Connecting Phonological Encoding To Articulation - Is Cascading Required? A Computational Investigation

H. Susannah Moat (h.s.moat@sms.ed.ac.uk)
Martin Corley (martin.corley@ed.ac.uk)
School of Philosophy, Psychology and Language Sciences, University of Edinburgh
7 George Square, Edinburgh, EH8 9JZ, United Kingdom

Robert J. Hartsuiker (robert.hartsuiker@ugent.be)
Department of Experimental Psychology, University of Ghent
Henri Dunantlaan 2, 9000 Ghent, Belgium

Abstract

Current psychological models of word production (e.g. Dell, 1986; Levelt, Roelofs, & Meyer, 1999) only detail how we plan the phonological content of words, and not how we articulate them. To understand how these models may be extended, we must determine how information flows from phonological encoding to articulation. For example, does activation cascade from unselected phonological representations? So far, the clearest support for cascading at this interface has come from the finding that erroneously produced phonemes exhibit characteristics of the intended phoneme (Goldrick & Blumstein, 2006). In this paper however, we use computational implementations to demonstrate that both cascading and non-cascading models can account for this result. An extension of model behaviour analysis to other speech error phenomena additionally shows that models based on the classic spreading activation account of word production (Dell, 1986) experience problems in replicating some key aspects of human error patterns.

Keywords: word production; phonological encoding; articulation; cascading; spreading activation

Introduction

What links the plan to say a word to its articulation? Although a number of models independently address planning (e.g., Dell, 1986; Levelt et al., 1999) or articulation (Browman & Goldstein, 1989; Guenther, 2003), a fuller model would explain the process of word production from conceptualisation through articulatory execution. As a preliminary move towards such a model, the present paper investigates the flow of information between phonological encoding and articulation.

To produce a word, the appropriate conceptual, lexical and phonological representations must be selected. Models such as those of Dell (1986) and Levelt et al. (1999) assume that at each level, similar representations may be partially activated. For example, communicating the concept of a couch should entail high activation and selection of the *couch* lexical entry, but the lexical entry for *sofa* may also become activated. Recently, studies such as that by Peterson and Savoy (1998) have concluded that activation from such unselected lexical entries cascades to phonological encoding. The question addressed in this paper is whether a similar cascade of activation occurs between the phonological level and articulation. We present a detailed computational investigation of this interface which shows that activation from unselected phonological representations need not necessarily cascade to the articulatory level in order to account for existing evidence.

Evidence for cascading

Various studies have suggested that articulation can reflect evidence of the activation of more than one phoneme at a time (Boucher, 1994; Frisch & Wright, 2002; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; Mowrey & MacKay, 1990). This has lead some authors to claim that activation from multiple phonemes passes through to articulation (Frisch & Wright, 2002; Mowrey & MacKay, 1990). However, it is possible that the evidence in these studies can be accounted for through the introduction of noise at the articulatory level, after phonological encoding is complete.

Goldrick and Blumstein (2006) addressed this issue by asking participants to produce tongue twisters such as *keff geff geff keff*, where words in the tongue twister differed only by the voicing of the onset. Their results demonstrated that when participants attempted to produce /g/s which were identified as sounding like /k/s, these /k/s were more voiced than intended /k/ onsets identified as /k/s. In other words, there was a trace of the intended voiced phoneme /g/ on an errorful production of the voiceless phoneme /k/. Traces were also observed for productions of voiced phonemes.

Goldrick and Blumstein argued that their findings could not be accounted for by articulatory noise, as there would be no reason for articulatory noise to systematically affect only phonemes which were selected in error. Therefore, they claimed, activation from the intended phoneme cascades to articulation, even when noise in the model causes another phoneme to be selected. When a /g/ in a tongue twister is pronounced as /k/, the phonemic representation of /g/ is relatively active (because it was the originally intended onset) and this activation cascades, affecting the articulation of the errorful /k/. If a /k/ is intended and selected, there will be relatively little activation to cascade from the /g/. In a model without cascading, the non-selected onset cannot affect articulation, and the resulting output for an errorful /k/ would not differ from an intended /k/.

However, there are two possible ways in which a model without cascading would be able to account for Goldrick and Blumstein's (2006) findings. First, the *activation level* of the selected phoneme may be lower if it has been selected in error. Second, *articulatory noise* may still be able to account for this evidence. We outline each of these arguments in more detail below.

Activation levels

Our key account of non-cascading traces is based on the fact that an intended phoneme will receive activation from higher-level processes, and will therefore on average be more activated when selected than a phoneme which is activated through noise. An erroneously selected and therefore less activated /k/ may activate its articulatory features less strongly than a correctly selected and therefore more activated /k/, such that an erroneously selected /k/ may result in a less voiceless production. By definition, a less voiceless production is more voiced, such that a trace of the intended /g/ would be present in the final articulation without any activation having been transmitted from the /g/ at the phonemic level.

Articulatory noise

An alternative account returns to articulatory noise. As observed by Goldrick and Blumstein (2006), there is no reason to believe that noise would affect incorrectly selected phonemes more than correctly selected phonemes. However, what is transcribed as an incorrectly selected phoneme may in fact be a correctly selected phoneme which has been affected by noise at the articulatory level. For example, a speaker may intend to produce a /g/ and correctly select the /g/ phoneme. Articulatory noise may then lead to the /g/ phoneme being realised as a /k/. Because the articulation level would retain activation from the /g/, the errorful /k/ would be more voiced than a correctly produced /k/.

One piece of experimental evidence speaks against this hypothesis. In a post-hoc analysis, Goldrick and Blumstein (2006) failed to find a significant difference in variability of voicing for tokens identified as correct between a tongue-twister and a control task. In this study we examine whether computational modelling can add to the evidence for or against articulatory noise as a source of traces.

To investigate these accounts, we implemented a model with optional cascading from phonological encoding to articulation, and evaluated its performance in producing word onsets. In Experiment 1 we aim to simulate evidence from Goldrick and Blumstein (2006). Experiment 2 investigates whether successful models generalise to other patterns of human speech errors from the literature. Below, we detail the model implementation before presenting our experimental findings.

Model Implementation

Model architecture

Our model was based on Dell's (1986) model of word production, and had word, phoneme, and articulatory feature levels. Figure 1 shows the portion of the model required for producing the words "gap" and "cap". The model had a further 48 CVC words in its vocabulary, together with the phonemes and features required to produce those words. We added these extra words as we wished to investigate contextual errors, and in a two word network, non-contextual errors would always look like contextual errors. The 48 words were chosen randomly,

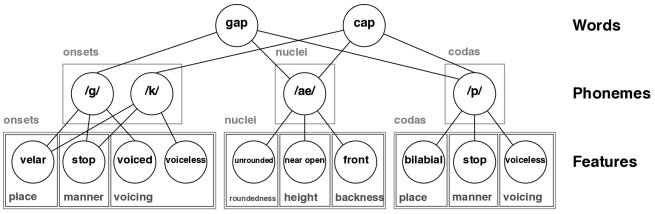

Figure 1: Architecture of the model

with the single constraint that at least one word beginning with /g/ and one word beginning with /k/ was found. This ensured that there was a possible influence of lexical onsets on the errors produced. The full list of words used is provided in the appendix.

Processing

Word production in the model occurs in two stages. First, the node corresponding to the word the model should produce is selected and its activation boosted. Any immediately upcoming words are primed with a smaller amount of activation to approximate processing at higher levels (cf. Dell, 1986). Activation then passes through the network, and the most activated phonemes are selected and their activation boosted. In the second stage, activation continues to pass through the network, after which the most activated features are selected for each feature dimension. For example, for the onset phoneme, the most activated features are selected for each of place, manner and voicing. Following production of the word, the activation levels of all previously selected nodes are set to 0. Production of any remaining words then ensues in the same manner.

The manner in which activation spreads through the network depends on the connectivity settings selected. The model could be configured to permit activation from unselected nodes in one layer to cascade to the layer directly below; e.g., from the word layer to the phoneme layer, or, as was of particular interest to us, from the phoneme layer to the feature layer. Additionally, activation from a lower layer could optionally feed back to connected nodes on a higher layer, i.e., from the phoneme layer to the word layer, or from the feature layer to the phoneme layer. In our model, we do not implement feedback without cascading.

Activation of nodes in the network was calculated in the same manner as Dell, Schwartz, Martin, Saffran, and Gagnon (1997). Following Hartsuiker (2002), we added an additional parameter such that downward and upward spreading rates could differ.

For a node in a layer where feedback of activation from the layer below was not enabled, upward spreading rate would be 0. A node in a layer where there was no cascading of activation from the layer above would only receive activation from connected selected nodes in the layer above, rather than all connected nodes in the layer above. If selection had not yet taken place in the layer above, that node would receive no activation.

Table 1: Activation parameter values used in simulations

Name	Values
Downwards spreading rate	0.05, 0.2, 0.35
Upwards spreading rate	0.05, 0.2, 0.35
Decay	0.4, 0.5, 0.6
Activation noise factor	0.05, 0.15, 0.25
Intrinsic noise standard deviation	0.005, 0.01, 0.05
Jolt	50, 100, 150, 200
Prime	10, 50, 100
Number of steps per stage	2, 5, 8

Parameters

In order to carry out the simulations, a number of parameter values must be determined. Firstly, parameters affecting the spreading of activation must be set. The empirical grounding of these parameters is not obvious. We therefore tested model behaviour across a number of parameter setting combinations, to help reveal the true effect of changing the phonological encoding to articulation connectivity. The parameter values we used, shown in Table 1, were based on values used in the previous literature (Dell, 1986; Dell et al., 1997; Goldrick, 2006; Hartsuiker, 2002; Rapp & Goldrick, 2000). This facilitates a linking of our results to those of previous word production modelling studies.

All possible combinations of activation parameter values were tested over all connectivity options, subject to two constraints, such that the prime always had to be less than the jolt, and the feedback connection strength was never greater than the forward connection strength.

As well as the activation spreading parameters, the connectivity settings must be decided. Following Peterson and Savoy (1998) we elected to enable word to phoneme cascading. However, we tested all our models both with and without phoneme to word feedback.

We configured our model to simulate three categories of connectivity between phonological encoding and articulation: non-cascading, cascading, and cascading with feedback. The latter variant was included to enable comparison with Dell's (1986) model. We tested each of these categories of model across all the parameter value combinations detailed above. This resulted in a total of 21,870 simulations: 6,561 non-cascading and 6,561 cascading simulations, and 8,748 cascading with feedback simulations (the feedback connection strength parameter could not be varied for models in which no feedback occurred between any levels).

Model output

Preliminary analyses focus on the onset of each word produced. The production of each onset was classified in two ways. The first approach involved resolving the selected features into a phoneme. For example, if the selected onset features were velar, stop, and voiced, the onset was classified as a /g/. This corresponds to classifying each production by *resulting articulation*, and is intended to reflect the way in which we normally determine what someone has said. The second approach followed previous work such as that of Dell (1986), classifying responses by *selected phoneme*. This was achieved by recording the phonemes selected at the phonological level. Note that the choice of classification method is critical when determining how any traces in errorful speech are generated. If the main cause of traces is articulatory noise, classification by resulting articulation will result in traces being identified (an intended /g/ is correctly selected at the phoneme level, but noise on the voicing nodes leads to it being identified as a /k/ with strong /g/ characteristics) but classification by the selected phoneme will not (a correctly selected /g/ phoneme should be articulated the same way as an incorrectly selected /g/ phoneme).

The simulations were implemented in Java 1.5. Simulations were carried out using the Edinburgh Compute and Data Facility (`www.ecdf.ed.ac.uk`).

Experiment 1

Experiment 1 was designed to test whether models in the three connectivity categories displayed traces of intended phonemes on erroneous productions.

Method

Each variant of the model produced the phrases "gap cap" and "cap gap" 5,000 times each, resulting in a total of 10,000 phrase productions per variant. Each onset production was classified as correct, a contextual error, or a non-contextual error (once using classification by resulting articulation and once using selected phoneme). For each production resulting in a /k/ or a /g/, voicing of the onset phoneme was calculated as the activation of the voiceless feature minus that of the voiced feature. This measure varies in a similar way to voice onset time, such that very voiceless productions result in a high value, and voiced productions result in a low value.

Voicing values were compared for correct and erroneous productions, assuming that at least two /k/ → /g/ and two /g/ → /k/ errors had been observed. If the values differed significantly in the right direction for correct compared to erroneous productions, we assumed that traces of the intended phoneme were observed when errors were produced. For the purposes of the present paper, only simulations which produced traces in both /k/ → /g/ and /g/ → /k/ errors were considered.

Results

Figure 2 shows the results of Experiment 1. The results are displayed as percentages of simulations to facilitate comparison between sets of simulations. 52.2% of simulations did not produce enough errors for trace analysis. When classifying productions by resulting articulation, we found 565 parameter setting combinations for the non-cascading model which displayed traces, out of 3,024 simulations where enough errors occurred to permit trace analysis. For the cascading model, we found 1,503 simulations which displayed traces out of 3,008 analysable simulations. 2,619 out of 4,414 analysable cascading with feedback simulations displayed traces.

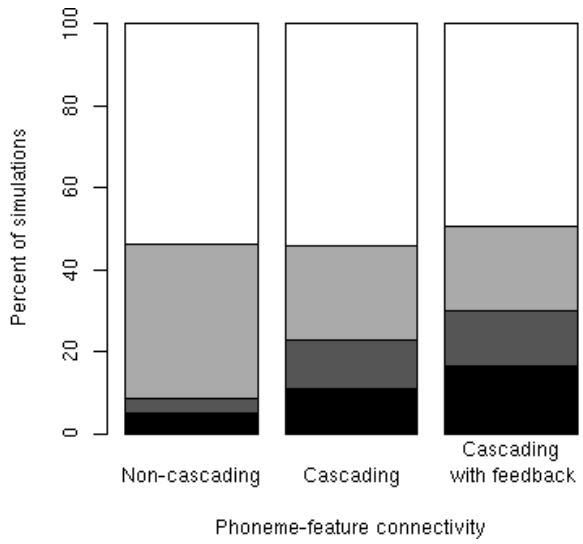

Figure 2: Results of Experiment 1. Portions of bars coloured in black denote simulations that showed traces when productions were classified by both resulting articulation and selected phoneme. Dark grey denotes simulations that showed traces only when productions were classified by resulting articulation. Light grey denotes simulations where enough errors were produced to carry out the trace analysis, but no trace was found. White denotes simulations where not enough errors were produced to carry out the trace analysis.

We analysed further the simulations which showed traces when productions were classified by resulting articulation to determine whether they also showed traces when productions were classified by selected phoneme. We found 331 non-cascading model simulations, 716 cascading model simulations, and 1,461 cascading with feedback model simulations which showed traces whichever way the productions were classified.

Discussion

Although the proportions of simulations which show traces are higher when models include cascading, our results show that there are parameter setting combinations for the non-cascading model which result in traces of the intended phoneme being observable on erroneous productions.

So how do models produce traces? A number of simulations in all connectivity categories only showed traces when productions were classified by resulting articulation, and not by selected phoneme. This would imply that articulatory noise is largely the cause of traces in these simulations. This helps confirm that non-cascading models can indeed generate traces through articulatory noise. However, such performance is not in line with Goldrick and Blumstein's (2006) post-hoc analysis.

Traces that are observed when productions are classified by selected phoneme cannot, on the other hand, simply be attributed to articulatory noise. Analysis showed that correctly selected phonemes were significantly more activated than erroneously selected phonemes in 98.8% of analysable non-cascading simulations, supporting the activation levels hypothesis. However, only 13.0% of analysable non-cascading simulations produce traces when productions are classified by selected phoneme. Further analysis is required to determine which parameter settings permit activation level differences in phonological encoding to pass down to articulation.

Experiment 2

Experiment 2 was designed to investigate whether the models tested generalised to other types of speech error. When exploring such a large parameter space, we were keen to verify that the parameter settings of simulations which produced traces did not cause the models to exhibit other non-human like behaviour. In Experiment 2 we used evidence from corpora to determine reasonable bounds on human behaviour for a number of types of speech error, and compared the behaviour of the previously tested models to our bounds using a random two word production task. We then re-evaluated the results of Experiment 1 in the light of the new findings.

Method

Determining constraints Several corpora of speech error data were examined to determine bounds on the observable rates of anticipations, perseverations, exchanges and non-contextual errors in everyday human speech (del Viso, Igoa, & Garcia-Albea, 1991; Garnham, Shillcock, Brown, Mill, & Cutler, 1981; Nooteboom, 1969, 2005; Pérez, Santiago, Palma, & O'Seaghdha, 2007; Stemberger, 1989). Where counts for incomplete anticipations (such as *barn door* → *darn...*) were listed separately, these errors were split between the anticipation and exchange categories in the same proportion as completed anticipations and exchanges, as suggested by Stemberger (1989).

For each statistic, we calculated the mean value across corpora, and the standard deviation of the mean. We created a set of inner bounds, one standard deviation below and above the mean, and a set of outer bounds, two standard deviations below and above the mean. Since not all literature includes details of numbers of correct productions, the inner bounds for these were based on the highest and lowest percentages of correct productions seen across all participants in Goldrick and Blumstein (2006). This range was doubled to create the outer bounds. Where bounds exceeded 100% or fell below 0%, they were replaced with these cutoff values. All bounds are given in Table 2.

Procedure 10,000 random two word phrases from the network's vocabulary were chosen, subject to the constraint that the words had different onsets. The same set of 10,000 phrases was used for each simulation.

Results

For brevity, all results reported use classification by resulting articulation. Results using classification by selected phoneme do not differ from the reported results in any significant respect.

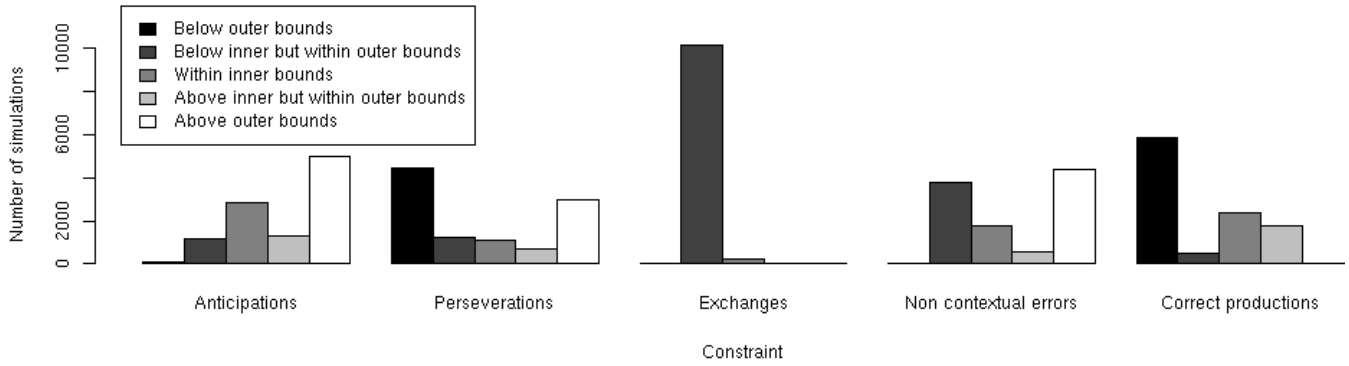

Figure 3: Results of Experiment 2. Constraints passed by all simulations which produced enough errors for trace analysis in Experiment 1. Note: 33 simulations did not produce any contextual errors for anticipation, perseveration or exchange analysis, and 5 simulations did not produce any errors for non-contextual error analysis.

Table 2: Percentage bounds on speech error production. The anticipation, perseveration and exchange constraints are given as percentages of all contextual errors. The non-contextual constraints are given as percentages of all errors. The correct production constraints are given as percentages of all productions.

	Outer lower bound	Inner lower bound	Inner upper bound	Outer upper bound
Anticipation	0.54%	23.96%	70.79%	94.20%
Perseveration	1.85%	13.22%	35.97%	47.34%
Exchange	0%	8.13%	47.93%	67.83%
Non-contextual	0%	5.22%	40.66%	58.38%
Correct productions	94.25%	96.0%	99.5%	100%

No simulations met the inner constraints. As is clear from Figure 3, a notable problem was exchanges. Only 247 simulations of the 10,446 model and parameter combinations which produced enough errors for trace analysis in Experiment 1 produced enough exchanges to fall within the inner bounds for exchanges, and 28 further simulations produced too many. The remaining 97% of the simulations produced too few exchanges.

373 simulations did however meet the outer constraints. 155 of these were non-cascading model simulations, while 114 were cascading model simulations, and 104 were feedback model simulations. Figure 4 shows the proportions of these simulations which displayed traces of the intended phoneme on erroneous productions in Experiment 1. When productions from Experiment 1 were categorised by resulting articulation, 44 non-cascading model simulations displayed traces, as well as 60 cascading simulations and 57 cascading with feedback simulations. When productions were categorised by selected phoneme, 43 non-cascading model simulations displayed traces, as well as 54 cascading simulations and 55 cascading with feedback simulations.

Discussion

Our results surprisingly showed that no parameter settings of the classical word production model (Dell, 1986) resulted in behaviour which met the inner constraints. It appears that low exchange error rates are a particular issue for this architecture. Closer examination of previous models shows that only 1 out of the 5 comparable simulations detailed in Dell (1986) and Hartsuiker (2002) produced sufficient exchanges (9% in the Dell, 1986 model replication; Hartsuiker, 2002). Part of the reason that this problem has previously gone unnoticed lies with researchers having frequently underestimated the proportions of exchanges in natural speech (see e.g. Nooteboom, 2005, for further details).

Of those simulations which pass the outer constraints, there are still non-cascading models which generate traces as in Experiment 1. This suggests that non-cascading models which produce traces need not exhibit behaviour which is less plausibly human than that of other architectures.

Across connectivity conditions however, there are now very few simulations which show traces when utterances are categorised by resulting articulation but not when utterances are categorised by selected phoneme. Closer examination of the data shows that 85.6% of the simulations previously thought to generate traces through articulatory noise produce too few correct utterances to fall within the outer bounds, and that 86.3% of these simulations produce too many non-contextual errors. This result supports Goldrick and Blumstein's (2006) conclusion that articulatory noise is not an adequate explanation for their data.

Conclusions

This paper has demonstrated that cascading from unselected phonemes to articulation is not necessary to explain voicing traces of intended phonemes on erroneous productions. This result underlines the importance of modelling to better understand the constraints placed on theories by empirical results. Investigation of model performance past this single phenomenon and across parameter settings has also shown that models based on the classic spreading activation account

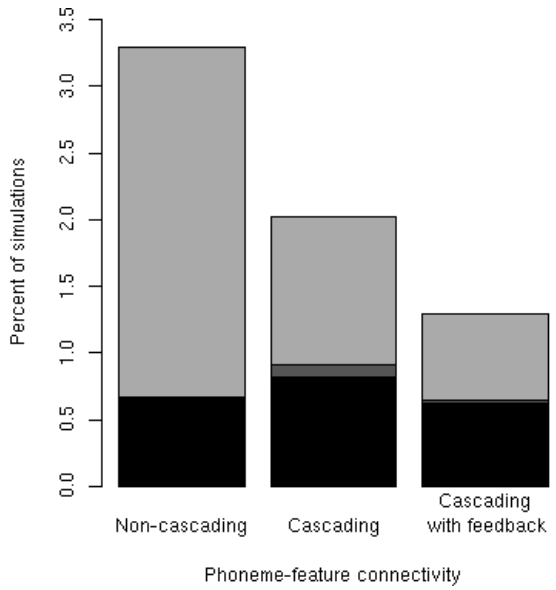

Figure 4: Results of Experiment 2. Simulations which met the outer constraints, as a proportion of all simulations run in each connectivity category. Portions of bars coloured in black denote simulations that passed the outer constraints and showed traces both when productions were classified by resulting articulation and by selected phoneme. Dark grey denotes simulations that passed the outer constraints but showed traces only when productions were classified by resulting articulation. Light grey denotes simulations which passed the outer constraints but which did not show traces.

of word production (Dell, 1986) may require modification in order for them to be able to accurately capture key human speech error patterns.

Acknowledgements

A portion of this work was carried out whilst the first author was visiting the University of Ghent. The first author is supported by a grant from the Economic and Social Research Council, United Kingdom.

References

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

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

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

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

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801–838.

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

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

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, *21*, 817–855.

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

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

Guenther, F. H. (2003). Neural control of speech movements. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities.* Berlin: Mouton de Gruyter.

Hartsuiker, R. J. (2002). The addition bias in Dutch and Spanish phonological speech errors: The role of structural context. *Language and Cognitive Processes*, *17*, 61–96.

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

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

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

Nooteboom, S. G. (2005). Listening to oneself: Monitoring in speech production. In R. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 167–186). Hove, UK: Psychology Press.

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

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

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

Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language*, *28*, 164–188.

Appendix: Words in the model's vocabulary

barb, bard, beep, bun, cad, cap, carl, cease, cell, come, cull, feat, fen, gap, gob, hack, hag, harsh, hock, horn, jam, jeff, kick, kit, knock, knot, lad, lash, lid, loll, love, luck, mart, park, pause, pig, pod, root, rot, seed, source, sup, tarn, teak, tech, teethe, thug, wring, youth, zoom