PHONOLOGICAL PLANNING FOR SPEECH PRODUCTION SPEECH ERROR EVIDENCE FOR WORD-BASED VS. SYLLABLE-BASED STRUCTURE

Stefanie Shattuck-Hufnagel

Speech Communications Laboratory, Research Laboratory of Electronics
Massachusetts Institute of Technology, Cambridge, Mass. USA 02139

ABSTRACT

Single-segment speech errors are more likely to occur between two word-onset consonants than between a word-onset consonant and one in word-medial position. Errors elicited from speakers of American English using tongue twister stimuli show that this onset-position similarity predominates even when the two consonants are followed by vowel nuclei with different degrees of lexical prominence or stress. This onset similarity constraint holds for tongue twisters in both word-list and phrasal form, suggesting that, for this language, word and morpheme structure forms a part of the representation that is in force at the point in production planning when segmental interaction errors occur.

INTRODUCTION

Phonological planning for speech production presumably involves the representation of organizational units larger than the individual segment. Candidates include elements which have proven useful in the statement of phonological, phonetic and metrical regularities, such as the morpheme, the word, the syntactic phrase, the syllable, the foot and the prosodic phrase, etc. Phonological errors in speech production can shed some light on the question of which of these larger units play a part in the planning process, because of the following characteristic: when two segments of an utterance interact in an error like an exchange, they tend to come from corresponding structural positions. For example, an initial consonant tends to interact with another initial consonant, rather than with a final consonant, giving e.g. "bate of dirth" for "date of birth", but not "thate of bird" [1]. Similarly, final consonants tend to interact with each other and not with initial consonants, giving "noth wort knowing" for "not worth knowing", but not "thot wern knowing."

By examining the nature of this position-similarity constraint on segmental interaction errors, we can begin to determine the identity of the larger elements that play a role in segmental processing. In this way, we can distinguish between proposed models of the cognitive process of production planning that rely on different representational elements. For example, one type of model for segmental processing might postulate that words and morphemes are part of the phonological planning frame, while a second type of model might hypothesize that by the time segmental errors occur, the word boundaries have been erased, leaving an organizational framework that consists only of prosodic structures like the syllable, foot, etc. A model of Type 1 provides for the possibility that two target segments that interact in an error tend to be from similar positions in their respective words, while a model of Type 2 cannot easily account for such a word-based position-similarity constraint. Instead, Type 2 models might predict a constraint based on position in the syllable or in the foot, or perhaps on syllable prominence.

What does the pattern of segmental speech errors suggest with respect to this issue? Segmental interaction errors collected from spontaneous American English speech [2] do not resolve the question clearly , because so many of the errors occur in monosyllabic words, or in words with lexical stress on the first syllable, so that both word-based and prosody-based models predict the same result [3]. We need to know the error distribution for consonants in pairs ofwords that do not both have lexical stress on the first syllable. By comparing the number of /p/-/f/ interaction errors in sequences like (a) "parade fad" and (b) "repeat fad", we can determine whether more errors occur (a) when the two target segments share position in their word onsets (but not before the stressed vowel), or (b) when the two target sements share position before the stressed vowel (but not in their word onsets.)

Shattuck-Hufnagel [4] has described an elicitation experiment that used tongue twisters of these two types. For both a reading and a recall task, results show substantially more interaction errors between two word-onset consonants than between one word-onset and one word-medial consonant, i.e. more /p/-/f/ errors for stimuli like "parade fad foot parole" than for stimuli like "repeat fad foot repair." This pattern supports Type 1 models, in which words and morphemes are still part of the planning representation at the point where segmental errors occur.

THE EXPERIMENT

One question that might be asked about the generalizability of the tongue-twister results to models of spontaneous speech planning concerns the list-like nature of the stimuli. Since each stimulus consisted of four lexical items presented visually, with spaces between them, it might be argued that the task somehow emphasized the first letter of each item and that this special emphasis turned the corresponding four word-onset segments of each twister list into a mutually-confusable set. The solution to this problem is to embed the four words of each twister in a larger string of words, so that the set of post-space segments does not isolate the four targets as a particularly-confusable set. In the experiment to be reported here, this embedding was accomplished by positioning each of the original twister words in a short phrase, so that "parade fad foot parole" became "The parade was a fad and the foot got parole," etc.

The decision to use four short locally well-formed phrases as the embedding material was motivated by the results of a different experiment, in which the effect of list-like vs. phrase-like stimulus structure was investigated for the initial and final consonants of CVC words [5]. In that elicitation experiment, changing from lists to phrases had a substantially larger effect on final consonants than on initial consonants. When word-list twisters like "peal tone pan tool" and "leap note nap lute" were turned into phrasal twisters like "From the leap of the note to the nap of the lute", initial consonant interaction errors declined only 16% but final consonant errors declined 78%. The resulting low rare of final-consonant errors elicited by phrasal twisters more closely resembles the pattern found in collections of spontaneous speech errors, suggesting that phrasal twisters may be more appropriate than word-list twisters for invoking the representations used in spontaneous speech planning. This provides further motivation for re-testing the word-onset similarity constraint in a phrasal context.

Stimuli

The original word-list experiment employed 24 sets of tongue twisters, each set constructed around a pair of consonantal segments which had a high error rate in a corpus collected from spontaneous speech [6]. For the present experiment, the words of these twisters were embedded in locally-grammatical phrases. Each set contained four twisters, developed according to the principles illustrated here for the target pair /p/-/f/.

Condition 1: Share Word Onset Position
Both members of target pair occur in word onset; only one occurs in position before stressed vowel
e.g. The parade was a fad
and the foot got parole.

Condition 2: Share Stress Position
Both members of target pair occur before stressed vowel; only one occurs in word onset
e.g. To repeat was a fad
and the foot could repair.

Condition 3: Share Both Positions
Both members of target pair occur in word onset and both occur before stressed vowel
e.g. To his peril the fad
was the foot of a parrot.

Condition 4: Share Neither Position
Only one member of target pair occurs in word onset and before stressed vowel
e.g. For a ripple a fad
has a foot and is rapid.

The twenty-four sets of twisters were built around 12 different pairs of target consonants: /f-p, r-l, b-g, l-y, b-p, m-n, r-w, d-g, p-k, l-n, d-t, j-d/. In addition, each twister contained a third "filler" consonant, like /r/ in the set above, which was phonologically dissimilar.

Speakers

Twenty MIT undergraduates participated as speakers; thirteen were male and seven were female. All were native speakers of American English with no known speech or hearing deficits; all were right-handed. Speakers were paid a nominal amount for their participation.

Presentation

The 24 sets of stimuli were divided into two groups of twelve sets each. Speakers were divided into two groups of 10 each. One group of speakers produced half the stimuli: Conditions 1 and 3 from the first 12 sets, and Conditions 2 and 4 from the remaining 12 sets. The second group of speakers produced the other half of the stimuli: Conditions 2 and 4 from the first 12 sets, and Conditions 1 and 3 from the remainder. Thus every speaker produced two of the stimuli from each of the 24 sets, or 48 stimuli in all. The 48 stimuli were presented in two sessions of 24 stimuli each; the two half-hour sessions were counterbalanced for order of presentation.

The tongue twisters were typed on 3 x 5 cards and presented visually to the speaker, who read each one aloud three times and then turned the card over and recited it three times from memory. After each card the speaker generated a sentence as part of another experiment, so that the twisters were separated from each other by a 15-60 second period and by a different kind of activity. Speech was tape-recorded for later transcription.

Scoring

Utterances were transcribed by ear with repeated listening, and scored for errors. Only segmental interaction errors between the two members of the target segment pair were scored, which means that the following error types were not included: substitution errors with no apparent source in the utterance, omission and addition errors, and errors that involved units larger than a single segment, such as a CV or VC. In addition, interactions between one of the two target segments and the third "filler" segment were scored separately. Results for the reading and recitation tasks were also scored separately.

Results

Like the findings for word-list twisters described earlier, the results for phrasal twisters show a stronger tendency for word-onset consonants to interact with each other than for word-onset consonants to interact with medial consonants, even when the onset-medial pairs shared pre-stress position. These results can be summarized as follows.

Condition	Number of Errors		
	Reading	Recall	Total
1) Share Word-Onset	49	85	134
2) Share Prestress	17	42	59
3) Share Both	58	144	202
4) Share Neither	5	10	15

Three aspects of the data in particular illustrate the predominance of onset-onset errors:

1. **Conditions in which the target consonants both occurred in the word onset provoked many more errors than conditions where one of the two target consonants appeared elsewhere in the word.** The two onset-onset conditions (1 and 3) elicited a total of 336 interactions, and the two onset-nononset conditions (2 and 4) only 74.
2. **The two conditions which directly pit word-onset similarity against pre-stress similarity show that shared word onset induces twice as many errors as shared prestress position.** That is, Condition 1 stimuli elicited 134 errors while Condition 2 stimuli elicited only 59. In Condition 1, the two target segments share word onset position but <u>differ</u> in the stress of the following vowel, while in Condition 2, the two targets share prestress position but <u>differ</u> in word position, providing a direct test of the strength of these two position similarity constraints. Under these circumstances, word onset similarity clearly can overpower dissimilarity in stress more often than stress similarity can overpower dissimilarity in onset position.
3. **Unexpected interactions between a member of the target consonant pair and the phonologically less-similar third "filler" consonant also favor interactions between word onsets.** The third "filler" consonant in the bisyllabic words of each twister was less similar to the target consonant in the monosyllabic word than the two targets were to each other, as measured by the number of shared features. On a simple categorization as same or different in Voicing, Manner or Place, the mean difference between the two members of the target pairs was 1.2 features out of 3, while the mean for the target-filler pairs was

2.3 features. Despite this feature dissimilarity, speakers produced 32 unexpected interactions between onset-onset pairs (like /r/-/f/ in "To repeat was a fad that the foot could repair"), and only 3 for onset-nononset pairs (like /r/-/f/ in "The parade was a fad and the foot got parole".) Although the numbers are much smaller, the unexpected errors clearly support the claim that shared word-onset status is more conducive to interactions than is shared prestress position.

DISCUSSION

The distribution of errors in the elicitation experiments described above suggests that the predominance of word-onset errors in corpora gathered from spontaneous speech is not the result of some unnoticed factor that is coincident with word-onset position, but rather can be interpreted as evidence that lexicomorphemic structure is part of the processing representation that speakers make use of as they plan utterances for production. This claim alone would certainly not be surprising; the new and further information here is concentrated in two points:

1) Word structure is part of the processing representation at the point when segmental errors occur. This observation rules out a class of models in which the segmental processing mechanism that is susceptible to serial ordering errors operates on a representation in which morphemic structure has been erased, and supports the class of models in which that structure is preserved. There are several ways in which this claim for the preservation of morphemic structure could be realized in a particular model, including (a) as part of the organizational framework of serially-ordered slots to be associated with target segments, or (b) as part of the lookup mechanism for retrieving the phonological forms of morphemes from long-term storage. For futher discussion of this issue, see Shattuck-Hufnagel [7].

2) The second summary point is that prosodic structure also plays a role in this processing representation. Despite the fact that shared word-onset position can overwhelm both stress dissimilarity and feature dissimilarity to provoke an interaction error, the experimental results provide evidence that prosodic factors are also at work. To see this, compare the results for Condition 2 with Condition 4. Suppose that lexical stress similarity and syllable position affiliation similarity played no role in conditioning errors. Then we might expect a similar rate of /p/-/f/ errors for Condition 2 and Condition 4. That is, Condition 2 stimuli like "To repeat was a fad that the foot could repair" (where /p/ and /f/ are both pre-stressed and syllable-initial), and Condition 4 stimuli like "For a ripple a fad has a foot and is rapid" (where only /f/ is prestressed and the syllable position affiliation of /p/ is unclear), could be expected to provoke about the same number of errors if syllable affiliation and lexical prominence play no role in constraining the error process. Instead, Condition 2 elicits substantially more errors than Condition 4. This suggests that both lexicomorphemic structure and prosodic structure play a role in the phonological planning representation at the point in that process where segmental errors occur. Further elicitation experiments are in progress to determine in more detail the nature of the prosodic component of this representation.

REFERENCES

[1] V.A. Fromkin, Speech Errors as Linguistic Evidence, The Hague: Mouton, 1973.
[2] M.F. Garrett, The analysis of sentence production, in G. Bower (Ed.), The Psychology of Learning and Motivation, Vol. 9, New York: Academic Press, 1975.
[3] S. R. Shattuck-Hufnagel, Sublexical units and suprasegmental structure in speech production planning, in P. MacNeilage (Ed.),The Production of Speech, New York: Springer-Verlag, 1983.
[4] S. R. Shattuck-Hufnagel, Context similarity constraints on segmental speech errors, in J.L. Lauter (Ed.), Proceedings of the Conference on the Planning and Production of Speech in Normal and Hearing-Impaired Individuals, American Speech-Language-Hearing Association ASHA Report No. 15, 1985.
[5] S. R. Shattuck-Hufnagel, Position of errors in tongue twisters and spontaneous speech, in Speech Group Working Papers, Research Laboratory of Electronics, MIT, Vol. 1, 1982.
[6] S. Shattuck-Hufnagel and D.H. Klatt, The limited use of distinctive features and markedness in speech production. Journal of Verbal Learning and Verbal Behavior 18, 41-55, 1979.
[7] S. Shattuck-Hufnagel, The role of word-onset consonants in speech production planning, in E. Keller and M. Gopnik (Eds.), Motor and Sensory Processes of Language, Hillsdale, N.J.: Lawrence Erlbaum Associates, in press.