Research Article

Towards the Quantification of Peggy Babcock: Speech Errors and Their Position within the Word

Christine Mooshammer[a, b]   Mark Tiede[b]
Stefanie Shattuck-Hufnagel[d]   Louis Goldstein[b, c]

[a]Institut für deutsche Sprache und Linguistik, Humboldt Universität zu Berlin, Berlin, Germany; [b]Haskins Laboratories, New Haven, CT, USA; [c]Department of Linguistics, University of Southern California, Los Angeles, CA, USA; [d]Speech Communication Group, Research Laboratory of Electronics, MIT, Cambridge, MA, USA

Abstract

Sequences of similar (i.e., partially identical) words can be hard to say, as indicated by error frequencies, longer reaction and execution times. This study investigates the role of the location of this partial identity and the accompanying differences, i.e. whether errors are more frequent with mismatches in word onsets (*top cop*), codas (*top tock*) or both (*pop tot*). Number of syllables (*tippy ticky*) and empty positions (*top ta*) were also varied. Since the gradient nature of errors can be difficult to determine acoustically, articulatory data were investigated. Articulator movements were recorded using electromagnetic articulography, for up to 9 speakers of American English repeatedly producing 2-word sequences to an accelerating metronome. Most word pairs showed more intrusions and greater variability in coda than in onset position, in contrast to the predominance of onset position errors in corpora from perceptual observation.   © 2018 S. Karger AG, Basel

Introduction

It is well known that certain utterances are more difficult to produce correctly (without producing errors) than others. A prime example is the tongue twister *Peggy Babcock* for which the following errors have been reported by Butterworth and Whitacker (1980): *Bagcock, Bagpock, Bagpop*. Examples of

Christine Mooshammer
Institut für deutsche Sprache und Linguistik
Humboldt-Universität Berlin
DE–10999 Berlin (Germany)
E-Mail christine.mooshammer@hu-berlin.de

tongue twisters (error-inducing utterances) have been collected from many languages (the First International Collection of Tongue-twisters, http://www.tongue-twister.net includes samples from 118 languages), and it seems likely that examples could be found for any language. Understanding why some utterance types should be error-prone and others not is a challenge for models of speech production. For example, in models which include both a planning and execution component (e.g., Goldstein et al., 2007; Levelt et al., 1999), it is not clear which types of errors occur during the operation of which of these two components. To meet this challenge, experiments have manipulated the structural phonological properties of target utterances and observed their relative difficulty in order to isolate the relevant phonological factors and thereby to test the predictions of particular models. This has been attempted by either measuring the amount of time required to produce the targets, e.g. Sevald and Dell (1994), O'Sheaghdha and Marin (2000), Damian (2003) or Cohen-Goldberg (2012), or by counting the number of errors talkers produce, e.g. Butterworth and Whittaker (1980), Dell (1984) or Sevald and Dell (1994). This paper extends these findings by addressing how errors are affected by manipulating the degree of similarity within repeated word sequences. We consider this first with a focus on the role of contextual similarity and how psycholinguistic and phonological theories that are concerned with phonemic speech errors deal with this effect, followed by an examination of theories which measure the gradient variation of phonemic speech errors (for a recent overview, see also Slis, 2018).

Speech Errors in Traditional Speech Production Models

It has often been observed that one major factor that emerges as causal from studies of speech errors is partial *similarity*. Repeatedly producing two similar, but not identical, words in succession results in more errors than two phonologically unrelated words or two identical words (Meyer & Gordon, 1985; Sevald & Dell, 1994). Word sequences can be similar in several different ways, including, e.g., containing similar sounds in similar positions with similar stress. The effects of similarity can be seen not only in experimentally induced errors (Meyer, 1992), but also in corpora of naturally occurring speech errors (Dell and Reich, 1981; Fromkin, 1973; MacKay, 1971; Shattuck-Hufnagel, 1979; Vousden et al., 2000). In both types of experiments, similar segments interact more frequently in errors. This effect has been addressed by a variety of models. Some suggest that the activation of a word in the lexicon, to ready it for production, spreads to words that share phonological units with the target word (Dell, 1984); such models are particularly well suited to account for interactions between words that are not targets for the current utterance. Other models propose that interactions between the target words for a planned utterance arise because these activated words are stored in a planning buffer, from which sounds are selected during a serial ordering process (Shattuck-Hufnagel, 1979, 2015). So, in either approach, when producing a sequence of similar words, this spreading of activation causes the two words to compete over insertion of their phonological segments into the evolving plan for production. The results of this competition can be seen in errors in which a segment is produced in the wrong position of the plan (anticipations, perseverations), or errors in which more than one segment is produced concurrently (gestural intrusions, discussed be-

low: Goldstein et al., 2007) or concurrently influences execution (Goldrick & Blumstein, 2006, McMillan & Corley, 2010). As will be discussed later, phonological theories such as Shattuck-Hufnagel's serial order model and Dell's spreading activation model are based on the assumption that complete phonemes are misselected resulting in phonemic errors. Therefore, they fail to account for gradient subphonemic errors that were found in a number of experimental studies (e.g., Goldrick & Blumstein, 2006; Goldstein et al., 2007).

Importantly, there are a number of different principles or factors that govern error interactions, e.g. element similarity, position similarity and contextual similarity (see Shattuck-Hufnagel, 1992, 2015). For example, in corpus studies word-initial consonants are found to substitute for other word-initial consonants as in *Tanadian from Coronto* (Slis, 2018), and word-final consonants for other word-final consonants as in *helf lapping* instead of *help laughing* (see Dell, 1986, and Shattuck-Hufnagel, 1979, for English; Berg, 1991, for Spanish; Vousden et al., 2000, for Dutch). Experimental elicitation of errors, therefore, has also controlled for the structural position of the target elements (Baars et al., 1975; Dell, 1986). It is unclear whether the relevant structural domain for position is the word or the syllable, as much experimental work has employed monosyllabic stimuli. However, the fact that in polysyllabic words exchanges involve word-initial syllable onsets more often than word-medial syllable onsets speaks against a pure syllable position effect for English and Dutch (for discussion, see Meyer, 1992; Shattuck-Hufnagel, 1992, 2015; Vousden et al., 2000). In Spanish, however, Berg (1991) found more frequent exchanges in syllable-initial word-medial position than in the absolute onset which he attributes to differences in stress placement in Spanish and German.

Comparison of the relative sensitivity of initial versus final position (onset vs. coda in a monosyllabic word) to similarity effects has the potential to inform about the temporal course of the speech production/planning process. Sevald and Dell (1994) found that producing a string of monosyllabic words with *identical* onsets but different codas (e.g., *pick pin*) is more difficult than the converse (e.g., *pick tick*), replicating and extending an earlier finding of Butterworth and Whittaker (1980). Sevald and Dell found a slower speaking rate and higher error rates for coda mismatches, and interpreted this as evidence that when a word is activated for production, the segments composing it are activated over time sequentially (as hypothesized by Meyer, 1992, and Houghton, 1990), rather than simultaneously as has been assumed in some phonological competition models (O'Seaghdha et al., 1992; Peterson, 1991; Peterson et al., 1989). In the Sevald and Dell *sequential cuing model* (henceforth SCM), activation of onsets immediately causes spreading activation to words with similar onsets. If those activated words have discrepant units later in the word, competition will arise between those discrepant units. In the converse case, by the time a final consonant is activated, the onset of the word has already been produced (or at least inserted into the plan for execution), so the activation of competing forms sharing that final consonant is too late to have an inhibitory effect. However, Wilshire (1998) found a word-initial effect in an error elicitation task similar to the Sevald and Dell task with real words, but no word position effects for nonsense words. In Wilshire's view the word-initial effect in real words comes about because the competition between phonemes of simultaneously activated words is lower at

the word onset and builds up after the onset is produced. Since there is no lexical access for nonsense words, exchange errors are equally frequent in initial and final position.

In contrast to these experimental studies, research using corpora of naturally occurring speech errors has reported that errors, particularly exchange errors such as *Baggy Pabcock*, are more common in word onset position than in other positions in the word (e.g., MacKay, 1970; Shattuck-Hufnagel, 1987). Vousden et al. (2000) additionally showed a higher probability of syllable onset errors, even after excluding the word onset effect. To the extent that these errors are at least partially triggered by similarity later in the word, this seems to contradict the predictions of the SCM. One possible reason for the discrepancy between these studies might be a bias for listeners to detect errors more readily at word onsets compared to other positions in a word (see Browman, 1978). Such a possibility is highlighted by recent work showing that there can be a considerable disparity between counting errors based on articulatory kinematics versus listener perception. Pouplier and Goldstein (2005) examined listeners' perception of gestural intrusions, in which both an intended and an erroneous gesture are coproduced. For example, in repeating a sequence like *cop top cop top* ... kinematic errors can be found in which the main constriction gestures for /k/ (tongue dorsum) and /t/ (tongue tip) are simultaneously produced. Listeners' identification of these coproductions was found to depend on the relative magnitude of the two gestures, but was also subject to an asymmetry, such that an intrusive /k/ gesture influenced listener judgments more readily than in an intrusive /t/ gesture. It is possible that a word position bias could also influence the perception of coproductions and therefore contribute to the reported error rates in initial versus final position. Sevald and Dell (1994) found a difference between shared onsets versus shared coda in production speed as well as error rate, but it is possible that a perceptual bias could influence the speaker's self-monitoring (Hartsuiker, 2006), thus interacting with the production process in some unknown way to produce the differences in speed. For these reasons, it would be desirable to replicate the Sevald and Dell results using kinematic measures on repetition of sequences with alternating onsets versus alternating codas, e.g., *cop top* versus *pock pot* which is one of the goals here.

Another possible limitation of the SCM, frame-based models and phonological similarity models generally, is raised by results from anecdotal observations that involve the repetition of a different type of sequence: two CVC words or syllables, in which the initial and final consonants of each syllable are identical, but the consonants differ in the two units, as in, e.g., *tot pop* or the tongue twister *Peggy Babcock* (Butterworth & Whittaker, 1980). Pilot work in our laboratory had shown that such sequences are very difficult to produce and are highly error-prone, perhaps more so than alternating onset versus codas. However, from the point of view of the SCM, these should not be particularly problematic. Their onsets are similar but not identical as they would be in, e.g., *cop cot*. More generally, considering the alternating onset and coda consonants in *pop tot* separately, the context for the alternating onset (e.g., –Vp –Vt) is less similar than in, e.g., *pock tock* (–Vk Vk), and likewise the context for the alternating coda (pV– tV–) is less similar than in, e.g., *cop cot* (kV– kV–). A possible cause for *pop tot* being more difficult to repeat than the alternating onsets and codas could be some in-

Mooshammer/Tiede/Shattuck-Hufnagel/
Goldstein

teraction between onset and coda. This, however, would contradict frame-based models which are based on the observation that errors that do not preserve syllable positions occur only rarely (see e.g., Shattuck-Hufnagel, 1979; Vousden et al., 2000). So, kinematic data on this type of sequence (which will be referred to as "double mismatch") hold the promise of revealing some novel properties of the speech production and planning system.

Variability and Speech Errors

Furthermore, many instrumental production studies with controlled stimuli found that partial similarity in word sequences induced gradient subphonemic errors that could not be detected impressionistically (e.g., Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Mowrey & MacKay, 1990; for a recent overview, see Slis, 2018). The dichotomy of discrete phoneme substitutions versus gradient intrusion and reduction errors is mirrored in phonological descriptions of allophones versus gradual variants due to gestural overlap and time-dependent target undershoot (see e.g., Kühnert & Hoole, 2004; Nolan, 1992; and recently Parrell & Narayanan, 2018). Whereas phonological theories assume discrete units with countable variants, experimental work of the last 30 years has suggested otherwise. This ongoing discussion has been addressed with respect to speech errors by Goldrick and Blumstein (2006), McMillan and Corley (2010) and others within the framework of the cascading activation model. It is assumed that the synchronous activation of two (or more) units in the lexical representation is trickling down to the articulatory level in a gradient manner and, depending on the activation level of each unit, inducing variability in the temporal and spatial domain. This increased variability is restricted to the phonetic features that alternate. For example, McMillan and Corley (2010) found that alternating *def tef* sequences increase the variability in voice onset time but less in tongue-palate contact (as measured by means of electropalatography). On the other hand, in *tef kef* sequences tongue-palate contact variability increased more than voice onset time variability. Speech errors in this view are instances of more extreme variability that have perceivable acoustic effects and are therefore identified as a different phoneme. However, to our knowledge, the cascading activation model does not address or predict whether competition on the planning level leads to different variation patterns regarding the position within the word or syllable or regarding low-level articulatory effects.

In controlled alternating sequences designed to elicit errors (Pouplier, 2003), a frequency mismatch exists between the gestures that alternate and those that occur in every stimulus unit. For example, in the sequence *"cop top cop top"* the lips constrict with each coda, while the dorsal and apical constrictions associated with /k/ and /t/ alternate each onset, resulting in a 2:1 frequency relationship between the bilabial gestures in the coda and the alternating dorsal and apical gestures in the onset, respectively. Pouplier observed that unintended coproduced constrictions (intrusions) or incomplete targeted constrictions (reductions) can arise as a consequence of this alternation, either of which may be incompletely realized. Goldstein et al. (2007) advanced the explanation that because the 2:1 (base:alternating) production frequency is less stable than a 1:1 pattern, coproduced constriction errors of this type reflect a tendency to prefer the more stable pattern (cf. Haken et al., 1985). In this view, once

repetitive production of a sequence is established, it constitutes an oscillating pattern of constriction time functions that form the consonants. For example, for *cop top*, there are oscillations of the lips, tongue tip and tongue dorsum. However, because /p/ occurs in every syllable, its oscillation frequency is twice that of the tongue tip or tongue dorsum. Entrainment of these oscillations over time would lead to a shift from 2:1 mode of frequency locking to a more stable 1:1 mode of frequency locking, with the tongue tip or tongue dorsum gesture occurring in every syllable. A series of studies by Kelso and colleagues (e.g., Kelso et al., 1993) make this clear in a different domain: when index fingers of opposing hands are wagged back and forth at an accelerating rate, a phase transition occurs such that initially antiphase movements transition to in-phase movements.

Additional evidence for competition in frequency modes was found in an ultrasound study by Pouplier (2008). By comparing CV CV with CVC CVC sequences, she showed that coda consonants play a crucial role for eliciting gradient speech errors. In sequences such as *taa kaa* intrusions and reductions occurred less frequently compared with *top cop* sequences. Apart from the onset-coda asymmetries, frequency modes can also account for the finding that gradient intrusions and reductions are much more frequent than phonemic substitution errors (see Goldstein et al., 2007).

Within the framework of task dynamics (Saltzman & Munhall, 1989) and articulatory phonology (e.g., Browman & Goldstein, 1988), spatial and temporal variability follows from coupling differences on the planning level and makes explicit predictions about the relation between (one kind of) planning difficulty and token-to-token variability. Recently, a new theoretical account, the *coupling graph model* (henceforth CGM, Nam & Saltzman, 2003; Nam et al., 2009), has been developed concerning why certain structural properties (such as being an onset vs. a coda consonant) can be considered as relatively less stable than others (Browman & Goldstein, 2000; Saltzman et al., 2006): by hypothesis, gestures in different structural positions enter a different number and different types (in-phase, antiphase) of coupling relations as specified by an utterance's coupling graph, and these coupling relations are assumed to exhibit different degrees of stability and planning stabilization time. Empirical support for this hypothesis has been presented by Mooshammer et al. (2012), who showed that syllables with onsets and no codas (CV) have shorter response latencies to initiate production (planning RT) than those with codas and no onsets (VC). They explain this difference as well as the difference in timing variability observed by Byrd (1996a). She found in an electropalatographic study of C1#C2 consonants, that the C1 coda consonant generally exhibited more variability and spatial reduction than the C2 onset consonant, and also that C1 was overlapped more by C2 than the other way round (see also Byrd & Tan, 1996).

The CGM by Nam and Saltzman (2003) and Nam et al. (2009) assumes that during the planning process, the relative phases of gestural planning oscillators settle into a stable pattern, and these stabilized relative phases are used to trigger the production of their associated gestures. The oscillators stabilize at their target values more quickly in onset than in coda, because of the different topologies of the coupling structures that have been hypothesized to govern syllable onset and coda positions (Browman & Goldstein, 1988, 2000; Byrd, 1996b). This model of planning stability can account for the latency findings and also for the

token-to-token variability findings, if we assume that the coda structures (which require a longer stabilization time due to their assumed antiphase coordination pattern) may be initiated before they fully stabilize, and thus their timing will vary from trial to trial because their planning is incomplete. Mooshammer et al. (2012) argued that longer planning times for VC than for CV(C) sequences might be caused by the coupling differences in a different manner. This could also have consequences for gradient speech errors and variability in general. For CV-initial sequences, consonantal and vowel gestures are initiated at the same time, and thus because more articulators are simultaneously recruited there is less scope for variability of the remaining articulators during the initial consonant. Due to the antiphase coupling for VC the articulators are less constrained during the coda consonant and thus have more degrees of freedom.

In a similar vein, Slis and van Lieshout (2013, 2016) and Slis (2018) argued that variability caused by the phonetic context is related to how many and which articulators are recruited for executing the consonant and the co-occurring vowel gesture in the onset. For alternations of onset consonants, they found that the tongue dorsum was less prone to intrusions and reductions, and the lower lip least affected in most vowel contexts. They conclude that the less restricted the articulator of an intruding gesture is, i.e. the fewer articulators it shares with other gestures, the "better [it is] able to maintain linguistic goals and counteract pressure from coupling forces to stabilize coordination patterns" (Slis and van Lieshout 2016, p. 14). Therefore, in their view the least involved articulator is likely to produce fewer intrusions and reductions. Furthermore, there is an interaction between position within the syllable and articulator. Coronal stops in the coda position are frequently glottalized and flapped in American English (e.g. Huffman, 2005; Warner & Tucker, 2011).

Aims of This Study

In this work we investigate the relationship between variability, speech errors and position within words systematically through three experimental conditions. The first aims at investigating whether the *position* of mismatch (onset vs. coda vs. double) influences production difficulty as measured by two complementary approaches to quantify error rates and variability from articulator kinematics (described below). As was detailed above, most corpus-based studies found more speech errors in the word and syllable onset than in the coda. Tongue-twister-like elicitation studies, however, found evidence for the opposite in support of the SCM (Sevald & Dell, 1994) and the coupled oscillator planning model (Nam & Saltzman, 2003; see also Mooshammer et al., 2012). Up to now, an instrumental investigation that also detects subphonemic variation and systematically varies onset and coda alternation is still missing. In addition to single mismatch in onset or coda, we introduce here the double mismatch condition (e.g., *pop tot*) that has not been investigated yet. As was pointed out above, the SCM would predict fewer errors and less variability than for coda mismatch because nonidentical onsets do not reactivate the most recent coda. According to the frequency locking approach (cf. Goldstein et al., 2007), the double mismatch condition is assumed to elicit more intrusion and reduction errors because the rhythmic organization of the executing articulators is more complicated than the 2:1 mode in the single mismatch condition.

The second experimental condition compares the alternation of filled and empty word slots, e.g. *top cop* versus *top op*. The question here is whether more errors are elicited if actual articulators alternate compared to the alternation between constriction gestures and empty syllable slots (i.e., onsets or codas). Put differently, in the first case the syllable structure is repeated (CVC for *top cop*), while in the second case the syllable structure alternates (CVC _VC for *top op* or CVC CV_ for *top ta*), whereas the phonological content differs for one position in both conditions. Sevald et al. (1995) argued that syllable frames are stored together with the segmental string and therefore repeating syllable frames is as beneficial as repeating strings for speech planning. They found that the production time for alternating word pairs was shorter if the syllable structure was identical (e.g. in *kil kil.per*) as compared to different syllable structures (e.g., *kilp kil.per*). Assuming that a less beneficial condition also means it is more difficult to plan and execute, it could follow that these sequences are also more variable and error-prone in multiple repetitions. Consequently, for the missing condition investigated in the current study, more errors and greater variability should occur for the CVC _VC and the CVC CV_ sequences as compared to the CVC CVC sequences. Furthermore, the alternation regarding the coda (CVC CV_) might again be more error-prone for the reasons already mentioned in the first condition. An alternative outcome, namely more errors in the alternating than in the missing condition, is predicted on the motor level. In the CVC condition three consonantal articulators are involved (e.g., *top cop*) with two alternating (e.g., tongue tip and tongue dorsum). For the frequency locking account, the actual movement is relevant. This is also supported by Pouplier (2008), who found fewer intrusions and reductions in alternating CV CV sequences compared to CVC CVC sequences.

The third experimental condition compares monosyllabic word pairs with bisyllabic word pairs, e.g. *tape cape* versus *taper caper*. The aim of this comparison is threefold: first, in the monosyllabic case it is not clear whether more frequent errors in the final consonant compared to initial consonants are an effect of the syllable or the word, i.e. whether syllable and word boundary are confounded. However, due to phonological and lexical restrictions the medial consonant in the bisyllabic words is not strictly a coda consonant but is either in the onset (for *taper caper*) or ambisyllabic (for *picky ticky*). Second, in the bisyllabic case we also varied the number of overlapping phonemes: e.g. in *pick tick* two segments /ɪ/ and /k/ overlap, while in *picky ticky* there is an additional identical segment. This should lead to more competition and therefore more errors. And third, by adding a syllable without changing the metronome rate, the speakers are under increased time pressure which could also lead to more errors.

Methods

Participants
Five female and 4 male native speakers of American English from the New Haven community participated in this experiment. They were between 20 and 30 years of age with a mean of 24.4 years. All participants read and signed an informed consent and were paid for their participation. None of the participants reported any neurological, speech or hearing disorders. This work was approved by the Yale University Institutional Review Board.

Recordings

Acoustic and articulatory movement data were recorded using electromagnetic articulography (Carstens AG500). Small movement-transducing sensors were attached to the speech articulators using dental adhesive. Three sensors were glued on the midsagittal tongue surface, one sensor as far back as the participant would tolerate (hereafter TR), one sensor 1 cm behind the tongue tip (TT) and one in between (TB). For tracking jaw movements, one sensor was attached to the gingiva below the lower front incisors in the midsagittal plane and one placed parasagittally below the left premolar. Two additional sensors were attached to the upper and lower lips at the vermillion border. Four reference sensors were used to correct for head movement: two placed on the left and right mastoid processes, one on the gingiva above the upper incisors and one on the nasion. Articulatory data were sampled at 200 Hz and acoustic data at 16,000 Hz. Movement data were low-pass-filtered at 20 Hz, corrected for head movement, and rotated and translated to the occlusal plane using reference biteplane data.

Speech Material

The speech material consisted of word pairs that were repeated in time with a metronome, one beat per word (condition 1; see below for details). The words had a CVC structure with voiced and voiceless stops as consonants. Voiced stops only occurred in the coda of a limited number of pairs. The word pairs always had the same vowels. They differed, however, in the place of articulation for the stops. In the *onset mismatch* condition, the onset consonants had different places of articulation, e.g. ***t**op **c**op*, while the other segments were identical. In the *coda mismatch* condition, the place of articulation for the coda varied, e.g. *to**p** to**ck***. For the *double mismatch* condition, both the onset and coda varied, but within each CVC the onset consonant was identical with the coda consonant, e.g. ***p**i**p** **k**i**ck***. A list of all word pairs, number of speakers and trials is given in Table 1 for the conditions tested here. Some participants produced the word pairs in two orders (see Table A1 in the Appendix for more details). For each of the words, control trials with simple repetitions of each target word were produced (e.g., *top top*).

As stated above, condition 2 probed whether more errors are elicited if two actual articulators are alternating or, more abstractly, the word frame alternates between filled and empty slots. Therefore, word pairs alternating in mismatch position were compared to word pairs alternating in missing positions, e.g. *top cop* versus *top op* or *top tock* versus *top ta*. Five of the 9 speakers produced the word pairs shown in Table 1 with missing positions. Condition 3 tested position within the word: by comparing error rates for monosyllabic word pairs with bisyllabic word pairs (e.g., *tip tick* vs. *tippy ticky*), the mismatch occurs word- and syllable-finally in the first case and word-medially in the second case. Three of the speakers also produced the bisyllabic word pairs. Due to lexical restrictions the stimuli could not be balanced for place of articulation and vowel combinations. All data for a particular speaker were collected within the same experimental session.

Procedure

Trials were cued with instructions presented on a computer monitor ("Get ready, breathe, GO" sequenced at 1-s intervals) together with the word pair under test. Participants were encouraged to avoid respiration during production because breathing has a phase-resetting effect (Goldstein et al., 2007). Some of the speakers were instructed to pronounce the first word with stress, the others were left free in their placement of stress, though all were consistent in their choice.

At the same time as the GO stimulus was presented, the participants also heard metronome clicks presented via an earpiece. These clicks were initially stable at a rate of 170 clicks/min over the first half of the trial (about 10 s) and then over the second half accelerated at a linear rate to 230 clicks/min under computer control. The reason for the variable rate was to elicit an initial, easy to produce baseline with minimal errors, followed by an increasingly difficult production task in which errors were increasingly likely. Participants were instructed to time the onset of each produced word to a click.

Table 1. Trials with alternating word pairs for the three sets and the conditions onset mismatch, coda mismatch and double mismatch

Set/speaker/trials	Condition	Articulators	C	V	Word pairs
Set 1 9 speakers 278 trials	Onset mismatch	lab/cor	p-t	/æ/	pack tack
			p-t	/ɑ/	pock tock
		dor/lab	k-p	/ɪ/	kit pit
			k-p	/ɑ/	cod pod
		dor/cor	k-t	/ɑ/	cop top
			k-t	/eɪ/	cape tape
	Coda mismatch	lab/cor	p-t	/eɪ/	cape Kate
			p-t	/ɑ/	cop cot
			b-d	/ɑ/	cob cod
		lab/dor	p-k	/eɪ/	tape take
			p-k	/ɑ/	top tock
			p-k	/æ/	tap tack
			p-k	/ɪ/	tip tick
		dor/cor	k-t	/æ/	pack pat
			g-d	/ɑ/	pog pod
7 speakers 177 trials	Double mismatch	lab/cor	p-t	/ɑ/	pop tot
		lab/dor	p-k	/ɪ/	pip kick
Set 2 5 speakers 68 trials	Onset missing	cor	t-0	/ɑ/	top op
		dor	k-0	/ɑ/	cop op
	Onset mismatch	cor/dor	k-t	/ɑ/	cop top
	Coda missing	lab	p-0	/ɑ/	top ta
		dor	p-0	/ɑ/	cop Kaa
	Coda mismatch	lab/cor	p-t	/ɑ/	cop cot
		lab/dor	p-k	/ɑ/	top tock
Set 3 3 speakers 43 trials	Onset mismatch, 2 syllables	cor/dor	t-k	/eɪ/	taper caper
		lab/cor	p-t	/ɪ/	picky ticky
	Onset mismatch, 1 syllable	cor/dor	t-k	/eɪ/	tape cape
		lab/cor	p-t	/ɪ/	pick tick
		lab/dor	p-k	/ɪ/	pit kit
	Coda mismatch, 2 syllables	lab/cor	p-t	/eɪ/	caper cater
		lab/dor	p-k	/eɪ/	taper taker
		dor/cor	k-t	/ɪ/	picky pity
		dor/lab	k-p	/ɪ/	ticky tippy
	Coda mismatch, 1 syllable	lab/cor	p-t	/eɪ/	cape Kate
		lab/dor	p-k	/eɪ/	tape take
		dor/lab	k-p	/ɪ/	tick tip

Most of the subjects also spoke the trials with word pairs in reversed order. lab, labial; cor, coronal; dor, dorsal.

_____

Mooshammer/Tiede/Shattuck-Hufnagel/
Goldstein

Measures

In this study two complementary measures were applied to quantify the effect of mismatch type and position within the word on articulatory behavior in word repetitions. The first method, *error rates*, uses normative distributions to establish thresholds for identifying deviant movement amplitudes (see Pouplier, 2008). The second method, *delta*, quantifies spatial variability across conditions by calculating Euclidean distances between mean positions and individual tokens, and is based on McMillan and Corley (2010).

Error Rate Measure

The *error rate* measure relies on establishing a threshold, computed separately for each trial, for identifying reductive or intrusive behavior of an articulator based on its non-errorful behavioral range. This is determined by first labeling the maximal constrictions for all consonants. For dorsal and apical stops this labeling used the vertical component of the TR and TT sensors, respectively. For bilabial stops the lip aperture signal was labeled, calculated as the Euclidean distance between the sensors on the upper and the lower lips. (In one instance the sensor on the upper lip failed during the experiment and so the vertical lower lip trajectory was used instead.) Figure 1 shows this labeling for tongue tip maxima during the /d/ in *cod* in an utterance of the word sequence *cod cob*.

These time points were used then for identifying the amplitudes of the alternating articulator at the unconstrained point of its cycle, i.e. at the point where the controlled articulator is achieving its target constriction: e.g., the lip aperture during /d/ for the word pair *cod cob* in Figure 1. In order to obtain more stable results, the samples within a 9-sample window (45 ms) around the measurement point were averaged for both the controlled and uncontrolled articulators, shown as gray vertical lines in Figure 1. These values were then used to calculate mean amplitudes characterizing normative behavior for controlled and unconstrained positions within that trial. To desensitize these averages against errors and outliers only values within the inner quartiles (25:75%) contributed to each mean. The error rate threshold was then determined by splitting the difference between these means (the "split-mean" criterion; Pouplier, 2008). The resulting values, shown as black separation lines in the lower panel of Figure 1, were used as thresholds to define several error types:

1   *Reductions* are defined as intended gestures below the threshold, e.g. tongue tip positions during /d/ that fail to rise to the expected positional range for that gesture. They are shown as small squares for reduced tongue tip positions in the upper left quadrant of the lower panel in Figure 1

2   *Intrusions* are defined as instances of the unconstrained articulator rising above the threshold, e.g. tongue tip positions during /b/ that are within range of an intended constriction, shown as small stars in the upper and lower panel of Figure 1

3   *Substitutions* occur when there is a full intrusion of the unconstrained articulator and at the same time a full reduction of the intended articulator; this type of error is closest to the phonemic errors that have been investigated in most corpus studies

Based on these definitions the error rates were calculated as percentages of errors per trial for each error type, normalized by the number of words per trial.

This procedure was applied for counting error types for the monosyllabic and the bisyllabic single mismatch condition. It did not deliver reasonable results for the double mismatch condition, e.g. word pairs such as *pop tot*, because of the overlap between the constriction gesture across word boundaries. The problem arises because only two articulators alternate and therefore the unconstrained articulator during one consonant is already in position for the following consonantal constriction; e.g. during the coda /p/ in *pop* the tongue tip is already in place for the onset of *tot*, so that an errorful /t/-gesture could not be distinguished from early timed intended gesture for the following /t/. Therefore, intrusion and substitution rates will be reported for the double mismatch condition. Reductions, however, do not depend on the timing between coda and the following onset consonants. Therefore, reduction rates of the double mismatch condition will be compared to the single mismatch condition in the Results section.
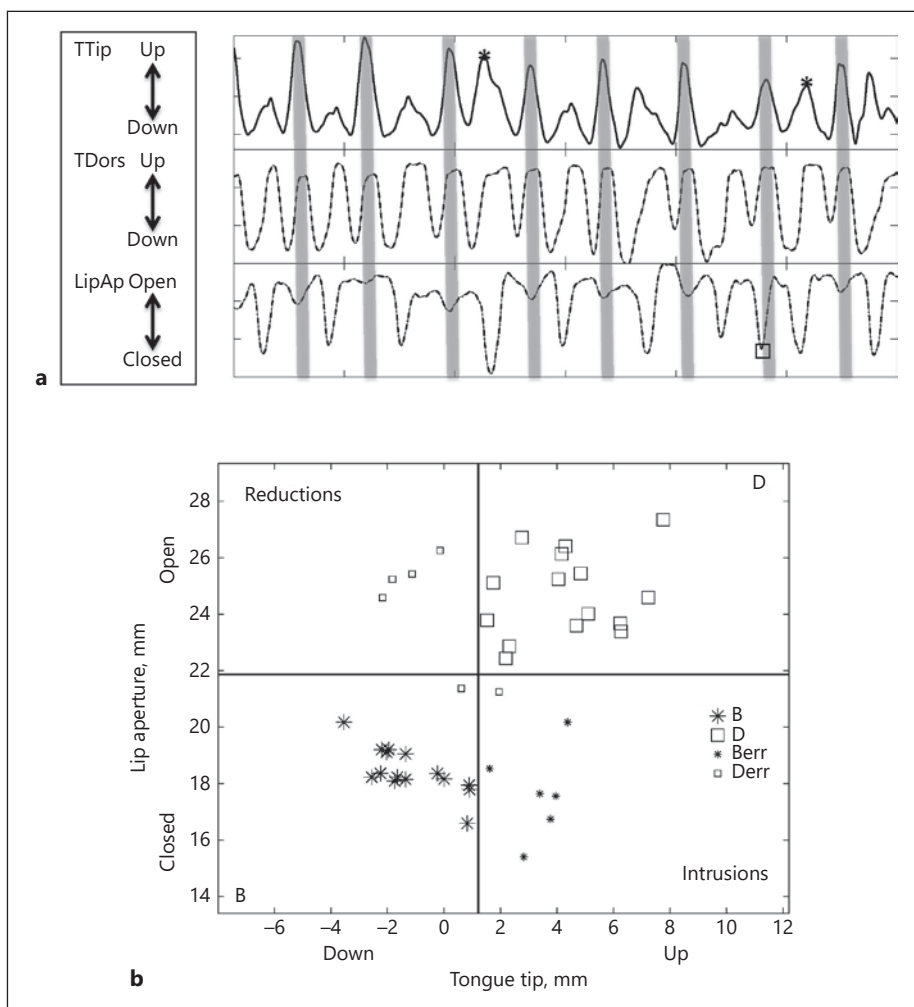
Fig. 1. Procedure for identifying errors, exemplified for repetitions of *cod cob*. **a** Movements of the vertical tongue tip (TTip; continuous line, upper panel), the vertical tongue dorsum (TDors; dashed line, middle panel) and lip aperture (LipAp; dashed dotted line, lower panel); gray vertical lines indicate measurement points for maximal tongue tip constrictions during /d/. The stars denote intrusions of the tongue tip and the square an intrusion of the lower lip during a constriction for /d/. **b** Scatterplot for vertical tongue tip (*x* axis) and lip aperture measures (*y* axis) during /d/ (D; denoted as squares in the upper two quadrants) and /b/ (B; denoted as stars in the lower two quadrants).

For the missing condition the procedure had to be adjusted because for the missing positions no maximal constriction could be labeled. In these cases, the smaller "bumps" of the unconstrained articulator were labeled, e.g. for *top op* a small maximum in the tongue tip excursion could be observed at the onset of *op* and was labeled as the unconstrained movement.

Mooshammer/Tiede/Shattuck-Hufnagel/
Goldstein

The thresholding method described above does not work correctly if the positions of the intended gesture and the unconstrained movement are too similar. Therefore, if more than 50% of the items were produced with either a reduction or an intrusion error these trials were removed from further analysis. Out of 280 alternating trials 2 were excluded. Both were *pat pack* alternations from the coda mismatch condition with large numbers of reductions of the alveolar stop.

Delta Value Measure

The *delta* measure, adapted from McMillan and Corley (2010), quantifies spatial variability of all articulators, including nontarget articulators, during some point of maximum constriction. First, the horizontal and vertical positions of all sensors are measured at the maximum constrictions of intended gestures within a trial, e.g. the TRy maxima for /k/ during repetitions of *pod cod*. Delta measures are computed as the Euclidean distance between sensor position at each of these target instances and the mean across all instances within that trial, resulting in a single delta value for each target (e.g., /k/ during *pod cod*). This quantifies the spatial deviation of a given measured instance from the reference configuration. Delta values were calculated for the alternating trials (e.g., *pod cod*) and for corresponding nonalternating control trials (e.g., *cod cod*). McMillan and Corley (2010) reported that delta values are larger for voice onset times and electropalatography contact patterns for trials with alternating words than for nonalternating controls. They attribute the larger delta values in alternating trials to coactivation of phonemes cascading to the phonetic level. Therefore, the difference between delta values of the alternating trials and the mean of the delta values of the matched control trials are used here; this quantifies increased variability through comparisons between the Euclidean distances of alternating and control trials.

For the *missing condition* in which empty onsets and codas are compared to filled ones, the maxima of the active consonantal gestures were used as time points for extracting the sensor positions. For the empty onsets and codas this was obviously not possible. Therefore, the maximum displacement of the articulator that is alternating was used. For example, in a sequence such as *top op* the first time point corresponds to the maximum constriction for the alveolar stop in *top*. The second maximum in the tongue tip movement occurs during or after the lip closure for the coda /p/ in *top*. These are generally smaller maxima that could not always be detected. In these cases nothing was measured. These two time points per word pair were used to strobe the articulatory positions for calculating the delta values.

Both the error rate and delta methods have certain advantages and disadvantages but provide complementary insight. The error rate method has the disadvantage of using a distributionally based but physiologically arbitrary threshold (see McMillan and Corley, 2010) to distinguish between errorful and "normal" variability. This is not the case for the delta method, which just quantifies variability independently of the cause. However, this is also the disadvantage of the delta method because it cannot distinguish between kinds or distinct causes of variability. As mentioned above, coronal consonants in particular tend to show reduction processes in syllable-final position (see e.g. Byrd, 1996a, b) which also leads to more variability compared to the syllable onset position. Since the thresholding method distinguishes between reduction and intrusion, these different types can be related to their causes. Another disadvantage of the delta method is that it quantifies the spatial variability of all included articulators, i.e. it does not distinguish between active and passive articulators or alternating and nonalternating articulators. As was found by Slis and van Lieshout (2016) intrusions in the onset are more frequent for intruding dorsal gestures compared to lower lip gestures, but the delta method is not sensitive to the articulator involved (see Slis, 2018, for a discussion of error measures).

Analysis

The hypothesis that spatial variability is larger in the coda than in the onset was tested using linear mixed effects models (see e.g. Baayen, 2008; Pinheiro and Bates, 2000) with

Table 2. Results of the linear mixed effects model for the incorrect-correct distributions for substitution, reduction and intrusion error types with the fixed effects position, articulator of intended gesture and the interaction

Error type	Condition	β	SE(β)	z value	p	Effect
Substitution	Intercept	−5.77	0.43			
	Pos	1.30	0.45	2.88	**	OD < CD
	Art	−0.05	0.18	−0.27		
	Pos × Art	−0.01	0.39	−0.01		
Reduction	Intercept	−4.30	0.23			
	Mismatch	0.80	0.40	2.00	*	single < double
	Pos	2.64	0.28	9.27	***	OD < CD
	Art	−0.18	0.30	−0.62		
	Mismatch × Pos	0.24	0.48	0.49		
	Mismatch × Art	−0.04	0.26	−0.14		CD lab, cor: single < double
	Pos × Art	−0.80	0.24	−3.36	***	CD: lab < dor < cor
	Mismatch × Pos × Art	−0.38	0.52	−0.73		
Intrusion	Intercept	−2.86	0.18			
	Pos	0.87	0.25	3.46	***	OD < CD
	Art	−0.06	0.06	−1.12		
	Pos × Art	0.36	0.12	3.15	**	lab, dor: OD < CD OD: lab, cor < dor CD: cor, dor < lab

The slope β corresponds to the increase from onset mismatch (OD) to coda mismatch (CD) for the factor position (Pos: onset OD vs. coda CD). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Art, articulator, labial (lab) vs. coronal (cor) vs. dorsal (dor). For the double mismatch condition, only reductions are included (see text for explanation).

delta values as the dependent variables, with random intercepts by speaker and item. To test the significance of structure on error rate, the proportion of errors to correct instances was used as dependent variable for logistic general mixed effects models separately for the error types intrusion, substitution and reduction. In order to avoid collinearity between factors and factor levels, the factors were coded and centered by subtracting the grand mean, following suggestions of Gelman and Hill (2007). All statistics were carried out using R 3.3.0 (see R Core Team, 2016) with the packages lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2016). As a preliminary, log-likelihood comparisons were used to assess whether the model fit improved by including random slopes by speaker. For significant interactions the data set was split accordingly. Statistical significances of the fixed effects are presented by the estimates of the regression coefficients of the model and their associated standard errors, with probabilities of their quotient (a t test) based on the Satterthwaite approximation for denominator degrees of freedom.

Because the distributions of the delta values were skewed by outliers, items with residuals exceeding 2.5 times the standard deviation were excluded (following Baayen, 2008). Based on this criterion 819 out of 28,929 delta values were excluded; this was unnecessary for tabulating error rates because the two trials with more than 50% errors were already excluded (see above, Error Rates). Test results on intercepts show whether the grand mean is significantly different from 0. This information is only of significance for the
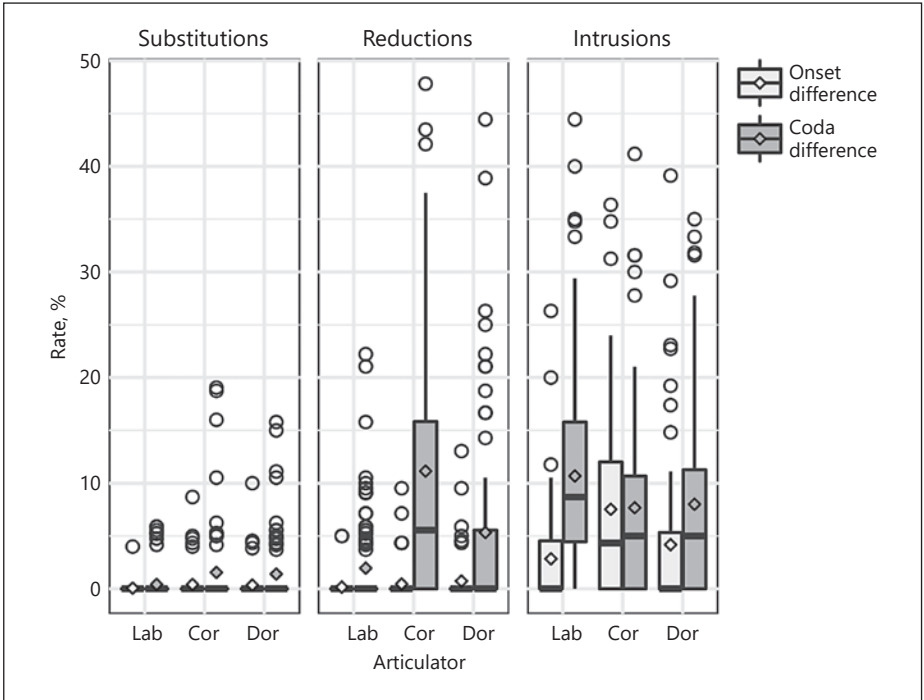
Mooshammer/Tiede/Shattuck-Hufnagel/
Goldstein

Fig. 2. Boxplots for rates for substitutions (left), reductions (middle) and intrusions (right), shown for trials with alternating onsets (light gray) and alternating codas (dark gray) and articulators during the consonants. Medians are indicated as black lines and means as diamonds. Lab, labial; Cor, coronal; Dor, dorsal.

difference between delta values in alternating and nonalternating trials. Delta differences significantly larger than zero denote more spatial variability in alternating trials than nonalternating trials which is predicted by Goldrick's cascading activation model (see Goldrick and Blumstein, 2006).

Results

Onset versus Coda Mismatch (Condition 1)

This condition tests whether the error rate depends on the position of mismatch, i.e. onset mismatch (e.g., **to**p **c**op) versus coda mismatch (e.g., to**p** to**ck**) or double mismatch (e.g., *pop tot*).

Error Rates

Table 2 and Figure 2 show the difference in error rates between onset mismatch and coda mismatch for the three error types (substitution, reduction and intrusion) and the articulators of the intended consonants. Since intrusion and substitution rates cannot be calculated for double mismatch, only alternating

word pairs with single mismatch are presented here. As was also shown by Pouplier (2003) and Goldstein et al. (2007), intrusions are the most frequent error type with 7% compared to reductions (3.3%) and substitutions (0.7%). All error types are significantly more frequent for coda mismatch than for onset mismatch. For intrusions the position effect is not significant for coronal consonants. In the onset intrusions are more frequent for dorsal consonants than for labial or coronal consonants. In the coda, intrusion rates are largest for labial consonants.

Reductions in the single mismatch condition show the largest increase from onset mismatch (0.5%) to coda mismatch (5.4%) with final /t/ being most frequently reduced in the coda (reductions for voiceless stop coda: /t/ 14.1%, /k/ 4.7%, /p/ 1.9%) (significant in the coda, see Table 2). Including speaker-specific slopes for articulator improved the model significantly for reduction rates. As was pointed out in the section "Error Rate Measure" above, reduction rates could also be calculated for the double mismatch condition (see Fig. 3 and Table 2). The reduction rates are very small in the onset: the mismatch condition increases reduction rates only slightly with a mean of 0.5% for single mismatch to 0.8% for double mismatch. The mismatch effect for the coda is considerably larger with 5.8% for single mismatch to 14.5% for double mismatch ($z = 2.1$, $p < 0.05$). There was a significant interaction with articulator in the coda position with an increase in reductions for labial and coronal consonants but not for dorsal ones.

Figure 4 shows boxplots for pooled intrusion and substitution errors in onset and coda positions, comparing single speakers. The increase in error rates for coda mismatch is generally consistent across speakers even though speakers differ considerably in this difference and in their absolute error rates; compare e.g. speakers F2 and F3.

Delta Values

Figure 5 shows the delta value (spatial variability changing from non-alternating controls to alternating trials) for single (Fig. 5a) and double mismatch (Fig. 5b) for onsets and codas grouped by articulators. Whether mismatch, position and articulator of the intended gesture had significant effects was tested by computing a number of linear mixed effects models (results are shown in Table 3). Two speakers, F2 and M4, were excluded from this analysis because they did not produce all of the control nonalternating trials.

Model comparison for the overall model with position, mismatch and articulator as fixed factors showed that including speaker-specific slopes for articulator of the intended gesture improved the model significantly. Alternating trials are more variable than controls shown by the significant effect for the intercept. The significant main effects for position and mismatch suggest that variability is larger in the coda than in the onset and larger for double mismatch than for single mismatch. The intended articulator did not show a significant main effect but there were significant interactions with position and mismatch. Therefore, the data were split for the mismatch condition. The results for single mismatch are shown in the second part of Table 3. Including the consonantal articulator as a random slope improved the model significantly ($p < 0.05$). Again, the intercept is significant, indicating larger variability for alternating than for
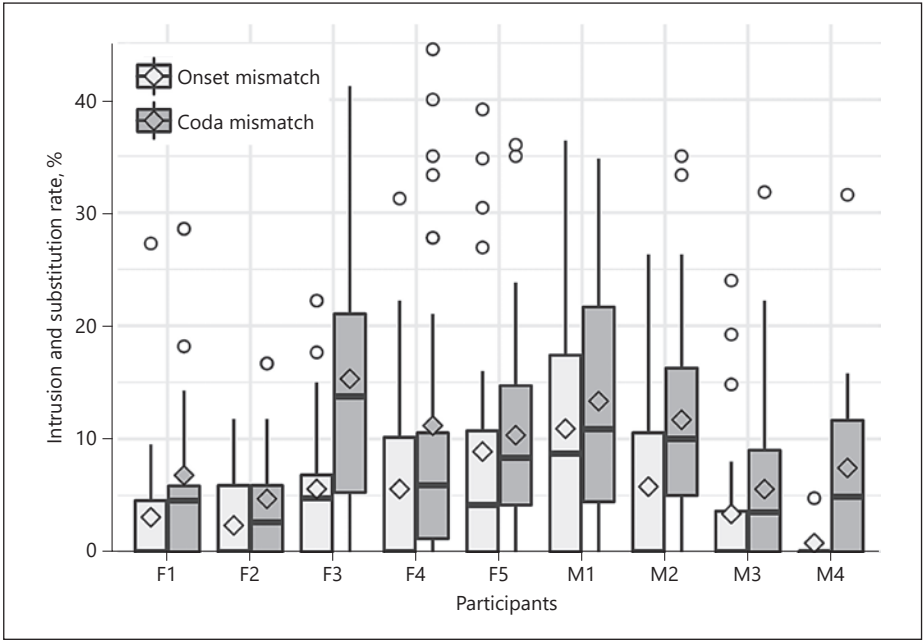
Fig. 3. Boxplots for pooled intrusion and substitution rates for onset mismatch (light gray) and coda mismatch (dark gray) per speaker.
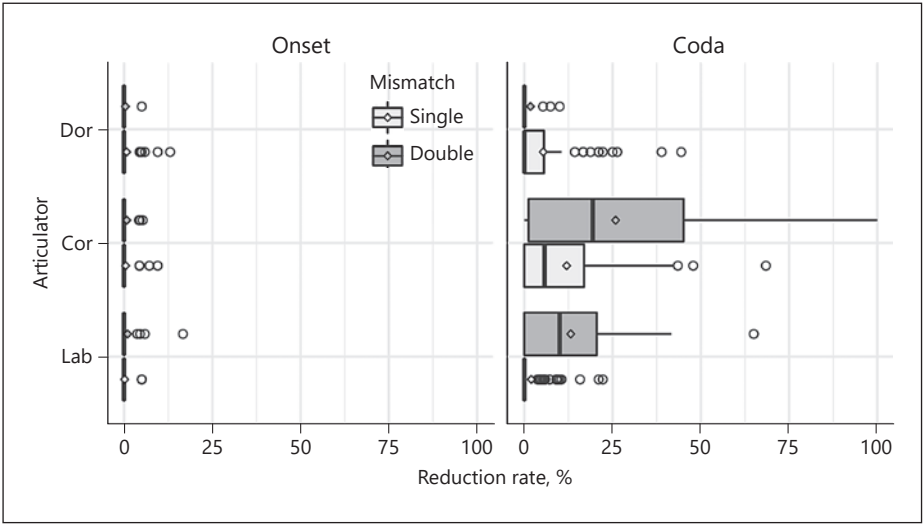

Fig. 4. Boxplots for reduction rates at the onset position (left) and the coda position (right), shown for trials with single mismatch (light gray) and double mismatch (dark gray) and articulators during the consonants (*y* axis). Medians are indicated as black lines and means as diamonds. Lab, labial; Cor, coronal; Dor, dorsal.
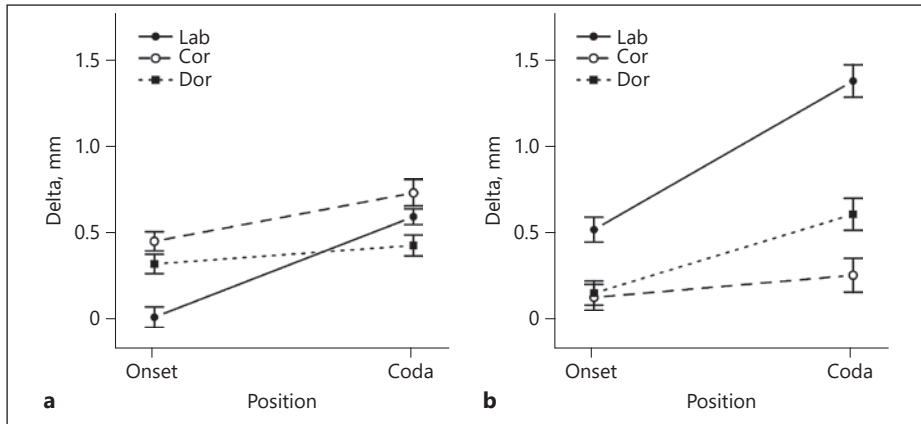
Fig. 5. Means and standard errors for delta values (changes from controls to alternating trials), comparing mismatch in the onset with mismatch in the coda for single mismatch (**a**) and double mismatch (**b**) for articulators of the intended gestures. Lab, labial; Cor, coronal; Dor, dorsal.

Table 3. Results of the linear mixed effects model for delta values with the fixed effects position, mismatch (single vs. double), articulator of intended gesture and the interactions

Subset	Condition	β	SE(β)	df	t value	p	Effect
All	Intercept	0.50	0.08	6.0	6.1	***	
	Pos	0.37	0.04	6,633.0	9.1	***	OD < CD
	Mismatch	0.13	0.04	6,635.0	2.9	**	single < double
	Art	−0.09	0.06	6.0	−0.15		
	Pos × mismatch	0.22	0.09	6,631.0	2.6	**	
	Pos × Art	−0.22	0.05	6,566.0	−4.5	***	
	Mismatch × Art	−0.34	0.05	6,576.0	−6.4	***	
	Pos × mismatch × Art	−0.07	0.10	6,625.0	−0.7		
Single	Intercept	0.47	0.11	6.0	4.2	**	
	Position	0.29	0.05	4,553.0	6.0	***	OD < CD
	Art	0.04	0.05	4,648.0	1.5		
	Pos × Art	−0.18	0.06	3,212.0	−3.1	**	lab, cor: OD < CD
Double	Intercept	0.52	0.09	6.2	5.6	**	
	Pos	0.50	0.07	2,065.8	6.7	***	OD < CD
	Art	−0.34	0.20	5.9	−1.7		
	Pos × Art	−0.26	0.09	2,066.2	−2.9	**	lab, dor: OD < CD

In the upper part (entitled "All") all factors are tested, in the lower parts the data were subset. OD, onset mismatch; CD, coda mismatch; Pos, position (onset OD vs. coda CD); Art, articulator, labial (lab) vs. coronal (cor) vs. dorsal (dor). ** $p < 0.01$, *** $p < 0.001$. Statistics are based on 7 speakers excluding speakers F2 and M4. 191 out of 6,949 values were excluded as outliers based on the criterion described in the section "Statistics."
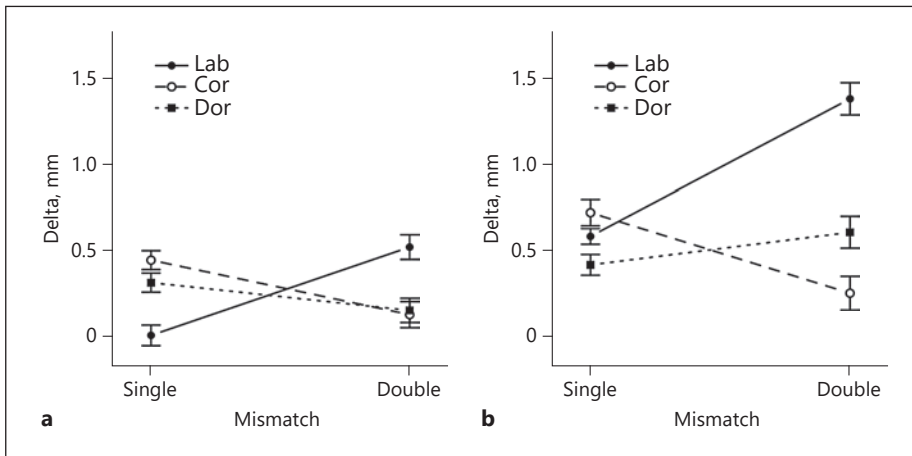
Fig. 6. Means and standard errors for delta values (changes from controls to alternating trials), comparing single and double mismatch in the onset (**a**) and the coda (**b**) for different articulators of the intended gesture. Lab, labial; Cor, coronal; Dor, dorsal.

nonalternating trials. Delta values are significantly larger in the coda than in the onset. This result is in agreement with the error analysis above, especially with the intrusion rates. The main effect of articulator for the single mismatch condition does not reach significance but the interaction with position is significant because for dorsal stops variability did not increase significantly from onset to coda.

For the double mismatch condition shown in the third part of Table 3 the model improved significantly by including speaker-specific slopes for articulator of the intended gesture. Spatial variability increased in the coda compared to the onset. The significant interaction between position and articulator comes about because only labial and dorsal consonants show an increase in the coda.

For testing whether single and double mismatch had different effects for different positions, the data were split for onset and coda (Fig. 6). In the onset there was a significant interaction with articulator ($t = -4.32$, $p < 0.01$) because only for labial consonants did double mismatch show an increase in variability ($t = 4.6$, $p < 0.001$). For coronal consonants the effect was in the opposite direction ($t = -3.33$, $p < 0.01$), and for dorsal consonants there was no significant change. In the coda position effects are similar to the onset position for the different articulators.

In summary, increased variability could be found for alternating versus control stimuli and in codas as compared to onsets. This was independent of the articulator for the single mismatch condition and only significant for labial and dorsal consonants in the double mismatch condition. The increase in variability for coda versus onset was less consistent in the double mismatch condition for different articulators (see Fig. 5b, coronal consonants).
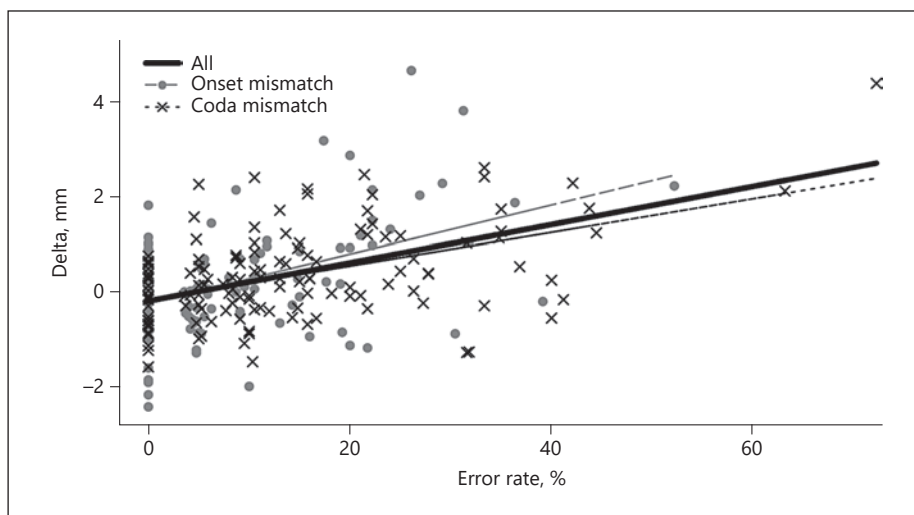
Fig. 7. Scatterplot of error rates (intrusion, substitutions and reductions) and the medians of delta values per trial. Data points for onset mismatch are plotted as gray circles, coda mismatch as black crosses. Superimposed are regression lines for all data (thick black), onset mismatch (thin dashed gray line) and coda mismatch (thin dotted black line).

Relationship between Error Rates and Delta Values

Qualitatively the delta values and the error rates lead to similar results, namely that generally more errors occur in the coda mismatch condition. In order to investigate whether this is also a statistically significant relationship, regression models were calculated for error rates and delta values. Because reduction rates also contribute to spatial variability, error rates were calculated as the sum of intrusion, substitution and reduction rates. Figure 7 shows averaged delta values plotted against error rates per trial. The slope of this relationship is highly significant ($F(1, 260) = 78.12$, $p < 0.001$, adjusted $R^2 = 0.23$). There was no significant effect of the mismatch position or the interaction between mismatch position and error rate as shown by the almost completely overlying regression lines in Figure 7. The Pearson product-moment correlation of 0.48 between the two measures was also highly significant ($t(1, 260) = 8.84$, $p < 0.001$). Separate correlation coefficients for reduction and intrusion error rates with the delta value were also significant, with a slightly larger coefficient for intrusions ($R = 0.41$, $t(1, 260) = 7.34$, $p < 0.001$) than for reductions ($R = 0.28$, $t(1, 260) = 4.77$, $p < 0.001$). Therefore, both types of errors are related to spatial variability.

However, since the relationship between error rates and delta values explains only 23% of the variance, we investigated in further detail whether a pattern could be found for individual test pairs. Figure 8 shows delta values and error rates for words and condition, averaged across speakers. Only words that were used for either the onset mismatch or the coda mismatch condition are considered for this figure. Data points from onset mismatch trials are printed as gray circles and data points from coda mismatch as black triangles. For example, values for the
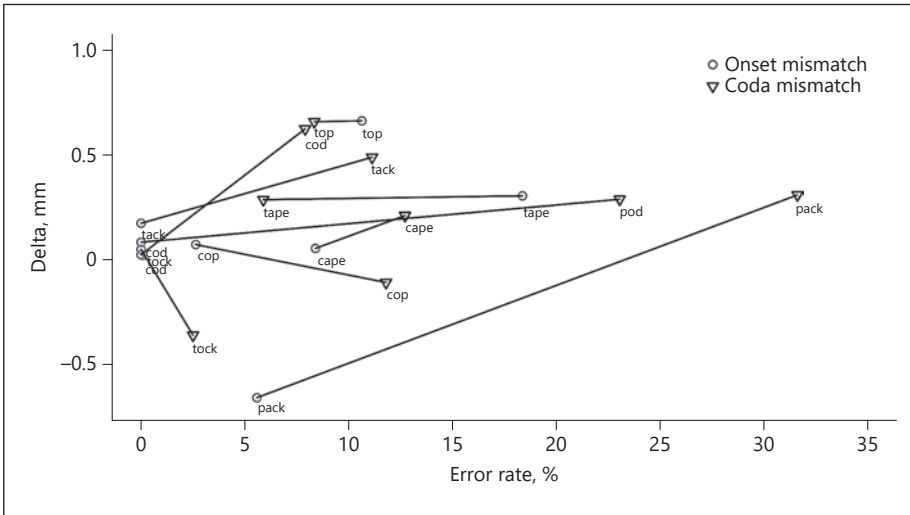
Mooshammer/Tiede/Shattuck-Hufnagel/
Goldstein

Fig. 8. Scatterplot of error rates (intrusion, substitution and reductions) and the delta measure averaged across speakers and reversals. Gray circles correspond to onset mismatch and black triangles to coda mismatch. The connecting lines are drawn for a better visualization.

word *cod* printed in gray with a circle are taken from trials with alternating onsets (*pod cod* and *cod pod*); values for the same word printed in black with a triangle were extracted from trials with alternating codas (*cod cob* and *cob cod*). The lines connecting values for the two conditions per word are plotted for better visualization. As can be seen almost all lines have a positive slope, indicating that items with larger error rates within a pair also have larger spatial variability. The only exceptions are the words *tock* and *cape* which show slightly lower delta values for larger error rates. Out of the remaining 7 word pairs, 5 go in the expected direction with larger error rates and delta values for coda mismatch as compared to onset mismatch. The other 2 (*top* and *tape*) go in the opposite direction with smaller values for coda mismatch as for onset mismatch. The fact that 5 of 9 words show a pattern consistent with the hypothesis that more errors are associated with larger spatial variability suggests that the two phenomena are related.

Missing Onsets and Codas (Condition 2)

The aim of this experimental condition is to test whether intrusion errors and spatial variability are triggered by the alternation of actual gestures or rather in a more abstract way by alternations of the syllable structure. Therefore, in the onset missing condition, pairs of CVC words with alternating onsets will be compared to CVC VC word pairs in which an onset consonant is alternating with an empty onset, e.g. *top cop* versus *top op*. In the coda missing condition, CVC word pairs with alternating codas will be compared to CVC CV word pairs in which the coda consonant is alternating with an empty coda, e.g. *top tock* versus *top ta*; 5 speakers produced stimuli from set 2 (Table 1).

Table 4. Results of generalized linear mixed effects models for the incorrect-correct distributions of reduction and intrusion + substitution error types with the fixed effects position, structure and the interactions for 5 speakers

Error type	Subset	Condition	β	SE(β)	t value	p	Effect
Reduction	Onset	Intercept	−4.14	0.39			
		Pos	1.73	0.61	2.85	**	OD < CD
		Struc	−1.91	0.50	−3.80	***	missing < mismatch
		Pos × Struc	−1.17	1.10	1.06		
Intr + Subs		Intercept	−2.72	0.18			
		Pos	−0.12	0.21	−0.59		
		Struc	−0.80	0.20	−4.09	***	missing < mismatch
		Pos × Struc	0.08	0.39	0.20		

Intr, intrusion; Subs, substitution; Pos, position, onset mismatch (OD) vs. coda mismatch (CD); Struc, structure, missing vs. mismatch. ** $p < 0.01$, *** $p < 0.001$.

Error Rates

Since in the missing condition substitutions do not involve a reduction of the other articulator (because it is missing), substitution and intrusion rates were pooled together for this analysis. The denominator for calculating the percentage of errors corresponded to the number of occurrences of onsets and codas, i.e. about half for the missing condition compared to the mismatch condition. For the reduction rates position and word structure showed significant main effects with more reduction in the coda compared to the onset and in the mismatch condition compared to the missing condition (Table 4; Fig. 9). The interaction was not significant. For the intrusions and substitutions, a significant effect was found for word structure (mismatch vs. missing) with less frequent errors in the missing condition. No significant effect was found for position for this subset of the data. This is contrary to the results from condition 1, in which a clear position effect was observed with larger intrusion rates in the coda. However, as can be seen in Figure 8 the word pair *top cop* that was used here as control for the missing condition had error rates as large as the word pair *top tock.* Therefore, we assume that this can be attributed to this particular choice of word pairs. One speaker (F3) could not produce open syllables and has very high error rates for coda-missing stimuli. Without her there is a general tendency for lower error rates in the missing condition compared to the mismatch condition. However, excluding her did not yield a significant effect for the position effect.

Delta Values

As explained in the Methods section, the delta values for the missing positions had to be calculated by using the maximum of the unconstrained articulator as the time point for measuring the delta value in the case of empty onsets
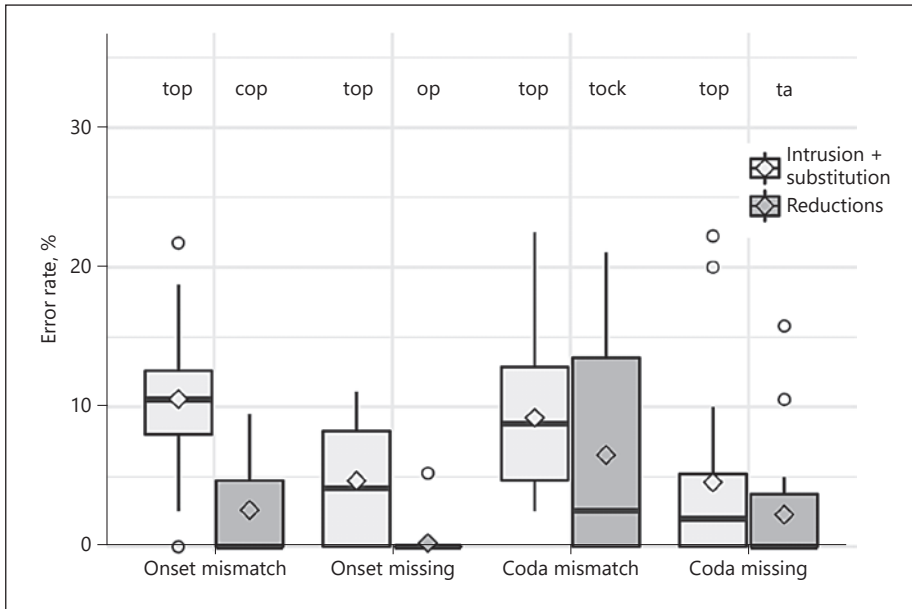
Fig. 9. Boxplots for rates of intrusions and substitutions (light gray) and reductions (dark gray). Medians are indicated as black lines and means as diamonds shown for trials with alternating onsets, missing onsets, alternating codas and missing codas.

or codas. 150 out of 5,338 values were excluded as outliers prior to calculating the difference between alternating and matched control trials. Figure 9 shows means and standard errors for mismatch versus missing and for onset versus coda. The results of a linear mixed effects model are presented in Table 5. Structure (mismatch vs. missing) reached significance, with more spatial variability in the mismatch condition than in the missing condition. Neither position nor the interaction was significant. As can be seen from Figure 10, delta values for the missing condition are negative on average; that means that the controls (e.g., *cop cop*) are produced with greater spatial variability than that of the alternating trials. It has to be kept in mind that only delta values for the CVCs of the missing alternating trials (e.g. **c**op in *cop op* alternations) could be calculated because for the missing consonant condition appropriate controls (e.g., *ta ta*, *op op*) no time point of the missing consonant could be detected. Therefore, the intercept of the delta values is not significant, meaning that variation for alternating trial is similar to the control trials.

*Monosyllabic versus Bisyllabic Word Pairs (Condition 3)*

Condition 3 compares monosyllabic word pairs with bisyllabic word pairs. The aim of this comparison was twofold: first, in the monosyllabic case it is not clear whether more frequent errors in the final consonant compared to initial consonants are an effect of the syllable or the word, i.e. syllable and word bound-

Table 5. Results of the linear mixed effects model for the delta values (difference between alternating trial and the mean of the corresponding control trial) based on 5 speakers

Condition	β	SE(β)	df	t value	p	Effect
Intercept	0.23	0.20	3.98	1.13		
Position	−0.06	0.31	3.98	−0.19		
Structure	−0.55	0.18	3.94	−2.98	*	missing < mismatch
Position × structure	0.15	0.49	3.89	0.30		

Including structure and position as random subject slopes improved the model significantly ($\chi(11, 15) = 22.3$, $p < 0.001$). * $p < 0.05$. 150 out of 5,338 values were excluded as outliers prior to calculating the difference between alternating and matched control trials.

Fig. 10. Means and standard errors for delta values, comparing word position (onset vs. coda) and mismatch with the missing condition with examples for word pairs (top).
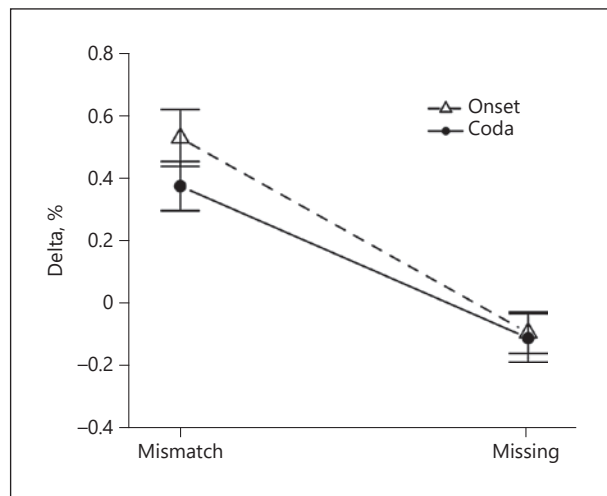
ary are confounded. Second, in condition 2 we found fewer errors for trials with alternation of syllable/word structure, but this could be due to a smaller number of adjacent consonant sequences across word boundaries. The bisyllabic words had a trochaic stress pattern, and the second syllable was open (*pity*) or ended on a rhoticized vowel (*caper*) (see Table 1 for further details). The medial consonant was ambisyllabic, and the word boundary always consisted of V#C sequences. Three speakers produced this subcorpus.

Error Rates
Figure 10 and Table 6 present the results for reduction, intrusion and substitution rates for the three speakers who produced the bisyllabic word pairs. Intrusion rates are significantly larger for monosyllabic than for bisyllabic

Mooshammer/Tiede/Shattuck-Hufnagel/
Goldstein

Table 6. Results of the generalized linear mixed effects models for the incorrect-correct distributions of reduction, intrusion and substitution rates and the fixed factors position and number and the interactions based on 3 speakers

Error type	Condition	β	SE(β)	z value	p	Effect
Reduction	Intercept	−4.17	0.23			
	Pos	1.30	0.50	2.63	**	onset < coda
	Num	−1.16	0.46	−2.53	*	2 < 1
	Pos × Num	−0.74	0.98	−0.76		
Intrusion	Intercept	−3.19	0.26			
	Pos	0.24	0.27	0.88		
	Num	−0.97	0.26	−3.76	***	2 < 1
	Pos × Num	−1.28	3.29	−0.38		
Substitution	Intercept	−5.33	0.36			
	Pos	0.47	0.75	0.63		
	Num	0.11	0.73	0.15		
	Pos × Num	1.54	1.50	1.02		

Pos, position (onset vs. coda); Num, number (1 vs. 2 syllables). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

words. Confirming condition 1 there tended to be more intrusion errors in the coda than in the onset for monosyllabic words, but this difference did not reach significance. Neither did the interaction between position and number of syllables. For the reduction rates the main effects of syllable number and position were significant with higher reduction rates in the coda and in monosyllabic words. Substitution rates were again very small and did not differ for either position or syllable number.

Delta Values

Since control trials for the *pick picky* alternations were not recorded, only trials from the *cape caper* subset are taken into consideration here. 59 out of 1,807 values were excluded as outliers prior to calculating the difference between alternating and control trials. Figure 11 shows the mean and standard errors for the delta values for onset versus coda position and Figure 12 those for monosyllabic versus bisyllabic words. Table 7 shows that there is a significant main effect of syllable number with larger variability in monosyllabic words, but neither position and nor the interaction were significant.

Summary of Results

In this section the most important results regarding error rates and spatial variability are summarized:
- For CVC word pairs, mismatch in the coda induces larger spatial variability and more substitutions, reductions and intrusions than mismatch in the onset
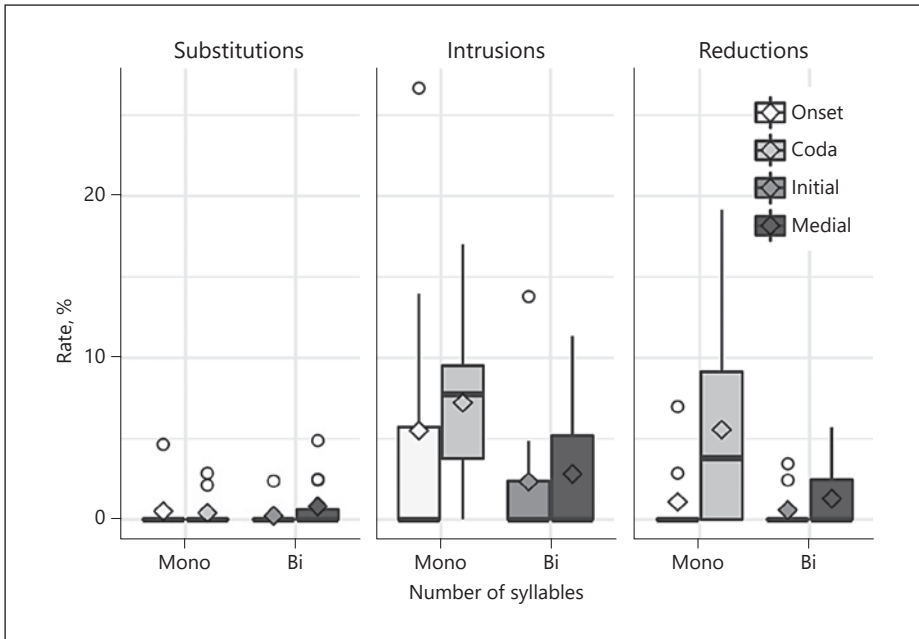
Fig. 11. Means and standard errors for rates for substitutions, reductions and intrusions, shown for trials with monosyllabic (mono) and bisyllabic (bi) words.
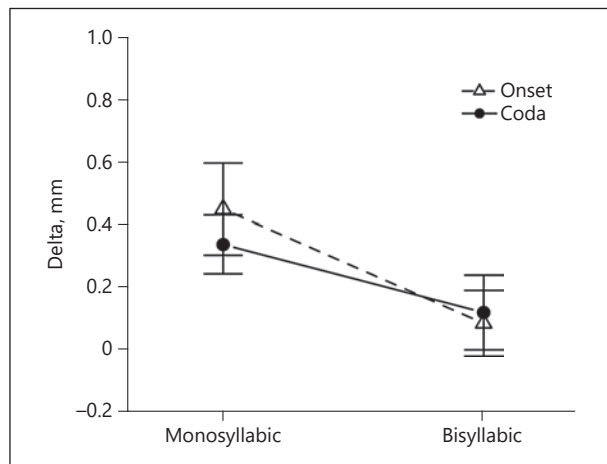

Fig. 12. Means and standard errors for delta values, comparing word position (onset vs. coda) and number of syllables (monosyllablic vs. bisyllabic).

- Substitutions occurred only rarely and reductions mainly in the coda; intrusions were the most frequent error type
- Double mismatch induced larger variability and more reduction errors compared to single mismatch. This increase was larger in the coda than in the onset

Table 7. Results of the linear mixed effects model for the delta values for the fixed effects position and number (1 vs. 2 syllables) based on 3 speakers

Condition	β	SE(β)	df	t value	p	Effect
Intercept	0.27	0.14	1.9	2.08		
Pos	−0.08	0.12	560.4	−0.68		
Number	−0.31	0.12	561.3	−2.56	*	2 < 1
Pos × number	0.18	0.24	561.0	0.74		

The slope β corresponds to the increase from onset mismatch to coda mismatch. Pos, position (onset vs. coda). * $p < 0.05$. 59 out of 1,807 values were excluded as outliers prior to calculating the difference between alternating and control trials.

- Error rates and spatial variability were smaller for the missing condition (e.g., *top ta*) than for the mismatch condition (e.g., *top tock*)
- For monosyllabic word pairs more errors and larger spatial variability were measured than for bisyllabic word pairs

Discussion

The aim of this study was to determine whether errors occur more frequently in the onset or in the coda of a word. Two measure types were used to quantify errors. The error rate measure is distributionally based and provides the frequency of three different error types: intrusions, reductions and substitutions. The second measure, the delta value, quantifies the difference in spatial variability during the maximal constriction for alternating and nonalternating trials. In general, both approaches provide evidence for errors occurring more frequently for coda mismatch (e.g., *top tock)* than for onset mismatch (e.g., *top cop*). Results on spatial variability were in the same direction. We found significantly higher intrusion and substitution rates and a larger spatial variability for coda mismatch than for onset mismatch. This is also in agreement with results from an earlier reaction time study (Mooshammer et al., 2015) in which we showed that the execution time for producing a word pair is significantly longer for coda mismatch than for onset mismatch. This lengthening mainly took place during the final rhyme of the word pair. Further evidence has been provided by Tiede et al. (2011), who showed in a study of head movements linked to alternating production that speakers nodded more for coda mismatch than for onset mismatch conditions, which was interpreted as a correlate of difficulty in planning or producing the speech stimuli.

Three additional conditions were included to extend these findings. In the first condition we included sequences with both onset and coda mismatch (e.g., *pop tot*). Results for this condition showed the highest degree of error-associated spatial variability across conditions, though this varied by alternating articulator; some of the reduced variability observed for coda /t/ was likely due to allophonic glottalization. In the second, error rates were compared for alternat-

ing CVC and (CV/VC) pairs (e.g., *top ta*, *top op*), with results showing about half the percentage of intrusion and substitution errors found for mismatched alternating codas. Thirdly, error rates were evaluated for alternating bisyllabic words (e.g., *ticky picky*), with results showing fewer errors despite the greater time pressure for production (two syllables produced within the same interval as a monosyllabic word). Delta values also showed greater spatial variability for monosyllabic word pairs compared to bisyllabic pairs.

The single mismatch results showing more errors and greater variability for coda alternation than for onsets is consistent with predictions of the SCM (Sevald & Dell, 1994), which attributes longer execution times and more errors in the coda mismatch condition to a reactivation of the first word in a pair triggered by identical initial consonants. According to this model this leads to competition of the differing coda consonants and consequently to more errors which – following Goldrick and Blumstein (2006) – are a consequence of increased variability. However, it is worth emphasizing that the data in the current study were elicited over multiple repetitions (approx. 20 per trial) using an accelerating metronome paradigm. With immediate repetition, all parts of the stimulus can be taken as reactivating everything in the sequence with the phonemes still being activated because decay is too slow (see Dell, 1986). Alternatively, repetition of this kind may involve execution from an activated planning buffer without lexical retrieval after the first instance, and thus without the potential for competition after initial planning. In addition, the SCM in its pure form predicts similar delta values for onsets and codas in the double mismatch condition (e.g., *pop tot*) because their different onsets do not reactivate the preceding words (as they would for *top tock*); however, our results show significantly increased variability for the codas relative to onsets and more frequent reductions. It is unclear how the SCM would account for the lower error rates observed for alternating closed and open/onsetless syllables (e.g., *top ta/op*), and bisyllabic words (e.g., *picky ticky*) relative to CVCs. Sevald et al. (1995) found a clear advantage for repeating the syllable structure with shorter execution times and lower perceived error rates compared to alternating syllable structures, which is opposite to what was found in the current study. This difference could either be due to differences in the stimulus material (Sevald et al., 1995, tested CVCs and bisyllabic words, e.g. *kil kil.per*), or to the different measures used.

One alternative explanation for the higher error rates for coda mismatches than for onset mismatches is based on the proposal that syllables (and by extension monosyllabic words) have a binary structure of onset and rhyme. On this view, the vowel is structurally more tightly tied to the coda consonant than to the onset consonant. As a result, the identity of the vowel from one word to the next may have a more powerful effect on the coda consonant, drawing it into more errors and provoking greater variability.

Another alternative explanation of asymmetric error effects in the single mismatch condition follows from the CGM of Nam and Saltzman (2003; see also Nam et al., 2009) in which onsets differ from codas in their phasing with respect to the vowel nucleus: onsets are timed to initiate *in-phase* with the vowel, while codas are timed to execute *antiphase* to the vowel. Because the antiphase relationship is inherently less stable, variability and thus the possibility for errors is increased in coda alternations. Accordingly, the larger in-phase coupling

Mooshammer/Tiede/Shattuck-Hufnagel/
Goldstein

strength between onset and vowel in a CV sequence reduces variability in the onset, accounting for lower error rates and spatial variability in the onset mismatch condition.

The highest error rates and greatest delta variability measures for onset alternation were observed when /p/ was in the coda (e.g., *top cop* and *tape cape*, gray circles in Fig. 7), predicted by neither the SCM nor the CGM. For these trials a physical explanation may be relevant, as the oscillation of the tongue apex and dorsum in producing the target constrictions may provide competing demands on tongue placement facilitating coconstriction errors (see also Slis and van Lieshout, 2013, 2016).

Other aspects of the findings can be explained using the error generation hypothesis in Goldstein et al. (2007). As explained in the Introduction, in this view intrusions occur because of a shift from a more complicated 2:1 mode of frequency for alternating articulators in *cop top* to a more stable 1:1 mode. The pressure to entrain in the monosyllabic conditions would expected to be high, because the lip oscillations are approximately in-phase with the tongue tip oscillations and with the tongue dorsum oscillations. This is so because in English, a coda consonant overlaps substantially with a following onset consonant (Goldstein and Pouplier, 2014). This entrainment and frequency-locking hypothesis can also account for the reduced error rate in the bisyllabic condition, e.g., *picky ticky*. In this condition, there is a 2:1 relation between the tongue dorsum oscillations and the lip oscillations, and also between the tongue dorsum oscillations and the tongue tip oscillations. However, because of the extra vowel, these oscillations are not in-phase (there is a substantial interval of time between the tongue dorsum gesture and a following lip or tongue tip gesture). Therefore, the coupling among the oscillators in this case would be expected to be weaker, and so the pressure to shift to a 1:1 mode of frequency locking would also be expected to be weaker, thus predicting fewer errors in this condition than in the monosyllabic conditions.

Reduced error rate in the missing condition (e.g., *top ta*) can also be explained by frequency mode-locking. This is due to the fact that, in this condition, for every second syllable only one articulator is active. Therefore, while this condition does display a 1:2 frequency relation between oscillators, it has only one such relation (as opposed to the two such relations in *cop top*). Fewer active 1:2 relations probably lead to less potential variability and cross-talk between the oscillators.

For the double mismatch condition, another element of dynamic instability comes into play, that of the relative phase of oscillating constrictions. For example, in *pop tot*, there are oscillations of the lips and tongue tip that are in a 1:1 frequency mode, which would be expected to be stable. However, the relative phase of those oscillations changes on every cycle, rather than being fixed. To see this, note that at the end of the word *pop*, the lip constriction and the tongue tip constriction for the following /t/ will overlap, with the lip constriction preceding. However, on the next cycle (end of the word *tot*), they overlap again, but now the tongue tip precedes. So, the relative phase of lip and tip oscillations changes on every cycle, which from this perspective is an unstable situation.

Finally, it is worth observing that the two approaches to error quantification used here – delta values and error rates – lead to complementary but simi-

lar results. Applying the delta measure has the advantage that it is not based on an arbitrary threshold (McMillan and Corley, 2010) distinguishing between errors and "normal" variation. Furthermore, it can also be calculated in cases for which the error counts are not applicable, due to constraints on the articulators, i.e. the double mismatch condition (*pop tot*). Error rates based on distributions, on the other hand, support distinguishing different types of errors, i.e. intrusions, reductions and substitutions, and therefore provide a more nuanced view of the results. One of the major results in this study is that mismatch in the coda induces more variability and a larger error frequency compared to mismatches in the onset. However, this is partly due to a phonetic allophonic process in English, namely that coda /t/ is often reduced, substituted by a glottal stop or deleted completely. Since the delta value does not distinguish between reduction and intrusions this could not have been uncovered. By a regression analysis we showed that both methods give similar results. Therefore, we suggest that – since both methods have advantages and disadvantages – it is best to apply both together because they provide complementary insight.

Appendix

	F1	F2	F3	F4	F5	M1	M2	M3	M4
Set 1: onset and coda mismatch									
pack_tack/tack_pack	2	1	2	2	2	1	2	2	0
pock_tock/tock_pock	2	2	2	2	2	1	2	2	1
pit_kit/kit_pit	2	2	2	2	2	1	2	2	0
cod_pod/pod_cod	2	2	2	2	2	1	2	2	2
cop_top/top_cop	2	2	2	2	2	2	2	2	2
cape_tape/tape_cape	2	1	2	2	2	1	2	2	0
cape_Kate/Kate_cape	1	1	1	2	2	1	2	2	0
cop_cot/cot_cop	2	1	2	2	2	1	2	2	2
cob_cod/cod_cob	1	1	2	2	2	1	2	2	2
take_tape/tape_take	1	1	2	2	2	1	2	2	0
tock_top/top_tock	1	1	2	2	2	1	2	2	2
tack_tap/tap_tack	1	1	2	2	2	1	2	2	0
tick_tip/tip_tick	2	2	2	2	2	2	2	2	0
pack_pat/pat_pack	1	1	2	2	2	1	2	2	0
pod_pog/pog_pod	0	0	2	2	2	0	2	2	2
Both									
pop_tot/tot_pop	4	0	4	3	4	3	4	2	0
pip_kick/kick_pip	4	0	4	4	2	2	4	4	0
Set 2									
top_op/op_top	0	0	2	2	2	0	2	2	0
cop_op/op_cop	0	0	2	2	2	0	2	2	0
cop_top/top_cop	0	0	2	2	2	0	2	2	0
top_ta/ta_top	0	0	2	2	2	0	2	2	0
cop_Kaa/Kaa_cop	0	0	1	2	2	0	1	2	0
cop_cot/cot_cop	0	0	2	2	2	0	2	2	0
tock_top/top_tock	0	0	2	2	2	0	2	2	0

Set 3									
caper_taper/taper_caper	2	2	0	0	0	1	0	0	0
picky_ticky/ticky_picky	2	2	0	0	0	1	0	0	0
cape_tape/tape_cape	2	1	0	0	0	1	0	0	0
kit_pit/pit_kit	2	2	0	0	0	1	0	0	0
caper_cater/taper_taker	2	2	0	0	0	2	0	0	0
picky_pity	1	1	0	0	0	1	0	0	0
ticky_tippy	1	1	0	0	0	1	0	0	0
cape_Kate	1	1	0	0	0	1	0	0	0
tape_take	1	1	0	0	0	1	0	0	0
tick_tip/tip_tick	2	2	0	0	0	2	0	0	0

Acknowledgments

Statement of Ethics

All participants read and signed an informed consent and were paid for their participation.

Disclosure Statement

The authors have no conflict of interest to declare.

References

Baars, B., Motley, M., & MacKay, D. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*(4), 382–391. doi:10.1016/S0022-5371(75)80017-X

Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511801686

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01

Berg, T. (1991). Phonological processing in a syllable-timed language with pre-final stress: Evidence from Spanish speech error data. *Language and Cognitive Processes, 6*(4), 265–301. doi:10.1080/01690969108406945

Browman C (1978). Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics, 42*, 1–149.

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*(2-4), 140–155. doi:10.1159/000261823

Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée, 5*, 25–34.

Butterworth, B., & Whittaker, S. (1980). Peggy Babcock's relatives. In G. Stelmach & J. Requin (Eds.), *Tutorials in Motor Behavior* (pp. 647–656). Hillsdale, N.J.: Erlbaum. doi:10.1016/S0166-4115(08)61976-7

Byrd, D. (1996a). Influences on articulatory timing in consonant sequences. *Journal of Phonetics, 24*(2), 209–244. doi:10.1006/jpho.1996.0012

Byrd, D. (1996b). A phase window framework for articulatory timing. *Phonology, 13*(02), 139–169. doi: 10.1017/S0952675700002086

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics, 24*(2), 263–282. doi: 10.1006/jpho.1996.0014

Cohen-Goldberg, A. (2012). Phonological competition within the word: Evidence from the phoneme similarity effect in spoken production. *Journal of Memory and Language, 67*(1), 184–198. doi:10.1016/j. jml.2012.03.007

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 29*(3), 416–431. doi:10.1037/0278-7393.29.3.416

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 10*(2), 222–233. doi:10.1037/0278-7393.10.2.222

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. doi:10.1037/0033-295X.93.3.283

Dell, G., & Reich, P. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*(6), 611–629. doi:10.1016/S0022-5371(81)90202-4

Frisch, S., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics, 30*(2), 139–162. doi:10.1006/jpho.2002.0176

Fromkin, V. (1973). *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Goldrick, M., & Blumstein, S. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes, 21*(6), 649–683. doi: 1080/01690960500181332

Goldstein, L., & Pouplier, M. (2014). The temporal organization of speech. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *Oxford handbook of language production*. Oxford, UK: Oxford University Press.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition, 103*(3), 386–412. doi:10.1016/j.cognition.2006.05.010

Haken, H., Kelso, J. A., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics, 51*(5), 347–356. doi:10.1007/BF00336922

Hartsuiker, R. (2006). Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes, 21*(7-8), 856–891. doi:10.1080/01690960600824138

Houghton, G. (1990). The problem of serial order: A neural network model of sequential learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current Research in Natural Language Generation* (pp. 287–319). London: Academic Press.

Huffman, M. (2005). Segmental and prosodic effects on coda glottalization. *Journal of Phonetics, 33*(3), 335–362. doi:10.1016/j.wocn.2005.02.004

Kelso, J., Buchanan, J., DeGuzman, G., & Ding, M. (1993). Spontaneous recruitment and annihilation of degrees of freedom in biological coordination. *Physics Letters [Part A], 179*(4-5), 364–371. doi: 10.1016/0375-9601(93)90692-S

Kühnert, B., & Hoole, P. (2004). Speaker-specific kinematic properties of alveolar reductions in English and German. *Clinical Linguistics & Phonetics, 18*(6-8), 559–575. doi:10.1080/02699200420002268853

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). *lmerTest: Tests in linear mixed effects models. R package version 2.0-32*. Retrieved from https://CRAN.R-project.org/package= lmerTest.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38. doi:1017/S0140525X99001776

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*(3), 323–350. doi:10.1016/0028-3932(70)90078-3

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition, 117*(3), 243–260. doi:10.1016/j.cognition.2010.08.019

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition, 42*(1-3), 181–211. doi:10.1016/0010-0277(92)90043-H

Meyer, D., & Gordon, P. (1985). Speech production: Motor programming of phonetic features. *Journal of Language and Memory, 24*(1), 3–26. doi:10.1016/0749-596X(85)90013-0

Mooshammer, C., Goldstein, L., Nam, H., McClure, S., Saltzman, E., & Tiede, M. (2012). Bridging planning and execution: Temporal planning of syllables. *Journal of Phonetics, 40*(3), 374–389. doi:10.1016/j.wocn.2012.02.002

Mooshammer, C., Tiede, M., Katsika, A., & Goldstein, L. (2015). Effects of phonological competition on speech planning and execution. *Proceedings of the 18th International Congress of Phonetic Sciences*.

Mowrey, R. A., & MacKay, I. R. (1990). Phonological primitives: Electromyographic speech error evidence. *The Journal of the Acoustical Society of America, 88*(3), 1299–1312. doi:10.1121/1.399706

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 299–328). Berlin, Germany: Mouton de Gruyter.

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of Phonetic Sciences*: 2253-2256.

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. In G. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 261–280). Cambridge, UK: Cambridge University Press.

O'Seaghdha, P., Dell, G., Peterson, R., & Juliano, C. (1992). Modelling form-related priming effects in comprehension and production. In R. Reilly & N. Sharkey (Eds.), *Connectionist Approaches to Language Processing* (pp. 373–408). Hillsdale, NJ: Erlbaum.

O'Seaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology. Human Perception and Performance, 26*(1), 57–73. doi:10.1037/0096-1523.26.1.57

Parrell, B., & Narayanan, S. (2018). Explaining Coronal Reduction: Prosodic Structure and Articulatory Posture. *Phonetica, 75*(2), 151–181. doi:10.1159/000481099

Peterson, R. (1991). *A phonological competition model of form-related priming effects* (Unpublished doctoral dissertation), University of Rochester, Rochester, NY.

Peterson R, Dell G, & O'Seaghdha P (1989): A connectionist model of form-related priming effects. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society* (pp. 196–203). Hillsdale, NJ: Erlbaum.

Pinheiro, J., & Bates, M. (2000). *Mixed effects models in S and S-PLUS*. New York: Springer-Verlag. doi:10.1007/978-1-4419-0318-1

Pouplier M (2003): The dynamics of error. *Proceedings of the 15th International Congress of the Phonetic Sciences* (pp 2245–2248).

Pouplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics, 36*(1), 114–140. doi:10.1016/j.wocn.2007.01.002

Pouplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics, 33*(1), 47–75. doi:10.1016/j.wocn.2004.04.001

R Core Team (2016). R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333–382. doi:10.1207/s15326969eco0104_2

Saltzman, E., Nam, H., Goldstein, L., & Byrd, D. (2006). The distinctions between state, parameter and graph dynamics in sensorimotor control and coordination. In M. L. Latash & F. Lestienne (Eds.), *Motor Control and Learning* (pp. 63–73). New York: Springer Publishing. doi:10.1007/0-387-28287-4_6

Sevald, C. A., & Dell, G. S. (1994). The sequential curing effect in speech production. *Cognition, 53*(2), 91–127. doi:10.1016/0010-0277(94)90067-1

Sevald, C. A., Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language, 34*(6), 807–820. doi:10.1006/jmla.1995.1035

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

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gupnik (Eds.), *Motor and Sensory Processes of Language* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition, 42*(1-3), 213–259. doi:10.1016/0010-0277(92)90044-I

Shattuck-Hufnagel, S. (2015). Prosodic frames in speech production. In Redford, M. (Ed.), *The handbook of speech production* (pp. 419–444). Hoboken, NJ: Wiley & Sons.

Slis, A. (2018). Articulatory variability and speech errors: An overview. *Toronto Working Papers in Linguistics, 40*, p. 15.

Slis, A., & Van Lieshout, P. (2013). The effect of phonetic context on speech movements in repetitive speech. *The Journal of the Acoustical Society of America, 134*(6), 4496–4507. doi:10.1121/1.4828834

Slis, A., & van Lieshout, P. (2016). The effect of phonetic context on the dynamics of intrusions and reductions. *Journal of Phonetics, 57*, 1–20. doi:10.1016/j.wocn.2016.04.001

Tiede M, Goldstein L, Mooshammer C, Nam H, Saltzman E, & Shattuck-Hufnagel S (2011): Head movement is correlated with increased difficulty in an accelerating speech production task Talk presented at the *9th International Seminar on Speech Production*, Montreal.

Vousden, J. I., Brown, G. D., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41*(2), 101–175. doi:10.1006/cogp.2000.0739

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America, 130*(3), 1606–1617. doi:10.1121/1.3621306

Wilshire, C. E. (1998). Serial order in phonological encoding: An exploration of the 'word onset effect' using laboratory-induced errors. *Cognition, 68*(2), 143–166. doi:10.1016/S0010-0277(98)00045-6