Errors in Speech Production: Explaining Mismatch and Accommodation

Andrea Gormley (agormley@carleton.ca)
Institute of Cognitive Science, Carleton University
Ottawa, ON K1S 5B6 Canada

Terrence C. Stewart (tcstewar@uwaterloo.ca)
Centre for Theoretical Neuroscience, University of Waterloo
Waterloo, ON N2L 3G1 Canada

Abstract

The study of errors allows researchers insight into the production of speech. Speech errors have been shown to accommodate in form to their erroneous environment, demonstrating that errors occur before the processing of the phonological rule component. That this configuration is a complete picture of the processing involved, however, has been called into question by the prevalence of non-accommodated errors that have been detected via instrumental analysis (Gormley 2008). This paper presents a model of speech production developed using Python ACT-R (Stewart & West, 2007a) that uses a noisy recall system and explicit encoding of phonological rules. This system produces both accommodated and mismatch speech errors at the same rates as observed in the empirical study.

Keywords: Speech errors; phonology; ACT-R; equivalence

Introduction

Transcription studies of speech errors demonstrate a phenomenon wherein a speech error tends to become phonologically accommodated in its erroneous context. For example, if a speaker makes the error of saying *an apple* instead of *a pear*, the determiner *an* is used before the vowel-initial *apple* even though consonant-initial *pear* was intended. This is in contrast to a non-accommodated, or mismatch error, where the speaker would say *a apple*. Accommodation is generally thought to be the highly prevalent across a variety of different types of speech errors (Boomer & Laver, 1968).

However, the rate of occurrence of mismatch errors may be significantly underestimated by transcription studies. By analyzing the wave forms of produced speech, instrumental analysis of speech errors shows that mismatch errors occur more often that accommodated ones (Gormley, 2008). This calls into question standard models of speech production, since they do not exhibit this effect. This paper presents a model of speech production developed using Python ACT-R (Stewart & West, 2007a) that uses a noisy recall system and explicit encoding of phonological rules. This system produces both accommodated and mismatch speech errors at the same rates as observed in the empirical study.

Speech Errors

An instrumental analysis of speech errors was conducted to re-address the question of phonological accommodation. Based on the methodology of Goldrick and Blumstein (2006), thirty-two non-word tongue twisters were designed that would induce voicing errors on the final consonant of a syllable (i.e. the coda). Sample tongue twisters are shown in Table 1, each of which follows an A B B A pattern, where A and B are identical other than one having a voiced final consonant and the other having a voiceless final consonant. Participants were recorded repeating each tongue twister three times.

Table 1: Sample non-word tongue twisters

tiff tivv tivv tiff
kess kezz kezz kess
tuzz tuss tuss tuzz
kavv kaff kaff kavv

In English, vowels are lengthened before voiced codas. This means that vowel length can be measured to see if phonological accommodation has occurred. To determine the expected length for voiced and voiceless codas, each participant also provided a control condition where the same word was repeated over and over.

Transcription studies on phonological accommodation are at a disadvantage given the difficulties in perceiving errors at a phonetic level. Acoustic analysis removes this perceptual issue, yielding a more objective result. Given that errors may occur at all levels of speech production; semantic, morphological and phonological, it would not be surprising to find that errors can occur after phonological processing, before the formation of an articulatory plan.

Unlike other speech error studies, all data were analyzed, not just those tokens that were perceived by the researcher as errors. This method has been used by Frisch and Wright (2002) and eliminates the perceptual bias inherent in the perception of speech. For each participant, measurements were made for all fricatives and vowels in the control condition. Measures of percent voicing were taken for all coda fricatives and durations were measured for all vowels for each participant. The initial consonant did not make a significant difference in vowel duration or voicing of codas and were not analyzed separately.

An error was defined as a situation where the participant produced a token with a vowel duration or fricative voicing more than two standard deviations away from the mean for the control condition. Each participant's statistics were evaluated separately. All tokens in the experimental

condition were categorized as normal or erroneous based on this range, as illustrated in Figure 1.

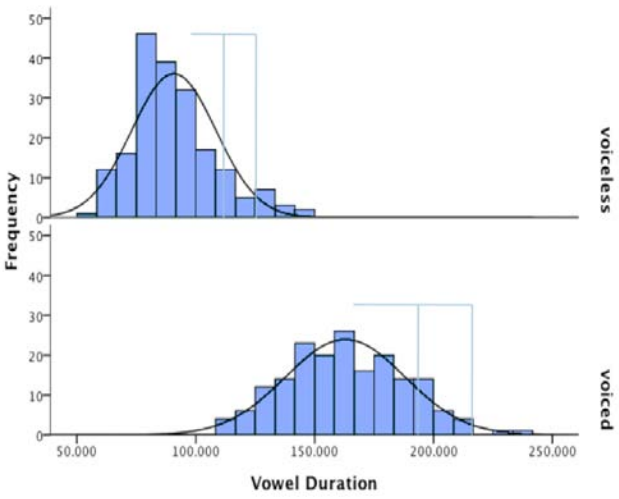

Figure 1: Example data where mean duration for vowels before voiceless codas is 91 ms (s.d. 18 ms). The mean vowel length before voiced codas is 163 ms (s.d. 25 ms).

Tokens with values that were more than two standard deviations from the mean were considered errors. Mismatch errors were defined as when the vowel length did not correspond to the voicing value of the coda fricative, such as when the token *tiff* was determined to have a long vowel and the coda fricative was determined to be voiceless. There was a considerable grey area in the determination of the range of normal values for each participant. Because values for the categories voiced and voiceless as well as vowel duration are on a continuum, a vowel's duration could fall within a normal range for either long or short, as is shown in Figure 2. These indeterminate cases were not classified.

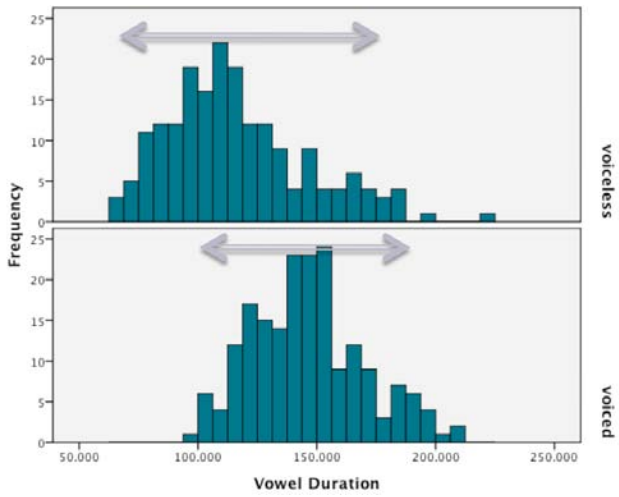

Figure 2: Example where mean duration for short vowels is 116 ms (s.d. 30 ms). Long vowel duration is 146 ms (s.d. 24 ms). There is considerable overlap.

As shown in Table 2, of the 4790 tokens analyzed, 355 are mismatch errors, 66 are accommodated and 451 are unclassified, yielding 872 total errors. Also reported are the 95% confidence intervals for these measures, which are used to evaluate our computational models.

Table 2: Error types

	N	%	95% Confidence Interval	
Correct	3918	81.8	80.7	82.9
Mismatch	355	7.4	6.7	8.2
Accommodated	66	1.4	1.1	1.7
Unclassified	451	9.4	8.6	10.3
Total Errors	872	18.2	17.1	19.3

The results of this experiment show that not all errors accommodate to the erroneous environment. By analyzing all tokens and determining the normal range of production for each participant, an unbiased view of speech errors is obtained. That errors can be phonologically inappropriate to their environment shows that phonological rules are not consistently processed after the error is made. This suggests that the common assumption made by psycholinguists that speech errors must occur before the phonological processing component is not the complete story.

To investigate possible mechanisms to explain these mismatch errors, we constructed two separate computational models. The first considers the possibility that errors can result from the retrieval of information after the phonological rules have been applied, and the second considers errors in applying the rules themselves.

Errors Before and After Rules

In the standard conception of speech production, the speech planning system produces an ordered sequence of phonemes to be uttered. A set of phonological rules are applied to these phonemes to produce the output. Any errors are due to problems with the speech planning system, such as the misordering of phonemes in a string. The phonological rules, however, are assumed to be applied perfectly.

Such a model only accounts for accommodation errors, not mismatch errors. Given the empirical data in Table 1, there must be a mechanism for introducing these errors. One possibility is that, after the rules are applied, errors are introduced while taking the results of the rules and producing an output.

For our model, the non-words are input into the system as a string of phonemes. Each phoneme's features are then retrieved from a feature store. These features are combined with the syllabic and order information in the input and placed in what we call the *first speech planning module*. This is an extremely short-term memory that merely stores the information and makes it available to the phonological rule module, which makes the relevant transformations. Any errors in retrieving information from the feature store

or in passing the information to the phonological module will result in accommodation errors, as the incorrect information will go through the phonological rule section yielding a form with the appropriate phonology.

To apply the rules, the phonology module retrieves each segment from the first speech planning module and performs the rule of vowel lengthening where applicable. The results of applying these rules are passed to the *second speech planning module*. Once it is stored here, it is outputted in sequence based on the stored order information. This retrieval from the second speech planning module is where mismatch errors occur. Since the phonological rule has already applied, the retrieval of an *s* instead of a *z* will yield the form *tiis* with a long vowel that is inappropriate before the voiceless *s*.

Incorrect retrievals from the feature store and speech planning are not random. Given that speech errors often confuse elements that are similar in form and structure such as two onsets or two alveolar stops, the model is constructed to favour this type of confusion. Two onsets are more likely to be confused than an onset and a nucleus.

Table 3: Example of correct processing

Step		Stored Value
1	Input phonemes	tiztis
2	Retrieve features	tiztis
3	Phonological rule	tiiztis
4	Output	tiiztis

Table 4: Example of processing an accommodation error

Step		Stored Value
1	Input phonemes	tiztis
2	Retrieve features (error: got features for *z* instead of *s*)	tiztiz
3	Phonological rule	tiiztiiz
4	Output	tiiztiiz

Table 5: Example of processing a mismatch error

Step		Stored Value
1	Input phonemes	tiztis
2	Retrieve features	tiztis
3	Phonological rule	tiiztis
4	Output (error: got *i* instead of *ii*)	tiztis

Cognitive Architectures

There are a vast selection of possibilities for developing a computational model for this task. While it would certainly be possible to develop a special-purpose model from scratch to exhibit the desired behaviour, we instead chose to base our model on existing cognitive theory. In particular, the general cognitive architecture ACT-R (Anderson & Lebiere, 1998) has been applied to a wide variety of psychological tasks. However, few of these applications have been in the domain of linguistics, and it has not been previously applied to speech production.

To apply this architecture to this novel domain, we followed the approach of Stewart and West (2007a). Here, the particular components of ACT-R are treated as general modules that can be re-purposed for different tasks. Instead of assuming there is just one central production system for ordering event, we consider that the ACT-R production system may be a suitable model for many different separate aspects of cognition, all happening in parallel. This is also consistent with the Massive Redeployment Hypothesis (Anderson, 2007), which argues that once a cognitive component has been developed, evolution is likely to redeploy that same component for multiple purposes, if it is a suitable system. This is much more efficient than evolving a new system for every new capability.

The first generic component from ACT-R is a storage system for symbolic information. Information (known as *chunks*) consisting of an ordered list of symbols can be placed into the memory and retrieved at a later time. This memory is not perfect; it will sometimes fail to retrieve a chunk and sometimes a different chunk than the one intended will be returned. This has been used for a broad range of explicit and implicit memory tasks, and is based on the general principle that the odds of a memory being needed decay as a power law over time. This principle is a close match for realistic human cognitive environments (Anderson & Schooler, 1991).

To implement this, each item i in memory is given an activation level A_i, calculated using Equation 1, where t_k is the amount of time since the k^{th} appearance of this item, d is the decay rate, and $\varepsilon(s)$ is a random value chosen from a logistic distribution.

$$A_i = \ln \sum_{k=1}^{n} t_k^{-d} + \varepsilon(s) \quad (1)$$

When attempting to recall an item that matches a given pattern, the activity level of each potential answer if calculated and the one with the highest A_i is selected.

The second generic component is a production system: a set of rules that identify what action should take place in a given condition. These actions are not overt physical actions. Instead, they are internal actions which may create new chunks, request the retrieval of chunks, change chunk values, and so on.

The ACT-R cognitive theory also specifies how long these components take to perform various tasks. Production rules in their normal context of the central executive for cognition are though to require 50 milliseconds to fire, while memory recall time is proportional to e^{-A}. For a phonological model we must assume much faster processing, but these details are not considered here.

Model Construction

To construct our model, we used the two basic components from ACT-R: the memory module for storing symbolic information and the production system for defining the sequence of events that should occur within the model. In contrast with standard ACT-R, four separate special-purpose memories were defined, each of which can act in parallel. The feature store is a long-term static memory holding the particular features of each phoneme. The first and second speech planning module are short-term memories holding the phonemes currently being processed. Finally, the rule memory stores the rules to be applied.

It should be noted that we are not assuming these components are identical to the standard ACT-R production system and declarative memory system. In particular, they must be much faster in order to produce speech effectively. We hypothesize that, while they maybe optimized to perform phonological tasks, they can still be thought of as special cases of these generic cognitive components.

For this model, there are two sources of randomness that can affect behaviour: the first and second speech planning modules. Each of these is hypothesized to store phonemes and their feature information. ACT-R provides two methods for configuring the recall error from such a memory. First, there is random fluctuation of the activity levels of a chunk (*s* in Equation 1). Second, a similarity score can be set between values to indicate that, for example, a retrieval attempt for *red* would only receive a small activation penalty for retrieving *pink* instead. We use this method here to indicate the level of similarity between phoneme pairs such as *z* and *s*.

This results in a complex model, but with only two free parameters. All other parameters are fixed due to the choice of using the ACT-R components as a general-purpose cognitive architecture. Importantly, the accommodation and mismatch error rates are not independently adjustable in our model. Instead, they arise from an interplay between these parameters and the overall system.

Modelling Results

To determine the accuracy of our model, we compared it to the results shown in Table 2. The first requirement is that the model's overall rate of error is comparable to that of the participants. From Table 2 we note that the proportion correct rate is between 80.7% and 82.9% (95% confidence intervals). However, there were also a large number of unclassified responses, due to the inability to distinguish phonemes via acoustic analysis. The most conservative possible assumption is that the actual proportion correct is between 80.7% and 93.2%.

The model's proportion correct is shown in Figure 3. This proportion changes as the two parameters for the model are adjusted. As can be seen, the model makes no errors when there are low amounts of noise, and more errors are seen with more noise and a higher similarity between matched pairs of phonemes (such as *s* and *z*).

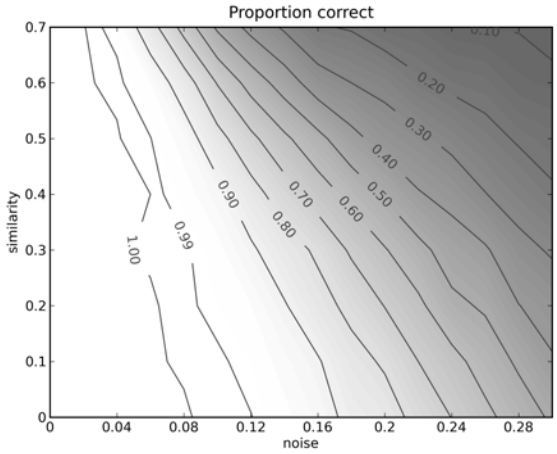

Figure 3: Proportion of responses for the first model that have no speech errors. Each point shows the behaviour of the model for different noise and similarity values.

The second requirement is that more mismatch errors be produced than accommodation errors. From Table 2, this value should be between 0.05% to 0.071%, although this does not take into account the unclassified errors.

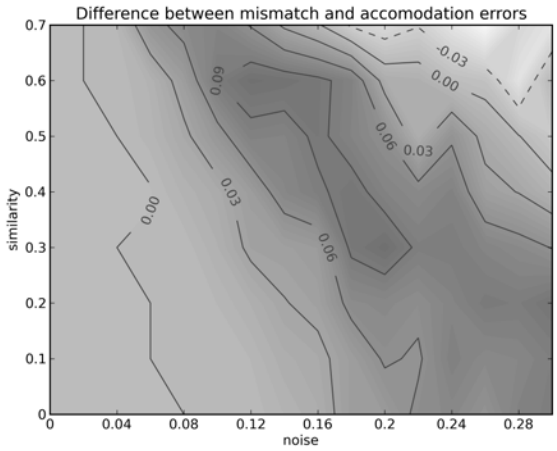

Figure 4: The proportion of mismatch errors minus the proportion of accommodation errors for the first model.

It is common practise in modelling research to identify a parameter setting that has the minimum mean squared error with respect to the human data. However, this merely measures prediction error rather than determining whether the model's behaviour is statistically distinguishable from the empirical data. Instead, we applied the Relativized Equivalence measure (Stewart & West, 2007b). This produces a value below 1 if the difference between the model's behaviour and the participant's behaviour is less than the size of the empirical confidence interval for every measure. In other words, if the relativized equivalence is below 1, then the model's results are statistically indistinguishable from the empirical data. This is an extremely conservative metric for evaluating a model.

The relativized equivalence for our first model is shown in Figure 5. All parameter settings inside the contour line labelled 1.0 are models whose behaviour matches that of the participants. To further distinguish these models, more accurate empirical results are needed.

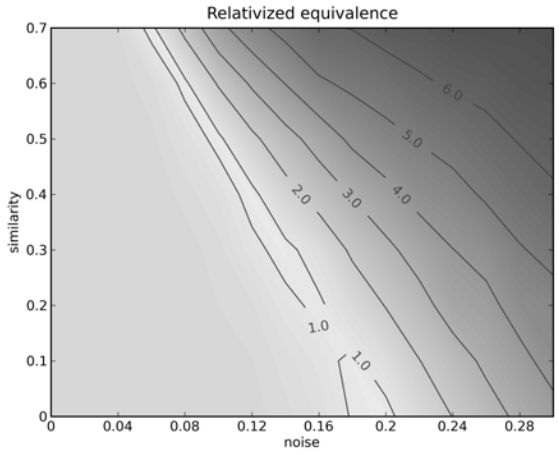

Figure 5: Relativized equivalence for the first model as noise and similarity are adjusted. All parameter settings inside the 1.0 contour are models that statistically match the participant's behaviour.

Errors During Phonological Processing

The second possible explanation for mismatch errors is that they can arise during the processing of phonological rules. Phonologists do not tend to study errors, and as a result the idea that errors can arise from within the phonological component itself has not been widely explored. What if the notion that errors can arise during the processing of phonological rules was considered? What would an error of this type look like? The answer is that an error of phonological rule processing, that is the failure of a phonological rule to apply, would look like a mismatch error. The second model presented here represents this option.

This model introduces error into the application of rules itself. It is identical to the first model, except that random noise has been added when retrieving the rule to be applied. If this retrieval fails, the phonological rule will fail to apply, leading to a mismatch error.

Rather than introducing a new parameter to adjust the probability of a rule failing, we used the same value as the random noise in the rest of the model. This keeps our model as simple as possible, rather than adding new independently adjustable parameters.

Table 6: Example of processing an accommodation error

Step		Stored Value
1	Input phonemes	tiztis
2	Retrieve features (error: got features for *z* instead of *s*)	tiztiz
3	Phonological rule	tiiztiiz
4	Output	tiiztiiz

Table 7: Example of processing a mismatch error

Step		Stored Value
1	Input phonemes	tiztis
2	Retrieve features	tiztis
3	Phonological rule (error: the rule fails to apply)	tiztis
4	Output	tiztis

It should be noted that this model introduces an asymmetry, since the errors generated by this model will appear as under-applying rules rather than over-applying rules. That is, there will not be situations where the rule applies when it should not (as in lengthening the vowel in *tis* to *tiis*). Interestingly, when the mismatch errors from the empirical study were examined, a strong tendency was found for *tiiz* to be pronounced *tiz* more often than *tis* was pronounced tiis (see Table 8).

Table 8: Asymmetry of vowel length mismatch errors among classified tokens.

	N	%	95% Confidence Interval	
tiiz as *tiz*	102	2.13	1.76	2.58
tis as *tiis*	38	0.79	0.58	1.09

Modelling Results

The same analysis was performed on the second model as on the first. Figure 6 shows the proportion correct, which is slightly lower that the first model at the same parameter settings, as there are three sources of noise.

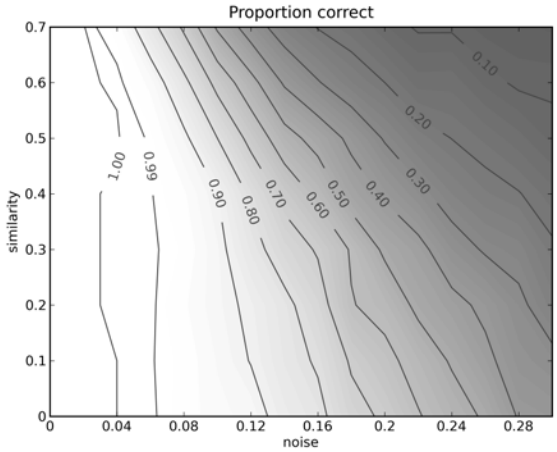

Figure 6: Proportion of responses for the second model that have no speech errors.

To examine the effect of the asymmetry where rules fail to apply, Figure 7 indicates the difference in occurrence rates between mismatch errors where *tiiz* is pronounced *tiz* and where *tis* is pronounced *tiis*. The first situation can be

caused by either a rule failure or a error after the rule, while the second situation is only caused by errors afterwards.

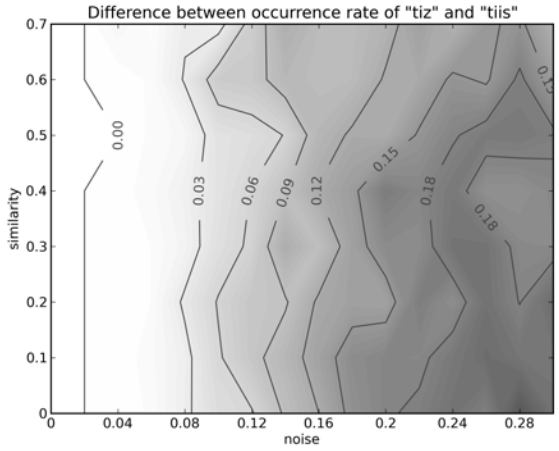

Figure 7: The rate of *tiiz* being pronounced *tiz* minus the rate of *tis* being pronounced *tiis* for the second model.

To evaluate the overall performance of the second model, we repeated the equivalence measure, with the inclusion of the rates of the two types of mismatches given in Table 8. The result is shown in Figure 8, indicating that the second model also has parameter settings which make it indistinguishable from the empirical data. This is the set of parameter values inside the 1.0 contour line. Interestingly, this is a wider range of parameter values than was seen in our first model.

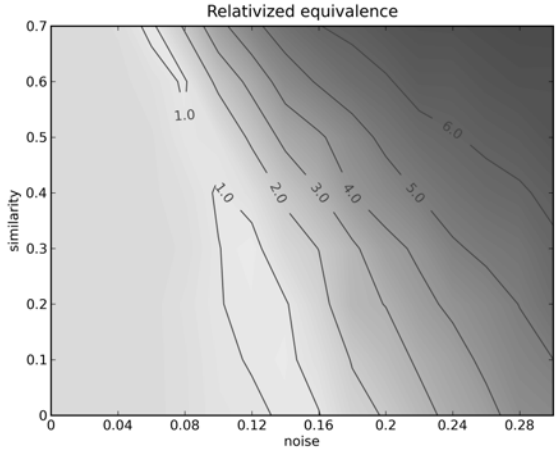

Figure 8: Relativized equivalence for the second model as noise and similarity are adjusted. All parameter settings inside the 1.0 contour are models that statistically match the participant's behaviour.

Conclusion

The tongue twister study presented here shows that the surface phonology does not always conform to its erroneous environment. This result can be interpreted in two ways. First, this could show that phonological rules can, contrary to previous assumptions, be processed before speech errors occur. Second, and more controversially, that the errors can be due to the phonological rule component itself as the result of a rule failing to apply.

Two models were created to evaluate these possibilities. While they both could match the overall pattern of human responses, only the second one (where rules could fail to apply and speech errors could occur after the rules) captured the asymmetry in the empirical results.

This result suggests a fruitful new area for speech production research. The failure of phonological rules to apply is needed to fully capture human performance in these domains.

References

Anderson, J. R. and Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396-408.

Anderson, M. (2007). Massive redeployment, exaptation, and the functional integration of cognitive operations. *Synthese, 159*(3), 329-345.

Boomer, D., & Laver, J. (1968). Slips of the tongue. *British Journal of Disorders of Communication, 3*, 1-12.

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.

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.

Gormley, A. (2008). The Relationship between Phonology and Speech Planning. Paper presented at *Fifth North American Phonology Conference*. Concordia University, Montréal, QC.

Stewart, T.C., West, R.L. (2007a) Deconstructing and Reconstructing ACT-R: Exploring the Architectural Space. *Cognitive Systems Research. 8*(3), 227-236.

Stewart, T. C., West, R.L. (2007b) Equivalence: A Novel Basis for Model Comparison. *29th Annual Meeting of the Cognitive Science Society*.