PRODUCTION PLANNING AND CORONAL STOP DELETION IN SPONTANEOUS SPEECH

James Tanner, Morgan Sonderegger, Michael Wagner

Department of Linguistics, McGill University, Canada
james.tanner@mail.mcgill.ca, morgan.sonderegger@mcgill.ca, chael@mcgill.ca

ABSTRACT

Coronal stop deletion in English word-final consonant clusters (CSD), one of the most studied variables in sociolinguistics, has been consistently shown to be affected by the preceding and following phonological context. Previous work has treated a following pause as one type of context, on par with a following consonant or vowel. Looking at deletion rates in a corpus of spontaneous speech, we instead treat pause as a quantitative proxy for boundary strength and test the hypothesis that, not only does pause by itself gradiently reduce deletion rate, but also modulates the effect of a following phonological segment, as predicted if phonological processes are constrained by the locality of production planning. We show that the influence of a following segment on CSD decreases with increasing boundary strength, but not that of a preceding segment: an asymmetry that provides direct support for the production planning locality hypothesis.

Keywords: Phonetics of conversation, coronal stop deletion, production planning

1. INTRODUCTION

This paper examines the realisation of word-final coronal stops in English consonant clusters as a method of addressing a larger issue: *what is the relationship between prosodic boundaries and segmental variation?* We argue that this relationship can be better understood by reference to production planning, the psycholinguistic process in which speech sounds are encoded online.

Coronal stop deletion (CSD, a.k.a t/d deletion) is one of the most studied cases of variable segmental realisation in English, with decades of work in the sociolinguistic and phonetic literatures showing that a variety of factors condition deletion rate, including surrounding segmental environment, speaking rate, word frequency and morphological class [5, 9, 16, 20]. The following phonological context consistently has the largest effect [12], with word-final t/d deleting more often before more similar segments (e.g. near-categorically before coronal stops) [5, 21]. Prosodic boundaries have long been recognised to affect CSD rate, operationalised in most work as a following context of "pause" (variously defined), and treated like a phonological context—each t/d is followed by *either* a pause *or* a vowel *or* a consonant. This conceptualisation of prosodic boundary is common in the wider literature on variable segmental realisation beyond CSD, such as analogous deletion of final /t/ in Dutch [13].

Studies of CSD have found very different effects of "pause" on deletion rate, which are usually ascribed to dialectal differences [4, 6, 8]. Another possibility, however, is that the effect of prosodic boundaries on CSD rate are richer than previously suggested by the binary analysis, in two ways. First, because prosodic boundaries coexist with segmental context, it makes sense to treat them as independent factors influencing deletion rate, rather than as mutually exclusive. Second, because boundaries of different strengths may have different effects, it makes sense to treat boundary strength as continuous, rather than binary. The first goal of this paper is to clarify the role of prosodic junctures in CSD through an analysis incorporating these methodological changes.

Turning to production planning, it is known that the planning window in which detailed phonological encoding takes place is very narrow. Early work [18] hypothesised that the phonological motor plan is subject to rapid decay, and hence only planned very locally, while Levelt's influential theory [10, 11] holds that the planning window for phonological encoding does not extend beyond a single prosodic word. This would rule out any influence of phonological material in an upcoming word on the realisation of the current one, and is hence incompatible with the strong dependence of CSD rate on the following segment. More recent work has shown that the planning window must extend beyond the current prosodic word, at least under certain circumstances [7, 15]. [22] argues that the locality of production planning has interesting and complex effects on the variability and locality of phonological processes, based on the plausible assumption that the likelihood with which any phonological information about an upcoming word will be have been planned is inversely correlated with the strength of the prosodic boundary separating the two words.

With respect to CSD, this means that information about the following phonological context is only probabilistically 'available' if the information about the first segment of the following word has been retrieved when the articulation of the cluster containing t/d is planned. We therefore expect the effect of the following segment but not of a preceding segment on CSD *to be gradiently modulated by the size of the boundary separating the two words*: the larger the boundary, the smaller the effect of the following segment on CSD rate. We call this prediction about how production planning mediates variable realisation the *production planning hypothesis* (PPH; [22]). So far, the PPH has been tested for cases of variable allomorphy (English *in-ing*) [22], and for rate of flapping, glottalization and release of word-final [t] following vowels [23]; the PPH makes important predictions about other kinds of variable process, and so additional investigation is of prime interest. Thus, a second goal of this paper is to test the PPH for the case of CSD.

This paper addresses these goals by testing several hypotheses. Based on the PPH, we expect prosodic boundaries to affect CSD by modulating the effect of the following segment, but not the effect of the preceding segment, as described above. Second, we predict a global gradient effect of boundary strength on CSD rate, independent of following context. We test these hypotheses in a corpus of spontaneous British English speech, using duration of the following pause as a proxy for boundary strength, and controlling for other variables which affect CSD rate.

2. DATA

The data comes from a subset of a corpus of speech from contestants on the 2008 season of *Big Brother UK* [16, 17]. The current dataset comes from 20 speakers, mostly of different varieties of British English. (One speaker each is from the US and Australia; the remaining 18 appear to be native speakers of British English varieties.) The dataset contains 6646 observations of word-final consonant clusters ending in an underlying /t/ or /d/ segment corresponding to 410 unique word types (per speaker: mean = 332.3, sd = 262.75; per word: mean = 16.5, sd = 135.87). Three research assistants transcribed the data, counting any phonetic realisation of the t/d segment (including burst and glottalisation) as non-deletion. Deletion of word-final /t/ or /d/ occurred in 4588 observations (token: 69%, type: 41%), comparable with previous studies of British English CSD [14, 20]. (We give both deletion rates averaged over all tokens, and averaged over word types, since types occur with very different frequencies.) The data was coded for surrounding segmental environment (PRECEDING CONTEXT and FOLLOWING CONTEXT), PAUSE LENGTH (log-transformed), FREQUENCY (CELEX wordform, log-transformed; [1]), SPEAKING RATE, and MORPHOLOGICAL CLASS (2 levels: past tense, other). As phonological environment and pause length are the variables directly related to our hypotheses, only they will be described in detail here.

2.1. Phonological Context

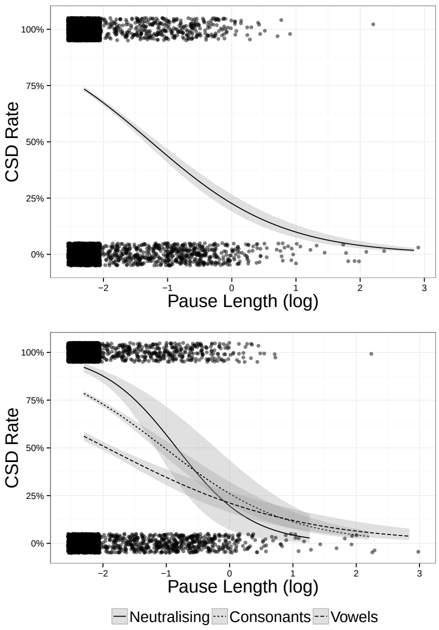
The following segmental environment was coded with 3 levels: *neutralising* consonants (i.e., coronal obstruents: /t/, /d/, /s/, /z/, /θ/, /ð/), non-neutralising *consonants*, and *vocalic* segments. (In contrast to previous work on CSD [5, 20, 19, 21], tokens in "neutralising" context were not excluded from the dataset; we instead accounted for high deletion rates in this context by including appropriate terms in the statistical model presented below [16].) The direction of deletion follows the observations of previous CSD studies, where neutralising environments induce the largest amount of deletion (type: 86%, token 89%), followed by consonants, (type: 53%, token: 75%), and vowels inducing the lowest deletion rates (type: 21%, token: 51%).

The preceding segmental environment was coded with 3 levels, based on previous work: sonorants, sibilants, and nonsibilant obstruents [20]. As expected, t/d segments before sibilants favoured deletion the most (type: 54%, token: 73%), followed by sonorants (type: 43%, token: 70%), whilst nonsibilant fricatives disfavoured deletion (type: 24%, token: 25%).

2.2. Pause

Fig. 1 (top) demonstrates that CSD rate is negatively correlated with the length of the pause between the CSD environment and the following segment (Spearman's $\rho = -0.275$). In this sense, pause length has a gradient effect, where deletion is partially influenced as a function of the length of the pause. Whilst pause seems to reduce the rate of deletion as the length of the pause between segments increases, the pause length also seems to modulate the relative differences *between* following phonological environments. Fig. 1 (bottom) demonstrates this relationship, where the different deletion rates after

Figure 1: CSD rate observed in the dataset, as a function of pause length (log-transformed) (top) and both following context and pause length (bottom). Solid line indicates logistic smooth to empirical data, with 95% confidence intervals (shaded area). Jittered points indicate the distibution of deletion and non-deletion.



each following segmental environment is mitigated before long pauses. These effects of pause duration are not directly comparable with previous work on CSD, which have treated pause as a binary variable (as discussed above).

3. MODEL AND RESULTS

The production planning hypothesis is concerned with the relationship between the length of a pause (as a proxy for determining the strength of a prosodic juncture) and the following phonological environment in conditioning the likelihood of deleting a word-final coronal stop. The fixed effects structure of the model contained the predictors of FOLLOWING CONTEXT, PRECEDING CONTEXT, PAUSE LENGTH (log-transformed), SPEAKING RATE, WORD FREQUENCY (log-transformed), and MORPHOLOGICAL CLASS. In addition, the following interaction terms were included: PAUSE LENGTH : FOLLOWING CONTEXT, PAUSE LENGTH : PRECEDING CONTEXT, SPEAKING RATE : FOLLOWING CONTEXT, and WORD FREQUENCY : FOLLOWING CONTEXT. These additional predictors were included to improve the accuracy of the model's estimates, as well as to examine the effect of pause length on other factors relevant to the planning of speech.

The data was fit as a mixed-effect logistic regression using the lme4 package in R [3]. Continuous variables were centred and divided by two standard deviations. The two-level factor (morphological class) was transformed to a numerical predictor and centred. The preceding and following phonological context were coded using helmert contrasts, e.g. neutralising environment versus non-neutralising consonants (contrast 1) and all consonants versus vowels (contrast 2) for following context. The model was fit with a maximal random-effects structure (full by-word and by-speaker intercepts and slopes) [2], but did not include any correlations between random effects.

3.1. Results

The fixed-effect coefficients of the model are shown in Table 1. Note that because of how predictors in this model are coded, fixed-effect coefficients for main effects can be interpreted as effects when all other variables are averaged over.

Table 1: Fixed effect coefficients ($\hat{\beta}$), standard errors (se($\hat{\beta}$)), z-scores, and p-values for all model predictors.

Fixed Effects	$\hat{\beta}$	se($\hat{\beta}$)	z	Pr($\|z\|$)
(Intercept)	−0.044	0.173	−0.257	0.797
Speaking Rate	0.757	0.188	4.035	< 0.0001
Pause (log)	−2.024	0.307	−6.597	< 0.0001
Following Context (Neut vs. C)	−0.810	0.177	−4.584	< 0.0001
Following Context (Neut/C vs. V)	−0.902	0.079	−11.354	< 0.0001
Morphological Class	0.077	0.171	0.449	0.654
Word Frequency (log)	0.456	0.179	2.553	0.011
Preceding Context (Sib vs. Son)	−0.374	0.116	−3.223	0.001
Preceding Context (Sib/Son vs. Obs)	−0.548	0.077	−7.107	< 0.0001
Pause (log) : Following Context (Neut vs. C)	0.455	0.220	2.070	0.039
Pause (log) : Following Context (Neut/C vs. V)	0.481	0.081	5.967	< 0.0001
Pause (log) : Preceding Context (Sib vs. Son)	0.288	0.215	1.340	0.18
Pause (log) : Preceding Context (Sib/Son vs. Obs)	0.101	0.205	0.496	0.62
Speaking Rate : Following Context (Neut vs. C)	0.163	0.212	0.771	0.44
Speaking Rate : Following Context (Neut/C vs. V)	0.350	0.114	3.058	0.002
Following Context (Neut vs. C) : Word Frequency (log)	0.001	0.207	0.048	0.961
Following Context (Neut/C vs. V) : Word Frequency (log)	0.243	0.099	2.458	0.014

Speaking rate and word frequency had strong and significant effects on deletion rates ($\hat{\beta} = 0.757$, $z = 4.035$, $p < 0.0001$; $\hat{\beta} = 0.456$, $z = 2.553$, $p = 0.011$), with t/d more likely to delete in faster speech and more frequent words (c.f. [24]). Confirming the finding of [20] for British English, morphological class did not significantly affect deletion rate ($p = 0.654$).

3.1.1. Phonological Context, Pause

We first consider how phonological context and pause *independently* affect deletion rate, by discussing the main effect terms of Table 1.

The effect of phonological context (i.e., at average pause duration) follow the findings of previous research. The following context has a strong and significant effect on the likelihood of deletion (rows 4–5 in Table 1), with vowels inducing less deletion than consonantal segments, and a large difference in deletion rates between neutralising and non-neutralising consonants, as expected given the near-categorical deletion rate for neutralising consonants [5, 6, 20]. The preceding context also significantly affects deletion rate (rows 8–9 in Table 1), with the effect of previous segment identity on deletion rate (sibilant obstruents > sonorants > non-sibilant obstruents) following that observed in previous studies [8, 20]. The size of the preceding context effect is smaller than that of the following context: changes in the type of preceding segment result in smaller changes in deletion rate.

Pause length (i.e., averaging over phonological contexts) has a large and significant effect on deletion rate ($\hat{\beta} = -2.024$, $z = -6.597$, $p < 0.0001$): the probability of deletion is reduced as pause length increases, *independently* of the phonological context.

3.1.2. Pause & Phonological Context: interaction

Whilst the model has shown that the length of the pause reduces CSD rates globally, the production planning hypothesis predicts that the differences observed between following phonological contexts should be minimized as the length of the pause increases. The model reports a strong and significant interaction effect (rows 10–11 of Table 1), which is very similar to the pattern seen in the empirical data, in Fig. 1 (bottom): as the pause between the deletion environment and the following segment increases, the over all CSD rate reduces, and the relative difference between deletion rates for different classes of following segment is reduced, resulting in similar deletion rates regardless of the following segmental environment (we do not show separate model predictions here, for lack of space).

A crucial prediction of the production planning hypothesis is that the length of pause should *not* condition the relative differences of preceding context on deletion. The corresponding terms in the model, for the interaction of preceding context and pause, are weak and non-significant (rows 12–13 of Table 1); in addition, a likelihood ratio test comparing models with and without this interaction confirms that it does not significantly affect model likelihood ($\chi^2_{(4)} = 2.6443$, $p = 0.619$). Thus, there is no evidence that the differences between preceding segmental environments are conditioned by the length of pause. Instead, segmental differences remain largely constant across differing pause lengths.

4. DISCUSSION

The model reported here confirms the patterns observed in the empirical data—in particular, the strong interaction between pause and following context. This result is directly predicted by the PPH, under which the conditioning of phonological processes is assumed to be constrained by the locality of production planning, and the assumption that prosodic boundary strength indeed inversely correlates with the availability and detail of upcoming phonological information. It also follows that the effect of word-internal segments preceding t/d, which should be reliably available independent of the strength of a following boundary, will not show the same interaction.

If phonological encoding is universally as local as it has been found to be in English and other languages studied so far, then the PPH makes the prediction that any phonological effect conditioned by segmental information across word boundaries is *necessarily* variable and modulated by prosodic boundary strength. The hypothesis is thus not only able to rationalise existing patterns of variability, but also makes predictions about whether a phonological process will be variable, depending on the information assumed to trigger them.

Our findings serve to clarify the effect of pause in CSD as well as other segmental reduction processes. First, the effect of pause length on CSD is gradient, where deletion globally reduces as a linear function of pause length. Second, pause has a *modulating* effect on other predictors, where, in the case of CSD, the effect of the following context is reduced as pause increases. These findings provide a new direction for approaching other kinds of variable phonological and sandhi processes.

Acknowledgements This work was partially funded by SSHRC #430-2014-00018 and FRQSC #183356 to MS, SSHRC #410-2011-1062 to MW, and SSHRC #435-2014-1504 to MW and Dr. Meghan Clayards. We thank L. Bassford, T. Knowles, M. Labelle, and M. Schwartz for research assistance, and M. Clayards for comments.

5. REFERENCES

[1] Baayen, R., Piepenbrock, R., Gulikers, L. 1996. *CELEX2 (CD-ROM)*. Philadelphia: Linguistic Data Consortium.

[2] Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. 2013. Random effects structure for confirmatory hypoth-

esis testing: Keep it maximal. *Journal of Memory and Language* 255–278.

[3] Bates, D., Maechler, M., Bolker, B., Walker, S. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.

[4] Coetzee, A., Pater, J. 2011. The place of variation in phonological theory. In: Goldsmith, J., Riggle, J., Yu, A., (eds), *The Handbook of Phonological Theory*. Wiley-Blackwell 401–434 2nd edition.

[5] Guy, G. 1991. Explanation in variable phonology: An exponential model of morphological constraints. *Lang. Var. Change* 3(1), 1–22.

[6] Hazen, K. 2011. Flying high above the social radar: coronal stop deletion in modern Appalachia. *Lang. Var. Change* 23, 105–137.

[7] Jescheniak, J. D., Schriefers, H., 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.

[8] Labov, W. 1989. The child as a linguistic historian. *Lang. Var. Change* 1, 85–97.

[9] Labov, W., Cohen, P., Robins, C., 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.

[10] Levelt, W. J. 1992. Accessing words in speech production: Stages, processes and representations. *Cognition* 42(1-3), 1–22.

[11] Levelt, W. J., Meyer, A. S. 2000. Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology* 12(4), 433–452.

[12] Schreier, D. 2005. *Consonant change in English worldwide: Synchrony meets diachrony*. New York: Palgrave MacMillan.

[13] Schuppler, B., van Dommelen, W., Koreman, J., 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. *J. Phon.* 40(4), 595–607.

[14] Smith, J., Durham, M., Fortune, L. 2009. Universal and dialect-specific pathways of acquisition: Caregivers, children, and t/d deletion. *Lang. Var. Change* 21(1), 69–95.

[15] Smith, M., Wheeldon, L. 2004. Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(3), 675.

[16] Sonderegger, M. 2012. *Phonetic and phonological dynamics on reality television*. PhD thesis University of Chicago.

[17] Sonderegger, M. 2014. Trajectories of phonetic variability in spontaneous speech on reality TV. Poster presented at LabPhon 14 (Tachikawa, Japan).

[18] Sternberg, S., Monsell, S., Knoll, R., Wright, C. 1978. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In: Stelmach, G. E., (ed), *Information processing in motor control and learning*. New York: Academic Press 117–152.

[19] Tagliamonte, S. 2006. *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

[20] Tagliamonte, S., Temple, R. A. M. 2005. New perspectives on an ol' variable: (t,d) in British English. *Lang. Var. Change* 17, 281–302.

[21] Temple, R. A. M. 2009. (t,d): the variable status of a variable rule. *Oxford University Working Papers in Linguistics, Philology and Phonetics* 145–170.

[22] Wagner, M. 2012. Locality in phonology and production planning. *McGill Working Papers in Linguistics* 22(1), 1–18.

[23] Wagner, M., Clayards, M. 2013. Syntactic effects on variable phonological processes and the locality of production planning. Talk presented at the workshop on 'Universality and variability in segment prosody interactions' at the 2013 Linguistics Institute of the LSA, Ann Arbor, MI, July 12.

[24] Walker, J. A. 2012. Form, function, and frequency in phonological variation. *Lang. Var. Change* 24(3), 397–415.