

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**  
Progress Report No. 22 (1998)  
*Indiana University*

**The Phonetics of Phonological Speech Errors:  
An Acoustic Analysis of /s/ and /z/ Errors by Four Talkers<sup>1</sup>**

**Stefan Frisch<sup>2</sup> and Richard Wright<sup>3</sup>**

*Speech Research Laboratory  
Department of Psychology  
Indiana University  
Bloomington, Indiana 47405*

---

<sup>1</sup> This work was supported by NIH-NIDCD Training Grant DC00012 to Indiana University. Parts of this paper were presented at the 133<sup>rd</sup> and 134<sup>th</sup> meetings of the Acoustical Society of America, the 72<sup>nd</sup> annual meeting of the Linguistic Society of America, and the University of Illinois at Urbana/Champaign. We would like to thank those audiences for their insightful questions, extensive comments and helpful suggestions.

<sup>2</sup> Now at the University of Michigan, Ann Arbor, MI.

<sup>3</sup> Now at the University of Washington, Seattle, WA.

## **The Phonetics of Phonological Speech Errors: An Acoustic Analysis of /s/ and /z/ Errors by Four Talkers**

**Abstract.** Phonological speech errors are a traditional source of evidence for phonological units like features and segments in both phonological and psycholinguistic theory. Assumptions about the unitary nature of speech errors have been brought into question by research that found evidence for sub-featural errors in speech production [Mowrey & MacKay (1990), Phonological primitives: Electro-myographic speech error evidence, *Journal of the Acoustical Society of America*, 88(3), 1299-1312]. Based on an articulatory study of speech errors, Mowrey and MacKay conclude that speech errors occur on a continuum of muscular activation, and that segmental errors are just the extreme cases that are perceptually anomalous. However, they analyzed electromyographic data from single muscles and thus were unable to determine whether single muscle activations occurred independently, or as part of larger gestural units. This paper examines the unitary nature of speech production acoustically. We find evidence for both featural and sub-featural errors in speech production. Our results support a model of speech production that hierarchically organizes individual gestural units into gestural constellations equivalent to linguistic segments. In addition, we find evidence for interaction between the lexical and gestural level, supporting connectionist models of speech error production.

### **Introduction and Background**

Speech errors have traditionally been used to provide evidence for models of speech production which utilize the constructs of linguistic theory as psychologically real components of linguistic performance (e.g., Levelt, 1989). While it is indisputable that speech errors do occur, few unambiguous conclusions about the mechanisms of speech production can be drawn from speech error data. In addition, several researchers have questioned the validity of phonological speech error data, which is traditionally recorded using phonetic transcription (Boucher, 1994; Ferber 1995; Laver, 1980; Mowrey & MacKay, 1990). In this paper, we undertake an acoustic analysis of speech produced by four talkers in a speech error elicitation experiment. Acoustic analysis circumvents the problems of perceptual bias introduced by phonetic transcription. Our analysis provides evidence for the psychological reality of phonological segments in speech production as a statistical tendency, supporting traditional analyses. However, we also find evidence for speech errors at a sub-featural or phonetic level which have not previously been attested. Our data support a model of speech production where individual gestures are organized into a loosely yoked hierarchy of linguistic units. Features, segments, phonemes, and words are levels of the hierarchy which influence the implementation of gestures in both erroneous and error-free productions. Overall, the model is compatible with dynamical systems approaches to speech production (e.g., Browman & Goldstein, 1990; Saltzman & Munhall, 1989) where gestural constellations corresponding to segments and words are attractor states of the system.

### **Background**

Phonological speech errors (also called sub-lexical errors) have been an important source of evidence for the psychological reality of phonological features and segments. In many speech errors, it appears that portions of the intended utterance are produced in an unintended order. It is claimed that the misordered portions frequently correspond to linguistic units such as clusters, phonemes, segments, and

features. The errors in (1) are given by Fromkin (1971) as support for the independent existence of segments, distinct from words or clusters.

- (1) a. frish gotto 'fish grotto'  
 b. blake fruid 'brake fluid'  
 c. spicky point 'sticky point'

In (1a), the /r/ in grotto is presumably misordered, and appears as part of the preceding word instead. In (1b) the /l/ and /r/ are exchanged, each appearing in the others' place. In (1c), the /p/ (or alternatively, the [+labial] feature) is anticipated, but also repeated in its proper place. As with (1c), it is often the case that any particular error can be interpreted in more than one way. Another error from Fromkin (1971), glear plue sky for clear blue sky, is claimed to unambiguously demonstrate that errors involving linguistic features are possible and thus that features are psychologically real.

Mowrey and MacKay (1990) argue against traditional speech error evidence as methodologically unsound. Based on electromyographic (EMG) recordings of tongue twister production they conclude "that errors which have been consigned to the phonemic, segmental, or feature levels could be reinterpreted as errors at the motor output level" (p. 1311). In this section, we review the methods of traditional speech error analysis and the results of the instrumental study of Mowrey and MacKay (1990).

### Traditional Approach

Traditional approaches to speech error analysis all use phonetic transcription to encode speech errors for future analysis. In 'naturally occurring' speech error corpora, errors that are observed in everyday speech are written down. In the early corpora (e.g., Fromkin, 1971; Shattuck, 1975) the error recorders were participants in the communicative event in which the error occurred. Stemberger (1983) collected naturally occurring errors only as an observer in an attempt to reduce the potential for perceptual bias. In some cases, recordings of naturally occurring speech are used, and suspected errors are listened to repeatedly to ensure accurate transcription (e.g., Garnham, Shillcock, Brown, Mill, & Cutler, 1982). Transcription is also normally used to encode errors in speech error elicitation experiments (e.g., Baars, Motley, & MacKay 1975; Dell, 1986), though usually the utterances themselves are recorded on tape or computer and listened to repeatedly.

In all cases where transcription is used, the recording of a speech error necessarily coincides with the hearer noticing an anomalous percept. Thus, in traditional approaches, a speech error can be defined as an utterance that produces an anomalous percept that would be recognized as anomalous by the speaker (e.g., Dell 1986). Mowrey and MacKay (1990, p. 1299) note that imperceptible speech errors may also exist and claim that "such production anomalies are errors if speech output differs from the speaker's intended output, however subtle the anomaly." Their claim raises the question of how articulatively detailed the speaker's intentions are, which we discuss below.

Speech error evidence recorded using transcription has been used to argue in favor of the psychological reality of many phonological units, including the feature, segment, phoneme, cluster, syllable, and word. Among sub-lexical errors it has been claimed that errors occur primarily at the level of the phoneme or feature (Wickelgren, 1965) and that erroneous utterances are phonetically/phonotactically grammatical (Fromkin, 1971; Wells, 1951). In other words, it is claimed that speech errors occur by misordering abstract units and the result is a phonetically normal segment and possible word according to the grammar of the language. Phonetic errors are often explicitly argued against (e.g., Fromkin, 1971) and it is claimed that when abstract units move to different locations, they phonetically accommodate to their new environment. It should be noted that there is disagreement on these conclusions even among

experimenters using the same collection techniques. For example, Stemberger (1983), based on his own corpus of naturally occurring errors, claimed that ungrammatical utterances do occur, though infrequently.

The use of transcription to encode speech errors has received widespread criticism and been the subject of some empirical research (Boucher, 1994; Ferber, 1995; Garnham, Shillcock, Brown, Mill, & Cutler, 1982; Laver, 1980; Mowrey & MacKay, 1990; Shattuck-Hufnagel, 1983). There are two primary criticisms. First, the use of phonetic transcription cannot capture sub-contrastive errors, below the level of a segment or feature, since the transcription system is inherently segmental. If non-contrastive errors do occur, careful transcription of repeatedly heard recordings of speech errors would probably discover some of them. However, speech errors heard in conversation are usually only broadly transcribed and the listener's full attention is not on phonetic detail. Thus the errors contained in naturally occurring corpora (Fromkin, 1971; Shattuck, 1975) may only represent a portion of the actual speech errors produced in natural dialogue, and any model based on transcription evidence is therefore unable to answer questions about the phonetic details of speech errors. Errors collected in speech error inducing experiments (e.g. Baars, Motley, MacKay, 1975; Dell, 1986; Shattuck-Hufnagel, 1992) might be more revealing of phonetic detail, since the errors are recorded and can be reviewed many times over. However, the design of these experiments is usually to produce a specific error. Thus, the experimenter's task is more like a forced-choice decision – is it an error or not – rather than a phonetic transcription task.

The second criticism of error collection using transcription is that the transcript is subject to the perceptual biases of the listener. It is well known from the literature on speech perception that speech is perceived in the context of the language system of the listener (see Wright, Frisch, & Pisoni, in press, for a recent review). For example, in the phenomenon of categorical perception, phonetically anomalous speech sounds which are acoustically intermediate between two categories are perceived by naive listeners as members of one category or the other, rather than a blend (see Liberman, 1997, for several articles). Categorical perception is also influenced by the lexical status of the outcome; intermediate speech sounds are more likely to be heard as a segment which creates a word rather than a non-word percept (Ganong, 1980). In another phenomenon, known as phonemic restoration, speech samples which have had segments replaced by noise or a cough are perceived as intact. Listeners, even when informed that there is a missing segment, are unable to accurately report which segment is missing or where in the word the disruption occurred (Samuel, 1981; Warren, 1970). Research on the detectability of mispronunciations of segments in running speech has found that the likelihood of detecting an error depends on the error's place within the word and sentence, and the predictability of the word in its sentential context (Cole, 1973; Marslen-Wilson & Welsh, 1978). In summary, speech error percepts are subject to systematic biases and it is unclear whether many of the patterns observed in traditional speech error analyses are informative of the speech production process or merely a reflection of the hearer's perceptual system.

### **Instrumental Data**

Mowrey and MacKay (1990) presented an electromyographic (EMG) study of sub-lexical speech errors elicited using tongue twisters. They used EMG recordings of the orbicularis oris muscle (lower lip) and lingual transversus/verticalis muscle (tongue blade) in combination with audio recordings to determine whether non-contrastive errors occur. The tongue twisters they used, shown in (2), crucially involved segments with lower lip (/Σ/ in 2a) or tongue blade (/λ/ in 2b, c) articulations in proximity to segments that did not contain such articulations (/s/ and /r/ or /O/, respectively).

- (2) a. She sells sea shells by the seashore

- b. Bob flew by Bligh bay
- c. Fresh fried flesh of fowl

Mowrey and MacKay interpreted unexpected muscular activity as evidence for an error in speech production, and so in their study the definition of an error is crucially different from traditional techniques that are based on perceptible contrast. They found a number of instances of inappropriate muscular activity where there was no perceptible anomaly. In addition, they found several cases of inappropriate muscular activity where there was a potentially intermediate percept that was not a clear member of either category. They conclude based on this evidence that sub-contrastive errors do occur, and they further claim such errors are quite frequent. In one set of 150 recordings of Bob flew by Bligh bay they report 48 tokens containing errors involving intrusive transversus/verticalis muscle activation, the majority of which involved an amount of activation intermediate between none and that appropriate for a normal [l] production. Their proposal, from an articulatory perspective, is that sub-lexical errors occur on a continuum of gestural activity, and are neither segmental nor grammatical under any reasonable definition of these terms.

This study of EMG activity during tongue twister production clearly reveals gradient levels of gestural activation that have imperceptible consequences. Mowrey and MacKay conclude that such evidence demonstrates that

“neither the speaker nor the listener may be aware of the true nature of the errors made or whether indeed an error has been made at all .... [which] contradicts any model claiming that ‘low-level’ phonetic processes are necessarily overseen by a higher-order segmental or featural planning unit.” (p. 1311)

While transcription is too coarse a means of encoding phonetic details of an utterance, EMG recordings of the activity a single muscle cannot reveal the gestural coordination that would be signs of higher level organization at the level of the feature or segment. There is no evidence as to whether other muscles required to produce a normal segment were active. It is possible these muscles were also gradually activated in a coordinated fashion. It is also possible their behavior was independent. Mowrey and MacKay’s evidence thus demonstrates that gradient speech errors can and do frequently occur at the level of the gesture, but their data have no bearing on the question of higher levels of organization in speech production.

Additionally, their definition of an error as unexpected muscle activity provides no allowance for what might be normal variation, by assuming that the speaker’s intention excludes such activity. They found many of these gradient errors to be imperceptible, even upon repeated listening with high quality equipment. A possibility that should be considered is that one goal when talking is to convey a distinct message to an observer. Thus, random variation that is non-contrastive may not be monitored as strictly and for this reason such productions might not be considered erroneous by the production mechanism.

Finally, the presence of activation for the particular muscle fiber that was measured does not ensure that any actual movement took place. If activation of the muscle was sporadic or inconsistent across different fibers, the muscle may not have moved at all or may have moved in some uncoordinated fashion. While most researchers would agree that uncoordinated movement is an articulatory error, if no such movement occurred it is difficult to assert that an error was made. The lack of movement may just be a part of the normal operation of the muscle that has the fortuitous side effect of inhibiting random muscle activation from disturbing normal production, and thus it could be considered part of the monitoring system.

## Overview of the Study: Acoustic Analysis of Speech Errors

Phonetic transcription provides a coarse coding of speech, filtered through the perceptual system of the listener. EMG analysis of a single muscle provides detailed information about activity levels of muscle fibers but little information about the actual movements of the articulators and their coordination. We propose to analyze experimentally elicited speech errors using acoustic measures. Acoustic analysis can examine both contrastive and sub-contrastive dimensions together, providing an intermediate level of analysis between transcription and EMG.

The following section contains a description of the data and analysis techniques used in an acoustic analysis of phonological speech errors. Our data support the conclusions of Mowrey and MacKay (1990) and of traditional error analysis. We find evidence for both categorical segmental/featural errors and gradient sub-featural errors. In addition, we find that the distribution of both gradient and categorical errors are influenced by the phonetic context of the error and the lexical status of the resulting word. Our results support a model of speech production where the traditional linguistic units are viewed as preferred or familiar modes of organization for constellations of speech gestures. Sub-lexical error data provide evidence for a dynamical systems model of speech production in which error-free and error-ful productions tend towards the same attractor states; grammatical sequences of segments. We conclude from this that the phonetics of phonological speech errors is an important topic for future research.

### Methods

#### Data

The data for our analysis come from a speech error-inducing experiment using tongue twisters that was part of the first author's dissertation research (Frisch, 1996). In the original experiment, 21 participants each produced 6 repetitions of 88 different tongue twisters. The participants were monolingual American English speaking undergraduate students at Northwestern University, who were paid for their participation. The tongue twisters were printed individually in large type on index cards. The participants read each tongue twister three times, and then repeated the tongue twister from memory three times. Participants were allowed to consult the index card between repetitions during the repeat from memory portion if they forgot the correct words. Each participant had a break after half of the experiment was completed in which they were engaged in normal conversation for about 5 minutes. The experiment lasted about 45 minutes.

The entire session was recorded on audiocassette using a Marantz portable cassette recorder. The participants wore an electret condenser microphone attached to their shirt in the upper chest area.

The original experiment was a psycholinguistic study designed to induce word onset consonant errors for a 22 different consonant pairs. Each consonant pair was used in four different tongue twisters. Each tongue twister consisted of four monosyllabic words. The /s/-/z/ tongue twisters we examined in detail for the present study are given in (3). Half of the tongue twisters created errors that were existing words (3a, c), and half created non-word errors (3b, d).

- (3)
- a. sit zap zoo sip
  - b. sung zone Zeus seem
  - c. zit sap sue zip
  - d. zig suck sank zilch

The words in the tongue twister and their error targets (if they were words) were balanced for lexical frequency within each tongue twister so that lexical frequency effects would not differentially influence the relative error rate between the consonants, from /s/ to /z/ or /z/ to /s/ in this particular case.

The acoustic analysis focuses on a single consonant pair, /s/-/z/. Data from the first four participants in the original experiment were analyzed. Twisters containing /s/-/z/ were selected for this study as the first author found some of the tokens in these twisters to be perceptually ambiguous when scoring the original experiment. It was assumed that these stimuli had good potential for finding a variety of error types and examining the existence of and relationship between categorical and gradient speech errors. The choice of the first four participants in the experiment was arbitrary. Two of these participants were male and two were female.

It should be noted that the use of nonsense tongue twisters to generate speech errors may introduce systematic patterns not observed in naturally occurring speech, and thus the results of this paper may not generalize to normal speech production. Shattuck-Hufnagel (1992) compared perceived error rates in nonsense tongue twisters with those in spontaneously generated utterances using the same words. There was a difference in the number of errors observed in the two conditions, but the relative patterns of errors were the same for both conditions. Since quantitative but not qualitative differences were found, it appears that speech in tongue twisters does not differ in important ways from spontaneously generated utterances. It may be that tongue twisters enhance the likelihood of observing gradient speech errors, given that they involve an unusually high level of repetition and alternation. However, we assume that this entails quantitative rather than qualitative differences, and that the same phonological encoding and phonetic implementation is involved in producing both tongue twisters and natural speech.

## Measurements

The onset /s/ or /z/ of each word was measured in a variety of ways, with the overall goal of finding quantitative differences between /s/ and /z/ along acoustic dimensions which were potentially independent, in that there was no necessary articulatory or acoustic interdependence between measures. In other words, the goal is to represent each production along several acoustic dimensions and then examine the extent to which variation along the dimensions is gradient or categorical. Three measures which differentiate /s/ and /z/ tokens produced by our talkers whose articulatory bases are sufficiently physiologically independent are used in this paper.

The first measure is the duration of the fricative. We considered the duration of the fricative to be the total duration of fricative noise, including overlap with the following vowel and the preceding vowel, if any. We determined whether or not fricative noise was present by examining the waveform in conjunction with a spectrogram. In general, /z/ overlapped with the surrounding vowels and /s/ did not. The expected pattern was for /s/ to be longer than /z/ (Klatt, 1976), which was true across all four talkers in our study (see figure 1, top panel).

-----  
 Insert Figure 1 about here  
 -----

The second measure is the amplitude of frication noise. Amplitude of frication noise was calculated based on the root mean square (RMS) amplitude of the signal after high pass filtering at 2kHz. For each fricative, RMS amplitude was computed for 10ms windows across the entire fricative. The



window containing the peak RMS amplitude was identified, and the amplitude of frication was defined to be the RMS amplitude of the fricative in a 50ms window around the peak. This generally corresponded to the middle of the fricative segment. The expected pattern was for /s/ to have more frication noise than /z/ (Stevens, 1960; Pickett, 1980), which was true across all four talkers in our study (see figure 1, middle panel).

The final measure is the percent age of voicing in the fricative. We used the waveform to determine when voicing was present, and considered portions of the signal which had evidence of clear periodicity as voiced. The percent of voicing was the fraction of the total duration that contained voicing. In most cases, there was clear evidence of a glottal closure phase that indicated the presence of voicing. In some cases there was very breathy voicing resulting in sinusoidal waveform overlaid with frication noise. Breathly voicing was considered voicing until the point where the pressure valleys marking each glottal cycle were too obscured by frication noise to be reliably identified. The expected pattern was for /z/ to have a larger percent of voicing than /s/, which was true across all four talkers in our study (see figure 1, bottom panel).

The measurement of duration, frication amplitude, and percent voicing for a sample case of /z/ is shown in figure 2. The top panel in figure 2 shows the original signal, with brackets indicating the overall length of frication and the portions which are voiced. This example is devoiced in the middle, a relatively common pattern for /z/ (Haggard, 1978). The bottom panel shows the signal filtered at 2kHz, which has removed most of the periodicity, and the 50ms window around the peak over which the amplitude of frication was computed.

-----  
 Insert Figure 2 about here  
 -----

While the goal was to consider measurements which are articulatorily and acoustically independent, it is clear that duration, amplitude of frication, and percent voicing are not completely unrelated. During voicing, amplitude of frication is necessarily reduced, as the closed phases of the glottal cycle stop the airflow that is crucial for maintaining frication. However, talkers have additional articulatory strategies for enhancing the noisiness of fricatives, such as grooving the tongue to create an air channel aimed at the teeth (Shadle, 1985). These strategies are independent of voicing and are selectively applied to voiceless fricatives to enhance their contrastiveness. Devoiced /z/ can be distinguished from /s/ in part by the difference in frication amplitude. Duration can also interact with voicing. Pressure buildup behind the fricative constriction equalizes transglottal pressure, making voicing more difficult. Thus, long duration voiced fricatives may have a tendency to devoice. However, the connection is not an absolute one, as it is perfectly possible to maintain voicing during a long fricative. Though rare, voiced geminate fricatives are attested in the world's languages (e.g., Italian). Note also that this dependence only applies to voicing in /z/. For /s/, which is voiceless, there is no necessary increase in duration or amplitude of frication associated with the lack of vocal fold vibration. Thus, to the extent that the same patterns are found both for /s/ to /z/ errors and /z/ to /s/ errors, the dependence between acoustic properties can be considered sufficiently weak to be treated as independent in the statistical analysis.

### **Description of the Data**

In the previous section, it was demonstrated that the three dimensions of duration, frication amplitude, and percent voicing differentiate /s/ and /z/ which these dimensions are different on average. It should not be surprising that there is considerable overlap in the individual tokens that were produced.



A sketch of the variability in production across the four talkers is presented in this section, along with a discussion of the problem of defining a speech error acoustically.

*Acoustic Description of the Data.* In describing the data acoustically, tokens are categorized as /s/ or /z/ based on the word the speaker was supposed to say if the tongue twister was produced correctly. This was assumed to be the speaker's intended utterance. Anomalous and non-anomalous tokens are treated identically, as a method of determining acoustically anomalous tokens has yet to be developed.

Duration of all tokens for each speaker is shown in figure 3. The data are grouped into intervals of 20ms and the aggregate number of tokens is shown for each group. All talkers have overlap, especially talkers 2 and 3. While the distributions of tokens for talker 4 look like normal, Gaussian variation, the other talkers show some signs of multi-modality. There are potential sources of systematic variability in the stimuli, including the position of the word in the phrase and the surrounding phonetic context, which are discussed below.

-----  
 Insert Figure 3 about here  
 -----

The amplitude of friction for all tokens is given in figure 4. Aggregate data in 200 unit intervals are given for talker 1 due to the large amount of variation for this talker, and in 100 unit intervals talkers 2, 3, and, 4. All talkers show clear evidence of multi-modality in their distributions of intended /s/ and /z/ productions. Note also that in most cases, peaks in occurrence rate for /s/ occur in the same location as peaks in occurrence rate for /z/. This suggests either that there are certain modes of articulation for each talker which result in common acoustic consequences for the level of friction noise, or that there is a massive error rate for sub-contrastive elements of articulation. A dynamical systems perspective on speech production can unify these two hypotheses, as preferred modes of articulation will be utilized for both error-free and erroneous productions, leading to similar multi-modal distributions in both cases.

-----  
 Insert Figure 4 about here  
 -----

The distribution of percent voicing for each talker is shown in figure 5. There is considerably less overlap between /s/ and /z/ for percent voicing. The data are aggregated into intervals of 20%, but the endpoints of 0% and 100% which occurred quite frequently, are presented separately. Since there is less variability, there is less evidence of multi-modality, but there are suggestive trends, especially for /z/.

-----  
 Insert Figure 5 about here  
 -----

### **The Acoustic Definition of a Speech Error**

Given that there is overlap in the distribution of tokens across acoustic dimensions, and it is probably not desirable to consider all of this overlap as indicative of a speech error, the question of how to define a speech error acoustically becomes crucial. Traditional analyses consider only perceptible changes to be errors. The EMG study of Mowrey and MacKay took any muscle activation to be an error, and it is uncertain whether that activation had any articulatory or acoustic effect. We propose to use an acoustic definition of speech error that is based on contrast. For example, an intended production of /s/







which results in acoustic properties characteristic of a /z/ can be thought of as a speech error. With this definition in mind, each dimension can then be analyzed independently, and the value along each dimension can be either more like the intended segment (e.g., /s/) or the other (e.g., /z/). The term category crossing will be used to denote a value on an acoustic dimension that is more like the unintended segment than the intended segment. A production which has category crossings on all dimensions would likely result in an error percept, since its acoustic characteristics are all indicative of the wrong category.

Two definitions of an error for a dimension will be considered, one ‘strict’ and one ‘lax’. The relevant values for the acoustic definition of a speech error on each dimension for each talker are given in table I. Table I shows mean values for each talker for each dimension, as well as the crossover point between categories. In the strict definition, the value on a dimension for an intended segment must be more extreme than the mean value for all productions of the other segment. For example, for an intended /s/ production by talker 1 to have an error on duration by the strict definition, the duration of the production would have to be less than 155.6ms. In the lax definition, the value on a dimension for an intended segment must be more like the other segment (a category crossing as defined above). For example, for an intended /s/ production by talker 1 to have an error on duration by the lax definition, the duration of the production would have to be less than 176.9ms.

**Table I**

**Mean values of acoustic parameters and category crossover points for duration, frication amplitude, and percent voicing for each talker.**

Talker	Mean /s/			Category crossing			Mean /z/		
	Duration	Percent voicing	Frication amplitude	Duration	Percent voicing	Frication amplitude	Duration	Percent voicing	Frication amplitude
1	200.2ms	9.8%	1210.8	176.9ms	51.2%	883.0	153.6ms	92.7%	555.3
2	199.6ms	7.8%	461.7	180.9ms	45.0%	391.2	162.2ms	81.2%	320.8
3	163.7ms	15.0%	893.8	147.0ms	47.7%	738.6	130.3ms	80.3%	583.5
4	161.6ms	3.1%	357.6	136.9ms	43.6%	298.0	112.2ms	84.1%	238.4

In the next two sections, evidence for both gradient and categorical errors is presented, by examining the distribution of category crossings over the three dimensions of duration, frication, amplitude, and percent voicing. Both strict and lax definitions are considered, and the same conclusions can be drawn in either case, suggesting that the details of the definition are not crucial to the overall results.

### **Gradient and Categorical Errors**

Acoustic evidence for gradient errors can be found by examining whether category crossings occur for a single dimension, while the other dimensions remain normal. If such errors are found, then non-contrastive aspects of the production of a segment can be involved in speech errors, and the conclusion of Mowrey and MacKay is supported. Acoustic evidence for categorical errors can be found

by examining the rate at which category crossings occur on all three dimensions simultaneously. If errors across dimensions are correlated in some way, then there is evidence for categorical behavior in the phonetic details of speech errors, suggesting there are higher levels of organization, such as the segment or word, which bind gestures together.

### Sub-Featural Errors

Gradient or sub-featural errors are found for all acoustic dimensions, though there is less variability and fewer instances of category crossings for percent voicing. The total number of single dimension errors for each dimension is shown in figure 6. Figure 6 shows the total number of occurrences in the corpus of single dimension errors under different criteria. The strictest criterion is at the extreme left. Under this criterion, the dimension in error was more extreme than the mean of productions of the unintended segment, and the other two dimensions were more extreme than the mean productions of the intended segment. Thus, the erroneous dimension was very atypical and the other two dimensions were very typical. The most lax criterion is at the extreme right. Under this criterion, the dimension in error was a category crossing, more like the unintended segment than the intended segment. The other dimensions were not category crossings, and thus more like the intended segment. The other two criteria use a mixture of strict and lax criteria for the erroneous dimension and the other dimensions. Across all four criteria, the patterns are qualitatively similar. There are a number of acoustic errors for the duration and frication amplitude dimensions. Many fewer errors are found on the voicing dimension, but they do occur at all criteria levels except the most strict. The paucity of percent voicing errors is not surprising, given that the most salient cue for the voicing contrast in /s/ and /z/ is periodicity. We suspect that in a larger corpus, such errors would be found even at the strictest criterion. Overall, then, support for gradient sub-featural speech errors is found in the acoustic measurements.

-----  
 Insert Figure 6 about here  
 -----

Recall also from the previous section that the distribution of acoustic characteristics within dimensions is multi-modal. This suggests that these non-segmental errors, at the sub-featural or gestural level, are not just random variation in the phonetic implementation of segments. Error-free and erroneous productions tend toward similar acoustic outcomes, suggesting that there is organization at the gestural level. These errors are not just ‘slips of the tongue’ which can be dismissed as normal variability.

Finally, in a descriptive analysis of the productions of talker 2, Frisch and Wright (1997) report an instance of an intended /s/ token which was very short, of low amplitude, but completely lacking in voicing. This token produced a /z/ percept despite the complete lack of voicing. While this analysis highlights the difficulty of sorting out normal variability from a gradient speech error, the systematic patterns which we have observed support the conclusion that non-contrastive errors in speech production do occur, in line with the findings of Mowrey and MacKay (1990).

### Segmental Errors

The statistical analysis undertaken to evaluate whether there are errors at the featural or segmental level is inspired by Mowrey and MacKay’s statement that “units such as features, segments, and phonemes may well exist; if it is found at some later time that blocks of motor commands behave as single entities, we should have good evidence of a higher level of organization” (p. 1311). Above, it was demonstrated that category crossings can occur for one dimension to the exclusion of others. If there are



categorical tendencies in the acoustic patterns of speech errors, then we would expect that such errors are exceptional, with category crossings across all dimensions much more common. Since there is overlap between categories on all dimensions, some effort must be made to determine whether categorical errors occur independently of normal variation and gradient errors. In other words, do instances of category crossings on all three dimensions occur more frequently than we would expect given the rate of category crossings on each dimension?

Table II shows the number of category crossings and their cooccurrence for all three dimensions. There are 11 instances of category crossings on all three dimensions simultaneously. Given the rate of category crossing across each dimension separately, we would expect 3.7 instances with category crossings on all three dimensions. The actual distribution of occurrences in Table II differs significantly from the distribution predicted by chance cooccurrence of category crossings ( $\chi^2(4) = 25.1, p < .001$ ). Most of the lack of fit comes from the disparity between actual and expected cooccurrence of all three category crossings. Note that if strict criteria are employed rather than lax criteria, the same results are found. Fewer instances of errors on all three dimensions are observed, but strict criteria make errors on each dimension less probable, and so the combination of errors on all dimensions becomes extremely rare. In fact, the predicted rate is about 0.2 in this corpus, so the chi-square test cannot be reliably used to assess significance for the strict criteria.

**Table II**

**Distribution of tokens across dimensions under lax criteria for all talkers.**

Frication Amplitude				
Duration	Voicing	Normal	Crossing	Total
Normal	Normal	187	64	251
	Crossing	12	6	18
Crossing	Normal	60	37	97
	Crossing	11	11	22
Total		270	118	388

The fact that there are more instances where category crossings occur on all three dimensions than would be expected by chance suggests that there are categorical tendencies to the observed variation in the data. When an error on one dimension occurs, it is also more likely that there will be errors on other dimensions. A second indication of categorical behavior would be the suppression of gradient errors. This tendency is also observed in the data. If the rate of occurrence of zero, one, two, or three category crossings is compared to the rate expected by chance, gradient errors, with one or two category crossings, are found less than expected by chance. Figure 7 shows the ratio of the observed number of productions with no error (no category crossing), a gradient error (with one or two category crossings), and a

categorical error (with three category crossings) to the number expected by chance combination of category crossings for all talkers. A tendency for categorical behavior is observed, as no error and categorical error cases are found more often than expected by chance (observed over expected greater than one), while gradient errors are found less often than expected (observed over expected less than one).

-----  
 Insert Figure 7 about here  
 -----

As noted earlier, there is some dependence between duration, frication amplitude, and voicing for /z/, and thus a statistical analysis that assumes independence may be inappropriate. However, the same categorical tendencies are found for both /s/ and /z/, if they are examined separately. No phonetic dependence between dimensions is predicted for /s/, so the categorical tendencies which are observed in /z/ to /s/ errors must be the results of tendencies in the organization of gestures, and not merely a low-level phonetic dependence.

Errors across acoustic dimensions are correlated, suggesting that there are higher level tendencies to organize gestures into constellations equivalent to linguistic segments or features. Gradient errors are found, but occur less frequently than would be expected if we assumed that such errors were the result of normal phonetic variation. The observed patterns are compatible with a dynamical systems model of speech production. Such a system is not rigidly organized, as one might expect from an entirely symbolic model of speech perception, but instead displays systematic tendencies. The greater than expected cooccurrence of non-errors and categorical errors suggests that segments can be thought of as attractor states in the space of possible gestural combinations. Non-segmental productions can occur if the system is perturbed, for example by a tongue twister or in normal speech by competition between alternative production plans reflecting different word choices. Under load free conditions, such as reciting a familiar sentence or phrase, such anomalies would not be expected.

### Individual Differences

The data for individual talkers are sparse, but are in all but one case consistent with the group data. Talker 4 produced no instances of three category crossings, and so trivially had fewer categorical errors than were predicted by chance. However, only 0.4 were predicted by chance, so the lack of observed instances is not surprising. The observed versus expected distribution of non-errors, gradient errors, and categorical errors for the individual talkers is given in figure 8. For this figure, expected rate of category crossing depends on the individual talker's category crossing rate for each dimension. Given the rate of category crossings for talker 4, more non-errors are found than expected, while less gradient errors are found than expected. Despite the lack of categorical errors, talker 4 still shows fewer gradient errors (one or two category crossings) than expected, and thus still demonstrates some evidence of categorical behavior. All other talkers show categorical tendencies consistent with the group data to a greater or lesser extent.

-----  
 Insert Figure 8 about here  
 -----





## Phonetic and Lexical Context

Speech errors involving segmental interactions in naturally occurring corpora and error inducing experiments are influenced by a variety of contextual factors, including shared segmental context (Dell, 1984), and shared word position and syllable stress (Shattuck-Hufnagel, 1992; Frisch, in press). In this section, the influence of phonetic and lexical context on the phonetics of speech errors is explored. There are systematic influences of context on the rate and cooccurrence of category crossings for the duration, frication amplitude, and percent voicing of /s/ and /z/ productions. Comparison with the original speech error experiment scoring (Frisch, 1996) shows that these factors are directly reflected in the observation of error percepts. This suggests that phonetic and lexical context are important factors in both the perception and production of speech errors.

### Phonetic Context

It is a well known fact that phonetic context influences the articulatory details of the production of a segment, through coarticulation and coproduction. The tongue twisters examined in this study contain some variation in segmental context for both /s/ and /z/. The vowel following the onset /s/ or /z/ varied across the twisters. However, no systematic influence of this vowel was observed. In this section, the effect of the preceding segment is examined. Systematic influences of the preceding segment can be seen in the acoustic properties of /s/ and /z/, suggesting that the acoustic error data are influenced by details of articulatory implementation such as coarticulation. Given that preceding context influences the acoustic properties of /s/ and /z/, it might also be the case that the observation of errors in a traditional speech error experiment is influenced by context. To explore this possibility, the acoustic data are compared to the error scoring performed by the first author in the original tongue twister study in Frisch (1996), henceforth referred to as the ‘transcribed error data’. The transcribed error data are based on the perceived errors in the productions of 21 participants. The acoustic error data are based on the productions of the first four participants from this population.

Intended /s/ occurs phrase initially in two twisters (sit zap zoo sip and sung zone Zeus seem) and intended /z/ occurs phrase initially in the other two twisters (zit sap sue zip and zig suck sank zilch). Intended /s/ and /z/ occur after stops in several places. There is one instance where intended /s/ occurs following a coda /s/, in sung zone Zeus seem. In this same twister, both intended /z/s occurs after a nasal. Intended /s/ and /z/ each occur intervocally in one twister (sit zap zoo sip for /s/ and zit sap sue zip for /z/).

There is systematic influence of the preceding context in the acoustic error data. In particular, the category crossing rate in the four talker’s productions for the different acoustic dimensions depends on context. Some of this variation does not appear to be indicative of an increase in rate of gradient errors in that context. For example, productions of both /s/ and /z/ had higher frication amplitude in initial position. This effect of prosodic position reduces the category crossing rate for frication amplitude for initial /s/ and increases the category crossing rate for initial /z/ in the acoustic error data. We feel that this is not a genuine speech error effect. It merely reveals that using the overall frication amplitude across all contexts is only an approximation of the true phonetic description of frication amplitude for /s/ and /z/, which is most likely subject to prosodically conditioned fortition (Fougeron & Keating, 1997, *inter alia*). In a similar case, the duration and amplitude of frication of /s/ was reduced following another /s/<sup>4</sup>. Duration and amplitude of frication may have been reduced following /s/ since articulatory effort could be reduced

<sup>4</sup> It should be noted here that in cases where there were not two distinct /s/ productions, with silence in between, duration for /s/ was not measured and such tokens were excluded from any analysis involving duration.

without seriously endangering the listener's percept of /s/ (Lindblom, 1983). This is another example of normal contextual variation rather than contextual effects on gradient speech error rates in the acoustic error data.

Contextual effects on category crossing rate, which are indicative of the influence of phonetic context on gradient and categorical speech error rates, are clearly found in the acoustic error data for the percent voicing dimension. Figure 9 shows category crossing rate for /s/ and for /z/ as a function of context for all four talkers, along with the perceived error rate from the transcribed error data. For /s/, the category crossing rate directly reflects the perceived speech error rate, suggesting that when an intended /s/ becomes mostly voiced, a /z/ percept results. The pattern of category crossings in the acoustic error data clearly reflects the effect of phonetic context on voicing. After a stop, in initial position, or after another /s/, the likelihood of erroneous intrusive voicing is quite low. These contexts themselves do not promote vocal fold vibration. It is also not surprising that post /s/ context would be the least likely to contain a category crossing or a perceived error.

-----  
 Insert Figure 9 about here  
 -----

For /z/, the relation between category crossing rate and perceived error rate is more complex, but also more revealing of the true nature of phonological speech errors. Overall, the qualitative patterns of category crossing and perceived error rate for /z/ are similar, except in the case of intervocalic /z/. In intervocalic context, category crossings for voicing in the acoustic error data are found quite frequently, but the perceived error rate in the transcribed error data is low. Intervocalic devoicing of /z/ is quite common (Haggard, 1978). In this context, category crossings for duration and amplitude of frication were also quite common. In intervocalic context, frication overlaps with both the preceding and following vowel. Thus, cues to voicing are quite strong, and the need to sustain voicing throughout the fricative is lessened. We believe the increase in duration and frication amplitude is a context-specific production strategy that enhances the robustness of the cues for frication. In this case, the information that the segment is a voiced fricative is distributed temporally, rather than occurring simultaneously with the steady state portion of the fricative. Setting aside the intervocalic context, the relative rates of category crossing in post-stop, initial, and post-nasal contexts in the acoustic error data are mirrored in the perceived error rates from the transcribed error data. However, unlike the perceived error rates for /s/, the perceived error rates for /z/ are much lower than the rate of category crossing. This suggests that there is a quantitative effect of phonetic context on the percent voicing of /z/, but that in most cases this effect is not extreme enough to produce a voiceless percept. As with /s/, the phonetic context has an influence on the acoustics of the intended /z/, and this influence in turn has an effect on the perceived error rate.

In the case of /s/, some voicing results in a /z/ percept, while in the case of /z/, devoicing does not necessarily result in an /s/ percept. Frisch (1996) reports a large asymmetry in the error rate from /s/ to /z/ versus from /z/ to /s/. Errors from /s/ to /z/ are much more common. This fact is straightforwardly explained by the phonetics of these segments. A partially voiced /s/ is perceived as a /z/, but a devoiced /z/ is still perceived as a /z/.

Analogous effects of segmental context on the voicing of /s/ and /z/ (as well as /f/ and /v/) were found by Pirello, Blumstein, and Kurowski (1997). Pirello et al examined the effect of preceding stops on the voicing of fricatives in nonsense phrases like his fav sips it. They found effects of voicing at the boundary between stop release and fricative onset. Fricatives following voiced stops were often initially voiced and fricatives following voiceless stops were often initially devoiced, regardless of whether they were supposed to be voiced or voiceless.



## Lexical Context

The phonetic details of a segment's environment have an effect on its acoustic characteristics, which in turn influences the likelihood of a perceived error. This suggests that there are low-level phonetic influences on the higher-level segmental categories. It might be expected that there would be influences in the other direction, with higher-level categories influencing the phonetics. A contextual factor explicitly manipulated in the tongue twister experiment is the lexical status of the error outcome. In half of the twisters, the error outcomes were all words. In the other half, the outcomes were all non-words. We found that the word status of the outcome influences the degree to which errors are categorical or gradient.

Categorical tendencies in the production of speech errors were demonstrated in section 5. Normal productions and categorical errors occurred more frequently than were expected by chance combination of category crossings, and gradient errors were found less often than expected. Figure 10 shows the ratio of observed to expected rate of non-errors, gradient errors, and categorical errors in the word outcome and non-word outcome contexts. The tendency toward categorical behavior, with higher relative rates of occurrence for non-errors and categorical errors and lower relative rates for gradient errors, is stronger in the word outcome case. The same patterns are observed, though less strongly, for the non-word case.

-----  
 Insert Figure 10 about here  
 -----

The fact that categorical segmental tendencies are found in both word and non-word outcome cases provides evidence for the segment as a unit of organization in the speech production process. Categorical tendencies are found more strongly in the word outcome case, providing evidence that there are top-down influences on speech production similar to the top-down influences on speech perception. In this case, there is a close parallel with the Ganong effect, where a phonetically ambiguous stimulus is more likely to be identified as a segment which creates a word percept rather than a non-word percept (Ganong, 1980). The quantitative top-down influences at the segmental and lexical level provide strong evidence for the dynamical systems approach to speech articulation. Familiar segments and familiar words are well-worn paths through the space of articulatory possibilities, and speech articulation tends to follow these familiar patterns. This conception has close ties to interactive activation models of speech production (Dell, 1986) which provide top-down influences from a variety of levels. The lowest level in these models is usually the linguistic feature. We propose to extend this class of models to encompass the individual articulatory gestures, which are the building blocks of the features and segments (Browman & Goldstein, 1990). The tendency toward categorical behavior in such a model is probabilistic, and there can be both normal and abnormal variation depending on the initial conditions of the system and external perturbations.

## Summary and Conclusion

This study presented an acoustic analysis of /s/ and /z/ speech errors by four talkers along three different acoustic dimensions. Cases of atypical productions along one dimension that coincided with normal productions along the other dimensions were found. In addition, the distribution of acoustic properties showed similar multi-modality for both /s/ and /z/ productions. Together these facts suggest that sub-lexical, non-contrastive errors can occur, in support of the claims of Mowrey and MacKay (1990). Analysis of the statistical cooccurrence of atypical productions along each dimension showed



that, while such productions are common, there is a tendency toward categorical behavior. Typical or categorically atypical productions were found more often than expected by chance, and intermediate productions were found less often than expected. The bias toward categorical productions interacted with the lexical status of the error, with stronger categorical tendencies for word outcomes, and weaker tendencies for non-word outcomes. An interaction with segmental phonetic context of the intended /s/ or /z/ was also found. Segmental context statistically influenced the acoustic properties of the productions, which in turn influenced the rate of occurrence of typical productions and the perception of these productions as speech errors. These patterns provide clear evidence for a set of higher-level units which group gestures together into gestural constellations at the level of the segment and word, agreeing with some of the observations of traditional speech error analyses based on transcribed data.

While some of the theoretical conclusions of transcription-based speech error analyses are supported, others must be rejected. Assertions that speech errors result in grammatically acceptable utterances are not supported (e.g., Fromkin, 1971). The detailed acoustics of some of the productions in our corpus, specifically those resulting in gradient errors, are distinctly anomalous. For example, errors where something like a devoiced [z] is produced in place of intended [s] are perceptually quite striking upon careful listening, and look highly unusual when a waveform display is consulted. Frisch and Wright (1997), in a detailed descriptive analysis of talker 2's productions, also found instances of blend errors or fluent corrections which produced [sz] or [zs] clusters. Such clusters are not phonotactically grammatical onset sequences. These findings support the claims of a growing number of researchers: that transcription is inadequate for complete error coding, and that transcription makes assumptions about the categorical and abstract nature of the data which are incorrect (e.g., Boucher, 1994; Ferber 1995; Laver, 1980). Transcription techniques are susceptible to perceptual bias, and so can be expected to be effective only when bias is minimized. Thus, data from experimentally elicited errors that are recorded and can be examined repeatedly are much more reliable than naturally occurring errors that are transcribed opportunistically while the 'experimenter' is engaged in conversation or otherwise not completely focused on the transcription task. But even these careful listening techniques will still be influenced by listener bias, and sub-lexical errors are likely to be missed.

The categorical tendencies in the acoustics of the four talkers' productions provide evidence that gestures tend to be organized into permissible segments. This tendency interacted with the lexical status of the outcome, suggesting that gestures tend to be organized into permissible words. Previous studies based on perceived errors found that errors which produce word outcomes are more frequent than errors which do not (Dell & Reich, 1980; Motley & Baars, 1975), providing converging evidence for this conclusion. Extending the tendency to produce phonetically normal segments and words to the more general case can explain the traditional claim that errors result in grammatical sequences of segments. Errors tend to result in grammatical sequences. Stemberger's (1983) very carefully collected corpus of naturally occurring errors supports the claim that grammaticality is a tendency, but not an absolute. He found some perceptually detectable but phonotactically anomalous productions in naturally occurring dialogue. The agreement between our analysis and Stemberger's corpus also suggests that our conclusions are not specific to tongue twisters, and that they apply to speech production in general.

Overall, our conclusions are incompatible with traditional models of speech production which involve selecting, organizing, and ordering abstract discrete representational units into a frame which is then implemented by the phonetic module (e.g., the scan-copier model of Shattuck-Hufnagel, 1987). The interaction between the articulatory details of speech production and organization at the lexical and segmental level suggests a hierarchically organized and interactive mechanism involving graded categories. Appropriate models involving discrete linguistic levels and spreading activation (Dell, 1986) or less explicitly organized connectionist architecture (Dell, Juliano, & Govindjee, 1993) have been previously been applied to speech errors at the level of features and segments. These models are

compatible if they are extended to include the level of individual articulatory gestures. Such models fit into the broader class of dynamical systems approaches to speech production and share many architectural properties with dynamical systems implementations (Saltzman, 1991, *inter alia*). Dynamical systems models of speech production have been applied to modeling context effects on articulation, including assimilation, coarticulation, and coproduction (Browman & Goldstein, 1990; Saltzman, 1986; Saltzman & Munhall, 1989) and have successfully accounted for the semi-categorical nature of normal speech production.

This paper demonstrated the semi-categorical nature of anomalous speech production using acoustic measures. Under the dynamical systems approach, anomalous speech production can be seen as no different from normal production; both lie within the same space of gestural possibilities. Anomalous speech production is merely in the remote regions of the gestural landscape, farther away from the well-worn paths of familiar segments, words, and utterances.

### References

- Baars, B., Motley, M., & MacKay, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382-391.
- Boucher, V. (1994). Alphabet-related biases in psycholinguistic enquiries: considerations for direct theories of speech production and perception. *Journal of Phonetics*, 22, 1-18.
- Browman, C. & Goldstein, L. (1990). Tier in articulatory phonology, with some implications for casual speech. In J. Kingston & M. Beckman (eds.), *Papers in laboratory phonology: Between the grammar and physics of speech*. Cambridge: Cambridge University Press. 341-376.
- Cole, R. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, 13, 153-156.
- Dell, G. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 222-233.
- Dell, G. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.
- Dell, G., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, 17, 149-195.
- Dell, G. & Reich, P. (1980). Toward a unified theory of slips of the tongue. In V. Fromkin (ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press. 273-286.
- Ferber, R. (1995). The reliability and validity of slip-of-the-tongue corpora: a methodological note. *Linguistics*, 33, 1169-1190.
- Frisch, S. (1996). Similarity and frequency in phonology. Unpublished Ph.D. Dissertation, Northwestern University, Evanston, IL.

- Frisch, S. (in press). Temporally organized lexical representations as phonological units. In M. Broe & J. Pierrehumbert, *Papers in laboratory phonology V: Language acquisition and the lexicon*. Cambridge: Cambridge University Press.
- Frisch, S. & Wright, R. (1997). A preliminary acoustic study of errors in speech production. *Research on spoken language processing: Progress report 21*. Pp. 503-516. Bloomington, IN: Speech Research Laboratory, Indiana University.
- Fougeron, C. & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America.*, 101, 3728-3740.
- Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47(1), 27-52.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology*, 6, 110-125.
- Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D., & Cutler, A. (1982). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics*, 19, 805-817.
- Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics*, 6, 95-102.
- Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*. 59, 1208-1221.
- Laver, J. (1980). Slips of the tongue as neuromuscular evidence for a model of speech production. In H. Dechert & M. Raupach (eds.), *Temporal variables in speech*. The Hague: Mouton. 21-26.
- Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Lieberman, A. (1997). *Speech: A special code*. Cambridge, MA: MIT Press.
- Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (ed.), *The production of speech*. New York: Springer-Verlag. 217-245.
- Marslen-Wilson, W. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.
- Mowrey, R. & MacKay, I. (1990). Phonological primitives: Electro-myographic speech error evidence. *Journal of the Acoustical Society of America.*, 88(3), 1299-1312.
- Motley, M. & Baars, B. (1975). Encoding sensitivities to phonological markedness and transition probability: evidence from spoonerisms. *Human Communication Research*, 2: 351-361.
- Pickett, J. (1980). *The sounds of speech communication*. Baltimore, MD: University Park Press.
- Pirello, K., Blumstein, S.E. & Kurowski, K. (1997). The characteristics of voicing in syllable-initial fricatives in American English. *Journal of the Acoustical Society of America*, 101, 3754-3765.
- Saltzman, E. L. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research*, 15, 129-144.

- Saltzman, E. L. (1991). The task dynamic model in speech production. In H. F. M. Peters, W. Hulstijn, and C. W. Starkweather (eds.), *Speech motor control and stuttering*. New York: Elsevier Science Publishers. 37-52.
- Saltzman, E. L. & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333-382.
- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110, 474-494.
- Shadle, C. (1985). *The acoustics of fricative consonants*. Cambridge, MA: MIT Press.
- Shattuck, S. (1975). Speech errors and sentence production. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (ed.), *The production of speech*. New York: Springer-Verlag. 109-136.
- Shattuck-Hufnagel, S. (1987). The role of word onset consonants in speech production planning. In E. Keller & M. Gopnick (eds.), *Motor and sensory processing in language*. Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42, 213-259.
- Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review*. Bloomington, IN: Indiana University Linguistics Club.
- Stevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, 3, 32-49.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 176, 392-393.
- Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine*, December, 9-12.
- Wickelgren, W. (1965). Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America*, 38, 583-588.
- Wright, R., Frisch, S., & Pisoni, D. B. (in press). Speech perception. In J. Webster (ed.), *Encyclopedia of electrical and electronics engineering*. New York: John Wiley & Sons.

**Figure 1.** Mean duration (top panel), frication amplitude (middle panel), and percent voicing (bottom panel) for each talker's productions of /s/ and /z/. Error bars indicate 95% confidence intervals for the mean.

**Figure 2.** Example measurement of /z/ in *zilch* showing waveform and high pass filtered waveform. Relevant regions for measuring duration, frication amplitude, and percent voicing are indicated by braces.

**Figure 3.** Distribution of duration values for /s/ and /z/ for each talker. Data are aggregated into groups of 20 ms.

**Figure 4.** Distribution of frication amplitude values for /s/ and /z/ for each talker. Data are aggregated into intervals of 200 units for talker 1 and 100 units for talkers 2, 3, and 4.

**Figure 5.** Distribution of percent voicing for /s/ and /z/ for each talker. Data are aggregated into intervals of 20%, but the endpoints 0% and 100% are displayed separately.

**Figure 6.** Single dimension errors for duration, frication amplitude, and percent voicing for four levels of criteria. Criteria are either strict or lax, for both the dimension in question and the other two dimensions (see text).

**Figure 7.** Ratio of observed to expected number of productions with no category crossings (non-errors), one or two category crossings (gradient errors), or three category crossings (categorical errors) for all talkers under lax criteria.

**Figure 8.** Ratio of observed to expected number of non-errors, gradient errors, and categorical errors for each talker under lax criteria.

**Figure 9.** Category crossing rate for percent voicing for all talkers along with perceived error rate from 21 participants in Frisch (1996) for /s/ (top panel) and /z/ (bottom panel).

**Figure 10.** Ratio of observed to expected number of non-errors, gradient errors, and categorical errors for all talkers for word outcome twisters (left panel) and non-word outcome twisters (right panel).