Monitoring for vowels in isolation and in a consonantal context

Brad Rakerd,[a] Robert R. Verbrugge, and Donald P. Shankweiler
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510, and Department of Psychology, University of Connecticut, Storrs, Connecticut 06268

(Received 25 March 1983; accepted for publication 19 March 1984)

The identifiability of isolated vowels (/V/) was compared to that of vowels in consonantal context (/pVp/) when subjects performed a monitoring task. On successive blocks of trials in a test series, the subjects listened for instances of one or another of nine monophthongal vowels (/i,ɪ,ɛ,æ,ʌ,ɑ,ɔ,ʊ,u/) and identified each test item as being an instance or not. On average, resulting false alarm errors occurred significantly less often in the /pVp/ condition, consistent with the previous finding that vowel perception may be aided by consonantal context. This beneficial effect of context was found to be restricted to the class of open vowels, however, with perception of the close vowels being somewhat hindered by context. The error data for misses also showed an interaction between context and vowel height. Various accounts of the interaction are considered.

PACS numbers: 43.70.Dn, 43.70.Ve

I. MONITORING FOR VOWELS

Of continuing interest in speech research is the question of whether vowel perception is affected by the consonantal context in which a vowel occurs. Perceivers might be expected to exhibit some context sensitivity because the acoustic correlates of a vowel often vary with changes in the identity of neighboring consonants (Stevens and House, 1963; Lindblom, 1963; Broad, 1976). Strong support for this hypothesis comes from studies in which vowels have been shown to be more identifiable in a consonantal context than in isolation (e.g., Strange *et al.*, 1976, 1979; Gottfried and Strange, 1980).

Recently, this evidence has been challenged on grounds that it is largely an artifact of the perceptual task subjects have been asked to perform. It has typically been required that subjects make a multiple-choice identification judgment by: (1) selecting the "best match" to a presented vowel from among a prescribed set of alternatives; and (2) indicating their choice by circling a written form of the alternative on an answer sheet. That written form can be orthographically related to a presented item in varying degrees. It can, for example, be an English spelling of the item itself (e.g., "pep" as the correct response for /pɛp/), or it can be a spelling of a word that contains the "same" vowel as the presented item (e.g., "bed" as the correct response for /pɛp/). The degree of relationship between item and response alternative has been shown to affect vowel identification performance (Macchi, 1980; Diehl *et al.*, 1981; Assmann *et al.*, 1982). This variable was not controlled in early studies of consonantal context (e.g., Strange *et al.*, 1976), therefore the significance of the obtained effect has been called into question (but see Strange and Gottfried, 1980).

The significance of the context effect has also been questioned on the argument that the typical response task—i.e., the searching for and circling of an appropriate alternative on an answer sheet—is itself somewhat biased in favor of the context condition. This is owing to the fact that such a task makes strong demands on short-term memory in that a stimulus trace must be held long enough to be compared with each of the alternatives. Vowels in context might be expected to be somewhat better remembered than isolated vowels for two reasons: (1) vowel–consonant combinations tend to make up words already represented in a subject's lexicon, or at least portions of such words; and (2) in English, the orthographic representations of vowels in context tend to be less ambiguous than those of isolated vowels (see Diehl *et al.*, 1981, for elaboration on this argument).

In light of these methodological concerns over past work, we thought it useful to make a comparison of the identifiability of vowels in and out of context with a different kind of perceptual task than has previously been employed. While such a task would, no doubt, have certain limitations of its own, it was felt that if these were sufficiently different from the limitations of the multiple-choice identification task, the results could speak to the methodological generalization of any effects of consonantal context. The specific task we set for subjects was that of monitoring lists of test items for instances of particular target vowels. Subjects simply checked "yes" on an answer sheet if a presented item (an isolated vowel or a vowel in context) was judged to be an instance of the target vowel being monitored and "no" if it was not.

This method has two virtues that are noteworthy: it is comparatively free from orthographic bias since there are no written vowel alternatives on the answer sheet, and it imposes minimal memory demands on the subject since a presented item can be immediately judged to match the target or not. Monitoring thus affords a good comparison with the identification method of past studies. Here, we strengthened that comparison further by examining vowel stimuli for which perceptual data had already been collected with the previous method (Strange *et al.*, 1976).[1]

A. Experimental methods

1. Stimuli

All /pVp/ and /V/ stimuli were produced by a single male talker who spoke an Upper Midwestern dialect of English. For each condition, he produced five tokens of each of

[a] Current address: The Department of Psychology, Michigan State University, Psychology Research Building, East Lansing, MI 48824.

 0001-4966/84/070027-05$00.80

the nine vowels. These were organized into /pVp/ and /V/ test series according to the following protocol: (1) 90 items (two repetitions of each token) were assembled in randomized order to make up a block; (2) monitoring instructions identifying the particular vowel to be listened for in that block were inserted at its beginning; (3) instructions reminding the subject of the target vowel were inserted after the 30th and 60th items; (4) steps (1) through (3) were repeated for a total of nine test blocks. There was a 2-s pause between test items and a 30-s pause between blocks.

The monitoring and reminder instructions were recorded by a male speaker with the same dialect as that of the speaker who had produced the test stimuli. For both experimental conditions, the monitoring instructions were given in the following form: "In this test block, you will be listening for the vowel (exemplar 1), as in (CVC 1), (CVC 2), (CVC 3). Listen for the vowel (exemplar 2), (exemplar 3), (exemplar 4)." The exemplars were isolated productions of the vowel. The CVCs were English monosyllabic words that contained the vowel.[2] The reminder instructions were as follows: "Remember, you are listening for the vowel (exemplar 5), (exemplar 6), (exemplar 7)."

The order in which vowels were monitored was varied across listeners; nine different orders were generated with the constraint that each of the nine vowels was monitored in each ordinal position.

a. Acoustic characteristics of the stimuli. These stimuli are a subset of the items employed in a previous study of vowel perception (Strange *et al.*, 1976). Their acoustic characteristics conform to generalizations reported in that study. The first of these generalizations is that the formant frequencies of all isolated vowels except /ɔ/ were comparable to normative values reported by Peterson and Barney (1952). The deviations in /ɔ/ reflect an idiosyncrasy of the speaker's dialect. Average first formant frequencies for the vowels in /pVp/ context were comparable to the values for isolated vowels. In contrast, the second formant frequencies of /pVp/ vowels were somewhat "reduced" (cf. Lindblom, 1963) relative to the isolated vowels. That is to say, they exhibited a somewhat smaller range of deviation about the average value for all vowels in the set.

The isolated vowels were, on average, considerably longer than /pVp/ vowels. Relative durations of vowels in the two conditions were roughly comparable, however. As might be expected on the basis of previous reports (e.g., Peterson and Lehiste, 1960), the vowels /ɪ,ɛ,ʌ,ʊ/ generally were the briefest in duration, /i,u/ were intermediate, and /æ,ɑ,ɔ/ were the longest. The exceptions to this were the vowel /u/ in the /pVp/ context and the vowels /ɑ,ɔ/ in isolation, all of which were somewhat shorter than expected.

2. Subjects

Thirty-six undergraduates, enrolled in an introductory psychology course at the University of Connecticut, participated in this experiment. They were randomly assigned to the /pVp/ and /V/ conditions, so that there were 18 subjects in each condition. All of the subjects were adult native speakers of English with normal hearing. They had no knowledge of the purpose of this study.

3. Procedure

Subjects were asked to monitor the lists of test items for occurrences of the monophthongal vowels /i,ɪ,ɛ,æ,ʌ,ɑ,ɔ,ʊ,u/. They reported their decisions by checking "yes" on an answer sheet if an item was judged to be an instance of the vowel being monitored on a trial and "no" if it was not.

Instructions and test materials were presented over headphones with the volume adjusted to a comfortable listening level, conditions comparable to those employed in the previous identification study conducted with these same stimuli (Strange *et al.*, 1976). Subjects were tested, two at a time, in a sound-attenuated room. Prior to the start of testing, they were familiarized with the stimulus and response materials in the following way. First, the testing procedure was described. It was explained that a number of different speech stimuli would be presented and that the task would be to monitor the vowels in the manner described above. Next, a randomly selected sample of the stimuli was presented. For the first few trials (approximately 15), subjects were asked to listen to the stimuli and make no response. Then, they were given a sample answer sheet and, for 30 trials, monitored the sample items for instances of a particular target vowel. This target was randomly varied across subjects. No feedback was given as to the accuracy of these practice responses; subjects were, however, allowed to ask questions of clarification about all aspects of the procedure. The test was begun only after all subjects expressed confidence that they completely understood the task.

B. Results

With this monitoring procedure, subjects could make errors of two types: false alarms and misses. False alarms were erroneous acceptances of vowels other than the target—responding "yes" when the correct choice was "no." Misses were failures to recognize actual instances of the vowel being monitored—responding "no" when the correct choice was "yes." Neither type of error was significantly related to the order in which the vowels were monitored; consequently, the data that will now be considered were pooled across monitoring orders.

1. False alarms

In the left half of Table I, *composite* false alarm error rates are summarized for each vowel category. A composite false alarm resulted whenever a presented vowel was erroneously taken to be an instance of any of the other alternatives. For example, in the isolated condition, listeners variously misheard the vowel /ʌ/ to be an instance of /æ/, /ɑ/, and /ʊ/. Together, these false alarms occurred on 8.4% of all trials in which /ʌ/ was the presented vowel but was not, in fact, the target. Since many vowel pairs (/ʌ-i/ for instance) were seldom if ever confused, averaging over all of the alternatives in this way generally resulted in rather low error rates. However, this measure of false alarms is perhaps the most comparable to the miss percentage to be considered below and it will be seen that the data exhibit a similar structure.

TABLE I. Average composite false alarm and miss error rates.

Vowel	Percentage of errors: Composite false alarms /pVp/	/V/	/pVp/-/V/	Misses /pVp/	/V/	/pVp/-/V/
i	0.5	0.2	+0.3	2.2	1.7	+0.5
ɪ	0.3	1.9	−1.6	6.1	1.1	+5.0
ɛ	1.7	7.7	−6.0	6.1	11.1	−5.0
æ	1.2	2.4	−1.2	3.9	11.7	−7.8
ʌ	3.3	8.4	−5.1	6.1	17.8	−11.7
ɑ	4.0	5.2	−1.2	26.1	41.1	−15.0
ʊ	4.7	2.7	+2.0	17.8	13.3	+4.5
u	2.2	0.6	+1.6	6.7	0.6	+6.1
Overall	2.2	3.6	−1.4	9.4	12.3	−2.9
ɔ	9.2	6.0	+3.2	9.4	3.9	+5.5

TABLE II. Average error rates for the major false alarm vowel pairs.

Vowel pair	Percentage of false alarm errors /pVp/	/V/	/pVp/-/V/
ɛ-æ	4.8	30.8	−26.0
ʌ-ɑ	14.2	31.6	−17.4
ʌ-ʊ	15.6	13.0	+2.6
ʊ-u	9.5	3.6	+5.9
Overall	11.0	19.8	−8.8
ɔ-ɑ	50.8	59.7	−8.9

The two leftmost columns of Table I report composite false alarm rates for the consonantal-context (/pVp/) and isolated (/V/) conditions, respectively. The difference in error rates between these two conditions is given in the third column (/pVp/-/V/). Results for the vowel /ɔ/ are reported separately in the table. This is because the acoustic characteristics of /ɔ/ proved to be abnormal and because this vowel behaved differently than the other open vowels, both here and in the comparison study of Strange *et al.* (1976). (For further consideration of this difference see the Discussion section below.) Arc sin transformations of the composite false alarm data shown in Table I, and of all other data to be discussed, were submitted to analysis of variance.[3]

Two of the findings regarding composite false alarms speak to the question of whether or not consonants exert a contextual influence on vowel perception. The first is that, overall, error rates in the consonantal condition were significantly lower than in the isolated condition [$F(1,34) = 4.20$, $p < 0.05$]. This indicates that when listeners monitor vowels, as when they perform other identification tasks (Strange *et al.*, 1976, 1979; Gottfried and Strange, 1980), their performance may be positively influenced by the presence of neighboring consonants. The second finding is that the beneficial effect of context was not in evidence for all vowels (see column three of Table I). Generally speaking, it was the perception of open vowels that was aided by context, with perception of close vowels proving to be somewhat *poorer* in the context condition. The only exceptions to this generalization were the vowels /ɔ/, which behaved anomalously throughout, and /ɪ/, which was seldom confused with the other vowels in either condition. This difference between the open and close vowels was reflected in a significant interaction between context and vowel height [$F(1,34) = 20.84$, $p < 0.001$]. *Post hoc* examination of this interaction revealed that the simple main effect of context was significant only for open vowels [$F(1,34) = 18.20$, $p < 0.001$].

As noted above, false alarm errors occurred only rarely for many of the vowel pairs. However, a few pairs did show false alarm rates that were rather high. These are summarized in Table II. Note that the mean false alarm rate for these vowel pairs was at least five times as great as the mean composite false alarm rate in both the /pVp/ and /V/ conditions. Hence, these *high-likelihood* false alarm pairs were the major contributors to overall error scores. The two observations made about the composite false alarm data apply to these high-likelihood false alarms as well. First, overall identifiability of the vowels was enhanced by context. There were significantly fewer errors in the /pVp/ condition [$F(1,34) = 8.88$, $p < 0.01$]. Second, there was a significant interaction between context and vowel height [$F(3,102) = 11.05$, $p < 0.001$] reflecting the fact that errors on open vowel pairs occurred significantly less often in consonantal context[4] [/ɛ-æ/: $F(3,102) = 25.64$, $p < 0.001$; /ʌ-ɑ/: $F(3,102) = 13.62$, $p < 0.001$] and those on the close pair (/ʊ-u/) occurred more often but not significantly so.

2. Misses

Miss errors are reported on the right half of Table I. It can be seen that their overall pattern parallels that of the false alarms. Subjects were, however, much more variable in exhibiting the pattern with misses. As a consequence, the main effect of context was not significant for these data [$F(1,34) < 1.0$]. There was a highly significant context-by-vowel height interaction [$F(1,34) = 15.54$, $p < 0.001$]. As before, this resulted from the fact that performance on the open vowels was significantly aided by context [$F(1,34) = 8.90$, $p < 0.01$], while that on the close vowels was hindered to a lesser and nonsignificant degree. Also as before, /ɔ/ behaved differently from the other open vowels. It was missed more frequently in the consonantal condition than in isolation.

3. The question of response biases

Although we have looked at false alarm and miss errors separately, the two are not strictly independent. Notice, for example, that if the subjects in this experiment had (for any reason) chosen to respond "yes" on all monitoring trials we would have observed no miss errors and 100% false alarms. Conversely, a bias toward "no" responses would have inflated misses and deflated false alarms. It is therefore reasonable to wonder whether some or all of the effects that we observed can be attributed to systematic response biases. Given the overall patterning of the two types of errors, this possibility can be confidently rejected. It has been noted throughout that the data structure of the miss and false alarm errors was roughly the same. If there were significant response biases, we should have expected the two types of errors to have *complementary* distributions, not comparable ones. For example, we should have expected that the observed interactions between context and vowel height would have been in

opposite directions for the two types of errors. They were not.

C. Discussion

We compared listeners' ability to identify vowels in and out of a consonantal context (/pVp/) when they performed a monitoring task and found that they made significantly fewer false alarm errors (both composite false alarms and high-likelihood false alarms) in the /pVp/ condition. This clearly supports the view that the contextual advantage for vowel perception observed here and elsewhere (Strange *et al.*, 1976, 1979; Gottfried and Strange, 1980) is a genuine perceptual effect and not simply a methodological artifact. At the same time, however, these monitoring results add to evidence indicating that the demonstrability of a contextual influence may be greatly affected by task variables. Pooling misses and false alarms, our subjects made an average of 4.1% errors in the /pVp/ condition and 5.3% in the /V/ condition. In the comparison identification study conducted with these same stimuli (Strange *et al.*, 1976), substantially different error rates were reported. In that instance, there were 9.7% errors in the /pVp/ condition and 33.1% in the /V/ condition. Clearly, absolute error rates can vary substantially with the method of assessment, and these form the baseline against which any relative influence of consonantal context must be measured.

There were two additional points of agreement with the study of Strange *et al.* (1976) that merit comment. The first involves the vowel /ɔ/, which did not behave like the other open vowels in the present instance. It turns out that perception of /ɔ/ was anomalous in that earlier study as well. This can be seen in Table III, which summarizes their multiple-choice identification data for our speaker's tokens (these data are excerpted from the segregated-talker condition of experiment I in Strange *et al.*, 1976). Notice that with their method Strange *et al.* observed a contextual advantage for the identification of all vowels *except* /ɔ/. It appears that the unusual perception of this vowel reflects some abnormality in its production. This conclusion is further supported by the fact that formant frequencies for /ɔ/—as produced in both conditions—were very different from population norms.

The second point of comparison with Strange *et al.* (1976) concerns the perceptual interaction between context and vowel height that we observed. Some analog to that interaction can also be seen in their data. Note in Table III that while all vowels in their /pVp/ condition (except /ɔ/) were identified more accurately than the isolated counterparts, open vowels were much more aided by context than the close vowels. The mean contextual advantage (/pVp/–/V/) for the open vowels was 40.3%, while that for the close vowels was only 12.7%. In both studies, then, we see some evidence that the presence of a /pVp/ context differentially affected perception of the open and close vowels.

TABLE III. Identification error rates determined by Strange *et al.* (1976). Data are for the single male talker in their segregated-talker condition of experiment I.

Vowel	Percentage of identification errors /pVp/	/V/	/pVp/-/V/
i	0.0	11.0	−11.0
ɪ	0.9	14.0	−13.1
ɛ	1.8	63.0	−61.2
æ	1.8	19.0	−17.2
ʌ	6.4	57.0	−50.6
ɑ	42.7	75.0	−32.3
ʊ	15.5	33.0	−17.5
u	1.8	11.0	−9.2
ɔ	16.4	15.0	+1.4
Overall	9.7	33.1	−23.4

Though the acoustic and/or articulatory origins of this effect are yet to be confidently determined, we can make some preliminary observations. First, we may note that no satisfactory explanation of it is likely to be advanced in terms of formant frequency differences among the vowels. Owing to the phenomenon of vowel reduction (Lindblom, 1963), those differences were in fact *less* great in the more perceptually distinctive /pVp/ condition. A more promising acoustic account is that the perceptual effect somehow results from the greater degree of spectral change associated with open vowels. In /pVp/ context, open vowels are typically marked by more extensive formant transitions out of and into the flanking consonants than are close vowels. Vowel height should be particularly related to transitions of the first formant. There have been speculations that acoustic dynamics of this sort positively influence vowel perception (Strange *et al.*, 1976, 1983).

The acoustics also provide evidence that vowels and consonants were coarticulated in the /pVp/ condition—vowel formant frequencies were reduced in this context. This has led us to consider an articulatory account of the perceptual effect. It may be that the beneficial influence of /pVp/ context was focused on the open vowels because those vowels are coarticulated with the consonants in some manner in which closed vowels are not. This could pertain particularly to articulatory movements of the jaw. The jaw lowering required for production of open vowels must be coordinated with jaw raising to achieve bilabial closure for the consonants. While production of the close vowels would likewise call for some jaw lowering (and hence for some articulatory coordination with the consonants), it is conceivable that this requirement differs in kind or degree from that for the open vowels. If listeners are aware of such a coarticulatory difference, it could affect their interpretation of the acoustic signal.

We plan to distinguish between these alternative accounts of the perceptual effect by looking at vowel monitoring performance in other consonantal contexts. While perception of the open vowels was particularly aided by /pVp/ context in the present study, we expect that rather different interactions will occur with consonants of some other place and manner of articulation.

ACKNOWLEDGMENTS

This research was supported by grants from NIH (HD01994) and BRS (RR05596) to Haskins Laboratories. We wish to thank Winifred Strange and Carol Fowler for their assistance in the design of this study and for their comments on earlier versions of the manuscript.

[1]Those earlier data are for the single male talker in the segregated-talker condition of experiment I in Strange *et al.* (1976).

[2]The CVCs in the monitoring instructions were as follows. For the vowel /i/: *green, peak, seal*; /ɪ/: *bit, tin, sick*; /ɛ/: *pen, wet, step*; /æ/: *hat, fan, map*; /ʌ/: *cup, gum, rut*; /ɑ/: *top, sock, dot*; /ɔ/: *fog, call, gone*; /ʊ/: *put, look, should*; /u/: *boot, cool, moon*.

[3]Because the error rates were often close to zero, they were transformed according to the following formula suggested by Winer (1962): $X' = 2 \text{ arc sin}\sqrt{X + 1/N}$, where X is the original score, X' is the transformed score, and N is the number of subjects in a condition. A one-within (vowel height), one-between (context) analysis of variance was performed on the transformed scores.

[4]In this instance /ɔ/ again proved to behave differently from the other open vowels. False-alarm confusions between members of the vowel pairs /ɛ-æ/ and /ʌ-ɑ/ were greatly reduced by /pVp/ context for both "directions"—i.e., with regard to confusions of the first member of the pair with the second *and* the second with the first. This was not the case with /ɔ-ɑ/, however. In the /pVp/ condition, /ɑ/ was misheard as /ɔ/ much less frequently than in isolation (47.2% false alarms versus 75.5%), but /ɔ/ was misheard as /ɑ/ *more* frequently in this condition (54.4% false alarms versus 43.9%). The data reported in Table II reflect the average of these two types of confusions.

Assmann, P. F., Neary, T. M., and Hogan, J. T. (**1982**). "Vowel identification: Orthographic, perceptual, and acoustic aspects," J. Acoust. Soc. Am. **71**, 975–989.

Broad, D. J. (**1976**). "Toward defining acoustic phonetic equivalence for vowels," Phonetica **33**, 401–424.

Diehl, R. L., McCusker, S. B., and Chapman, L. S. (**1981**). "Perceiving vowels in isolation and in consonantal context," J. Acoust. Soc. Am. **68**, 239–248.

Gottfried, T. L., and Strange, W. (**1980**). "Identification of coarticulated vowels," J. Acoust. Soc. Am. **68**, 1626–1635.

Lindblom, B. E. F. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.

Macchi, M. J. (**1980**). "Identification of vowels spoken in isolation and in consonantal context," J. Acoust. Soc. Am. **68**, 1636–1642.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Stevens, K. N., and House, A. S. (**1963**). "Perturbation of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**, 111–128.

Strange, W., Edman, T. R., and Jenkins, J. J. (**1979**). "Acoustic and phonological factors in vowel identification," J. Exp. Psychol.: Human Percept. Perform. **5**, 643–656.

Strange, W., Jenkins, J. J., and Johnson, T. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.

Strange, W., and Gottfried, T. L. (**1980**). "Task variables in the study of vowel perception," J. Acoust. Soc. Am. **68**, 1622–1625.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). "Consonantal environment specifies vowel identity," J. Acoust. Soc. Am. **60**, 213–224.

Winer, B. J. (**1962**). *Statistical Principles in Experimental Design* (McGraw-Hill, New York).