*Perception & Psychophysics*
*1973, Vol. 1, 153-156*

# Listening for mispronunciations: A measure of what we hear during speech*

RONALD A. COLE
*University of Waterloo, Waterloo, Ontario, Canada*

Ss heard a passage from Lewis Carroll's *Through the Looking Glass* and were asked to indicate, as quickly as possible, whenever they heard a mispronunciation. Mispronunciations were produced by changing one consonant sound in a three-syllable word by one, two, or four distinctive features (e.g., busily to "pizily," "visily," or "sizily"). Mispronunciations involving a single feature change were seldom detected, while two and four feature changes were readily detected. The syllable in which a mispronunciation occurred did not affect the probability of detecting a mispronunciation. However, reaction times to mispronounced words were at least a third of a second slower when they occurred in the first syllable of the word. The results were taken to support the notion that words are identified by their distinctive features.

Speech may be adequately described as a series of phonemes. The word "bit," for example, is composed of three phonemes—/b/, /I/, /t/—and we use this phonemic information in order to discriminate "bit" from "pit," "bet," and "bid." Whereas words may be described in terms of their component phonemes, the phonemes within a word may be described in terms of their distinctive features—distributions of acoustic energy which accompany a phoneme in any syllable context (Jakobson, Fant, & Halle, 1952).

Experiments using isolated syllables (usually consonants spoken with /a/) have shown that phonemes are perceived (Miller & Nicely, 1955), compared (McInish & Tikofsky, 1969; Cole & Scott, 1972a), and remembered (Wickelgren, 1965, 1966) in terms of their distinctive features. Recently, Scott (1971), Cole and Scott (1972b), and Eimas (1972) reported direct evidence for phoneme feature detectors. These studies demonstrated that repeated presentation of a consonant phoneme (paired with /a/) caused individual features to satiate, which resulted in predictable changes in the perception of the syllable.

Experiments with ongoing speech suggest that features may be identified directly at the syllable or word level. Warren (1971) reported that Ss can detect the presence of a target syllable in ongoing speech faster than they could detect the presence of a target phoneme. The faster identification of syllables suggests that the phonemes in a syllable may be identified in parallel, a notion advocated by several investigators (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Massaro, 1972).

Although a listener must attend to certain acoustic features in order to understand speech, it is also known that we do not require a complete listing of acoustic features in order to perceive individual phonemes in a word-level context. Warren (1970) and Warren and Obusek (1971) reported that listeners are able to "fill in" a missing phoneme on the basis of its linguistic context. In these experiments, Ss were presented with a sentence in which the initial /s/ was removed from the word "legislatures" and replaced by a "cough." Ss reported "hearing" the missing phoneme as clearly as if it was actually present, and usually localized the cough several phonemes away from its actual site. Studies of such "phonemic restorations" demonstrate that, under noisy conditions, a listener can generate a phoneme from its surrounding linguistic context. In this case, a phoneme is heard in the absence of any particular acoustic feature which could signal its presence.

The present study attempts to examine the role of individual acoustic features in the perception of ongoing speech. The procedure involves asking Ss to detect mispronounced words which have been embedded in ongoing speech. Words are mispronounced by changing one phoneme in the word by one, two, or four distinctive features. If Ss need only a limited number of acoustic features in order to identify a word, then words mispronounced by a single feature may not be detected as a mispronunciation. Thus, changing a word by one distinctive feature (confusion-gunfusion) should result in fewer detections than changing the word by four distinctive features (confusion-sunfusion). In addition, by varying the syllable in which a mispronunciation occurs, we may determine whether Ss attend differentially to different syllables in a word.

## METHOD

### Subjects

Forty-five undergraduate students from an introductory psychology course served as Ss. All Ss spoke English as their first

*This research was supported in part by a grant from the National Research Council of Canada. I thank Jim Schaefer for his invaluable assistance at all stages of this experiment, and Dominic Massaro, Brian Scott, M. Phillip Bryden, and Loretta Cole for their comments on an earlier draft of the paper.

Table 1
Distinctive Feature Composition of Consonant Sounds Used in This Experiment According to Keyser and Halle's (1968) Distinctive Feature System

| | p | b | m | f | v | k | g | t | d | TH | th | n | s | z | ch | j | sh | zh |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Peripheral | + | + | + | + | + | + | + | – | – | – | – | – | – | – | – | – | – | – |
| Back | – | – | – | – | – | + | + | – | – | – | – | – | – | – | + | + | + | + |
| Hissing-Hushing | – | – | – | + | + | – | – | – | – | – | – | – | + | + | + | + | + | + |
| Nasal | – | – | + | – | – | – | – | – | – | – | – | + | – | – | – | – | – | – |
| Continuing | – | – | – | + | + | – | – | – | – | + | + | – | + | + | – | – | + | + |
| Voiced | – | + | + | – | + | – | + | – | + | – | + | + | – | + | – | + | – | + |

*Note–The symbol + means having the relevant feature; – means lacking the relevant feature.*

language. Each S served in only one of three groups of 15 Ss in a session lasting approximately 30 min.

### Stimuli

The stimulus material was a passage from Lewis Carroll's *Through the Looking Glass* entitled "The Lion and the Unicorn." Forty-five three-syllable words were selected randomly from this chapter. Each word was mispronounced by changing a single consonant phoneme in the word to a new phoneme differing from the original by one, two, or four distinctive features. Thus, while a mispronounced word always differed from the original word by a single phoneme, changes involved one, two, or four distinctive features.

Mispronunciations occurred equally often in the first, second, or third syllable of the word. Mispronunciations in the first syllable always involved the first phoneme in the syllable (e.g., suggested-zuggested), while a mispronunciation in the second or third syllable always involved the final phoneme in the syllable (e.g., Messenger-messemger: introduce-introdush). In addition to the stimulus words already described, 10 monosyllabic words were also mispronounced in each tape. These words were included so that Ss would not become aware that mispronunciations involved three-syllable words.

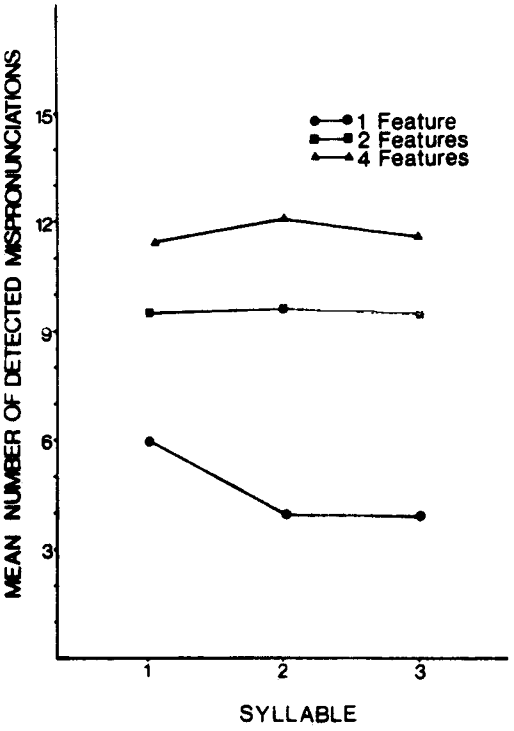

Fig. 1. Mean number of words detected in each feature-change condition as a function of the syllable in which the mispronunciation occurred.

All phonemic changes were made according to the distinctive feature system shown in Table 1. Previous research (e.g., Cole & Scott, 1972) has shown that this distinctive feature system provides a valid measure of the perceptual similarity of different phonemes.

Three stimulus tapes were recorded by a male speaker with a southern Ontario dialect. These tapes were identical in all respects except that all of the mispronounced words in the first tape were mispronounced by one distinctive feature, words in the second tape were changed by two distinctive features, while words in the final tape were changed by four distinctive features. The 15 Ss in each group heard the same experimental tape. Therefore, all Ss in a given group heard words mispronounced by the same number of distinctive features.

At the onset of a mispronounced phoneme, a 300-msec tone was placed on the second channel of the tape. This tone was used to start a Hunter 100-msec timer. The onset of each phoneme was located by manually drawing the tape over the playback head of the tape recorder and monitoring the output via headphones. With practice, this technique allows one to locate the onset of a particular acoustic segment with a standard error of approximately ±5 msec.[1]

In order to insure that words were actually mispronounced as intended, those syllables which were judged difficult to detect in a mispronounced word were removed from their context and presented to listeners in isolation. Disagreements occurred on only four syllables, all of which involved a one-feature change. Sentences containing these syllables were recorded a second time and spliced onto the master tape.

### Procedure

All stimulus material was presented on a Sony Model TC 630 tape recorder connected via a Dynaco power amplifier to a Dynaco loudspeaker. The S was seated in front of the loudspeaker with the index finger of his right hand on a microswitch. He was told that he would hear a story in which some words were mispronounced. The S was instructed to press the button in front of him as quickly as possible whenever he heard a mispronunciation.

## RESULTS

Figure 1 displays the mean number of mispronounced words that were detected in each feature change condition as a function of the syllable in which the mispronunciation occurred. This figure shows that Ss detected fewer than 30% of words mispronounced by one distinctive feature. Words changed by two or four distinctive features were detected with 60% and 75% accuracy, respectively. Analysis of variance revealed a significant effect of distinctive features ($F = 64$, df = 2/42, $p < .001$) and a significant interaction between distinctive features and the syllable in which the mispronunciation occurred ($F = 3.10$, df = 4/84, $p < .02$). This interaction reflects the greater number of detections for words changed by one distinctive feature in the first syllable.

False alarms were also compared in each feature change condition in order to insure that detections were not influenced by the S's willingness to report a possible mispronunciation. Analysis of variance revealed that there was no difference in the number of false alarms made in the different feature change conditions.

Reaction times to mispronunciations in each syllable are displayed in Fig. 2 for the three groups. This figure reveals that (a) RTs were approximately 300 msec longer when a mispronunciation was detected in the first syllable of a word ($F = 80.1$, df = 2/42, $p < .001$), and (b) RTs were approximately 90 msec longer for words mispronounced by a single distinctive feature ($F = 4.66$, df = 2/42, $p < .01$).

## DISCUSSION

This experiment demonstrates that distinctive features are involved at some stage in the recognition of words during ongoing speech. The fact that words mispronounced by one distinctive feature were rarely heard as mispronounced suggests that Ss do not attend to all of the acoustic information that is present in the speech wave.

Failure to detect a mispronounced phoneme in a word was clearly dependent upon hearing the altered phoneme embedded in a larger word-level context. When mispronounced phonemes were removed from their word-level context and presented in isolation, in a CV or VC syllable, Ss always identified the phoneme correctly (i.e., as it was mispronounced).

We may view speech perception as the continuous matching of a set of features identified in the speech wave with a set of features stored in memory for a particular word. It is likely that a listener will recognize a particular word when a certain minimal number of acoustic features are present in the speech wave. We may assume that there is a certain amount of "noise" tolerated in this recognition process, so that a word altered by a single distinctive feature may fall within the normal limits of acceptability for that word. When this occurs, the listener will fail to hear a mispronunciation. When a phoneme in a word is altered by several acoustic features, the resulting distribution of features is not accepted as a normal word, and a mispronunciation is heard.[2]

A second result of this experiment is that words mispronounced by two or four distinctive features were detected equally often in the first, second, or third syllable of the word. This suggests that S attends equally to all syllables of a three-syllable word during ongoing speech.

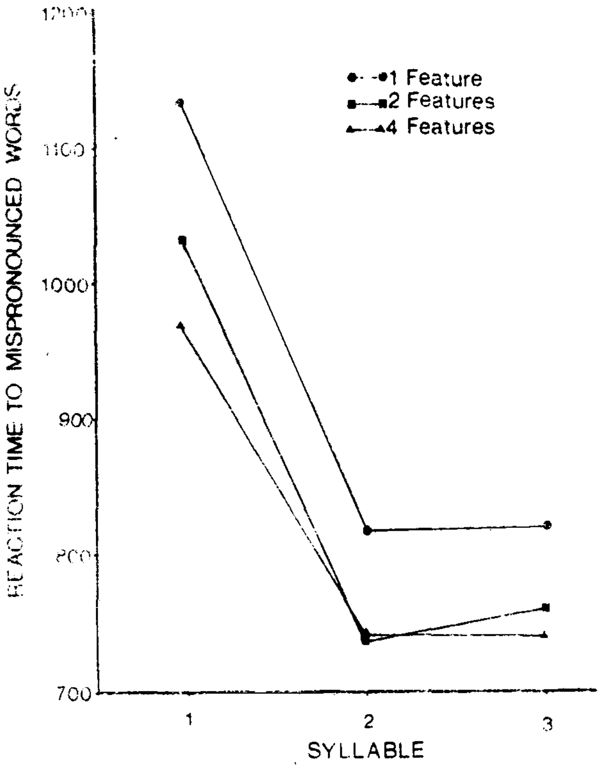

Fig. 2. Reaction time to words detected in each feature-change condition as a function of the syllable in which the mispronunciation occurred.

Finally, it was found that S takes longer to detect a mispronunciation that occurs in the first syllable of a word. This could reflect the fact that S generally needs more information than is provided by the first syllable in a word (and the preceding linguistic context) in order to decide that a mispronunciation has occurred. If S must identify an entire word before he is able to identify a mispronunciation within the word, then RTs should be longer for syllables in the beginning of the word. However, RTs should also be faster for mispronunciations occurring at the end rather than in the middle of the word—and the present data showed no difference in RTs for these syllables. Moreover, the magnitude of the effect—300 msec—is much greater than would be expected if S was simply waiting for additional syllables in order to identify a word.

An alternative explanation is that S "generates" or hypothesizes a word from the preceding linguistic context after hearing its first syllable. When the first syllable has been mispronounced, S may generate an incorrect word, and he will have to change his hypothesis upon hearing additional syllables. Since changing an icorrect hypothesis takes time, RTs would be longer for words mispronounced in the first syllable, but no difference in RTs would be expected for words mispronounced in the second or third syllables.

In addition to the specific results of this experiment, the present research demonstrates that it is possible to use mispronunciations to systematically examine the relationship between sound and meaning. By varying mispronunciations as a function of syntactic, semantic, or other linguistic variables, it is possible to gain a more precise understanding of information processing strategies during ongoing speech.

## REFERENCES

Cole, R., & Scott, B. Distinctive feature control of decision time: Same-different judgments of simultaneously heard phonemes. Perception & Psychophysics, 1972a, 12, 91-94.

Cole, R., & Scott, B. Phoneme feature detectors. Paper presented at meeting of the Eastern Psychological Association, Boston, 1972b.

Eimas, P. Selective adaptation of linguistic feature detectors. Cognitive Psychology, 1972, in press.

Jakobson, R., Fant, C. G. M., & Halle, M. *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge: M.I.T. Press, 1952.

Keyser, S. J., & Halle, M. What we do when we speak. In P. Kolers and M. Eden (Eds.), *Recognizing patterns.* Cambridge: M.I.T. Press, 1968.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-459.

Massaro, D. W. Preperceptual images, processing time, an;perceptual units in auditory perception. Psychological Review, 1972, 79, 124-145.

McInish, J. R., & Tikofsky, R. S. Distinctive features and response latency: A pilot study. Perception & Psychophysics, 1969, 6, 267-268.

Miller, G. & Nicely, P. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 1955, 27, 338-352.

Scott, B. The verbal transformation effect as a function of embedded sounds. Unpublished Master's thesis, University of Waterloo, 1972.

Warren, R. M. Perceptual restoration of missing speech sounds. Science, 1970, 167, 392-393.

Warren, R. M. Identification times for phonemic components of graded complexity and for spelling of speech. Perception & Psychophysics, 1971, 9, 345-349.

Warren, R. M., & Obusek, C. J. Speech perception and phonemic restorations. Perception & Psychophysics, 1971, 9 (3B), 358-362.

Wickelgren, W. A. Distinctive features and errors in short-term memory for English consonants. Journal of the Acoustical Society of America, 1965, 38, 583-588.

Wickelgren, W. A. Distinctive features and errors in short-term memory for English vowels. Journal of the Acoustical Society of America, 1966, 39, 388-398.

## NOTES

1. When the onset of a particular phoneme was difficult to locate (such as nasals which follow a vowel), speech spectrograms were made of the mispronounced word and the distance was measured from the nearest stop consonant to the target phoneme. The mispronounced phoneme was then located on magnetic tape by first finding the stop consonant (which is always preceded by silence and easily localized) and then measuring the distance to the mispronounced phoneme. Speech spectrograms weremade after tone placement to insure that tones were located at the onset of the mispronounced phoneme.

2. Instructions to listen for mispronunciations during speech clearly change the listener's criterion. In an informal demonstration, 200 students were asked to listen to the passage in which words were mispronounced by a single distinctive feature. After listening to the entire tape, fewer than 20 students reported hearing a single mispronunciation.

(Received July 14, 1972;
revision received October 2, 1972.)