

Categorical discrimination of vowels produced in syllable context and in isolation

TERRY L. GOTTFRIED

State University of New York, Purchase, New York

and

JAMES J. JENKINS and WINIFRED STRANGE

University of South Florida, Tampa, Florida

An innovative experimental paradigm that avoids certain problems of response bias in speech perception studies is presented. The paradigm was tested in a replication of an important finding in the perception of American English vowels. The problem was the relative identifiability of vowels in different syllable contexts, /t/-vowel-/t/ (TVT) and isolated vowels (V). The traditional ABX discrimination procedure was converted to a *categorical* discrimination task by having the three stimuli on each trial spoken by different people. This task requires a match according to vowel category, not acoustic identity. The technique eliminates the response-alternative problems of keyword identification tasks. Although overall error rates were low, the original findings were replicated: Listeners were more accurate when discriminating some vowels in TVT than in V syllables. Results are interpreted as support for a theory that considers dynamic acoustic information important for vowel perception.

In a series of studies, Strange and her colleagues (see Shankweiler, Strange, & Verbrugge, 1977) have hypothesized that acoustic information from (at least) the entire coarticulated syllable is used by a listener to determine a vowel's identity. This hypothesis was supported by perceptual experiments that showed that vowels coarticulated with consonants in English were identified with fewer errors than were vowels produced in isolation, even though the former were *more* ambiguous acoustically with respect to target formant frequencies (Gottfried & Strange, 1980; Strange, Edman, & Jenkins, 1979; Strange, Verbrugge, Shankweiler, & Edman, 1976; Verbrugge, Strange, Shankweiler, & Edman, 1976).

Some subsequent studies have not replicated these findings. For example, Macchi (1980) pointed out that there was a stimulus-response incompatibility between the isolated vowel stimuli and the orthographic response alter-

natives (CVC syllables) used in the Strange et al. (1976) study, whereas no such incompatibility existed for the consonantal context condition. Using a different identification task, Macchi did not find a significant difference between context conditions. She hypothesized that stimulus-response incompatibility might have spuriously increased the errors for the isolated vowel stimuli in the Strange et al. study. Strange and Gottfried (1980), however, showed that Macchi's results were attributable to factors other than the task differences. They suggested that differences in subjects' dialect, dialect homogeneity, and manner of stimulus presentation might account for the differences between the two studies, rather than stimulus-response compatibility.

Diehl, McCusker, and Chapman (1981) also addressed the problem of stimulus-response compatibility by comparing groups of listeners who used several different methods of responding in identification tests of isolated and coarticulated vowels. Using CVC keywords, they replicated the Strange et al. (1976) results: Isolated vowels were more difficult to identify than coarticulated vowels. However, when they used isolated vowel symbols (i.e., $\bar{E}\bar{E}$, \bar{I} , \bar{A} , \bar{E} , AE, AH, UH, \bar{O} , $\bar{O}\bar{O}$, $\bar{O}\bar{O}$), all errors greatly increased, and coarticulated vowels were not significantly easier to identify than isolated vowels. It can thus be concluded that naive subjects cannot readily learn to use the isolated vowel symbols appropriately to report their perceptions.

In another condition, Diehl et al. (1981) asked listeners to mimic the vowel stimulus, rather than choose an or-

This research was supported by the Graduate School of the University of Minnesota, by National Institute of Mental Health Grant MH-21153 to J. J. Jenkins and W. Strange, and by National Institute of Child Health and Human Development Grants HD-01136 and HD-07151 and National Science Foundation Grant BNS 77-22075 to the Center for Research in Human Learning. We wish to thank Bruno H. Repp and Vincent J. van Heuven for their helpful suggestions on earlier drafts of the manuscript. We gratefully acknowledge the help of Lenief Heimstead, James Nead, and Karen Siegel in the collection and analysis of the data. Reprint requests may be sent to Terry L. Gottfried, Division of Natural Sciences, State University of New York, College at Purchase, Purchase, NY 10577.

thographic response. They hypothesized that such a mimicking task would decrease task difficulty and eliminate the problem of stimulus memorability differences between isolated and coarticulated vowels. According to their hypothesis, isolated vowels are less familiar to listeners than CVC syllables, and therefore fewer "mnemonic tags" are available to aid in "holding" the isolated vowels in short-term memory. Diehl et al. reported lower overall error rates and no differences between isolated and coarticulated vowels using the mimicking task.

However, the lack of significant differences between the isolated vowels and the coarticulated vowels may have been due to performance ceiling effects as well as to the control of stimulus memorability factors. A more important limitation of the result with the mimicking is the fact that "correct" imitation performance does not provide direct evidence for correct perceptual categorization of the imitated phoneme. A study by Nemes (1971) showed that the kinds of errors made by Hungarian learners of English in an imitation task were different from those made in identification and discrimination tasks of unfamiliar English consonant contrasts. Thus, whether a mimicking task can be used to assess phonemic categorization of speech sounds is questionable.

The purpose of the present study was to examine further the effects of stimulus-response compatibility and memorability on vowel perception. In order to eliminate the problems of response alternatives, a discrimination task that retained the categorial nature of the identification task used in previous studies of vowel perception was devised. Discrimination tasks, in general, put fewer demands on the listener in terms of response complexity. Such discrimination tasks are therefore less subject to the criticism that they are affected by the form of the response alternatives. An ABX discrimination test in which the three stimuli of each trial were spoken by three different people was constructed. Hence, all discrimination judgments required that the listeners judge the stimuli as *same* or *different* on the basis of their vowel category, not on the basis of acoustic identity.

METHOD

Stimulus Materials

Two adult male and two adult female speakers who were natives of Minnesota were recorded individually in a sound-attenuating room, using an Electro-Voice microphone (660SP) and a ReVox tape recorder (A77). Speakers produced two tokens each of 10 American English vowels (/i, I, e, ε, æ, a, ʌ, ɔ, U, u/) in each of two different syllabic forms: TVT and V. Speakers were told to produce each token briskly, "as you would say it in a sentence." The two different tokens for each vowel in a consonantal context were not produced consecutively. (See Gottfried, 1982, for further details about these stimuli.)

After a master series had been obtained from every speaker, all tokens were low-pass filtered (4800-Hz cutoff) and digitized at 10,000 samples/sec by a PDP-11/34 computer. Four different listening tests were constructed: a test consisting of the first TVT tokens of each speaker (TVT1); a test of the second TVT tokens (TVT2); a test of the first V tokens (V1); and a test of the second V tokens (V2).

Each trial consisted of three tokens spoken by three different speakers. The first two tokens of a trial were different from each other (A and

B); the third token was of the same vowel category as the first token (A) or the second token (B). Eight pairs of vowels were tested using this categorial discrimination task: /i, e/, /I, ε/, /e, æ/, /ε, æ/, /a, ʌ/, /a, ɔ/, /ʌ, U/, and /U, u/. These pairs were selected because the vowels of each pair had often been confused with each other in previous studies (see Gottfried, 1982). Each vowel pair appeared 12 times on a test, for a total of 96 trials in random order. The stimuli within each trial were separated by a 1-sec interval. Trials were separated by 3 sec. Trials were blocked in groups of 10, with a 6-sec interval separating blocks.

Subjects and Procedure

Four groups of subjects were tested on the syllable context conditions in the following combinations and orders: TVT1, V1; V1, TVT1; TVT2, V2; and V2, TVT2. There were 40 subjects, 10 in each of the groups. All subjects, volunteers from an introductory psychology class, were native speakers of American English and reported no hearing abnormalities; most were from the Upper Midwest region of the United States.

The subjects were instructed to write "1" next to the trial number if the last of three stimuli was the same as the first stimulus; they were told to write "2" if the last stimulus was the same as the second stimulus. Examples of the trials and five practice trials with feedback were presented before each condition. After the instructions and practice trials, the first 96-trial ABX discrimination test was presented to the subjects. When the first test had been completed, the subjects were given instructions, practice, and the second 96-trial discrimination test.

RESULTS

The overall percentages of errors for the TVT and V conditions are presented in Table 1. (Errors were averaged over the TVT1 and TVT2 and the V1 and V2 conditions.) Overall error rates for the TVT stimuli (5.3% errors) were significantly lower than error rates for the V stimuli (8.2% errors) [$F(1,36) = 25.97$, $MSe = 6.37$, $p < .001$], although the error rates for both groups were quite low. The groups that were given different orders of testing (TVT syllables first or isolated vowels first) did not significantly differ from each other on overall error rates [$F(1,36) = 3.39$, $MSe = 19.65$, $p > .05$]. However, order did significantly affect how well subjects performed in the two different context conditions [$F(1,36) = 14.86$, $p < .001$]. The error rate was higher for the TVT condition when that condition was presented first (7.3% errors) than when it was presented second (3.1%). To a lesser extent, the same was true for the V condition (8.4% vs. 8.0% errors). There were, therefore, significant practice effects during the two discrimination tests.¹

Table 1
Mean Percent Errors for Vowel Pair Comparisons

Vowel Pair	Syllable Context	
	TVT	V
/i, e/	3.6	3.6
/I, ε/	4.4	5.4
/e, æ/	3.2	1.3
/ε, æ/	5.7	15.0
/a, ʌ/	3.8	18.3
/a, ɔ/	15.8	12.3
/ʌ, U/	1.7	6.3
/U, u/	4.2	3.2
Overall Percent Error	5.3	8.2

Table 1 also displays errors for each vowel pair in each condition to show which vowel comparisons were difficult for listeners to discriminate. An analysis of variance on the subject averages, in which vowel comparison pair (e.g., /i, e/, etc.) was a fixed factor and the particular trial with these comparison pairs was a random factor, was performed. Although there was still a significant difference between the TVT and V conditions in discrimination accuracy [$F(1,352) = 7.30$, $MSe = 4.56$, $p < .01$], only some of the vowel comparison pairs contributed to the effect [i.e., the effect of vowel comparison was significant, $F(7,352) = 8.03$, $p < .01$, and the context \times vowel comparison interaction was significant, $F(7,352) = 4.03$, $p < .01$].

In the TVT condition, there were significantly more errors on trials that paired /a/ and /ɔ/ than on trials with any other pairing ($p < .01$). About 40% of the errors in the TVT condition were on pairs of this type. In the V condition, there were significantly more errors on pairs /ε, æ/, /ɑ, ʌ/, and /ɑ, ɔ/ than on the other pairs ($p < .05$). The /ε, æ/ pair contributed 23% of the total errors in the V condition; /ɑ, ʌ/, contributed 28%, and /ɑ, ɔ/ contributed 19%.

Errors on pairs /ε, æ/ and /ɑ, ʌ/ were significantly greater in the V condition than in the TVT condition ($p < .01$ using the Dunn correction for multiple comparisons). All other vowel pairs did not differ significantly in error rates across conditions. These results are similar to the findings of Gottfried and Strange (1980), Strange et al. (1979), and Strange et al. (1976) with respect to the vowel pairs that are particularly confusing.²

DISCUSSION

The results of this experiment confirm the findings of previous experiments (Gottfried & Strange, 1980; Strange et al., 1979; Strange et al., 1976) that isolated vowels are often more difficult to identify than vowels coarticulated in CVC syllables. The differences between the isolated vowels and coarticulated vowels cannot be attributed to experimental artifacts attributable to the response alternatives. In addition, the vowel pairs that caused the greatest number of errors for listeners in discrimination were the same pairs that were often confused in identification tests (e.g., Gottfried, 1982; Gottfried & Strange, 1980). Thus, these particular vowel confusions are not attributable merely to effects of stimulus-response incompatibility or other peculiarities of the response-sheet alternatives.

The fact that only the /ε, æ/ and /ɑ, ʌ/ vowel comparisons contributed to the significant difference between context conditions may be attributed to ceiling effects. Error rates for the discrimination task were very low, and this may have masked differences between the context conditions that are revealed by a more difficult task (e.g., the identification task of Strange et al., 1979, and Strange et al., 1976). Increasing the inter-stimulus interval or introducing noise in the categorial ABX task might increase the task difficulty and allow differences between context conditions to be seen for vowel comparisons other than /ε, æ/ and /ɑ, ʌ/.

This study suggests that there are genuine differences in the information available for the identification of isolated and coarticulated vowels in English. A study by Rakerd (1984) showed that the similarity judgments of American English listeners may be more consistently scaled (using an individual-differences scaling procedure) according to linguistic dimensions when the stimuli are coarticulated vowels rather than isolated vowels. He found that this difference between coarticulated and isolated vowels was not based on previously acquired knowledge, but

was attributable to the acoustic characteristics of the vowel stimuli themselves.

Target theories that assume vowels are cued exclusively by information present in a single spectral cross section (e.g., Joos, 1948; Ladefoged, 1967) are inadequate to account for the results of many of the vowel-identification studies and Rakerd's (1984) experiments. If the information for vowel identity is the target, then the isolated vowels would provide clearer information for vowel perception than would vowels in the context of consonants. However, these presumably more canonical isolated vowels are in fact often identified less accurately and are perceived less linguistically, according to Rakerd's analysis, than are coarticulated vowels.

Strange and her colleagues (Gottfried & Strange, 1980; Strange et al., 1979; Strange, Jenkins, & Johnson, 1983; Strange et al., 1976) have emphasized the role of dynamic acoustic information in specifying the identity of the vowel. Indeed, Assmann, Nearey, and Hogan (1982) found that programs to differentiate vowels acoustically were most successful when spectral information from more than one spectral cross section was utilized. The precise nature of the important acoustic parameters that contribute to the more "linguistic" nature of coarticulated vowels has yet to be delineated, although it is clear that relative duration is very important (e.g., Jenkins, Strange, & Edman, 1983; Peterson & Lehiste, 1960; Rakerd, 1984; Strange et al., 1979; Strange et al., 1983). A detailed specification of the important acoustic parameters in vowel identification also depends in part on the discovery of more adequate acoustic descriptions of the dynamic information that is present in the transition portion of the syllables (see Strange et al., 1983).

REFERENCES

- ASSMANN, P. F., NEAREY, T. M., & HOGAN, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustical aspects. *Journal of the Acoustical Society of America*, **71**, 975-989.
- DIEHL, R. L., MCCUSKER, S. B., & CHAPMAN, L. S. (1981). Perceiving vowels in isolation and in consonantal context. *Journal of the Acoustical Society of America*, **69**, 239-248.
- GOTTFRIED, T. L. (1982). *Perception of French and American English vowels: A cross-language study*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.
- GOTTFRIED, T. L., & STRANGE, W. (1980). Identification of coarticulated vowels. *Journal of the Acoustical Society of America*, **68**, 1626-1635.
- JENKINS, J. J., STRANGE, W., & EDMAN, T. R. (1983). Identification of vowels in "vowelless" syllables. *Perception & Psychophysics*, **34**, 441-450.
- JOOS, M. A. (1948). Acoustic phonetics. *Language Supplement*, **24**, 1-136.
- LADEFOGED, P. (1967). *Three areas of experimental phonetics*. New York: Oxford University Press.
- MACCHI, M. H. (1980). Identification of vowels spoken in isolation versus vowels spoken in consonantal context. *Journal of the Acoustical Society of America*, **68**, 1636-1642.
- NEMSER, W. (1971). *An experimental study of phonological interference in the English of Hungarians*. Bloomington, IN: Indiana University Press.
- PETERSON, G. E., & LEHISTE, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, **32**, 693-703.
- RAKERD, B. (1984). Vowels in consonantal context are perceived more linguistically than are isolated vowels: Evidence from an individual differences scaling study. *Perception & Psychophysics*, **35**, 123-136.
- SHANKWEILER, D. P., STRANGE, W., & VERBRUGGE, R. R. (1977). Speech and the problem of perceptual constancy. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing* (pp. 315-345). Hillsdale, NJ: Erlbaum.
- STRANGE, W., EDMAN, T. R., & JENKINS, J. J. (1979). Acoustic and phonological factors in vowel identification. *Journal of Experimental Psychology: Human Perception and Performance*, **5**, 643-656.
- STRANGE, W., & GOTTFRIED, T. L. (1980). Task variables in the study of vowel identification. *Journal of the Acoustical Society of America*, **68**, 1622-1625.

- STRANGE, W., JENKINS, J. J., & JOHNSON, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, **74**, 695-705.
- STRANGE, W., VERBRUGGE, R. R., SHANKWEILER, D. P., & EDMAN, T. R. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, **60**, 213-224.
- VERBRUGGE, R. R., STRANGE, W., SHANKWEILER, D. P., & EDMAN, T. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, **60**, 198-212.

NOTES

1. Because there were significant asymmetrical transfer effects from one context condition to another, another study that tested different subjects in the two context conditions was conducted. There were still significant differences between the TVT and isolated vowel conditions when the comparison was between subjects rather than within subjects (see Gottfried, 1982).

2. Diehl et al. (1981) suggested that poor articulation could account for the large number of errors in the isolated vowel condition. Unfortunately, they did not provide acoustical analyses of their vowel tokens, but used the perceptual judgments of phonetically trained listeners to eliminate "misproduced" tokens from their analysis. There is reason to believe, however, that even phonetically trained listeners are susceptible (to at least some degree) to the same effects of context on vowel perception as untrained listeners (W. Strange & N. Niccum, personal communication, 1978). Spectrographic analyses of the tokens used in this study (see Gottfried, 1982) showed that, although there were some vowel tokens produced in isolation and in consonantal context that deviated from expected target formant frequencies and durations, these "misproduced" tokens could not account for the higher error rates in the identification of isolated vowels.

(Manuscript received for publication November 11, 1984.)