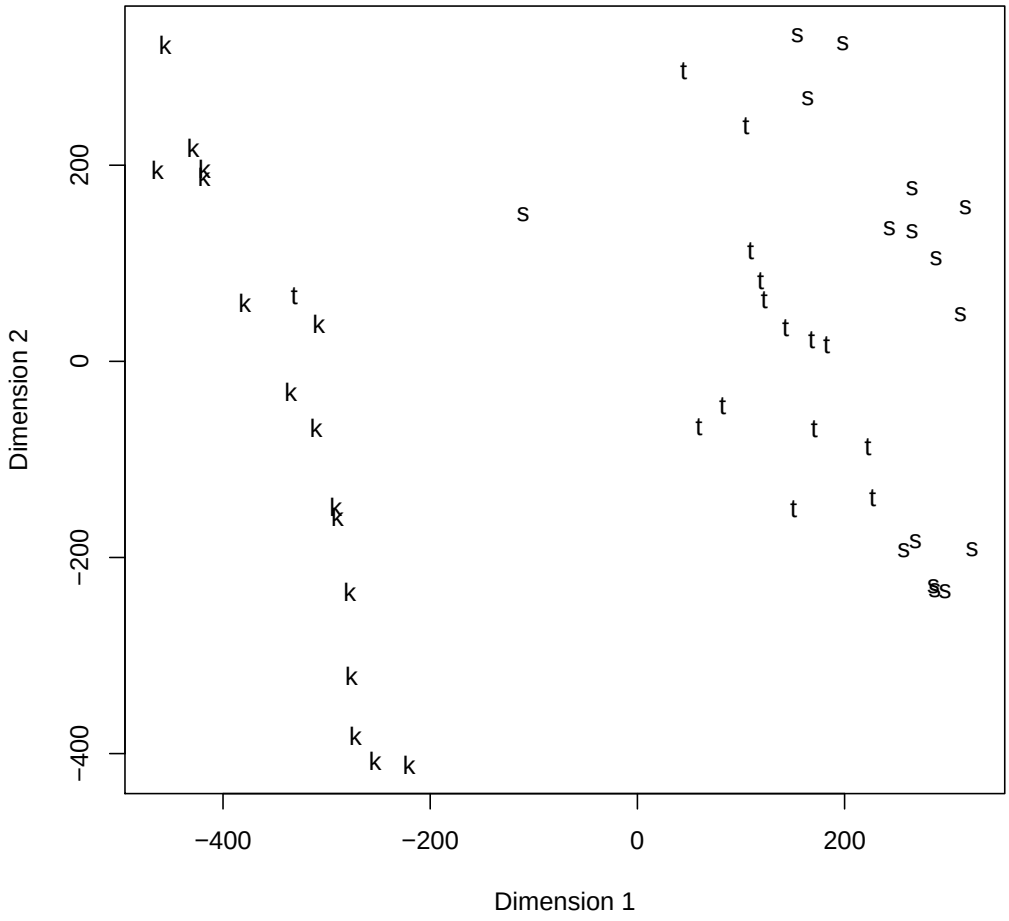


Figure 7.4: A multidimensional scaling plot for the comparisons of 16 /k/, 16 /t/, and 16 /s/ articulations recorded with ultrasound. Δ was calculated for each articulation relative to every other articulation. The plot contains a clear cluster of velar (/k/) articulations and a cluster of alveolar (/t/, and /s/) articulations

articulation is most similar to the /t/ reference, and the /s/ articulation is most similar to the /s/ reference. It is also clear that the alveolar articulations (/t/, /s/) differ most from the velar (/k/) reference. This pattern suggests that the Delta method can successfully capture relative (dis)similarities between different types of articulations.

Lastly, to demonstrate the range of Δ values from the ultrasound recordings we calculated Δ for each articulatory record relative to every other articulatory record and generated a (dis)similarity matrix containing all of the values. We then used a multidimensional scaling algorithm (Cailliez, 1983; Cox & Cox, 1994) to visualise

the results. Multidimensional scaling takes a set of similarity values (e.g., Δ values) and returns a set of points on a scatter plot so that the distance between the points of the plot are approximately equal to similarity values between the points. The results from the multidimensional scaling analysis are presented in Figure 7.4. There are three features of this plot that are important. First, all of the /k/ articulations form a cluster that is clearly separate from /t/ and /s/ articulations. This provides evidence that the Delta method successfully captures relative differences between velar and alveolar articulations. Second, the alveolar articulations (/t/, /s/) cluster together, but within this cluster there is one sub-cluster of /t/ articulations and several sub-clusters of /s/ articulations. This suggests that the Delta method can capture similarities in alveolar articulations and at the same time can capture the subtle differences between /t/ and /s/ articulations. These findings are important because it establishes evidence that the Delta method can be a useful tool for measuring relative differences in the variability of articulation recorded with ultrasound.

In summary, the Delta method can be used for measuring relative differences between articulations recorded with ultrasound. The use of the Delta method for ultrasound has benefits over previously used ultrasound analysis methods. It is a relative measurement, that does not require assumptions about spatial properties of articulation. Since it is based on the full ultrasound image, rather than traced tongue contours, it can also be used for images that contain visual artefacts. Lastly, the Delta method provides a value that accounts for changes in articulation over time, rather than measurements that only evaluate a single frame of ultrasound data.

In the following section we present an experiment designed to investigate whether features are required for models of production. Similar to Experiments 1 and 3, we use the Delta method to compare articulation in error-invoking conditions (e.g., tongue-twisters in which phonological representations are competing) to articulation in conditions that do not invoke speech errors (e.g., tongue-twisters with noncompeting representations).

7.5 Experiment 4: Ultrasound Tongue-Twisters

To further investigate the role of features in models of production we conducted an ultrasound investigation investigating phonological similarity. The experiment attempts to provide additional evidence for feedback by investigating variability of

articulation measured with ultrasound. The experiment takes the form of a tongue-twister investigation similar to the one used in Experiment 3. Using a similar design across experiments allows us to test for a replication of the phonological similarity effects from Experiment 3 using a different articulatory imaging technique.

A benefit of using ultrasound, rather than EPG, is that partial movements which reflect activation of phonological representations can be recorded. Previous ultrasound analyses from speech error investigations have suggested that partial "errors" are produced during repetition tasks (Pouplier, 2004; Frisch, 2007; Stearns, 2006). In one investigation Pouplier (2004) traced the contour of the tongue on ultrasound recordings of alternating velar and alveolar repetitions (e.g., *cop top*). She then measured tongue-dorsum height and tongue-tip slope for each recording and observed a continuum of values for each measure ranging from normal to abnormal. For example, the tongue-tip slope for the production of alveolar consonants ranged from being typical for an alveolar production to being typical for a velar production. Similarly, Stearns (2006) identified several occurrences of "gradient errors", which were defined as articulations that differed by two standard deviations from the mean tongue-dorsum height and the mean tongue-tip height. These findings by Pouplier (2004) and Stearns (2006) are consistent with other articulatory speech error investigations which have demonstrated a continuum of values using EMMA (Goldstein et al., 2007; Pouplier, 2003, 2007).

However, the sources of ultrasound evidence that suggest partial errors can occur (Pouplier, 2004; Stearns, 2006) were based on analysis methods with limitations. First, only one frame of the ultrasound data for each onset was analysed which does not account for temporal characteristics of the articulations. Second, both analyses relied on tracing the contour of the tongue which resulted in the exclusion of a high proportion of data. In one analysis Pouplier (2004) had to exclude one of two participants because the tongue contours for all /k/-initial targets could not be traced reliably. In addition, once the tongue contours were traced the analyses were restricted to two arbitrary points on the tongue surface. Lastly, and most importantly, the responses from each analysis were assigned to discrete categories.

For the current investigation we use the Delta method to investigate articulatory variability recorded with ultrasound. The Delta method allows us to measure variability between articulations with competing features relative to articulations without competing features. By analysing the data in this way we can quantify how feature competition influences articulation without assigning responses to categories.

Additionally, the Delta method does not require tongue contour tracing and can quantify both spatial and temporal properties of the recorded articulations.

The current experiment innovates from Experiment 3 by introducing additional feature competition by including the phonemes /s/ and /z/. The inclusion of additional phonemes allows us to manipulate whether Place (change, no change), Voice (change, no change), and/or Manner (change/no change) features were competing during the repetition of simple four syllable phrases (e.g., *gom som som gom*) and test the influence of having three competing features. Previous speech error investigations on phonological similarity in error production have demonstrated that errors are more likely to occur if the competing phonological representations differ by one or two features (Shattuck-Hufnagel, 1979, 1983; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1982). Since we are investigating articulatory variability, rather than categorising errors, it is important to determine the influence of a third competing feature on tongue contact variability.

In addition to having a third competing feature the use of /s/ and /z/ also introduces a different type of feature to the experimental design. Our previous analyses have only focused on stop consonants. However, it has been suggested that different types of features may interact differently (Shattuck-Hufnagel, 1979). Shattuck-Hufnagel and Klatt (1979, see also Shattuck-Hufnagel & Klatt, 1975) generated a confusion matrix of all errors observed in the MIT corpus. In an analysis of the confusion matrix they demonstrated that place of articulation substitutions are much more common than manner or voicing substitutions. By adding manner to our analyses we can extend our previous research to include fricatives, and evaluate whether different types of features make different contributions to articulatory variability.

In addition to adding a third competing feature to the experimental design, we refined the methodology of the tongue-twister experiment. The materials for the current experiment are more precisely controlled. All stimulus items contain the same vowel (/ɒ/) and final-consonant (/m/) and thereby only differ in phonological onset. This was done to control for movement of the tongue to the vowel position after production of the initial onset. In ultrasound any measurement after palatal contact will be clearly influenced by the following vowel position since the entire contour of the tongue is being measured. This is unlike the previous EPG experiments in which the vowels chosen for stimulus items typically yield zero or minimal palatal contact. Likewise, we used the same final consonant so that any variability caused by transition from the final onset of one word to the initial-onset of the following word would be consistent across all conditions. Additionally the chosen

final consonant differed in all phonological features from the target except voice+ for some stimulus items, though there is no voice- pair for /m/ in the languages of the world.

Another methodological modification is the use of a slower speaking rate (100 syllables/minute compared to 150 syllables/minute in Experiment 3). A slower speaking rate was chosen primarily for practical reasons: the acquisition rate for ultrasound recording in the laboratory at Queen Margaret University is only 25 frames per second compared to EPG which was acquired at 100 frames per second. By using a slower speaking rate we can increase the duration of articulation and therefore include more frames per token in the data analysis. Interestingly, a replication of the phonological similarity effect at a slower speaking will establish that variability in a tongue-twister task can not solely be attributed to rapid repetitions. This finding would compliment Wilshire's (1999) evidence that tongue-twister errors are not due to errors in articulatory execution due to rapid repetition.

In summary, the primary goal of this experiment is to replicate the phonological similarity effects observed in Experiment 3. The first analysis we report is an ultrasound analysis of variability in articulation. This analysis only includes the target phonemes used in Experiment 3 (/k/, /g/, /t/, /d/). The second analysis is an acoustic analysis of VOT that was also designed to replicate the VOT results from Experiment 3. The final analysis is an ultrasound analysis that includes the additional phonemes (/s/, /z/). This analysis will investigate the influence of a third competing feature, manner of articulation, on phonological similarity.

7.5.1 Methods

Participants

Ten native-English speakers from the University of Edinburgh and Queen Margaret University communities participated in the experiment. All participants reported no speech or hearing impairments and were treated in accordance with the University of Edinburgh and Queen Margaret University ethical guidelines. Two speakers were excluded from all analyses due to poor ultrasound image quality in comparison to the other eight speakers.

Materials

The tongue-twisters were designed to include Place of Articulation (alveolar+ vs. velar+), Voicing (voice+ vs. voice–), and Manner (plosive vs. fricative) contrasts. To achieve this a set of 15 tongue-twisters were designed with each possible combination of two onsets (/t/, /d/, /k/, /g/, /s/, /z/). The onset phonemes were selected so that the shape of the midsagittal spline of the tongue would differ across phonemes, except perhaps for the voicing contrasts. Each pair of onsets was combined with the same vowel (/ɒ/) across onset pair combinations. Every sequence also ended with /m/ which differed in manner from the stop consonants and fricatives. Each of these 15 syllable pairs was then used to create 15 ABBA tongue-twisters (e.g., *tom gom gom tom*) and 15 BAAB tongue-twisters (e.g., *gom tom tom gom*). In addition to these 30 tongue-twisters another 6 tongue-twisters were created that only contained repetitions of one of each onset phoneme (e.g., *tom tom tom tom*).

Apparatus

The experiment was conducted in a sound-treated recording suite at Queen Margaret University. Ultrasound data was collected using a Concept M6 Digital Ultrasonic Diagnostic Imaging System (Dynamic Imaging: Livingston, UK) together with an endocavity transducer probe (Model 65EC10EA; Mindray: Shenzhen, China). The probe was secured at an approximately 90° angle beneath the chin with a custom manufactured lightweight helmet (Articulate Instruments Ltd.; Edinburgh, UK). Ultrasound images were acquired with a 6.5MHz image frequency, 120° image field sector, and a 25Hz acquisition rate. The axial resolution, when measured in water, was 0.5mm with a penetration depth of 95mm.

Acoustic recordings of participants' responses were recorded at 22,050Hz using an Audiotechnica ATM10a microphone. The acoustic and ultrasound data were synchronised using Articulate Assistant Advanced software (Articulate Instruments Ltd.: Edinburgh, UK). The entire video file for each stimulus item was exported from Articulate Assistant into AVI format using a MPEG-4 (mp42) Video Codec. Lastly, each frame of the video file was converted to a PNG still frame using Mplayer (http://www.mplayerhq.hu) software.

To control speaking rate participants were presented with an auditory metronome beat at a rate of 100 beats per minute. The metronome signal was played through

stereo headphones and participants were given the choice to listen binaurally or monaurally (using their preferred ear).

Procedure

Once participants had been seated they were fitted with a lightweight helmet designed to hold the ultrasound transducer stationary relative to the head position. Then the ultrasound transducer was secured to the helmet and beneath the chin with a pressure as firm as comfortable. Participants were instructed to read out loud two randomly selected tongue-twisters. During these repetitions the transducer probe was adjusted to yield the highest quality ultrasound image. Participants were given feedback about their vowel pronunciation to ensure that they were pronouncing the vowel (/ɒ/) correctly.

After setup was completed participants were fitted with headphones for the metronome signal. Participants were instructed to repeat each sequence on the monitor four times at a rate of one word per metronome beat. The experimenter, who was in another room, controlled the presentation rate and indicated to the participant to begin speaking by changing the background of the visual display to green. After the speaker was finished with each tongue-twister sequence the display colour was made white. There was a short pause of approximately 7 seconds between each tongue-twister to allow data to be saved. Participants were instructed to let the experimenter know if they required a longer break. All 36 tongue-twisters were presented in random order.

Data Treatment

Following the experiment, measurements were performed on the ultrasound data and a subset of the acoustic data. Each initial onset from the tongue-twisters was treated as an independent item. The only items excluded were those items not collected due to recording failure (66 out of 4608 possible responses).

Ultrasound Data Each recorded item was identified in Praat (Boersma & Weenink, 2006) using the acoustic signal. The key time point identified for each item was the onset of the acoustic release. The ultrasound record was then defined as the duration of 0.3s before the release to 0.3s after the release. The video frames for these records were extracted from the ultrasound video files and converted into PNG still images using Mplayer (http://www.mplayerhq.hu) software. Once the ultrasound records were extracted they were analysed using the Delta method described in Section 7.4.

A mean reference articulation was created for each onset phoneme (/t/, /d/, /k/, /g/, /s/, /t/) for each speaker. The items used to create the references were from the control tongue-twister sequences (e.g., *tom tom tom tom*). We then calculated Δ difference scores for each recorded item relative to the relevant mean reference. The higher the difference score the greater the variability of the target from the mean reference.

Acoustic Data The VOT for each stop consonant target item (/t/, /d/, /k/, /g/) was measured using Praat (Boersma & Weenink, 2006) software. This was defined as the duration from the onset of the acoustic burst through to the onset of periodicity of the associated vowel. Mean reference VOTs were created by averaging the VOT duration of each stop consonant from the control tongue-twisters (e.g., *tom tom tom tom*) for each speaker. For the acoustic analysis we report VOT difference scores calculated as the absolute value of the target VOT minus the relevant mean reference VOT.

7.5.2 Replication Results

To investigate whether features are required in models of speech production we conducted two independent analyses on the ultrasound and VOT data. Both analyses only included the stop-consonant articulations (/k/, /g/, /t/, /d/) and were designed to test for a replication of the phonological similarity results from Experiment 3 (Chapter 6). The recorded responses of ultrasound and VOT data were analysed independently using a Generalised Linear Mixed-Effects model, with the lme4 (Bates & Sarkar, 2007) and languageR (Baayen, in press) packages in R. These analyses include Place (change, no change) and Voice (change, no change) as fixed effect factors and item and participant as random effect factors. Each tongue-twister sequence was treated as an experimental item for the analyses. For both analyses the fixed effect factors were centred to reduce multi-collinearity in an unbalanced design (Landsheera et al., 2006) and the interaction term between factors was included because the interaction was the primary contrast of theoretical interest.

The *t*-values for each contrast are reported along with probabilities based on 10,000 Markov Chain Monte Carlo (MCMC) samples. Using a MCMC probability estimate has been suggested because it can be difficult to calculate the accurate degrees of freedom corresponding with each *t*-value for the intercepts (Bates & Sarkar, 2007). All reported mean values and confidence intervals are based on MCMC

mean estimates of each intercept. The intercepts were calculated in the 2 × 2 design by modelling each factor separately. The confidence intervals correspond to the effect size within each of the respective models.

It should be noted that there is an important distinction between Linear Mixed-Effects models and ANOVA statistical tests. In an ANOVA analysis the mean and variance within each experimental condition are compared relative to one another. However, a Linear Mixed-Effects model is based on whether the variance contributed by one condition is greater than the variance contributed by another condition. For example, in the current experiment a main effect for Place (change, no change) indicates that a change in place of articulation has an influence on the dependent variable (e.g., articulatory variability). Likewise, a significant Place × Voice interaction indicates that a change in both place of articulation and voicing significantly influences the dependent variable. For analyses of the present experiment, a significant positive effect for one factor (e.g., Place) plus a significant *negative* effect for the combination of factors (Place × Voice) can be interpreted as the equivalent of a traditional ANOVA interaction in which a change in Place causes greater articulatory variability than changes in Place and Voice combined.

Articulation Analysis (Ultrasound)

The articulation analysis included 2023 ultrasound recordings for 8 speakers and 16 items. The analysis revealed a significant main effect of Place [t=11.72, p<0.0001] in which articulation was more variable when onsets included a place of articulation change (642.85) compared to no change (560.99; 95%CI ±14.91). We also observed a significant main effect of Voice [t=5.04, p<0.0005] with more variability in articulation when voice changed (596.22) compared to when voice did not change (560.99; 95%CI ±14.86). Lastly, the Place × Voice interaction was also significant [t=5.13, p<0.0005]. When only place of articulation changed (642.85) articulation was more variable than when both place of articulation and voice changed (606.39; 95%CI ±29.83). Refer to Figure 7.5 for MCMC mean estimates for each condition.

Acoustic Analysis (VOT)

The acoustic analysis included 2023 recorded onsets for 8 speakers and 16 items. The same mixed effects model was used as above with Place (change, no change) and Voice (change, no change) as fixed effects factors and participant and item as random effects factors. We observed a significant main effect of Voice [t=2.46, p<0.05]: VOT variability was greater when voice changed (11.96) compared to

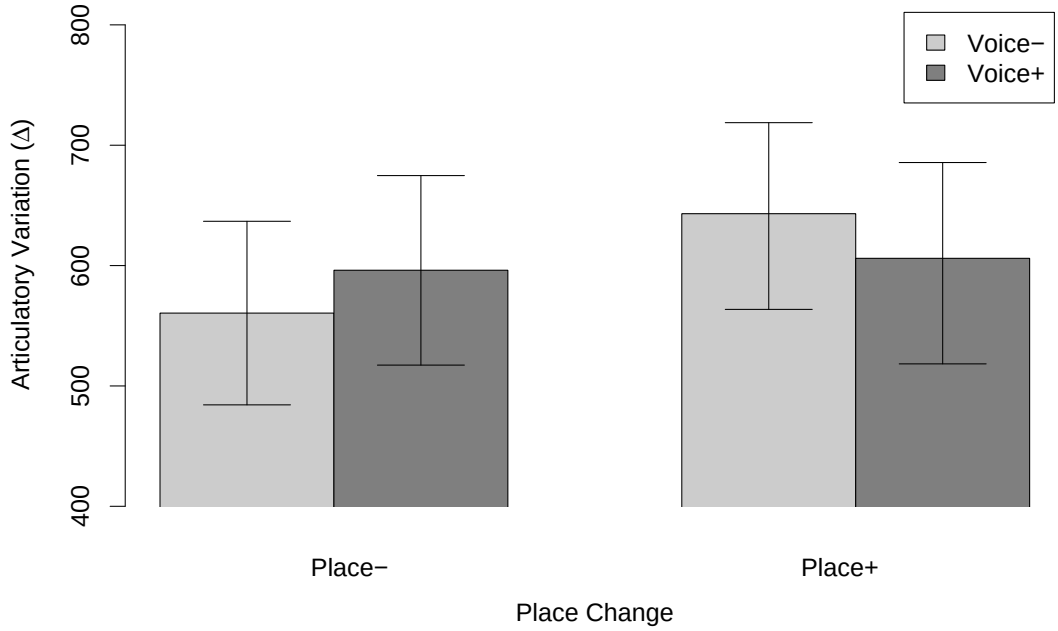


Figure 7.5: MCMC mean estimates for articulatory variation (Δ) recorded with ultrasound for tongue-twisters from Experiment 4 with 95% confidence intervals.

when it did not change (9.3ms; 95%CI ±2.35). We did not observed a significant main effect of Place [t<1] or a significant Place × Voice interaction [t=1.52, p>0.1]. Refer to Figure 7.6 for MCMC mean estimates for each condition.

7.5.3 Replication Discussion

The analysis of ultrasound data establishes experimental evidence for phonological similarity. For the ultrasound analysis we compared articulation when there were similar (e.g., /k/-/t/) and dissimilar (e.g., /k/-/d/) competing phonological representations relative to articulations when there were no competing phonological representations (e.g., /k/-/k/). This analysis demonstrated that variability in articulation was greatest when only place of articulation was competing compared to when place of articulation and voicing were competing. This pattern of results provides a direct replication of the EPG analysis from Experiment 3: articulation is most variable when similar phonological representations are competing. This firmly establishes that feature representations are required in models of speech production.

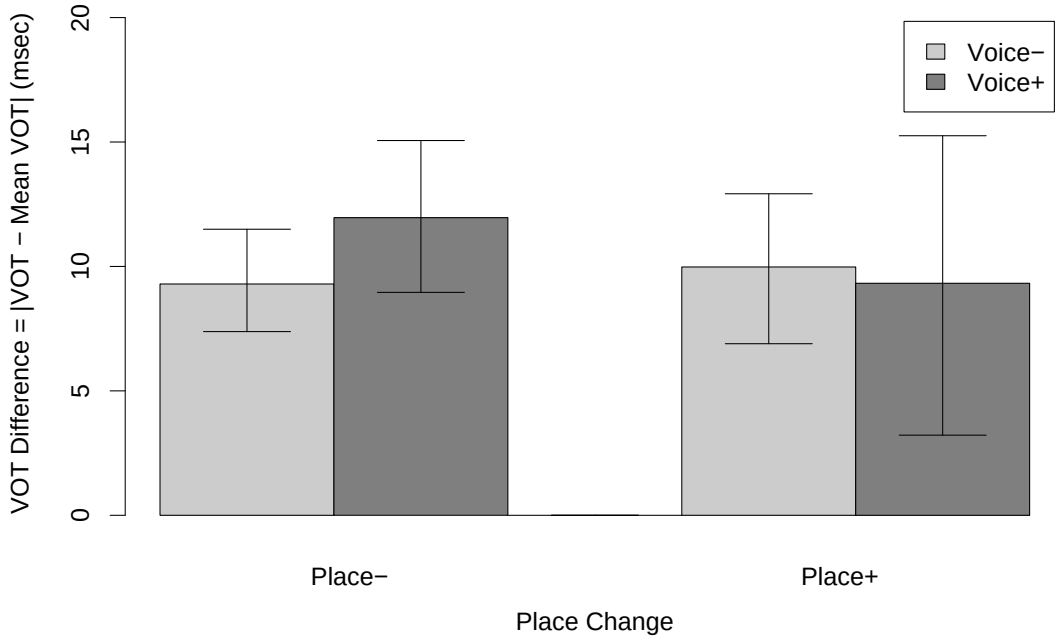


Figure 7.6: MCMC mean estimates for VOT variation (msec) in the acoustic analysis of tongue-twisters from Experiment 4 with 95% confidence intervals.

Importantly, this replication also provides evidence that the Delta method can be used to quantify articulatory data for EPG and ultrasound imaging.

The analysis of the VOT data did not reveal significant evidence for phonological similarity. For this analysis we compared the VOT of phonemes spoken when there were similar (e.g., /k/-/g/) and dissimilar (e.g., /k/-/d/) competing phonological representations relative to the VOT of phonemes spoken when there were not competing phonological representations (e.g., /k/-/k/). The mean VOT difference score was more variable when there was voicing competition compared to when voicing and place of articulation were competing. However, this difference was not statistically significant. The acoustic results, therefore, do not provide significant evidence for feature representations.

Before we discuss the theoretical implications of these results we report an additional ultrasound analysis. The final analysis investigates phonological similarity when there are three competing feature representations. This analyses includes all of the recorded tongue-twisters, which contain /k/, /g/, /t/, /d/, /s/, and /z/

phonological onsets. The addition of /s/ and /z/ also introduces a different type of feature to the analysis: manner of articulation.

7.5.4 Three Competing Features Results

To investigate phonological similarity when there is a third competing feature representation and different type of feature competing (manner of articulation) we performed an additional analysis on the ultrasound data. The complementary acoustic analysis for VOT was not performed because different measures of voicing are required for stop consonants (/k/, /g/, /t/, /d/) and fricatives (/s/, /z/). The ultrasound data was analysed using a Generalised Linear Mixed-Effects model, with the lme4 (Bates & Sarkar, 2007) and languageR (Baayen, in press) packages in R. The analyses is based on every recorded ultrasound observation and included Place (change, no change), Voice (change, no change), and Manner (change, no change) as fixed effect factors and item and participant as random effect factors. Each tongue-twister sequence was treated as an independent item for the analysis. The fixed effect factors were centred to reduce multi-collinearity in an unbalanced design (Landsheera et al., 2006). The interaction term between each factor (e.g., Place × Voice × Manner) was included since all of the interactions were the primary contrasts of theoretical interest. Similar to the previous analyses, the *t*-values for each contrast are reported along with probabilities based on 10,000 Markov Chain Monte Carlo (MCMC) samples. All reported mean values and confidence intervals are based on MCMC mean estimates of each intercept. The intercepts were calculated in the 2 × 2 × 2 design by modelling each factor separately.

This ultrasound analysis included 4542 ultrasound recordings for 8 speakers and 36 items. Refer to Figure 7.7 for MCMC mean estimates for each condition. All of the main effects were significant with more variability when there was a change within each factor compared to no change: Place [t=17.44, p<0.0001; 650.73 vs. 569.90; 95%CI ±83.35], Manner [t=8.64, p<.0001; 609.96 vs 569.90; 95%CI ±9.58], and Voice [t=4.01, p<.001; 588.39 vs. 569.90; 95%CI ±9.44].

All of the two way interactions were also significant. Articulation was more variable when only place changed (650.73) compared to when place and voice changed (623.89; t=4.89, p<0.0001; 95%CI ±18.72). Similarly, there was more variability in articulation when only manner changed (609.96) compared to when both manner and voice changed (572.98; 95%CI ±18.72; t=5.98, p<0.0001). However, a change

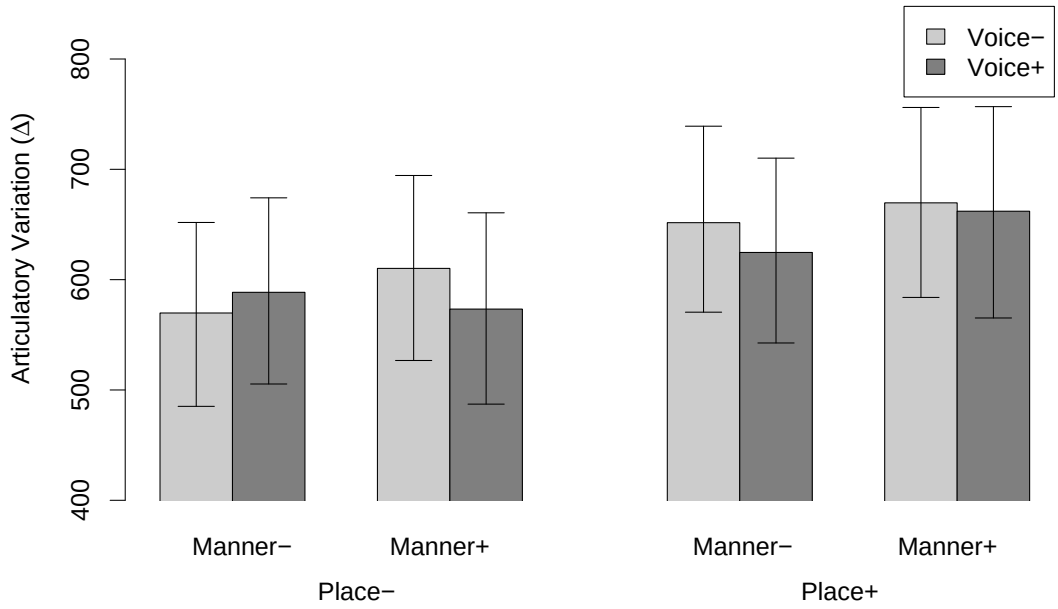


Figure 7.7: MCMC mean estimates for articulatory variation (Δ) recorded with ultrasound for tongue-twisters from Experiment 4 with 95% confidence intervals.

in both place and manner resulted in more variability than when only place changed (669.83 vs. 650.73; t=2.25, p<0.05; 95%CI ±18.84).

Lastly, we observed a significant three way interaction for Place × Manner × Voice [t=4.03, p<.001; 95%CI ±37.98]: variability was greatest when place and manner changed regardless of whether voice changed or did not change.

7.5.5 Three Competing Features Discussion

The analysis of ultrasound recordings established additional evidence for phonological similarity influences of articulatory variability. We compared the variability of ultrasound recordings of phonemes spoken when there was one competing feature representation (e.g., /k/-/t/), two competing representations (e.g., /k/-/d/), or three competing feature representations (e.g., /k/-/z/) to the articulations when there were no competing representations (e.g., /k/-/k/). This analysis revealed that articulation was more variable when only a place of articulation feature was competing compared to when a place of articulation and voicing feature were competing. This pattern of articulatory variability replicates the pattern of results observed in

the previous ultrasound analysis and the EPG analysis from Experiment 3: articulation is more variable when similar phonological representations are competing. This is an important finding because it establishes additional evidence for phonological similarity and therefore implicates a role for feature representations in models of speech production.

The results of the current analysis also demonstrated that there was more variability in articulation when only manner of articulation representations were competing compared to when manner of articulation and voicing representations were competing. This pattern provides additional evidence for phonological similarity. Since this analysis included fricatives (e.g., /s/-/t/ variability compared to /s/-/d/ variability) it establishes that that phonological similarity effects of articulatory variation are not unique to place of articulation competition and are not unique to the articulation of stop consonants.

Two patterns of articulatory variability were observed that were not consistent with phonological similarity effects. First, articulation was more variable when both manner of articulation and place of articulation features were competing (e.g., /k/-/s/) than when only place of articulation was competing (e.g., /k/-/t/). Second, the greatest overall amount of variation was observed when both place of articulation and manner of articulation feature representations were competing. We discuss potential reasons for this unpredicted pattern in the following section.

7.6 General Discussion

This chapter presented an extension of the articulation investigations reported earlier in this thesis. The primary goal was to extended our previous EPG investigations on articulation to include an alternative articulatory imaging method. Ultrasound is a useful articulatory imaging technique that provides a midsagittal image of the tongue during articulation. The ability to measure tongue shape during articulation complements the previously reported EPG research that is limited to tongue-to-palate contact information.

In this chapter we presented a review of ultrasound analysis methods and identified key limitations of these methods. The primary limitation is that ultrasound analysis methods often rely on tracing the contour of the tongue. Restricting analysis to the tongue contour has a consequence that spatial assumptions are required to

analyse the tongue position. These assumptions include measuring the tongue position relative to arbitrary points on the ultrasound image, which do not correspond to absolute physiological landmarks. Moreover, there are practical limitations for tongue tracing including the identification of the tongue contour when there are visual artefacts.

In Section 7.4 we demonstrated how the Delta method (originally developed in Chapter 4) can be used for ultrasound data analysis. This method involves calculating differences between articulations based on the entire ultrasound image. Using a relative measurement does not require assumptions about the spatial properties of articulation. Also, since the analysis uses the entire ultrasound image, assumptions are not required about where the tongue contour should be when it is not fully visible. Importantly, we demonstrated that quantifying ultrasound data with the Delta method is a successful approach for identifying (dis)similarities between different types of articulations. In particular it can identify dissimilarities between alveolar and velar patterns of articulation.

The ability to use the Delta method for different articulatory imaging methods is a victory for articulatory imaging analysis. Most articulatory analysis methods are specific to each articulatory imaging technique: EPG analysis is usually restricted to spatial indices (Byrd et al., 1995; Hardcastle et al., 1991, see also Section 4.3 for a detailed discussion) not relevant for ultrasound analysis; and ultrasound analysis is usually restricted to tongue contour measurements (Stone, 2005) which are not relevant for EPG. Our research is focused on whether articulation is more variable in one condition compared to another, not the physio-acoustic details of how different patterns of palatal contact or tongue shapes yield different sounds. The Delta method is able to capture this variability independent of the articulatory imaging technique used. Therefore, the ability to use the Delta method for ultrasound and EPG allows for replicative investigations to be undertaken.

In Experiment 4 we used the Delta method to analyse tongue-twisters designed to investigate phonological similarity. The primary goal of this experiment was to test for a replication of the phonological similarity results from EPG recordings from Experiment 3. In an ultrasound analysis, we demonstrated that articulation is more variable for stop consonants when only a place of articulation feature is competing compared to when a place of articulation and voicing features are competing. This pattern directly replicates the results from Experiment 3: articulation is most variable when phonologically similar representations are competing.

Additional ultrasound analyses also support the pattern of phonological similarity. In an analysis that included fricatives and stop consonants, we demonstrated that articulation is more variable when manner of articulation is competing compared to when manner of articulation and voicing features are competing. This finding supports the phonological similarity pattern of Experiment 3 and the previous analysis. This evidence for phonological similarity is not limited to stop consonant articulation or the place of articulation features. This is an important finding because some researchers have argued that different types of features interact differently (Shattuck-Hufnagel & Klatt, 1979).

The most straightforward account for the phonological similarity results is by including featural representations in models of production and by allowing a feedback flow of information between feature and phonological representations (Dell, 1986; Dell et al., 1993; Stemberger, 1982, 1985a). According to this account, activation from the phonological representations flows to activate feature representations, which feed back and in turn reinforce the phonological representations. Phonological representations that are reinforced are more likely to cascade to articulation. This type of a model is illustrated in Figure 6.1(C) and (D). Despite the evidence for phonological similarity and the implications for the inclusion of features in a model, we cannot discriminate between an account that proposes cascading phonological representations and an account with cascading feature representations. Future research is required to distinguish between these accounts.

Our claim about the importance of features comes with some caveats. First, we observed in our ultrasound analysis that when place of articulation and manner of articulation representations were competing, variability was greater compared to all other conditions. This pattern suggests that variability is greater when dissimilar phonological representations are competing, which is contrary to the previously observed phonological similarity effects.

This unpredicted result is likely a limitation of the experimental design. The design of Experiment 4 included alveolar stop consonants (/t/, /d/), velar stop consonants (/k/, /g/), and alveolar fricatives (/s/, /z/). It did not include velar fricatives (/x/, /ɣ/), because these sounds are not used in English. As a result the comparison of the onsets for the condition with manner and place of articulation competition included the phonemes /k/, /g/, /s/, and /z/ and the condition with only manner competing included the phonemes /t/, /d/, /s/, and /z/. Therefore, any articulation that includes properties of the competing representation in the former condition (e.g., /k/ has some properties of /s/) will yield a higher Δ value than in the latter

condition (e.g., /t/ has some properties of /s/). This is because the tongue shapes for /k/ and /s/ are dissimilar, while the tongue shapes for /t/ and /s/ are similar. Future research is required to resolve this issue. One potential solution is to conduct a fully counterbalanced experiment with speakers of languages such as Dutch, which use all of the required phonemes.

Another source of caution is our failure to replicate the variation in VOT reported in Experiment 3. We conducted an analysis on VOT that was designed to replicate the acoustic results from Experiment 3. In Experiment 4 we observed that variability in VOT was greatest when only voicing feature representations were competing compared to when voicing and place of articulation features were competing. However, this pattern was not significant. One possible account of this failure to replicate is our measurements of voicing were limited to VOT. A range of acoustic cues can be measured that represent differences in voicing for stop consonants, including: onset of the first formant (Summerfield & Haggard, 1977), amplitude of the burst (Repp, 1979), and post-obstruent vowel duration (Kessinger & Blumstein, 1998; G. E. Peterson & Lehiste, 1960). Moreover, it is well established that a slower speaking rate for voiceless stop constants yields more lengthening in the vowel duration than VOT duration (Kessinger & Blumstein, 1998; Miller, Green, & Reeves, 1986; Port, 1981; Volaitis & Miller, 1992). Since Experiment 4 was conducted at a slower speaking rate it is possible that vowel duration would be a more sensitive measure of acoustic variation. Note, however, the results of Goldrick and Blumstein's (2006) acoustic investigation of non-canonical errors revealed that the VOTs of errors and not the vowel durations were sensitive to competing representations.

The most striking result of Experiment 4 is that the ultrasound analysis designed to replicate the EPG analysis in Experiment 3 was successful. Both of these analyses used the Delta method for calculating how (dis)similar articulations are in conditions with competing representations relative to conditions without competing representations. The observation of a replication, using different stimulus items and different articulatory imaging techniques, suggests that the Delta method is a useful tool for investigating articulation.

In conclusion, this chapter provides an extension of theoretical and methodological contributions from previous chapters. Using ultrasound we established further evidence for the role of features in models of speech production. Moreover, this evidence supports models that include feedback between featural and phonological

representations. Lastly, the results of the Delta method demonstration and Experiment 4 suggest that the Delta method can be used to investigate variability across different articulatory imaging techniques.

7.7 Chapter Summary

This chapter extended the articulatory research presented in earlier chapters to a different articulatory imaging technique. We demonstrated that the Delta method can be used for ultrasound data analysis. The Delta method is useful because it can be used to perform relative comparisons of ultrasound recordings without relying on tongue contour tracing, categorisation or spatio-temporal assumptions. Importantly, in this chapter we presented a replication of the phonological similarity effects observed in Chapter 6. The replication provides firm evidence for a role for phonological features in a cascading model of production that includes feedback. Moreover, it establishes that the Delta method can be used across articulatory imaging techniques.

Chapter 8

Conclusions

The experimental work presented in this thesis was designed to investigate the consequences of assuming a cascading model of speech production. We focused on whether feedback or feedforward interactivity is required and whether feature representations must be incorporated. In this chapter we discuss the theoretical implications for models of production, the methodological implications for speech error research and avenues for future research.

8.1 Theoretical Implications

Phonological speech error investigations have provided a primary source of evidence for psychological (e.g., Garrett, 1975) and linguistic (e.g., Fromkin, 1971) models of speech production. However, these errors have traditionally been viewed as canonical substitutions of one phonological representation by another. More recently, researchers have demonstrated with articulatory and acoustic investigations that phonological errors can also be non-canonical (Frisch, 2007; Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein et al., 2007; Laver, 1980; Mowrey & MacKay, 1990; Pouplier, 2003, 2007; Stearns, 2006). An utterance may contain properties of both the intended phoneme and a competing phoneme. The transcription investigation reported in Chapter 3 adds to the growing articulatory evidence challenging the traditional view of speech errors. An articulatory transcription of EPG records revealed a high proportion of 'other' errors which included properties of both intended and competitor phonological representations.

The occurrence of non-canonical errors presents a problem for stage-based models of production. Specifically, staged models of production have attributed canonical errors to the misselection of a competing representation (e.g., Dell, 1986; Levelt et al., 1999) and non-canonical errors to a breakdown in articulatory implementation

(e.g. Laver, 1980; Levelt et al., 1999). Throughout this thesis we have assumed an alternative account for speech errors based on a cascading model of speech production. A cascading model allows partially activated phonological representations to cascade to articulation (Goldrick & Blumstein, 2006). As a consequence articulation can reflect continuous levels of activation of competing phonological representations. Importantly, a cascading model can provide a parsimonious account of both canonical and non-canonical speech errors: a canonical error reflects high activation of a competitor phonological representation; a non-canonical error reflects activation of both intended and competitor phonological representations.

A cascading model which incorporates feedback between phonological and lexical representations provides a straightforward account for influences of lexical competitors on articulation observed in Chapters 3 and 5. Feedback models posit that the activation of phonological representations feeds back to activate lexical representations. This feedback flow of information increases the activation of target representations through reinforcement and additionally yields activation of competitor representations (Dell, 1986; Hartsuiker et al., 2005; Humphreys, 2002). In cases where erroneously activated phonological representations can yield real words, lexical competitors become activated, but in cases where phonological representations can not yield real words there is no increase in competitor activation. The finding in Chapter 5 that targets with real word competitors, as opposed to non-word competitors, are articulated more similarly to the competitor representation clearly fits into a feedback account. Additional support for feedback comes from the observed correlation of articulatory variation with neighbourhood size, and the observation in Chapter 3 of a context-independent lexical bias effect for canonical substitution errors.

In addition to phonological to lexical feedback, a cascading model which incorporates feature to phonological feedback can account for the phonological similarity influences on articulation reported in Chapters 6 and 7. According to a feedback account, activation from feature representations feeds back to reinforce the activation of phonological representations. The more features that competing phonological representations have in common, the more reinforcement the competitor phonological representation will receive (Dell, 1986; Dell et al., 1993; Stemberger, 1982, 1985a).

Together, the evidence for feedback between phonological and lexical representations and for feedback between feature and phonological representations can be

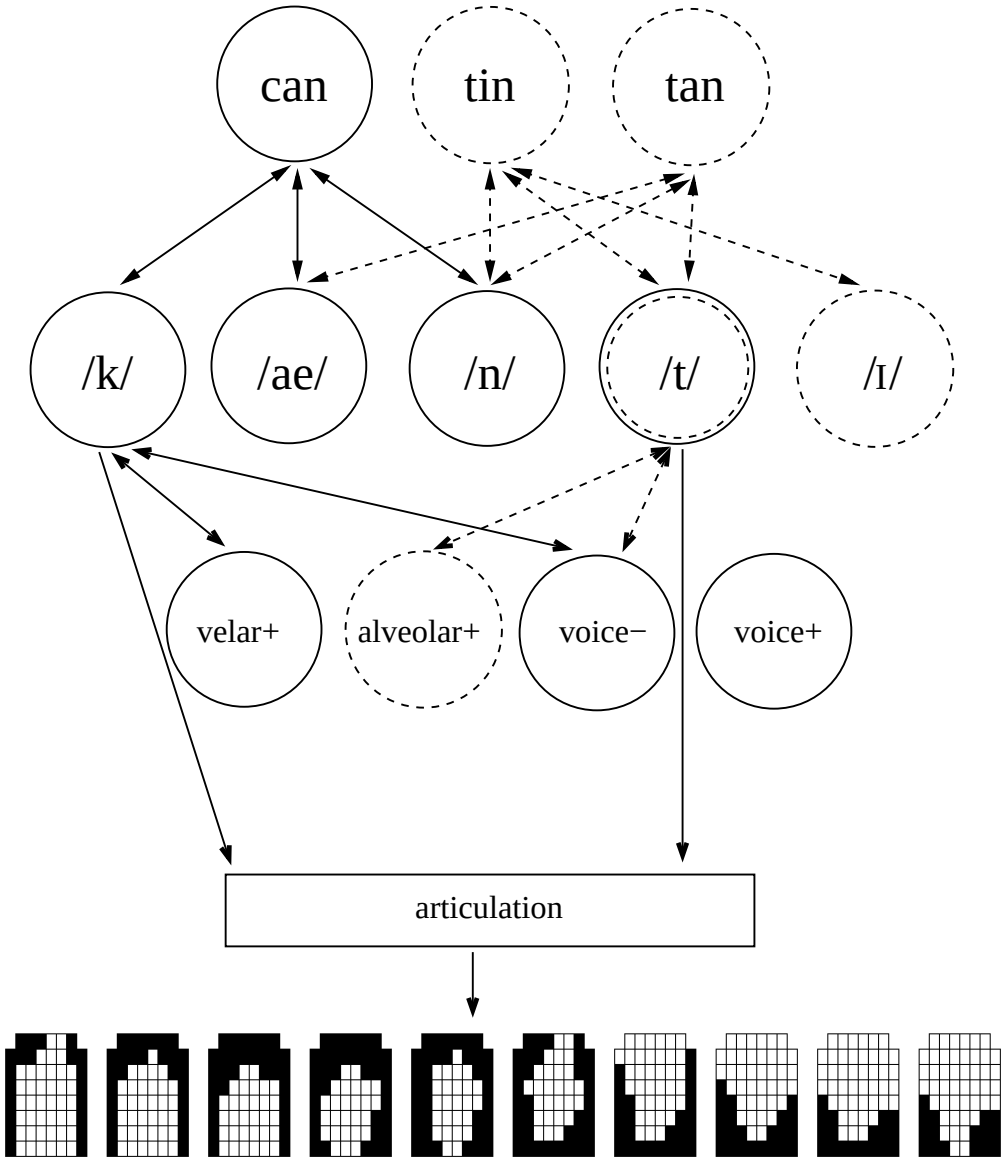


Figure 8.1: A cascading model of production with feedback between phonological and lexical representations and feedback between feature and phonological representations: solid lines represent strong activation, dotted lines represent weak activation, and a dotted line within a solid line represents initially weak activation that was reinforced and became strong. Only strong activation can cascade to articulation. In this model the target utterance is *can* from the phrase "tin can". Activation from CAN spreads to the corresponding phonological representations (/k/, /æ/ and /n/) and TIN, which is weakly activated, yields weak activation of its phonological representations (/t/, /ɪ/, and /n/). Activation of the phonological representations then feedback to weakly activate TAN. Since TAN is weakly activated, the activation of /t/ is reinforced. Both /k/ and /t/ then activate their corresponding features and, because they share the <voice-> feature, /t/ activation becomes further activated through feedback reinforcement. Finally, the strongly activated phonological representations /k/ and /t/ cascade to articulation. Since both representations cascade, the articulatory signal contains properties of both phonological representations.

integrated into a cascading model of production. In Figure 8.1 we present an illustration of a cascading model for the intended utterance *can* in the phrase *tin can*. The model follows the same rules as previously described models (see Figures 2.1 and 6.1): solid lines represent strong activation; dotted lines represent weak activation; a dotted line within a solid line represents initially weak activation that was reinforced and became strong; and only strong activation can cascade to articulation.

In the model, the representation CAN receives an initial jolt of activation which spreads to its corresponding phonological representations (/k/, /æ/ and /n/). The activation of the lexical representation TIN is also weakly activated since it is part of the intended phrase. Since TIN is weakly activated, its corresponding phonological representations (/t/, /ɪ/ and /n/) are also weakly activated. The activation of /t/, /æ/ and /n/ then feeds back to activate the lexical representation TAN, which in turn reinforces the activation of /t/.

In addition to activation flowing between the phonological and lexical levels, activation flows between the phonological and feature levels. Since /k/ and /t/ are phonologically similar, they both activate the <voice–> feature representation. The <voice–> representation feeds back to further reinforce activation of /t/. Since /k/ and /t/ have received strong activation, both representations cascade to articulation[1]. As a result the articulatory output of the model includes velar and alveolar closure patterns.

While this model can account for the lexical bias effect and phonological similarity evidence reported in this thesis, three components of the model require further investigation. First, we have assumed that representations can cascade to articulation, however, there is some evidence which suggests that planning is staged (see, for example, Damian, 2003; Damian & Dumay, 2007). Second, we have assumed that representations below the level of the phoneme are featural rather than gestural. Third, the current model does not include any form of monitoring, which may be required to account for the context effects observed in Chapter 5. In Section 8.3 we identify future directions for research to address these points. But, first, we discuss the methodological implications for our research.

[1]In this model, phonological representations cascade to articulation, but it is also possible to modify the model so that feature representations cascade instead. This would be achieved by allowing the strong activation from phonological representations to activate their corresponding feature representations. For example, strong activation from /k/ and /t/ would yield strong activation of <velar+,alveolar+,voice–>. Therefore, a model which allowed cascading feature representations could yield the same output as the model illustrated in Figure 8.1.

8.2 Methodological Implications

One of the most significant contributions of the work discussed in this thesis is that we reported variability in articulation, rather than categorising responses as "errorful" or "correct". In experiments designed to elicit speech errors, EPG and ultrasound recordings were indiscriminately assigned *a priori* to the 'error' category and therefore all items were included in the analysis. In order to measure the degree to which they deviated from 'normal' articulations, they were compared to averaged onset phonemes obtained under comparable but (by definition) non error-producing circumstances. This approach has several clear advantages over more traditional methods of investigating speech production.

The primary advantage, and main motivation, of this approach is that responses do not have to be assigned to discrete categories. The categorisation of responses suffers from the fact that any boundary created between categories must be arbitrary. As a result, an assumption must be made about what articulatory properties constitute each category. Based on our observations of articulatory patterns in Chapter 3, it is not clear how such boundaries should be defined. Moreover, given the well established evidence for categorical perception (between boundary phonemes are perceived as belonging to one category or another; Liberman, 1997), speech error investigations which rely on auditory transcription are further restricted by categorisation. Responses including the acoustic properties of two phonological representations may only be perceived as having the properties of one phonological representation.

An additional advantage of the approach is analyses can include all recorded responses rather than a potentially noisy subset. Nooteboom and Quené (in press) recently noted that studies which rely on categorisation of speech errors regularly suffer from sparse data, rendering statistical analyses unreliable. This criticism is clearly relevant to the transcription analysis presented in Chapter 3; although we presented the analysis for comparison with previous work, we noted that only 1.1% of transcribed items were entered into the analysis. This is equivalent to only approximately 1 'error' out of 96 responses for each of 47 participants. By focusing on variation, we were able to include 599 word onsets, equal to approximately 85 out of 96 responses for each of 7 participants. Relatedly, the Delta method takes all aspects of each articulatory recording into account: for EPG, analysis is not restricted to a pre-defined region on the artificial palate; and for ultrasound, the entire video image is included in analyses which allows the data to be analysed

to its fullest potential without the intermediate reductionist step of tracing tongue contours or identifying arbitrary reference points.

Perhaps the most important consequence of analysing articulatory variation, however, is that it provides a direct link between the cognitive and the resulting motor movements that produce the speech. Like other newly-emerging paradigms such as mouse tracking (see Spivey, Richardson, & Dale, in press, for a review), articulatory analysis shows that motor movements can give us a fine-grained insight into the cognitive processes that drive them.

A potential source of caution in interpreting the present results is that the speech analysed in this thesis was obtained using 'error' elicitation tasks. This methodology might lead one to question whether the variation we report is 'representative'. Essentially, there are two answers to this question. First, observations of acoustically or articulatory deviant speech have been reported using a variety of laboratory methods including repetition tasks (Goldstein et al., 2007), a SLIP task (Pouplier, 2007), the WOC task (see Chapters 3 and 5) and tongue-twisters (Frisch & Wright, 2002; Goldrick & Blumstein, 2006, see also Chapters 6 and 7). Similar articulatory variation has been observed where the experiment was not designed to elicit errors (Boucher, 1994). Second, even if the systematic differences we report here were not to be found in everyday speech, we believe that they constrain the set of potential models that can be used to account for speech production.

8.3 Directions for Future Research

Articulatory imaging techniques have been used for decades to investigate the phonetic detail of speech (Ladefoged, 1957; MacMillan & Kelemen, 1952; Perkell et al., 1992). However, it has only been recently that articulatory and acoustic investigations have emerged to investigate psychological aspects of speech production (e.g., Frisch, 2007; Goldrick & Blumstein, 2006). This exciting new approach opens several future directions for psycholinguistic research. In this section we address three avenues for future research that provide a direct extension of the research reported here.

8.3.1 Cascading in Production

Throughout this thesis we have assumed a model of speech production which allows partial activation of phonological representations to cascade to articulation. We

demonstrated that a cascading model has to incorporate feedback during phonological encoding. While the research reported is consistent with a cascading account, future work is required to determine whether representations activated during phonological encoding can cascade to articulation.

An approach that will prove to be useful for investigating staged versus cascading models of production is computational modelling. Computational models have formed the foundation for several accounts of speech error patterns (e.g., Dell, 1986). To investigate whether cascading is required for the theoretical account presented in Section 8.1, it would be useful to develop a model that accounts for articulatory variation rather than speech errors. Preliminary work by Moat, Hartsuiker, and Corley (2007) has focused on how to formalise such a model. The aim of their work is to conduct systematic investigations of cascading and staged model behaviour and to quantify the probability of each model given the VOT data observed by Goldrick and Blumstein (2006). Preliminary results from the model suggest that a cascading model is much more likely to account for Goldrick and Blumstein's (2006) non-canonical VOT data than a staged model of production.

8.3.2 Is there a role for monitoring?

The occurrences of non-canonical errors raise the possibility that monitoring may be required to attribute the errors to some form of rapid repair. The model proposed in Section 8.1 does not include a monitoring mechanism. While the model can account for the observed lexical bias and phonological similarity effects, it may be necessary to extend the model to incorporate some form of monitoring. In particular monitoring may be required to, at least partially, account for the context influence on articulation observed in Chapter 5. Moreover, several sources of evidence, such as reports of rapid repair (Blackmer & Mitton, 1991; Levelt, 1983, 1989), suggest that speakers monitor their speech plan.

To investigate the potential role for monitoring in production future research must reevaluate the traditional definitions of repair. In Chapter 5 we highlighted that precise measurements of articulation make criteria such as the "0ms repair" (Blackmer & Mitton, 1991) less tenable. Moreover, the distinction between "overt" and "covert" repairs (Levelt, 1983) becomes unclear: a double articulation is detectable with articulatory imaging techniques, but may not be detectable by listeners. Lastly,

it is not clear whether a double articulation, if interpreted as a repair, reflects "internal" or "external" monitoring. Focusing on the time course of articulatory variation may be the most fruitful approach for reevaluating repairs. By investigating temporal variability, models can be refined to more accurately simulate the speech production process (Hartsuiker & Kolk, 2001; Levelt et al., 1999).

8.3.3 The nature of lower level representations

The research reported throughout this thesis has focused on processing during speech production and has not addressed the specific nature of the representations involved in speaking. The phonological similarity evidence reported in Chapters 6 and 7 establishes that lower-level representations are required in a cascading model of production and that feedback is required between lower-level representations and phonological representations. Our discussion was agnostic with respect to the nature of the lower-level representations, though we discussed representations in terms of features.

The representations which feed back to phonological representations could be featural in nature. Feature representations are abstract units which represent contrast between phonological representations (e.g., Chomsky & Halle, 1968). However, since features do not specify motor commands some additional "action unit" may be required in the proposed model. For example, a full model would have to account for the way in which representations are integrated together to create continuous speech through, for example, coarticulation.

The lower-level representations could also be gestural "action units" which specify spatio-temporal motor commands for articulation (e.g., Browman & Goldstein, 1989). The articulatory speech error evidence reported by Goldstein et al. (2007, see also Pouplier, 2003, 2007) has been explicitly used to argue for gestural units in production. Additionally, Levelt et al.'s (1999) model provides a basic account of articulatory processing including gesture representations to guide speech-motor control. However, given the processing constraint of feedback presented in this thesis, a gestural model must incorporate feedback between gestural representations and segmental representations to be successful.

Goldrick (2004, see also Goldrick & Blumstein, 2006) has suggested that speech error investigations cannot discern between the nature of lower-level representations. We align with this view since both a featural and gestural model can account for the observed findings. Indeed the features (<voice–>, <voice+>,

<alveolar+>, <velar+>) that we discussed in previous chapters directly map onto gestures (<glottal aperture open>, <glottal aperture closed>,<tongue tip constriction>,<tongue-dorsum constriction>). While it is not possible to conclude from the current work about the nature of representations, future investigations on the articulatory and acoustic properties which can integrate the psychological and linguistic aspects of production may be able to resolve this complex question.

8.4 Conclusions

Investigating the articulatory and acoustic variation of speech, without using categorisation, allows us to investigate the consequence of cascading activation for models of speech production. A cascading model of production requires feedback between phonological and lexical representations. A cascading model also requires lower level (e.g., feature) representations which feed back to phonological representations.

Appendix A

Experiment 1 Stimulus Items

Real Word Competitors	Nonword Competitors
gim dulp	gib dulm
gome dasp	gofe dasb
gope doof	gobe doove
dap gime	dalf gipe
dape gam	dabe galf
dulf gamp	duf galve
tave gub	tafe gup
tum gop	tup golve
timp giff	tib gilf
garp tiv	garm tirve
guff tob	gulb tov
gube tolf	goove tolm
keff darve	kem darf
kip doff	kiv dolf
koom darp	coob dalp
dop kuv	dolb kulve
duff cump	dulve culp
dup kive	dulp kife
toop kerm	toove kurp
tove kemp	tofe keb
tome kipe	tobe kive
keam turve	keeb turp
curf talm	kerp talb
kime turb	kibe turp

Appendix B

Experiment 3 Stimulus Items

Tongue Twister	Place	Voice
duv duv duv duv	-	-
giv giv giv giv	-	-
kef kef kef kef	-	-
tuf tuf tuf tuf	-	-
gef kef kef gef	-	+
gev kev kev gev	-	+
gif kif kif gif	-	+
giv kiv kiv giv	-	+
guf kuf kuf guf	-	+
guv tuv tuv guv	-	+
kef gef gef kef	-	+
kev gev gev kev	-	+
kif gif gif kif	-	+
kiv giv giv kiv	-	+
kuf guf guf kuf	-	+
kuv guv guv kuv	-	+
def gef gef def	+	-
dev gev gev dev	+	-
dif gif gif dig	+	-
div giv giv div	+	-
duf guf guf duf	+	-
duv guv guv duv	+	-
gef def def gef	+	-
gev dev dev gev	+	-
gif dif dif gif	+	-

Continued on next page

Table B.1 – continued from previous page

Tongue Twister	Place	Voice
giv div div giv	+	-
guf duf duf guf	+	-
guv duv duv guv	+	-
kef tef tef kef	+	-
kev tev tev kev	+	-
kif tif tif kif	+	-
kiv tiv tiv kiv	+	-
kuf tuf tuf kuf	+	-
kuv tuv tuv kuv	+	-
tef kef kef tef	+	-
tev kev kev tev	+	-
tif kif kif tif	+	-
tiv kiv kiv tiv	+	-
tuf kuf kuf tuf	+	-
tuv kuv kuv tuv	+	-
def kef kef def	+	+
dev kev kev dev	+	+
dif kif kif dif	+	+
div kiv kiv div	+	+
duf kuf kuf duf	+	+
duv kuv kuv duv	+	+
gef tef tef gef	+	+
gev tev tev gev	+	+
gif tif tif gif	+	+
giv tiv tiv giv	+	+
guf tuf tuf guf	+	+
guv tuv tuv guv	+	+
kef def def kef	+	+
kev dev dev kev	+	+
kif dif dif kif	+	+
kiv div div kiv	+	+
kuf duf duf kuf	+	+
kuv duv duv kuv	+	+
tef gef gef tef	+	+

Continued on next page

Table B.1 – continued from previous page

Tongue Twister	Place	Voice
tev gev gev tev	+	+
tif gif gif tif	+	+
tiv giv giv tiv	+	+
tuf guf guf tuf	+	+
tuv guv guv tuv	+	+

Appendix C

Experiment 4 Stimulus Items

Tongue Twister	Place	Voice	Manner
dom dom dom dom*	–	–	–
gom gom gom gom*	–	–	–
kom kom kom kom*	–	–	–
som som som som	–	–	–
tom tom tom tom*	–	–	–
zom zom zom zom	–	–	–
dom zom zom dom	–	–	+
som tom tom som	–	–	+
tom som som tom	–	–	+
zom dom dom zom	–	–	+
dom tom tom dom*	–	+	–
gom kom kom gom*	–	+	–
kom gom gom kom*	–	+	–
som zom zom som	–	+	–
tom dom dom tom*	–	+	–
zom som som zom	–	+	–
dom som som dom	–	+	+
som dom dom som	–	+	+
tom zom zom tom	–	+	+
zom tom tom zom	–	+	+
dom gom gom dom*	+	–	–
gom dom dom gom*	+	–	–
kom tom tom kom*	+	–	–
tom kom kom tom*	+	–	–
gom zom zom gom	+	–	+

Continued on next page

Table C.1 – continued from previous page

Tongue Twister	Place	Voice	Manner
kom som som kom	+	–	+
som kom kom som	+	–	+
zom gom gom zom	+	–	+
dom kom kom dom*	+	+	–
gom tom tom gom*	+	+	–
kom dom dom kom*	+	+	–
tom gom gom tom*	+	+	–
gom som som gom*	+	+	+
kom zom zom kom	+	+	+
som gom gom som	+	+	+
zom kom kom zom	+	+	+

* denotes the item was included in the replication analyses

References

Atal, B. S., Chang, J. J., Mathews, M. V., & Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique. *Journal of the Acoustical Society of America*, *63*, 1535–1555.

Baars, B. J. (1992). A dozen competing-plans techniques for inducing predictable slips in speech and action. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 129–150). New York: Plenum Press.

Baars, B. J., & MacKay, D. G. (1978). Experimentally eliciting phonetic and sentential speech errors: Methods, implications, and work in progress. *Language in Society*, *7*, 105–109.

Baars, B. J., & Motley, M. T. (1976). Spoonerisms as sequencer conflicts: Evidence from artificially elicited errors. *American Journal of Psychology*, *89*(3), 467–484.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behaviour*, *14*, 382–391.

Baayen, R. H. (in press). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge University Press.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*, 362–365.

Barry, M. (1985). A palatographic study of connected speech processes. *Cambridge Papers in Phonetics and Experimental Linguistics*, *4*, 1–16.

Bates, D., & Sarkar, D. (2007). *Lme4: Linear mixed-effects models using S4 classes.*

Blackmer, E., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*, 173–194.

Bock, K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575–586.

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review*, *3*, 395–421.

Boersma, P., & Weenink, D. (2006). *Praat: doing phonetics by computer (version 4.5.01).* Retrieved October 28, 2006, from http://www.praat.org/.

Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication*, *3*, 2–12.

Boucher, V. J. (1994). Alphabet-related biases in psycholinguistic inquiries: Considerations for direct theories of speech production and perception. *Journal of Phonetics*, *22*(1), 1–18.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 201–251.

Buckingham, H., & Yule, G. (1987). Phonemic false evaluation: Theoretical and clinical aspects. *Clinical Linguistics and Phonetics*, *1*, 113–125.

Butterworth, B., & Whittaker, S. (1980). Peggy babcock's relatives. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 647–656). Amsterdam: North-Holland Publishing Company.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*, 209–244.

Byrd, D., Flemming, E., Mueller, C. A., & Tan, C. C. (1995). Using regions and indices in EPG data reduction. *Journal of Speech and Hearing Research*, *38*, 821–827.

Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika*, *48*, 343–349.

Celex English database - release e25 [on-line]. (1993). Available: Nijmegen: Centre for Lexical Information [Producer and Distributor].

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper and Row.

Cox, T. F., & Cox, M. A. A. (1994). *Multidimensional scaling.* London: Chapman and Hall.

Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 109–162). Berlin: Walter de Gruyter/Mouton.

Cutler, A. (1982). The reliability of speech error data. In A. Cutler (Ed.), *Slips of the tongue and language production.* Berlin: Walter de Gruyter/Mouton.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 416–431.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*, 195–209.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, *20*, 445-464.

Davidson, L. (2004). *Assessing tongue shape similarity: Comparing norms, area, and average distance.* Talk presented at Ultrafest II. Vancouver.

del Viso, S., Igoa, J. M., & Garcia-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research*, *20*, 161–185.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124–142.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*, 313–349.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and model. *Psychological Review*, *104*, 123–147.

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Similarities and differences.* Berlin: Mouton de Gruyter.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*, 149–195.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review*, *98*, 604–614.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in speech production. *Cognition*, *42*, 287–314.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*, 611–629.

Dell, G. S., & Repka, R. J. (1992). Errors in inner speech. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 237–262). New York: Plenum.

Ferber, R. (1991). Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research*, *20*, 105–22.

Flege, J. E. (1986). Plasticity in adult and child speech production. *Journal of the Acoustical Society of America*, *79*, S54.

Fletcher, S. (1989). Palatometric specification of stop, affricate and sibilant sounds. *Journal of Speech and Hearing Research*, *32*, 736–748.

Friel, S. (1998). When is a /k/ not a [k]? EPG as a diagnostic and therapeutic tool for abnormal velar stops. *International Journal of Language and Communication Disorders*, *33*(Suppl), 439–44.

Frisch, S. A. (1996). *Similarity and frequency in phonology.* Unpublished doctoral dissertation, Northwestern University.

Frisch, S. A. (2007). Walking the tightrope between cognition and articulation: The state of the art in the phonetics of speech errors. In C. T. Schütze & V. S. Ferreira (Eds.), *The state of the art in speech error research* (pp. 155–172). MIT Working Papers in Linguistics, Volume 53).

Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*, 139–162.

Fromkin, V. A. (1968). Speculations on performance models. *Journal of Linguistics*, *4*, 47–68.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27–52.

Fromkin, V. A. (Ed.). (1973). *Speech errors as linguistic evidence.* The Hague: Mouton.

Fromkin, V. A. (1980). Introduction. In V. A. Fromkin (Ed.), *Errors in linguistic performance.* New York: Academic Press.

Fry, D. B. (1977). *Homo loquens: Man as a talking animal.* London: Cambridge University Press.

Garnham, A., Shillcock, R., Brown, G. D. A., Mill, A. I. D., & Cutler, A. (1981). Slips of the tongue in the London–Lund corpus of spontaneous speech. *Linguistics*, *19*, 805–817.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 133–175). San Diego: Academic Press.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–256). Amsterdam: North Holland Publishing Company.

Garrett, M. F. (1979). Levels of processing in sentence production. In B. L. Butterworth (Ed.), *Language production* (Vol. Volume 1: Speech and Talk). New York: Academic Press.

Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, *51*, 586–603.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649 - 683.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*, 386–412.

Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language*, *82*, 113-145.

Hardcastle, W. J., & Gibbon, F. (1997). Instrumental clinical phonetics. In M. J. Ball & C. Code (Eds.), (pp. 149–193). London: Croom Helm.

Hardcastle, W. J., Gibbon, F., & Nicolaidis, K. (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics*, *19*(3), 251–266.

Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*, *8*, 291–309.

Harley, T. A., & Brown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology*, *89*, 151–174.

Harshman, R., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, *62*, 693–713.

Hartley, T., & Houghton, G. (1996). A linguistically restrained model of short-term memory for non-words. *Journal of Memory and Language*, *35*, 1–31.

Hartsuiker, R. J. (2006). Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes, 21, 856-891*, *21*, 856-891.

Hartsuiker, R. J., Antón-Méndez, I., Roelstraete, B., & Costa, A. (2006). Spoonish spanerisms: A lexical bias effect in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 949–953.

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). *Journal of Memory and Language*, *52*, 58–70.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*(2), 113–157.

Hedrick, W. R., Hykes, D. L., & Starchman, D. E. (1995). *Ultrasound physics and instrumentation* (3rd ed.). St. Louis, MO: Mosby.

Hewlett, N., Gibbon, F., & Cohen-McKenzie, W. (1998). When is a velar and alveolar? Evidence supporting a revised psycholinguistic model of speech production in children. *International Journal of Language and Communication Disorders*, *33*(2), 161–176.

Hoole, P. (1999). On the lingual organization of the german vowel system. *Journal of the Acoustical Society of America*, *106*, 1020–1032.

Humphreys, K. R. (2002). *Lexical bias in speech errors*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Jackson, M. (1988). Analysis of tongue positions: Language-specific and cross-linguistic models. *Journal of the Acoustical Society of America*, *84*, 124–143.

Kawamoto, A. H., Kello, C. T., Higareda, I., & Vu, J. V. Q. (1999). Parallel processing and initial phoneme criterion in naming words: Evidence from frequency effects on onset and rime duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 362–381.

Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initital phoneme duration. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 862–885.

Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 719–750.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task-dependence of staged versus cascaded processing: An empircal and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General*, *129*, 340–360.

Kessinger, & Blumstein, S. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, *26*, 117–128.

Kupin, J. J. (1982). *Tongue-twisters as a source of information about speech production*. Bloomington, USA: Indiana University Linguistics Club.

Ladefoged, P. (1957). Use of palatography. *Journal of Speech and Hearing Disorders*, *22*, 764–774.

Landsheera, J. A., van den Wittenboerb, G., & Maassena, G. H. (2006). Additive and multiplicative effects in a fixed 2 × 2 design using ANOVA can be difficult to differentiate: Demonstration and mathematical reasons. *Social Science Research*, *35*, 279–294.

Laver, J. (1980). Slips of the tongue as neuromuscular evidence for a model of speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler.* The Hague: Mouton.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, *22*, 1–75.

Levelt, W. J. M., Schriefer, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*, 122-142.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 122–142.

Levitt, A. G., & Healy, A. F. (1985). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory and Language*, *24*, 717–733.

Li, M., Kambhamettu, C., & Stone, M. (2005). Automatic contour tracking in ultrasound images. *International Journal of Clinical Linguistics and Phonetics*, *19*, 545–554.

Liberman, A. M. (1997). *Speech: A special code.* Cambridge: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–460.

Liker, M., & Gibbon, F. (2007). EPG characteristics of velar stops in normal adult English speakers. In *Proceedings of the 16th ICPhS.* Saarbrücken.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295–327.

Lundberg, A., & Stone, M. (1999). Three-dimensional tongue surface reconstruction: Practical considerations for ultrasound imaging. *Journal of the Acoustical Society of America*, *105*, 2858–2867.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, *8*, 323–350.

MacKay, D. G. (1980). Speech errors: Retrospect and prospect. In V. A. Fromkin (Ed.), *Errors in linguistic performance.* New York: Academic Press.

MacKay, D. G. (1982). Problems with flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, *89*, 483–506.

MacKay, D. G. (1987). *The organisation of perception and action.* New York: Springer.

MacMillan, A. S., & Kelemen, G. (1952). Radiography of the supraglottic speech organs; A survey. *Archives of Otolaryngology*, *55*, 671–688.

Maeda, S. (1990). Speech production and speech modeling. In W. Hardcastle & A. Marchal (Eds.), (pp. 131–150). Dordrecht: Kluwer Academic Publishers.

Marchal, A. (1988). Coproduction: Evidence from EPG data. *Speech Communication*, *7*, 287–295.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

McCutchen, D., Bell, L. C., France, I. M., & Perfetti, C. A. (1991). Phoneme-specific interference in reading: The tongue-twister effect revisited. *Reading Research Quarterly*, *26*, 87–103.

McLeod, S. (2006). Australian adults' production of /n/: An EPG investigation. *Clinical Linguistics and Phonetics*, *20*, 99–107.

McMillan, C. T., Corley, M., & Lickley, R. (in press). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes.*

Meringer, R., & Mayer, C. (1895). *Versprechen und verlesen: eine psychologisch-linguistiche studie.* Stuttgart: G. J. Göschen'sche Verlagshandlung.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of succesive syllables. *Journal of Memory and Language*, *29*, 524–545.

Meyer, A. S. (1992). Investigations of phonological encoding through speech error analyses: Achievement, limitations, and alternatives. *Cognition*, *42*, 181–211.

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*, 131–147.

Michi, K., Suzuki, N., Yamashita, Y., & Imai, S. (1986). Visual training and correction of articulation disorders by use of dynamic palatography: Serial observation in a case of cleft palate. *Journal of Speech and Hearing Disorders*, *51*, 226–238.

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, *43*, 106–115.

Moat, S., Hartsuiker, R. J., & Corley, M. (2007). *Cascading from phonological encoding to articulation? A computational investigation.* (Poster presented at AMLaP'07 Conference)

Morrish, K., Stone, M., Shawker, T., & Sonies, B. C. (1985). Distinguishability of tongue shape during vowel production. *Journal of Phonetics*, *13*, 189–204.

Morrish, K., Stone, M., Stonies, B., Kurtz, D., & Shawker, T. (1984). Characterization of tongue shape. *Ultrasonic Imaging*, *6*, 37–47.

Motley, M. T., Baars, B. J., & Camden, C. T. (1981). Syntactic criteria in prearticulatory editing: Evidence from laboratory-induced slips of the tongue.journal of psycholinguistic research, 10, 503522. *Journal of Psycholinguistic Research*, *10*, 503–522.

Motley, M. T., Baars, B. J., & Camden, C. T. (1983). Experimental verbal slip studies: A review and an editing model of language encoding. *Communication Monographs*, *50*, 79–101.

Mowrey, R. A., & MacKay, I. R. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, *88*(3), 1299–1312.

Nolan, F. (1992). Papers in laboratory phonology II: Gesture, segment, prosody. In G. J. Docherty & D. R. Ladd (Eds.), (pp. 261–289). Cambridge: Cambridge University Press.

Nooteboom, S. G. (1969). Speech errors slip into patterns. In A. J. van Essen & A. A. van Raad (Eds.), *Leyden studies in linguistics and phonetics* (pp. 114–132). The Hague: Mouton.

Nooteboom, S. G. (2005a). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication*, *47*(1–2), 43–58.

Nooteboom, S. G. (2005b). Listening to one-self: Monitoring in speech production. In R. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech.* Hove, UK: Psychology Press.

Nooteboom, S. G., & Quené, H. (2007). The SLIP technique as a window on the mental preparation of speech: Some methodological considerations. In M. J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 339–350). Oxford: Oxford University Press.

Nooteboom, S. G., & Quené, H. (in press). Self-monitoring versus feedback: a new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language*.

Pérez, E., Santiago, J., Palma, A., & O'Seaghdha, P. G. (2007). Perceptual bias in speech error data collection: Insights from Spanish speech errors. *Journal of Psycholinguistic Research*, *36*(3), 207–235.

Perkell, J. S., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992). Electo-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, *92*, 3078–3096.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693–703.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*, 539–557.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons with and across phonetic categories. *Perception and Psychophysics*, *15*, 285–290.

Port, R. F. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America*, *69*, 262–274.

Pouplier, M. (2003). *Units of phonological coding: Empirical evidence*. Unpublished doctoral dissertation, Yale University.

Pouplier, M. (2004). An ultrasound investigation of speech errors. *Working Papers and Reports of the Vocal Tract Visualization Laboratory*, *6*, 1–17.

Pouplier, M. (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech*, *50*(3).

Pouplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics*, *33*, 47–75.

Pouplier, M., & Hardcastle, W. (2005). A re-evaluation of the nature of speech errors in normal and disordered speakers. *Phonetica*, *62*, 227–243.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–499.

Repp, B. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, *22*, 173–189.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. A., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology*, *24*, 121–133.

Roelofs, A. (1992). A spreading activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–42.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47*, 59–87.

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, *35*, 854–876.

Roelofs, A. (1997a). Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes*, *12*, 659–696.

Roelofs, A. (1997b, Sep). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249–284.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review*, *110*(1), 88–125.

Roelofs, A., & Hagoort, P. (2002, Dec). Control of language use: cognitive modeling of the hemodynamics of Stroop task performance. *Cognitive Brain Research*, *15*(1), 85–97.

Saito, S., & Baddeley, A. D. (2004). Irrelevant sound disrupts speech production: Exploring the relationship between short-term memory and experimentally induced slips of the tongue. *Quarterly Journal of Experimental Psychology: Section A*, *57*, 1309-1340.

Schiller, N. O., Costa, A., & Colomé, A. (2002). Phonological encoding of single words: In search of the lost syllable. In C. Gussenhoven & N. Warner (Eds.), *Papers in laboratory phonology* (pp. 35–59). Berlin: Mouton de Gruyter.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E prime 1.0.* Pittsburgh, PA: Psychological Software Tools.

Schwartz, M. F., Saffran, E. M., Bloch, D. E., & Dell, G. S. (1994). Disordered speech production in aphasic and normal speakers. *Brain and Language*, *47*, 52–88.

Searl, J., Evitts, P., & Davis, W. J. (2006). Perceptual and acoustic evidence of speaker adaptation to a thin pseudopalate. *Logopedics Phoniatrics Vocology*, *31*(3), 107–116.

Shallice, T., & Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia*, *15*, 729–735.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shattuck-Hufnagel, S. (1982). Three kinds of speech error evidence for the role of grammatical elements in processing. In L. Obler & L. Menn (Eds.), *Exceptional language and linguistics.* New York: Academic Press.

Shattuck-Hufnagel, S. (1983). The production of speech. In P. F. MacNeilage (Ed.), (pp. 109–136). New York: Springer-Verlag.

Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook*, *3*, 117–149.

Shattuck-Hufnagel, S. (1987). Motor and sensory processes of language. In E. Keller & M. Gopnik (Eds.), (pp. 17–51). Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (1992). The role of word order in segmental serial ordering. *Cognition*, *42*, 213–259.

Shattuck-Hufnagel, S., & Klatt, D. H. (1975). An analysis of 1500 phonetic errors in spontaneous speech. *Journal of the Acoustical Society of America*, *58*, 66 (A).

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*, 41–55.

Slud, E., Stone, M., Smith, P., & Goldstein, M. (2002). Principal components representation of the two–dimensional coronal tongue surface. *Phonetica*, *59*, 108–133.

Spivey, M., Richardson, D. C., & Dale, R. (in press). Movements of eye and hand in language and cognition. In E. Morsella & J. Bargh (Eds.), *The psychology of action, vol. 2.* New York: Oxford University Press.

Stearns, A. M. (2006). *Production and perception of place of articulation errors.* Unpublished master's thesis, University of South Florida.

Stemberger, J. P. (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua*, *56*, 235–259.

Stemberger, J. P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology*, *1*, 281–313.

Stemberger, J. P. (1985a). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 143–186). London: Erlbaum.

Stemberger, J. P. (1985b). The reliability and replicability of speech error data: A comparison with experimentally induced errors. *Research on Speech Perception Progress Report*, *11*, Bloomington, Indiana University.

Stemberger, J. P. (1991a). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language*, *30*, 161–185.

Stemberger, J. P. (1991b). Radical underspecification in language production. *Phonology*, *8*, 73–112.

Stemberger, J. P., & McWhinney, B. (1986). Frequency and the lexical storage of regulary inflected forms. *Memory & Cognition*, *14*, 17–26.

Stemberger, J. P., & Stoel-Gammon, C. (1991). The special status of coronals: Internal, and external evidence. In C. Paradis & J. F. Prunet (Eds.), (pp. 181–199). San Diego: Academic Press.

Stemberger, J. P., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language*, *25*, 163–180.

Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and X-ray beam data. *Journal of the Acoustical Society of America*, *87*, 2207–2217.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*, *19*, 455–502.

Stone, M., & Lundberg, A. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America*, *99*, 3728–3737.

Stone, M., Shawker, T., Talbot, T., & Rich, A. (1988). Cross-sectional tongue shape during the production of vowels. *Journal of the Acoustical Society of America*, *83*, 1586–1596.

Stone, M., Sonies, B., Shawker, T., Weiss, G., & Nadel, L. (1983). Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system. *Journal of Phonetics*, *11*, 207–218.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in intial stop consonants. *Journal of the Acoustical Society of America*, *62*, 435–448.

Tent, J., & Clark, J. E. (1980). An experimental investigation into the perception of slips of the tongue. *Journal of Phonetics*, *8*, 317–325.

Unser, M., & Stone, M. (1992). Automatic detection of the tongue-surface in sequences of ultrasound images. *Journal of the Acoustical Society of America*, *91*, 3001–3007.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, *128*(3), 442–472.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 735–747.

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighbourhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*(4), 491–504.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, *92*, 723–735.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, *41*, 101–175.

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 387–398.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*, 392–393.

Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine*, *3*, 9–30.

Wilshire, C. E. (1999). The "tongue twister" paradigm as a technique for studying phonological encoding. *Language and Speech*, *42*(1), 57–82.

Wrench, A. (2003). *Articulate assistant user guide, version 1.3a.* http://www.articuleinstruments.com.