

Production Planning and Coronal Stop Deletion in Spontaneous Speech

James Tanner*, Morgan Sonderegger, Michael Wagner

Department of Linguistics, McGill University, Canada

Abstract

Many phonological processes can be affected by segmental context spanning word boundaries, which often lead to variable outcomes. This paper tests the idea that some of this variability can be explained by reference to production planning. We examine coronal stop deletion (CSD), a variable process conditioned by preceding and upcoming phonological context, in a corpus of spontaneous British English speech, as a means of investigating a number of variables associated with planning: prosodic boundary strength, word frequency, conditional probability of the following word, and speech rate. From the perspective of production planning, (1) prosodic boundaries should affect deletion rate *independently* of following context; (2) given the locality of production planning, the effect of the following context should decrease at stronger prosodic boundaries; and (3) other factors affecting planning scope should modulate the effect of upcoming phonological material above and beyond the modulating effect of prosodic boundaries. We build a statistical model of CSD realization, using pause length as a quantitative proxy for boundary strength, and find support for these predictions. These findings are compatible with the hypothesis that the locality of production planning constrains variability in speech production, and have practical implications for work on CSD and other variable processes.

1 Introduction: Variability and Coronal Stop Deletion

Research within sociolinguistics, within the framework of prosodic phonology, within phonetics, and within the literature on probabilistic reduction has uncovered rich patterns of variability in the application of phonological processes. Less is known about *why* certain types of processes are variable, and what determines the structure of this variability, such as in ‘sandhi’ processes that span word boundaries, which tend to be more variable than word-internal processes.

This paper examines the realization of word-final coronal stops in English (coronal stop deletion: CSD, a.k.a. t/d deletion). Our main interest is the effect of the phonological content of an upcoming word on CSD, e.g. the fact that CSD is more likely to apply if a consonant follows compared to when a vowel follows, and that it is affected by whether a word is followed by a pause. The effect of the upcoming phonological context is probabilistic rather than categorical, which has been identified as a recurring property of across-word phonological interactions in a number of literatures interested in variability.¹

The hypothesis we explore in this paper is that the effect of the phonological content of a following word is *necessarily* variable because of the way speech production planning is constrained: speakers do not reliably plan out the phonological and phonetic detail beyond the current word ahead of time. As

*Corresponding author address: james.tanner@mail.mcgill.ca

a result of this flexibility in planning scope, the phonological details of an upcoming word may or may not be available yet at the time when the current word is being planned. Only if the details of upcoming word are already known at the time of planning of the current word can they exert their conditioning effect on a phonological process. We hypothesize that the locality of production planning is one source of variability in sandhi processes (cf. discussion in Wagner, 2012).

We will refer to this as the Production Planning Hypothesis (PPH). This basic intuition is clear: the phonological content of a word can only exert its influence if it has actually been planned out sufficiently to make that content available. If the application of phonological processes across words is indeed directly constrained by the locality of production planning, we expect that factors that have been shown to affect the planning of words should interact with the application of sandhi processes. This perspective suggests a close relationship between the conditioning environment of a process and whether and to what degree it will be variable, and predicts that sandhi processes that rely on detailed phonological information about an upcoming word should always be variable.

This paper examines the predictions of the Production Planning Hypothesis for CSD, using a corpus of spontaneous British English speech, by addressing three research questions: (1) How does duration of a pause following the coronal stop (which serves as a proxy for boundary strength) affect deletion rate? (2) How does boundary strength (i.e. pause duration) modulate the effects of surrounding segments on deletion rate? (3) How do other factors influencing the size of the planning window in production planning, such as measures of word predictability and speech rate, modulate the effects of the following phonological context on deletion rate?

Our hypothesis is that information about the following context is probabilistically available only if the word following the CSD environment has been planned. If the likelihood of planning is indeed inversely correlated with the strength of the prosodic boundary, the hypothesis related to (1) is that as the strength of a prosodic boundary increases, the probability of deletion decreases. The prediction with respect to (2) is that the effect of the following context on CSD should be *gradiently modulated by the strength of the prosodic boundary*. As the preceding context always falls in the same local planning domain as the CSD target, we also predict that the strength of the prosodic boundary does *not* condition the effect of the preceding segment. With respect to (3), we expect that other factors affecting planning scope, such as the frequency of the target word, bigram probability of the word and the following word, and the conditional probability of the following word, should modulate the effect of the following word, as well as speech rate. We empirically test these predictions of the PPH on a corpus of spontaneous British English speech, using pause length as a quantitative proxy for prosodic boundary strength.

Variability of sandhi processes, and the variability CSD more specifically, have figured prominently in a number of separate strands of research. In the remainder of this section, we review some of these findings, focusing mostly on the effect of the upcoming phonological context on CSD, and previous ideas on how it can be accounted for. In section 2, we show how the PPH provides a new rationale for some of the observed patterns, and also makes new predictions for the structure of the observed variability, before turning to our new data in section 3. In section 4, we report on the results of the statistical model, which are interpreted and discussed in section 5.

1.1 Observations about Variability in Sociolinguistics

CSD involves the deletion of the final t/d in a word-final cluster ending in a coronal stop, resulting in pronunciation of words like *mist* or *bold* without an audible final stop. Decades of work in variationist sociolinguistics and phonetics have studied CSD and other variable processes (‘variables’) in particular languages in detail and explored the linguistic and social factors conditioning a variable’s rate of ap-

plication. CSD is one of the best-studied variables in the sociolinguistic literature of the past decades (beginning with Labov et al., 1968; Fasold, 1972), and the factors which condition CSD rate are well understood after decades of analysis.

CSD has been shown to be conditioned by a number of linguistic and non-linguistic factors (reviewed by Schreier, 2005; Hazen, 2011; Tagliamonte and Temple, 2005), most of which have qualitatively similar effects on deletion rate across dialects of English. Amongst these, understanding how the phonological environment affects the likelihood of application for processes like CSD is a major concern in the variationist sociolinguistic literature. The rate of deletion has been shown to vary depending on the type of segment at the start of the upcoming word (Guy, 1980b): this ‘following context effect’ is usually the conditioning factor which has the greatest effect on deletion rate (versus e.g. speech rate, preceding context), with deletion occurring more often before more similar segments, and consonants inducing higher rates of deletion than vocalic segments (Guy, 1980b, 1991a; Hazen, 2011; Schreier, 2005; Tagliamonte and Temple, 2005; Temple, 2009).

Despite the robustness of the following consonant > vowel ordering on CSD rate, previous studies have reached different conclusions about the rate of deletion before pauses. Guy (1980a) found that, for Philadelphia speakers, pauses patterned more similarly to vowels, whilst more similarly to consonants for New York speakers. And Tagliamonte and Temple (2005), amongst many others, observed pauses to induce the least deletion of following contexts, while Hazen (2011) observed higher deletion rates before pauses than before vowels. A potential cause of the discrepancies found for the effect of pause in previous CSD studies may be methodological differences in defining the presence of a pause, which has usually been defined structurally as the lack of a segment following the t/d environment (Kendall, 2013). Instead of viewing pause as one out of several possible following contexts, we will consider pause as an *independent* factor, coded as a *gradient* variable (pause duration), that can modulate the effect that the phonological context following the pause has on the previous word. Whether pause has such a modulating effect has not to our knowledge been tested in previous work.

Compared with the following context, the preceding context has often been shown to play a consistent but weaker role in conditioning CSD. Generally speaking, deletion occurs more frequently after sonorants than obstruents, but sibilant fricatives often induce the highest rates of deletion (Hazen, 2011; Tagliamonte and Temple, 2005). Differences between the effect of the preceding context and the following context are of interest here: Preceding segments are necessarily planned at the time at which a speaker decides whether or not to produce a [t,d], and our hypothesis therefore predicts no modulation by factors affecting production planning. Speech rate has also been long thought to affect CSD, where deletion increases in likelihood in faster speech (Guy, 1980a; Guy et al., 2008),² but whether speech rate modulates the effect of the following phonological context, which is of relevance here, has not been explored.

Another conditioning factor of considerable interest in the sociolinguistic literature is morphological class, which actually *has* been examined with respect to its interaction with the effect of following phonological context (Guy, 1991a). The general observation about morphological class is that past-tense forms show deletion less frequently than non-past-tense forms, and that weak past-tense (e.g., *missed*, *walked*) forms show less deletion than strong (irregular) past tense (e.g., *kept*) forms (Coetzee and Pater, 2011; Guy, 1991a). However, Tagliamonte and Temple (2005) found morphological class did not significantly condition CSD in York (UK) English, and argued that the differing deletion rates by morphological class are instead due to preceding phonological context (which is correlated with morphological class) (see Hazen, 2011 for a similar interpretation). We will include morphological class as a control variable in our model, but will not discuss it any detail.

Within the sociolinguistic literature more generally, variable processes are often thought of as rules

can be indexed with a probability of application, which determines ‘the ratio of cases in which the rule actually does apply to the total population of utterances in which the rule can possibly apply’ (Labov, 1969). Our main interest here, however, is not variability per se, but the variable influence of phonological information that spans across word boundaries. This formalization of rule variability by itself does not, however, help account for why cross-word processes should be more variable than within-word processes—in fact, the *same* processes that are essentially invariable within words are often (more) variable in their application across word boundaries, e.g. flapping in North American English (Nespor and Vogel, 1986).

Relevant to our present discussion, Guy (1991b,a) proposed an account of the morphological effects observed in CSD, which provides a partial account for the asymmetry between within-word applications and across-word applications of rules when it comes to their degree of variability. Combining Labov’s variable rules with the model of lexical phonology, Guy argues that if the phonological environment for a process is met at each step in derivational theory, it will get multiple opportunities to apply. This ‘exponential’ model of variable rule application captures the differences between mono-morphemes, and irregular and regular past-tenses in their CSD rate. Guy (1991a) discusses some predictions of this model for the effect of following conditioning environment, but this model makes no prediction about how pauses or other factors affecting the availability of a following word should interact with the effect of a following word.

Other approaches to variability within phonological theory have modelled variability by positing that speakers have internalized multiple grammars, grammars with partially ranked constraints, probabilistically ranked constraints (Boersma and Hayes, 2001), constraints indexed to certain lexical items or lexical strata, or weighted constraints (see Anttila, 2007 and Coetzee and Pater, 2011 for review). Work in this vein develops formal grammatical models of how different factors condition a variable’s rate of application, often to address higher-level questions such as: what the set of possible patterns of variation are for a given variable, across dialects (e.g. why is deletion rate never higher before vowels than before consonants), and why variability occurs in some contexts but not others (e.g. in codas but not in onsets, for CSD). For example, Coetzee and Kawahara (2013) propose an account for why high frequency words are more likely to undergo deletion in CSD, by weighting faithfulness constraints depending on the lexical frequency of the words involved. The idea that the effect of phonological environment external to a word (like the onset of a following word in the case of CSD) might interact with factors like pause duration and word frequency has not been explored, however, with the notable exception of Coetzee (2009), who reports that frequency does not interact with the effect of the phonological environment in an experimental test of which factors affect intuitions about the likelihood of t/d deletion. Such interactions between measures of frequency, or more complex probabilistic measures (like the conditional probability of a following word), are not part of what existing formal models can capture. If they turn out to be real, this would require some modification of these models.

1.2 Variable Processes in Prosodic Phonology

Variable phonological processes have also been central in another, largely separate literature, that of prosodic phonology (Kaisse, 1985; Nespor and Vogel, 1986; Selkirk, 1986). Prosodic phonology is concerned with the representation of prosodic phrasing in sentences, and one source of evidence is sandhi processes. In prosodic phonology, particular processes are associated with specific prosodic domains, and effectively serve the purpose of encoding phonological domains. Some processes have been characterized as ‘fortition’, such as the strengthening observed at the beginning of prosodic domains and related phonological phenomena (see Keating, 2006, for a review), other have been characterized as ‘lenition’ within prosodic domains (Kingston et al., 2008; Katz, 2016).

A recurring observation in this literature is that sandhi rules tend to be inherently variable, and affected by speech rate (e.g. Hasegawa, 1979; Kaisse, 1985; Kiparsky, 1985). The reason *why* phonological processes that span word boundaries tend to be more variable than word-internal processes is usually not discussed. Some models assume that phonological processes apply categorically within particular types of prosodic domains, for example, a process might apply within the phonological phrase, but not across phonological phrase boundaries (e.g. Nespor and Vogel, 1986; Selkirk, 1986). Nespor and Vogel (1986) try to explain some types of variability as being due to variability in the choice between different phrasing options. The reason tapping across word boundaries, for example, is variable, is linked to variability in phrasing options. If different levels of prosodic structure differ in how variable they are, this perspective could in principle account for some of the structure in the variability that we observe: The greater variability of sandhi processes compared to within-word processes could be seen as a consequence of the greater level of variability in the assignment of higher level prosodic structure. Such an explanation would be non-circular if independent criteria to establish phrasing can be established. For the most part, however, the variability of cross-word phonological processes is taken as a given in this literature. We know of no model in this domain that would predict gradient modulating effects of the effects of the phonological content of upcoming words, that is, the types of effects we examine here.

1.3 Variability in Phonetics and Articulatory Phonology

Some sandhi processes clearly seem categorical, for example gemination in Sardinian (Ladd and Scobbie, 2003), or liaison in French (Post, 2000). But others may be inherently gradient, such as place assimilation across word boundaries, which tends to be variable and involve various degrees of assimilation (cf. Nolan, 1992; Niebuhr et al., 2011, for review). CSD might also fall into this class of gradient process. Even when it sounds like [t] or [d] is deleted, the underlying gestures are often still present, but either overlap with adjacent gestures or are not fully realized, leading to the *appearance* of deletion. Such ‘hidden’ gestures have been reported for assimilation and coarticulation patterns (Hardcastle, 1985; Barry, 1985, 1992), but also for CSD (Browman and Goldstein, 1990). Browman and Goldstein (1990) discuss renditions of “perfect memory”, in which the words were either separated by an intonational phrase or produced as a single prosodic phrase. In the latter type of rendition, the [t] was often not audible (and would therefore likely be transcribed as deleted), and yet articulatory evidence suggests that the gesture was partially realized. More evidence that CSD is better characterized as a gradient phenomenon rather than a variable categorical rule is presented in Temple (2014).

Browman and Goldstein (1990, 1992) proposed that all examples of ‘fluent speech alternations’ are due to such gradient changes in the articulatory gestures involved, rather than due to the application of categorical phonological processes. Relevant for our discussion here, they argue that in certain prosodic configurations, gestures can gradiently change with respect to their degree of overlap, and with respect to the degree in gestural magnitude. These effects, they argue, will be modulated by the strength of prosodic boundaries and other prosodic factors: The lower degree of lengthening at weaker prosodic boundaries will lead to greater gestural overlap; the lower amount of time for articulatory movement in weak prosodic positions will make it more likely that segments will appear to be reduced or even deleted. This latter effect of a decrease in magnitude of gestures in weak positions is related to the very common idea that hypoarticulation is a form of effort reduction (Lindblom, 1990; Kirchner, 1998). This account predicts that, more generally, prosodic structure can interact with the gestural score and cause greater or lower overlap or greater or lower gestural magnitude under certain circumstances. Prosodic boundaries induce a slow down of the gestural movements, which within Articulatory Phonology is often thought of as a separate π gesture (Byrd and Saltzman, 2003). The greater the magnitude of the π gesture, the greater the modulation of the gestural score.

The view from Articulatory Phonology therefore provides a way to interpret modulating effects of prosodic boundaries on sandhi processes, and makes predictions that partially overlap with those of the PPH explored here. For example, AP predicts that pause duration should have an effect on CSD: The greater slow-down at prosodic boundaries should decrease gestural overlap and hence make the appearance of CSD less likely (Browman and Goldstein, 1990, 1992). Although not discussed in their papers, it also predicts that pause duration should modulate the effect of the following environment on CSD, since greater pause duration will necessarily come with smaller degree of gestural overlap. However, even if CSD often involves gestural overlap or gestural undershoot, this does not mean that CSD is not planned, and that the locality of production planning should not matter. Whalen (1990), for example, argues that coarticulation is largely planned and argues against the idea that it can be explained as an automatic effect of the temporal overlap of gesture in production, as is sometimes assumed in AP and related overlap accounts. The PPH makes predictions about the locality of planning effects, be they categorical as in the case of deletion or gradient as in the case of gestural overlap. If it is correct, we should be able to see additional effects of production planning factors when holding the durational and prosodic factors constant that AP predicts to affect gestural realization, as we will outline in section 2.

1.4 Variability and probabilistic reduction

Finally, variable phonological process have more recently figured prominently in the literature on probabilistic reduction. While a basic relationship between reduction and probability has long been established (e.g. Jespersen, 1922; Zipf, 1929), this relationship has more recently become a central concern in the field, dubbed the ‘Probabilistic Reduction Hypothesis’ in Jurafsky et al. (2001). The idea is supported by findings that show that frequent words and words that are highly predictable in a certain context tend to be reduced in terms of their phonetic duration, and/or with respect to their segmental content. Such effects are often seen as a rational use of resources from an information-theoretic perspective: Highly probable words may be easier to retrieve for speakers during planning and easier to recover for listeners in perception (cf. Jurafsky et al., 2001).

CSD has been found to be affected by word frequency in several studies that explore probabilistic reduction in word-final coronal stops (e.g. Gregory et al., 1999; Jurafsky et al., 2001; Bybee, 2000; Coetzee and Kawahara, 2013) and word-medial coronal stops (Raymond et al., 2006). A recurring finding is that CSD is more likely to apply in frequent words compared to infrequent words (see also Pluy-maekers et al., 2005; Gahl and Garsney, 2006; Guy et al., 2008). However, Walker (2012), in a study of CSD in Canadian English, found that word frequency does not significantly impact deletion once lexically-specific effects and interactions are accounted for. Gregory et al. (1999) and Jurafsky et al. (2001) explored the effects of various other measures of predictability on CSD and vowel reduction, in addition to word frequency, such as conditional probability or mutual information in two-word sequences. To our knowledge, this literature has not yet explored whether frequency or other measures of predictability modulate the effect of a following environment. We return to why such effects would be expected in section 2.3.³

1.5 Summary

The variability of CSD has figured prominently in a several different strands of research. The goal of this paper is to relate aspects of this variability to the locality of production planning. Considering factors affecting planning scope can help rationalize some of the previous findings, and produces new predictions about the structure of the observed variability. One main focus in exploring these pre-

dictions how factors affecting production planning modulate the effect of the phonological content of upcoming words on CSD.

2 Variability and production planning

This paper explores a simple idea about one source of the variability of CSD: The effect of the phonological content of an upcoming word on the likelihood or degree of deletion should be modulated by factor affecting the likelihood that the upcoming phonological context could be planned in time to exert its effect. More generally, we argue that one source of variability in the application of sandhi processes is the locality of phonological encoding.

At least since the 1970s, it has been well known from studies of speech errors (Fromkin, 1971; Garrett, 1988) and experimental studies of production latency (Sternberg et al., 1978) that syntactic and semantic information is often planned over a relatively wide window, while phonological detail is planned in a much narrower planning window. The size of this narrow planning window remains controversial. Levelt (1989, 1992); Wheeldon and Lahiri (1997) claim that the planning window for phonological encoding comprises only about one phonological word, some authors hold that it can be as narrow as a syllable (Schriefers, 1999; Rastle et al., 2000), or even a single segment (Kawamoto et al., 2015). Whatever the minimal planning unit may be, many findings suggest that speakers preferentially plan phonological detail in small increments rather than across many words or even an entire utterance (cf. Griffin, 2003).

The recent literature converges on the idea that there may not be a fixed window size for the scope of phonological planning. Rather, the window size appears to be *flexible*, and varies by time pressure (Ferreira and Swets, 2002) and task (cf. Costa and Caramazza, 2002; Wagner et al., 2010; Konopka, 2012; Wheeldon, 2012; Fuchs et al., 2013). Crucially, the window for planning of phonological encoding can, under certain circumstances, include multiple words (Ferreira and Swets, 2002; Griffin, 2003; Jescheniak et al., 2003; Wagner et al., 2010)—as the very existence of sandhi processes suggests.⁴ The idea is then that the reason sandhi processes are variable might simply reside in the variability of planning scope in speech planning. Whether an upcoming or preceding word will be available at the time the current word is planned will depend on the factors that have been shown to affect planning scope. In this paper, we explore whether these same factors modulate the effect of upcoming phonological information on CSD. In the following section, we review the major factors that are considered to condition the window of production planning.

2.1 Prosodic Boundaries and Constituency

It is now often assumed that there are two counteracting pressures determining planning scope: one is to initiate speaking as soon as possible and pre-plan less, perhaps to relieve working memory load or to avoid missing one's turn; the other is to talk fluently, and failing to plan ahead sufficiently can result in pauses and disfluencies (cf. Fraundorf and Watson, 2008; Swets et al., 2013). If an upcoming word has not been sufficiently planned after the offset of the first, a pause or other type of disfluency will result (e.g. Fox Tree and Clark, 1997). Also, Griffin (2003) reports that when planning a two-word sequence, the latency before the utterance reflects a speaker's effort to articulate the two words without a pause separating them, at least under certain circumstances, suggesting that pauses (in this case a pre-utterance pause) reflect planning (cf. Ferreira, 1991).

Compatible with this is the observation that pauses tend to align with constituent boundaries (Cooper and Paccia-Cooper, 1980, and many others), and that one factor decisive for whether two words are planned together is whether they form part of a single syntactic constituent (Wheeldon, 2012; Fuchs

et al., 2013), which correlates with forming a meaning unit. The corpus we will examine here is not syntactically annotated, hence we cannot test for such effects directly. However, syntactic constituency and semantic cohesiveness correlate with prosodic phrasing, so our measures of the strength of the prosodic boundary between two words will include some information about their relationship.

The prosodic strength of the boundary between two words might not simply reflect planning scope, but also influence planning scope, such that planning an upcoming word is less likely across stronger boundaries (cf. Ferreira, 1993). Independent of whether prosodic boundaries directly constrain planning scope or simply reflect other factors that do, the strength of a prosodic boundary between two words can serve as a proxy measure to gauge the likelihood that the phonological information in the second was available when the first one was planned, and hence should modulate the effect of the phonological content of an upcoming word in CSD. The PPH predicts no such effect of the strength of a following boundary on the effect of the *preceding* environment.⁵

There are many measures of prosodic boundary strength, and for simplicity we will focus in this paper on the presence and duration of pauses (Goldman-Eisler, 1972; Grosjean and Collins, 1979; Price et al., 1991; Krivokapić, 2007; Kendall, 2013). In sociolinguistic literature on CSD, a following pause is treated as an environment on par with that of a following consonant or a following vowel. The perspective we take here looks at pauses very differently: Pauses modulate the distances to the following consonantal environment. Of course, beyond a certain pause length, it is very unlikely that an upcoming word would be planned, and in this limiting case *pause*, or rather, the null environment, is indeed a separate type of context. But treating *all* pauses as a separate environment might lead to misleading results, and obscure the fact that pauses might modulate the effect of the phonological material that follows them.

2.2 Speech Rate

One relevant factor affecting planning scope is speech rate. Wagner et al. (2010) used a direct manipulation of cognitive load to show that planning scope is reduced under cognitive load, and that a more incremental production strategy with a lower speech rate is cognitively less demanding than full planning. If a lower speech rate is associated with more incremental planning and a smaller planning window, we might expect the influence of the following environment to be reduced with a lowering of the speech rate. Lower speech rate has also been shown to lead to more boundaries (Turk, 2010), and also to stronger phonetic cues to boundary strength (Beckman and Edwards, 1990; Sugahara and Turk, 2009). A reviewer points out that it is not clear whether speech rate will affect planning scope or vice versa: a greater planning scope might simply enable faster speech. While we are mostly interested in whether speech rate modulates the effect of the following phonological environment, we note that higher speech rates are generally associated with more casual speech and greater reduction, so CSD should be more common for faster speech (cf. Guy, 1980a; Guy et al., 2008).

2.3 Measures of Word Probability

The likelihood of whether a speaker begins planning an upcoming word while planning the fine detail of the current word will depend on the accessibility of the two words. A first factor that might be relevant is the probability of the word that potentially undergoes CSD. One simple proxy measure for this probability is the frequency of the word in a large corpus, which has already been observed to correlate with CSD rate (Jurafsky et al., 2001). Of interest to us is whether word frequency plays a role in conditioning the rate of deletion, but also in modulating the effect of the phonological context of an upcoming word.

The effect of frequency on the production of single word utterances is relatively straightforward. The latency of producing the names of objects that have a high frequency is shorter than that of low frequency names (Oldfield and Wingfield, 1965; Goodglass et al., 1984), presumably because frequent words are easier to retrieve and plan. Other studies have found that measures of familiarity and age of acquisition, while correlated with frequency, are actually a better predictor of naming latency (Carroll and White, 1973; Morrison et al., 1992; Snodgrass and Yuditsky, 1996).⁶ The level of representation responsible for frequency effects in speech production remains controversial, however. Most models of phonological planning distinguish at least two stages, the retrieval of a general lemma corresponding with general lexical information and syntactic information associated with words, and a second stage of phonological retrieval and planning (Dell and O'Seaghdha, 1992; Levelt, 1992; Goldrick and Rapp, 2007). Some models locate the role of frequency at the point of phonological retrieval; that is, more frequent phonological word forms are retrieved faster (Levelt et al., 1999). Other models locate the role of frequency at 'lemma selection', when a certain concept is retrieved (Alario et al., 2002). A recent study trying to arbitrate between these views found that frequency effects operate at all levels, while age of acquisition might be specific to the phonological level (Kittredge et al., 2008).

The level of representation at which frequency takes effect turns out to be crucial for the predictions of the PPH for *multiple* word utterances, such as those containing a t/d-final word followed by another word. If frequency effects operate at the point of phonological retrieval, but do not affect the relative timing of retrieval at the lemma level, then we might expect that in a two-word sequence a higher frequency of the first word will have the effect that its phonological form is planned earlier relative to the phonological retrieval of the second lemma. The likelihood of the phonological form of the second form being available at the point when that of the first word is planned should then be *lower*. Concurrent with this prediction is the observation by Miozzo and Caramazza (2003) that in single word utterances, frequent distractors have a *smaller* interference effect on production latency than low frequency distractors. The proposed explanation is that frequent words are planned earlier and hence suppressed earlier relative to the target word, and therefore interfere less with its realization. The prediction of the PPH would then be that CSD should be *less* likely to be affected by the phonological shape of the upcoming word in more frequent (t/d-final) words.

It is not obvious, however, that the effect should go in this direction. A higher frequency of the first word has been associated with an increase of semantic interference due to the second word in multi-word utterances (Konopka, 2012). This would suggest that the lemma of the second word is retrieved more quickly relative to the first word as frequency of the first word increases, in line with other studies showing frequency effects for phonological planning at higher levels of representations (Alario et al., 2002; Kittredge et al., 2008). If this is true, then it could be that the *phonological* shape of the second word is also sooner available relative to that of the first word. Under this scenario the PPH would make the opposite prediction, that CSD should be *more* likely to be affected by the phonological shape of the upcoming word, the higher the frequency of the t/d-final word.

We therefore do not have a clear prediction with respect to the modulating effect of frequency. However, we note that whether phonological retrieval of a word happens sooner or later relative to a previous word is independently testable; there simply has not (to our knowledge) been a study which provides direct evidence on this point. Examining frequency effects in cross-word applications of a variable process, such as CSD, might provide evidence as to the level of representation at which frequency effects operate in general.

Whether a following word will be planned at the same time as a preceding word may also depend on the predictability of the second word. Given our interest in the effect of the phonological shape of the following word, we also want to include a measure of predictability for that word. There are

multiple ways to quantify the local predictability of a word in its context, and several have been related to degree of reduction in the prior literature.

A simple measure of this would be the frequency of the upcoming word, which serves as an estimate of its prior probability. However, from the point of view of production planning a following word with a high frequency might not necessarily be likely to be part of the same planning domain. Suppose, for example, that the following word is the first word of a new syntactic/semantic constituent or of a new clause. Since production planning is constrained by syntactic constituency (Wheeldon, 2012; Fuchs et al., 2013), it might actually be less likely to be anticipated than a low frequency word within the same constituent. In English, the first words of new constituents are often high frequency words (function words like determiners or prepositions), and words that by themselves form a separate constituent are also often high frequency words (e.g., pronouns or adverbs). In other words, in the absence of having a way to control for syntactic structure (which is not annotated in our corpus), the frequency of the following word (which we in principle predict should facilitate with the degree of influence of the following phonological environment) might actually be correlated with constituency breaks (which we predict should reduce the effect of the following environment).

A more sophisticated measure is joint probability of two words, also called their ‘bigram frequency’ or ‘string frequency’, which can be estimated by the frequency of the two-word string in a large corpus, and has been shown to correlate with reduction (Pan and Hirschberg, 2000). According to Jurafsky et al. (2001), most researchers interested in effect of word cohesion use measures that control for the frequency of one or both words instead of taking the overall bigram frequency. We followed the suggestion of a reviewer and looked at the conditional probability of the following word given the first word (calculated as described in Sec. 3), as an index of the availability of the following phonological context. This conditional probability is equivalent to the bigram probability divided by the probability of the first word, and should only be high if there is a predictive relationship between the preceding and the following word.

2.4 Other measures

There are several potentially relevant factors influencing production planning that we do not explore in this paper. The first is the length of the adjacent words, which has been argued to correlate with prosodic boundary strength (cf. Ferreira, 1991; Watson and Gibson, 2004; Krivokapić, 2007). Griffin (2003) reports that when planning a two-word sequence, the latency before the utterance increases when the first word is *shorter*. One interpretation of this result is that speakers spend more time planning the second word when the first word is short in order to avoid pauses and disfluencies between the words. However, this result was obtained in a task in which participants were explicitly instructed to say the words avoiding to pause between them, and it is not obvious that this strategy would also be used in other tasks or even in spontaneous discourse. Since we will not look at the effect of the size of adjacent words, we will not discuss this further here. A second factor we will not consider is neighborhood density. Gahl et al. (2012, 793) present evidence that words in dense neighborhoods are shorter and contain more reduced vowels, suggesting a facilitative effects of dense neighborhoods on production planning—in contrast to the greater difficulty in processing words in dense neighborhoods in perception. Finally, there are individual differences in both planning efficiency (Mortensen et al., 2008) and scope (Schriefers, 1999). Swets et al. (2014) show that planning scope correlates with working memory as measured in a reading span task, and found that speakers with high or low working memory showed very similar utterance initiation times when looking at utterances of similar length and complexity, but different planning scopes. The planning scope was evaluated by looking at measures of eye-gaze. Speakers with high working memory showed more evidence for advance planning than

speakers with low working memory. Wagner et al. (2010) used a direct manipulation of cognitive load to show that planning scope is reduced under cognitive load, and that a more incremental production strategy is cognitively less demanding than full planning. Our data set does not lend itself to explore the effects of individual differences or of cognitive load, but it is clear that the PPH would predict correlations with the application rate of sandhi processes.

2.5 Summary

Our main research questions, already anticipated in the introduction, are: (1) How does duration of a pause following the coronal stop (which serves as a proxy for boundary strength) affect deletion rate? (2) How does boundary strength (i.e. pause duration) modulate the effects of surrounding segments on deletion rate? (3) How do other factors influencing the size of the planning window in production planning, such as measures of word predictability and speech rate, modulate the effects of the following phonological context on deletion rate? The Production Planning Hypothesis for CSD makes several predictions. First, the PPH predicts that the length of the following pause will gradually reduce the probability of deletion. Second, the size of the prosodic boundary will also modulate the relative effect of upcoming segments, where the influence of the following segment will be neutralised before long pauses. Finally, words with higher predictability may be planned faster, and thus reduce the influence of upcoming phonological material. In the following, we first describe the dataset used in this study to explore these questions, and examine empirical plots (Sec. 3); fit a statistical model to test our predictions (Sec. 4); and then discuss the model’s results with respect to our research questions and the broader issues raised at the outset (Sec. 5).

3 Data

We first describe the dataset of coronal stop realization used in this study, then describe the factors which are included in our statistical model of CSD rate: those which relate to our research questions (phonological context, pause length, speech rate, word frequency), and other factors affecting CSD rate which are included as controls.

3.1 Dataset

The dataset used for analysis was taken from a corpus of spontaneous speech from the 2008 season of *Big Brother UK* (Sonderegger et al., 2016), consisting of 14259 tokens of words with consonant clusters containing underlying word-final t/d segments from 21 speakers, recorded over three months. Orthographic transcriptions of the corpus were force-aligned with the audio files using FAVE (Rosenfelder et al., 2011). Further details of the annotation process and dataset are given in Sonderegger et al. (2016).

The dataset was manually annotated by four phonetically-trained research assistants for surface realisation of word-final t/d, phonological context surrounding the t/d segment (surface realizations of preceding and following phones), and presence and duration of any pause following the t/d. To annotate the coronal stop, annotators chose from eight possible realisations of t/d, including ‘burst’, ‘glottal stop’, ‘glottalized vowel’, ‘stop closure’ and ‘none’ (no acoustic cues to t/d presence or perception of presence). Annotators used evidence from the spectrogram, waveform, and perceptual judgement. For the purposes of the analysis in this paper, any instance of surface realization (i.e. any annotation besides ‘none’) was taken as a case of non-deletion.

In the cases where the t/d was followed by a coronal stop (e.g., *want to*), t/d was taken to be realised if a separate t/d was present following the closure. In this sense, t/d segments in this context were only treated as realised if two separate t/d segments could be clearly observed. Since we do not have any articulatory data to match the acoustic data, we cannot test whether what our annotators marked as deleted [t,d]s might in fact involve gestures that were not fully realized or overlapped with adjacent gestures to give our annotators the impression that they were deleted. These annotations therefore have to be treated as abstractions over a range of degrees of reductions, as is standard in corpus studies of CSD.

Pauses were annotated by manual adjustment of pauses inserted by the forced aligner. Each force-aligned pause of less than 30 msec was set to be in fact absent (0 msec duration), as pauses shorter than this duration are likely to reflect aligner errors or low-amplitude periods of speech, such as stop closures. For each force-aligned pause of greater than 30 msec, annotators were instructed to correct the boundaries to line up with the end of the previous segment and the beginning of the next segment. More precise criteria were not given (for example, where the right edge of a pause preceding a stop closure should be placed), but annotators discussed problem cases as they arose, and attempted to keep their criteria synchronized. The goal of this semi-automatic pause annotation method was to improve on simply using force-aligned pause boundaries, which we believe it does. However, the noise introduced by various sources in this process (forced aligner, 30 msec-cutoff, manual annotation) could skew the distributions of pause durations in the dataset, and may be in part responsible for the heavily right-skewed distribution of pause durations.

Speech rate was calculated using the force-aligned transcriptions as the number of syllables per second within a phrase, defined as an interval of speech by the speaker, bounded on each side by at least 60 msec of silence (e.g. force-aligned pauses) or non-speech.⁷ Word frequencies were calculated as the count per million of a wordform in the full corpus (21 speakers).

Three sets of tokens were excluded. One L2 English speaker was excluded due to near-categorical deletion, presumably due to first language influence (L1 Thai). As most speakers used non-rhotic dialects, all tokens of words with word-final /rt/ and /rd/ clusters were removed (Tagliamonte and Temple, 2005). Finally, 1804 points were excluded where the conditional probability measure could not be reliably calculated, due to no bigrams (combination of the t/d-final word and the following word) being present in the corpus used to estimate conditional probability (see further discussion below: Sec. 3.2.4). 11504 tokens from 20 speakers were used in the final analysis.

These 11504 tokens correspond to 397 word types, of which 161 correspond to only one token, and 135 correspond to at least 5 tokens. The overall deletion rate across all word tokens is 70.1%; the type-level deletion rate, averaging across the observed deletion rates for each word type, is 41.3%.

3.2 Predictors

We model CSD rate as a function of a number of factors: those which relate to our research questions—phonological context, boundary strength (represented as pause duration), word frequency, conditional probability (of the next word, given the target word), speech rate—and additional factors which are included as controls: morphological class (the main additional factor which is expected to affect CSD rate based on previous work) and annotator identity. We first describe each predictor and the empirical trend of how it affects CSD rate, with particular attention to effects of interest for testing the PPH.⁸ In section 4 we report on the statistical modelling of the predictors and their relationships to the empirical observations.

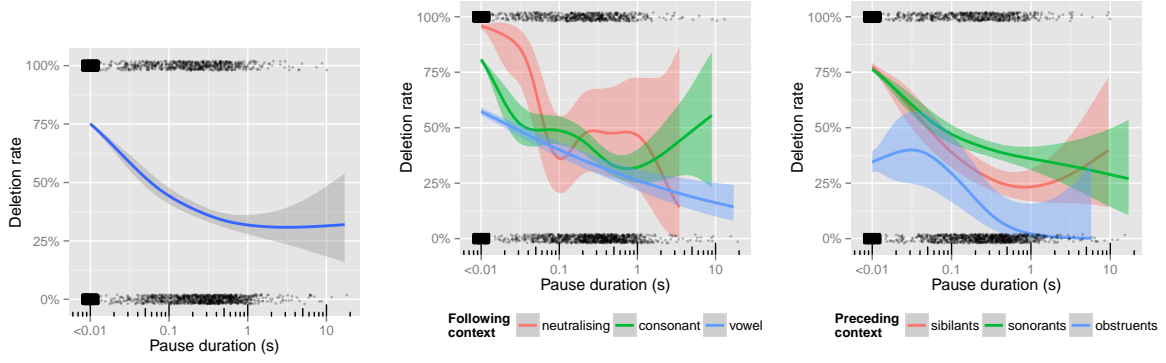


Figure 1: Empirical plot of deletion rate as a function of (log-transformed) pause duration across the whole dataset (left), and by following (middle) and preceding (right) phonological context. Dots are deleted (at 100%) and non-deleted (at 0%) tokens, with jittered positions. Solid lines and shading are non-parametric smooths (GAM, logit link) with 95% confidence intervals.

3.2.1 Phonological Context

FOLLOWING CONTEXT was coded using three levels: neutralising segments (i.e., coronal stops: /t/, /d/, where we expected the highest deletion rate given previous findings), other consonants (besides coronal stops), and vowels ($n = 895, 7137, 3472$).⁹ In contrast to most previous CSD studies, observations occurring in ‘neutralising’ environments were not excluded from the analysis. Whilst high, deletion in these positions were not categorical (type: 83.9%, token: 91.1%), and their high rate of deletion is captured in the statistical model (Sec. 4.1) as FOLLOWING CONTEXT=neutralising. The order of deletion environments follows the pattern observed across previous CSD studies: deletion rate is higher before consonants (type: 55.8%, token: 75.7%) than before vowels (type: 18.9%, token: 53.1%). Following Tagliamonte and Temple (2005), PRECEDING CONTEXT was similarly coded using three levels—sibilant obstruents, sonorants, and non-sibilant obstruents ($n = 2750, 8318, 436$)—which are expected to show progressively lower deletion rates. In our data, deletion rates were similar for sibilants (type: 51.8%, token: 71.2%) and sonorants (type: 44.5%, token: 71.7%), and substantially lower for non-sibilant obstruents (type: 25.9%, token: 31.6%).

We note that previous studies of CSD often use parametrizations of preceding and following phonological context that are different in two ways from those used here. First, they are typically more complex (with > 3 levels). The relatively simple coding used here allows us to best address our research question of how the effect of phonological context is *modulated* by boundary strength, with maximum statistical power, given that there is relatively little data for each phonological context as pause length increases. Second, our coding for preceding context differs from much previous work on CSD in the sociolinguistic and phonological variation literatures, which follow influential work such as Guy and Boberg (1997); Labov (1989) largely based on North American varieties. We follow the coding of Tagliamonte and Temple (2005) for our sample of largely British speakers, because this is the most authoritative CSD study to date on a British variety. The particular coding scheme chosen for preceding and following context could skew our results, by not accounting for further distinctions among preceding and following segments. This possibility should be mitigated by our inclusion of random intercepts for both the t/d-final word and the bigram (see Sec. 4.1), which account for aspects of the t/d-final word and the following word beyond terms included in the model.

3.2.2 Pause

Of primary interest for our research questions are (1) whether pause duration, acting as a proxy for boundary strength, gradiently affects CSD rate, independently of the effect of following context; (2) whether it modulates the effect of phonological context. Figure 1 (left) shows the trend of deletion rate as a function of pause length (using a generalized additive model-based smoother: GAM), across the dataset. There is a clear gradient effect, where longer pauses correlate with lower deletion rate (Spearman's $\rho = -0.278$). The middle and right panels of Figure 1 similarly show smoothers for deletion rate as a function of pause length, for tokens in each phonological context. The effect of following context is clearly modulated by pause length: longer pauses reduce the relative difference between the deletion rates in different contexts, eliminating them for pauses of about 100 msec or longer (in the sense of overlapping confidence intervals). The mitigating effect of pause length is especially clear for following consonant and vowel contexts; the more variable pattern for neutralizing is likely due to the far smaller amount of data in this context.

The effect of preceding context is not clearly modulated by pauses: the difference in deletion rate between different contexts does not consistently increase or decrease as a function of the duration of the following pause. We note that although Figure 1 shows empirical trends in probability space, all the same qualitative observations made here based on Figure 1 also hold in log-odds space (which is more relevant for the logistic regression models used below).

As can be observed in Figure 1, the distribution of pause length is right-skewed, and the data is unbalanced as a function of pause length (mean = 65 msec, median = 0 msec, SD = 393 msec): 85% of tokens ($n=9844$) have no following pause, and the remaining 15% of tokens are spread over a large range,¹⁰ with only half ($n = 830$) having pause duration of more than 250 msec, a commonly used minimal cutoff for (binary) 'pause' in sociolinguistic studies (Kendall, 2013, Sec. 6.3). Tokens with pause length in a given range are also unevenly distributed among phonological contexts, making estimation of the effect of pause length on CSD rate less certain: this is why the smoothers in Figure 1 have large confidence intervals in some regions, especially for longer pauses. We found that the unbalanced distribution of pause length led to problems with overfitting to short- and no-pause data (in the statistical model), if pause length is coded as a continuous variable. However, since our research questions crucially involve the *gradient* effect of pause duration on deletion rate (and its interaction with phonological context), we could not discretize pause duration as a binary variable. Instead, we coded pause duration as an ordered factor, denoted PAUSE CLASS, with four levels: no pause ($n = 9844$), $0 < \text{pause} < 105$ msec ($n = 423$), $105 \text{ msec} \leq \text{pause} < 362$ msec ($n = 618$), and $362 \text{ msec} \leq \text{pause}$ ($n = 619$).¹¹ These cut points were chosen automatically, using the `cut2` function in the `Hmisc` package in R (Jr et al., 2015). This coding allows us to examine gradient effects of pause (increasing PAUSE CLASS), without the estimates of these effects being skewed by the distribution of pause durations.

3.2.3 Speech rate

Both speech rate and frequency might be expected to affect CSD rate, based on previous work. For our research questions, we are particularly interested in whether either factor modulates the effect of following context, as is predicted if they correlate with the likelihood that the next segment is 'available' when production of the t/d is planned.

Speech rate was separated into two predictors: the mean speech rate for each speaker (SPEECH RATE MEAN), and the difference between speech rate for a given token and the speaker's mean (across all tokens), for each token from a given speaker (SPEECH RATE DEVIATION). This coding allows us to differentiate between increased deletion rate for 'faster speakers' from increased deletion rate for 'faster speech' (within a speaker) (Snijders and Bosker, 2011); both might be expected to positively correlate

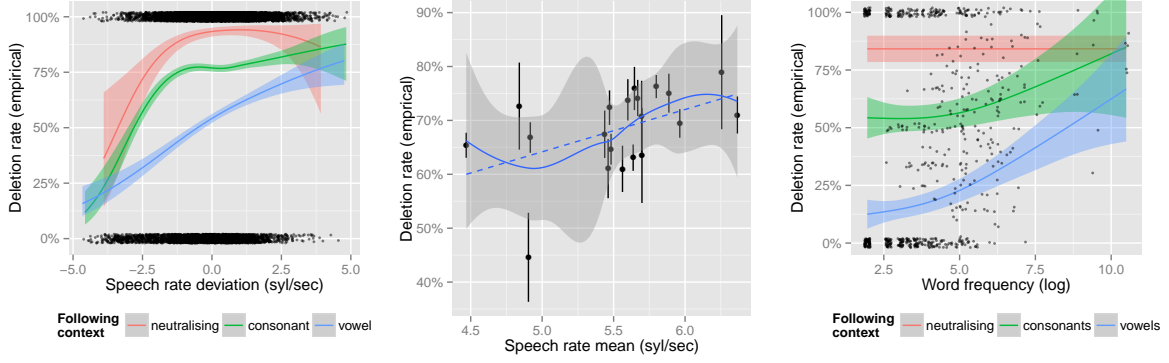


Figure 2: Empirical plots of deletion rate as a function of speech rate and word frequency. Left: speech rate deviation versus deletion rates, for each following phonological context (one dot per token, as in Fig. 1). Middle: mean speech rate versus deletion rate (one point per speaker, errorbars are 95% CIs on proportions). Right: word frequency versus deletion rate, for each phonological context (one point per word/context pair). Solid lines and shading are GAM nonparametric smoother and 95% CIs for left panel (logit link) and right panel (linear link); for middle panel, nonparametric smoother is LOESS with 95% CIs, and dotted line is a linear smooth.

with deletion rate.

SPEECH RATE DEVIATION is positively correlated with deletion, across tokens ($\rho = 0.124$): Figure 2 (left) shows that deletion rate generally increases for greater speech rates, as expected, but the slope becomes less pronounced at higher speech rates. In order to allow for the observed nonlinear effect, SPEECH RATE DEVIATION was coded as a restricted cubic spline with three knots (= 2 components), which intuitively allows the fitted curve to have one ‘bend’, based on visual inspection (Figure 2 left).¹² The effect of speech rate deviation also appears to differ by following context, in such a way that the effect of following context is modulated by speech rate: in very slow speech (within a given speaker), deletion rate differs minimally between following contexts. As speech rate increases, the difference in deletion rate between contexts rapidly increases; around -1.25 syllables/second, this trend reverses, and deletion rates in different contexts gradually become more similar for ‘normal’ and ‘fast’ speech rates (deviation ≥ 0), showing a possible ceiling effect.

MEAN SPEECH RATE is positively correlated with mean deletion rate, across the 20 speakers ($\rho = 0.52$), with a roughly linear relationship — as in Figure 2 (middle), the line of best fit lies within the confidence intervals of the nonparametric smoother. We thus code MEAN SPEECH RATE as a single continuous variable. The effect of MEAN SPEECH RATE does not appear to strongly depend on following context, based on exploratory plots (not shown).

3.2.4 Frequency, Bigram Frequency, Conditional Probability

Frequency FREQUENCY was measured as log-transformed corpus frequency (tokens per million) of the t/d-final word in the Big Brother corpus. A token count of 1 in the corpus corresponds to 7.05 per million, and a log-transformed frequency of 1.953. Frequency is positively correlated with deletion rate: t/d is deleted more often for higher-frequency words ($\rho = 0.31$). Figure 2 (right) illustrates that the effect of frequency on deletion rate appears to differ by following context: the effect is stronger (higher slope) before vowels ($\rho = 0.48$) than before (non-neutralizing) consonants ($\rho = 0.26$), and is near-absent in neutralizing context ($\rho = 0.030$). As a result, the following context effect is heavily modulated by word frequency: deletion rate differs maximally between contexts for very infrequent words, and deletion rates in different contexts are progressively more similar for more frequent words. The relationship between frequency and deletion rate within a given context is approximately linear;

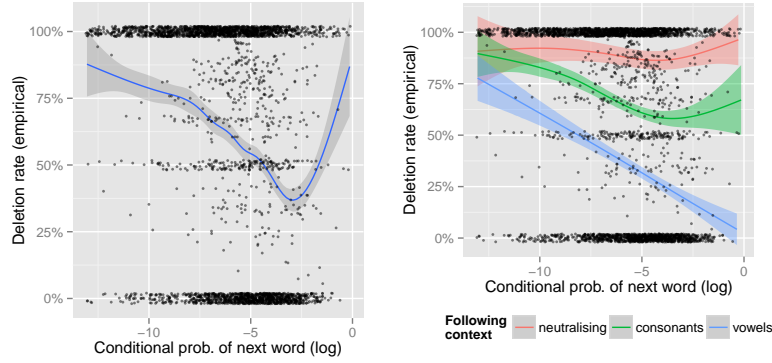


Figure 3: Empirical plot of deletion rate as a function of conditional probability across the whole dataset (left), and by following phonological context (right). One point per bigram. Solid lines and shading are GAM nonparametric smoothers and 95% CIs.

we thus code FREQUENCY as a single continuous variable.¹³

Conditional Probability To estimate the conditional probability of the following word given the previous word for observations in the dataset, we fitted a trigram language model using the `lmp1z` function in the KenLM language model toolkit (Heafield et al., 2013), which estimates the conditional probability of n -grams using Kneser-Ney smoothing without pruning. The language model was fitted to the spoken portion of the British National Corpus (BNC, ~ 10 million words; British National Corpus, 2007), and hence estimates the cooccurrence probabilities of words in spoken British English, matching the nature of our spoken British English corpus. We did not use the Big Brother dataset itself to calculate conditional probabilities because it was deemed too small to give reliable estimates (0.14 million words). (In particular, most bigrams in the CSD dataset analyzed here only occur once.) The language model was used to assign a conditional probability of the following word given the t/d -final target word, for each data point for which the corresponding bigram (combination of target word and next word) occurred at least once in the BNC. The 1804 tokens corresponding to bigrams which did not occur in the BNC were excluded from the dataset.

In empirical plots (Fig. 3 left), conditional probability shows a nonlinear relationship with deletion rate across the dataset. The effect of conditional probability differs by following context (Fig. 3 right), and this interaction shows a clear pattern: deletion rate decreases for words in all following contexts for low to mid conditional probabilities, with progressively greater slope before neutralizing consonants, other consonants, and vowels. The relationship between conditional probability and deletion rate becomes flat for higher-frequency words preceding a consonant, but continues to be negative for higher-frequency words preceding a vowel. The overall effect is that the conditional probability effect is heavily modulated by following context: deletion rate differs minimally between contexts for bigrams with a low conditional probability, and deletion rates in different contexts are progressively more different for bigrams with higher conditional probability. The relationship between conditional probability and deletion rate for a given following context is roughly linear; we thus code CONDITIONAL PROBABILITY as a single continuous variable, after log-transforming it to bring its distribution closer to normality.¹⁴

3.2.5 Other variables

MORPHOLOGICAL CLASS was coded using two levels: past-tense forms (past) and all other words (non-pasts). More deletion is observed in non-past-tense forms (type: 44.1%, token: 72.6%) than in

past-tense words (type = 38.3%, token = 43.2%), as expected from previous CSD studies. This variable is important to include in our analysis as a control, because it is collinear with PRECEDING CONTEXT and FREQUENCY, which are crucial to our research questions: non-past forms disproportionately have preceding segments which favor deletion and tend to have higher frequencies, compared to past tense forms (e.g. Temple, 2013).

A number of other properties of the word are found to affect t/d rate in previous work, including final consonant identity (t vs. d), whether the preceding consonant has the same voicing status as the final t/d, and stress of the word-final syllable (e.g., Hazen, 2011). In contrast to MORPHOLOGICAL CLASS, these effects are not directly related to our research questions; they are also typically weak. Accordingly, we account for them, as well as any other idiosyncratic differences between words which affect deletion rate, by including a by-word random intercept in the statistical model.

Properties of speakers may also affect CSD rate. Although CSD rate is generally found to exhibit only weak effects of ‘social factors’ (speaker gender, social class, etc.), different English dialects show markedly different deletion rates. This may underly some of the differences between speakers in the Big Brother dataset, which are visible in the vertical spread of points in Figure 2 (middle): these speakers come from a range of dialect regions. We account for any properties of speakers which affect deletion rate (beyond SPEECH RATE MEAN), as well as idiosyncratic differences between speakers, by including a by-speaker random intercept in the statistical model.

4 Analysis

4.1 Model structure

A mixed-effects logistic regression of coronal stop realisation (deleted vs. realised) as a function of the predictors described above was fit to the data using the `lme4` package in *R* (Bates et al., 2014). In order to address our research questions, the model contained fixed effects for PAUSE CLASS, FOLLOWING CONTEXT, PRECEDING CONTEXT, SPEECH RATE MEAN, SPEECH RATE DEVIATION, word FREQUENCY, and CONDITIONAL PROBABILITY; as well as interaction terms (for the possible modulation of phonological context effects by factors associated with planning): PAUSE CLASS : FOLLOWING CONTEXT, PAUSE CLASS : PRECEDING CONTEXT, FREQUENCY : FOLLOWING CONTEXT, CONDITIONAL PROBABILITY : FOLLOWING CONTEXT, and SPEECH RATE DEVIATION : FOLLOWING CONTEXT. Fixed-effect terms for MORPHOLOGICAL CLASS and ANNOTATOR identity (4 levels) were also included, as controls. Continuous predictors (SPEECH RATE MEAN/DEVIATION, FREQUENCY, CONDITIONAL PROBABILITY) were centered and divided by two standard deviations. (For SPEECH RATE DEVIATION, this standardization was done within tokens for a given speaker, to control for interspeaker differences in the *range* of speech rate.) MORPHOLOGICAL CLASS was converted into a numeric predictor with range 1, then centered. Discrete predictors with multiple levels (FOLLOWING CONTEXT, PRECEDING CONTEXT, ANNOTATOR) were coded with helmert contrasts; e.g., for FOLLOWING CONTEXT, the interpretations of the two contrasts are (1) neutralising segments vs. other consonants (2) all consonants (neutralizing and non-neutralizing) vs. vowels, and for PRECEDING CONTEXT, the interpretations of the two contrasts are (1) sibilants vs. sonorants (2) {sibilants and sonorants} vs. obstruents.¹⁵ PAUSE CLASS (4 levels) was coded using (three) orthogonal polynomial contrasts, corresponding to linear, quadratic, and cubic trends.

The model was fit with the following random effects structure: (1) by-speaker, by-word, and by-bigram intercepts, to control for differences in CSD rate beyond the effects of predictors included in the model, and obtain accurate estimates of fixed-effect coefficients for properties of speakers, words, and bigrams (i.e., CONDITIONAL PROBABILITY); (2) all possible by-speaker random slopes for all terms re-

lated to our research questions, to control for variability in these effects across speakers: PAUSE CLASS, FOLLOWING CONTEXT interactions with {PAUSE CLASS, SPEECH RATE DEVIATION, FREQUENCY, CONDITIONAL PROBABILITY}, PRECEDING CONTEXT interaction with PAUSE CLASS (as well as all subset terms, e.g. main effect of FREQUENCY); (3) by-word random slopes, to do the same for variability across words, for PAUSE CLASS only. (Models with further by-word random slope terms did not converge, presumably due to there being too little data per word: 1–2 observations for many word types.) Correlations between random-effect terms were not included, to avoid model over-parametrisation and aid convergence. The random-effect structure in the model is as close as possible to ‘maximal’, in the sense of Barr et al. (2013): as many random-slope terms as possible are included, such that the model still converges, with terms prioritized which guard against Type I error in the fixed-effect coefficients related to our research questions. (For example, by-speaker random slopes for interactions of PAUSE CLASS with FOLLOWING CONTEXT help guard against significant effects driven by particular speakers, which would spuriously support the predictions of the PPH.)

4.2 Results

The fixed-effects coefficients of the model are summarized in Table 1, with significances based on Wald tests. We do not discuss the random-effect terms, which can be seen in Table 2. The coding system used for the predictors means that the main effect coefficients can be interpreted at an ‘average value’ of other predictors, for an ‘average’ speaker and word. For example, the PRECEDING CONTEXT 2 coefficient’s interpretation is “the difference in deletion rate between non-sibilant obstruents and other consonants, at average word frequency and speech rate, averaged across all pause classes, following segment types, morphological classes, and annotators”. The intercept term thus predicts the ‘average’ rate of deletion to be 19.5% (inverse logit of -1.415).

We also assess the importance of each term in the regression related to our research questions (PAUSE, PAUSE:FOLLOWING CONTEXT, etc.: rows of Table 3), in a second way. For each such term, we conduct a likelihood ratio test between the full model and a model with all fixed and random effects for this term dropped; χ^2 and the corresponding significance for each test is reported in Table 3. These χ^2 tests assess whether adding information about a variable improves the model, taking into account *both* fixed and random effects and assessing all corresponding regression terms at once, thus giving complementary information to the Wald test results (Table 1), which only assess individual fixed effects.

We first discuss each of the effects relevant to our research questions, then briefly summarize other effects. Because the effects of primary interest are difficult to interpret from the model table (due to nonlinear terms and multi-level factors), we use the partial-effect plots in Figures 4–5 to visualize the model’s predictions. These plots show predictions as a subset of predictors are varied, with other continuous predictors held at 0 (average value) and other discrete predictors held constant (at PRECEDING CONTEXT = *sonorant*, FOLLOWING CONTEXT = *consonant*, PAUSE CLASS = *none*). Predictions (solid lines, dots) and 95% confidence intervals (vertical lines, shading) are computed based on fixed effects alone (without taking random-effect variances into account), and can be thought of as the prediction for an ‘average’ word and speaker.

4.2.1 Pause

Pause duration has a strong negative effect on deletion rate ($\chi^2(9) = 147.5, p < 0.0001$), : for the main effect of PAUSE CLASS, the linear trend is negative ($\hat{\beta} = -2.086$), highly significant ($p < 0.001$), and has a much greater effect size than the quadratic or cubic trends (which do not reach significance). The

Table 1: Summary of fixed-effect coefficients in the logistic regression model of coronal stop deletion: coefficient estimates, standard errors, z , and corresponding p -value (Wald test). Coefficients are in log-odds.

| Predictor | $\hat{\beta}$ | SE($\hat{\beta}$) | z -value | Pr(> z) |
|--|---------------|---------------------|------------|--------------|
| Intercept | -1.415 | 0.288 | -4.917 | <0.001 |
| Annotator 1 | 0.373 | 0.048 | 7.847 | <0.001 |
| Annotator 2 | 0.015 | 0.023 | 0.643 | 0.52 |
| Annotator 3 | -0.071 | 0.036 | -2.002 | 0.045 |
| Morphological Class | -0.037 | 0.179 | -0.204 | 0.839 |
| Speaking Rate (mean) | 0.759 | 0.259 | 2.932 | 0.003 |
| Pause Class (linear) | -2.086 | 0.296 | -7.045 | <0.001 |
| Pause Class (quadratic) | 0.283 | 0.307 | 0.922 | 0.356 |
| Pause Class (cubic) | -0.356 | 0.269 | -1.324 | 0.185 |
| Preceding Context 1 (sibilants vs sonorants) | -0.179 | 0.129 | -1.383 | 0.167 |
| Preceding Context 2 ((sib/son) vs obstruents) | -0.350 | 0.137 | -2.554 | 0.011 |
| Following Context 1 (neutralising vs consonants) | -0.774 | 0.302 | -2.560 | 0.01 |
| Following Context 2 ((neut/cons) vs vowels) | -0.845 | 0.130 | -6.522 | <0.001 |
| Conditional probability (log, standardized) | 0.331 | 0.246 | 1.345 | 0.179 |
| Word Frequency (log, standardized) | 0.523 | 0.162 | 3.228 | 0.001 |
| Speaking Rate Deviation 1 | 0.198 | 0.044 | -4.479 | <0.001 |
| Speaking Rate Deviation 2 | -0.937 | 0.165 | -5.679 | <0.001 |
| Pause Class (linear) : Preceding Context 1 | 0.197 | 0.122 | 1.615 | 0.106 |
| Pause Class (quadratic) : Preceding Context 1 | -0.288 | 0.196 | -1.471 | 0.141 |
| Pause Class (cubic) : Preceding Context 1 | -0.028 | 0.272 | -0.103 | 0.918 |
| Pause Class (linear) : Preceding Context 2 | -0.245 | 0.259 | -0.947 | 0.344 |
| Pause Class (quadratic) : Preceding Context 2 | -0.579 | 0.245 | -2.358 | 0.018 |
| Pause Class (cubic) : Preceding Context 2 | -0.034 | 0.242 | -0.139 | 0.889 |
| Pause Class (linear) : Following Context 1 | 0.630 | 0.242 | 2.603 | 0.009 |
| Pause Class (quadratic) : Following Context 1 | -0.335 | 0.248 | -1.350 | 0.177 |
| Pause Class (cubic) : Following Context 1 | -0.046 | 0.254 | -0.180 | 0.857 |
| Pause Class (linear) : Following Context 2 | 0.561 | 0.103 | 5.428 | <0.001 |
| Pause Class quadratic) : Following Context 2 | -0.156 | 0.100 | -1.565 | 0.118 |
| Pause Class (cubic) : Following Context 2 | 0.053 | 0.106 | 0.498 | 0.618 |
| Conditional probability : Following Context 1 | -0.354 | 0.342 | -1.037 | 0.3 |
| Conditional probability : Following Context 2 | 0.021 | 0.165 | 0.127 | 0.899 |
| Word Frequency : Following Context 1 | 0.225 | 0.128 | 1.753 | 0.08 |
| Word Frequency : Following Context 2 | 0.208 | 0.055 | 3.792 | <0.001 |
| Speaking Rate Deviation 1 : Following Context 1 | 0.147 | 0.059 | -2.491 | 0.013 |
| Speaking Rate Deviation 1 : Following Context 2 | 0.103 | 0.023 | -4.553 | <0.001 |
| Speaking Rate Deviation 2 : Following Context 1 | 0.371 | 0.209 | 1.778 | 0.075 |
| Speaking Rate Deviation 2 : Following Context 2 | 0.071 | 0.086 | 0.832 | 0.405 |

interpretation of this effect is that longer pauses gradiently decrease deletion rate, in general (averaging across phonological contexts). Figure 4 illustrates that this interpretation generally holds within each different phonological context as well: deletion rate gradually decreases for progressively longer pauses. (A possible exception is before vowels, where there may be a floor effect (deletion rate is always very low).) This result is in line with previous work showing that binary ‘pause’, coded as a following environment, affects CSD rate; it is novel in showing a *gradient* effect of pause duration, *independent* of following context, as discussed further below.

4.2.2 Phonological Context

We first consider the main effect terms for phonological context. Confirming the empirical observations and previous CSD research, following phonological context was a highly significant indicator of deletion (FOLLOWING CONTEXT 1: $\hat{\beta} = -0.774$, $p = 0.01$; FOLLOWING CONTEXT 2: $\hat{\beta} = -0.845$, $p < 0.001$). At average pause duration, vowels induce less deletion than consonants, with neutralising segments

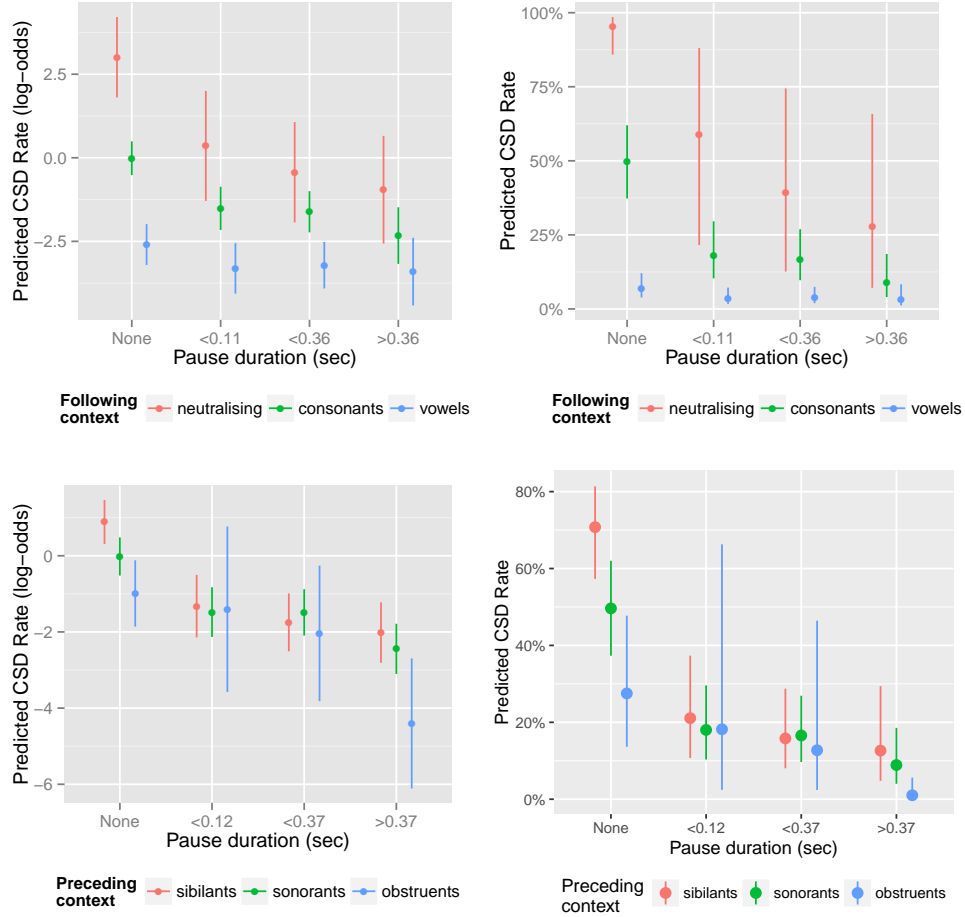


Figure 4: Partial effect plots for phonological context and pause duration: predicted deletion rate by PAUSE CLASS and PRECEDING CONTEXT (bottom row) or FOLLOWING CONTEXT (top row), in log-odds (left column) and probability (right column) space. Dots and errorbars indicate predictions and 95% CIs with other predictors held constant.

inducing much higher rates. This is expected, given the near-categorical rate of deletion in the dataset for neutralising environments.

The effect of preceding context also follows the expected pattern based on the empirical data and previous work: sibilants, sonorants, and non-sibilant obstruents induce progressively less deletion, at average pause duration. However, the effect size is notably smaller than that of following context and is only weakly significant (PRECEDING CONTEXT 1: $\hat{\beta} = -0.179$, $p = 0.167$; PRECEDING CONTEXT 2: $\hat{\beta} = -0.350$, $p = 0.011$), illustrating that preceding context has less influence on deletion rates than following context, as expected (Schreier, 2005).

Of primary interest for our research questions is how pause duration modulates the effect of phonological context, corresponding to the FOLLOWING CONTEXT:PAUSE CLASS and PRECEDING CONTEXT:PAUSE CLASS interactions. The conditioning effect of PAUSE CLASS on FOLLOWING CONTEXT is highly significant ($\chi^2(12) = 103.1$, $p < 0.0001$) due to the terms corresponding to the linear trend of PAUSE CLASS (interaction with: FOLLOWING CONTEXT 1: $\hat{\beta} = 0.630$, $p = 0.009$; FOLLOWING CONTEXT 2: $\hat{\beta} = 0.561$, $p < 0.001$). The difference in deletion rate between different following contexts gradiently decreases as pause duration increases, as shown in Figure 4 (top row). For each pause class, the effect of following context shows the predicted ordering (vowels < consonants < neutralizing), but the *size* of this effect greatly decreases. This modulation is predicted by the PPH, since the planning window should be less

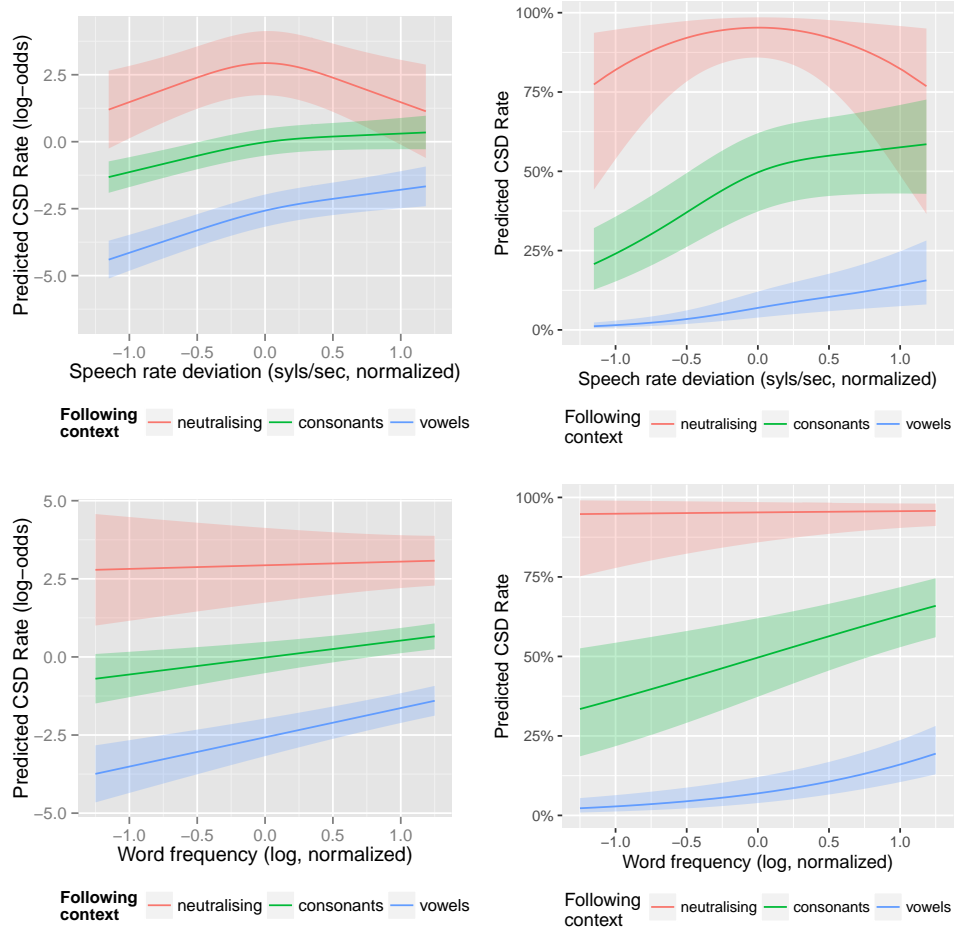


Figure 5: Partial effect plots for speech rate and word frequency: predicted deletion rate by SPEECH RATE DEVIATION (top row) and FREQUENCY (bottom row) for different following phonological contexts, in log-odds (left column) and probability (right column) space. Lines and shading as in Figure 4.

likely to contain the following segment at stronger boundaries.

The interaction of PAUSE CLASS and PRECEDING CONTEXT is marginal at the $p = 0.05$ level ($\chi^2(12) = 21.0$, $p = 0.051$). The interpretation of this effect is less straightforward. Based on Figure 4 (bottom row), we see that preceding context has a small effect on deletion rate when there is no pause, in the expected direction, but does not significantly affect deletion rate when there is a pause (of any duration)—corresponding to overlapping confidence intervals for PAUSE CLASS != none. At face value, this effect is inconsistent with the prediction of the PPH that preceding context should not be conditioned by pause duration (as preceding context should always fall into the same local planning window). However, it is also notable that the effect does *not* show gradient modulation of the preceding context effect as a function of pause duration (as for following context), suggesting that factors besides production planning may be at play, as discussed further below. We note that an account of CSD in terms of gestural overlap, as in Articulatory Phonology, might predict a modulation of the effect of the preceding environment—as segments ‘stretch out’ more due to slow down at stronger boundaries, the effect of the preceding segment might decrease. The observed trends might be an indication of such a gestural effect.

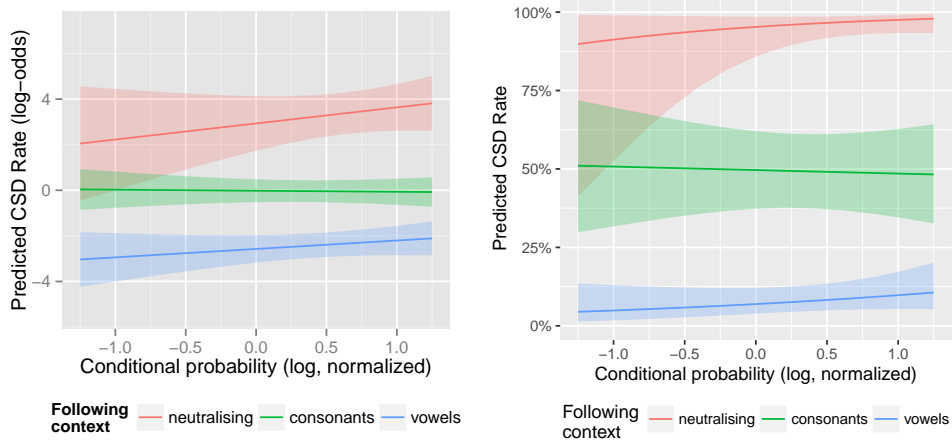


Figure 6: Partial effect plots for conditional probability: predicted deletion rate by CONDITIONAL PROBABILITY for different following phonological contexts, in log-odds (left) and probability (right) space. Lines and shading as in Figure 4.

4.2.3 Speaking Rate

Both measures of speech rate have significant effects on CSD rate. Speakers who speak faster on average delete more frequently (SPEECH RATE MEAN: $\hat{\beta} = 0.759$, $p = 0.003$). In addition, faster speech (within a speaker) generally leads to more deletion (SPEECH RATE DEVIATION 1, 2: $\hat{\beta} = 0.198$, -0.937 ; $p < 0.001$, $p < 0.001$), although the effect differs significantly depending on following context ($\chi^2(8) = 24.9$, $p = 0.0016$; also the SPEECH RATE DEVIATION:FOLLOWING CONTEXT rows of Table 1). Figure 5 (top row) shows that faster speech leads to more deletion up to about a speaker’s average rate (SPEECH RATE DEVIATION = 0), after which the effect is less pronounced for following consonants and vowels, and reverses direction in neutralizing context. How speech rate modulates the effect of following context depends on whether we interpret the model’s predictions in log-odds or probability space. In log-odds space, the difference in deletion rate between contexts seems to decrease at high speech rate. In probability space, the difference between contexts decreases for progressively lower speech rates, as is the case in the empirical data (Figure 5). Regardless of the interpretation of this effect, there are strong speech rate effects on CSD rate, and speech rate deviation modulates the effect of following context.

4.2.4 Frequency

Word frequency has a significant positive effect on deletion rate: higher frequency words induce higher rates of deletion (FREQUENCY: $\hat{\beta} = 0.523$, $p = 0.001$), averaging across phonological contexts. This finding replicates earlier results (e.g. Jurafsky et al., 2001; Coetzee, 2009). Due to the inclusion of a by-word random intercept, the significant frequency effect cannot be attributed to the influence of particular lexical items (e.g. *and*, *just*). This result contrasts with claims that frequency does not affect CSD once other factors and are controlled for (Walker, 2012).

Of interest for our research questions is the significant interaction of FREQUENCY with FOLLOWING CONTEXT ($\chi^2(4) = 15.9$, $p = 0.0032$). The rows of Table 1 corresponding to this effect give its interpretation: as word frequency increases, the effect size of following context reduces, resulting in following context conditioning CSD to a lesser degree (Fig. 5, bottom row). This modulation of the following context effect by word frequency mirrors the pattern seen in the empirical data (Figure 2), and has a natural explanation in terms of production planning, under the assumption that higher-frequency words are more likely than lower-frequency to be planned before the following phonological context is available.

Coetzee (2009) found no such interaction in an experimental test of which factors affect intuitions about the likelihood of t/d deletion.

4.2.5 Conditional Probability

Conditional probability does not significantly affect deletion rate (CONDITIONAL PROBABILITY: $p = 0.179$), averaging across phonological contexts. Of primary interest for our research questions is the *interaction* of CONDITIONAL PROBABILITY with FOLLOWING CONTEXT, which the likelihood ratio test suggests significantly contributes to model likelihood ($\chi^2(4) = 12.7$, $p = 0.012$). However, the corresponding fixed effect coefficients are not significantly different from zero (CONDITIONAL PROBABILITY:FOLLOWING CONTEXT 1: $\hat{\beta} = -0.354$, $p = 0.3$; CONDITIONAL PROBABILITY:FOLLOWING CONTEXT 2: $\hat{\beta} = 0.021$, $p = 0.899$). We interpret this discrepancy in the two ways of assessing the interaction as follows. First, note that the effect size of the CONDITIONAL PROBABILITY:FOLLOWING CONTEXT 1 coefficient is actually relatively large—in particular, larger than the coefficients for the FREQUENCY interaction with FOLLOWING CONTEXT—and in the direction predicted by the empirical plots (Figure 3), where we see a clear pattern suggesting that the influence of the following phonological context increases with the conditional probability of the following word. The interpretation of this coefficient (see Figure 6) is that the difference in deletion rate between a following t/d and other consonants increases as conditional probability increases, as predicted in terms of production planning. (The coefficient for CONDITIONAL PROBABILITY:FOLLOWING CONTEXT 2 has a very small effect size, and we interpret it as effectively zero.) The ‘reason’ that the CONDITIONAL PROBABILITY:FOLLOWING CONTEXT 1 is not significant, despite its large effect size, is its high standard error. This standard error, in turn, is likely large due to the high degree of variability between participants: the size of this variability (0.427, 0.318: the standard deviation for the by-speaker random slopes for FOLLOWING CONTEXT:CONDITIONAL PROBABILITY terms in Table 2) is comparable to the effect size of the CONDITIONAL PROBABILITY:FOLLOWING CONTEXT 1 term. Intuitively, this means that the model cannot reliably detect a group-level effect given the degree of variability across speakers. This high inter-speaker variability also explains why the χ^2 test is significant: it is the random effect terms that make CONDITIONAL PROBABILITY:FOLLOWING CONTEXT significantly contribute to model likelihood. We do not have an explanation for the high degree of inter-speaker variability, but note that future work could better test for the CONDITIONAL PROBABILITY:FOLLOWING CONTEXT in a corpus with more speakers (rather than just 20) and more data per speaker (the data in the Big Brother corpus is very unbalanced), to better estimate differences between speakers. We conclude that the high effect size, direction, and significance (using the likelihood ratio test) of the CONDITIONAL PROBABILITY:FOLLOWING CONTEXT effect provide some tentative support for our hypothesis about how conditional probability modulates the following context effect, but that this interaction merits more investigation in future work.

4.2.6 Other effects

The effect of MORPHOLOGICAL CLASS is in the expected direction (non-past-tense forms > past tenses: MORPHOLOGICAL CLASS $\hat{\beta} = -0.037$), but is not significant ($p = 0.839$). This finding follows similar observations for British English (Tagliamonte and Temple, 2005) and Appalachian English (Hazen, 2011), where morphological class did not significantly condition the rate of deletion. The effect of ANNOTATOR is significant (rows 2–4 in Table 1), especially the effect of ANNOTATOR 1 ($\hat{\beta} = 0.373$, $p < 0.001$), indicating that one annotator marked tokens as deleted less frequently than others. Although this effect is unwelcome (ideally, annotators would be statistically indistinguishable from each other), we hope to have controlled for the first annotator’s different behavior by including the ANNOTATOR term in the model.¹⁶

5 Discussion

This study has examined coronal stop deletion in a corpus of spontaneous British English. We discuss our results in reference to our three research questions, which also bear on the broader motivations for this study, repeated here: (1) How does duration of a pause following the coronal stop (which serves as a proxy for boundary strength) affect deletion rate? (2) How does boundary strength (i.e. pause duration) modulate the effects of surrounding segments on deletion rate? (3) How do other factors influencing the size of the planning window in production planning, such as measures of word predictability and speech rate, modulate the effects of the following phonological context on deletion rate?

Our results also have practical implications for studying coronal stop deletion and other segmental processes which can apply across word boundaries, and make predictions which can be tested in future work.

5.1 Boundary strength

Our first research question was how boundary strength, here approximated by duration of the pause following the t/d, affects deletion rate. We found that pause duration has a strong negative effect on deletion rate, across phonological contexts. Crucially, this effect is *gradient* and *independent* of following phonological context.

With PAUSE CLASS coded as a factor independently from following phonological context, larger PAUSE CLASS in the statistical model steadily decreases deletion rate, corresponding to the clearly gradient trend in the empirical data (Fig 1). If pause duration is interpreted as a proxy for boundary strength, this result is unsurprising, given that gradient effects of boundary strength on segmental realization at prosodic boundaries are common cross-linguistically (e.g. Byrd and Saltzman, 1998; Cho and Keating, 2001; Fougeron and Keating, 1997; Fougeron, 2001).

Treating pause duration as gradient and independent of following context contrasts with previous work on CSD, where ‘pause’ (coded using perceptual or acoustic criteria) is a binary variable coded as one possible following context (e.g. as an alternative to a consonant or a vowel). Different studies have found inconsistent effects of a pause coded in this way, with some finding that pauses pattern more like consonants (higher deletion rate; African American English in Fasold, 1972; Wolfram et al., 2000), or more like vowels (lower deletion rate; Philadelphia speakers in Guy, 1980a), or induce the least deletion of any following context (Tagliamonte and Temple, 2005). These different effects are generally attributed to dialectal differences, which is unusual, given that most factors influencing CSD rate (such as preceding context and morphological class) have been found to have strikingly similar qualitative effects across many dialects (Schreier, 2005).

Our results suggest an alternative possibility: different studies may have found different effects of ‘pause’ by discretizing the gradient effect of pause duration in different ways, or because of different correlations of the presence of a pause with the identity of the following segment in the datasets used in different studies. Previous studies of CSD have not interpreted ‘pause’ as raising or lowering deletion rate, per se; we found that longer pauses markedly decrease deletion rate, when considered independently of segmental context. This result suggests that pause duration (and any other correlates with boundary strength) should be treated as independent of the following segment, and ideally as a gradient variable, in future work.¹⁷ These methodological changes may help to clarify the interplay between segmental and prosodic factors in conditioning deletion rate, in line with Kendall’s (2013) suggestion that variable processes can be better understood by a more detailed consideration of the role of prosodic information (pauses and speech rate). The suggested methodological change applies more generally to

sociolinguistic, phonetic, and phonological studies of any variable process that can take place across word boundaries, such as final [t]-deletion in German and Dutch (closely related to English CSD) or Spanish /s/-lenition; in these literatures, ‘pause’ (as a proxy for boundary strength) is often treated as a possible following context (e.g. File-Muriel and Brown, 2011; Schuppler et al., 2012).¹⁸

In the phonological variation literature on CSD in particular, ‘pause’ is typically treated as a following environment independent of following context (e.g. Coetzee and Kawahara, 2013; Coetzee, 2004), and this assumption informs the structure of the phonological grammar which is postulated to account for the data: formal mechanisms (in previous work, Optimality Theoretic constraints) penalize deletion before consonants, vowels, and pauses to different degrees. For the facts reported in this paper to be accommodated in a grammatical account, a different kind of grammar would need to be developed to account for the independent and gradient effect of pause on deletion rate, and how pause duration modulates the following context effect, and it would have to be powerful enough to accommodate interactions of phonological context with external factors such as frequency and conditional probability. While Coetzee and Kawahara (2013) discuss how to accommodate a general effect of frequency on CSD, their model does not allow for interactions of frequency and following phonological environment. A more detailed discussion of the formal options is beyond the scope of this paper.

5.2 Modulation of contextual effects and the production planning hypothesis

5.2.1 Phonological context

Our second research question was how boundary strength modulates the effect of surrounding segments. We found that the length of the pause between the t/d and the following word strongly and gradiently modulates the effect of the first segment of the following word: for longer pauses, there is progressively less difference in CSD rate between words beginning with t/d, another consonant, or a vowel; for pauses of above 362 msec, there is almost no effect of following context. This is particularly striking for words in ‘neutralizing’ context: without a pause, deletion is near categorical (hence the exclusion of these tokens from most studies of CSD); when there is even a short pause, deletion rate drops to the rate expected before any other segment. This effect supports the PPH, which predicts that due to the locality of production planning, the availability of upcoming information (i.e., the following context) will be conditioned by boundary strength, and hence should have *no* effect after sufficiently strong boundaries.

The following context effect is in a sense obvious—for a long enough pause, following material must be invisible. However, it is not obvious a priori what ‘long enough’ means, and the fact that modulation occurs even for very short pauses is consistent with the very local nature of production planning. Following phonological context often has a large effect on whether variable segmental processes apply (in sociolinguistic and phonetic studies). The prediction based on our results, if the PPH is right, is that for any variable segmental process that can take place across a word boundary, prosodic boundary strength should modulate the following context effect, in the same direction as observed here, given the locality of production planning.

This modulation of the context effect by the duration of pauses is compatible with the predictions of the production planning hypothesis, which predicts that due to the locality of production planning, the availability of upcoming information (i.e., the following context) will be conditioned by boundary strength, and hence should have *no* effect after sufficiently strong boundaries. It also receives a natural explanation, however, under the view that the greater ‘slow down’ at stronger prosodic boundary will result in less gestural overlap. Under the Articulatory Phonology analysis of t/d deletion, this predicts a lower rate of cases in which it will appear that t/d deletion has taken place. This result is therefore

compatible with other interpretations.

The PPH does not predict a similar modulation of the *preceding* context effect by boundary strength (decreasing effect as boundary strength increases), since the identity of the preceding segment will always be available when the t/d is planned. Importantly, it does not predict that there will be *no* interaction between preceding context and boundary strength, which could be due to sources besides production planning. We already discussed before that such effects could be due to decreased gestural overlap with increased lengthening preceding prosodic boundaries, and predicted by the account of t/d deletion in Articulatory Phonology. Indeed, after discretizing pause duration into four classes, we found a weak but significant interaction: preceding context has the expected effect when there is no pause, but does not show an effect before pauses of any length. This interaction does not look like the following context interaction—the effect of preceding context does not gradually decrease for stronger boundaries—but why does this occur? To some extent, it may be an artefact of how we discretized pause duration: in the empirical data (Fig 1), where all pause lengths are considered, it is clear that preceding context does affect deletion rate for arbitrarily long pauses (in contrast to the following context effect). There may also simply not be enough data before pauses (15% of the data; $n=1660$) to resolve the effect size of preceding context (consistently found to be small in previous studies), or there may be a psycholinguistically-motivated explanation unrelated to production planning, such as listeners making less use of the acoustic cues associated with preceding context (e.g. formant transitions) in environments where a stop burst is likely (before a pause) (e.g. Steriade, 2009). Regardless of its source, we argue that the observed weak modulation of the preceding context effect by pause duration does not offer evidence against the production planning hypothesis.

5.2.2 Word frequency and speech rate

Our third research question was how factors beyond boundary strength modulate the effect of following context on deletion rate. We consider word frequency and speech rate here, and turn to conditional probability below. These three factors should affect the probability that the following segment is ‘available’ when the articulation of the final t/d is planned; thus, the PPH predicts that they should modulate the effect of following context on deletion rate.

Frequency significantly affects deletion rate across phonological contexts, with t/d more likely to delete in higher-frequency words. As noted above, this finding contrasts with some previous CSD studies (Walker, 2012), but is unsurprising if CSD is viewed as a case of segmental reduction, given that word frequency (or predictability) is often positively correlated with reduction probability in such processes cross-linguistically (e.g. Jurafsky et al., 2001; Schuppler et al., 2012; Bell et al., 2009; Zipf, 1929; Ernestus et al., 2006). What is important for our purposes is not the existence of an overall frequency effect, but the fact that it significantly modulates the effect of following context: the higher the frequency of the t/d-final word, the less its deletion rate depends on the identity of the following segment. This effect is expected under the PPH if we assume that higher-frequency words are planned earlier relative to the phonological retrieval of a following word.¹⁹ Based on our discussion in the introduction, the direction of the effect is as expected if frequency effects operate at the level of phonological retrieval, and leaves the relative timing of the lemmas unaffected. In this case, the phonological form of the first word with a higher frequency will be planned earlier relative to the phonological retrieval of the second word. As noted above, it remains controversial whether frequency effects arise only at the level of phonological retrieval, and different assumptions about frequency effects might lead to a different prediction about how word frequency modulates the following context effect. Our finding provides additional motivation for a psycholinguistic study probing the availability of the phonological content of a following word, as a function of the frequency of the first word; to our knowledge, such a study

has not been done.

The PPH also predicts a modulation of the effect of following context by speech rate. If increased speech correlates with a wider planning window, which we argued is plausible given prior results, we should find an increase of the probability that the following segment identity is available when the t/d is planned in faster speech, and we would expect the opposite effect in slower speech. These predictions are not clearly borne out in our data (Fig. 5, top row). Although speech rate significantly modulates the following context effect, the direction of this modulation (larger vs. smaller difference in deletion rate between following contexts) differs for lower and higher speech rates, and depending on whether we think in terms of log-odds or in probabilities. The effect can be roughly described as capturing the pattern in the empirical data (Fig. 2 mid): different contexts have maximally different deletion probabilities around average speech rate, and progressively more similar deletion probabilities as speech rate is either increased or decreased. The pattern as speech rate is decreased (from mid to low speech rate) is compatible with the PPH. As is apparent in the plot in Fig. 2, the deletion rate at the fast end of the spectrum is very high indeed. The observed decrease of the context effect is not in line with the predictions of the PPH, and we are not sure how to interpret this at this point.

5.2.3 Conditional Probability of the Following Word

We found only limited support for the hypothesis that the following context effect is modulated by the conditional probability of the following word, possibly due to high interspeaker variability in the size of the effect. We note, however, that we might be underestimating the effect of conditional probability, which is confounded with two important variables: word frequency and syntactic constituency. As observed in Jurafsky et al. (2001), estimates of bigram frequency are not very accurate compared to estimates of word frequency, given the sparsity of even relatively frequent bigrams, even in a large corpus (such as the BNC, used here).

Conditional probability is bigram probability divided by the first word’s probability (proportional to its frequency). Since the first word’s probability is more accurately estimated, the non-random variance in the conditional probability measure may be due more to variability in the numerator than in the denominator, leading the model to attribute some of the variance actually due to conditional probability to word frequency instead. We would then expect the *word frequency* interaction with following context to show the *opposite* pattern of what is expected for conditional probability (higher word frequency \Rightarrow less following context effect)—which is exactly what we observe (Figure 5, bottom). (Indeed, Jurafsky et al., 2001 similarly attribute the lack of a conditional probability effect to it being possibly masked by word frequency.) In sum, the modulating effect we observe of the word frequency of the first word might actually be partially due to an underlying effect of the conditional probability of the second word given the first word. As discussed above, there is also a plausible explanation based on production planning for the observed directionality of word frequency’s modulating effect, without reference to conditional probability. Thus, the observed word frequency effect is compatible with either pattern.

Another limitation of our data set is that syntax is not annotated. This means that our measure of conditional probability serves double duty: It works as a proxy measure for being part of the same constituent, and at the same time it serves as a proxy measure of how likely the second word given the first once we hold syntactic constituency constant. To some extent, our bigram random effect as well as pause duration will control for this—given the correlation of pause duration with syntactic constituency—but this correlation is far from perfect (Watson and Gibson, 2004). A more richly annotated corpus would allow us to test more sophisticated hypotheses and might lead to much clearer results. We leave disentangling the modulating effects of word frequency, conditional probability, and

syntactic constituency to future work.

5.3 Production planning as an explanatory factor

The high-level goals of this study were a better understanding of the relationship between prosodic boundaries and segmental variability, and what factors determine whether particular processes are variable and the structure of this variability. We examined whether reference to production planning could address these issues, with respect to coronal stop deletion. To what extent have we found evidence for production planning as an explanatory factor, and what are the broader implications?

One effect we observed, showing how following context is modulated by pause duration, was exactly as predicted if production planning constrains whether following material can affect the application of CSD. However, this modulating effect of pauses and the lower rate of deletion at pauses can also be accounted for based on Articulatory Phonology (AP): Greater temporal compression tends to lead to greater gestural overlap and more gestural undershoot—both of which can lead the perception of CSD even if a coronal closing gesture is still present, as discussed. The lengthening or articulatory slow-down associated with pauses will mean that gestures can be realized to their full magnitude, making it less likely that a [t] or [d] will be incompletely articulated and perceived as deleted. The PPH therefore makes overlapping predictions with AP for these effects.

However, our data provides good support for a role of planning locality in explaining CSD patterns. First of all, some proportion of CSD might involve complete deletion of /t/ rather than mere coarticulation with a following word. Several studies have found that in assimilatory sandhi processes, both categorical and gradient effects may be at play (Barry, 1985, 1992; Nolan et al., 1996; Kochetov and Pouplier, 2008; Niebuhr et al., 2011), suggesting that the variability of sandhi rules may not entirely due to gradient gestural overlap or lessened gestural magnitude (see also Bermúdez-Otero, 2010, for discussion). Moreover, there is good evidence that gestural overlap itself is often planned, rather than just be a surface result of temporal compression of gestures. For example, Whalen (1990) conducted experiments in which part of what a speaker needed to say was variably only revealed when a speaker had already initiated speaking or was known to the speaker at an earlier point. Speakers failed to show gestural overlap to the same degree when they did not have the opportunity to plan coarticulation ahead of time. The results suggest that in fact coarticulation is largely planned, and not an automatic result of temporal compression. So even if most instances of apparent CSD deletion in fact do not involve full deletion, as Purse and Turk (2016) recently reported based on a corpus study, this does not mean that the PPH does not play any role in accounting for the data: Whether or not upcoming phonological material is available at the time of planning might be just as important in planning gestural overlap as it is in making categorical decisions such as deleting a coronal gesture altogether.

Coordinating speech gestures should be impossible if the precise gestures of the following word are not available yet at the time of planning. We should then be able to see the effect of production planning factors *when holding the prosodic and temporal factors constant*. In our model, we aimed to control for this by including relevant predictors such as speech rate and proxy measures for prosodic boundary strength in our model. The results suggest that the observed CSD patterns cannot be explained as an automatic result of temporal compression. As expected under the PPH, we found that word frequency of the present word and the conditional probability of the upcoming word modulate the effect of following phonological context, after controlling for prosodic boundary strength. We also tested for effects of speech rate, which did not clearly pattern as predicted if the source of these effects were in production planning. While this must lend a cautionary note to our other findings, it is important to note that even if the PPH is right, the locality of production planning is only one of many factors affecting the structure of variability in a particular case of segmental realization. This would be the case especially in

spontaneous speech, such as the data considered here, suggesting that further work in more controlled speech could better test the PPH. The PPH can also be teased apart better from alternative explanations by looking at processes that, given their phonetic substance, cannot possibly be explicable in terms of coarticulation. For a recent example, see Kilbourn-Ceron (2016), who explores the predictions of the PPH looking at liaison in French—a process where segments (and gestures) are inserted rather than deleted depending on the phonological shape of an upcoming word.²⁰

If the PPH is right, it makes interesting predictions and suggests ways that work in phonology and sociolinguistics can draw on the rich production planning literature to inform investigations of segmental variability (cf. Wagner, 2012). From the perspective of phonology, the PPH predicts that *any* phonological process that is conditioned by information across a word boundary must be variable in nature, and modulated by prosodic boundary strength. From the perspective of sociolinguistic and phonetic studies of variable segmental realization, the PPH predicts that for *any* conditioning factor (such as following phonological context) involving information that is less likely to be available (in the planning sense) when another conditioning factor (such as boundary strength) is increased, the second factor should negatively modulate the strength of the first one. While these predictions are likely too strong, we believe they suggest interesting possibilities for future work.

6 Conclusion

This paper is a first step towards understanding the relative role of production planning and other factors in explaining segmental variability. It fits into the broader goal of recent work of *explaining* the sources of variation in spontaneous speech by reference to cognitive factors about which much is independently known—such as priming, memory, how speakers use pauses, and speech perception (e.g. Kendall, 2013; Labov, 2010; Tamminga, 2014)—building on the rich literatures charting the extent of this variation as a function of linguistic, extralinguistic, and social factors.

7 Acknowledgements

This paper is an extended version of an earlier report on this project in Tanner et al. (2015). Thanks for helpful comments to Meghan Clayards, Oriana Kilbourn-Ceron, Tyler Kendall, Meredith Tamminga, Stefanie Shattuck-Hufnagel, Kristine Yu, three anonymous reviewers, and audiences at *ICPhS 18* and *ETAP 3*. This project benefitted from funding from SSHRC grant 435-2014-1504 and the SSHRC CRC program to Michael Wagner, and SSHRC grant #430-2014-00018 and FRQSC grant #183356 to Morgan Sonderegger.

Notes

¹Whether the relevant notion of ‘word’ in the definition of what counts as a sandhi process constitutes prosodic words or morphological words varies between authors, and it will not play a role in this article.

²A similar effect was found for word-medial t/d deletion (Raymond et al., 2006).

³Coetzee and Kawahara (2013, Figure 3) shows empirical plots of the effect of frequency for different following contexts (consonants, vowels, pauses) which suggest a slight difference in the slope of the frequency effect by following context. However, whether the frequency effect differs by following context is not explicitly evaluated (i.e. statistically) or discussed. Coetzee (2009) reports that native speaker intuitions about CSD rate in different phonological contexts were not affected by word frequency.

⁴In fact, some argue that it can be as big as a phrase, or even an entire transitive sentence (e.g. Oppermann et al., 2010).

⁵Note that accounts of CSD in terms of gestural overlap do predict such an interaction, since the slow down associated with stronger boundaries should lead to more time to fully realize the gestures associated with coronal stops.

⁶A third important factor is codability, that is the number of potential labels of a picture (Lachman et al., 1974; Snodgrass and Vanderwart, 1980).

⁷Since the forced aligner uses a pronunciation dictionary based on reference pronunciations, ‘number of syllables’ means number of underlying syllables rather than the number of syllables actually realized.

⁸Note that although we will discuss empirical trends in probability space, all the same qualitative observations we make based on Figures 1–2 also hold in log-odds space (which is more relevant for the logistic regression models used below).

⁹We write predictor names in SMALL CAPS, and level names in teletype.

¹⁰Note that the cluster of points around “<0.01” in Figure 1 corresponds almost completely to data points followed by no pause: only 11 points have Pause duration annotated as between 1 and 10 msec, compared to 9844 with pause duration annotated as 0 msec. The appearance of variability in pause duration is due to the points being “jittered” to give a sense of the data’s distribution.

¹¹Note that the ‘no pause’ cutoff here is set to 0 msec, rather than 30 msec, despite the 30 msec cutoff in pause annotation described above. This is because a small number of hand-corrected pauses were given values below 30 msec by annotators.

¹²Splines were computed using `rCs` in the `rms` package (Harrell, 2014) applied to principal components. The number of knots was chosen by examining exploratory plots, as in Figure 2 (left).

¹³Although it appears in Figure 2 (right) that the frequency /deletion rate slope becomes markedly steeper around FREQUENCY = 4.0, this is not the case in log-odds space, where lines of best fit are within the confidence intervals for the nonlinear smooths. Note also that the preponderance of words with empirical deletion rates near 0% and 100% for lower word frequencies in this figure is expected, given that most word types in the corpus (70%) occur 5 or fewer times, at frequencies for which categorical deletion or non-deletion would be likely observed by chance, assuming for example an overall deletion rate of 19.5% (the rate implied by the intercept of the statistical model).

¹⁴The empirical plot in Fig. 3 suggests the possibility of a nonlinear effect; we found that adding a nonlinear term for conditional probability did not significantly improve model likelihood, so we code conditional probability as linear.

¹⁵Helmert contrasts were used because they are orthogonal, which minimizes multicollinearity, and are ‘centered’ in the sense that the intercept is the mean of factor levels, which facilitates interpretation of the main effect coefficients for predictors which interact with FOLLOWING CONTEXT and PRECEDING CONTEXT.

¹⁶Following a reviewer’s suggestion, we attempted to fit a model including interaction terms between ANNOTATOR and each predictor of interest, to evaluate whether our results depend on annotator identity. This model did not converge, presumably due to the many sparse or zero-count cells once the data is divided up by annotator. Empirical plots split up by annotator corresponding to Figs. 1–3 did not suggest any qualitative difference between annotators in the effects of interest for our research questions.

¹⁷However, for sufficiently long pauses or at e.g. turn ends, ‘pause’ should still be coded as a following context.

¹⁸Though not always, e.g. Zimmerer et al. (2014).

¹⁹A reviewer suggests an alternative interpretation: Since more frequent words are less likely to be realized with the final [t,d] in general, the question which segment follows may become less and less relevant with an increase in frequency. However, it’s not clear that this would be sufficient to explain the interaction between frequency and following environment—which is predicted by the PPH.

²⁰Production planning has been shown to be constrained by syntax and semantics, in that production windows tend to be syntactic constituents (cf. Wheeldon, 2012). If true, we might see gestural overlap be constrained by syntactic structure in ways that again are not entirely reducible to the prosodic modulation of gestures. Since our corpus is not syntactically annotated, we cannot test this prediction.

References

- Alario, F.-X., Costa, A., and Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, 17(3):299–319.
- Anttila, A. (2007). Variation and optionality. In De Lacy, P. V., editor, *The Cambridge handbook of phonology*, pages 519–536. Cambridge University Press, Cambridge.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, pages 255–278.
- Barry, M. (1985). A palatographic study of connected speech processes. *Cambridge papers in phonetics and experimental linguistics*, 4(1-16).
- Barry, M. C. (1992). Palatalisation, assimilation and gestural weakening in connected speech. *Speech communication*, 11(4):393–400.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.
- Beckman, M. E. and Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In Beckman, M. E. and Kingston, J., editors, *Papers in laboratory phonology I—Between the grammar and physics of speech*, page 152. Cambridge Univ Pr, Cambridge, UK.
- Bell, A., Brenier, J., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1):92–111.
- Bermúdez-Otero, R. (2010). Currently available data on English t/d-deletion fail to refute the classical modular feedforward architecture of phonology. Paper presented at the 18th Manchester Phonology Meeting.
- Boersma, P. and Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic inquiry*, 32(1):45–86.
- British National Corpus (2007). *The British National Corpus, version 3 (BNC XML Edition)*. Distributed by Oxford University Computing Services on behalf of the BNC Consortium.
- Browman, C. P. and Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, pages 341–376.
- Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.
- Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In Barlow, M. and Kemmer, S., editors, *Usage-based models of language*, pages 65–85. CSLI, Stanford.
- Byrd, D. and Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, 26(2):173–199.
- Byrd, D. and Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31:149–180.
- Carroll, J. B. and White, M. N. (1973). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*, 12(5):563–576.
- Cho, T. and Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, 29(2):155–190.
- Coetzee, A. and Pater, J. (2011). The place of variation in phonological theory. In Goldsmith, J., Riggle, J., and Yu, A., editors, *The Handbook of Phonological Theory*, pages 401–434. Wiley-Blackwell, 2nd edition.
- Coetzee, A. W. (2004). *What it means to be a loser: Non-optimal candidates in Optimality Theory*. PhD thesis, University of Massachusetts Amherst.
- Coetzee, A. W. (2009). Phonological variation and lexical frequency. In Anisa Schardl, Martin Walkow, M. A., editor, *Proceedings of the North East Linguistic Society 38*, pages 189–202, Amherst. GLSA.
- Coetzee, A. W. and Kawahara, S. (2013). Frequency biases in phonological variation. *Natural Language & Linguistic Theory*, 31(1):47–89.
- Cooper, W. E. and Paccia-Cooper, J. (1980). *Syntax and Speech*. Harvard University Press, Cambridge, Mass.

- Costa, A. and Caramazza, A. (2002). The production of noun phrases in english and spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory and Language*, 46(1):178–198.
- Dell, G. and O'Seaghdha, P. (1992). Stages of lexical access in language production. *Cognition*, 42(1-3):287–314.
- Ernestus, M., Lahey, M., Verhees, F., and Baayen, R. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, 120(2):1040–1051.
- Fasold, R. W. (1972). *Tense marking in Black English: A linguistic and social analysis*, volume 8. Harcourt College Pub.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30(2):210–233.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100:233–253.
- Ferreira, F. and Swets, B. (2002). How incremental is language production? evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46:57–84.
- File-Muriel, R. J. and Brown, E. K. (2011). The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change*, 23(02):223–243.
- Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, 29(2):109–135.
- Fougeron, C. and Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, 101(6):3728–3740.
- Fox Tree, J. and Clark, H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62(2):151–167.
- Fraundorf, S. and Watson, D. G. (2008). Dimensions of variation in disfluency production in discourse. *Semantics and Pragmatics of Dialogue*, page 124.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, pages 27–52.
- Fuchs, S., Petrone, C., Krivokapić, J., and Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, 41(1):29–47.
- Gahl, S. and Garsney, S. (2006). Knowledge of grammar includes knowledge of syntactic probabilities. *Language*, 82:405–410.
- Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4):789–806.
- Garrett, M. F. (1988). Processes in language production. In Newmeyer, F. S., editor, *Linguistics: The Cambridge Survey: Volume 3, Language: Psychological and Biological Aspects*, volume 3, pages 69–96. Cambridge University Press.
- Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, 15(2):103–13.
- Goldrick, M. and Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, pages 2198–260.
- Goodglass, H., Theurkauf, J. C., and Wingfield, A. (1984). Naming latencies as evidence for two modes of lexical retrieval. *Applied Psycholinguistics*, 5(02):135–146.
- Gregory, M., Raymond, W., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In *Proceedings of the Chicago Linguistic Society*, volume 35, pages 151–166.
- Griffin, Z. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin and Review*, 10:603–609.
- Grosjean, F. and Collins, M. (1979). Breathing, pausing, and reading. *Phonetica*, 36:98–114.
- Guy, G. (1980a). Variation in the group and the individual: The case of final stop deletion. In Labov, W., editor, *Locating language in time and space*, pages 1–36. Academic Press, New York.
- Guy, G. (1991a). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, 3(1):1–22.

- Guy, G., Hay, J., and Walker, A. (2008). Phonological, lexical, and frequency factors in coronal stop deletion in early New Zealand English. Poster at LabPhon 11.
- Guy, G. R. (1980b). Variation in the group and the individual: The case of final stop deletion. In Labov, W., editor, *Locating language in time and space*. Academic, New York.
- Guy, G. R. (1991b). Contextual conditioning in variable lexical phonology. *Language variation and change*, 3(02):223–239.
- Guy, G. R. and Boberg, C. (1997). Inherent variability and the obligatory contour principle. *Language Variation and Change*, 9(02):149–164.
- Hardcastle, W. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, 4:247–263.
- Harrell, Jr, F. E. (2014). *rms: Regression Modeling Strategies*. R package version 4.2-0.
- Hasegawa, N. (1979). Casual speech vs. fast speech. In Clyne, P. R., Hanks, W. F., and Hofbauer, C. L., editors, *Papers from the fifteenth regional meeting of the Chicago Linguistics Society*, pages 126–137, Chicago, IL.
- Hazen, K. (2011). Flying high above the social radar: coronal stop deletion in modern Appalachia. *Language Variation and Change*, 23:105–137.
- Heafield, K., Pouzyrevsky, I., Clark, J. H., and Koehn, P. (2013). Scalable modified kneser-ney language model estimation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 690–696, Sofia, Bulgaria.
- Jescheniak, J. D., Schriefers, H., and Hantsch, A. (2003). Utterance format effects phonological priming in the picture-word task: Implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2):441.
- Jespersen, O. (1922). *Language: its nature, development and origin*. George Allan & Unwin.
- Jr, F. E. H., with contributions from Charles Dupont, and many others. (2015). *Hmisc: Harrell Miscellaneous*. R package version 3.17-0.
- Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, *Frequency in the Emergence of Linguistic Structure*, pages 229–254. John Benjamins, Amsterdam.
- Kaisse, E. M. (1985). *Connected Speech. The interaction between syntax and phonology*. Academic Press, Orlando, Flo.
- Katz, J. (2016). Lenition, perception and neutralisation. *Phonology*, 33(1):43–85.
- Kawamoto, A. H., Liu, Q., Lee, R. J., and Grebe, P. R. (2015). The segment as the minimal planning unit in speech production: Evidence based on absolute response latencies. *The Quarterly Journal of Experimental Psychology*, 67(12):2340–2359.
- Keating, P. A. (2006). Phonetic encoding of prosodic structure. In Harrington, J. and Tabain, M., editors, *Speech Production: Models, Phonetic Processes and Techniques*, Macquarie Monographs in Cognitive Science, chapter 11, pages 167–186. Psychology Press, New York.
- Kendall, T. (2013). *Speech Rate, Pause, and Sociolinguistic Variation: Studies in Corpus Sociophonetics*. Palgrave Macmillan, New York.
- Kilbourn-Ceron, O. (2016). Speech production planning affects variability in connected speech. Talk presented at the Annual Meeting on Phonology, University of Southern California.
- Kingston, J., Diehl, R. L., Kirk, C. J., and Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of phonetics*, 36(1):28–54.
- Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonology*, 2:85–138.
- Kirchner, R. (1998). *Lenition in phonetically-based Optimality Theory*. PhD thesis, PhD dissertation, UCLA.
- Kittredge, A. K., Dell, G. S., Verkuilen, J., and Schwartz, M. F. (2008). Where is the effect of frequency in word production? insights from aphasic picture-naming errors. *Cognitive neuropsychology*, 25(4):463–492.
- Kochetov, A. and Pouplier, M. (2008). Phonetic variability and grammatical knowledge: An articulatory study of korean place assimilation. *Phonology*, 25(03):399–431.

- Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, 66(1):143–162.
- Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of phonetics*, 35(2):162–179.
- Labov, W. (1969). Contraction, deletion, and inherent variability of the english copula. *Language*, pages 715–762.
- Labov, W. (1989). The child as a linguistic historian. *Lang. Var. Change*, 1:85–97.
- Labov, W. (2010). *Principles of linguistic change. Volume 3: Cognitive and cultural factors*. Blackwell, Oxford.
- Labov, W., Cohen, P., Robins, C., and Lewis, J. (1968). A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City. *Project Literacy Reports*, 8:66–84.
- Lachman, R., Shaffer, J. P., and Hennrikus, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of verbal learning and verbal behavior*, 13(6):613–625.
- Ladd, D. R. and Scobbie, J. M. (2003). External sandhi as gestural overlap? counterevidence from sardinian. In Local, J., Ogden, R., and Temple, R., editors, *Phonetic interpretation: Papers in laboratory phonology VI*, pages 164–182. Cambridge University Press.
- Levelt, W. (1989). *Speaking. From Intention to Articulation*. MIT Press.
- Levelt, W., Roelofs, A., and Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(01):1–75.
- Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42(1-3):1–22.
- Lindblom, B. (1990). On the communication process: Speaker-listener interaction and the development of speech. *Augmentative and Alternative Communication*, 6(4):220–230.
- Miozzo, M. and Caramazza, A. (2003). When more is less: a counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, 132(2):228.
- Morrison, C. M., Ellis, A. W., and Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, 20(6):705–714.
- Mortensen, L., Meyer, A. S., and Humphreys, G. W. (2008). Speech planning during multiple-object naming: Effects of ageing. *The Quarterly Journal of Experimental Psychology*, 61(8):1217–1238.
- Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*. Foris, Dordrecht.
- Niebuhr, O., Clayards, M., Meunier, C., and Lancia, L. (2011). On place assimilation in sibilant sequences—comparing french and english. *Journal of Phonetics*, 39(3):429–451.
- Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In Docherty, G. and Ladd, D. R., editors, *Papers in Laboratory Phonology II: gesture, segment, prosody*, pages 261–280. Cambridge: Cambridge University Press.
- Nolan, F., Holst, T., and Kühnert (1996). Modelling [s] to [ʃ] accommodation in english. *Journal of Phonetics*, 24(1):113–137.
- Oldfield, R. C. and Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4):273–281.
- Oppermann, F., Jescheniak, J. D., and Schriefers, H. (2010). Phonological advance planning in sentence production. *Journal of Memory and Language*, 63(4):526–540.
- Pan, S. and Hirschberg, J. (2000). Modeling local context for pitch accent prediction. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 233–240, Morristown, NJ. Association for Computational Linguistics.
- Pluymaekers, M., Ernestus, M., and Baayen, R. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62(2-4):146–159.
- Post, B. (2000). Pitch accents, liaison and the phonological phrase in French. *Probus*, 12:127–164.
- Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90(6):2956–2970.

- Purse, R. and Turk, A. (2016). '/t,d/ deletion': Articulatory gradience in variable phonology. Ms. Poster presented at LabPhon 15, Cornell University.
- Rastle, K., Harrington, J., Coltheart, M., and Palethorpe, S. (2000). Reading aloud begins when the computation of phonology is complete. *Journal of Experimental Psychology: Human Perception and Performance*, 26(3):1178.
- Raymond, W., Dautricout, R., and Hume, E. (2006). Word medial /t, d/ deletion in spontaneous speech: modeling the effects of extra-linguistics, lexical, and phonological factors. *Language Variation and Change*, 18:55–97.
- Rosenfelder, I., Fruehwald, J., Evanini, K., and Yuan, J. (2011). FAVE (Forced Alignment and Vowel Extraction) Program Suite [Computer program].
- Schreier, D. (2005). *Consonant change in English worldwide: Synchrony meets diachrony*. Palgrave MacMillan, New York.
- Schriefers, H. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology*, 11(1):17–50.
- Schuppler, B., van Dommelen, W., Koreman, J., and Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, 40(4):595–607.
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, 3:371–405.
- Snijders, T. and Bosker, R. (2011). *Multilevel analysis*. Sage, London, 2nd edition.
- Snodgrass, J. G. and Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology*.
- Snodgrass, J. G. and Yuditsky, T. (1996). Naming times for the snodgrass and vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, 28(4):516–536.
- Sonderegger, M., Bane, M., and Graff, P. (2016). The medium-term dynamics of accents on reality television. Manuscript to appear in *Language*. Available at <http://people.linguistics.mcgill.ca/~morgan/>.
- Steriade, D. (2009). The phonology of perceptibility effects: The p-map and its consequences for constraint organization. In Hanson, K. and Inkelas, S., editors, *The Nature of the Word: Studies in Honor of Paul Kiparsky*, pages 151–179. Cambridge: MIT Press.
- Sternberg, S., Monsell, S., Knoll, R., and Wright, C. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. *Information processing in motor control and learning*, pages 117–152.
- Sugahara, M. and Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology*, 26(03):477–524.
- Swets, B., Jacovina, M. E., and Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. *Discourse Processes*, 50(1):23–51.
- Swets, B., Jacovina, M. E., and Gerrig, R. J. (2014). Individual differences in the scope of speech planning: evidence from eye-movements. *Language and Cognition*, 6(01):12–44.
- Tagliamonte, S. and Temple, R. A. M. (2005). New perspectives on an old variable: (t,d) in British English. *Language Variation and Change*, 17:281–302.
- Tamminga, M. (2014). *Persistence in the Production of Linguistic Variation*. PhD thesis, University of Pennsylvania.
- Tanner, J., Sonderegger, M., and Wagner, M. (2015). Production planning and coronal stop deletion in spontaneous speech. In *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK. University of Glasgow.
- Temple, R. (2013). Rethinking (t,d). Ms., Oxford University.
- Temple, R. (2014). Where and what is (t,d)? a case study in taking a step back in order to advance sociophonetics. In Calamai, C. C. S., editor, *Advances in Sociophonetics*, pages 97–136. John Benjamins, Amsterdam.
- Temple, R. A. M. (2009). (t,d): the variable status of a variable rule. *Oxford University Working Papers in Linguistics, Philology and Phonetics*, pages 145–170.

- Turk, A. (2010). Does prosodic constituency signal relative predictability? a smooth signal redundancy hypothesis. *Laboratory Phonology*, 1(2):227–262.
- Wagner, M. (2012). Locality in phonology and production planning. *McGill Working Papers in Linguistics*, 22(1):1–18.
- Wagner, V., Jescheniak, J. D., and Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2):423.
- Walker, J. A. (2012). Form, function, and frequency in phonological variation. *Language Variation and Change*, 24(3):397–415.
- Watson, D. G. and Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*.
- Whalen, D. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18:3–35.
- Wheeldon, L. (2012). Producing spoken sentences: The scope of incremental planning. In Fuchs, S., Weirich, M., Pape, D., and Perrier, P., editors, *Speech production and perception: Vol. 1. Speech planning and dynamics*, pages 97–118. Peter Lang, Frankfurt, Germany.
- Wheeldon, L. and Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37:356–381.
- Wolfram, W., Childs, B., and Torbert, B. (2000). Tracing language history through consonant cluster reduction: Comparative evidence from isolated dialects. *Southern Journal of Linguistics*, 24:17–40.
- Zimmerer, F., Scharinger, M., and Reetz, H. (2014). Phonological and morphological constraints on German /t/-deletions. *Journal of Phonetics*, 45:64–75.
- Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, 15:1–95.

Table 2: Summary of all random-effect terms included in the statistical model of coronal stop realization: variances and corresponding standard deviations.

| Predictor | Variance | Standard Deviation |
|---|----------|--------------------|
| Speaker | | |
| Intercept | 0.150 | 0.387 |
| Following Context 1 | 0.019 | 0.137 |
| Following Context 2 | 0.000 | 0.000 |
| Pause Class (linear) | 0.021 | 0.145 |
| Pause Class (quadratic) | 0.048 | 0.218 |
| Pause Class (cubic) | 0.000 | 0.000 |
| Preceding Context 1 | 0.047 | 0.216 |
| Preceding Context 2 | 0.010 | 0.101 |
| Word Frequency (log) | 0.043 | 0.206 |
| Conditional probability (log) | 0.008 | 0.092 |
| Speaking Rate Deviation 1 | 0.006 | 0.075 |
| Speaking Rate Deviation 2 | 0.085 | 0.291 |
| Pause Class (linear) : Following Context 1 | 0.131 | 0.361 |
| Pause Class (linear) : Following Context 2 | 0.031 | 0.175 |
| Pause Class (quadratic) : Following Context 1 | 0.113 | 0.336 |
| Pause Class (quadratic) : Following Context 2 | 0.004 | 0.064 |
| Pause Class (cubic) : Following Context 1 | 0.183 | 0.428 |
| Pause Class (cubic) : Following Context 2 | 0.000 | 0.000 |
| Pause Class (linear) : Preceding Context 1 | 0.047 | 0.216 |
| Pause Class (linear) : Preceding Context 2 | 0.000 | 0.000 |
| Pause Class (quadratic) : Preceding Context 1 | 0.015 | 0.121 |
| Pause Class (quadratic) : Preceding Context 2 | 0.000 | 0.000 |
| Pause Class (cubic) : Preceding Context 1 | 0.051 | 0.225 |
| Class (cubic) : Preceding Context 2 | 0.104 | 0.322 |
| Word Frequency : Following Context 1 | 0.000 | 0.000 |
| Word Frequency : Following Context 2 | 0.000 | 0.000 |
| Following Context 1 : Speaking Rate Deviation 1 | 0.005 | 0.071 |
| Following Context 2 : Speaking Rate Deviation 1 | 0.000 | 0.000 |
| Following Context 1 : Speaking Rate Deviation 2 | 0.000 | 0.000 |
| Following Context 2 : Speaking Rate Deviation 2 | 0.007 | 0.083 |
| Following Context 1 : Conditional probability (log) | 0.182 | 0.427 |
| Following Context 2 : Conditional probability (log) | 0.101 | 0.318 |
| Word | | |
| Intercept | 0.361 | 0.601 |
| Pause Class (line) | 0.000 | 0.000 |
| Pause Class (quadratic) | 0.233 | 0.483 |
| Pause Class (cubic) | 0.057 | 0.239 |
| Bigram | | |
| Intercept | 0.775 | 0.881 |

Table 3: Results of likelihood ratio tests for sets of terms in the logistic regression model related to research questions. The row for each variable reports the χ^2 test (with degrees of freedom in parentheses) and significance for comparing models with and without all fixed- and random-effect terms corresponding to the variable.

| Variable | χ^2 (df) | Pr(> χ^2) |
|---|---------------|-----------------|
| Pause Class | 147.5 (9) | <0.0001 |
| Pause Class : Preceding Context | 21.0 (12) | 0.051 |
| Pause Class : Following Context | 103.1 (12) | <0.0001 |
| Conditional Probability : Following Context | 12.7 (4) | 0.012 |
| Word Frequency : Following Context | 15.9 (4) | 0.0032 |
| Speaking Rate Deviation : Following Context | 24.9 (8) | 0.0016 |