Perception & Psychophysics
1983, 34 (5), 470-474

Voicing, vowel, and stress mispronunciations in continuous speech

Z. S. BOND and LARRY H. SMALL
Ohio University, Athens, Ohio

The purpose of this study was to investigate the perceptual effects of three types of mispronunciations, affecting the voicing of obstruents, the front-back dimension of stressed vowels, and the stress pattern of words. Subjects were instructed to shadow prose passages containing mispronunciations. Words containing voicing mispronunciations typically were repeated in their original form; words with vowel and stress pattern mispronunciation led instead to other response types.

Recent studies of the perception of fluent speech suggest that listeners use both phonetic ("bottom up") and contextual ("top down") information in word recognition. Much of these data are derived from an experimental paradigm in which subjects are asked to respond in various ways to speech containing mispronounced words. This research paradigm was first introduced by Bagley (1900) and reintroduced more recently by Cole (1973) (see also Cole & Rudnicky, 1983).

Considerable research has shown that "top down" and "bottom up" information both contribute to word recognition. For example, Marslen-Wilson and Welsh (1978) found that subjects shadowing prose tended to correct mispronunciations without hesitation, particularly if the mispronunciations were relatively minor (changes of a single distinctive feature), if the mispronounced words were highly predictable from context, and if the mispronunciation occurred late rather than early in a word. Cole, Jakimik, and Cooper (1980) found that reaction time in detecting mispronunciations was influenced by the contextually suggested segmentation of a target word. For example, mispronunciations of "drift" were detected more quickly in "snowdrift," where they occurred in the second syllable of a single word, than in "snow drift," where they occurred in the second of two monosyllabic words. In addition, Cole and Jakimik (1978) found that mispronunciation detection was faster when target words were predictable from context—either from words explicitly present or from the general theme of a passage.

The relative usefulness to a listener, or the salience, of various types of phonetic information has received somewhat less attention, even though researchers have hypothesized that not all phonetic information is equally useful in understanding fluent speech. On the basis of an extensive series of experiments examining the detectability of voicing, place, and manner mispronunciations (Cole, Jakimik, & Cooper, 1978), Cole and Jakimik (1978) have suggested that word recognition involves "perceptual anchors," invariant phonetic features to which listeners attend in the recognition of words. Pisoni (1981) has made a similar suggestion, terming the properties of the stimulus input that can be used to access various sources of knowledge "islands of reliability." Cole and Jakimik (1978) suggest that the consonant portion of stressed CV syllables may serve as one of a set of perceptual anchors. Pisoni (1981) mentions "the presence of stressed syllables, the beginnings and ends of words, and the locations of various spectral changes indicating shifts in the source function" (p. 255).

The purpose of this study was to investigate the salience of different types of phonetic information, using the following working assumption: if a phonetic property is relatively useful, then its distortion would impair word recognition[1] more extensively than distortion of a less useful phonetic property. As a point of departure, we selected quite different phonetic properties of words, the voicing of obstruents, the front-back dimension of stressed vowels, and the stress pattern. Our selection of these three phonetic variables requires some explanation.

Whether the suprasegmental stress is involved in word recognition is not uncontroversial. Fay and Cutler (1977) examined speech errors involving the substitution of one word for another of a different meaning, for example, "equivocal" for "equivalent." The target and the substitute shared a stress pattern in 98% of the errors. From these data, Fay and Cutler suggest that the "mental lexicon" is organized, in part, in terms of the stress pattern of words. Brown and McNeill (1966) also mention the stress pattern as a property of words, on the basis of words recalled by subjects in the "tip of the tongue" state. On the other hand, Garrett (1980) proposes,

A preliminary version of this paper was presented at the spring meeting of the Acoustical Society of America, Chicago, 1982. We want to thank Randall R. Robey for assistance with statistical analysis. The authors' mailing address is: School of Hearing and Speech Sciences, Ohio University, Athens, Ohio 45701.

also on the basis of speech-error data, that a phrasal stress pattern is "calculated" independently of lexical insertion in sentence production; Garrett's view, apparently, is that syllables carry phrasal stress only when they may potentially carry lexical stress. However, Garrett does not specify the details of stress assignment. If Garrett's proposal is true, then stress patterns might not provide information useful for word recognition in fluent speech. Given that the function of lexical stress is far from clear, we decided that the perceptual effects of stress mispronunciations were worth investigating.

The phonetic properties of vowels have not received much investigation in studies dealing with continuous speech. Vowel quality tends to be highly variable (see, for example, Balchak, 1980) and, also, varies across regional dialects of American English more than does consonantal quality. On the other hand, vowels are often more intense and of longer duration than surrounding consonants. In addition, there is some rather fragmentary evidence that stressed vowels may provide reliable information in fluent speech perception. Bond and Garnes (1980) report that very few stressed-vowel errors are found among errors in the perception of fluent conversational speech.

In comparison with the other two variables, changes in the voicing of obstruents have been investigated fairly often. Voicing mispronunciations seem to be readily detectable, although their detectability is influenced by numerous contextual factors.

METHOD

Materials

The test materials consisted of recordings of three prose passages taken from a popular novel, one for each of the three experimental conditions. Each passage contained approximately 600 words, including the 20 two-syllable test words being altered.

All the test words in the three experimental conditions were equated for predictability from context and for frequency of occurrence in English. To determine predictability from context, three groups of 10 subjects each were given written sentences from the prose passages with 80 two-syllable words omitted. The subjects simply filled in the blank with the word they deemed most appropriate. Predictability from context, or contextual constraint, for each word was determined by using a scoring scheme based on Marslen-Wilson and Welsh (1978). If the subject's written response was the omitted word, it was scored "1"; a synonym was scored "2," a related word was scored "3," and an unrelated word was scored "4." The 40 words with the lowest mean ratings, that is, the words most predictable from context, were selected. The grand mean ratings for the 40 words from each of the three conditions were: 1.98 for voicing, 2.04 for vowels, and 2.44 for stress.

Each of the 120 words was then examined for frequency of occurrence in English, utilizing the Kučera and Francis (1967) norms. A one-way analysis of variance was performed to determine whether the frequency of occurrence of the words selected differed among the three conditions. The mean frequencies of occurrence for the 40 words were: 182 for voicing, 204 for vowels, and 200 for stress. The F ratio obtained was not significant.

Twenty words were selected as test words from each 40-word list, depending on the phonetic structure of the words required for each condition.

The three experimental tapes and a 1,000-word practice tape were recorded by a male speaker at the rate of approximately 140 words/min. There were no mispronounced words on the practice tape. On the first experimental tape, the voicing condition, obstruents in 10 of the test words were changed from voiced to voiceless and 10 were changed from voiceless to voiced, for example, "business" to /pɪznəs/ and "kitchen" to /gɪtʃən/. The mispronounced consonants occurred in the initial position of a stressed syllable in 13 of the words and in the initial position of an unstressed syllable in the remaining 7 words. Twelve mispronunciations were word-initial; eight were medial.

On the second experimental tape, the vowel condition, vowels in 10 of the test words were altered from front to back, and the remaining 10 from back to front. The substitute was each vowel's "mirror image" according to the traditional vowel quadrilateral; for example, /i/ was substituted for /u/, /oɑ/ for /eɪ/, and so forth. All mispronunciations occurred in the stressed syllable of the test words, in the first syllable of 15 words, and in the second syllable of the remaining 5 words.

On the third experimental tape, the stress condition, 10 words normally stressed on the first syllable were mispronounced, with stress given on the second syllable. Stress was shifted from the second to the first syllable for the other 10 words. Because there are no clearly defined stressed vowel counterparts of unstressed vowels, the procedure for altering the stress pattern of words cannot be as clearly defined as that creating the other two types of mispronunciations. In arriving at the mispronounced form of the test words, we reduced stressed vowels to /ə, ɪ, or ɚ /. The unstressed vowels that were pronounced with stress received their form primarily on the basis of English spelling. For example, the test word "people" was pronounced /p ə' pʌl/; "decide" became /'disɪd/. The speaker produced all mispronunciations while reading the test passages after considerable practice with the material. We should add that the stress condition may have presented the subjects with more misleading phonetic information than did the other two conditions.[2]

Subjects

Thirty[3] English-speaking students at Ohio University with no history of hearing and speech problems served as subjects. The subjects participated in the study in partial fulfillment of course requirements.

Procedure

All subjects were told that they would be listening to four passages of a story; they were instructed to repeat, as rapidly as possible, everything they heard the speaker say. The instructions stressed repetition rather than paraphrasing the context of the passages. Each subject received the practice passage first. The three experimental passages were presented in random order to all subjects. Subjects were offered a rest period after shadowing the practice passage and after the first experimental passage.

The subjects were tested in an IAC No. 402 double-walled, sound-treated room; they listened to the passages presented at 60 dB SPL on headphones (Grason-Stadler TDH 39). The tapes were played on a Pioneer tape recorder (RT-707) with a Sansui integrated amplifier (A-40). The signal delivered to the headphones was presented binaurally.

One channel of the Pioneer recorder was connected to one channel of a Dokorder (No. 4000) four-channel recorder. The subjects wore a Sony electret condensor microphone (ECM-150) coupled to the second channel of the Dokorder recorder. Each subject's responses, as well as the spoken experimental passages, were recorded simultaneously for later analysis.

RESULTS

Classification of Responses

Subjects' responses to the test words were grouped into three major categories: restorations, defined as

corrections of the mispronunciations; repetitions; and omissions. Restorations were further subdivided into four types: (1) Fluent complete restoration—the subject corrected the mispronounced word to its original form. (2) Fluent partial meaningful restoration—the subject substituted a real word for the mispronunciation, but not the word in the original passage. (3 and 4) The other restoration types were identical to the first two, but were produced with a noticeable hesitation, that is, the responses were no longer fluent. The rationale for counting all four of these response types as restorations is that in all cases subjects arrive at a meaningful lexical item, that is, they recognize a word.

When the subjects repeated the test word as mispronounced, the responses were classified as repetitions, either fluent or hesitant. Also counted as repetitions were meaningless phonetic sequences in place of the mispronounced word. An omission was either an omission of the mispronounced word or a completely unintelligible response. The percentages of response types observed for all three experimental conditions are summarized in Table 1. As shown in the table, only a very small proportion of the responses were hesitant. Consequently, all hesitant responses are grouped with their fluent counterparts in further discussion.

The majority of the responses were either complete restorations or repetitions. The percentage of restorations, repetitions, and other responses for all three experimental conditions are given in Figure 1.

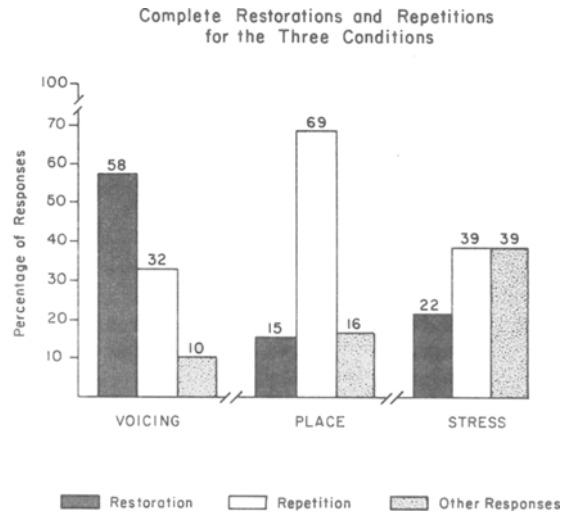

Figure 1. Major response categories to voicing, vowel, and stress mispronunciations.

Restorations

Restorations are the most interesting response type, because they give the clearest indication that subjects have recognized an intended word from the misleading phonetic information presented to them. The percentage of restorations was different when subjects were shadowing in the three conditions. For voicing mispronunciations, 58% of the responses were restorations, counting both fluent and hesitant responses. In the vowel condition, on the other hand, only 15% of the responses were restorations. In the stress condition, 22% of the responses were restorations. These differences were significant beyond the .001 level [min $F'(2,115) = 22.778$] (Clark, 1973).

Table 1
Percent of Response Types Found for the Three Mispronunciation Conditions

Response Type	Voicing	Vowels	Stress
	Restorations		
Fluent Complete	57	13	20
Hesitant Complete	1	2	2
Fluent Partial*	3	6	10
Hesitant Partial*	**	**	1
	Repetitions		
Fluent	31	64	35
Hesitant	1	5	4
Fluent Partial	2	3	11
Hesitant Partial	**	**	1
	Omissions		
	4	8	17

*Meaningful. **Less than 1%.

Examining restorations suggests that the mispronounced segment, within each condition, had little effect on the ease of restoration. Voiced segments pronounced as voiceless were restored 58% of the time; the reverse mispronunciations were restored 59% of the time. Sixteen percent of the mispronounced front vowels and 14% of the back vowels were restored. Words in which stress had been shifted from the first to the second syllable were restored 19% of the time; the reverse stress shift was restored 25% of the time.

In a small number of cases, the mispronunciation of the target word created a possible English word, although these words were semantically highly inappropriate. For example, the word *stable* was pronounced as *staple* in a context dealing with housing animals. Apparently, the subjects were not influenced by the creation of words. The one voicing mispronunciation was restored—pronounced *stable*—by 20 of the 30 subjects, 67% restorations. The four words with vowel mispronunciations which could be interpreted as English words were restored 17% of the time. There were no stress mispronunciations that lead to possible English words.

The location of a mispronunciation within a word may have had some effect on the ease of restoration. Fifty-one percent of the word-initial mispronounced stops and 70% of the medial stops were restored, as were 16% of vowel mispronunciations occurring in the first-syllable and 11% of the second-syllable mis-

Table 2
Percent of Meaningful Restorations in Various Mispronunciation Conditions

	N	Fluent	Hesitant	Total*
	Voicing			
Voiceless to Voiced	10	57	2	59
Voiced to Voiceless	10	55	2	58
Stops	14	47	2	49
Fricatives, Affricate	6	78	1	79
Word Initial	12	48	3	51
Word Medial	8	68	1	70
Stressed Syllable	13	50	3	53
Unstressed Syllable	7	68	1	69
	Vowel			
Front to Back	10	13	2	16
Back to Front	10	12	1	14
First Syllable	15	14	2	16
Second Syllable	5	10	1	11
	Stress			
First to Second Syllable	10	15	4	19
Second to First Syllable	10	24	1	25

**The total is not always the sum of the fluent and hesitant responses because of rounding.*

pronunciations (see Cole & Jakimik, 1980). Cole et al. (1978) report a series of experiments investigating the detectability of mispronunciations of various phonetic targets. Whenever their results can be compared with ours, the findings are quite similar: the targets that Cole et al. found to be easier for subjects to detect, we found more difficult for them to restore. Stop mispronunciations are more difficult to restore than fricative mispronunciations; voiced to voiceless mispronunciations and the reverse are approximately equally easy to restore; and initial syllable mispronunciations are more difficult than medial syllable mispronunciations. These data are presented more fully in Table 2.

DISCUSSION

Although shadowing is not completely congruent with normal speech perception, it does impose two requirements on subjects. In order to comply with their instructions, subjects have to respond fairly quickly to keep pace with the signal presented to them and they have to give a verbal response. Clearly, both repetitions and omissions imply that subjects are either aware of a mispronunciation or at least of something being wrong with a phrase or sentence. However, we are not willing to claim that fluent restorations are possible only when subjects are unaware of a mispronunciation. The speaking rate of the test passages was quite leisurely; it is possible that subjects noticed a mispronunciation but had sufficient time to think of the intended target word and to say it, consciously correcting the mispronunciation. We would argue, however, that a restoration implies that the target word is readily available in spite of its phonetic degradation. Subjects can restore the target words, therefore, only if the words come readily to mind from their mispronounced forms.

Apparently, voicing alterations are not phonetically misleading enough to preclude restorations, since subjects supplied the target words 58% of the time.

For vowel and stress mispronunciations, restorations appear to be more difficult. The mispronunciations provide the subjects with phonetic information that is misleading enough for the target words to be less than readily recoverable. In the vowel condition, the target word is not obvious (15% of the responses are restorations), so the subjects simply repeat the mispronunciation (69% of the responses).

The stress condition presents listeners with mispronounced words that lend themselves neither to easy restoration (22% of the responses) nor to repetition (39% of the responses). Subjects' responses are considerably more varied.

Voicing mispronunciations are clearly less disruptive of easy word recognition than are vowel mispronunciations. Hence, listeners may consider stressed vowels as reliable phonetic information.

The effects of stress mispronunciations, however, are rather less clear. In altering the stress pattern of target words, we necessarily also altered vowel quality; hence, subjects were presented with two phonetic changes. Furthermore, stress mispronunciations may have destroyed the rhythmic patterns of phrases that subjects expected on syntactic or semantic grounds (cf. Cutler, 1976; Martin, 1972). We can only say that a target word with a disrupted stress pattern is difficult to either repeat or restore.

REFERENCES

Bagley, W. C. The apperception of the spoken sentence: A study in the psychology of language. *American Journal of Psychology*, 1900-1901, **12**, 80-130.

Balchak, G. *An analysis of the acoustic structure and intelligibility of conversational speech*. Unpublished master's thesis, Ohio University, 1980.

Bond, Z. S., & Garnes, S. Misperceptions of fluent speech. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J: Erlbaum, 1980.

Brown, R., & McNeill, D. The 'tip of the tongue' phenomenon. *Journal of Verbal Learning and Verbal Behavior*. 1966, **5**, 325-337.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335-359.

Cole, R. A. Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, 1973, **1**, 153-156.

Cole, R. A., & Jakimik, J. Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing*. London: Academic Press, 1978.

Cole, R. A., & Jakimik, J. How are syllables used to recognize words? *Journal of the Acoustical Society of America*, 1980, **67**, 965-970.

Cole, R. A., Jakimik, J., & Cooper, W. E. Perceptibility of

phonetic features in fluent speech. *Journal of the Acoustical Society of America*, 1978, **64**, 44-56.

COLE, R. A., JAKIMIK, J., & COOPER, W. E. Segmenting speech into words. *Journal of the Acoustical Society of America*, 1980, **67**, 1323-1332.

COLE, R. A., & RUDNICKY, A. I. What's new in speech perception? The research and ideas of William Chandler Bagley, 1874-1946, *Psychological Review*, 1983, **90**, 94-101.

CUTLER, A. Phoneme-monitoring reaction times as a function of preceding intonation contour. *Perception & Psychophysics*, 1976, **20**, 55-60.

FAY, D. A., & CUTLER, A. Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, 1977, **8**, 505-520.

GARRETT, M. F. Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1): Speech and talk. London: Academic Press, 1980.

KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence, R. I: Brown University Press, 1967.

MARSLEN-WILSON, W. D., & WELSH, A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 1978, **10**, 29-63.

MARTIN, J. G. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 1972, **79**, 487-509.

PISONI, D. B. Some current theoretical issues in speech perception. *Cognition*, 1981, **10**, 249-259.

NOTES

1. By word recognition, we mean simply recovering the phonological shape of a word.

2. The acoustical effect of these three changes is clearly not equivalent. To our knowledge, the relationship between acoustic and phonemic changes has remained unexplored in continuous speech perception. Even when a change involves a single phonetic feature, such as voicing, the acoustical effects of the change may be quite varied for different classes of segments and even within one segment class in different word positions. It is not obvious to us how acoustic changes, correlated with phonemic changes. are to be equated.

3. Thirty-one subjects were tested; one subject was excluded from the study because she was unable to perform the shadowing task.

(Manuscript received February 7, 1983;
revision accepted for publication August 25, 1983.)