Running head: CASCADING INFLUENCES ON THE PRODUCTION OF SPEECH

Cascading Influences on the Production of Speech: Evidence from Articulation

Corey McMillan

University of Pennsylvania Medical Centre

Department of Neurology

Martin Corley

University of Edinburgh

School of Philosophy, Psychology, and Language Sciences

Martin Corley

Psychology

School of Philosophy, Psychology, and Language Sciences

University of Edinburgh

Edinburgh EH8 9JZ, UK

(tel) +44 131 650 6682; (fax) +44 131 650 3461; Martin.Corley@ed.ac.uk

Abstract

Recent investigations have supported the suggestion that phonological speech errors may reflect the simultaneous activation of more than one phonemic representation. This presents a challenge for speech error evidence which is based on the assumption of well-formedness, because we may continue to perceive well-formed errors, even when they are not produced. To address this issue, we present two tongue-twister experiments in which the articulation of onset consonants is quantified and compared to baseline measures from cases where there is no phonemic competition. We report three measure of articulatory variability: changes in tongue-to-palate contact using electropalatography (EPG, Experiment 1), changes in midsagittal spline of the tongue using ultrasound (Experiment 2), and acoustic changes manifested as voice-onset-time (VOT). These three sources provide converging evidence that articulatory variability increases when competing onsets differ by one phonological feature, but the increase is attenuated when onsets differ by two features. This finding provides clear evidence, based solely on production, that the articulation of phonemes is influenced by cascading activation from the speech plan.

Cascading Influences on the Production of Speech: Evidence from Articulation

A long tradition of psycholinguistic research has maintained that the words we produce are occasionally affected by the insertions, deletions, or substitutions of well-formed phonemes (e.g., Dell, 1986; Garrett, 1980; Meringer & Mayer, 1895/1978; Shattuck-Hufnagel & Klatt, 1979). Based on this assumption, the patterns with which such errors occur have been used to determine further properties of the language production system. Substitutions, for example, are more likely to occur when there is *phonemic similarity* between the phoneme that is intended by the speaker and the phoneme that is eventually produced (Dell & Reich, 1981; Butterworth & Whittaker, 1980; Kupin, 1982; Levitt & Healy, 1985; MacKay, 1970; MacKay, 1980; Nooteboom, 2005a, 2005b; Shattuck-Hufnagel, 1986; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1982, 1985; del Viso, Igoa, & Garcia-Albea, 1991; Vousden, Brown, & Harley, 2000; Wilshire, 1999). One interpretation of this effect is that the production of phonemes in speech is influenced by the activation of subsegmental representations, such as phonological features, prior to articulation. As a consequence of feedback from these feature-level representations, misactivated phonemes which share features with an intended phoneme are likely to accrue activation through reinforcement (Dell, 1986). The phonemes which have the highest level of activation after a set period of time are selected and used to drive the process of articulation (Dell, 1986; cf. Levelt, Roelofs, & Meyer, 1999; Shattuck-Hufnagel & Klatt, 1979).

However, recent evidence has challenged the view that the articulatory plan is driven by selected phonemic representations. Articulatory and acoustic investigations have shown that many speech errors do not necessarily consist of simple substitutions of one phoneme for another (Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; McMillan, Corley, & Lickley, 2009; Mowrey & MacKay, 1990; Pouplier, 2003, 2007, 2008), but are instead ill-formed, in the sense that aspects of the production of more than one phoneme are

observed simultaneously. The present paper takes these observations as a starting point, and reevaluates the phonemic similarity effect in this context. We employ a tongue-twister task to demonstrate that the phonemic similarity of adjacent onset phonemes influences their articulation in predictable ways, without requiring the assumption that phonological errors derive from the substitutions of whole phonemes. We argue that our evidence reflects an organizing principle of planning, showing that the ways in which articulation varies are affected by the phonological properties of what is said. This rules out a view that distortions in articulation are 'motoric' in nature, pointing to the tight coupling between speech plans and their execution.

The structure of the paper is as follows. First, we review evidence suggesting that the phonemic similarity effect constrains models of speech production. Second, we turn our attention to evidence that speech errors may not involve simple substitutions of phonemes. We present two experiments in which we manipulate the phonemic similarity of onset phonemes in tongue-twisters and measure the resultant articulation. These experiments make use of a novel measure of articulatory variation, previously used for an electropalatographic analysis of articulation as affected by lexical status (McMillan et al., 2009). Here, the method is used to show the influence of phonemic similarity on tongue-to-palate contact (Experiment 1), and extended to the analysis of ultrasound images showing the midsagittal tongue contour (Experiment 2). Evidence from the two experiments is then considered in the context of the linkage between speech planning and articulation.

This paper does not address the distinction between potential types of subphonemic representation. One possibility is that these are phonemic features, in the sense of Chomsky and Halle (1968, e.g., Dell, 1986). An alternative proposal is that they consist of articulatory gestures (e.g., Goldstein et al., 2007). For present purposes, what concerns us is how representations in the speech plan are reflected in articulation. Because features are more widely discussed in the psychological literature, we have chosen a feature-based rather than a gesture-based approach, but

for much of the theoretical discussion which follows, the term 'feature' can also be interpreted as 'gesture'.

Phonemic Similarity in Speech Production

The similarity between two phonemes can be measured in several ways (see Frisch, 1996, for a discussion), but the most common method is to count the numbers of features by which two phonemes differ. By this definition, /k/ and /t/ only differ by one feature (place of articulation: velar vs. alveolar) and are therefore more similar than /k/ and /d/ which differ by two features (place of articulation as above; voicing: voiceless vs. voiced). Using this metric, phonemic similarity has been shown to have effects in a wide range of cognitive tasks, including working memory tasks (Baddeley, 1966) and picture naming (Bock, 1986). Here, our discussion focuses on speech errors.

Corpora of speech errors are a major source of evidence that phonemic substitutions are affected by phonemic similarity (Dell & Reich, 1981; MacKay, 1970; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1982; del Viso et al., 1991; Vousden et al., 2000). Shattuck-Hufnagel and Klatt (1979) carried out a confusion matrix analysis of the 1977 MIT Corpus and observed that in many consonant exchanges the phonemes differed by only one feature. Similarly, Vousden et al. (2000) examined all of the consonant exchanges in a corpus of 6,753 speech errors, and showed that phonemes which differed by only one feature were more likely to be exchanged than would be predicted by chance. These analyses did not take anticipations or perseverations into account, but in an analysis of 2,177 phonological speech errors, Stemberger (1982) reported high rates of single feature substitutions in such errors whether they were between-word (such as "pig pocket", 510 of 736) or within-word (70 of 136). Phonemic similarity effects have also been demonstrated in languages other than English. In an analysis of a German speech error corpus, MacKay (1970) demonstrated that over 55% of substituted phonemes differed by one distinctive feature, while less than 5% differed by four distinctive features (cf. del Viso et al., 1991, for Spanish).

In addition to evidence from corpora, there have been a number of experimental demonstrations of the effects of phonemic similarity using tongue-twisters (Butterworth & Whittaker, 1980; Kupin, 1982; Levitt & Healy, 1985; Wilshire, 1999). In these experiments the similarity of phonemes to be uttered can be directly manipulated. For example, Levitt and Healy (1985) demonstrated that stimuli including onsets which differed by one feature yielded more substitution errors than those with onsets which differed by more than one feature. Using a confusability metric of phoneme similarity based on Shattuck-Hufnagel and Klatt (1979), Wilshire (1999) showed that participants were nearly four times as likely to errorfully substitute similar compared to dissimilar phonemes. Investigations using silent tongue-twisters have reinforced the view that these phonological errors are representative of the speech planning process (Dell & Repka, 1992; Postma & Noordanus, 1996; Corley, Brocklehurst, & Moat, in press). Importantly, the addition of articulation (and hence motor processes) does not appear to significantly affect the pattern of errors observed (Postma & Noordanus, 1996).

Models of speech planning which can account for the effects of phonemic similarity found experimentally and in corpora must incorporate some way for similar phonemes to interact with one another more than those which are dissimilar. Under the assumption that the errors investigated result from the insertion or substitution of well-formed phonemes, the most straightforward way to achieve this is to include a lower level of representation for phonological features (Dell, 1986; Stemberger, 1982, 1985). Consider, for example, Dell's (1986) model of production. This model incorporates both phoneme and feature representations, as well as syllable representations which are not discussed here. Phonemic encoding for the articulation of a given syllable is completed when a predetermined time has elapsed. At this point, the phonemes with the highest activations are selected and passed to the articulation system. Because the output from Dell's (1986) model is driven by selected phonemes, the featural level exerts its influence through feedback. With feedback from activated

features, the likelihood that an unintended phoneme receives more activation than one which was intended, and so is selected and produced in error, will be greater to the extent that there are features in common between the competing phonemes.

Noncanonical Errors in Speech Production

In Dell's model, phonemes which are not selected are explicitly prevented from influencing the articulation of the present syllable (although they may affect the likelihood that a phoneme is later produced in error, since they retain activation). Similar mechanisms are found in other models of speech production (e.g., Levelt et al., 1999). However, this leads to what Dell (1986) identifies as a "featural paradox": The units that the model outputs are phonemes, but features are still required to account for phonemic similarity effects. As a first step towards addressing this paradox, we turn our attention to evidence for the existence of phonemic speech errors which are not phonemically well-formed, or *canonical*.

Several recent articulatory and acoustic investigations of speech errors have demonstrated the existence of non-canonical speech errors (Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldstein et al., 2007; Mowrey & MacKay, 1990). In an electromyographic (EMG) investigation, Mowrey and MacKay (1990) observed transversus/verticalis muscle movement normally associated with /l/ production during the production of "bay" in repetitions of *Bob flew by Bligh bay*. Using electromagnetic articulometry (EMA), Goldstein et al. (2007) reported that repetitions of *cop top* yielded articulations which include overlapping tongue-dorsum and tongue-tip raising during /k/ and /t/ articulations. In an acoustic investigation of /s/–/z/ errors, Frisch and Wright (2002) observed that both phonemes were produced with a continuum of percent voicing ranging from 0–100%. Goldrick and Blumstein (2006) similarly demonstrated that the voice onset times (VOTs) of stop consonants produced in error were not canonical: For example, a /g/ produced where a /k/ was intended has a different VOT from intentional productions of either /k/ or /g/.

In models with selection at the phonemic level, non-canonical errors such as those observed above must be attributed separately to 'noise' in the articulatory implementation. In their assessment of the evidence leading to their model, for example, Levelt et al. (1999) point out that the observation of inappropriate muscle movements by Mowrey and MacKay (1990) could be attributed to a late motor execution stage. Similar observations have been made about tongue-twister studies (e.g., Laver, 1980). Many of these studies have used relatively fast repetition rates (e.g., 180–210 syllables/minute: Kupin, 1982) which has led to the criticism that the task may require faster-than-usual articulatory movements, and any errors observed, whether canonical or not, may reflect motoric rather than planning difficulties. In fact, estimates of the speed at which English is typically spoken suggest that this criticism is unfounded (e.g., 240 syllables/minute for British English: Tauroza & Allison, 1990). Moreover, Wilshire's (1999) tongue-twister experiment showed phonemic similarity effects using a very slow speaking rate of 100 syllables/minute. Thus it would appear that at least the canonical errors observed in tongue-twister studies cannot be attributed to abnormally high repetition rates.

When non-canonical errors are instrumentally measured, it has been repeatedly observed that they 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), and that they cannot be attributed to a fast speech rate in tongue-twisters (Goldrick & Blumstein, 2006). This suggests that planning is involved in the production of these errors. Thus it appears unlikely that canonical and non-canonical errors can be attributed to different sources, and it may be more parsimonious to assume a single underlying mechanism. A candidate mechanism is a model of production in which information *cascades* from one level to the next (Goldrick, 2006; Goldrick & Blumstein, 2006; McMillan et al., 2009), resulting in articulation which can include properties of more than one phoneme at a

given moment. In such a model activation at the phonemic level directly influences articulation without (early) selection. Non-canonical errors can be attributed to the partial activations of representations that are competing during phonemic encoding. The movement of the articulators exhibits properties corresponding to the activation levels of each of the competing representations. According to this view, there is no distinction between non-canonical and canonical errors: If the properties of an unintended phoneme dominate the activation which is passed to articulation, the resultant speech is only likely to reflect observable qualities of that phoneme, and the error will appear to be canonical.

Suggesting that activation cascades from the phonemic to the articulatory level in speech production has three important consequences. First, it provides an answer, if not a solution, to Dell's featural paradox. Articulation that is driven by partially activated phonemes is likely to be indistinguishable from articulation which is driven by (partially) activated features. Although it remains easier to describe a cascading model in terms of partially activated phonemes (and that is what we do here) there is no longer any *requirement* that the articulatory plan is driven by phoneme, but not feature, activation: indeed several proposals include subsegmental, but not phoneme-level, representations (e.g., Browman & Goldstein, 1989; Pouplier, 2007). Hence the paradox is no longer a paradox, but simply an open empirical question.

Second, a cascading view forces us to abandon the concept of an error in speech production (see also McMillan et al., 2009). Because activation cascades from the phonemic level to articulation, articulation (and the resulting acoustic output) will vary along a continuum (as has been repeatedly observed in instrumental studies: Frisch & Wright, 2002; Frisch, 2007; Goldstein et al., 2007; Pouplier, 2007). Although the classification of an acoustic continuum may tell us about the *perception* of phonemes (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), assigning articulations to categories such as 'error' is unlikely to tell us about *production*, because it imposes an essentially arbitrary threshold below which the influence of

competing phonemes is not considered. Moreover, empirical studies in which responses are categorized frequently exclude instances which could not be readily classified, but these may be just as likely to show influences of competing phonemes as those which are included. Reviewing the SLIP task literature, McMillan et al. (2009) reported that in some experiments as few as 2.3% of responses were classified as 'errors' and analyzed, compared to up to 13.9% discarded 'other' responses (see also Nooteboom & Quené, 2007). Rather than investigating whether or not errors are produced, within a cascading framework it is more productive to determine the extent to which articulations vary in situations which are designed a priori to cause relevant competition in the speech plan.

The third consequence of adopting a cascading framework is the focus of the present paper. Because existing evidence of the effects of phonemic similarity relies on phoneme-level transcription and categorization, it cannot be used to rule out a noise-based explanation of articulatory variance. We elaborate on this point below, before introducing two studies designed to show that there is systematicity in articulatory variation.

Phonemic Similarity in Articulation

Merely introducing noise to articulatory processes will make it more likely that similar-sounding phonemes are mistaken for each other. A noisy /t/ is more likely to be mistaken for a /d/ than it is for a /g/, leading wrongly to a conclusion that /t/ is more likely to be erroneously *substituted* with /d/. Of course, acoustic confusability and featural similarity are not the same thing, although they tend to be highly related (Frisch, 1996). To the extent that this relationship holds, existing evidence based on the categorization of responses cannot be used to support the contention that articulation reflects properties of the speech plan.

In the remainder of this paper we show that articulation does reflect the speech plan, by establishing that the effects of phonemic similarity can be observed directly in

the articulatory record. We report three different measurements of articulation recorded during two tongue-twister tasks. In Experiment 1 we report tongue-to-palate contact over time recorded using electropalatography (EPG), in Experiment 2 we report the midsagittal contour of the tongue over time using ultrasound, and in both experiments we report VOT. Together, these three sources of evidence provide a quantitative measurement of the extent to which competing speech plan representations affect articulation during repetitions of phonemically similar and dissimilar tongue-twisters. Additionally, instrumental measurements of articulation provide a source of evidence which is not confounded by perceptual limitations and does not require the categorization of responses.

Experiment 1

Experiment 1 took as its starting point a cascading model of speech production, and was designed to measure the degree to which articulation was affected in situations where competition between phonemes would be likely. Following Wilshire (1999), we used a tongue-twister design: Rather than transcribe or categorize responses, however, we measured the variability of onset phoneme articulation in tongue-twisters (such as *kef def def kef*) relative to control sequences in which competition was unlikely (*kef kef kef kef*).[1]

We derived variability metrics from two measures. First, we analyzed the acoustic signal using voice onset time (VOT), a robust measure of voicing for onset stop consonants (Lisker & Abramson, 1964; Goldrick & Blumstein, 2006). We reasoned that VOT would be most affected when the competing phonemes in a tongue-twister differed in voicing (e.g., *kef gef gef kef*). Second, we analyzed the tongue's movements over time using electropalatography (EPG), which uses an artificial palate with an array of microswitches to measure tongue-to-palate contact (cf. McMillan et al., 2009). Since EPG clearly reflects the contact made to produce stop consonants, we expected EPG variability to be greatest when there was

competition for place of articulation (e.g., *kef tef tef kef*).

Of critical interest was the dissimilar case, in which both place and voicing differed (*kef def def kef*). If articulatory differences are attributable to noise, there is no reason to suppose that any variability in articulation would be attenuated in this case, and each measure should therefore reflect competition in the relevant dimension. If planning is implicated, however, then the phonemic similarity effect found in categorical studies might be expected to hold, such that dissimilar phonemes would be less likely to interfere than in cases where only one feature is varied.

Method

Participants. Seven native speakers of English from the Edinburgh research community participated in the experiment. Two speakers were excluded from the analysis because of technical failure during recording. All participants were experienced in speaking while wearing an EPG artificial palate. In this and in the following experiment, participants reported no speech or hearing impairments and were treated in accordance with Queen Margaret University and University of Edinburgh ethical guidelines.

Materials. Tongue-twisters were created using pairs of onsets selected from the four stop consonants /k, g, t, d/, resulting in sequences in which the onsets differed by place of articulation, voicing, or both. The onsets were chosen to yield firm tongue contact with the EPG palate, and were combined with nucleus vowels and coda consonants which were chosen to minimize subsequent EPG contact. One vowel was selected for each tongue-twister from the set /ɪ, e, ʌ/; four versions of the resultant sequence were created, one each in ABBA and BAAB order, and one each with the coda /f/ or /v/. Sequences were orthographically transcribed. For example, the onsets /k, t/ and the vowel /ɪ/ were used to create the four sequences *kif tif tif kif*, *tif kif kif tif*, *kiv tiv tiv kiv*, and *tiv kiv kiv tiv*. In all, 15 onset-vowel combinations were used, resulting in 60 tongue-twisters. Additionally, a control sequence, in which there was no

alternation, was created for each of the onsets, consisting of four repetitions of the onset together with an arbitrary vowel and coda (e.g., *kef kef kef kef*). Appendix A lists all 64 items.

Apparatus. The experiment took place in a sound-treated recording studio. Prior to testing, each participant was fitted with a custom electropalatography (EPG) palate (manufactured by Incidental, Newbury, UK or Grove Orthodontics, Norfolk, UK) molded to fit a dental cast from an impression of the hard palate. Each EPG palate was made of acrylic and contained 62 embedded silver contacts on the lingual surface, organized in eight rows of eight contacts (except the most anterior row, which had six).

A desktop computer, to which an Audio Technica ATM10a microphone and a WinEPG system (Articulate Instruments Ltd: Edinburgh, UK) were attached, was used to record participants' responses with Articulate Assistant (Articulate Instruments Ltd, 2007b) software. EPG data was recorded at rate of 100Hz using the WinEPG system, which connected the palate to a multiplexer unit that transferred the data to an EPG3 scanner and then to the serial port of the computer. Simultaneously, the acoustic signal of participants' responses was recorded to a single auditory channel at 22,050Hz.

Stimuli were presented on a 15" LCD monitor using 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 mono headphone (worn on participants' preferred ears) and to a direct audio line into the EPG computer, where it was recorded to a second auditory channel.

Procedure. After fitting and testing of the EPG palate, participants were instructed to read each experimental word (e.g., *kef*) aloud once, to make sure that they used the anticipated vowel. Feedback was given about their pronunciation, and if necessary, they were asked to repeat each word 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, at a rate of one word per metronome beat. Following the recording of each sequence the experimenter pressed a key which caused the display to advance to the next sequence after a short pause (approximately 3s). Participants were allowed to take a longer break by notifying the experimenter. The first four items were the four control sequences (e.g., *kef kef kef kef*). These were followed by the 60 experimental sequences, presented in random order.

Data Treatment

Following the experiment, we performed measurements on both the acoustic and EPG recordings. Each word onset was measured independently. The only items excluded were those items not collected due to technical failure of the recording equipment (83 items out of 5120 possible responses).

Acoustic Data. The VOT for each target item was measured from the acoustic signal using Praat (Boersma & Weenink, 2006). VOT was defined as the duration (in milliseconds) between the acoustic burst of the onset and the onset of the periodicity associated with the following vowel.

Next, we created a deviance score for each observation. First we calculated a mean reference VOT for each speaker and onset from the control sequences. The deviance between a tongue-twister VOT and the relevant reference VOT was then calculated by taking the absolute difference between the two measures. A higher value represents a VOT which differs more from the relevant mean reference VOT. This method of calculating deviance is equivalent to calculating the Euclidean distance between two VOT values, and is therefore equivalent to the EPG and ultrasound measures reported below.

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 word (or for the first word, a time point 150ms prior to the onset release) and the onset of the vowel in the word under consideration. The EPG record for this duration was extracted for preliminary inspection.

The EPG record for a given onset showed contact at each of the 62 palate microswitches (represented as 0 or 1), sampled every 10ms. Each record was trimmed to include the first palate before full closure through to the first palate after full closure, where full closure was defined as any continuous path across the lateral axis of the palate (such that the tongue was presumed to be blocking airflow). 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.

Figure 1 shows two example closures: (a) a trimmed velar item with full closure; (b) a trimmed velar item without full closure.

Insert Figure 1 about here

Once the EPG records had been extracted and trimmed they were standardized using an averaging algorithm which expanded or contracted the number of observed onsets to yield 10 data frames. This entailed treating each EPG contact as a continuous value, where 0 represents no contact and 1 represents continuous contact over each period of time. Once 10-frame versions of each EPG record were obtained, reference EPG records were calculated from the control sequences by averaging EPG contact for each frame at each of the 62 EPG contact points. Once again we calculated deviance scores between each EPG record and the relevant reference record. Deviance was defined as the mean of the Euclidean distances between each of the 10 corresponding pairs of frames, which were treated as 62-dimensional vectors. This method of comparing EPG records, referred to here as the Delta method (see also McMillan, 2008; McMillan et al., 2009), results in a single number representing the

deviance between a given EPG record and the reference: Higher values (in arbitrary 'Delta units') represent records which differ more from the relevant mean reference record.

Results

To investigate the influence of phonemic similarity on production we independently analyzed the VOT and EPG deviance scores using Generalized Linear Mixed-Effects models, with the lme4 (Bates, Maechler, & Dai, 2008) and languageR (Baayen, 2008) packages in R (R Development Core Team, 2008). Both analyses include every recorded observation and include Voice (change, no change) and Place (change, no change) as fixed factors, where item and participant can randomly vary the intercept. Each model included the interaction of Voice and Place, since this was the effect of primary theoretical interest. Each tongue-twister sequence was treated as an independent experimental item. Prior to model fitting, the fixed factors were centred to reduce multicollinearity (Dunlap & Kemery, 1987); for independent variables with two levels, this is conceptually equivalent to using sum coding, but the weights assigned are appropriate to unbalanced cell sizes, such that the model intercept represents the grand mean.

The t-values for each coefficient are reported along with estimated probabilities based on 10,000 Markov Chain Monte Carlo (MCMC) samples (p_{mc}). Using MCMC estimates to evaluate a fitted model has been suggested because it can be difficult to determine the degrees of freedom corresponding to each t-value for the model coefficients (Bates et al., 2008). The reported coefficient values and confidence intervals are also MCMC estimates of differences from the grand mean.

Acoustic Analysis (VOT). A mixed effects approach was used to model 5037 VOT deviance scores across five speakers and 64 items, resulting in a model with log likelihood of -18657. Results are reported relative to a baseline of 10.84ms, representing grand mean VOT deviance. Compared to this baseline, articulation

reliably varied by an additional 1.16ms when the tongue-twister included a change in Voice: $t = 2.18$; $p_{mc} = 0.03$. As predicted, a change in Place did not increase deviance: estimated effect 0.01ms; $t = 0.02$; $p_{mc} = 0.98$. Importantly, when both Voice and Place changed, there was a negative effect: VOT deviance was *reduced* by 2.64ms: $t = 2.02$; $p_{mc} = 0.04$. Table 1 shows the means and standard deviations for each condition and Figure 2 shows the coefficient estimates relative to the baseline, together with the attendant 95% confidence intervals, for each effect. Taken together, the model estimates show that when both Voice and Place change, the expected average deviance is $10.84 + 1.16 + 0.01 - 2.64$ or 9.37ms. In other words, there is a clear interaction effect such that the effect on VOT of a change in Voice is reduced when Place changes too.

Insert Table 1 about here

Insert Figure 2 about here

Articulation Analysis (EPG). A mixed effects approach was used to model 5037 EPG deviance scores across five speakers and 64 items, resulting in a model with log likelihood of -11988. Results are reported relative to the baseline of 2.86 Delta units, representing grand mean articulatory deviance. When there was a Voice change, articulation varied by a further 0.10 units, although this effect was not significant: $t = 0.79$; $p_{mc} = 0.43$. As predicted, however, deviance increased by 0.83 units when the tongue-twister included a Place change: $t = 5.67$; $p_{mc} = 0.0001$. As in the acoustic analysis, when Voice and Place change simultaneously, there was a negative effect. Deviance was reduced by 1.09 units: $t = 3.54$; $p_{mc} = 0.0004$. Table 2 shows the means and standard deviations for each condition and Figure 3 shows the coefficient estimates relative to the baseline, together with 95% confidence intervals. The EPG

analysis shows that a change in Place increases articulatory deviance, but only in cases where Voice remains unchanged.

Insert Table 2 about here

Insert Figure 3 about here

Discussion

In two analyses we calculated the deviations of articulatory measurements of phonemes obtained when there was competition between tongue-twister onsets from measurements obtained when competition was minimized in a control sequence. We then compared the deviation scores obtained when the tongue-twister onsets differed by either one or two features from the baseline. As predicted, voice onset time was more variable when there were changes in voicing between onsets (e.g., *kef gef gef kef*) than when place of articulation changed (*kef tef tef kef*). Conversely, tongue movements, as measured using EPG, became more variable when place changed, but were not affected by changes in voicing. These differences establish that articulation varies in predictable ways when there is phonemic competition. This is important because it demonstrates that it is not necessary to classify responses as 'errors' in order to observe the effects of phonemic competition in speech production. Moreover, it shows that competing phonemes cause relevant changes in articulation, contrary to the view that misarticulation is caused by 'noise' (e.g., Levelt et al., 1999).

Critically, when both voicing and place of articulation changed between onsets, the increases in deviance observed for single-feature changes were significantly reduced: In other words, for both VOT and EPG, variability increased significantly more relative to the baseline when the relevant single features were in competition than

when the competing phonemes differed by two features. Thus this experiment clearly demonstrates that similarity between competing phonemes has direct consequences for articulation. Since the demonstration does not depend on the transcription or categorization of responses, it provides prima facie evidence, obtained from two different dependent measures of articulation, that variability in the articulation of speech reflects differences in the speech plan.

Before concluding, however, we should note two limitations with the present experiment. First, our investigations are limited to differences of two features or less, in contrast to previous work, which has included differences of three or more features. A frequent conclusion is that substitution errors are more likely when the difference between competing phonemes is one or two features (Shattuck-Hufnagel, 1979, 1983; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1982). Second, our investigations of tongue movements are limited to observations of contact with the hard palate, despite evidence that the influences of competing phonemes may be expressed in 'partial' tongue movements that do not involve palate contact (Frisch, 2007; Goldstein et al., 2007; Pouplier, 2003, 2004, 2007; Stearns, 2006). These problems are interrelated: The limitations of EPG mean that changes in manner of articulation or nasality are hard to detect, and for this reason, Experiment 1 focuses on stop consonants. In Experiment 2, we avoid this limitation by using ultrasound rather than EPG to measure tongue movements. This gives us the opportunity to investigate the effects of phonemic similarity using a new material set in which onsets vary by up to three features, as well as to replicate our findings using a different articulatory imaging technique.

Experiment 2

Experiment 2 was a tongue-twister investigation in which, in addition to VOT measures, articulatory variation was measured using ultrasound imaging of the midsagittal contour of the tongue (see Stone, 2005, for a discussion of oral ultrasound physics). A benefit of using ultrasound to measure articulation was that partial tongue

movements which reflect activation of phonemic representations could be recorded. Previous ultrasound analyses have shown that the influences of competing elements in the speech plan can be observed even in cases where there is no palate contact (Pouplier, 2004; Frisch, 2007; Stearns, 2006). For example, Pouplier (2004) traced the contour of the tongue on single frames selected from ultrasound recordings of alternating /k–t/ repetitions (e.g., *cop top*). She then measured tongue-dorsum height and tongue-tip slope for each frame and observed a continuum of values, ranging from fully /k/-like to fully /t/-like, for each intended onset. Similar continua have also been found using EMA (Goldstein et al., 2007; Pouplier, 2003, 2007). These findings suggest that measurements based on EPG recordings of stop consonants may underestimate the degree of variability in articulation, since they are most likely to be influenced by events at each end of the continuum, where there is palate contact.

The design of the experiment was similar to that of Experiment 1, allowing us to test for a replication of our EPG findings using a different imaging technique. However, the use of ultrasound allowed us to introduce additional feature competition, by including the onset phonemes /s/ and /z/ as well as /k, g, t, d/. The inclusion of two fricative onsets allowed us to systematically vary whether voicing, place of articulation, or manner of articulation were competing in tongue-twister onsets (the design was not fully orthogonal because the velar fricatives /x, ɣ/ do not typically occur in British English). Varying tongue-twister onsets by up to three features allowed us to check the generality of our findings, as well as to explore the suggestion that different types of features may interact differently (e.g., Shattuck-Hufnagel, 1979; Stemberger, 1991). In line with Experiment 1, our general prediction was that phonemes with more similar competitors (i.e., those which varied by fewer features) would cause greater relevant articulatory variation.

Two additional changes were made to Experiment 2 as a consequence of the use of ultrasound. First, all words in all materials ended in the rime /-ɒm/. Since ultrasound imaging records the position of the tongue where there is no palatal

contact, it was important that the post-onset portions of each word should result in movements which were as similar as possible. The coda /m/ was selected because it shared minimal features with the onsets under investigation. Second, the speaking rate was slowed to 100 syllables per minute (compared to 150 syllables per minute in Experiment 1). The prime motivation for this was that we were only able to sample ultrasound images at 25Hz (compared to 100Hz for EPG) and hoped to encourage participants to articulate more slowly. However, it also served as an additional check that the results of Experiment 1 were not dependent on speech rate, by using the same slow rate as Wilshire (1999).

The primary goal of Experiment 2 was to replicate the phonemic similarity effects observed in Experiment 1. The first analysis we report is a VOT analysis based on the phonemes /k, g, t, d/, designed to replicate the analysis from Experiment 1. Second, we report an ultrasound analysis based on the same four onsets. The final analysis is an ultrasound analysis that includes the additional phonemes (/s, z/), to investigate the influence of a third competing feature, manner of articulation, on phonemic similarity.

Method

Participants. Ten native speakers of English from the Edinburgh research community participated in the experiment. Two speakers were excluded from all analyses due to poor ultrasound image quality in comparison to the other eight speakers.

Materials. Tongue-twisters were created using each of the 15 possible pairings of onsets selected from the six consonants /k, g, t, d, s, z/. Onset pairs were combined with the rime /-ɒm/ and used to create 15 ABBA and 15 BAAB tongue-twisters. Sequences were orthographically transcribed. For example, the onsets /k, s/ were used to create the two sequences *kom som som kom* and *som kom kom som.* Additionally, a control sequence, in which there was no alternation, was created for each onset (e.g.,

kom kom kom kom). Appendix A lists all 36 items.

Apparatus. The experiment took place in a sound-treated recording studio. 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. 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.

Stimuli were presented on a 15" LCD monitor. Acoustic recordings of participants' responses were recorded at 22,050Hz using an Audiotechnica ATM10a microphone. The acoustic and ultrasound data were synchronized using Articulate Assistant Advanced software (Articulate Instruments Ltd, 2007a). The entire video file for each stimulus item was exported from Articulate Assistant into AVI format using an MPEG-4 (mp42) Video Codec.

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 of listening binaurally, or monaurally with their preferred ear.

Procedure. Participants were fitted with the lightweight helmet and the ultrasound transducer was adjusted to fit beneath the chin with a pressure as firm as comfortable. Participants were instructed to read two randomly selected tongue-twisters aloud. During these repetitions the transducer probe was adjusted to yield the highest quality ultrasound image. Participants were given feedback about their pronunciation to ensure that they were pronouncing the vowel (/ɒ/) correctly.

Once setup was complete, each tongue-twister was presented individually and participants were instructed to repeat each phrase four times, at a rate of one word per

metronome beat. Following the recording of each sequence the experimenter pressed a key to advance to the next sequence. There was a pause of approximately 7s between items to allow data to be saved, which was indicated to the participant with a white blank screen. Participants were instructed to let the experimenter know if they required a longer break. All 36 items (30 tongue-twisters and 6 control sequences) were presented in random order.

Data Treatment

Following the experiment, we performed measurement on both the acoustic and the ultrasound data. Each word onset was measured independently. The only items excluded were those items not collected due to recording failure (66 out of 4608 possible responses).

Acoustic Data. The acoustic analysis did not include items which began with the fricatives /s/ or /z/, since VOT is a measurement derived from stop consonants. VOT measurements were made and deviance scores were derived in the same way as was for Experiment 1.

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 video sequence from 0.3s before the release to 0.3s after the release, equivalent to 15 data frames. A detailed inspection of a subset of ultrasound recordings suggested that this time window yielded a representative sampling of tongue-raising, constriction, and tongue-lowering for each articulatory token. Previous work revealed that the analysis method yielded similar results across different sized time windows (McMillan, 2008). The individual video frames of each record were extracted from the ultrasound video files and converted into PNG still images using Mplayer (http://www.mplayerhq.hu) software.

The ultrasound record for a given onset consisted of a sequence of black and

white video frames at a resolution of 640 × 480 pixels. In the initial stage of the analysis, we excluded regions of the image representing control information, extracting a rectangular region corresponding to the imaged tongue, where each of 216,720 pixels ranged in luminance value from 0 (black) to 255 (white). We reasoned that similar tongue positions should result in similar distributions of pixel values. To make the analysis more tractable, we then pixelized each frame by taking the average luminance of each 12 × 12 pixel grid, resulting in a 1,505-pixel image. Figure 4 shows an arbitrary example frame of recorded ultrasound.

Insert Figure 4 about here

Mean reference ultrasound records were calculated from the control sequences by averaging the luminance of each resultant pixel for each frame. We defined deviance between each ultrasound record and the relevant reference record in the same way as for EPG, as the sum of Euclidean distances between each corresponding pair of (here, video) frames, which were treated as 1,505-dimensional vectors (see also McMillan, 2008). Higher values (in arbitrary 'Delta units', here larger than for EPG because the input values range between 0 and 255 rather than 0 and 1) represent records which differ more from the relevant mean reference record. Note that due to the nature of ultrasound recordings a number of pixels in each image are more-or-less randomly grey (see, e.g., Figure 4). However, pixels at clear physiological junctures such as the lingual surface tend to result in pixels of deterministic hues, and there are likely to be a number of similarities in patterns of light and shade across the ultrasound image for similar tongue positions. Similarities between pixels will tend to reduce Delta values, allowing us to distinguish signal from noise.

Validation. To demonstrate that the ultrasound method was sensitive to relevant differences between articulations, we evaluated 16 /k/, 16 /t/ and 16 /s/ onsets, taken from one speaker's control recordings, independently. First, we calculated the Delta

deviance between each pair of items (120 deviance scores). We then used a multidimensional scaling algorithm (Cox & Cox, 1994) to visualize the results in two dimensions. Multidimensional scaling takes a set of similarity values (e.g., deviance scores in Delta units) and returns a set of points on a scatter plot arranged such that the distances between the points of the plot are approximately equal to the similarity values between the points. Figure 5 shows the results of this analysis. Two features of the plot are important. First, the /k/ articulations are clearly separate from the /t, s/ articulations, capturing the difference in place between velar and alveolar articulations. Second, the /t/ articulations are clustered together and distinct from /s/, capturing the difference in manner between stops and fricatives. This analysis shows that the Delta method usefully measures the differences between individual ultrasound records of phoneme production.

Insert Figure 5 about here

Results

Analyses were carried out in the same way as for Experiment 1. We report three separate analyses. The first two are a VOT and an ultrasound analysis based on a subset of the data, consisting of all of the observations that come from tongue-twisters which do not include either /s/ or /z/. For the VOT analysis, this is necessary because VOT can only be measured for stop consonants. In the case of ultrasound, the subset analysis allows us to make a direct comparison with the EPG analysis reported for Experiment 1. Finally, we report an ultrasound analysis over the whole data set, in which the model includes the factors of Place, Voice, and Manner, as well as all two-way and three-way interactions.

Acoustic Analysis (VOT). This analysis was restricted to experimental items which did not include /s/ or /z/. A mixed effects approach was used to model 2023

VOT deviance scores across 8 speakers and 16 items, resulting in a model with log likelihood of −7356. Results are reported relative to a baseline of 9.31ms, representing grand mean VOT deviance. Means and standard deviations for each condition are reported in Table 3. Compared to the baseline, VOT varied by an additional 2.63ms when Voice changed: $t = 2.42$; $p_{mc} = 0.03$. When there was a Place change, VOT was increased by 0.66ms, but this effect was not significant: $t = 0.60$; $p_{mc} = 0.56$. Although the model estimate of an effect of a change in both Voice and Place was negative, as in Experiment 1, the effect of −3.32ms failed to reach significance in the present experiment: $t = 1.52$, $p_{mc} = 0.16$. Figure 6 shows the estimates relative to the baseline, together with 95% confidence intervals.

Insert Table 3 about here

Insert Figure 6 about here

Articulation Analyses (ultrasound). The first ultrasound analysis was restricted to items which did not include /s/ or /z/ onsets. A mixed effects approach was used to model 2023 ultrasound deviance scores across 8 speakers and 16 items, resulting in a model with log likelihood of −11293. Results are reported relative to a baseline of 561.0 Delta units, representing grand mean articulatory deviance. Per-condition means and standard deviations are reported in Table 4. When there was a Voice change, articulation varied by an additional 34.9 units: $t = 5.01$; $p_{mc} = 0.0001$. Deviance also increased by 82.1 units when Place changed: $t = 11.79$; $p_{mc} = 0.0001$. When both Voice and Place changed, articulatory variance was reduced by 71.8 units: $t = 5.16$; $p_{mc} = 0.0001$. Figure 7 shows the estimates relative to the baseline, together with 95% confidence intervals.

Insert Table 4 about here

Insert Figure 7 about here

The final analysis was based on the entire dataset of ultrasound deviance scores. We fitted the data with a Generalized Linear Mixed-Effects model which included Place (change, no change), Voice (change, no change), and Manner (change, no change) as fixed factors, and item and participant as random factors. The model also included all two-way and three-way interactions between the fixed factors. We modelled 4542 ultrasound recordings across 8 speakers and 36 items, resulting in a model with log likelihood -25904. See Table 5 for means and standard deviations for each condition. Relative to a baseline of 569.9 Delta units, a change in Voice increased deviance by 18.3 units: $t = 3.95$; $p_{mc} = 0.0010$. A change in Place increased deviance by 81.2 units: $t = 17.36$; $p_{mc} = 0.0001$. A change in Manner (for example, where /t/ and /s/ were competing) also increased deviance, by 40.2 units: $t = 8.6$; $p_{mc} = 0.0001$. When any two factors changed, there was a significant reduction in deviance. Simultaneous changes in Voice and Place reduced deviance by 45.5 units: $t = 4.86$; $p_{mc} = 0.0001$. When Voice and Manner changed, deviance was reduced by 55.6 units: $t = 5.94$; $p_{mc} = 0.0001$. When Place and Manner changed, the reduction was 20.5 units: $t = 2.19$; $p_{mc} = 0.03$. Finally, a three-way interaction resulted in a 76.0 unit increase in variability: $t = 4.04$; $p_{mc} = 0.0004$. Refer to Figure 8 for estimates of each effect together with 95% confidence intervals. Taken together, these results show that variability is increased more by changes in any one factor than in cases when two factors change simultaneously. When three factors change, there is again an addition to the observed deviance, so that the model estimate of deviance for such cases (e.g.,

when /k/ and /z/ compete) is 663.1 Delta units.

Insert Table 5 about here

Insert Figure 8 about here

Discussion

Analyses including only the phonemes /k, g, t, d/ gave rise to similar patterns of results to Experiment 1. A single-feature difference increased articulatory variability in the relevant dimension, suggesting that competition between similar phonemes caused articulatory competition. When the difference was two features, the summed variation attributable independently to each feature was reduced, although this reduction was not reliable for VOT, perhaps as a consequence of the smaller number of materials in this experiment (which was necessitated by capacity limitations in storing ultrasound video). It should also be noted that, unlike the EPG analysis, the ultrasound analysis showed that articulatory deviance increased when there was a single-feature change in voicing (e.g., when /k/ competed with /g/). This may reflect the additional sensitivity of ultrasound to tongue movements which do not involve palatal contact: In the voiced case, the tongue root is likely to lower sooner to produce the nucleus vowel /ɒ/. Taken together, these findings are entirely compatible with Experiment 1: The phonemic similarity effect established in Experiment 1 using VOT and EPG analyses is here replicated using ultrasound. Importantly, this replication additionally demonstrates that the Delta method previously used to analyze EPG records (McMillan et al., 2009) can be used for quantitative analysis of speech recorded using ultrasound.

When /s/ and /z/ were included in a three-way analysis of the ultrasound deviance scores which included factors of voicing, place of articulation, and manner of

articulation, the results were broadly compatible with the previous analyses. A difference of any single feature between competing phonemes reliably increased articulatory variability, as predicted, demonstrating that the ultrasound record was sensitive to changes in all three dimensions. When competing phonemes differed by any two features, there was a reliable decrease in Delta score, showing that the net variation in articulation was always less in this case than the summed variation attributable to each individual feature. This is a clear interaction effect, such that interference is greater when similar phonemes compete.

One unexpected finding was that where competing phonemes differed by three features, there was a reliable increase in articulatory variation. We do not have a theory-based explanation for this increase. We note, however, that in the experimental design, observations from only two consonant pairs (/z-k/ and /s-g/) contributed to this effect (compared to seven for 1-feature and six for 2-feature differences). Because observations were based on differences in articulation, and the physiology of the mouth presumably limits the potential for deviant movement at different loci, the increase when 3 features change may represent particular aspects of the consonants involved. These idiosyncrasies are better controlled in the conditions with larger numbers of different phoneme comparisons. In order to investigate 3-feature differences further, Experiment 2 would require replication and extension in a language with a fuller contrastive set of consonants. For the present, we note that the effects when competing phonemes differ by either one or two features confirm the findings from Experiment 1 and extend them by showing that, when they can be measured appropriately, differences in manner of articulation (e.g., /t/ vs. /s/) cause variation but, again, only in cases where the competing phoneme does not differ by a second feature.

General Discussion

In this paper, we have provided clear evidence from two experiments that the articulation of onset phonemes in tongue-twisters is affected by competition with other

onset phonemes; and that when the competing phonemes differ by one feature, the effect on articulation is greater than when the difference is two features. Moreover, these findings are not based on the categorization of recorded sounds as 'errors'.

A standard account of this effect incorporates representations of phonemes and features, and suggests that there is feedback from the feature to the phoneme levels (Dell, 1986; Stemberger, 1982, 1985). According to this account activation from phonemic representations flows forward to activate featural representations, which in turn feeds back to reinforce the phonemic representation. Competing phonemic representations receive more reinforcement if they share feature representations and less reinforcement if they do not share feature representations. Note that an entailment of this view is that activation from unintended, but mistakenly activated, phonemes must be allowed to feed forward to features; in other words, even in models which presuppose whole-phoneme substitutions, such as Dell's (1986) model, there is a limited form of cascading prior to selection and articulation. Where our findings differ from previous demonstrations of phonemic similarity effects is that they are compatible with the view that there is no selection stage, such that the articulation recorded represents the combined influences of activated and partially activated phonemes (cf. Goldrick & Blumstein, 2006; McMillan et al., 2009). In other words, the perturbations in articulation measured in our experiments are predictable on the basis of a planning mechanism that predicts that similar phonemes are more likely to interfere with one another.

One important consideration is that the effects reported here may depend on overt articulation. Oppenheim and Dell (2008) report that phonemic similarity effects are not found in inner speech, but only when participants are asked to speak aloud (in contrast to lexical bias effects which are found in both cases; but cp. Corley et al., in press). In Oppenheim and Dell's case it is clearly impossible to measure 'inner articulation' and the comparison is of self-reported canonical errors. Even given this caveat, however, the paper raises the intriguing possibility that the 'features' to which

we have been referring throughout are in fact 'gestures' (e.g., Browman & Goldstein, 1989), which are only activated (and thus can only feed back) when the speech plan results in articulatory movements. Proponents of the gestural view have suggested that, rather than being affected by feedback, speech errors are the outcome of coordination relations between gestures (Goldstein et al., 2007; Pouplier, 2008), such that gestural 'atoms' are combined to form 'molecules'. These relations are based on the timings of gestures in the articulatory plan, and can therefore be seen as distinct from the phonemic level we have proposed as the organizing principle for features. Underlying the timing-based account is the general observation that executing repeating actions in phase with each other is easy; and that coupled oscillators will tend towards rhythmic synchronization (see Pikovsky, Rosenblum, & Kurths, 2001). To the extent that onsets in tongue-twisters share gestures (here defined as local constrictions within the vocal tract by articulators such as the tongue tip, velum, or larynx), the tendency for these gestures to propagate in phase will be increased. When participants repeat phrases such as *tom kom*, gestures associated with /t/ are likely to be repeated when /k/ is produced (and vice-versa) because there is a reasonable degree of gestural overlap between /t/ and /k/ (more than /t/ and /g/).

A defining aspect of this view is that similarity is not associated with the production of individual phonemes, but rather with the production of gestural scores. Thus an additional contributor to the /t-k/ interference in *tom kom* is the fact that the phrase includes two /m/s. To produce the phrase correctly participants must produce either /t/ or /k/ before each /m/, but there may be a tendency to synchronize on the faster frequency of /m/ repetitions. Pouplier (2008) tested this hypothesis by comparing the articulation of phrases such as *top cop* with that of *taa kaa*, where there is no competing frequency of articulatory gesture. Using a technique in which ultrasound was recorded and articulations were categorized on the basis of traced tongue contours, Pouplier showed that articulatory intrusions were more common for *top cop*, as predicted within this framework (results from /s-ʃ/

alternations were broadly in line with this pattern). Here the emphasis is not on the overall error rate (which replicates the well-known phoneme repetition effect, e.g., Nooteboom, 1969), but on the type of error: Intrusions are interpreted as wrongly-synchronized gestures, in line with the timing-based account.

In fact the tendency towards intrusion errors does not distinguish a timing-based from a feedback-based account. In a cascading framework gestural intrusion can be characterized as the (partial) performance of those gestures which are associated with the (partial) activations of competitor phonemes. The bias towards gestural intrusion (as opposed to deletion or substitution) follows naturally from the fact that activation cascades to production. Any organizing principle for subphonemic representations, whether gestural or phonemic, would ensure that gestures associated with competitor phonemes which shared subphonemic elements were more likely to be produced (and a principle which linked groups of gestures, or phonemes, to syllables would provide a similar account of the phoneme repetition effect, as in Dell, 1986).

Laboratory elicitations of speech errors tend to be based on competition in alternating sequences such as tongue-twisters (see Baars, 1992, for a variety of techniques). In this context, a view based on synchronization of oscillators is difficult to rule out. We have tended to focus on feedback between representational levels because it provides a parsimonious framework in which to consider the influences of phonemes (this paper), syllables (Nooteboom, 1969), and words (e.g., Goldrick & Blumstein, 2006; McMillan et al., 2009) on articulation. What the timing-based and feedback accounts share, however, is the notion that articulatory gestures are not produced independently, but are related to each other via some other mechanism; and this mechanism is embedded in pre-motoric representations, be they phonemes or gestures.

Before continuing, it is worth considering one other type of feedback mechanism which may account for phonemic similarity. As Rapp and Goldrick (2004) have pointed out, any model in which later stages of processes affect earlier ones can be considered

to be a feedback model. A potential source of such an effect would be the self-monitor. According to Levelt (1989), self-monitoring of a speaker's intended speech is possible because the speech plan is represented in phonemic units that can be parsed by the comprehension system (see also Levelt et al., 1999). Monitoring accounts of the lexical bias effect (that phoneme substitutions are more likely to result in real words than would be predicted by chance) propose that the perceptual system can detect nonwords and edit the speech plan to remove them if necessary. This type of monitor cannot account for the effects of phonemic similarity: Even if errors are canonical, *meat puppets* mispronounced as "peat muppets" yields real word outcomes, and the perceptual system would not detect an error. To account for the effects of phonemic similarity in a monitoring-based framework, Nooteboom (2005a, 2005b) proposed that the monitor must have access to the intended utterance. According to this account, speakers are less likely to detect an error in a speech plan that if it sounds similar to the intended utterance. However, this raises two issues. First, in a cascading framework, it would be necessary to define a 'cutoff' below which articulatory deviation was not considered to be an error. Second, even if this problem were soluble, it is not clear how a monitor with access to the intended utterance would be implemented. How can a correct intended speech plan be maintained for comparison and, if this is possible, why is an incorrect plan generated? Although refinements to a monitoring approach may circumvent these problems, it seems unlikely that a simple and plausible monitoring-based account of phonemic similarity effects in articulation is possible.

The view that activation cascades to articulation has direct repercussions for the investigation of speech errors. First, the concept of canonical errors, in which whole phonemes are errorfully substituted for one another, must be revised. In the context of a cascading model, all articulatory output represents the activation levels of competing representations. If competition is weak, there may be no discernible competitor activation and articulation may appear to be 'correct'. As competition increases, noncanonical articulations representing the combined activations of target and

competitor will be produced. Since there are no clear distinctions in the articulatory output, the category boundary for a 'canonical error' will have to be operationally defined (possible definitions include cases where competitor activation exceeds target activation, or cases where there is no discernible target activation).

Because of the boundary problem, we should consider carefully the relevance to research questions of interest of methods in which spoken responses are transcribed by researchers (in itself a form of categorization) and categorized as errors. We should be equally wary of some articulatory methods. For example, in Pouplier's (2008) study, errors are identified by first tracing the tongue's contour, and then measuring the tongue dorsum height (and tip slope, where this is possible). Articulations are then categorized as 'errorful' or 'correct', based on inner-quartile means of the measured attributes. Despite the continuous nature of tongue movements, this is essentially a categorical study, and very few (around 4%) of the observed articulations are categorized as errors. Different boundaries between categories may have resulted in different distributions of errors, directly affecting the conclusions drawn. This is not to say that these methods are not useful, for example, in determining the kinds of tongue movements that contribute to errorful articulations. In drawing conclusions from the distributions of 'errors' that are observed, however, their utility may be limited; and any method which quantifies numbers of errors based on a category distinction is subject to similar criticism.

Addressing these issues requires the development of methods which allow for variability in the articulation of speech. Accordingly, an important aspect of the present paper has been the development of a general approach to the measurement of articulatory similarity. Although we report three measures in this paper, the underlying principles of each quantification method are the same. In each case, we measure the Euclidean distance between a particular instance of phoneme articulation and an averaged reference sample. This approach makes few assumptions, other than the basic assumption that similar articulations will result in similar recorded data

patterns (be they onset latencies in the acoustic record, records of contact over time in EPG, or pixel luminances in ultrasound video). A potential limitation of an analysis method that makes no assumptions about the spatial or temporal properties of articulation is that the end result is an abstract measure of articulation. That is, it is not possible to interpret x delta units as a meaningful measure without performing some relative comparison. Moreover, the method does not allow us to capture the details of individual articulations, and must therefore be seen as complementary to established phonetic and articulatory methods. Despite these limitations, however, the Delta method allows us to test experimental hypotheses that have previously not been testable. Rather than trying to identify instances where 'errors' have been produced, we simply compare *all* recorded articulations under different experimental conditions. This has the advantage that conclusions are drawn from all of the available recorded data (in the present case consisting of 98.7% of recorded items in Experiment 1, and 98.6% in Experiment 2). A direct comparison of variability across conditions removes the need to examine the distributions of small numbers of responses which have been selected according to an arbitrary categorization.

The usefulness of this approach is clearly demonstrated in the present paper. However, it is important to note that the Delta method is predicated on the assumption that articulation reflects the simultaneous activations of more than one representation. It is conceivable that this is not the case. Mean Delta for a given condition could reflect a bimodal distribution of 'correct' and 'error' responses. More 'error' responses would result in higher Delta, but the analysis would fail to capture the categorical nature of participants' responses. In order to gain a clearer picture of the articulatory variance which contributed to our analysis, we examined the distributions of a subset of the data, consisting of all those cases in which voiceless phonemes competed with other phonemes in Experiment 2. Figures 9 and 10 show how VOT was affected by the differences between phonemes in the experimental tongue-twisters. They show the distributions of recorded VOTs separately for the

cases in which /t/ and /k/ competed with each of the phonemes /t,d,k,g/. (VOT differences for analysis were derived from the depicted VOTs by subtracting each participant's mean control VOTs, as described above.) Note that the classification of responses was entirely determined by experimental condition, and was therefore insensitive to the particular articulations produced by participants. For this reason the figures cannot be directly compared to typical VOT values for /t, k/ or for any other phoneme; importantly, however, they show that there was no systematic pattern in the VOT differences which contributed to the measures we reported.

Insert Figure 9 about here

Insert Figure 10 about here

Insert Figure 11 about here

Insert Figure 12 about here

Insert Figure 13 about here

Figures 11, 12 and 13 show how Delta was affected for intended /t, k/ and /s/ onsets respectively, compared with each of the phonemes /t,d,k,g,s,z/. Again, there is no evidence to suggest that canonical phonemes are being produced in error. Taken

together, these figures, together with figures 9 and 10, confirm the assumption that the articulation of phonemes is variable, lending weight to the Delta method as an analytical tool.

Throughout this paper we have argued that quantifying articulation, rather than categorizing it as errorful or correct, provides a means to empirically evaluate the extent to which information from a phonemic level influences resulting articulation. A potential limitation of such a theoretical framework is that the articulatory signal, as measured with EPG or ultrasound, is constrained by the degrees of freedom in which the tongue can freely move. For example, our analyses make no concessions to the fact that the tongue may be more likely to make contact with particular palate regions than others, or more generally, may have more freedom to move when articulating particular phonemes than when articulating others, and as a consequence, evidence based on a small number of phoneme pairings (such as where 3 features compete in Experiment 2) may be more difficult to interpret. This may have consequences for investigating whether phonemic similarity effects are asymmetric (Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1991). On the other hand, the method is robust enough to capture these differences in variability: Mean ultrasound deviance for all phonemes in Experiment 2 with alveolar contact (/t,s,d,z/) is 564.4, whereas for velar phonemes (/k,g/) it is 582.4, reflecting the fact that with velar contact the tongue tip is more free to move. Statistically, a mixed-effects model with subjects and items as random factors and place (alveolar or velar) as a fixed factor established that this difference was significant: log likelihood -25973, $t = 5.05$, $p_{mc} = 0.001$, effect = 16.0 (95% CI: 9.3–21.8).

As has been demonstrated throughout this paper, the strength of the Delta method is that it is able to characterize articulatory variation across experimental conditions and groups of participants, without requiring the categorization of responses. This allows us to investigate the general effects of phonemic competition on the ways in which speech segments are articulated. Within a new conceptual framework, with new theoretical consequences, the old and previously well-established

phonemic similarity effect has been re-established, showing that there is a tight coupling between articulation and the mental processes which drive it.

References

Articulate Instruments Ltd. (2007a). Articulate Assistant Advanced user guide: Version 2.07 [Computer software manual]. Edinburgh, UK.

Articulate Instruments Ltd. (2007b). Articulate Assistant user guide: Version 1.16 [Computer software manual]. Edinburgh, UK.

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.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge, UK: 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.

Bates, D., Maechler, M., & Dai, B. (2008). Lme4: Linear mixed-effects models using S4 classes [Computer software manual]. (R package version 0.999375-25)

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

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

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

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.

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

Corley, M., Brocklehurst, P. H., & Moat, H. S. (in press). Error biases in inner and overt speech: Evidence from tonguetwisters. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

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

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

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.

Dunlap, W. P., & Kemery, E. R. (1987). Failure to detect moderating effects: Is multicollinearity the problem? *Psychological Bulletin*, *102*, 418–420.

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: Proceedings of the 2005 LSA workshop* (pp. 155–172). Cambridge, MA: MIT Working Papers in Linguistics, vol.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.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production, vol.1, Speech and talk* (pp. 177–220). New York, NY: Academic Press.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*,

21(7–8), 817–855.

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.

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

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. (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.

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.

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

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

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.

McMillan, C. T. (2008). *Articulatory evidence for interactivity in speech production.* Unpublished doctoral dissertation, University of Edinburgh.

McMillan, C. T., Corley, M., & Lickley, R. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, *24*, 44–66.

Meringer, R., & Mayer, C. (1978). *Versprechen und verlesen: eine psychologisch-linguistiche studie* [Misspeaking and misreading: A psycholinguistic study]. Amsterdam: John Benjamins. (Original work published 1895)

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

Nooteboom, S. G. (1969). The tongue slips 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*, 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.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*, 528–537.

Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization: A universal concept in the nonlinear sciences.* Cambridge, UK: Cambridge University Press.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language*

and Speech, *39*, 375–392.

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*, 311–341.

Pouplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics*, *36*, 114–140.

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org`

Rapp, B., & Goldrick, M. (2004). Feedback by any other name is still interactivity: A reply to Roelofs (2004). *Psychological Review*, *111*, 573–578.

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. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (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., & 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.

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. (1985). 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. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language*, *30*, 161–185.

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

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics*, *11*, 90–105.

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.

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.

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

Appendix

Experimental Stimuli

Experiment 1

Control Items. duv duv duv duv, giv giv giv giv, kef kef kef kef, tuf tuf tuf tuf,

Experimental Items. 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 dif, div giv giv div, duf guf guf duf, duv guv guv duv, gef def def gef, gev dev dev gev, gif dif dif gif, 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, tev gev gev tev, tif gif gif tif, tiv giv giv tiv, tuf guf guf tuf, tuv guv guv tuv

Experiment 2

Control Items. dom dom dom dom, gom gom gom gom, kom kom kom kom, som som som som, tom tom tom tom, zom zom zom zom

Experimental Items. 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,

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

Author Note

Portions of this work were presented at the AMLaP'07 conference in Turku, Finland. This work was partially funded by an NRSA Kirschstein Fellowship Award (F31 DC07282) from the National Institute on Deafness and Communication Disorders (NIH/NIDCD) while the first author was a PhD student at the University of Edinburgh. The authors are grateful to Robin Lickley for help and advice in the execution of the experiments reported here, to Suzy Moat for discussion of theoretical aspects of the work, and to three anonymous reviewers who provided helpful feedback on an earlier version of this manuscript.

Footnotes

[1]Note that the control sequences also eliminate potential effects of coarticulation (between /k/ and /d/ in the example tongue-twister). We considered using ABAB control sequences (e.g., *kef def kef def*) but decided against this as Wilshire (1999) reports that ABAB and ABBA sequences give rise to equal numbers of phoneme substitution errors, even at slow speech rates.

Table 1

Grand means (M) and standard deviations (SD) of VOT deviance by experimental condition in Experiment 1 (ms)

	Place of Articulation			
	No Change		Change	
Voicing	M	SD	M	SD
No Change	9.09	8.78	10.56	9.77
Change	12.21	12.00	11.06	10.87

Table 2

Grand means (M) and standard deviations (SD) of articulatory variation (Delta) by experimental condition in Experiment 1 (arbitrary units)

	Place of Articulation			
	No Change		Change	
Voicing	M	SD	M	SD
No Change	1.72	1.91	3.16	3.38
Change	2.63	1.86	2.99	2.24

Table 3

Grand means (M) and standard deviations (SD) of VOT deviance by experimental condition in Experiment 2 (subset excluding /s/ and /z/) (ms)

	Place of Articulation			
	No Change		Change	
Voicing	M	SD	M	SD
No Change	6.82	6.42	9.13	8.63
Change	11.11	11.16	10.12	10.77

Table 4

Grand means (M) and standard deviations (SD) of articulatory variation (Delta) by experimental condition in Experiment 2 (subset excluding /s/ and /z/) (arbitrary units)

	Place of Articulation			
	No Change		Change	
Voicing	M	SD	M	SD
No Change	482.84	96.34	602.58	115.32
Change	555.34	104.49	601.86	107.48

Table 5

Grand means (M) and standard deviations (SD) of articulatory variation (Delta) by experimental condition in Experiment 2 (all data) (arbitrary units)

	Manner of Articulation							
	No Change				Change			
	Place of Articulation				Place of Articulation			
	No Change		Change		No Change		Change	
Voicing	M	SD	M	SD	M	SD	M	SD
No Change	471.05	101.74	601.92	115.60	572.15	124.84	637.77	135.75
Change	550.28	116.42	601.42	107.47	561.39	121.22	623.42	126.50

Figure Captions

Figure 1. Example EPG recordings, trimmed to include the palates immediately before and after closure: (a) a velar articulatory record with full closure; (b) a velar articulatory record which does not contain visible full closure and is therefore trimmed to include the palate before maximal closure through to the palate after maximal closure.

Figure 2. Markov Chain Monte Carlo (MCMC) estimates of the effects of changes in Place and Voice on VOT deviance (ms) in the EPG analysis of tongue-twisters from Experiment 1. Error bars show 95% confidence intervals.

Figure 3. Markov Chain Monte Carlo (MCMC) estimates of the effects of changes in Place and Voice on articulatory variation (Delta) in the EPG analysis of tongue-twisters from Experiment 1. Error bars show 95% confidence intervals.

Figure 4. Example frame from recorded ultrasound, showing the region representing tongue activity (a) as recorded and (b) pixelized prior to the calculation of Delta. The tongue root is on the left and the tongue-tip on the right of the image. Note that Depth information and other indicators in the image are invariant, and the corresponding pixels therefore contribute zero to calculations of Delta.

Figure 5. Comparison of 16 /k/, 16 /t/, and 16 /s/ articulations recorded with ultrasound. Delta was calculated for each articulation relative to every other articulation. The dimensionality of the resulting deviance scores was reduced using multidimensional scaling.

Figure 6. Markov Chain Monte Carlo (MCMC) estimates of the effects of changes in Place and Voice on VOT deviance (ms) in the EPG analysis of tongue-twisters from Experiment 2 (subset excluding /s/ and /z/). Error bars show 95% confidence intervals.

Figure 7. Markov Chain Monte Carlo (MCMC) estimates of the effects of changes in Place and Voice on articulatory variation (Delta) in the EPG analysis of tongue-twisters from Experiment 2 (subset excluding /s/ and /z/). Error bars show 95% confidence intervals.

Figure 8. Markov Chain Monte Carlo (MCMC) estimates of the effects of changes in Place, Voice and Manner on articulatory variation (Delta) in the EPG analysis of tongue-twisters from Experiment 2 (all data). Error bars show 95% confidence intervals.

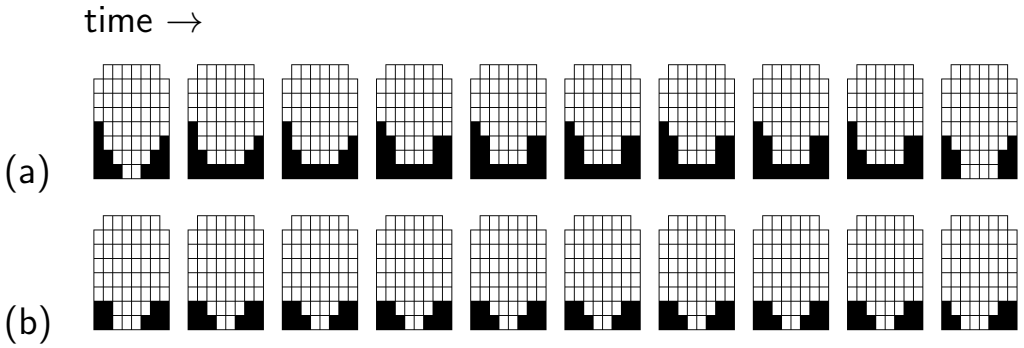
Figure 9. Distributions of VOTs for cases in which participants attempted to produce /t/ onsets (8 participants, 505 observations).

Figure 10. Distributions of VOTs for cases in which participants attempted to produce /k/ onsets (8 participants, 509 observations).

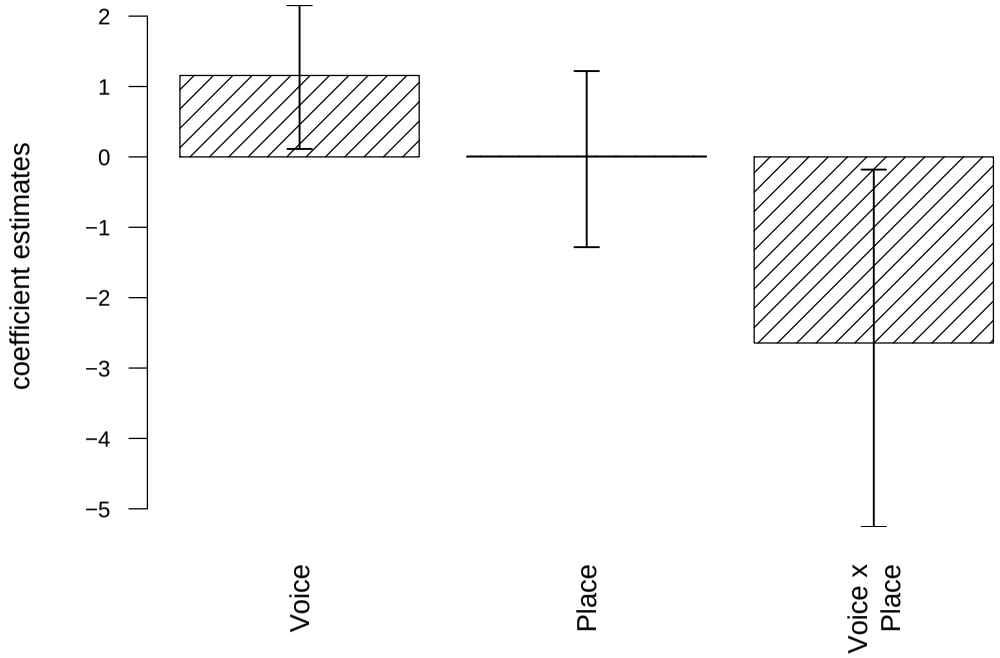
Figure 11. Distributions of Delta for cases in which participants attempted to produce /t/ onsets (8 participants, 767 observations).

Figure 12. Distributions of Delta for cases in which participants attempted to produce /k/ onsets (8 participants, 768 observations).

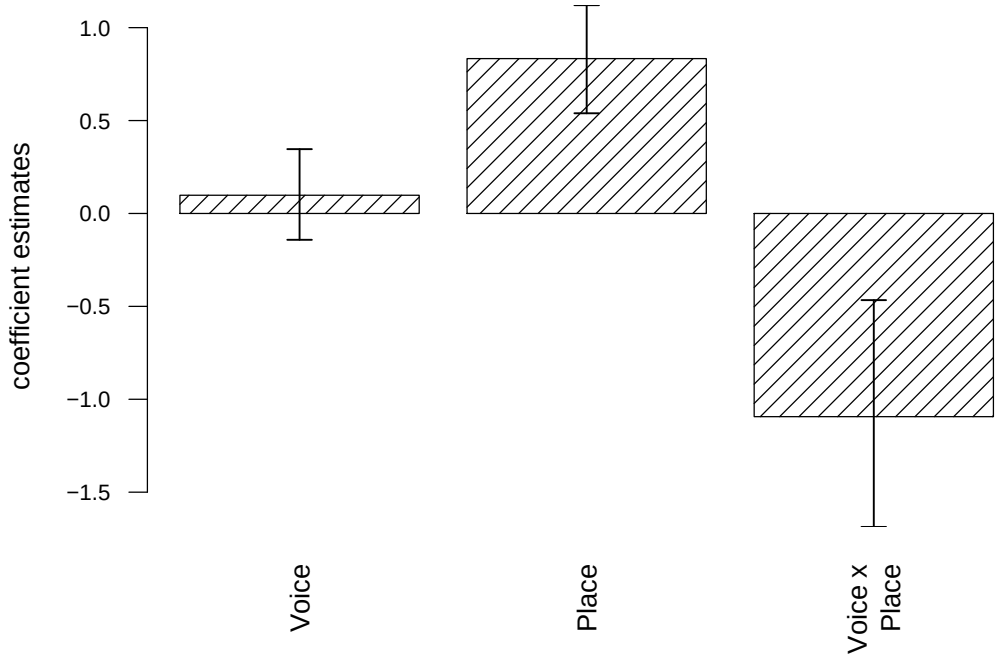
Figure 13. Distributions of Delta for cases in which participants attempted to produce /s/ onsets (8 participants, 752 observations).


time →
(a)
(b)

Cascading Influences on the Production of Speech, Figure 2

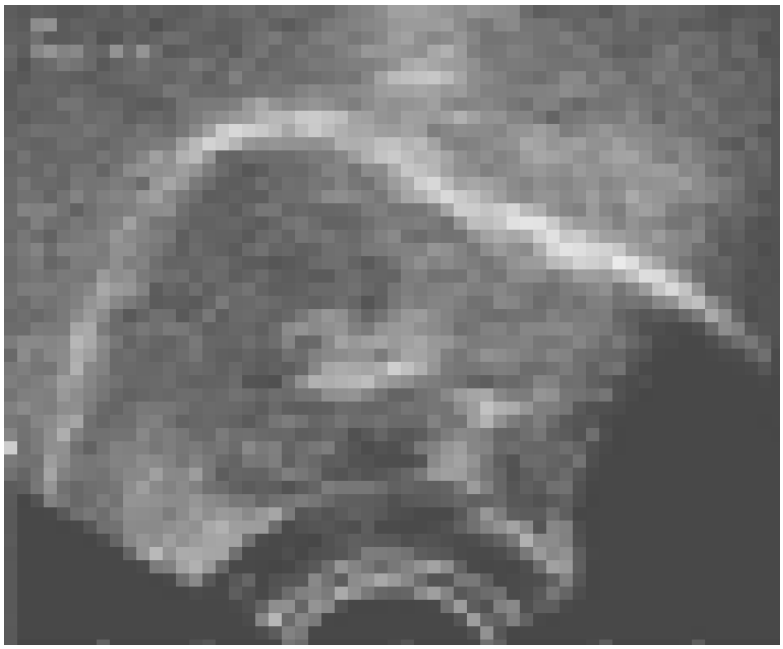


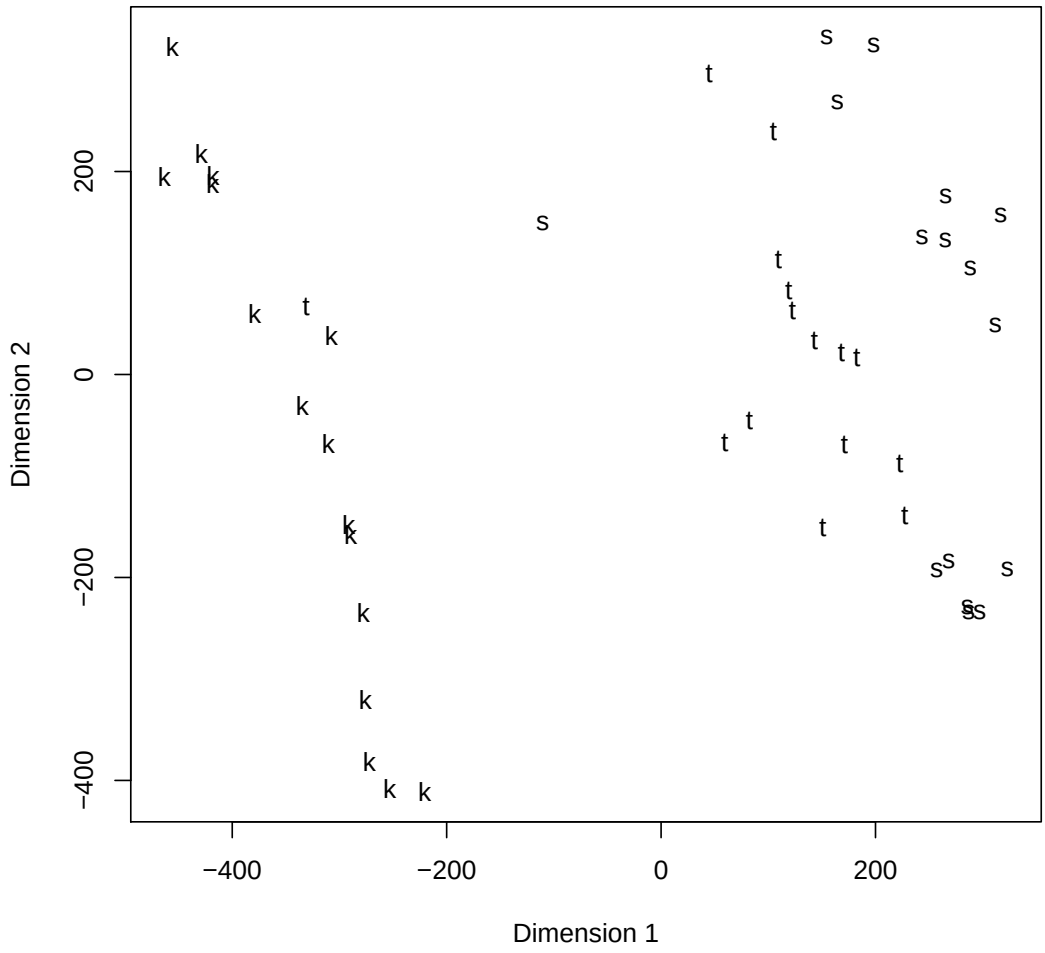
Cascading Influences on the Production of Speech, Figure 3

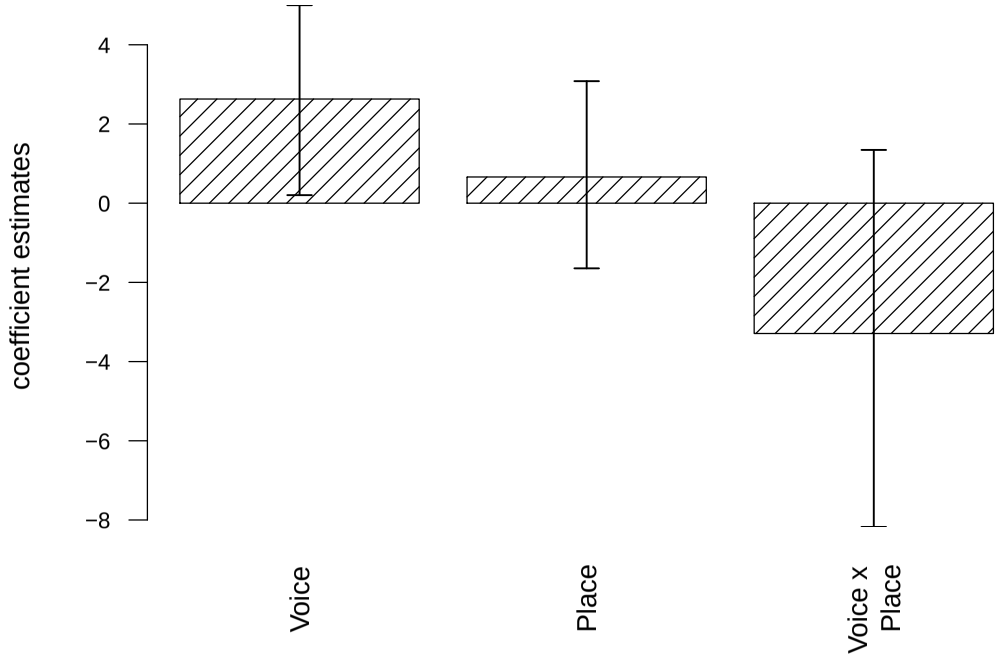


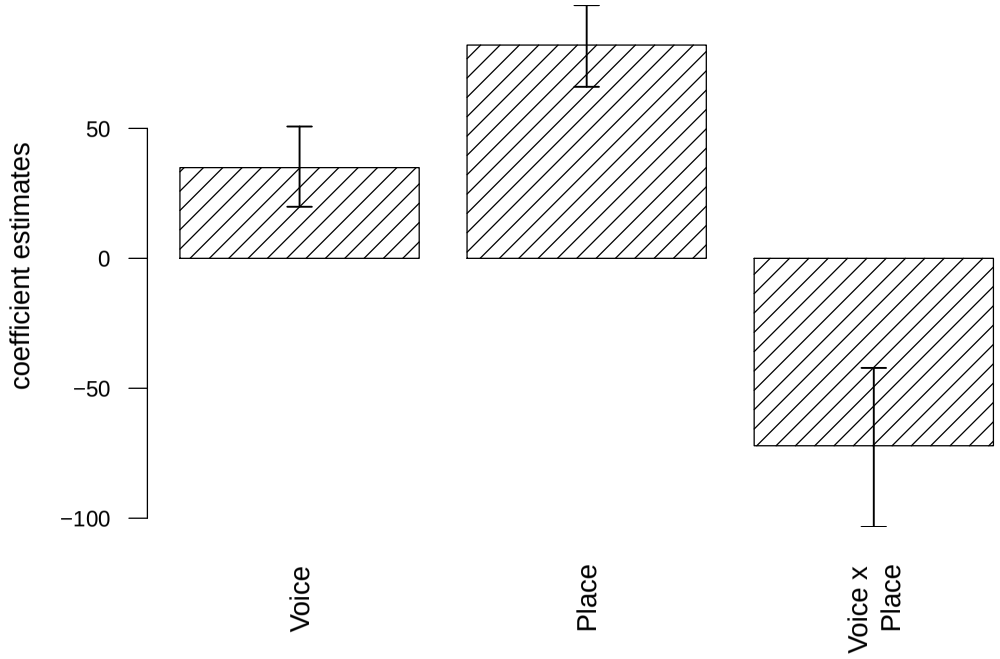
(a)

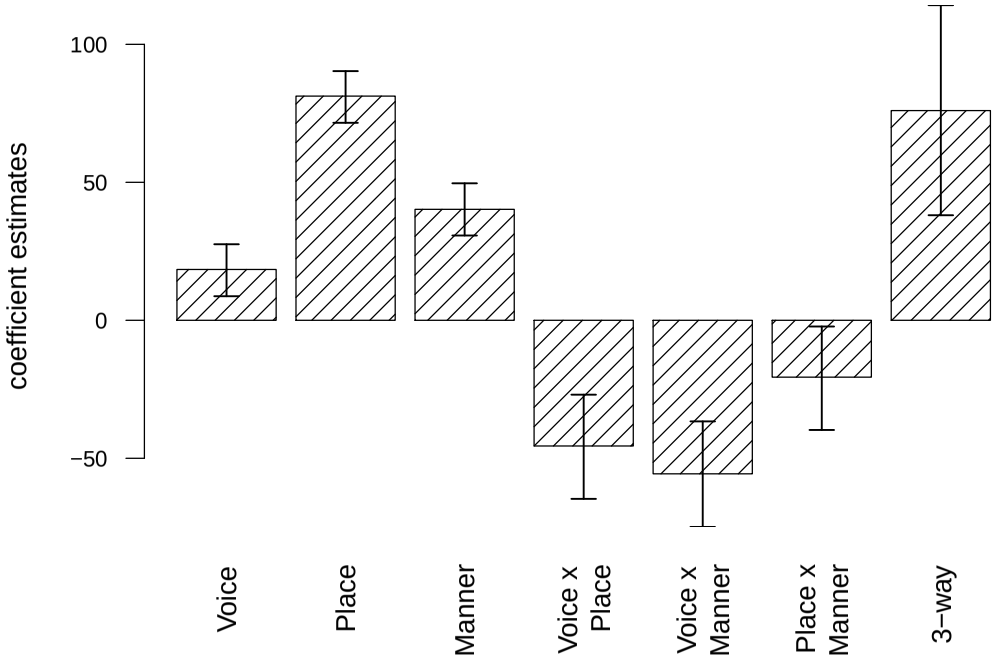
(b)

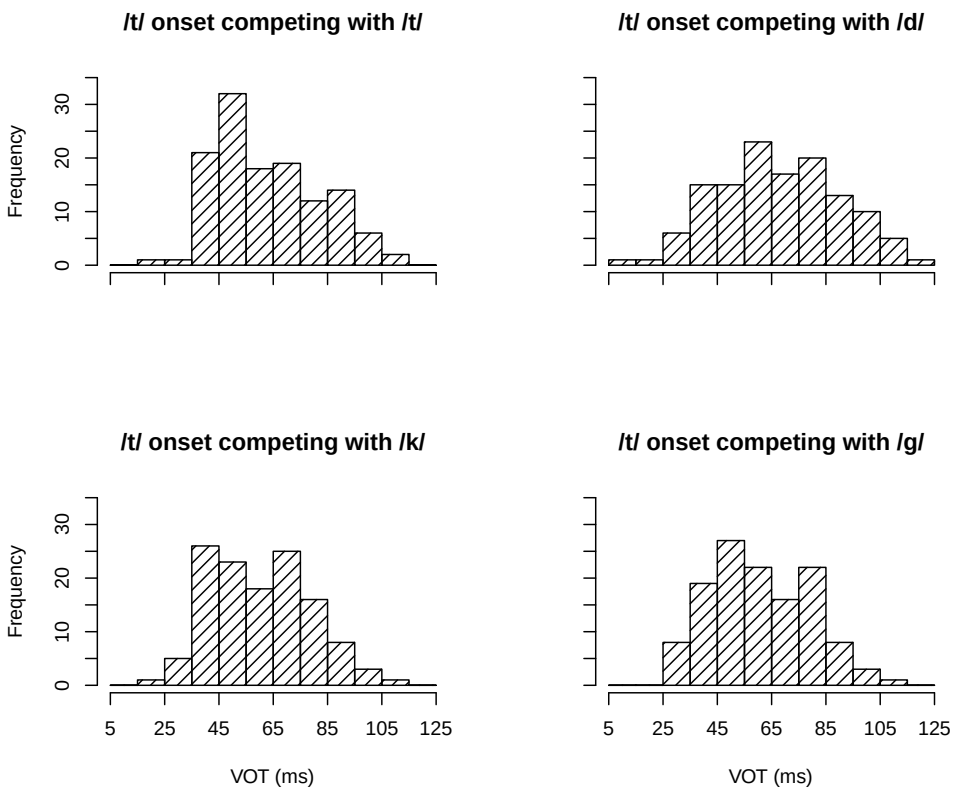


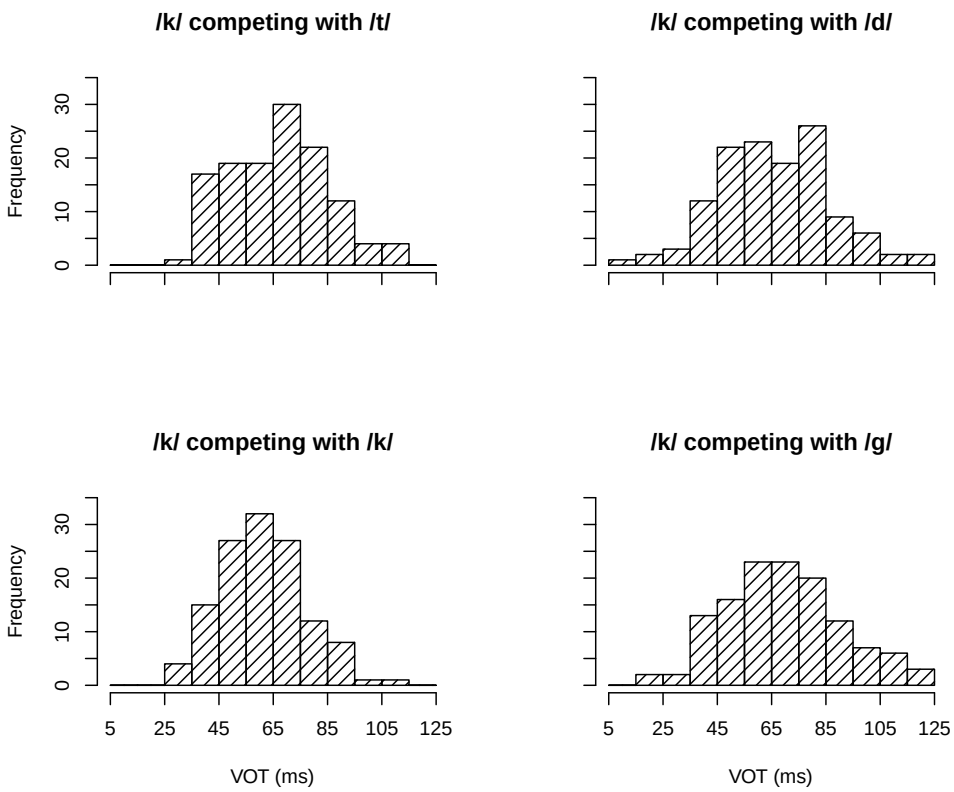

Dimension 2
Dimension 1
−400
−200
0
200
−400
−200
0
200

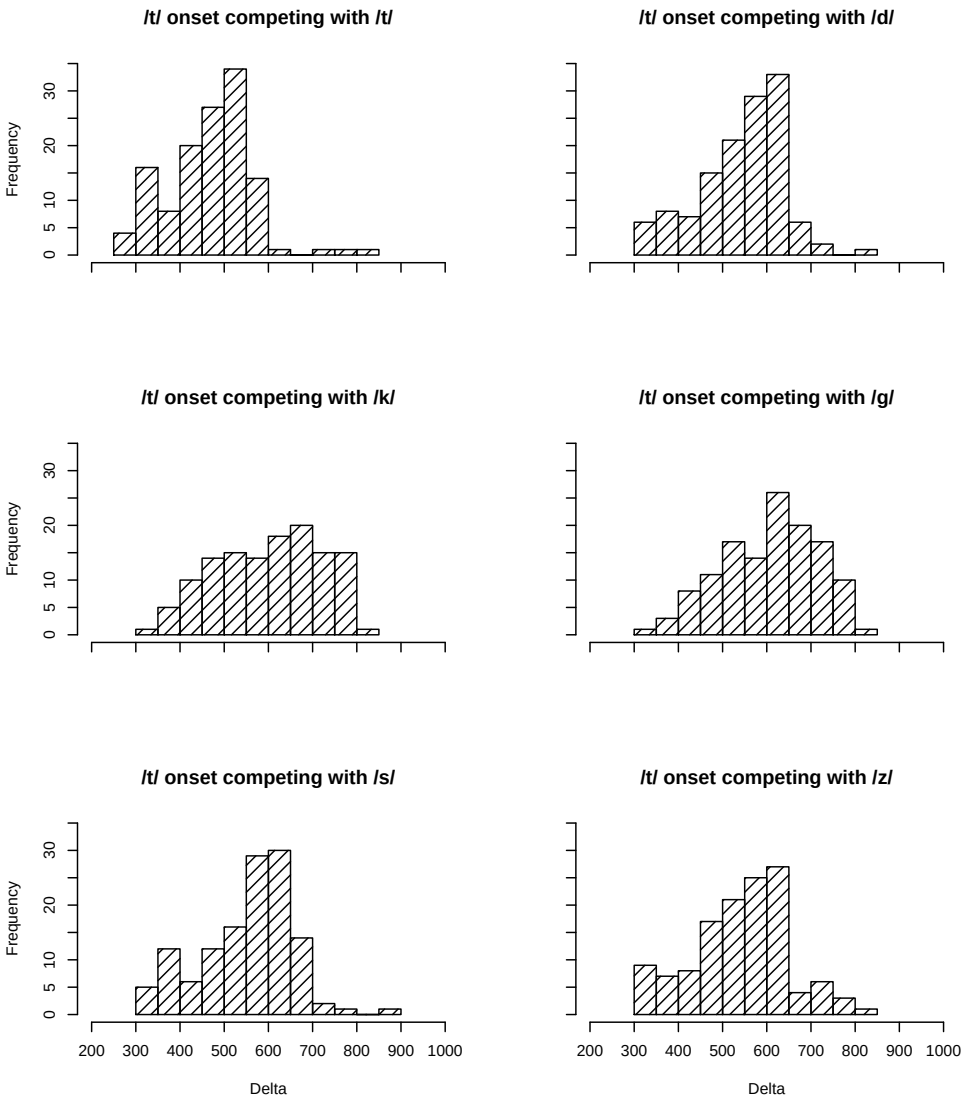

coefficient estimates
4
2
0
−2
−4
−6
−8
Voice
Place
Voice x
Place

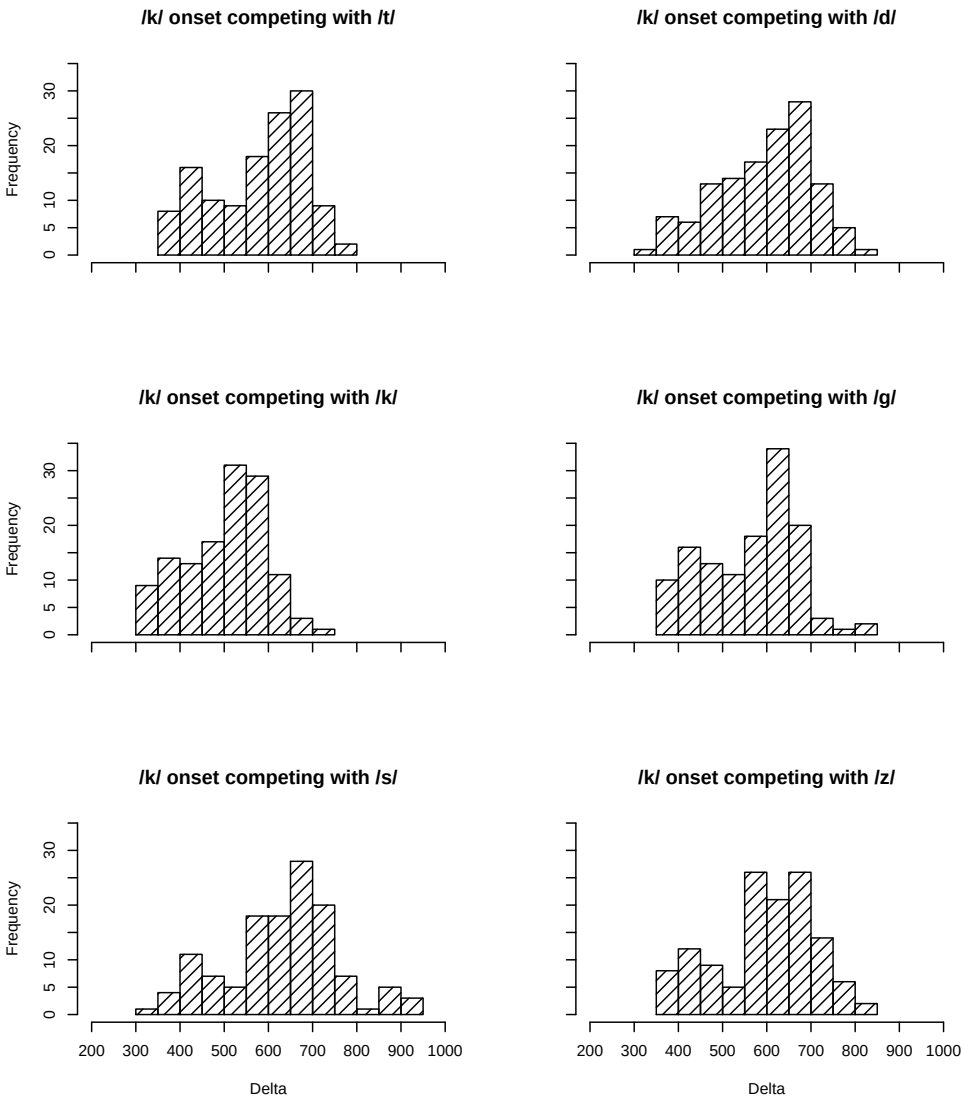

coefficient estimates
50
0
−50
−100
Voice
Place
Voice x
Place

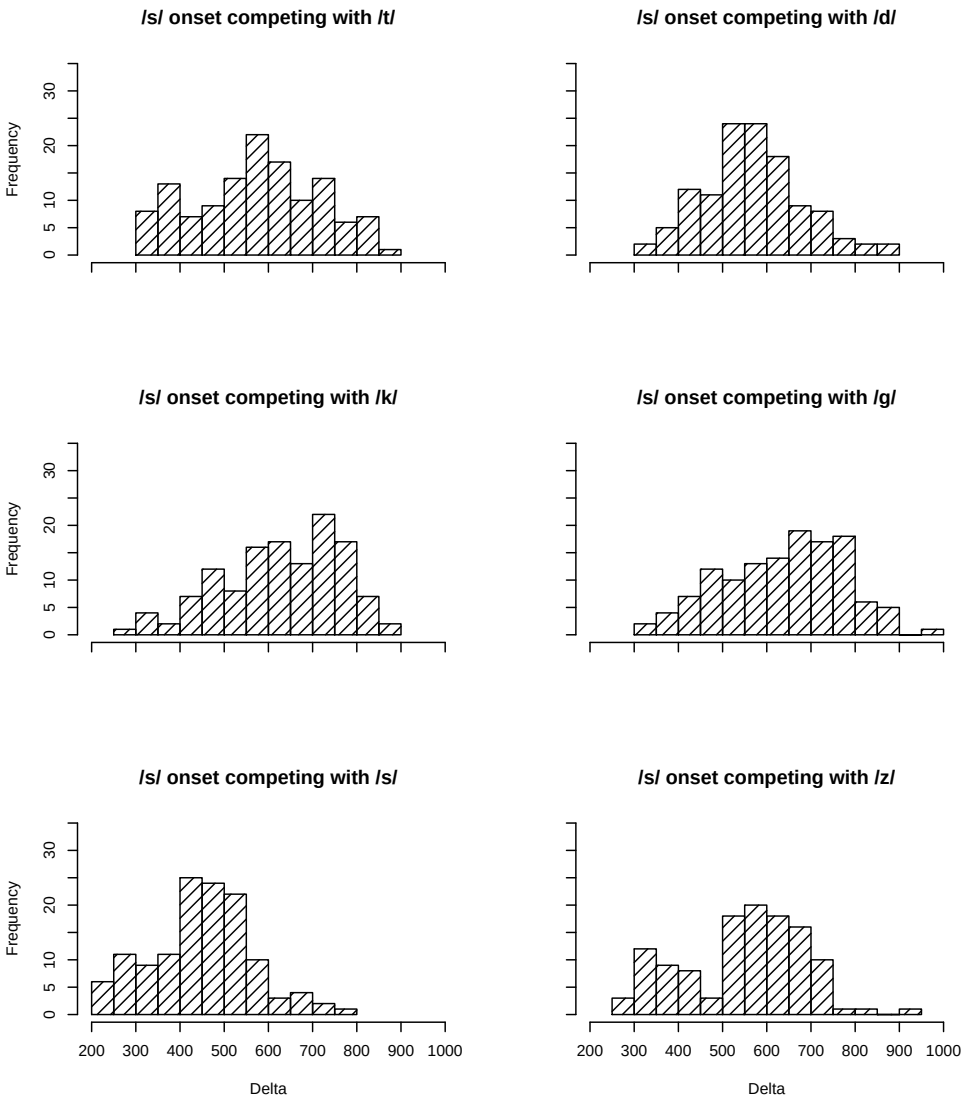

coefficient estimates
100
50
0
−50
Voice
Place
Manner
Voice x Place
Voice x Manner
Place x Manner
3−way


/t/ onset competing with /t/
/t/ onset competing with /d/
/t/ onset competing with /k/
/t/ onset competing with /g/
Frequency
0
10
20
30
5
25
45
65
85
105
125
VOT (ms)


/k/ competing with /t/
/k/ competing with /d/
/k/ competing with /k/
/k/ competing with /g/
Frequency
0
10
20
30
5
25
45
65
85
105
125
VOT (ms)


/t/ onset competing with /t/
/t/ onset competing with /d/
/t/ onset competing with /k/
/t/ onset competing with /g/
/t/ onset competing with /s/
/t/ onset competing with /z/
Frequency
0
5
10
20
30
200
300
400
500
600
700
800
900
1000
Delta


/k/ onset competing with /t/
/k/ onset competing with /d/
/k/ onset competing with /k/
/k/ onset competing with /g/
/k/ onset competing with /s/
/k/ onset competing with /z/
Frequency
0
5
10
20
30
200
300
400
500
600
700
800
900
1000
Delta


/s/ onset competing with /t/
/s/ onset competing with /d/
/s/ onset competing with /k/
/s/ onset competing with /g/
/s/ onset competing with /s/
/s/ onset competing with /z/
Frequency
0
5
10
20
30
200
300
400
500
600
700
800
900
1000
Delta