Holographic Reduced Representations for Oscillator Recall: A Model of Phonological Production

Harlan D. Harris (hharris@uiuc.edu)
University of Illinois at Urbana-Champaign
Department of Computer Science, MC-258
Urbana, IL 61801

Abstract

This paper describes a new computational model of phonological production, Holographic Reduced Representations for Oscillator Recall, or HORROR. HORROR's architecture accounts for phonological speech error patterns by combining the hierarchical oscillating context signal of the OSCAR serial-order model (Vousden, Brown, and Harley 2000; Brown, Preece, and Hulme 2000) with a holographic associative memory (Plate 1995). The resulting model is novel in a number of ways. Most importantly, all of the noise needed to generate errors is intrinsic to the system, instead of being generated by an external process. The model features fully-distributed hierarchical phoneme representations and a single distributed associative memory. Using fewer parameters and a more parsimonious design than OSCAR, HORROR accounts for error type proportions, the syllable-position constraint, and other constraints seen in the human speech error data.

Introduction

The phonological production subsystem is the part of the language production apparatus that sequences the sounds in individual words and groups of words. Phonological production is the mapping from lexical units, morphemes and words, to sequences of phonological units, phonemes. This paper presents a new model of the phonological production system, a model that offers a new explanation for errors and serial order in speech.

Speech Error Effects

Numerous constraints and patterns have been observed in speech error patterns, including error type proportions (see Table 1), the syllable position constraint, the C-V category constraint, the distance constraint, the phonological similarity effect, and the phonotactic regularity effect (Fromkin 1971). Unless otherwise specified, the numbers in the descriptions below are from the (Vousden et al. 2000) analysis of the (Harley and MacAndrew 1995) error corpus.

A strong constraint on movement errors (the first three error types in Table 1) is the *syllable position constraint*, or SPC. 89.5% of movement errors retain their position in the syllable (onsets move to onsets, vowels to vowels, etc.).

Type	Rate	Example
anticipations	35.1%	**d**et the dog
perseverations	26.0%	pet the **p**og
exchanges	10.6%	**d**et the **p**og
non-contextual slips	17.3%	pet the **l**og
mixed errors	11.0%	**l**et the **p**og

Table 1: Error type proportions. Target utterance is "pet the dog." Mixed errors include any error not in the other categories.

An even stronger constraint is the *consonant-vowel category constraint*, or C-V constraint. Errors very rarely involve the replacement of a consonant by a vowel or vice versa. A superset of these errors, those that violate language-specific rules (the *phonotactic regularity effect*), occur in less than 1% of errors (Stemberger 1983).

The *distance constraint* is the observation that phonemes tend to move only short distances (one or two syllables) in movement errors.

When movement errors occur, they are more likely than chance to involve phonemes that share phonetic features. For example, "pig bull" for the intended "big pull" is a more likely exchange than "bill pug," since [p] and [b] are more similar than are [g] and [l]. This is the *phonological similarity effect*.

Language Sequencing Models

Phonological production models can be categorized by how they generate serial order. I follow Vousden et al. (2000) and use the terms *associative chaining model*, *frame-based model*, and *control signal model*.

Associative chaining models account for serial order by having each subsequent phoneme be triggered by a combination of the pattern of previous phonemes and a representation of the target utterance (Dell, Juliano, and Govindjee 1993). These models successfully account for phonotactic regularity effects and the C-V constraint, but they do not generate anticipations and exchanges well, nor do they account for SPC effects.

Frame-based models (Dell 1986; Roelofs 1997) use strict phonological frames to slot phonemes into pre-

specified positions, such as the onset, nucleus, and coda positions of a syllable. These models often use chains of sequencing nodes to activate the slots sequentially (Eikmeyer and Schade 1991). Although frame-based models are influential, sequencing nodes are often criticized as being poorly motivated.

To address this point, control signal models (Burgess and Hitch 1992; Hartley and Houghton 1996; Vousden et al. 2000) replace discrete syllable frames with continuous time-varying signals. Prior to production, different parts of the word are associated with different parts of the signal. Then, as the signal changes during production, the associated phonemes are output sequentially. Simple control signal models explain how phonemes could be produced in order, but don't account for SPC effects.

The OSCAR model (Vousden et al. 2000), described below, is a complex control signal model that accounts for SPC effects by using a multidimensional control signal with biological motivation. It contains an implicit frame in the way that the control signal is structured, but does not require explicit slots or sequencing nodes for production.

Building Blocks

The HORROR model combines elements of two previously existing models: the OSCillator-based Associative Recall (OSCAR) model of serial-order and phonological production (Brown et al. 2000; Vousden et al. 2000), and the Holographic Reduced Representations (HRR) model of hierarchical associative memory (Plate 1995). Prior to describing HORROR, I review its two ancestral models.

OSCAR

OSCAR works by associating item vectors (phoneme representations) and phonological context vectors (PCVs) in a Hebbian associative memory. The PCVs are inspired by oscillating signals in the brain, and have an important hierarchical self-similarity pattern, described below. As the PCVs are iteratively presented to the associative memory, the original item vectors are recalled and become available for production. The self-similarity pattern generated by the oscillators, when combined with noise, generates patterns of errors that previously required the use of syllable frames.

In OSCAR, there are 30 oscillators in two groups of 15. In the non-repeating group, the oscillators generate sinusoidal values at frequencies ranging from very slow to very fast. Initial phases and frequencies are generated with sufficient randomness that the non-repeating group's state does not repeat for many steps. In the repeating group, the initial phases of the oscillators are random, but the frequencies are identical. The state of this group repeats precisely every three time steps, representing the period of a three-segment CVC syllable.

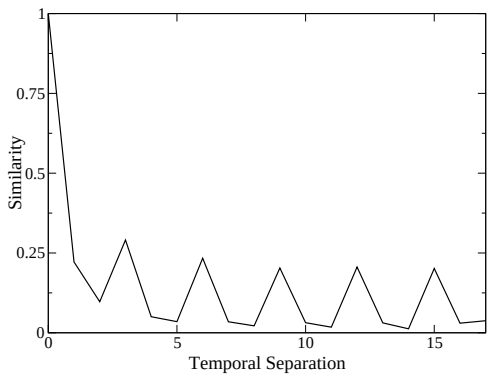

Figure 1: HORROR's PCV self-similarity function.

The PCV itself is generated by multiplying together selected oscillator signals to form a 32-element vector. Each element is a product of four oscillator signals, all of which are selected from the same group (repeating or non-repeating). The pattern of multiplications results in an automatically normalized PCV, allowing easy comparisons for similarity.

A key feature of OSCAR is that the PCV is self-similar in a hierarchical manner. Each state of the PCV is most similar to states that are multiples of three time-steps away, but nearby states are also somewhat similar (see Figure 1).

The process for producing a "word" (a randomly generated 18-segment sequence of six CVC syllables) is as follows. A PCV is initialized, and starts to change with time. At each time step, the PCV is associated with a phoneme feature vector in a Hebbian weight matrix. Each time step uses a *separate* weight matrix. This entire process is performed nine times (in parallel), to create a total of 81 weight matrices, nine replications of nine time steps. To produce the sequence, the PCV is re-instated to its initial state, then sequentially re-produces each step's state. The PCV is usually associated with the correct weight matrices to generate an approximation of the phoneme feature vector. In addition, a probabilistic process is used to generate errors. 70% of the time, segments which are associated with PCVs that are similar to the current PCV are combined with the output from the correct weight matrix. The result is an output vector, a potentially noisy approximation to the correct phoneme. Also, a post-output suppression mechanism is used to reduce excessive perseveration and facilitate exchange errors. The generated output vectors for each of the nine replications are compared to an item memory containing each phoneme, such that each phoneme is activated to an extent proportional to the similarity with the nine output vectors. The most active phoneme is then produced in a winner-take-all process.

OSCAR's Pros and Cons In many ways, OSCAR is important work in the literature of phonological production and speech error modeling, but it has significant problems that may limit its applicability. Its contributions include making good use of an independently-motivated context signal to create serial order, accounting for SPC effects without position-specific phonemes, and using an implicit rather than explicit syllable frame. Overall, it accounts for various error patterns better than do chaining models.

However, several limitations lead me to question the extent of the model's successes. Most importantly, the noise-addition procedure is unprincipled. As well, the artificial words did not include repeated phonemes, the associations between the context and phonemes are stored separately, and there are a concerning number of parameters.

Consider the noise-addition procedure. Cognitive models should use reasonable sources of noise to generate error phenomena. Many models add Gaussian noise, while others use intrinsic noise from distributed representations and imprecise network computation. Although OSCAR uses well-motivated oscillator signals to provide serial-order effects, its noise-generation procedures are much more weakly motivated. As described above and in Appendix C of Vousden et al. (2000), phonemes associated with states of the PCV that are selected by their similarity to the correct PCV are recalled in parallel and used to corrupt the winner-take-all process.

That this procedure generates impressive error results is not surprising. The noise in OSCAR is generated only by interference from particular phonemes in the current sequence, not by any sort of random numerical noise or other natural interference. OSCAR claims to explain why most errors are movement errors – in their model, it's because the generated noise *is* movement noise.

A related concern with OSCAR is that the associations between the PCV and phoneme vectors are stored separately. Although it is reasonable to use Hebbian learning to associate a PCV signal with phoneme representations, it is difficult to explain why each segment need be stored in entirely separate sets of weights. A more parsimonious solution would use a single set of weights and would treat the resulting noise as an asset, not a weakness.

HORROR adopts the oscillating PCV system from OSCAR, but replaces the movement-based noise-creation system with the noise inherent in an associative memory system with overlaid weights. It also uses a more parsimonious unified memory system, allows repeated phonemes within a sequence, and requires fewer free parameters[1].

[1]In addition to the five listed in Table 7 of Vousden et al. (2000), there are these four: the ratio of correct-to-incorrect activation, 0.6; the number of redundant associations, 9; the similarity threshold for allowing a phoneme to be added as noise, 0.5; and the similarity exponentiation factor in the item memory, 3.4.

	$\mathbf{T_1} = \mathbf{a} * \mathbf{b}$
$* : \mathbf{I} \times \mathbf{I} \Rightarrow \mathbf{T}$	$\mathbf{a} \# \mathbf{T_1} \rightarrow \mathbf{b} + \text{noise}$
$\# : \mathbf{I} \times \mathbf{T} \Rightarrow \mathbf{I}$	$\mathbf{T_2} = \mathbf{a} * \mathbf{b} + \mathbf{c} * \mathbf{d} + \mathbf{e} * \mathbf{f}$
$+ : \mathbf{T} \times \mathbf{T} \Rightarrow \mathbf{T}$	$\mathbf{d} \# \mathbf{T_2} \rightarrow \mathbf{c} + \text{noise}$
	$\mathbf{T_3} = \mathbf{g} * \mathbf{T_1} + \mathbf{h}$
	$\mathbf{g} \# \mathbf{T_3} \rightarrow \mathbf{T_1} + \text{noise}$

Figure 2: Holographic Associative Memory. $\mathbf{a} - \mathbf{h}$ are item vectors; $\mathbf{T_i}$ are memory vectors. $*, \#$, and $+$ symbolize circular convolution (encoding), correlation (decoding), and addition (composition).

Distributed Associative Memories

For several decades, mathematical psychologists have looked at distributed representations for models of memory (Murdock 1982; Eich 1982), and have accounted for many recognition and recall effects. Compared to localist connectionist models, where representations consist of features and microfeatures, distributed representations use long quasi-random vectors. These vectors are generated and manipulated such that similarity between two representations is defined by the dot product or cosine. Distributed representations can be combined in various ways. Two symbols may be associated by operations such as convolution or the outer product, resulting in another large vector. Retrieval from memory vectors is performed by inverting the association operation, correlation. Distributed memories can store a number of associations at once, simply by adding the vectors together. As vectors are overlaid, the amount of noise increases. This intrinsic noise is part of the model, and resulting simulations can account for list-length and item-similarity effects.

A limitation in much of the work on distributed memories is that the operations that generate associations greatly expand the size of the vector, with the result that hierarchies of associations are impractical. HORROR utilizes one of several approaches that overcome this problem, the Holographic Reduced Representations (HRRs) of Plate (1995).

With HRRs, the representations and associations are always fixed-length vectors. A circular version of convolution is used to associate vectors. The resulting memory vectors are the same length as the input vectors, at the cost of increased noise. The greatest benefit is that hierarchies of associations can be easily generated and stored. See Figure 2 for simple examples and notation. An auto-associative item memory (a Hopfield network or a nearest-neighbor search through a list) is necessary to identify the result of each correlation.

HORROR

HOlographic Reduced Representations for Oscillator Recall (HORROR) is a model of serial-order processing that combines the self-similar context vectors of OSCAR with the hierarchical representations and memory of HRRs. The result is a fully-distributed phonological production model that accounts for errors in the serial-order part of the system by using the intrinsic noise from the associative-memory part of the system.

OSCAR and HRRs fundamentally both represent similarity by distance between vector representations. In OSCAR, the extent to which pairs of context vectors which are near in time are also near in space determines retrieval accuracy and error patterns. With HRRs, capacity and noise levels are determined by the extent to which composed vectors are near (not orthogonal) to each other. In OSCAR these similarity metrics can be complex and hierarchical, determined by the oscillator patterns, and in HRRs, the similarity metrics can also be hierarchical, by the process of overlaying associations. In both models, item memories are used to clean up and to select a single item.

HORROR is a new model based on the general framework of OSCAR. It takes a variation of the PCV from OSCAR, and combines it with an HRR associative memory, replacing the simple associative memory used by OSCAR. In addition, the feature-vector phonological representations used in OSCAR are replaced with fully-distributed hierarchical representations in HORROR. A critical aspect of HORROR is that all of the phoneme-context pairs that make up a sequence are stored together in a single large vector, rather than in OSCAR's many separate weight matrices. The noise in this memory vector, combined with representational similarity and the PCV structure, provide sufficient opportunities for appropriately distributed error patterns to arise.

Experiments

A major goal of this work is to account for the same human speech error data as does OSCAR, using a simpler structure, more parsimonious procedures, and fewer parameters.

As with OSCAR, PCVs are generated sequentially and convolved with phoneme representations to form memory traces. Unlike OSCAR, these traces are summed to form a single vector representing the entire sequence. To produce the sequence, the vector is correlated with the PCVs in order, resulting in noisy versions of the phonemes. The phonemes are cleaned up in an item memory, and the results are analyzed for various types of errors.

The oscillators used to generate the PCV were the same as used by OSCAR. HORROR additionally includes a parameter, *nrep*, that specifies the proportion of repeating versus non-repeating oscillators. The procedure of generating the PCV from the oscillators in HORROR is very similar to the procedure used in OSCAR, but since HORROR's PCV is very wide (2048 elements), the process was repeated with different random initial phases and frequencies in order to fill up the vector, which was then normalized. See Table 2 for the list of PCV and other parameters used in the experiments described below.

Param.	Value	Description
nrep	17	# of non-repeating oscillators
vw	2048	Representation vector width
cc	3	Repeating oscillator inv. freq.
D	4	Speech-rate (larger = slower)
Inhib	.121	Post-activation inhibition level
InDec	.5	Inhib. decay (lower = faster)
ds	3	Phoneme dis-similarity factor

Table 2: Free parameters in HORROR

Vousden et al. (2000) use an articulatory-feature representation of phonemes. Each phoneme is 17 elements long, with binary features representing place and manner of articulation, nasality, voicing, and vowel position and tenseness. We converted these localist features into distributed features for the fully-distributed representations used in HORROR.

Phonological representations were built in a fully-distributed manner by generating random Gaussian vectors (of width *vw*) for each feature, then summing the appropriate features together and normalizing. Each vector thus has an intrinsic similarity metric, defined by the number of shared features. In order to partially "drown out" the similarity between otherwise very-similar phonemes, additional random vectors (*ds*) were added to each phoneme vector.

Decoding consists of sequentially correlating each time-step of the PCV with the single stored memory vector. The result is a series of approximations to the target phonemes, corrupted by the noise intrinsic to a holographic memory. Each recalled vector is compared to an item memory containing possible phonemes. The phoneme that is most similar to the recalled vector is then produced.

The item memory has three features that help it best account for the error patterns. First, each item in the item memory has a persistent activation level, a. Activation is added to similarity to determine which phoneme is selected. At each step, each item's a is increased by the item's distance from the recalled vector, weighted by the $Inhib$ parameter. Second, after a phoneme is selected, it is suppressed by setting a to be the negation of $Inhib$. Post-output suppression is a common feature of this type of model (Vousden et al. 2000; Dell 1986). Finally, at every time step, activation decays toward zero according to the decay constant $InDec$.

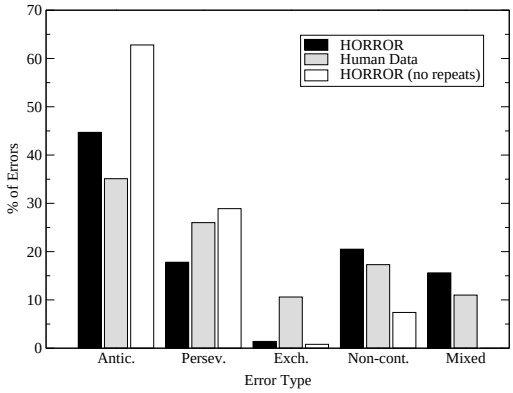

Figure 3: HORROR's error type proportions.

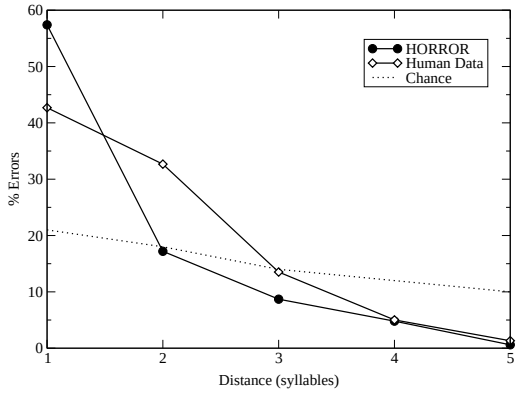

Figure 4: Distance gradients of the anticipation errors produced by HORROR, compared with the human data and chance baseline of Vousden et al. (2000). Adjacent syllables have a distance of 1. Same-syllable errors (separation 0) are not shown.

Experimental Results

2000 6-syllable "words" were generated, associated, and output. Errors were determined by an automatic categorization process. Error type proportions, SPC violations, distance constraint statistics, and phonetic similarity constraint statistics were counted.

Error proportions 1538 errors occurred during production of 36,000 segments, resulting in an overall error rate of 4.3%. Figure 3 shows the proportions of error types. The results compare fairly well with the human data reported in Vousden et al. (2000). Exchanges, however, were under-represented in the model, raising the question of whether HORROR's exchanges are true exchanges or merely the joint event of independent anticipations and perseverations. Other speech error models (Dell et al. 1993; Roelofs 1997) are unable to produce true exchanges, and are therefore seen as incomplete.

To address this, I calculated the expected number of exchanges, assuming that they are coincidental. This number, 0.33 per 2000 sequences, was more than sixty times smaller than the number of exchanges actually observed (22), demonstrating a true tendency for exchanges. Exchanges in HORROR occur because post-activation inhibition helps to prevent an erroneously anticipated phoneme from then appearing in its correct location. Instead, the earlier, replaced phoneme may be triggered via the PCV, turning an anticipation into an exchange.

Distance constraint The model's movement gradients parallel the distance constraints seen in human data, with disproportionately small separations. Exchange errors shifted least, an average 0.95 syllables, followed by anticipations, averaging 1.4 syllables, and perseverations, averaging 2.9 syllables. Figure 4 shows a comparison on anticipations between HORROR and human data. The shorter movements made by exchanges, compared to other error types, has been observed in human error data (Nooteboom 1973).

Error type	n	Mean shared features
Movements	895	2.4
Exchanges	22	3.1
Non-contextual	306	3.1
Chance		1.9

Table 3: Average similarity for consonant errors. Movements are anticipations and perseverations.

Phonetic similarity constraint Vousden et al. (2000) concentrate their analysis of the phonetic similarity constraint on consonant exchanges. HORROR produced only 22 consonant exchanges, 20 of which shared 75% or more of their phonetic features. Table 3 compares the categories of consonant errors to chance. Chance was determined by randomly selecting 1000 pairs of consonants, and counting the number of shared phonemes for each pair. The phonetic similarity constraint is clearly present in these results. Note that exchange errors were significantly more similar than were other movement errors. This is true for human exchanges, and also lends further support to the observed exchanges being real.

Syllable-position constraint 29.0% of the model's errors violated the SPC, compared to 10.5% of errors in human data (Vousden et al. 2000). To confirm that this number still reflects a constraint, and is not just the chance rate of violations, it's necessary to look at the probabilities of errors being in each syllable position. In this set of data, 50.7% of errors were in the onset, 8.6% in the vowel, and 40.8% in the coda. To

calculate the expected rate of SPC violations, assume that the consonant-vowel constraint is never violated, and that consonant errors have a 50% chance of movement from onsets and codas. Therefore, the expected SPC violation rate is $1 - (.086 + .507 * .5 + .408 * .5) = 45.7\%$. Although the SPC is violated more often by the model than it is in human data, it is still a real effect.

Consonant-vowel constraint Only 2.3% of the errors violated the C-V constraint, showing that the model is generally respecting the consonant-vowel categorical distinction seen in natural errors.

Repeated phonemes In order to investigate the role of repeated phonemes in the model, the same experiment was re-run with repeated phonemes disabled. Since repeated items are known to strongly affect performance in distributed memories, it was expected that the effects on HORROR would be significant as well. The error rate without repeated items was reduced to 1.3%, and the proportion of non-contextual errors was greatly reduced (see Figure 3). HORROR is more error prone when repetitions occur, as in human data (Dell 1986). Repeated phonemes appear to be an important trigger for speech errors, including non-contextual errors.

Discussion

The HORROR model combines the best features of OSCAR, a serial-order phonological model with a hierarchical context signal, and HRRs, a holographic associative memory using hierarchical representations. Its aim is to account for speech error patterns using more parsimonious mechanisms than previous related models.

HORROR succeeds in a number of ways. It allows repeated phonemes in the sequences, it combines associative memory traces into a single distributed association vector, and its error mechanism relies entirely on the intrinsic noise from the associative memory with no generated noise at all. It uses fewer parameters than does OSCAR, and accounts for a number of error patterns in human data. Specifically, the model's error type proportions, distance constraint, phonological similarity constraint, and C-V category constraint results were largely similar to human data. The SPC results were real, if modeled less accurately. HORROR accounts for these major speech error patterns by using fully-distributed hierarchical representations, a single intrinsically-noisy associative memory, and an oscillating phonological context signal.

Acknowledgments

I appreciate the advice and contributions of Gary Dell, Janet Vousden, and Franklin Chang. This work was supported by NSF grant SBR 98-73450 and NIH grant DC-00191.

References

Brown, G. D. A., T. Preece, and C. Hulme (2000). Oscillator-based memory for serial order. *Psychological Review 107*(1), 127–183.

Burgess, N. and G. J. Hitch (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language 31*, 429–460.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review 93*(3), 283–321.

Dell, G. S., C. Juliano, and A. Govindjee (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science 17*, 149–195.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review 89*(6), 627–661.

Eikmeyer, J.-J. and U. Schade (1991). Sequentialization in connectionist language-production models. *Cognitive Systems 3*(2), 128–138.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language 47*(1), 27–52.

Harley, T. A. and S. B. G. MacAndrew (1995). Interactive models of lexicalization: Some constraints from speech error, picture naming, and neuropsychological data. In D. Bairaktaris, J. Bullinaria, and D. Cairns (Eds.), *Connectionist models of memory and language*. London: UCL Press.

Hartley, T. and G. Houghton (1996). A linguistically restrained model of short-term memory for non-words. *Journal of Memory and Language 35*, 1–31.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative recall. *Psychological Review 89*(6), 609–626.

Nooteboom, S. (1973). The tongue slips into patterns. In V. A. Fromkin (Ed.), *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Plate, T. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks 6*(3), 623–641.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition 64*, 249–284.

Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review*. Bloomington: Indiana University Linguistics Club.

Vousden, J. I., G. D. A. Brown, and T. A. Harley (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology 41*, 101–175.