Journal of Phonetics (2002) **30**, 139–162
doi:10.1006/jpho.2002.0176
Available online at http://www.idealibrary.com on IDEAL®

The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue

Stefan A. Frisch

Department of Communication Sciences and Disorders, University of South Florida, 4202 E Fowler Ave, PCD1017, Tampa, FL 33620, U.S.A.

Richard Wright

Department of Linguistics, University of Washington, Box 354340, Seattle, WA 98195-4340, U.S.A.

Received 22nd November 2000 and accepted 7th February 2002

Acoustic analysis was used to examine whether speech errors involve lexical, segmental, or sub-featural errors in speech production. Nine participants produced tongue twisters that induced errors between /s/ and /z/ word onsets in contexts where the error outcomes were either words (e.g., *sit* to *zit*) or nonwords (e.g., *suck* to **zuck*). Three measurements of the /s/-/z/ contrast were made: (1) percent voicing, (2) duration of frication, and (3) amplitude of frication. The tokens were also transcribed under careful listening conditions. Gradient and categorical errors were found for all acoustic dimensions. The errors might or might not be detected by careful listening, depending on the extent to which there were errors along all three dimensions. These data support previous articulatory studies that found speech errors at a sub-featural level. However, cases where /s/ and /z/ are realized with a categorical change in voicing are more common than would be expected if categorical changes in voicing were merely extreme examples of gradient voicing errors. Also, both gradient and categorical error rates were higher when the error outcomes were words. Thus, our study also provides evidence for the psychological reality of phonological segments and words as units in the speech production process.

1. Introduction

Speech errors have traditionally been used to provide evidence for models of speech production that 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

Address correspondence to: S. A. Frisch. E-mail: frisch@chuma1.cas.usf.edu

0095–4470/02/$ - see front matter

of speech production can be drawn from speech error data. In addition, several researchers have questioned the validity of phonological speech error data that has been recorded using phonetic transcription (Laver, 1980; Mowrey & MacKay, 1990; Boucher, 1994; Ferber, 1995). In this paper, we undertake an acoustic analysis of speech produced by nine 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 transcriptional analyses. However, we also find evidence for speech errors at a sub-featural or gradient phonetic level that have not previously been attested. Our data support a model of speech production where individual gestures are organized into gestural constellations at the level of the segment (Saltzman & Munhall, 1989; Byrd, 1996). Segments are further organized into words. We find that the segment and word levels influence the implementation of gestures in both erroneous and error-free productions.

1.1. *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 in the speech error literature that the misordered portions correspond to abstract linguistic units such as onsets, codas, phonemes, segments, and features. The errors in (1) are given by Fromkin (1971) as support for the psychological reality of segments, distinct from words or syllable onsets.

(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 involve an exchange of the voicing feature. This demonstrates that errors involving linguistic features are possible and thus that features are psychologically real units of processing as well.

Mowrey & MacKay (1990), using electromyographic (EMG) recordings of tongue twister production 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 the remainder of this section, we review the transcriptional methods of speech error analysis and the results of the instrumental study of Mowrey & MacKay (1990).

1.2. *Transcriptional approach*

Traditional approaches to speech error analysis use phonetic transcription to encode speech errors at the time they are heard. In "naturally occurring" speech error corpora, errors that are observed in everyday speech are written down opportunistically. In the early corpora (e.g., Fromkin, 1971; Shattuck, 1975) the error recorders were usually 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 & Reich, 1980), though usually the utterances themselves are recorded on tape or computer and listened to repeatedly.

In all cases where transcription is used, the noting of a speech error necessarily coincides with the hearer noticing an anomalous percept. Thus, in transcriptional approaches, a speech error is defined to be an utterance that produces an anomalous percept that would be recognized as anomalous by the speaker (Dell, 1986). Mowrey & 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 articulatorily detailed the speaker's intentions are, which we discuss below.

Transcribed speech error evidence 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 and phonotactically grammatical (Wells, 1951; Fromkin, 1971). In other words, it is claimed that speech errors occur by misordering abstract phonological 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 some disagreement on these conclusions among experimenters using the same collection techniques. For example, Stemberger (1983), based on his own corpus of naturally occurring errors, claimed that phonologically 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 (Laver, 1980; Garnham *et al.*, 1982; Shattuck-Hufnagel, 1983; Mowrey & MacKay, 1990; Boucher, 1994; Ferber, 1995). There are two primary criticisms. First, the use of phonetic transcription cannot capture sub-contrastive or gradient errors, below the level of a segment or feature, since the transcription system is inherently segmental. If gradient 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 *et al.*, 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 transcription task is a forced-choice decision—is it an error or not—rather than an unconstrained 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, 1999, for a recent review). For example, in the phenomenon of categorical perception, phonetically anomalous speech sounds that 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). In another phenomenon, known as phonemic restoration, speech samples that 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 (Warren, 1970; Samuel, 1981). 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 transcriptional speech error analyses are informative of linguistic biases in the speech production process or merely a reflection of the hearer's perceptual system.

1.3. *Instrumental data*

Mowrey & MacKay (1990) present 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 noncontrastive errors occur. They acted as their own subjects. The tongue twisters they used, shown in (2), crucially involved segments with lower lip (/ʃ/ in 2a) or tongue blade (/l/ in 2b, c) articulations that were in proximity to segments that did not contain such articulations (/s/ and /r/ or nothing, 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 a perceptual definition based on transcription. 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 gradient errors do occur, and they further claim that 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 which is 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. Preliminary data from a recent study of one speaker using magnetometry supports the claim that gradient activation results in partial or incomplete gestures (Pouplier, Chen, Goldstein & Byrd, 1999).

These articulatory studies of gestural activity during tongue twister production clearly reveal that there can be 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 the phonetic details of an utterance, these instrumental studies have yet to explore how these gradient errors fit into the broader pattern of speech production. It is possible that other muscles were also gradiently activated in a coordinated fashion during these errors. Patterns of muscular coordination in speech errors would provide evidence of organization at the level of the feature or segment in speech production. If physically independent articulations for a segment are found to coordinate in gradient errors, such as lip rounding and tongue blade raising for /ʃ/, then the most plausible explanation would be the partial activation of a higher-order segmental planning unit. It is also possible that the behavior of other muscles in speech errors is independent. The available articulatory evidence to date demonstrates that gradient speech errors can and do frequently occur at the sub-segmental level, but this data does not rule out a role for higher levels of organization in speech production as a constraint on speech errors.

It is also worth noting that defining 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. These articulatory studies have found many of these gradient errors to be imperceptible, even upon repeated listening with high-quality equipment. It may be that these gestures have no significant acoustic effect. A possibility that should be considered is that one goal when talking is to convey a distinct, contrastive message to an observer. Thus, random variation that is noncontrastive may not be monitored as strictly and for this reason such productions might not be considered erroneous by the producer or detected as erroneous by the monitoring mechanisms in the production system of the producer.

1.4. *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 an individual articulator but little information about the coordination of articulators or the acoustic effects of the articulation. 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. In addition, acoustic analysis may reveal systematic biases in the perception of speech errors. Acoustic analysis has the further benefit that larger numbers of experimental subjects can be studied. The following section contains a description of the data and analysis techniques used in our acoustic analysis of phonological speech errors.

2. Methods

2.1. *Data*

The data for our analysis come from a recorded corpus of utterances from a speech error experiment that induced errors using tongue twisters (Frisch, 1996, Chapter 9). In the 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 experiment was self-paced and the speech rate of the participants was not controlled. The tongue twisters were printed individually in large type on index cards. The participants read each tongue twister aloud three times, and then repeated the tongue twister from memory three times. If they forgot the correct words, participants were allowed to consult the index card between repetitions during the repeat-from-memory portion. Each participant had a break after half of the experiment was completed in which they were engaged in normal conversation for about 5 min. The experiment lasted about 45 min in total. 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 22 different consonant pairs. Each consonant pair was used in four different tongue twisters. Each tongue twister consisted of four monosyllabic words. We examined the /s/-/z/ tongue twisters from the experiment, given in (3). Two of the tongue twisters created target errors that were existing words (3a, c), and two created nonword target 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. Thus, lexical frequency effects should not differentially influence the relative error rate between the consonants, from intended /s/ to [z] or intended /z/ to [s] in this particular case.[1]

The acoustic analysis focuses on the single consonant pair, /s/-/z/. Data from nine participants in the original experiment were analyzed. They were the first nine that participated in the experiment. 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 distribution of both categorical and gradient speech errors. The choice of the first nine participants in the experiment was arbitrary. Five of these participants were male and four were female. These participants produced a total of 448 /s/ tokens and 446 /z/ tokens. The number of productions of each segment in each word by each participant is not necessarily equal, as the participants sometimes repeated individual words or the entire tongue twister after producing an error. All productions were measured, including repetitions.

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. Note that Shattuck-Hufnagel's (1992) study used transcribed speech error evidence. 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 in the action of the production mechanism, and that the same phonological encoding and phonetic implementation is involved in producing both tongue twisters and natural speech.

2.2. *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 that can be plausibly associated with independent aspects of the articulation. In describing the data acoustically, tokens are categorized as intended /s/ or /z/ based on the word the speaker was supposed to say if the tongue twister was to be produced correctly. This was assumed to be the speaker's intended utterance. Both perceptually normal and erroneous tokens are included in the figures and tables that follow.

[1]For convenience, we borrow the slash and bracket notation of phonology to differentiate the intended production from the recorded utterance. The "underlying" /s/ and /z/ represent the intended productions of the participants, and the "surface" [s] and [z] (or other transcriptions in square brackets) represent the actual output.

Three measures are reported in this paper. These measures differentiate /s/ and /z/ tokens produced by our talkers and have articulatory bases that are sufficiently physiologically independent for our analysis. In addition to measuring each token, we transcribed each token based on repeated careful listening to digitized recordings of the individual repetitions of the tongue twisters. In most cases, the result was a clear /s/ or /z/ percept. However, there were several tokens of /z/ that were transcribed as devoiced.

The twister *sung zone Zeus seem* contained a sequence of coda /s/ followed by onset /s/. In 40 of the 55 repetitions of this twister, the participants did not produce two distinct /s/ segments (though in most cases there were two distinct amplitude peaks). These "geminate" /s/ tokens were not included in the analysis as they could not be unambiguously measured. In addition, there were 10 instances of "concatenation" errors where the participant produced a long [sz] or [zs] sequence, three instances where [ʒ] was produced in place of [z], and three instances that were perceptually bizarre and untranscribable. We discuss these tokens in Section 3.3, and they are excluded from our analysis in order to provide the clearest possible picture of normal /s/ and /z/ production by our participants. Given these criteria, data for 397 /s/ productions and 435 /z/ productions are reported in this section.

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 any measurable overlap of the frication noise with the formant structure of the following vowel and the preceding vowel or sonorant, if any. In general, this overlap was common for /z/ but lacking for /s/. We determined whether or not fricative noise was present by examining the waveform for the fuzziness of aperiodic noise and a spectrogram for evidence of broadband noise at high frequencies. The expected pattern was for /s/ to be longer than /z/ (Klatt, 1976), which was true across all nine talkers in our study. Fig. 1 shows means and error bars representing one standard deviation for each participant for our three measures of /s/ and /z/. Average duration is given in the top panel of Fig. 1, with the average for /s/ shown by the shaded bar, and the average for /z/ shown by the clear bar.

The second measure is the amplitude of frication noise. Amplitude of frication noise was calculated based on the RMS amplitude of the signal after high-pass filtering at 2 kHz. For each fricative, RMS amplitude was computed for 10 ms 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 50 ms 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/ (Strevens, 1960; Pickett, 1980), which was true across all nine participants in our study (see Fig. 1, middle panel).

The final measure is the percent of the fricative that contained voicing. We used the waveform to determine when voicing was present, and considered portions of the signal that 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. Breathy 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 nine talkers in our study (see Fig. 1, bottom panel).

The measurement of duration, frication amplitude, and percent voicing for a sample case of /z/ is shown in Fig. 2. This is a production of *zone* where fricative noise overlaps with the coda nasal of the previous word (*sung*). The fricative is devoiced in the middle, but voicing begins again before the vowel onset. This pattern is relatively common for /z/ (Haggard, 1978; Smith, 1997). The top panel in Fig. 2 shows the original signal, with brackets indicating the overall length of frication and the portions that are voiced. The bottom panel shows the signal filtered at 2 kHz, which has removed most of the influence of periodicity, and the 50 ms window around the peak over which the RMS amplitude of frication was computed.

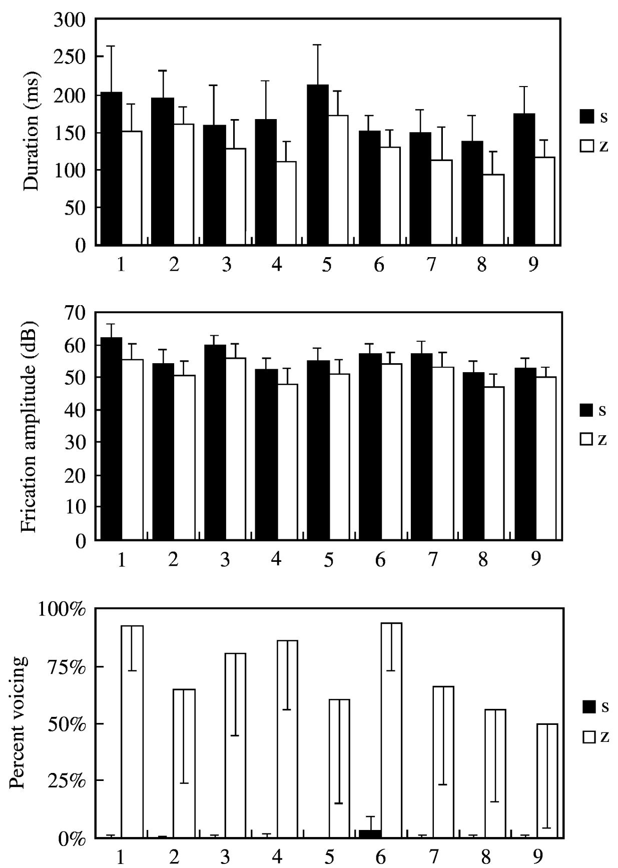

Figure 1. Mean duration (top panel), frication amplitude (middle panel), and percent voicing (bottom panel) for each participant's productions of /s/ and /z/. Error bars indicate one standard deviation around the mean.

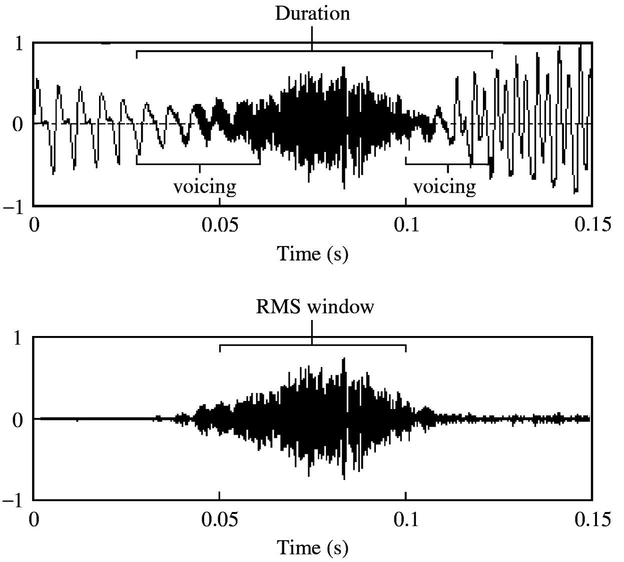

Figure 2. Example measurement of /z/ in *zone* showing waveform and high-pass filtered waveform. Relevant regions for measuring duration, frication amplitude, and percent voicing are indicated by braces.

While our goal was to consider measurements that are articulatorily independent, it is clear that duration, amplitude of frication, and percent voicing are not unrelated aerodynamically (Ohala, 1983). 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, many speakers produce /z/ with a voiceless portion in the middle of the fricative. During this devoiced portion, airflow is not constrained by glottal closure. Our measure of frication amplitude over the loudest 50 ms window would allow the frication amplitude of /z/ to be determined by the devoiced portion, if there is one. In addition, talkers also actively increase the pulmonic airflow in /s/ to produce a noisier fricative (Shadle, 1985). In an error articulation, the oral and glottal gestures for /s/ may be in place, but if this is not combined with increased pulmonic airflow, the resulting frication amplitude will be less than normal for /s/.

Duration may also interact with voicing. During a fricative, pressure buildup behind the oral constriction equalizes transglottal pressure, making voicing more difficult. Thus, long duration voiced fricatives may have a stronger 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 predicts that long /z/ will devoice, not that a very long fricative is necessarily entirely voiceless. We feel that the dependence between acoustic properties measured in our study cannot wholly be explained by aerodynamic interdependence between our measures, and variation is to some extent under the active control of the talker. Thus, there is room for more abstract levels of linguistic organization to influence the interrelation of our measures.

3. Categorical, gradient, and ungrammatical errors

In the previous section, it was demonstrated that the three dimensions of duration, frication amplitude, and percent voicing differentiate /s/ and /z/ for our participants. Of these, the percent voicing measure most directly reflects the voicing contrast. There is considerable overlap between /s/ and /z/ in the distribution of frication amplitude and duration, even among tokens that were perceived to be correctly produced. As a result, the effects of frication amplitude and duration are likely to be secondary in influencing the percept, and less definitive in determining when an error has occurred (Baum & Blumstein, 1987). In this section, we first investigate variation in percent voicing, and then consider the variation in frication amplitude and duration. Finally, we briefly discuss the other disfluent productions found in our corpus. Overall, the presence of these strange utterances and the considerable variability in production even for percent voicing supports the claim that many ungrammatical speech error outputs occur (Mowrey & MacKay, 1990).

3.1. *Variation in percent voicing*

The percent voicing of all tokens for each participant is shown in Fig. 3. The data are grouped to illustrate productions that are usual and unusual for each participant. Both 0 and 100% voicing are given separately. The other groups are 0–5, 5–10, 10–30, 30–60, and 60–100%. Each panel of Fig. 3 shows the number of occurrences for one participant. Within each panel, the shaded bars indicate occurrences for intended /s/ and the clear bars indicate occurrences for intended /z/. The data in Fig. 3 are summarized in Table I, where the aggregate number of occurrences for each percent voicing category for all participants is given for both /s/ and /z/.

Given that these tongue twisters were recorded in an experimental laboratory setting, we assume the participants were hyperarticulating and attempting to produce clear speech. Thus the the target percent voicing is likely to be 0% for /s/ and 100% for /z/. In normal speech, partial devoicing of /z/ results in productions that are voiceless in the middle, though these productions are usually still more than 60% voiced (Haggard, 1978; Smith, 1997). Fig. 3 and Table I show that many productions did not fall into the percent voicing groups that would normally be predicted for /s/ and /z/. Utterances where the percent voicing was between 5 and 30% are certainly anomalous for both /s/ and /z/. These productions with intermediate amounts of voicing provide evidence that gradient, noncategorical errors are made (Mowrey & MacKay, 1990).

Overall there are fewer cases of intermediate voicing for /s/ than for /z/.[2] Among the utterances with intermediate voicing, /s/ also shows a clear pattern of variability. As the percent voicing increases, fewer and fewer tokens of /s/ are found that have that degree of voicing. This is the pattern we might expect for phonetic variability if

[2]Participant 6 is a curious exception to the generalizations about the likelihood of gradient errors in /s/ and /z/. For participant 6, /s/ is more variable, and was overlapped with the following vowel (and the preceding vowel if there was one) a small amount in many utterances. On the other hand, /z/ for participant 6 is much less variable, with almost all instances realized with 100% voicing. Participant 6 is also atypical in the production of categorical errors, and does not follow the /s/ or /z/ pattern of the majority of the participants as discussed below.

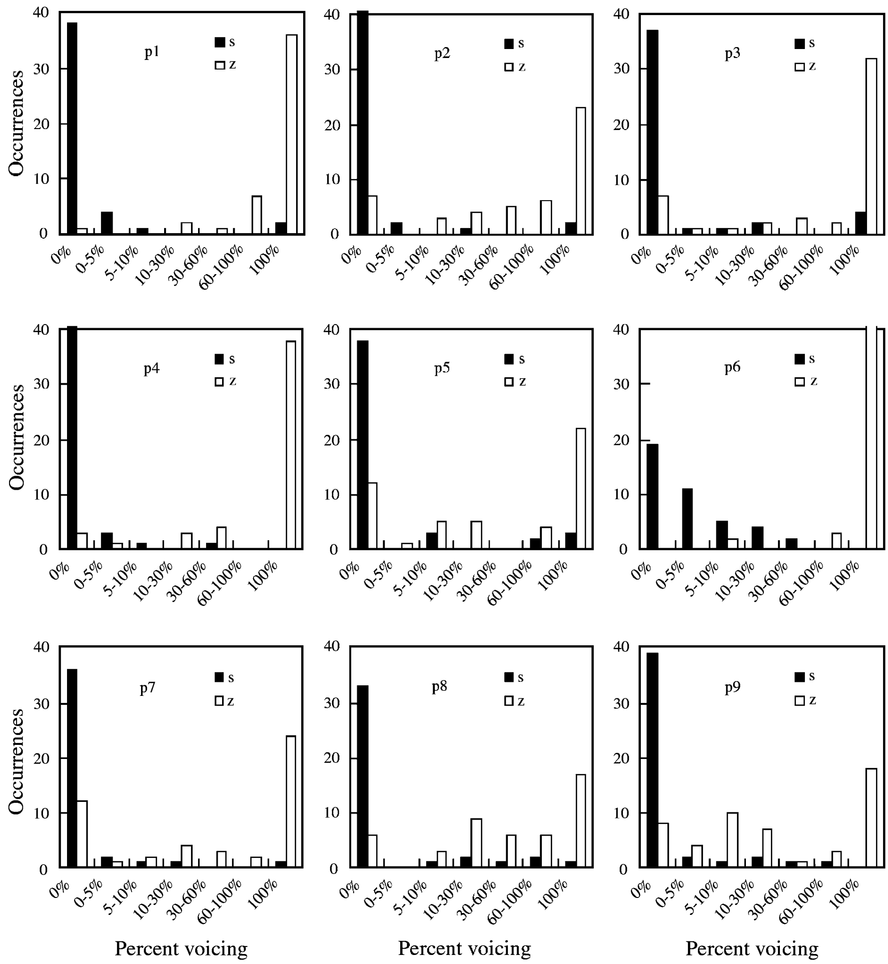

Figure 3. Distribution of tokens among percent voicing groups for /s/ and /z/ for each participant.

Table I. Distribution of intended /s/ and /z/ tokens by percent voicing for all participants

	Intended segment	
Percent voicing	/s/	/z/
0%	324	56
0–5%	25	8
5–10%	13	24
10–30%	12	39
30–60%	4	23
60–100%	6	33
100%	13	252
Total	397	435

0% voicing is intended. Productions that are farther away from the intended are relatively less likely, as they lie on the tail of the distribution. By contrast, the devoicing of /z/ was relatively common and does not clearly reflect degrees of variation away from an intended production of 100% voicing. Levels of voicing between 5 and 100% occur with relatively equal frequency and there are no apparent systematic patterns. It may be that, since the devoicing of /z/ is common in normal productions, the data here are mixture of normally devoiced /z/ and cases where /z/ is partially devoiced in error.

Fig. 3 and Table I also provide distributional evidence for categorical errors. Across all participants, there are 18 cases where /s/ was produced with more than 60% voicing (out of 397 total). Compared to the five cases with 30–60% voicing and 12 cases with 10–30% voicing, the number of completely voiced tokens is relatively large. For the individual data of six of nine participants the number of tokens with more than 60% voicing was greater than the number of tokens with 10–60% voicing. This suggests that in addition to the peak in the distribution of /s/ at 0% voicing, there is a second peak in the distribution of /s/ at 100% voicing. We interpret this second peak to be the result of categorical changes from intended /s/ to [z].

There is also phonetic evidence that the tokens of /s/ with greater than 60% voicing are categorical errors. These tokens have shorter duration and lower frication amplitude than the typical /s/. In addition, these tokens have overlap between the frication noise and the onset of the formant structure of the following vowel that is typical of /z/ in our data. Finally, these fricatives have amplitude envelopes that gradually increase from the amplitude of the fricative toward the amplitude peak of the following vowel, which is also typical for /z/ in our data.

There are 56 cases where /z/ was produced with 0% voicing (out of 435 total). The number of completely voiceless cases is larger than the number of cases with low levels of intermediate voicing (0–10%) for the individual data of seven of the nine participants. In about half of these cases, these also appear upon inspection to be categorical errors, with /z/ produced as [s]. In these cases, there is /s/-like duration and frication amplitude. There is a small positive VOT between the offset of /z/ and the onset of voicing and formant structure for the following vowel. There is also an abrupt decrease in frication amplitude of the /z/ followed by an abrupt rise in vowel amplitude to the vowel amplitude peak that is characteristic of /s/ in our data. However, there are other cases of 0% voicing that lack these additional /s/-like features where the result may be better transcribed as [z̥].

One example of an apparent categorical intended /z/ to [s] error is given in Fig. 4. The top panel of Fig. 4 shows [s] in the production of *zap* by participant 5. The bottom panel shows an error-free production of *sap* by the same participant. In both utterances, there is a relatively large amount of frication noise, and the offset of the fricative is complete before the onset of the following vowel begins.

An example of an error where intended /z/ is realized as [z̥] is given in Fig. 5. The top panel of Fig. 5 shows [z̥] in the production of *zap* by participant 9. The bottom panel of the figure shows an error-free production of *sap* by the same participant. In contrast with the categorical example above, the [z̥] has a much lower frication amplitude, and the frication noise offset is at about the same time as the vowel onset. So while there is no measurable voicing during the fricative, there is no delay

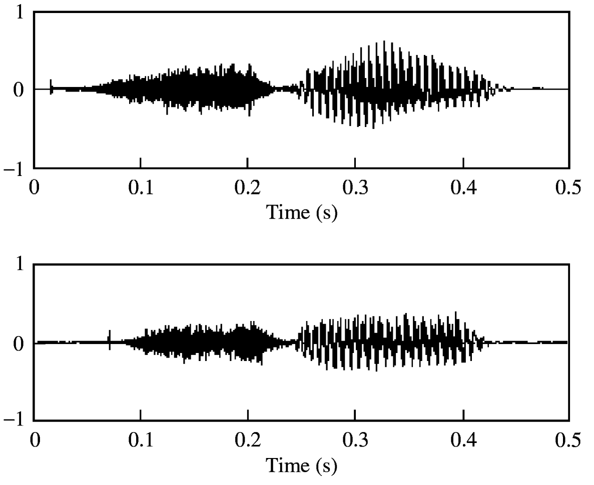

Figure 4. Example categorical error in the production of *zap* (top panel) and normal production of *sap* (bottom panel) for participant 5.

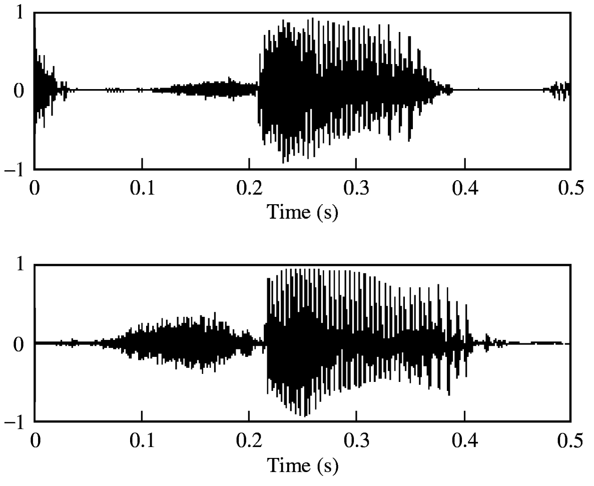

Figure 5. Example devoicing error in the production of *zap* (top panel) and normal production of *zap* (bottom panel) for participant 9.

in voice onset as there is with /s/. Thus, it appears that even though there is no voicing, this error is not a categorical change from /z/ to [s].

The variation in percent voicing suggests that more errors (whether gradient or categorical) were from /z/ toward [s] than from /s/ toward [z]. This pattern is expected, as /s/ is higher in frequency of occurrence than /z/ and in general low-frequency items are replaced by high-frequency items in speech errors (Stemberger, 1983). However, it has been reported in the speech error literature that /s/ and /z/ errors are one of several cases where the normal frequency effect is reversed (Stemberger, 1991). Stemberger found that /s/ to [z] errors were more common than /z/ to [s] errors in a transcribed speech error experiment designed to examine asymmetries in segmental errors.

Though the phonetic variability suggests that /z/ errors are more common than /s/ errors, variability in voicing in /z/ does not affect the perception of /z/ nearly as much as it affects the perception of /s/. Fig. 6 shows our judgement of the error rate based on our percept of each token. Tokens are aggregated into the same percent voicing groups as above, and the percent that were perceived as errors is plotted against the average percent voicing of tokens in the group. Our judgement of the percept was determined informally by consensus, based on repeated listening through headphones at a computer workstation. We found our level of agreement on the percept to be high. A token was considered to be an error if it was an /s/ that was not perceived as [s] or if it was a /z/ that was not perceived as [z]. While we did perceive some cases as [z̥], many cases that would be best described as [z̥] were not perceived to be anomalous. These cases were only identified as [z̥] by visual inspection of the waveform. In the perceptual data, only the cases of perceived [z̥] were counted as errors.

Fig. 6 shows a pattern of error detection that is strikingly similar to categorical perception with a boundary in the vicinity of 5% voicing (Liberman, 1997). Note that there is a heavy asymmetry in the perception of errors in /s/ and /z/ productions. While an /s/ with almost any voicing was perceived as [z], few cases of devoiced /z/ were perceived to be [s] or [z̥] . Even when there was no voicing at all in the /z/, it often sounded normal. Thus, while the production data suggest that there were more anomalies in the /z/ tokens than the /s/ tokens, the perception of this variation was heavily biased to detect /s/ errors but not /z/ errors. This perceptual asymmetry explains the apparent reversal of the frequency effect that was found by Stemberger (1991) and provides further evidence that speech error transcription is an unreliable method for examining the speech production process.

3.2. *Variation in frication amplitude and duration*

Unusual variation in percent voicing reveals that there are many gradient errors in the production of /s/ and /z/ in our corpus. It is less clear how the dimensions of frication amplitude and duration might reveal gradient errors, since these dimensions are not fully contrastive. Even though the distributions of frication amplitude and

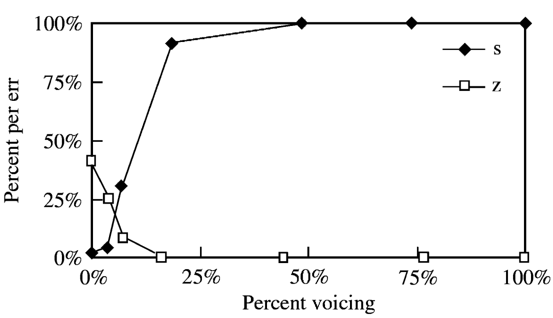

Figure 6. Perceived error rate based on the authors' percepts. Tokens of /s/ and /z/ are grouped across all participants for the same percent voicing groups in Table I and Fig. 3. Each point shows the mean percent voicing for tokens in the group (*x*-axis) and the percentage of tokens in the group that were perceived as errors (*y*-axis).

duration do not categorically differentiate /s/ from /z/, they do provide secondary cues to the voicing distinction (Cole & Cooper, 1975; Shadle, 1985). We have shown that tokens with 30% or more voicing were reliably perceived as [z]. Cases with less voicing were much less reliably categorized. Closer examination of the data reveal variation in frication amplitude and duration, in addition to voicing, affected the perception of errors. In particular, for tokens with small amounts of voicing, high frication amplitude and long duration were cues for /s/, while low frication amplitude and short duration were cues for /z/.

Table II shows mean frication amplitude and duration for /s/ and /z/ depending on whether they were perceived as errors or not. The top portion of the table presents aggregate data for tokens with 0–5% voicing, the middle portion presents aggregate data for tokens with 5–30% voicing, and the bottom portion presents aggregate date for tokens with 30–100% voicing. While very few /s/ with 0–5% voicing were perceived as errors, those that were appear to have lower frication amplitude and shorter duration than those that were not (for amplitude $t(347)=1.7$, $p=0.09$; for duration $t(347)=1.6$, $p=0.06$). For /z/ with 0–5% voicing, the pattern was similar. Tokens with high frication amplitude and long duration were perceived as [s], those with low frication amplitude and short duration were perceived as [z] (for amplitude $t(62)=4.0$, $p<0.01$; for duration $t(62)=2.5$, $p<0.01$). For /s/ and /z/ with 5–30% voicing, frication amplitude and duration again appear to have an effect on whether the tokens were perceived as errors or not, though the only statistically significant difference is for frication amplitude for /s/ ($t(22)=2.9$, $p<0.01$). Finally, in cases of 30–100% voicing, there were no tokens that were not perceived as [z], so

Table II. Mean frication amplitude and duration for low, intermediate, and high percent voicing groups as a function of intended segment and perceived error for all participants

Group	*N*	Frication (dB)	Duration (ms)
0–5% *voicing*			
/s/			
Perceived error	7	53.0	137.1
Non-error	342	56.3	173.2
/z/			
Perceived error	25	54.5	163.2
Non-error	40	49.1	133.0
5–30% *voicing*			
/s/			
Perceived error	15	52.2	159.9
Non-error	10	57.1	148.8
/z/			
Perceived error	2	53.0	131.2
Non-error	61	51.0	127.6
30–100% *voicing*			
/s/			
Perceived error	23	51.3	161.6
Non-error	0	—	—
/z/			
Perceived error	0	—	—
Non-error	307	52.0	131.5

it appears that variation in frication amplitude and duration are not sufficiently powerful cues to overwhelm the perceptual effect of periodicity.

If we compare the fully voiced /z/ tokens with the /s/ tokens with 0–5% voicing that were perceived as errors, we see evidence that these voiceless /s/ tokens were produced with frication amplitude and duration appropriate for /z/. These /s/ tokens could reasonably be considered to be errors on the frication amplitude and duration dimensions but not on the voicing dimension. In other words, these may be cases where /s/ was produced as [z̥]. If this is the correct analysis, these tokens provide additional evidence for gradient errors in production.

3.3. *Other variation*

3.3.1. *Concatenation errors*

There were 10 tokens, six for intended /s/ and four for intended /z/, where the speakers appeared to correct their articulations without interrupting their production. In all of these cases, the token begins as the unintended segment and is corrected to the intended segment. In all cases, the resulting concatenation of onset consonants is 300–500 ms long, and thus long enough to be equivalent to the production of two separate segments. The resulting [zs] or [sz] onset cluster is, of course, phonotactically illegal in English. Fig. 7 shows two examples of concatenation errors. The top panel shows [zs] in a production of *sue* by participant 6. The bottom panel shows [sz] (or perhaps [z̥z]) for a production of *zig* by participant 3.

3.3.2. *Other disfluencies*

There are six other anomalous productions that are interesting to note. Three of these productions appear to be linguistically explicable speech errors, but not ones that were intended to be elicited by the experiment. All three cases involved

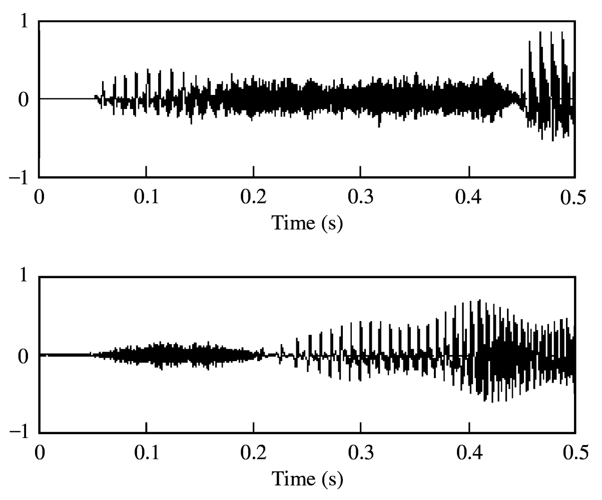

Figure 7. Example concatenation errors producing [zs] for *sue* (top panel) and [sz] for *zig* (bottom panel).

anticipation of the palatal articulation from the coda of the intended word *zilch* onto the onset of the same word. In two cases, one by participant 6 and one by participant 9, the result was [ʒ]. Inspection of the spectrograms of these two examples confirms that a voiced fricative was produced with frication noise in the same region as the coda [tʃ]. Participant 6 interrupted his production mid-word and immediately repeated the word correctly. In addition to being an unusual anticipation, the phonotactic legality of the [ʒ] onset is questionable in English. The third palatal intrusion, also by participant 6, was combined with an error in voicing. In this case, the palatal fricative was largely voiceless and of low amplitude. The frication noise overlapped with the onset of the following vowel, resulting in a short period of voiced frication. The perceptual result was [ʒ]. In this case, the participant once again interrupted his production mid-word, and then repeated the word correctly after a 1 s delay.

The three other productions are a nonhomogenous set of disfluencies. These cases are perceptually anomalous and could be considered speech production errors that are gross gestural mis-coordinations. One case is a low amplitude [z̥] like sound that lasts for about 500 ms before becoming a relatively normal amplitude and duration [s] onset. A second involves an intended /s/ where the alveolar constriction appears to end about 100 ms before the onset of the following vowel, resulting in an intrusive [h] percept. The third involves an intended /z/ that begins with about 100 ms of unfricated prevoicing followed by a 300 ms long fully voiced [z] onset.

3.3.3. *'Geminate' /s/*

As noted above, one of the tongue twisters contained the sequence *Zeus seem* in which a coda /s/ is followed by an onset /s/. There were 55 tokens of this sequence across all participants. In 40 cases, the two fricatives were produced without a discrete juncture. It is interesting to note that none of these long /s/ productions show signs of being voiced and none were perceived to be errors. In addition, the 15 cases where there were two distinct /s/ productions also show very little variation in voicing in the /s/ productions. Only one of the 15 distinct onset /s/ tokens has any voicing, and that case involves just 3% voicing from a small amount of overlap with the following vowel. Given that more than 15% of the other /s/ tokens contain some amount of erroneous voicing, the doubled articulation of /s/ appears to have somehow "protected" these segments from intrusion by /z/.

More general effects of segmental context on voicing of /s/ and /z/ (as well as /f/ and /v/) were found by Pirello, Blumstein & Kurowski (1997). They examined the effect of segmental context on the voicing of fricatives in nonsense phrases like *his fav sips it*. They found that voiceless fricatives following voiced segments were often initially voiced, and voiced fricatives following voiceless segments were often initially devoiced.

3.4. *Summary*

The concatenation errors and the many cases of intermediate amounts of voicing found in all nine participants in our study supports the conclusions of Mowrey & MacKay (1990). A significant number of tokens in our corpus are phonetically or phonotactically abnormal in English. Thus, it appears that gradient errors do occur frequently, and the process of phonetic accommodation does not regularize all speech errors to be acoustically normal English sounds. We also found that the gradient errors might or might not be detected as errors by careful listening.

On the other hand, there is also evidence for categorical effects in errors. Cases where /s/ and /z/ are realized with a categorical change in voicing are more common than we would expect if categorical changes in voicing were merely extreme examples of gradient voicing errors (i.e., the tails of the variation in the distribution of voicing). In addition, there is clear evidence that many cases of 0% voicing for /z/ also have duration and frication amplitude that is appropriate for /s/. It seems likely that these are instances of categorical errors, where a normal articulation appropriate to the wrong segment is produced. While there is some aerodynamic interdependence between voicing, frication amplitude, and duration, their interdependence is not so strong that a fricative of the length and amplitude of a normal /s/ must necessarily be completely voiceless. In fact, our corpus contains many productions of /z/ that have the length and frication amplitude of /s/ but are voiced throughout.

We conclude that a linguistic level of the segment or feature does have an influence on the phonetic details of speech error production. This effect can be captured in models of language processing that use graded activation and competition between linguistic units to explain speech errors in phonological encoding (e.g., Dell, 1986). Errors occur when competition results in the accidental mis-selection of an incorrect word, segment, or feature. In order to account for gradient speech errors, these models must be extended to include activation and competition among phonetic articulatory plans. While it is certainly not a trivial task to develop a working connectionist model of phonetic implementation, the conceptual extension is a simple one. If there is activation and competition in phonetic planning, there is the possibility for articulatory plans that mix gestures for different articulators from different segments. Complete devoicing of /z/ with a [z̥] result could be explained by a combination of the glottal abduction of /s/ with normal pulmonic effort associated with /z/. It is also conceivable that erroneous articulatory plans that blend articulatory parameters for the same articulator could be created. Partial voicing of /s/ could result from a combination of the articulatory gestures for /s/ with the timing of oral and laryngeal gestures for /z/ in which the fricative constriction is overlapped with the onset of voicing of the following vowel.

Activation and competition may also explain why "geminate" /s/ appeared to be unaffected by /z/. Elements of articulatory planning, such as the laryngeal abduction, may be more strongly encoded when onset and coda /s/ productions come together, as each segment would independently reinforce the activation of this gesture. Correct /s/ voicing articulation will be more successful in competing with /z/ as the result of this additional activation.

4. Lexical effects

In the previous section, we argued that there are observable segmental effects in the phonetics of speech errors. In our corpus, two of the tongue twisters were designed so that the elicited errors would result in words and two of the tongue twisters were designed so that the elicited errors would result in nonwords. We, therefore, also have the opportunity to examine whether the productions were influenced by the presence of a word error outcome. In the speech error literature, it has been reported that word error outcomes are more likely than nonword outcomes in both naturally occurring error corpora and in experiments that elicit speech errors (Motley & Baars, 1975; Dell & Reich, 1980; Stemberger, 1983).

Table III shows the tokens divided into groups by percent voicing, as in Table I, but with the tokens further divided by the lexical status of the error outcome of the tongue twister. In the case of /s/, it appears that all types of error in voicing were more common in the case of lexical outcomes, and the two distributions deviate significantly from a common distribution ($\chi^2(6) = 19.7$, $p < 0.01$). This provides clear evidence that the lexicality of the outcome affects both gradient and categorical aspects of speech errors.

In the case of /z/, the pattern is less clear, though the distributions are significantly different ($\chi^2(6) = 13.9$, $p < 0.05$). It appears that the number of tokens with 0–30% voicing is somewhat higher in the lexical case than in the nonlexical case, and the pattern is reversed for voicing from 30% up to 100%. As mentioned in Section 3, it may be the case that the distribution of percent voicing for /z/ reflects a mixture of normal devoicing and error-induced devoicing. The number of cases with highly reduced voicing in the lexical case may reflect a greater amount of error-induced devoicing on those tokens. The fact that there are fewer lexical cases than nonlexical cases with smaller amounts of devoicing is more difficult to explain. One possibility is that normally devoiced productions of /z/ are somehow more susceptible to additional error-induced devoicing than cases where /z/ is 100% voiced. If error-induced devoicing is stronger in the lexical case, then the number of tokens with only a small amount of devoicing would be differentially reduced.

Table III. Distribution of intended /s/ and /z/ tokens by percent voicing and lexical status of potential error outcome for all participants

Percent voicing	Lexical	Nonlexical
/s/		
0%	162	162
0–5%	20	5
5–10%	7	6
10–30%	11	1
30–60%	3	1
60–100%	5	1
100%	9	4
Total	217	180
/z/		
0%	34	22
0–5%	4	4
5–10%	15	9
10–30%	23	16
30–60%	7	16
60–100%	10	23
100%	129	123
Total	222	213

Given the clear lexical effects on /s/ production and some indication of lexical effects on /z/ production, we conclude that productions in which there is a competing lexical outcome produce greater numbers of errors. The effect of a competing lexical item on speech errors is found in an increase in both gradient errors and categorical errors. Together with the evidence from Section 3 that categorical errors in voicing are more common than expected based on the pattern of gradient errors, we conclude that the segment and word are both units of organization in the speech production process that affect the phonetic details of articulation.

The lexical effect we found in speech production further supports a connectionist approach to phonological encoding in speech production. If the articulatory plans for segments compete with one another during encoding, then competing segments that are reinforced by corresponding word nodes will be enhanced. For competing word outcomes, more errors will occur. When there is no word competitor, the correct segment will encounter less competition, and fewer errors will occur. Assuming that errors can involve both blending of articulatory plans and wholesale substitution, an increase in both gradient and categorical errors would be predicted in cases where there is a lexical competitor.

5. Discussion and conclusion

This study presented an acoustic and perceptual analysis of onset /s/ and /z/ speech errors by nine talkers. In support of the claims of Mowrey & MacKay (1990), we found that gradient, noncontrastive errors can occur, and that such errors are actually common. In addition, we found that categorical errors also occur at rates that are higher than would be expected if the only source of errors was from noncontrastive variation that happened to extend into another phonetic category. Finally, we demonstrated a lexical effect on both gradient and categorical errors. These patterns provide evidence for a set of higher level units that organize phonetic gestures 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 intermediate voicing errors, are distinctly phonetically anomalous. For example, errors where 10% of an intended /s/ is voiced by overlapping with the following vowel are clearly anomalous, and by no means similar to cases of normally devoiced /z/. In addition, the concatenation errors are instances of phonotactically ungrammatical onset sequences. While self-monitoring and editing certainly takes place in speech production (Levelt, 1989), it is not the case that the majority of speech errors in our corpus are best analyzed as mis-selection of phonological segments that have been phonetically encoded after the error occurred.

These findings support the claims of a growing number of researchers that transcription is inadequate for complete error coding, as transcription makes incorrect assumptions about the wholly categorical and abstract nature of the data (Laver, 1980; Boucher, 1994; Ferber 1995). 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 careful listening techniques will still be influenced by listener bias, and gradient errors are likely to be missed. In our own perception of the tokens in our corpus, there was a heavy perceptual bias to detect errors in /s/ production and not errors in /z/ production. The acoustically determined error pattern suggests that there were in fact more /z/ errors than /s/ errors.

The categorical tendencies in the acoustics of the productions in our corpus 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. 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 findings are incompatible with models of speech production that involve selecting, organizing, and ordering abstract discrete representational units into a frame that is then implemented by the phonetic module (e.g., the scan-copier model of Shattuck-Hufnagel, 1979). The interaction between the articulatory details of speech production and organization at the lexical and segmental levels suggests that there is a hierarchically organized and interactive production mechanism that manipulates phonetic units directly. Appropriate models that use discrete linguistic levels and spreading activation (Dell, 1986) or less explicitly organized connectionist architecture (Dell, Juliano & Govindjee, 1993) have previously been proposed as models of phonological encoding that sometimes generate speech errors at the level of words and segments. These models are compatible with our findings if they are extended to include competition at the level the articulatory plans of segments. We conclude from this that the phonetics of phonological speech errors is an important topic for future research in speech production.

Parts of this research were presented at the 133rd and 134th meetings of the Acoustical Society of America, the 72nd 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 questions and suggestions. We would also like to thank Dani Byrd, Terry Nearey, and two anonymous reviewers for helpful comments on an earlier draft of this paper

This work supported in part by NIH Training Grant DC 00012 to Indiana University and by the *Language Learning* Research Assistant Professorship at the University of Michigan.

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.

Baum, S. & Blumstein, S. (1987) Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English, *Journal of the Acoustical Society of America*, **82,** 1073–1077.

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

Byrd, D. (1996) A phase window framework for articulatory timing, *Phonology*, **13,** 139–169.

Cole, R. (1973) Listening for mispronunciations: a measure of what we hear during speech, *Perception & Psychophysics*, **13,** 153–156.

Cole, R. & Cooper, W. (1975) The perception of voicing in English affricates and fricatives, *Journal of the Acoustical Society of America*, **58,** 1280–1287.

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 *Errors in linguistic performance: slips of the tongue, ear, pen, and hand* (V. Fromkin, editor), New York: Academic Press. pp. 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 PhD dissertation, Northwestern University, Evanston, IL.

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

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 *Temporal variables in speech* (H. Dechert & M. Raupach, editors), The Hague: Mouton. pp. 21–26.

Levelt, W. (1989) *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Liberman, A. (1997) *Speech: a special code*. Cambridge, MA: MIT Press.

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.

Ohala, J. J. (1983) The origin of sound patterns in vocal tract constraints. In *The production of speech* (P. F. MacNeilage editor), New York: Springer-Verlag. pp. 189–216.

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.

Pickett, J. (1980) *The sounds of speech communication*. Baltimore, MD: University Park Press.

Pouplier, M., Chen, L., Goldstein, L. & Byrd, D. (1999) Kinematic evidence for the existence of gradient speech errors, *Journal of the Acoustical Society of America*, **106,** 2242.

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 PhD dissertation, Massachusetts Institute of Technology.

Shattuck-Hufnagel, S. (1979) Speech errors as evidence for a serial-ordering mechanism in sentence production. In *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (W. E. Cooper & E. C. T. Walker, editors), Hillsdale, NJ: Erlbaum. pp. 295–342.

Shattuck-Hufnagel, S. (1983) Sublexical units and suprasegmental structure in speech production planning. In *The production of speech* (P. F. MacNeilage, editor) New York: Springer-Verlag. pp. 109–136.

Shattuck-Hufnagel, S. (1992) The role of word structure in segmental serial ordering, *Cognition*, **42,** 213–259.

Smith, C. L. (1997) The devoicing of /z/ in American English: effects of local and prosodic context, *Journal of Phonetics*, **25,** 471–500.

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

Stemberger, J. P. (1991) Apparent anti-frequency effects in language production: the addition bias and phonological underspecification, *Journal of Memory and Language*, **20,** 161–185.
Strevens, 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. (1999) Speech perception. In *Encyclopedia of electrical and electronics engineering* (J. Webster, editor), New York: John Wiley & Sons. pp. 175–195.