

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 21 (1996-1997)
Indiana University

A Preliminary Acoustic Study of Errors in Speech Production¹

Stefan Frisch and Richard Wright

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

¹ This work supported by NIH-NIDCD Training Grant DC00012 to Indiana University. This paper was presented as a poster at the 133rd meeting of the Acoustical Society of America, June 1997.

A Preliminary Acoustic Study of Errors in Speech Production

Abstract. Phonological speech errors provide important psycholinguistic evidence for the representations of phonological theory. In an electromyographic (EMG) study of experimentally induced phonological speech errors, Mowrey and MacKay (1990) found that speech errors frequently occur at a sub-lexical, gestural level, with no apparent effect on the percept of the word. Based on these gradient errors, they argue against speech errors as evidence for the segmental unit. Mowrey and MacKay's study considered the activity of a single muscle, and thus was unable to determine whether single gestures acted independently of gestural constellations, which may be equivalent to traditional segmental units. This study is a preliminary report from an ongoing acoustic analysis of speech errors. The data are tape recordings of an error inducing experiment using nonsense tongue twisters. Recordings of a single speaker producing four different tongue twisters targeting /s/ and /z/, e.g. *sit zap zoo sip*, were digitized and analyzed. Some errors involved multiple changes in acoustic properties, including simultaneous changes in periodicity, amplitude of friction, and duration, while others involved a subset of these properties. This evidence suggests that errors can occur at both the single gesture level, affecting non-contrastive acoustic properties, and at the level of the gestural complex or segment, creating a perceptible, linguistically contrastive change.

Introduction

This study is an investigation of sub-lexical phonological speech errors using acoustic-phonetic measures. By using acoustic methods, we intended to evaluate evidence for phonological segments based on speech errors, and to uncover new characteristics of the organization of the speech production mechanism which would not be discovered using traditional transcriptional evidence.

Traditionally, speech error data is collected and analyzed using only phonetic transcriptions. Speech errors are collected either opportunistically, in 'natural error corpora,' or experimentally, from speech error inducing procedures such as the SLIPS priming technique (Baars, Motley, & MacKay, 1975) or tongue twisters (e.g., Shattuck-Hufnagel, 1992). Based on traditional data collection, researchers have claimed that most errors occur at the level of the phoneme or feature (Wickelgren, 1965, Fromkin, 1971). In addition, one 'law' of speech errors is that erroneous utterances are phonotactically grammatical (Wells, 1951; Fromkin, 1971).

Mowrey and MacKay (1990) analyzed electromyographic (EMG) recordings of tongue twisters to evaluate these two claims. They found that EMG activity during tongue twisters showed gradient speech errors. There was a range of muscle activation from none to that equivalent to a normal production for an intruding segment. For example, in tongue twisters such as *Bob flew by Bligh bay*, they found gradient activation of the lingual transversus-verticalis complex, used in the lingual gesture of [l], in the productions of *bay*. Muscle activation was found in productions which were perceptually normal, indicating that a gradient error may occur which is auditorily undetectable.

They conclude that gradient errors which would not be detected by traditional error collection techniques often occur (see also Laver, 1979; Boucher, 1994). Errors which occur on a continuum of muscle activation undermine arguments that the majority of speech errors occur at the phonological level of

the phoneme or feature. In addition, such errors violate the law that speech errors obey phonotactic grammaticality under any non-trivial interpretation of this generalization.

Mowrey and MacKay (1990) studied single muscle fiber activation, and did not examine the acoustic properties of their anomalous utterances. Thus, it is impossible to determine whether the gradient activation they found occurred in all fibers of the muscles involved in a linguistically significant gesture, or whether some higher level monitoring in the production component utilized agonistic or antagonist fibers to insure an auditorily normal outcome in cases of gradient errors. We propose that an acoustic analysis of speech errors which considers several dimensions upon which a linguistic contrast is based can reveal the full range of variation in the speech error data. Our finding is that errors may occur on a single acoustic dimension as the result of a single gestural error, or errors may have simultaneous changes on several independent dimensions, involving an entire constellation of gestures.

Methods

Data

Recordings from a speech error experiment (Frisch, 1996) were analyzed acoustically and compared to the transcriptions of the experimental session which were used to score productions as errors. The original experiment had 88 tongue twisters involving a variety of target consonants, all of which were onsets of monosyllables. The data we analyzed acoustically consisted of four tongue twisters targeting [s] and [z], each repeated six times. The twisters, in the order they were presented in the experiment, are given in (1).

- (1) sit zap zoo sip
 sung zone Zeus seem
 zit sap sue zip
 zig suck sank zilch

We chose tongue twisters involving [s] and [z] as they generated many errors in the experiment and were difficult to transcribe.

We initially conducted a qualitative analysis of 6 participants of the original 21 participants in the experiment (the first six participants). The results we present here are detailed measurements for one representative participant (Participant 2). To date, we have made measurements of two other participants (Participants 1 and 3) and the overall patterns observed are analogous to those for participant two. All measurements were made from waveforms, with accompanying spectrograms for reference. Table 1 shows the general acoustic characteristics of [s] and [z] we analyzed.

Table 1**Acoustic characteristics of [s] and [z]**

Characteristic:	[s]	[z]
Duration (Klatt 1976)	Long	Short
Periodicity	Aperiodic	Periodic
Frication amplitude (Stevens 1960, Pickett 1980)	Greater	Lesser
Vowel onset	Sharp	Gradual

Measurements

Four measurements intended to capture the major differences between [s] and [z] were made:

1. DURATION - the duration of the fricative noise, including overlap with the preceding or following vowels.
2. %VOICING - the fraction of the total duration which contained voicing.
3. WINDOW AMPLITUDE - the RMS amplitude of fricative noise of a 50ms window surrounding the amplitude peak of the fricative noise. The signal was high-pass filtered at 2kHz to remove the energy contributed by the periodic signal.
4. VOWEL RISE TIME - the time from the end of the fricative to the first vowel amplitude plateau.

These measurements are demonstrated for [s] and [z] in Figure 1.

 Insert Figure 1 about here

Results**Gross Characteristics of the Data**

Overall, the four measurements differentiate [s] and [z] for Participant 2. Means for Participant 2's productions for all measurements, with error bars showing standard error, are given in Figure 2.

 Insert Figure 2 about here

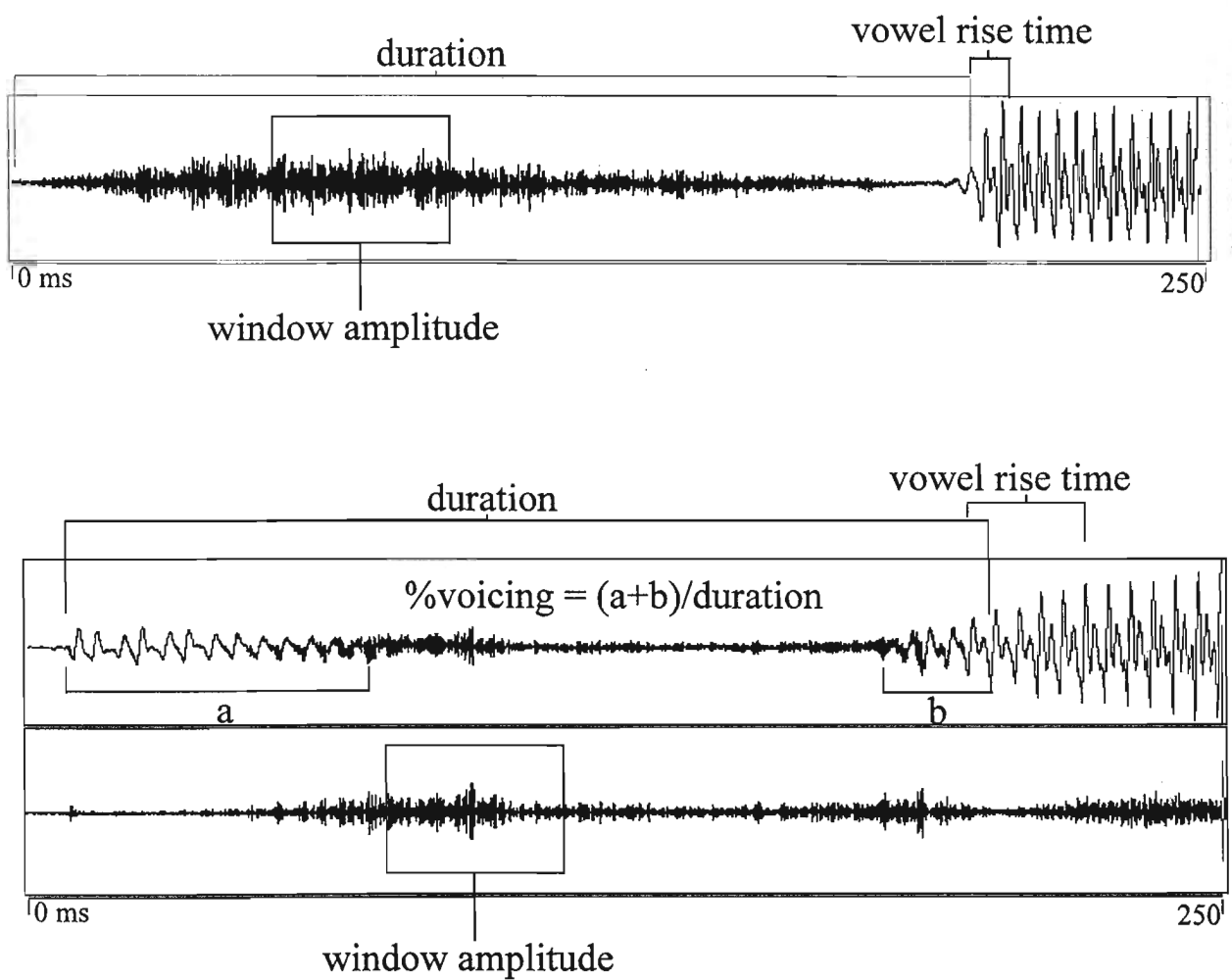


Figure 1. Sample measurements of [s] in *sip* (top) and [z] in *zilch* (bottom).

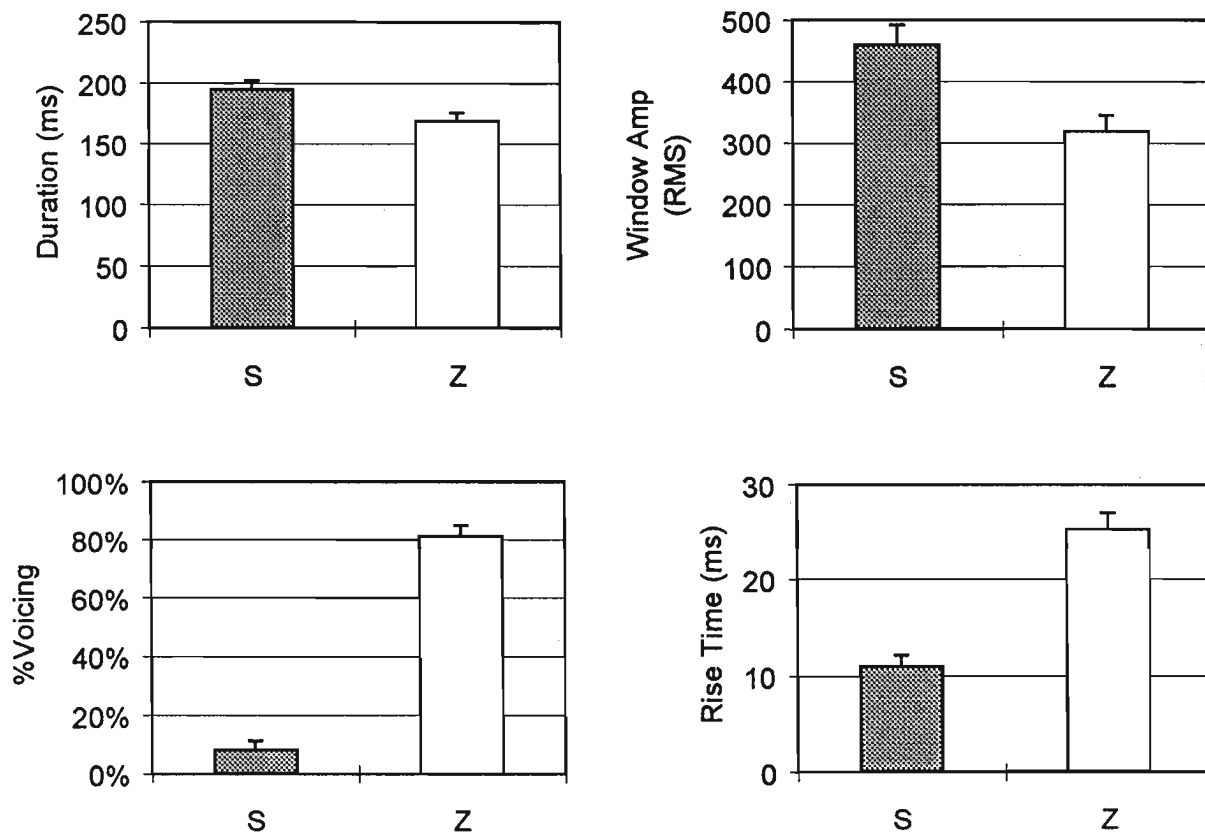


Figure 2. Mean and standard error for DURATION, WINDOW AMPLITUDE, %VOICING, and VOWEL RISE TIME for all productions.

Error Data

Categorical Errors

Two intended productions of [s] were [z]-like on every measure, and therefore appear to be categorical switches from [s] to [z]. Both of these productions were scored as errors in Frisch (1996). Acoustic characteristics for the categorical errors and the productions deemed to be completely normal are shown in Figure 3. The symbol indicates the intended production, [s] or [z]. The errors are marked by boxes around their symbols.

 Insert Figure 3 about here

Voicing Errors

Several productions of [s] and [z] had abnormal %VOICING and/or VOWEL RISE TIME. However, they have normal WINDOW AMPLITUDE and DURATION. These errors are shown in Figure 4. Intended [s] which were partially voiced, but transcribed as [s], are marked by boxes around their symbols. These productions contained voicing at the onset of the fricative, but did not overlap with the following vowel. The intended [s] which were mostly voiced and were transcribed as [z] in the coding of the speech error experiment are marked by diamonds around their symbols. These productions contained fricative noise overlapping with the following vowel.

 Insert Figure 4 about here

There was one intended [z] which was almost completely devoiced, indicated by the arrow. This production was, however, transcribed as [z]. It is entirely voiceless except at its very end where there is fricative noise overlapping with the following vowel onset. Intended [z] which were mostly devoiced are marked by circles around their symbols. These productions were transcribed as [z] with some difficulty in the original error experiment, and judged by the authors to be the most ambiguous tokens produced by this participant. They were voiced primarily in their beginning portions. One of these productions was corrected by the speaker, suggesting that the speaker thought the production was anomalous. This token was transcribed as [z] in the experimental coding, however, and thus not scored as an error.

Amplitude and Duration Errors

Several productions of [s] and [z] had abnormal WINDOW AMPLITUDE and/or DURATION, but normal %VOICING and VOWEL RISE TIME. Figure 5 shows amplitude and duration errors. Intended [s] with [z]-like duration and amplitude are marked by boxes around their symbols. One of these productions was transcribed as [z], it had higher vowel rise time than most other intended [s] productions (19.9ms).

 Insert Figure 5 about here

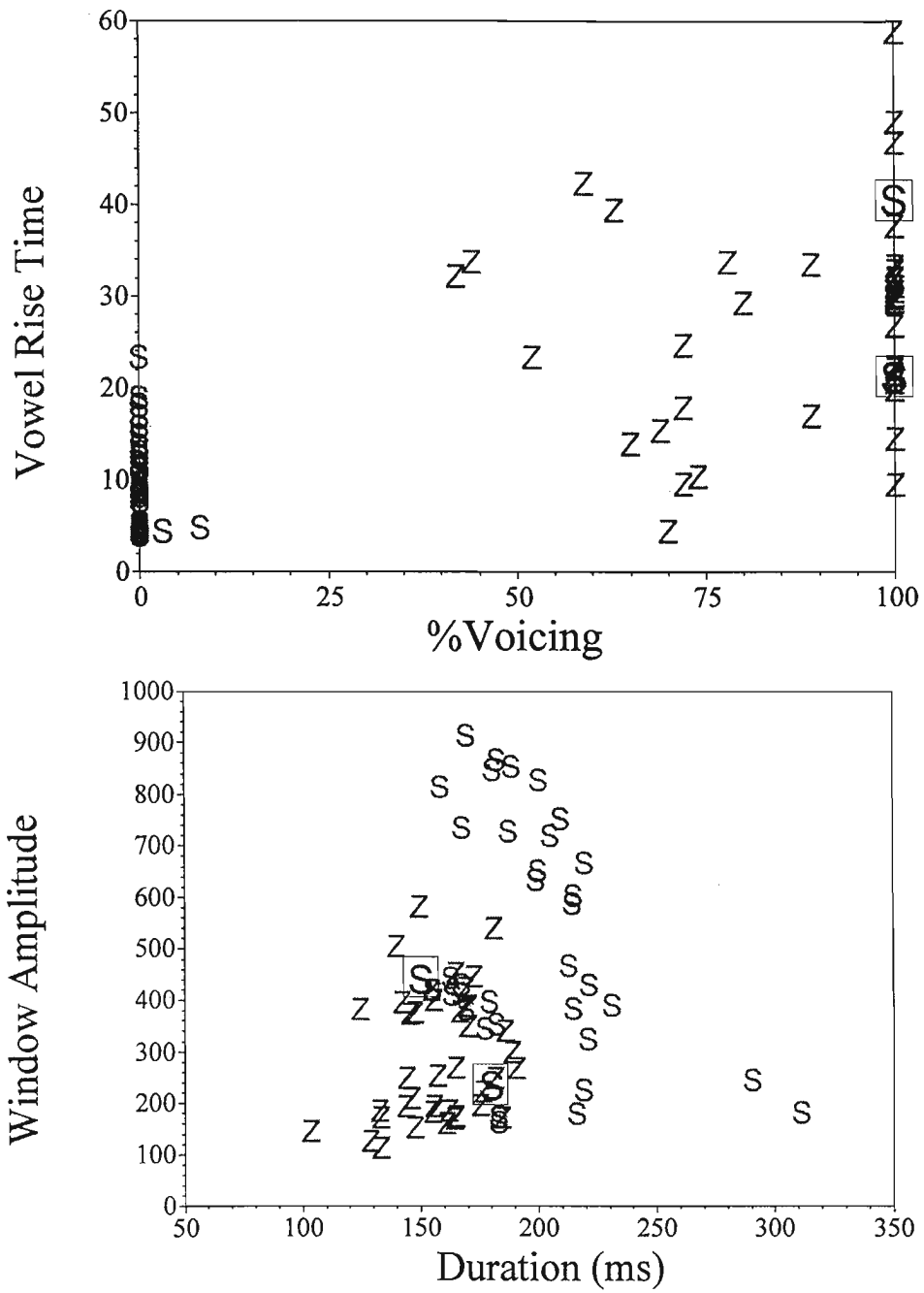


Figure 3. Categorical Errors.

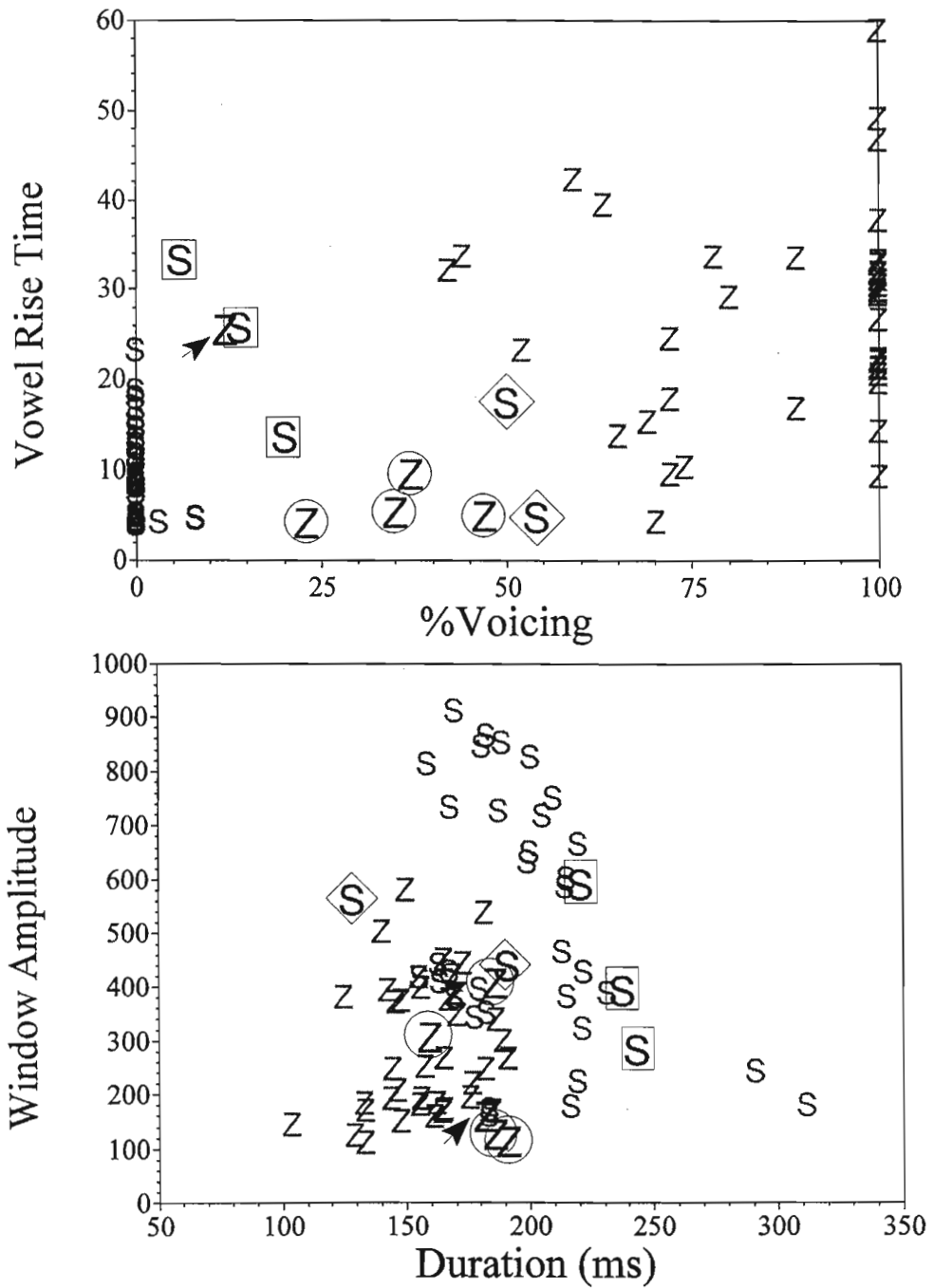
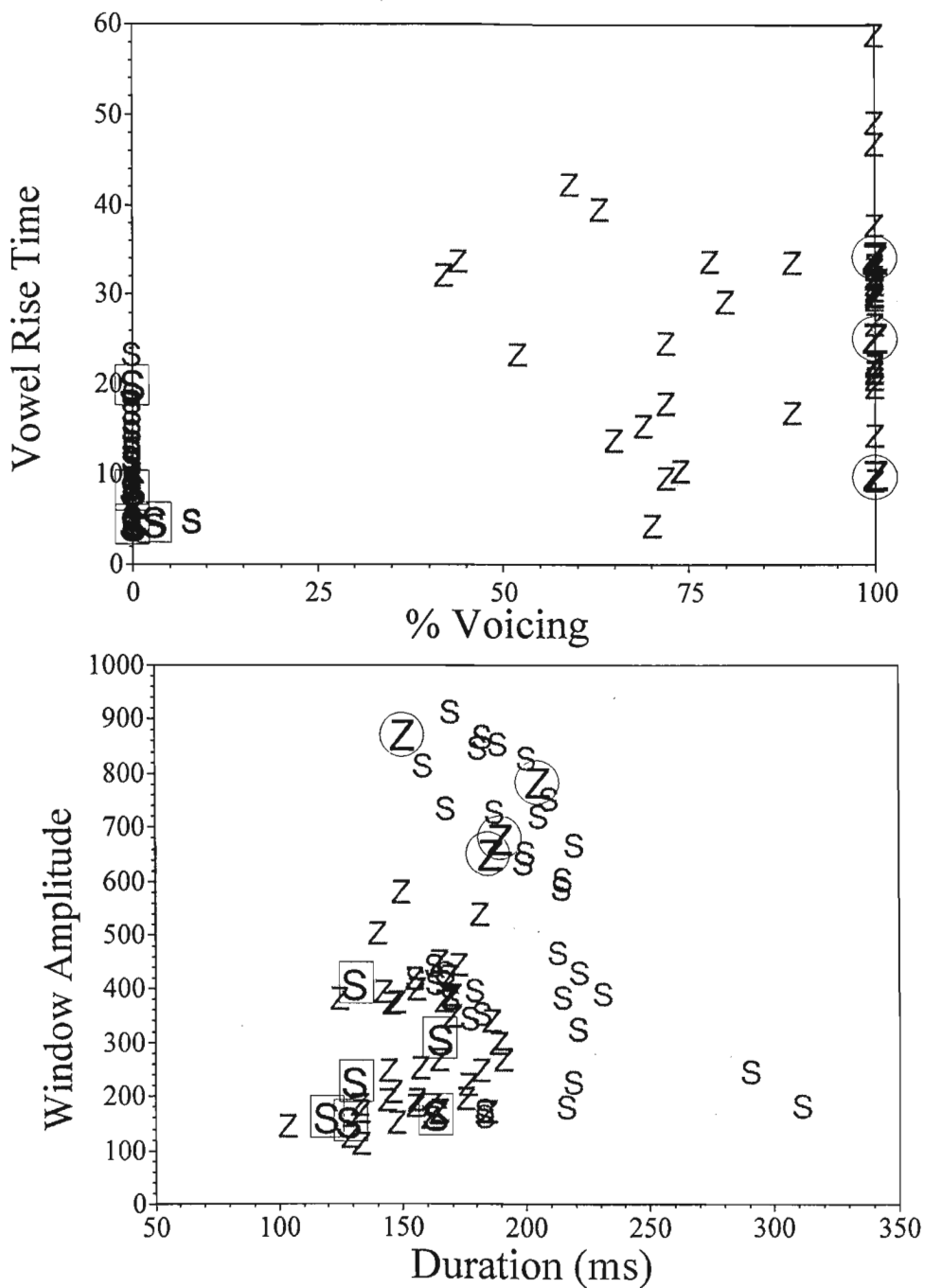


Figure 4. Voicing and Rise Time Errors.



The intended [z] with [s]-like amplitude are marked by circles around their symbols. These productions were all fully voiced, and were transcribed as [z] in the original experiment. Qualitatively, these sound to the authors like excellent examples of [z], regardless of their [s]-like duration. Examples such as these make it clear that some of the different acoustic cues for [s] and [z] are not perceptually equivalent.

Concatenated Errors

Two productions are extreme cases of errors which were corrected without hesitation, repetition, or cessation of the fricative noise. Waveforms for these productions are shown in Figure 6. As can be seen in the Figure, they have extremely long duration, appropriate for two independent productions. In both cases, the participant began with the incorrect production ([s] in the first case and [z] in the second) and ended with the correct production. In both cases, during the transition from the beginning to the end there was a period of reduced frication noise appropriate for [z], but with an absence of voicing. This is indicated in the figure as [z̄].

 Insert Figure 6 about here

Conclusion

Acoustic analysis of the speech errors in this study reveals acoustically categorical errors, presumably consistent with an error in a representational unit at the level of the segment or gestural constellation (Browman & Goldstein, 1986). However, a number of sub-featural errors were found. These errors may be the result of a single gestural error in line with the findings of Mowrey and MacKay (1990). Additionally, we have shown that sub-featural errors may or may not be auditorily contrastive, even if they are acoustically erroneous.

Instrumental investigation of speech errors shows that sub-lexical errors lie on a continuum between categorical and sub-featural errors. This pattern of errors can be revealed only by instrumental methods, as the errors are often not phonologically contrastive, and may not even be auditorily detectable. Individual gestures (as reflected in their acoustic consequences) do not obey constraints on phonotactic grammaticality (Mowrey & MacKay, 1990). This fact is most clearly apparent in the concatenation errors, which effectively produced [sz] and [zs] clusters word initially, a pattern not found in English. These violations were potentially sub-featural errors, however, so it may be the case that phonotactic regularity is upheld at the level of the segment. In other words, it may be the case that categorical segment errors do result in phonotactically regular words.

An analogous situation has been observed at other levels of language processing. Garrett (1975) found increased violations of the syntactic class constraint for units smaller than the word. Thus, the syntactic class constraint (e.g., nouns replace nouns) is upheld for words, but sometimes violated for morphemes, and frequently violated for segments. Dell and Gupta (1997) found the syllable position constraint (Boomer & Laver, 1968) is gradually violated depending on the domain of the syllable position generalization. In general, for example, syllable onsets interact with onsets, obeying their syllabic affiliation. Dell & Gupta found that this law is sometimes violated within the scope of an utterance, but if there is a distributional constraint within the language (e.g., /h/ only appears as an onset, /N/ only appears

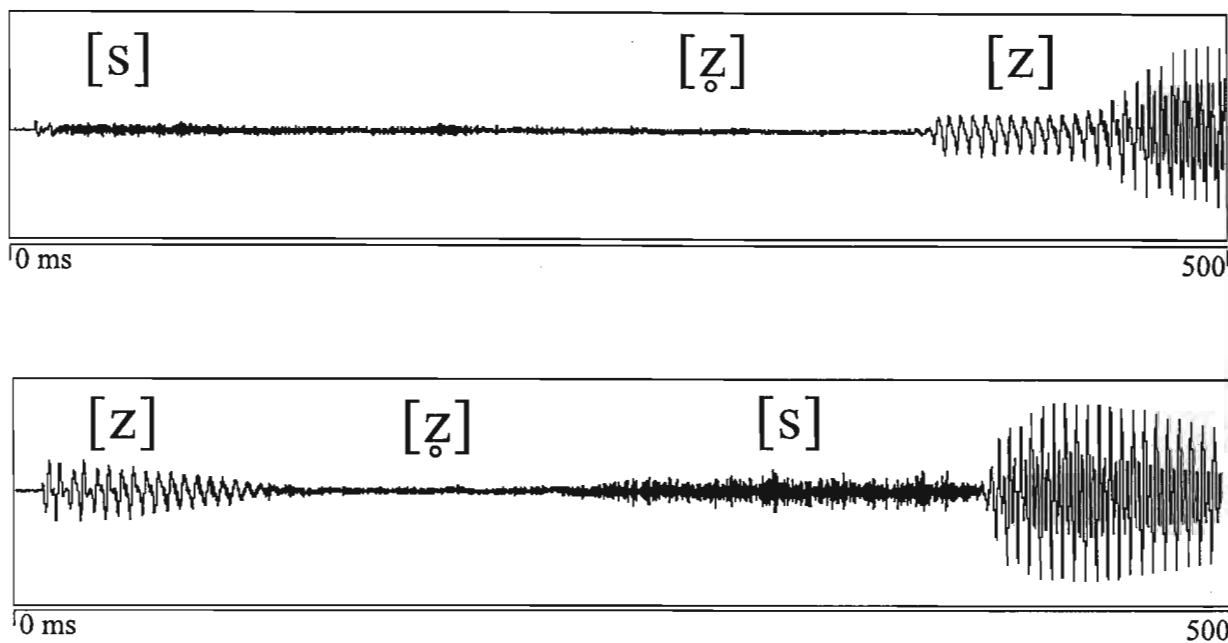


Figure 6. Concatenation Errors: [sz] with [z] intended (top) and [zs] with [s] intended (bottom).

as a coda in English) the syllable position constraint is never violated by those segments. Interestingly, they found that if the experiment is designed so that there is a distributional constraint within the experimental stimuli which is not found in the ambient language, that constraint is obeyed, and fewer violations of the syllable position constraint are found which also violate the distributional constraint within the experiment. Thus, they demonstrated gradient degrees of violation of the syllable position constraint, depending on the generality of the constraint in the participant's linguistic experience.

Future Work

Our current research plan is to combine the acoustic measurements presented in this study with analogous measurements for five other participants. Our long term goal is to find quantitative, rather than qualitative, evidence for or against a segmental level of organization in speech production. With the combined data, we plan to examine the statistical distribution of productions on each of the four dimensions presented here for [s] and [z]. If speech errors are solely the result of individual muscle mis-articulations (as proposed by Mowrey & MacKay) then we would expect to find distributions with a single mode for each dimension for each speaker for each consonant. Categorical errors would be instances where, by random variation, values of all dimensions co-occur in a single production which are appropriate for the other consonant. If, on the other hand, there is a segmental unit of organization which coordinates several gestures, then we expect to find a disproportionate number of categorical errors where all dimensions take extreme values simultaneously. So extreme values across dimensions will be correlated, and the distributions on each dimension for each speaker for each consonant will be bimodal. Since there is a great deal of variation over a relatively small sample of tokens for each speaker, we are also investigating statistical methods for combining data across speakers. Some patterns may not be reliably identified in every speaker, in which case group data may provide insight that an individual speaker analysis, such as the one presented here, might miss.

We also plan to investigate the perceptual side of speech errors, using the corpus of errors studied here in a series of playback experiments. Our measurements provide a quantitative scale on which errors can occur to different degrees. The degree to which errors of different degrees along different dimensions can be detected by naive listeners will be tested using tokens from this corpus. We are interested both in the reliability of error detection across listeners as well as individual differences in perceptual boundaries between listeners. These results bear directly on the reliability of speech error data collected in the laboratory or opportunistically.

References

- Boomer, D. & Laver, J. (1968). Slips of the tongue. *British Journal of Disorders of Communications*, 3, 1-12.
- 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. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219-252.
- Dell, G. & Gupta, P. (1997). Producing and representing serial order. Paper presented at the Carnegie Symposium on Emergentist Approaches to Language, May 1997, Pittsburgh, PA.

- Frisch, S. (1996). *Similarity and frequency in phonology*. Unpublished Ph.D. Dissertation, Northwestern University, Evanston, IL.
- Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47(1), 27- 52.
- Garrett, M. (1975). The analysis of sentence production. In G. Bower (ed.), *The Psychology of Learning and Motivation* (pp. 133-177). New York: Academic Press.
- 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. (1979). Slips of the tongue as neuromuscular evidence for a model of speech production. In H. Dechert & M. Raupach (eds.), *Temporal variables in speech* (pp. 21-26). The Hague: Mouton.
- Mowrey, R. & MacKay, I. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, 88(3), 1299-1312.
- Pickett, J. (1980). *The sounds of speech communication*. Baltimore, MD: University Park Press.
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42: 213-259.
- Stevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, 3, 32-49.
- 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.