Auditory contrast and speaker quality variation in vowel perception

Robert Allen Fox
Speech and Hearing Science, The Ohio State University, 154 North Oval Mall, Columbus, Ohio 43210

(Received 11 May 1984; accepted for publication 29 November 1984)

Selective adaption and anchoring effects in speech perception have generated several different hypotheses regarding the nature of contextual contrast, including auditory/phonetic feature detector fatigue, response bias, and auditory contrast. In the present study three different seven-step [hɪd]-[hɛd] continua were constructed to represent a low F_0 (long vocal tract source), a high F_0 (long vocal tract source), and a high F_0 (short vocal tract source), respectively. Subjects identified the tokens from each of the stimulus continua under two conditions: an equiprobable control and an anchoring condition which included an endpoint stimulus from one of the three continua occurring at least three times more often than any other single stimulus. Differential contrast effects were found depending on whether the anchor differed from the test stimuli in terms of F_0, absolute formant frequencies, or both. Results were inconsistent with both the feature detector fatigue and response bias hypothesis. Rather, the obtained data suggest that vowel contrast occurs on the basis of normalized formant values, thus supporting a version of the auditory-contrast theory.

PACS numbers: 43.71.Es, 43.66.Mk

INTRODUCTION

As has often been noted in the literature, the identification of a particular vowel sound can be significantly affected by the surrounding phonetic context (e.g., Ladefoged and Broadbent, 1957; Eimas, 1963; Fry *et al.*, 1962; Thompson and Hollien, 1970). In addition, some data have suggested that this contextual effect can extend to vowel discrimination as well (Sawusch *et al.*, 1980). In general, this influence has been found to be one of contrast.

In the last decade many studies have been concerned with understanding the nature of this contextual effect on vowel identification using such techniques as pairwise stimulus presentation (e.g., Repp *et al.*, 1979; Healy and Repp, 1982; Crowder, 1982; Crowder and Repp, 1984), selective adaptation (Morse *et al.*, 1976), or anchoring (Sawusch and Nusbaum, 1979; Sawusch *et al.*, 1980). In each of these methods stimuli from an acoustic continuum are presented in the context of other items whose effect on the perception of the test stimuli is assessed. The primary difference between the procedures lies in (1) the distribution and number of the contextual stimuli and (2) whether or not the subject is required to identify the contextual stimulus item. The typical result in all such procedures is that the phoneme (category) boundary of the stimulus continuum is shifted relative to the baseline boundary (i.e., the boundary obtained when the contextual stimuli are not present) in the direction of the context. That is, relatively ambiguous stimuli are perceived as contrasting with the other stimuli in their vicinity.

There are several possible (and competing) theoretical explanations for such vowel contrast. These different hypotheses include: (1) feature-detector fatigue—proposed primarily to explain contrastive effects in consonant identification (Cooper, 1975; Eimas and Miller, 1978); (2) response bias (e.g., range-frequency theory, Parducci, 1963, 1965, 1975); (3) auditory contrast and/or changes in adaptation level (Simon and Studdert-Kennedy, 1978; Diehl *et al.*, 1978; Sawusch and Nusbaum, 1979); and (4) recurrent lateral inhibition in auditory memory (Crowder, 1981). These hypotheses differ in terms of whether the contextual effect is caused by changes in input processing, changes in labeling strategy (i.e., response bias), and/or changes in auditory memory.

The feature-detector fatigue hypothesis has enjoyed much popularity in the past decade, but has come under increasing criticism in the last several years. This hypothesis is primarily connected with studies utilizing the selective adaptation procedure on consonantal distinctions (e.g., Eimas *et al.*, 1973; Eimas and Corbit, 1973; Cooper, 1974; Diehl, 1975; Cooper *et al.*, 1976; see Ades, 1976; and Eimas and Miller, 1978, for reviews) and is based on the assumption that there are detectors in the perceptual system specialized for phonetic (or perhaps auditory) features. However, this hypothesis has rarely been mentioned in connection with vowel perception, presumably because the large number of vowel categories and the relatively noncategorical perception of the stimuli made explanations in terms of discrete detectors seem unattractive. Also, while feature-detector fatigue is a plausible mechanism for explaining selective adaptation effects, it cannot account for pairwise contrast where only a single contextual item is presented.

A second explanation for contrast effects is response bias. This hypothesis suggests that adaptation/anchoring effects may be a product of changes in subjects' labeling strategy rather than sensory fatigue or modifications of internal perceptual referents. For example, subjects may simply use response categories other than that of the adjacent or more frequently occurring stimulus (cf. the range-frequency theory of Parducci, 1963, 1965, and 1975). However, this hypothesis cannot explain all the contextual effects found with consonants and vowels. First, contextual stimuli not drawn from the test series continuum may also produce category

 0001-4966/85/041552-08$00.80

shifts (Eimas and Corbit, 1973). Second, the degree of contrast depends, in part, on the degree of spectral overlap between the context and test series stimuli (Sawusch, 1977). Finally, alerting subjects to the nature of the stimuli (e.g., specifically warning subjects that a certain stimulus would occur more frequently) does not reduce the contrastive effects (Ladefoged and Broadbent, 1957; Sawusch and Nusbaum, 1979). As Simon and Studdert-Kennedy (1978) have pointed out, the weight of the relevant evidence seems to rule out response bias as a major determinant of adaptation/anchoring effects.

The third explanation of contextual contrast proposes that such effects arise from interactions in auditory memory. This explanation has been applied specifically to vowel perception because vowels, unlike most consonants, leave a strong trace in auditory memory (Crowder, 1971, 1973). For example, Simon and Studdert-Kennedy (1978) hypothesize that contrast occurs when an adaptor establishes an *auditory ground* against which subsequent test stimuli are compared. A similar response contrast hypothesis has been proposed by Diehl *et al.* (1978). Sawusch and Nusbaum (1979) argue for a variation of adaptation-level theory (Helson, 1964, 1971) proposed by Restle (1978), according to which stimulus identity is determined partially on the basis of contrast with (1) the immediate context and (2) the auditory background or adaptation level, representing an appropriately weighted measure of the acoustic characteristics of the stimuli previously presented. However, this theory makes no assumptions about the mechanism responsible for the contrast effects. Feature-detector fatigue is a possible mechanism; however, we have already dismissed this hypothesis. More recently, a process-oriented model of auditory contrast has been developed by Crowder (1981) based on his theory of inteference in auditory memory (Crowder, 1978).

Crowder (1978, 1981) assumes that auditory events are represented in memory on a grid which encodes the time of arrival and the physical channel of the event. The memorial representation is considered to be similar to a smudged wideband spectrogram. The physical channel distinction represents "the traditional selective-attention sense of communication channel—the dimension on which two voices of the same sex are moderately discrepant, two voices of the opposite sex are more discrepant, and on which a speech signal and a noise are extremely discrepant" (Crowder, 1981, p. 174). Contrast effects are assumed to result from recurrent lateral inhibition among the representations in auditory memory in a manner similar to that found in the retina of the horseshoe crab (cf. Cornsweet, 1970). Mutual inhibition occurs if two representations are spectrally similar, close in time, and "similar or identical" in physical channel. This lateral inhibition is considered to involve only those spectral components which the auditory representations have in common. For example, if a prototypic version of [ɪ] and a token ambiguous between [ɪ] and [ɛ] mutually inhibit one another, then they would "cancel out" the formant areas they have in common (see also Crowder and Repp, 1984). Since the prototypic [ɪ] would have a lower first formant ($F1$) than the ambiguous token, after mutual inhibition occurs the mean $F1$ of the ambiguous token (which would be based on the formant areas *not* in common) would be raised so that the token would more likely be identified as [ɛ] (i.e., in contrast to the prototypic [ɪ]).

This model seems to explain the published vowel contrast data quite well, but one important aspect of the assumption of different *physical channels* needs further clarification. A strict interpretation of Crowder's (1981) model would suggest that very little lateral inhibition would occur between auditory representations from relatively discrepant physical channels (such as between two very distinct voice qualities), but no relevant contrast data are available. For example, are the physical channel distinctions based on fundamental frequency, overall formant frequency range, or both?

In one of the earliest studies of the vowel contrast effect, Ladefoged and Broadbent (1957) demonstrated that rescaling the formant pattern of a short introductory phrase ("please say what this word is") has a significant impact on the identification of a following speech token. In terms of Crowder's model one might consider the precursor phrase and the following token to represent different physical channels, yet a significant contrast effect was produced. However, this result might have been obtained because some cross-channel inhibition occurred, or because contrast effects stem from a different source (as described above).

In selective adaptation study, Morse *et al.* (1976, experiment II) used a 13-step [i-ɪ-ɛ] synthetic vowel continuum with a woman's naturally produced [gig] and [gɛg] as adaptors. The formant-frequency (F pattern) differences between the test series and the adaptors (plus, we assume, the fundamental frequency difference—Morse *et al.* did not give the relevant F_0 measurements) would seem to represent two distinct physical channels and would provide a test of Crowder's (1981) hypothesis in terms of vowel contrast. Morse *et al.* found that the [gig] but not the [gɛg] adaptor produced a significant category boundary shift. They argued that these results demonstrated that a boundary shift found in an earlier experiment (experiment I, which used stimulus 13 [ɛ] from the test series as an adaptor) was a result of *auditory* feature-detector fatigue. Since [gig] *did* produce a significant boundary shift, they argue that "*more complex auditory*" (perhaps *phonetic*) feature detectors governed the perception of [i] but not [ɛ], citing the special status of [i] as a "point" vowel. Before these arguments are accepted on their face value, one should more closely examine the stimuli. The approximate formant frequency measurements given ("approximate" because the adaptor was naturally produced and diphthongized) suggest that formants of [gɛg] (particularly $F1$) did *not* overlap with the formants of the test stimuli. The adaptor [gig], on the other hand, had an $F1$ (300 Hz) which was very close to that of stimulus 3 of the 13-step [i-ɪ-ɛ] continuum (298 Hz). One could reasonably argue that the [gig] adaptor shifted the category boundary because of *auditory* contrast in the 300-Hz frequency range. The [gɛg] adaptor would *not* produce the same effect on the [ɛ] end of the test continuum since the first formant of [gɛg] ($F1 = 700$ Hz) did not overlap with even the highest $F1$ (stimulus 13, $F1 = 530$ Hz) of the test series.

Despite the possibility that spectral overlap played a role, this type of auditory contrast is not compatible with the

hypothesis of strict separation between discrepant physical channels. However, it is difficult to determine the role which different physical channels played in the Morse *et al.* (1976) study because there were so many confounded variables. In particular, the test series differed from the adaptors in at least four different ways: (1) The test series were isolated vowel tokens but the adaptors were in [g_g] environment; (2) the test vowels were synthetic and monophthongal, and the adaptors were naturally produced and diphthongal; (3) one must assume a fundamental frequency difference between the test stimuli (synthesized with a contour typical of a male speaker) and the adaptors (produced by female speakers), although Morse *et al.* gave no $F0$ values; and (4) there were formant-pattern differences present between the "male" test stimuli and the female-produced adaptors.

The present study examines how variations in speaker quality affects the vowel contrast phenomenon. The results should help determine the extent to which contrast effects are auditory (involving precategorical acoustic representations) as opposed to phonetic (involving phonetically categorized representations) in nature and whether such contrast can take place across identical, moderately discrepant, or strongly discrepant physical channels (i.e., different sources). To address these issues, an experiment utilizing the anchoring procedure was designed to discover the extent to which listeners' identifications of a set of vowel stimuli would be differentially shifted by anchors which, although of the same phonetic value, would differ in terms of $F0$ or absolute formant frequency. These acoustic variations parallel those found in natural vowels produced by male (low $F0$, long vocal tract) versus female (high $F0$, short vocal tract) speakers.

I. METHODS

A. Subjects

The subjects were 144 undergraduates at The Ohio State University solicited through the student paper and paid for participating in the experiment. None of the subjects had previously participated in a speech experiment. All subjects were native speakers of English with no known hearing impairment.

B. Stimuli

Three different seven-step vowel continua were constructed in a [h_d] context. The stimuli were generated using the Klatt cascade/parallel speech synthesis program implemented on a PDP 11/23 computer. The stepwise variations within each of the vowel continua were produced by varying the frequencies of the first three formants. Formant frequencies for the first vowel continuum (series A) are shown in Table I. All vowels were steady state (no change in formant frequencies over time) and each stimulus was 420 ms in duration. The formant bandwidths for $F1$, $F2$, and $F3$ were 50, 100, and 104 Hz, respectively, for the steady-state portions of all three vowel continua. $F0$ for each token began at 130 Hz and fell linearly, reaching 100 Hz after 360 ms and remaining at 100 Hz until the end of the token. The formant transitions to the final [d] were 30 ms in duration and went from the formant values specified in Table I to 200, 1600, and 2600 Hz for $F1$, $F2$, and $F3$, respectively. The series A continuum was designed to represent a formant frequency pattern and $F0$ contour characteristic of a speaker with a low $F0$ and long vocal tract (both characteristic of a male speaker). The second vowel continuum (series B) was identical to the series A continuum except that the $F0$ contour for each token began at 220 Hz and fell linearly to 190 Hz, remaining at this level for the last 60 ms of the token. Series B was designed to represent a speaker with a long vocal tract (characteristic of a male speaker) but having a relatively high fundamental frequency (characteristic of a female speaker). These stimuli sounded essentially like tokens produced by a male with a high-pitched (or falsetto) voice. The third vowel continuum (series C) was designed to represent a speaker with a shorter vocal tract than series A and B (as typically found in female speakers) with a relatively high fundamental frequency. Series C thus had the same $F0$ contour as did series B, but with higher overall formant frequencies. The frequencies of the first three formants of the series C continuum are shown in Table II. The formant frequencies chosen for the endpoints of this continuum were based on measurements obtained by the author in a different study (Fox, 1982) and modified until the required phonentic qualities were obtained. The formant transitions to the final [d] were 30 ms in duration and went from the formant values specified in Table II to 400, 2000, and 2900 Hz for $F1$, $F2$, and $F3$, respectively. The three [ɪ]-[ɛ] vowel continua constructed thus differed in terms of $F0$, absolute formant frequencies, or both, while representing the same range of phonetic qualities as determined by the experimenter.

There were six anchors used in this experiment. Anchors A1 and A7 represented 1 and 7 of series A, respectively; anchors B1 and B7 represented stimuli 1 and 7 of series B, respectively; and anchors C1 and C7 represented stimuli 1 and 7 of series C, respectively.

TABLE I. First, second, and third formant frequencies for the vowel stimuli in the series A and B continua.

Stimulus	$F1$	$F2$	$F3$
1	400	1800	2570
2	422	1780	2558
3	443	1760	2547
4	465	1740	2535
5	487	1720	2523
6	508	1700	2511
7	530	1680	2500

TABLE II. First, second, and third formant frequencies for the vowel stimuli in the series C continuum.

Stimulus	$F1$	$F2$	$F3$
1	400	2300	3000
2	406	2297	2995
3	417	2294	2989
4	434	2292	2984
5	585	2289	2978
6	618	2286	2973
7	650	2283	2889

All stimuli were converted into analog form and recorded on seven test tapes for each of the three continua. One of the tapes served to determine the baseline identifications of the vowels in each of the continua. On these tapes the seven stimuli from a continuum occurred 20 times in random order. The six anchor tapes for each continuum contained 60 presentations of an anchor plus 20 presentations of the seven test stimuli. Each anchor tape thus contained 200 stimuli. The order of stimuli on the tapes was constrained such that no single stimulus occurred more than three times in succession. All tapes were recorded with a 4-s interstimulus interval.

C. Procedure

The subjects were randomly divided into 18 groups of eight subjects each. Each subject was run individually in one 45-min session. The stimulus tapes were reproduced on a high-quality stereo cassette tape deck (BIC T-2M) and presented binaurally to subjects in an anechoic chamber at a comfortable listening level via Sennheiser HD 420 headphones. Each group of subjects listened first to the control tape and then to one of the six anchor tapes. The subjects were informed that they would be listening to speech tokens that would sound either like *hid* or *head*. They were asked to indicate their identification response in prepared booklets by circling the appropriate word on the answer sheets. In addition, each subject was required to rate each identification in terms of how confident they were that the response was correct. A 4-point scale was used with a 1 indicating that when the subject was positive the identification response was correct, a 2 indicating the response was probably correct, a 3 indicating the response was possibly correct, and a 4 indicating that the response was a guess. There was a 3-min break between the control tape and the anchor tape but no new instructions were given concerning any differences between the two tapes.

D. Results

The identification plus rating responses were converted into an 8-point scale (cf. Sawusch and Nusbaum, 1979) with 1 indicating a very positive *hid* response and 8 representing a very positive *head* response. Category boundaries were then determined by linear interpolation between points on either side of the boundary in both control and anchor conditions.

The results for the [ɪ] anchor groups are shown in Fig. 1, while the relevant mean category boundaries appear in Table III. It is clear from an inspection of the rating functions that in the anchor condition category boundaries are shifted toward the continuum endpoint corresponding to the anchor stimulus. However, category boundaries seem to be differentially shifted depending on the degree of discrepancy between the anchor stimulus and the test continuum. This was confirmed by analyzing the boundaries using a $3 \times 3 \times 2$ repeated-measures analysis of variance (Winer, 1971) with the between-group factors Test Series and Anchor Series, and the within-group factor of Anchoring (control versus anchor condition). The analysis revealed a very significant main effect of Anchoring, $F(1,63) = 140.92$, $p < 0.001$, and a significant effect of Test Series,

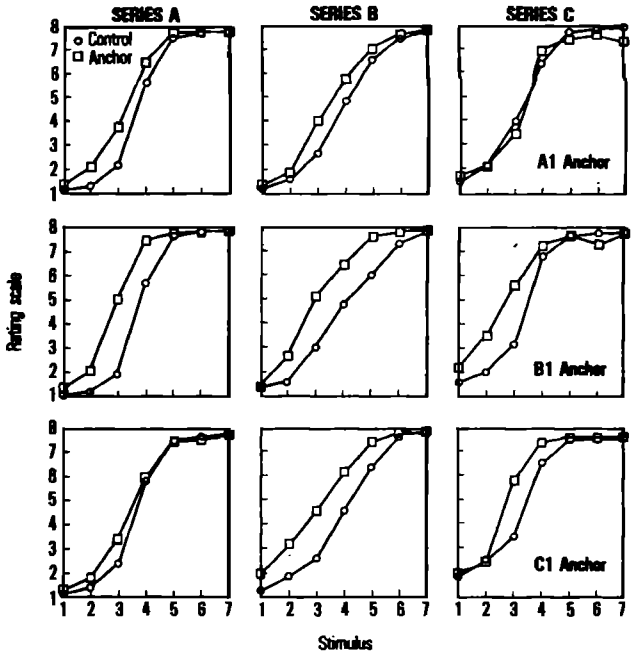

FIG. 1. Baseline and [ɪ]-anchored rating functions for the A1, B1, and C1 anchors using the series A, B, and C continua (from left to right, respectively).

$F(2,63) = 4.21$, $p < 0.02$, but no significant effect due to Anchor Series, $F(2,163) = 1.44$ $p > 0.24$. The Test Series means were compared using a Duncan multiple range test which showed that test series A and B differed from series C at the 0.05 level. The significant test series effect only indicates that the boundaries from the series C continuum, even in the control condition, were shifted toward the [ɪ] endpoint relative to the series A and B continua. There were significant interactions between Anchoring and Anchor Series, $F(2,63) = 7.58$, $p < 0.002$, and between all three factors, $F(2,63) = 2.99$, $p < 0.025$. This confirms that the amount of category shift was determined, in part, by the amount of discrepancy between the anchor and the baseline continuum. The interactions between Test Series and Anchor Series, $F(4,63) = 0.76$, $p > 0.56$, and between Anchoring and Test Series, $F(2,63) = 3.11$, $p > 0.05$, were nonsignificant.

The results for the [ɛ] anchor groups are shown in Fig. 2, while the relevant mean category boundaries appear in

TABLE III. Mean phoneme boundaries for control and [ɪ] anchor conditions.

Condition	[ɪ] Anchors A1	B1	C1	Mean
			Series A	
Control	3.67	3.70	3.69	3.69
Anchor	3.10	2.78	3.44	3.11
			Series B	
Control	3.80	3.87	3.99	3.89
Anchor	3.35	2.87	2.80	3.01
			Series C	
Control	3.42	3.39	3.23	3.35
Anchor	3.24	2.49	2.62	2.78
Mean	3.43	3.18	3.29	3.30

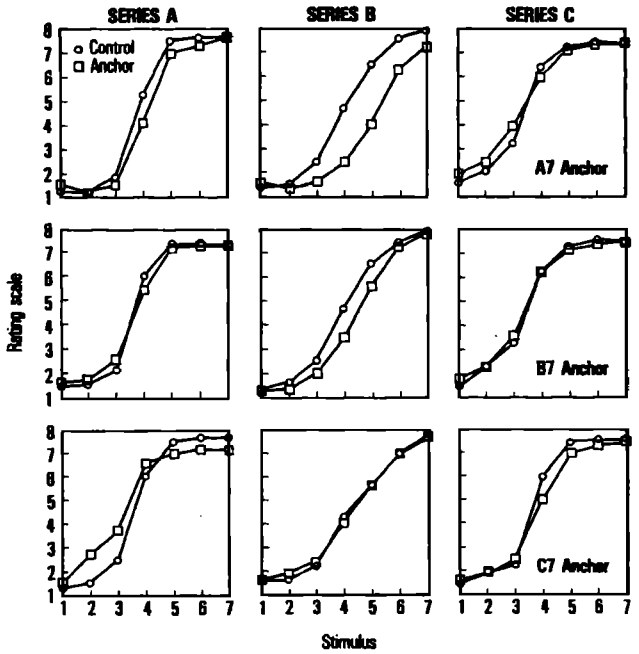

FIG. 2. Baseline and [ɛ]-anchored rating functions for the A7, B7, and C7 anchors using the series A, B, and C continua (from left to right, respectively).

Table IV. Inspection of the rating functions again reveals that differential shifts of the category boundary are obtained depending upon the degree of discrepancy between the anchor and test continuum. These category boundaries were, as above, analyzed using a $3 \times 3 \times 2$ repeated-measures analysis of variance. This analysis revealed significant main effects of both Anchoring, $F(1,63) = 17.83$, $p < 0.001$, and Test Series, $F(2,63) = 18.29$, $p < 0.001$, but no significant effect due to Anchor Series, $F(2,63) = 1.16$, $p > 0.32$. Duncan's multiple range test showed that, within the Test Series factor, series A, series B, and series C differed significantly between themselves at the 0.05 level. There were significant interactions between Anchoring and Test Series, $F(2,63) = 6.52$, $p < 0.003$, between Anchoring and Anchor Series, $F(2,63) = 16.04$, $p < 0.001$, and between all three factors, $F(4,63) = 4.55$, $p < 0.003$. These interaction effects again arise from the differential ability of a cross-series anchor to shift the category boundary of the test continuum.

The main effect of Anchoring with the [ɛ] anchors was not nearly as large as with the [ɪ] anchors. This indicates an asymmetry in the anchoring results, similar to that found in Sawusch *et al.* (1980) and Crowder and Repp (1984). In particular, the [ɪ] anchor seems to be more effective in producing a category boundary shift than [ɛ]. However, the category boundary data, as analyzed above, do not lend themselves

TABLE IV. Mean phoneme boundaries for control and [ɛ] anchor conditions.

Condition	[ɛ] Anchors A7	B7	C7	Mean
		Series A		
Control	3.81	3.62	3.53	3.65
Anchor	4.12	3.64	3.42	3.73
		Series B		
Control	4.04	4.02	4.02	4.03
Anchor	5.24	4.50	4.14	4.63
		Series C		
Control	3.31	3.41	3.65	3.46
Anchor	3.18	3.32	3.79	3.43
Mean	3.95	3.75	3.76	3.82

easily to an evaluation of this asymmetry. To directly address this issue, the magnitudes of boundary shifts were calculated by subtracting a subject's anchor condition boundary from his/her control condition. Since we know that, in general, the boundary shifts obtained with the [ɛ] anchors will be in the opposite direction to the [ɪ] anchors, and that it is the *extent* of shift rather than *direction* that interests us in checking for asymmetry, the signs of the [ɛ] anchor boundary shifts were reversed. These mean boundary shifts are shown in Table V.

These boundary shifts were analyzed using a $3 \times 3 \times 2$ analysis of variance (with no repeated measures) with the factors Test Series, Anchor Series, and Anchor Vowel. This analysis revealed a significant effect due to Anchor Series, $F(2,126) = 16.48$, $p < 0.001$, and Anchor Vowel, $F(1,126) = 32.38$, $p < 0.001$, but no significant effect due to Test Series, $F(2,216) = 1.19$, $p > 0.30$. The significant effect of Anchor Vowel supports the claim that [ɪ] was more effective in producing shifts than [ɛ]. In addition, Duncan's multiple range test showed that boundary shifts for series B (mean shift = 0.737) were significantly greater (at the 0.05 level) than those for series A (mean shift = 0.314) or series C (mean shift = 0.262), while the latter two were not significantly different. This result stems from the fact that series B was more susceptible to cross-series anchoring effects. There were significant interactions between Test Series and Anchor Series, $F(4,126) = 3.76$, $p < 0.006$, Test Series and Anchor Vowel, $F(2,126) = 11.33$, $p < 0.001$, and between all three factors, $F(4,126) = 5.80$, $p < 0.001$. There were no significant interactions between Anchor Series and Anchor Vowel, $F(2,126) = 1.10$, $p > 0.33$.

The results of this experiment can be briefly summarized as follows:

TABLE V. Mean category boundary shifts. The signs for the [ɛ] anchors have not been reversed in this table. Asterisks indicate levels of significance as determined by t tests (two tailed): * = <0.05; ** = <0.01; *** = <0.001.

Test continuum	Anchor A1	B1	C1	A7	B7	C7
Series A	−0.58*	−0.94***	−0.25	0.31**	0.06	0.16
Series B	−0.46**	−1.00***	−1.18**	1.20***	0.48**	0.12
Series C	−0.18	−0.78**	−0.50**	0.13	−0.09	0.23

(1) In the within-series anchoring conditions, significant boundary shifts were obtained with both [ɪ] and [ɛ] anchors, with the [ɪ] anchors producing larger shifts than the [ɛ] anchors.

(2) In the cross-series anchoring conditions with [ɪ], significant boundary shifts were obtained only when the anchor differed from the test series in terms of either $F0$ or formant frequencies alone, but not both.

(3) Cross-series anchoring with [ɛ] produced no significant boundary shifts except for a very large one of A7 on test series B.

E. Discussion

The discovery of asymmetrical contrast is consistent with both Sawusch *et al.* (1980) and Crowder and Repp (1984), although there is no complete explanation for such effects. Sawusch *et al.* (1980) found that an [ɪ] anchor produced a significantly smaller category shift than did an [i] anchor when both the anchors and the target continuum were placed in a CVC context. Sawusch *et al.* suggested that these results might indicate that anchoring with the [ɪ] vowel is mediated by auditory memory, whereas anchoring with the [i] vowel (a "point" vowel) is related to early perceptual processing, possibly the retuning of a vowel prototype from long-term memory. The CVC context would reduce the available auditory memory for vowel information and thus reduce the contrast effect for [ɪ]. Reduction of auditory memory for [i] would have no such effect on contrast stemming from early perceptual processing. Crowder and Repp (1984, experiment I) obtained similar results utilizing a paired-vowel identification task, except that they obtained no contrast at all from the [ɪ] anchor. Such results are unexplained by Crowder's (1978,1981) theory, but, as we shall see, neither is the explanation suggested by Sawusch *et al.* entirely adequate.

Crowder and Repp (1984, experiment II) used vowels from an [ɛ]-[æ] continuum and examined contrast effects to the [ɛ] and [æ] endpoints. They obtained significant contrast from the [ɛ] direction, but none from the [æ] direction. It is difficult to account for this asymmetry in terms of the prototype retuning hypothesis (Sawusch *et al.*, 1980), primarily because there is no reason to expect that anchoring with [ɛ] as opposed to [æ] results from early perceptual processing. Both vowels are nonhigh, front, lax vowels and do not represent one of the three point vowels ([i], [a], [u]).

The present experiment, utilizing the anchors [ɪ] and [ɛ] represents the completion of the possible adjacent pairings of the vowels [i]-[ɪ]-[ɛ]-[æ] in the context of a vowel contrast experiment. Here we have found that the [ɪ] anchor produces significantly greater category shifts than the [ɛ] anchor. Again, it would appear that the hypothesis suggested by Sawusch *et al.* (1980) will not account for this contrast asymmetry. Crowder (personal communication) has suggested the generalization that contrast effects tend to be larger when the context has a lower absolute formant frequency than the target as opposed to the case in which the target has the lower value. However, as Crowder and Repp (1984) point out, it is too early to propose any general hypothesis concerning the asymmetrical contrast results, and more relevant experimentation needs to be conducted. Since the main thrust of this paper concerns the efficacy of cross-series contrast, it is to this topic that we now turn.

The shifts in the [ɪ]-[ɛ] category boundaries were, in general, consistent with the results of both Sawusch and Nusbaum (1979) and Morse *et al.* (1976) with the within-series anchoring condition representing essentially a replication of those studies. The pattern of within-series anchors is explicable in terms of either phonetic and/or auditory contrast. However, the results with cross-series anchors do *not* support any simple auditory or phonetic contrast explanation. Rather, the pattern of results obtained strongly suggests that the ability of an anchor to shift the category boundary of a test series depends not only upon phonetic quality and/or spectral overlap, but a complex combination of the two.

For example, the A1 and A7 anchors differ from the B1 and B7 anchors in terms of $F0$ alone, yet this variation produces a significant difference in the category shift obtained with both the Test Series A and B continua. In particular, the B1 anchor shifted the [ɪ]-[ɛ] category boundary more toward the [ɪ] endpoint than did the A1 anchor for both test series A, $t(14) = 2.87, p < 0.02$, and test series B, $t(14) = 2.86, p < 0.02$ (all between-group comparisons were 2-tailed, uncorrelated t-tests). Similarly, the A7 anchor shifted the boundary more toward the [ɛ] endpoint than did the B7 anchor with both test series, B, $t(14) = 3.91$, $p < 0.002$ and A, $t(14) = 2.29$, $p < 0.038$. It is thus clear that the anchoring (contrast) effect cannot be explained on the basis of formant-frequency specific inhibition alone since both A1 and B1, on the one hand, and A7 and B7, on the other hand, have the same formant frequencies. In addition, the hypothesis of strict separation of distinct input channels is not supported since it would predict larger contrast effects of A anchors on the series A test tokens than on series B, and vice versa; this is not the pattern of results obtained.

As several studies (Miller, 1953; Slawson, 1968; Fujisaki and Kawashima, 1968) have shown, a given formant pattern may be perceived as representing different vowel qualities depending on its fundamental frequency. This difference is often explained by suggesting that listeners rescale or normalize the formant structure of a vowel to eliminate speaker quality variation (such as that due to vocal-tract length differences). Since $F0$ is highly correlated (negatively) with vocal tract length (at least across sexes), if a listener utilizes $F0$ to estimate vocal-tract length and then rescales the formant frequencies on the basis of this estimate, the perceived phonetic quality of a given formant pattern will covary with $F0$, at least within a limited range of vowel qualities. If we assume that this occurs before the process(es) responsible for the contrast effect, we can suggest that the vowel in the B1 anchor, for example, is perceived as a more prototypical version of [ɪ] than the A1 anchor, that is, its rescaled first formant would be lower. This would result in having the category shifts produced by the B1 anchor greater than that for the A1 anchor if the contrast effects were *phonetic* (rather than purely auditory) in nature. A similar argument can be constructed for the B7 anchor representing a less prototypical version of the [ɛ] anchor.

By this line of argument (i.e., that the contrast effects are essentially *phonetic* in nature) we should then expect that the C1 and C7 anchors might produce category boundary shifts similar to those found for the B1 and B7 anchors, but such is not the case. For example, neither the C1 anchor, $t(7) = 1.22, p > 0.25$, nor the C7 anchor, $t(7) = 0.92, p > 0.38$, produced a significant boundary shift with the series A continuum. This fact would seem to rule out a strictly phonetic basis for the contrast effect. One might be tempted to suggest that if there is no spectral overlap (in terms of formant-frequency) between the anchor and the test series no contrast effect will occur on either a phonetic or auditory level. However, this will not adequately account for the test series B and C data. For example, there was no significant difference between the B1 and C1 anchors in terms of amount of boundary shift for either test series B, $t(14) = 0.95, p > 0.30$, or test series C, $t(14) = 0.93, p > 0.35$.

II. GENERAL DISCUSSION

The present results strongly suggest that the process(es) responsible for vowel contrast effects are neither channel specific, frequency specific, nor necessarily occur only at a very early level of processing. In particular, we have shown that contrast effects may extend across at least moderately discrepant physical channels. In addition, even across these different channels, contrast effects can occur utilizing anchors with a phonetic quality similar to one endpoint of the test continuum but which does not overlap spectrally with that endpoint.

At least some of the results are consistent with the hypothesis that the contrast effects take place after a process of speaker normalization (Slawson, 1968). For example, one could explain the differential shifts produced by the A and B anchors by suggesting that the vowels were normalized (by reference to $F0$; see Fujisaki and Kawashima, 1968; Slawson, 1968) before lateral inhibition (or some other contrast-generating process) occurred. However, this suggestion does not readily explain why the A1 and C1 anchors do not produce significant boundary shifts in the series C and series A test stimuli, respectively. It seems clear that no appeal to a single level of simple auditory or phonetic processing will account for these and related contrast effects found in the literature. Rather, vowel contrast must arise at several different levels of speech representation.

This view has been supported at least partially by the recent work of Sawusch and Jusczyk (1981) and Sawusch and Nusbaum (1983) on consonant contrast effects. Sawusch and Jusczyk (1981) constructed an eight-step [ba]-[p^ha] VOT test continuum and a [spa] adaptor. The [pa] portion of the [spa] adaptor was acoustically identical to the [ba] end of the VOT continuum but was phonemically identical to the [p^ha] end of the continuum. Using a selective adaptation procedure, Sawusch and Jusczyk found that the [spa] adaptor had the same effect as did the [ba] stimulus when used as an adaptor. These results suggest that the contrast effects produced by selective adaptation are *auditory* in nature. Sawusch and Jusczyk also used a paired-comparison procedure (Diehl *et al.*, 1978, 1980) which paired the [spa] adaptor with a relatively ambiguous test item. This experiment produced a contrast effect in the opposite direction from that found during selective adaptation, that is, the phonemic status of the bilabial stop in the token [spa] governed the direction of the contrast effect. These data suggest that the contrast effect (at least for the paired-comparison task) is *phonetic* in origin. Sawusch and Nusbaum (1983) obtained a similar pattern of results using a [da]-[ga] place of articulation test continuum and a [ska] adaptor whose spectral characteristics (i.e., formant transitions) matched that of the [da] endpoint (cf. Mann and Repp, 1981). Sawusch and Nusbaum (1983) argue that these results can be explained only with reference to at least two distinct levels of processing: auditory and phonetic. Furthermore, the work of Sawusch (1977) suggests that there are at least two different levels of auditory processing involved in a selective adaptation of consonants: a spectrally specific level of auditory analysis which is monaurally driven and possibly peripheral in nature, and an integrative level of auditory processing which is binaurally driven and central.

Sawusch (1977) suggested the possibility that the central integrative level of processing may not integrate information over all possible frequencies, but may represent an intermediate level of integration that would allow for vocal tract length differences. Such an integrative level could thus correspond to a process of "speaker normalization" and could explain many of the results presented here. Note, however, that the pattern of results in Sawusch (1977) is really quite different from those found here. In particular, Sawusch found that "high" adaptors (which differed from "low" adaptors and the test series in terms of absolute formant frequencies only) produced 33%–40% of the contrast effects produced by the low adaptors. By contrast, our study demonstrated no significant difference in mean category boundary shifts between the B1 and C1 anchor groups on the series B and series C test stimuli, the conditions most similar to those found in Sawusch (1977). We will argue that these results demonstrate that contrast effects are occurring only at the intermediate level of analysis—a level which could conceivably be represented by Crowder's auditory memory.

In particular, the B1 and C1 anchors do not overlap spectrally in terms of formant frequency; therefore, we do not expect the contrast effects to arise from the peripheral auditory level. Since we do not find any contrast effects using the A1 and C1 anchors with the series C and series A stimuli, respectively, we must assume that the contrast effects are *not* taking place at a more abstract phonetic level. This study did find a significant difference between the A1 and B1 anchor groups using the series A and series B test stimuli, but these can also be explained in terms of contrast at an integrative level alone. If one assumes that the intermediate integrative level is normalizing (erroneously) the anchor differing from the test series in terms of fundamental frequency (partially on the basis of $F0$), then the obtained contrast effects could arise from the intermediate level alone.

There is at least one source of difficulty in accepting this explanation: Why should there *not* be normalization and subsequent contrast when the test stimuli differ from the anchors in terms of *both* formant frequencies and $F0$? Perhaps such radical normalization need not occur when the

two sets of stimuli are very discrepant in terms of physical channel, or perhaps the *normalized* values are mapped into an auditory similar to that suggested by Crowder (1981) and these two physical channels are too discrepant mutually to inhibit one another.

Clearly, more experimentation is needed to understand the vowel contrast phenomenon, and to evaluate proposed models of auditory and phonetic processing. For example, to test directly the proposal that the vowel contrast effects occur at an intermediate level of processing, we need to discover the extent to which monotic versus diotic anchoring would produce differential boundary shifts. In addition, one could vary the fundamental frequency and formant pattern differences between the anchors and test tokens in a more gradual fashion and discover the extent to which the contrast effects covary in a linear fashion. Research along these and other lines will further our understanding of the number and nature of the mechanisms underlying vowel perception.

ACKNOWLEDGMENTS

The author greatly appreciates the financial support of the Graduate School of The Ohio State University. The author is also indebted to Ilse Lehiste and Lida Wall for comments on an earlier version of this manuscript as well as to Bruno Repp, Robert Crowder, and an anonymous reviewer for suggestions on manuscript revisions. Results using the A1, B1, and C1 anchors were reported at a meeting of the Acoustical Society of America in Cincinnati OH, 6–10 May 1983.

Ades, A. E. (1976). "Adapting the property detectors for speech perception," in *New Approaches to Language Mechanisms*, edited by R. J. Wales and E. Walker (North-Holland, Amsterdam).

Cooper, W. E. (1974). "Adaptation of phonetic feature analyzers for place of articulation," J. Acoust. Soc. Am. **56**, 617–627.

Cooper, W. E. (1975). "Selective adaptation to speech," in *Cognitive Theory*, edited by R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (Erlbaum, Hillsdale, NJ), Vol. 1.

Cooper, W. E., Ebert, R. R., and Cole, R. A. (1976). "Perceptual analysis of stop consonants and glides," J. Exp. Psychol. Human Percept. Perform. **2**, 92–104.

Cornsweet, T. N. (1970). *Visual Perception* (Academic, New York).

Crowder, R. G. (1971). "The sound of vowels and consonants in immediate memory," Verb. Learn. Verb. Behav. **10**, 587–596.

Crowder, R. G. (1973). "Representation of speech sounds in precategorical acoustic storage," Verb. Learn. Verb. Behav. **10**, 587–590.

Crowder, R. G. (1978). "Mechanisms of auditory backward masking in the stimulus suffix effect," Psychol. Rev. **85**, 502–524.

Crowder, R. G. (1981). "The role of auditory memory in speech perception and discrimination," in *The Cognitive Representation of Speech*, edited by T. Myers, J. Laver, and J. Anderson (North-Holland, Amsterdam, The Netherlands).

Crowder, R. G. (1982). "Decay of auditory memory in vowel discrimination," J. Exp. Psychol. Learn. Mem. Cog. **8**, 153–162.

Crowder, R. G., and Repp. B. H. (1984). "Single formant contrast in vowel identification," Percept. Psychophys. **35**, 372–378.

Diehl, R. L. (1975). "The effects of selective adaptation on the identification of speech sounds," Percept. Psychophys. **17**, 48–52.

Diehl, R. L., Elman, J. L., and McCusker, S. B. (1978). "Contrast effects in stop consonant identification," J. Exp. Psychol. Human Percept. Perform. **4**, 599–609.

Diehl, R. L., Lang, M., and Parker, E. M. (1980). "A further parallel between adaptation and contrast," J. Exp. Psychol. Human Percept. Perform. **6**, 24–44.

Eimas, P. D. (1963). "The relationship between identification and discrimination along speech and nonspeech continua," Lang. Speech **6**, 206–217.

Eimas, P. D., and Corbit, J. D. (1973). "Selective adaptation of linguistic feature detectors," Percept. Psychophys. **13**, 247–252.

Eimas, P. D., Cooper, W. E., and Corbit, J. D. (1973). "Some properties of linguistic feature detectors," Cogn. Psychol. **4**, 99–109.

Fimas, P. D., and Miller, J. L. (1978). "Effects of selective adaptation on the perception of speech and visual patterns," in *Perception and Experience*, edited by H. L. Pick and R. D. Walk (Plenum, New York).

Fox, R. A. (1982). "Individual variation in the perception of vowels: implications for a perception-production link," Phonetica **39**, 1–22.

Fry, D., Abramson, A., Eimas, P. D., and Liberman, A. M. (1962). "The identification and discrimination of synthetic vowels," Lang. Speech **5**, 171–189.

Fujisaki, H., and Kawashima, T. (1968). "The roles of pitch and higher formants in the perception of vowels," IEEE Trans. Audio Electroacoust. **AU-16**, 73–77.

Healy, A. F., and Repp, B. H. (1982). "Context independence and phonetic mediation in categorical perception," J. Exp. Psychol. Human Percept. Perform. **8**, 68–80.

Helson, H. (1964). *Adaptation Level Theory* (Harper and Row, New York).

Helson, H. (1971). "Adaptation-Level Theory: 1970—and after," in *Adaptation-Level Theory*, edited by M. H. Appley (Academic, New York).

Ladefoged, P., and Broadbent, D. E. (1957). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Mann, V. A., and Repp, B. H. (1981). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

Miller, R. L. (1953). "Audiology tests with synthetic vowels," J. Acoust. Soc. Am. **25**, 117.

Morse, P., Kass, J. E., and Turkienicz, H. (1976). "Selective adaptation of vowels," Percept. Psychophys. **19**, 137–143.

Parducci, A. (1963). "Range-frequency compromise in judgment," Psychol. Monogr. **77**, 1–50.

Parducci, A. (1965). "Category judgment: A range-frequency model," Psychol. Rev. **72**, 407–418.

Parducci, A. (1975). "Contextual effects: A range-frequency analysis," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. II.

Repp, B. H., Healy, A. F., and Crowder, R. G. (1979). "Categories and context in the perception of isolated steady-state vowels," J. Exp. Psychol.: Human Percept. Perform. **5**, 129–145.

Restle, F. (1978). "Assimilation predicted by adaptation-level theory with variable weights," in *Cognitive Theory*, edited by N. J. Castellan and F. Restle (Erlbaum, Hillsdale, NJ), Vol. 3.

Sawusch, J. R., (1977). "Peripheral and central processes in selective adaptation of place of articulation in stop consonants," J. Acoust. Soc. Am. **62**, 738–750.

Sawusch, J. R., and Jusczyk, P. (1981). "Adaptation and contrast in the perception of voicing," J. Exp. Psychol. Human Percept. Perform. **7**, 408–421.

Sawusch, J. R., and Nusbaum, H. C. (1979). "Contextual effects in vowel perception I: Anchor-induced contrast effects," Percept. Psychophys. **25**, 292–302.

Sawusch, J. R., and Nusbaum, H. C. (1983). "Auditory and phonetic processes in place perception for stops," Percept. Psychophys. **34**, 560–568.

Sawusch, J. R., Nusbaum, H. C., and Schwab, E. C. (1980). "Contextual effects in vowel perception II: Evidence for two processing mechanisms" Percept. Psychophys. **27**, 421–434.

Simon, H. J., and Studdert-Kennedy, M. (1978). "Selective adaptation and anchoring of phonetic and nonphonetic continua," J. Acoust. Soc. Am. **43**, 1338–1357.

Slawson, A. W. (1968). "Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency," J. Acoust. Soc. Am. **43**, 87–101.

Thompson, C. L., and Hollien, H. (1970). "Some contextual effects on the perception of synthetic vowels," Lang. Speech **13**, 1–13.

Winer, B. J. (1971). *Statistical Principles in Experimental Design* (McGraw-Hill, New York).