

# Higher formant normalization results from auditory integration of F2 and F3

KEITH JOHNSON

*Indiana University, Bloomington, Indiana*

Two experiments were conducted to test the hypothesis that higher formant normalization results from the auditory integration of F2 and F3 when they are within 3 Bark of each other. In the first experiment, Formants 3-5 were manipulated in both a "hid"- "head" continuum (in which F2 and F3 are within 3 Bark of each other) and a "hood"- "HUD" continuum (in which F2 and F3 are not within 3 Bark of each other). It was found that there was a shift in identification consistent with the higher formant normalization effect only in the "hid"- "head" continuum. In the second experiment, F3 alone was manipulated in a "hood"- "HUD" continuum. The amplitude of F3 in this continuum was increased (as compared with the F3 in the "hood"- "HUD" continuum used in Experiment 1) and a pretest indicated that the shift in F3 could be detected. As in the first experiment, there was no shift in identification associated with shifting F3 frequency in a back-vowel continuum. The results of these experiments are not consistent with an explanation of higher formant normalization in which hearers adjust an internal vowel space in response to higher formant information; rather, the present findings indicate that higher formant normalization results from auditory integration of F2 and F3.

Fujisaki and Kawashima (1968) investigated the role of higher formants in vowel perception and found that shifts in Formants 3-5 can produce boundary shifts in the identification of vowel continua. This effect was greatest in a [u]-[e] continuum and was not so large in a continuum from [o] to [a]. They also found that by using a non-periodic noise source (with which the higher formants had greater amplitude), the effect of shifting the higher formants in the [o]-[a] continuum was increased. The largest shift of vowel-identification boundaries occurred when both higher formants and F0 were manipulated.

The shift in vowel identification that was found by Fujisaki and Kawashima (1968) is consistent with the hypothesis that hearers use the higher formants as an indication of the length of a speaker's vocal tract and that hearers then (using this estimate of the speaker's vocal-tract length) rescale an internal vowel space which is used in the perceptual identification of vowels.<sup>1</sup>

However, not all researchers who have manipulated the higher formants have found the same effect. Slawson (1968), using a scaling method of testing for vowel-perception change, found that variation in higher formants alone did not result in a shift of vowel perception. He concluded that although higher formants correlate well with vocal-tract length, they are not used by hearers to adjust a perceptual vowel space. Karnickaya, Mushnikov, Slepokurova, and Zhukov (1975) also failed to observe any effect of shifting F3 and F4 in either an [o]-[a] or an [e]-[a] continuum.

An explanation of this divergence of experimental findings was suggested by Carlson, Fant, and Granström (1975). They found that when F3 is within 3 Bark of F2, the perceived F2 (F2') is determined by both F2 and F3. (Perceived F2 in their experiment was measured by having their subjects match a two-formant stimulus to a five-formant stimulus.) It may be that in the studies of Slawson (1968) and Karnickaya et al. (1975), the distance between F2 and F3 was large enough that F3 did not have an effect on the perceived F2, whereas in Fujisaki and Kawashima's (1968) study, the F3 and F2 of the [e] tokens were close enough to result in auditory integration to a single "center of gravity" (see Chistovich, Sheikin, and Lublinskaja, 1979).

Thus, rather than being used to estimate vocal-tract length in a speaker-normalization process, it may be that the role of higher formants in vowel perception is purely auditory in nature—having to do with the integration of spectral information in the speech signal in the process of estimating spectral peaks. If we assume a spectral integration function similar to that proposed by Chistovich et al. (1979), we can predict that higher formant normalization will occur only when the distance between F2 and F3 is less than about 3 Bark.

The experiments reported here were conducted as a test of this hypothesis. If we find that there is a shift in identification functions as a result of a shift in higher formants in a vowel continuum in which F3 is within 3 Bark of F2, and that there is not such a shift in identification in a vowel continuum in which F3 is not within 3 Bark of F2, then the hypothesis that higher formant normalization results from the auditory integration of F2 and F3 will be supported. Such a result also rules out an interpretation of the "higher formant normalization" effect that re-

This research was supported by the Linguistics Department at Ohio State University and appeared as Chapter 3 in my 1988 doctoral dissertation. Many thanks to my doctoral dissertation members: Mary Beckman, Rob Fox, Ilse Lehiste, and Neal Johnson.

lies on a cognitive-level process in which hearers adjust an internal vowel space on the basis of their calculation of the length of the speaker's vocal tract.

### EXPERIMENT 1

Two continua were synthesized: [hɔd]–[hʌd], and [hɪd]–[hɛd]. In the [ɪ]–[e] continuum, F3–F2 is less than 3 Bark for both men and women in the vowel-production data reported by Peterson and Barney (1952), whereas in the [hɔd]–[hʌd] continuum, F3–F2 is greater than 3 Bark for both men and women.

The auditory explanation of higher formant normalization predicts that there will be a shift in identification only in the front series (due to the shift of F2' as a function of the frequencies of F2 and F3 when F3–F2 is less than 3 Bark). If higher formants are used to estimate a speakers' vocal-tract length in a cognitive-normalization process, we expect to find that higher formant normalization will not be limited to the front series but will be present in the back series as well.

### Method

**Subjects.** Ten native English-speaking undergraduate students (2 males and 8 females) at Ohio State University served as subjects in this experiment. None of the subjects reported any history of speech or hearing disorders. They were recruited through the university newspaper and through introductory linguistics classes, and received \$3.50 for participating in a single 1-h session.

**Materials.** Four [hVd] continua were synthesized. Two of these continua were front-vowel continua from "hid" to "head" and two were back-vowel continua from "hood" to "HUD" (the abbreviation of Housing and Urban Development). One front-vowel continuum and one back-vowel continuum were synthesized with relatively high higher formants, whereas the other front and back series had relatively low higher formants (for ease of reference, the higher formants in these five-formant tokens will be referred to as F3–F5).

The values chosen for F1 and F2 were ambiguous between the values for men and women in Peterson and Barney's (1952) data.

The formants and bandwidths for all four vowel continua are listed in Table 1. The /h/ portion of the tokens was 50 msec in duration. Each token had a vocalic portion of 220 msec, which included a 30-msec transition to /d/. The tokens had a total duration of 270 msec. Alveolar transitions for the "hid"–"head" continuum ended at 300, 1600, and 2600 Hz, respectively, for the first, second, and third formants. Transitions for the "hood"–"HUD" series went to 300 (F1), 1600 (F2), and dipped to 2150 then rose to 2570 Hz (F3). All tokens had F0 which was steady-state at 200 Hz for 115 msec and then fell linearly over 100 msec to 160 Hz. The voice source for all of these tokens was given a gradual spectral tilt such that frequency components above 3 kHz were attenuated by 28 dB relative to the default voice spectrum of the synthesizer. The subjects responded to 15 trials for each token. The total number of observations per subject was 420 (seven tokens × two higher formant levels × two continua × 15 repetitions).

**Procedure.** Experimental sessions were conducted online using a New England Digital Able 60 computer at Ohio State University. The subjects were seated in an anechoic chamber and the tokens were played through a 4.5-kHz low-pass filter and Sennheiser HD-420 headphones at a comfortable listening level.

The subjects indicated their response by pressing buttons on a response box labeled "hood" and "HUD" or "hid" and "head." The test items were randomized separately for each subject. Vowel-identification response and reaction-time data were collected by the computer, and the subjects were given reaction-time feedback via a computer screen for each item presented. Only the identification data were analyzed.

The subjects were instructed to identify the test tokens as either "hood" or "HUD," or as either "hid" or "head," and to disregard all other differences between the tokens.

The independent variables in the experiment were the level of the higher formants and front- versus back-vowel continuum. Both were treated as within-subjects variables. The vowel-continuum variable was blocked and the order of presentation was counterbalanced among subjects (half of the subjects heard the front-vowel continua first; half heard the back-vowel continua first). There were 15 practice trials for both vowel series. The dependent variable was the number of "hood" or "hid" responses.

Table 1  
Formant Values (and Bandwidths) for Tokens in Experiment 1

	Token							Bandwidth
	1	2	3	4	5	6	7	
	"hid"–"head"							
F1	432	448	464	480	496	512	528	55
F2	2208	2192	2176	2160	2144	2128	2112	95
High F3	2845	same	→					170
F4	3700	same	→					300
F5	4550	same	→					300
Low F3	2701	same	→					170
F4	3300	same	→					300
F5	3850	same	→					300
	"hood"–"HUD"							
F1	474	491	509	526	543	561	578	110
F2	1111	1124	1137	1150	1163	1176	1189	75
High F3	2512	same	→					110
F4	3700	same	→					300
F5	4550	same	→					300
Low F3	2368	same	→					110
F4	3300	same	→					300
F5	3850	same	→					300

## Results

The number of "hood" or "hid" responses to each continuum by the 10 subjects was analyzed in a 2×2 repeated-measures ANOVA. The results for Experiment 1 are presented in Figure 1.

As the figure indicates, there was an interaction between the continuum condition ([ɪ]-[ε] vs. [ω]-[ʌ]) and higher formants condition (low vs. high F3-F5) [ $F(1,9) = 9.67, p < .02$ ]. For the "hood"- "HUD" continuum, the percentage of "hood" responses for low F3-F5 was almost identical to the percentage of "hood" responses for high F3-F5 (52 and 51.9, respectively). On the other hand, for high F3-F5 in the "hid"- "head" continuum, there were 47% "hid" responses, whereas when F3-F5 was low, only 38% of the tokens were identified as "hid." There was a significant main effect for the higher formants condition [ $F(1,9) = 9.68, p < .02$ ]. Vowel identification as "hood" or "hid" was 50% with low F3-F5, and 45% with high F3-F5. There was no main effect for vowel continuum [ $F(1,9) = 4, p = .077$ ].

The difference between the back-vowel continuum and the front-vowel continuum is illustrated further in Figure 2. Panel (a) shows the percentage of "hood" responses for each of the tokens in both the high and the low F3-F5 "hood"- "HUD" continua. Category boundaries in the "hood"- "HUD" continuum for each subject were computed by linear interpolation. There was no difference between the high and the low F3-F5 conditions in a correlated-observations *t* test [ $t(9) = .456, n.s.$ ]. Panel (b) in Figure 2 shows the percentage of "hid" responses for each of the tokens in both higher formant conditions in the "hid"- "head" continua. As is clearly indicated in the figure, there was no effect of higher formant shift in the back-vowel continuum, whereas there was an increased tendency toward "hid" responses throughout the "hid"- "head" continuum in the high F3-F5 condition.

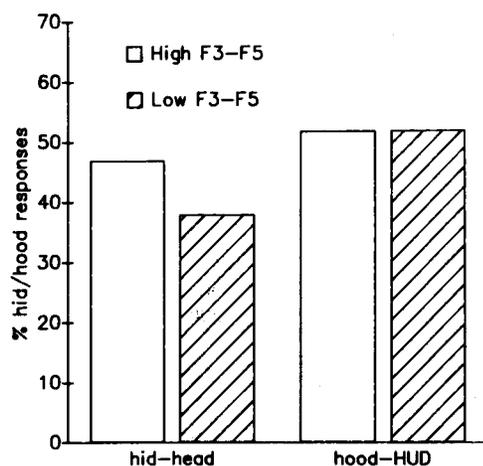


Figure 1. Results of Experiment 1. Identification data are classified by vowel continuum ("hid"- "head" vs. "hood"- "HUD") and higher formant condition (high vs. low).

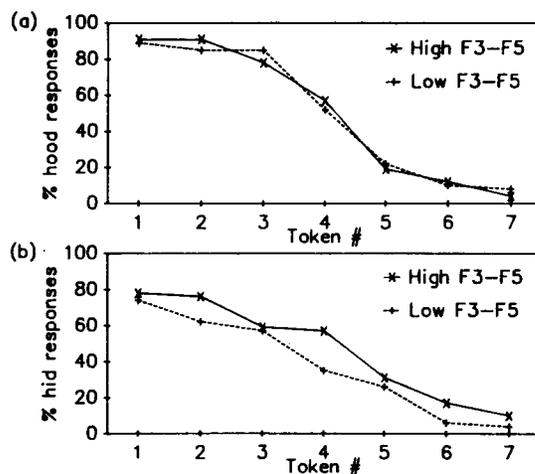


Figure 2. Results of Experiment 1 presented by token. (a) "hood"- "HUD" continua. (b) "hid"- "head" continua.

## Discussion

The results of this experiment seem to support the hypothesis that higher formant normalization is auditory in nature and depends upon the proximity of F2 and F3 in the vowel spectrum. The main effect for higher formants noted above was entirely the result of the strong difference between the high and low F3-F5 in the "hid"- "head" continuum. These data give no reason to suspect that there is a cognitive-level adjustment-to-talker normalization process utilizing higher formants in the perception of vowel formants; rather, they indicate that the influence of higher formants on vowel identification has to do with the role of F3 in the perception of F2 when F2 and F3 are within 3 Bark of each other.

However, one additional consideration motivated a follow-up experiment. Fujisaki and Kawashima (1968) found that amplitude of F3 played a role in the higher formant normalization effect. Nearey (1987) has suggested that back vowels may exhibit less higher formant normalization than do front vowels because the amplitude of the higher formants is lower in vowels with low F2. This provides an alternative interpretation of the results of Experiment 1. It is possible that the shift of F3-F5 did not influence the perception of the "hood"- "HUD" continuum because the higher formants were too low in amplitude for the shift to be detectable. In Experiment 2, this consideration was addressed by increasing the amplitude of F3 in the back-vowel tokens.

## EXPERIMENT 2

In this experiment, the amplitude of the F3 of the "hood"- "HUD" tokens that had been used in Experiment 1 was increased. The results of a preliminary signal-detection test indicated that F3 in the tokens was audible. Thus, the possibility that the lack of a higher formant effect in the back-vowel continuum in Experiment 1 was

due to low F3 amplitude was addressed by the use of this modified continuum. The F3 amplitude manipulation in this experiment ensured that F3 was audible; what is at issue is whether, when F3 is auditorily available, it will be used in a speaker-normalization process.

### Method

**Subjects.** Thirteen native English-speaking undergraduate students (4 males and 9 females) at Ohio State University served as subjects in this experiment. None of the subjects reported any history of speech or hearing disorders. They were recruited through introductory linguistics classes and received \$3.50 for their participation in a single 1-h session.

**Materials and Procedure.** Two "hood"- "HUD" continua were synthesized using the Klatt (1980) formant synthesizer. These stimuli had the same formant values for F1-F3 that were used in Experiment 1. F4 and F5 were set to 3500 and 4200 Hz, respectively, in all of the "hood"- "HUD" tokens. The two "hood"- "HUD" continua were different from each other only by virtue of the fact that F3 was 2512 Hz in the high F3 continuum and 2368 Hz in the low F3 continuum.

A change in the amplitude of F3 in these tokens, as compared with those used in Experiment 1, was effected by narrowing the bandwidth of F3 from 110 Hz (as it was in Experiment 1) to 30 Hz. This change in bandwidth (together with a change in spectral tilt of the voice source) produced a change in the amplitude of F3 of 4 dB for a vowel sound with formants at 474, 1111, 2368, 3500, and 4200 Hz (token 1 in the "hood"- "HUD" continuum). Flanagan (1957) found that the just-noticeable difference for F2 in [a] is 3 dB. The spectrum of the voice source was given half as much tilt as had been present in the tokens used in Experiment 1.

To verify that the shift of F3 in the "hood"- "HUD" continuum was audible, a preliminary listening test was conducted using 2 listeners (K.J. and M.W.). The task was an AX discrimination task in which items from the high and low F3 series were compared. F1 and F2 in the pairs were identical; in half of the pairs, F3 was different. A' scores for the two listeners were .61 and .80, respectively, for K.J. and M.W., with .5 indicating chance performance and 1.0 indicating perfect performance. These scores indicate that the change from high to low F3 in the "hood"- "HUD" continuum was detectable.

The procedure of the vowel-identification experiment was identical to that of Experiment 1.

### Results and Discussion

The identification functions for the "hood"- "HUD" continuum at the two different levels of F3 are presented in Figure 3. The data were analyzed in a two-way repeated-measures ANOVA in which the factors were F3 level (high or low) and token number. As is obvious in the figure, there was a significant effect of token number on percentage of "hood" responses [ $F(6,78) = 119.2, p < .001$ ]. There were no other significant effects. Observed  $F$  values for the F3 main effect and the F3  $\times$  token interaction were both less than 2 ( $p > .25$ ). In the high F3 condition, the percentage of "hood" responses to the "hood"- "HUD" continuum was 53.8, whereas in the low F3 condition, it was 55.7.

Although the F3  $\times$  token interaction was not significant, there does appear to have been a slight shift in the category boundary. Category boundaries were computed by linear interpolation for each subject in the two F3 conditions for the "hood"- "HUD" continuum. The ob-

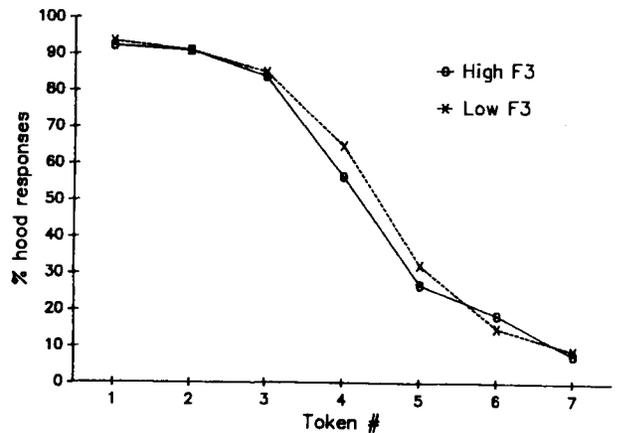


Figure 3. Results of Experiment 2. Identification functions for "hood"- "HUD" continua with two levels of F3.

served boundary shift in the "hood"- "HUD" continuum was marginally significant in a correlated observations  $t$  test [ $t(12) = 2.345, p < .05$ ].

It is not clear how the shift in the "hood"- "HUD" boundary can be attributed to a cognitive speaker-normalization process. In a speaker-normalization process, F3 would be used by the hearer as an indication of vocal-tract length, and thus the hearer's expectation for vowel-formant ranges would be adjusted. If F3 is low, this would be taken as an indication of a long vocal tract and thus of generally lower formants. So the formants in the "hood"- "HUD" continuum would not have to be as high to be identified as "HUD." Therefore, a cognitive-level process of speaker normalization based on F3 should result in more "HUD" responses when F3 is low. The boundary shift observed here indicates that when F3 was low, there were slightly fewer "HUD" responses—the opposite of what is predicted by positing a cognitive-level normalization effect. Therefore, neither the auditory-integration hypothesis nor a cognitive-range normalization hypothesis can account for the shift in the back-vowel continuum. It is necessary to hypothesize that some other perceptual process resulted in this shift in identification (if indeed it is a reliable effect).

Thus, F3 appears to play no role in a speaker-normalization process for back vowels. In this experiment, F3 was made assurably audible in a "hood"- "HUD" continuum, yet shifting F3 by almost 150 Hz in these tokens had virtually no effect on the category boundary, and the small effect that was found was not consistent with a speaker-normalization hypothesis.

### GENERAL DISCUSSION

The results of these experiments have been discussed in terms of a model of vowel normalization that includes reference to Chistovich et al.'s (1979) "center-of-gravity" effect in a process of formant-peak extraction. It has been suggested here that higher formant normalization occurs at this stage and not at the stage of vowel identification.<sup>2</sup>

In the paragraphs below, two recent models of vowel normalization and how they might account for the data reported here will be discussed.

The view taken by the present author of how Chistovich et al.'s (1979) center-of-gravity effect may play a role in vowel perception is quite different from the view that has been adopted by Syrdal and Gopal (1986). In Syrdal and Gopal's model of vowel normalization, vowels are represented in terms of the differences (in Bark) among vowel formants and fundamental frequency. Thus, the difference between F1 and F0 corresponds to the traditional vowel feature "openness" (see also Traunmüller, 1981). If F1 - F0 is greater than 3 Bark, the vowel is assigned the value [+open]; if the difference is less than 3 Bark, it is assigned the value [-open]. In a similar manner, the difference between F3 and F2 corresponds to the feature "frontness." If the difference between F3 and F2 is greater than 3 Bark, the vowel is classified as [-front]; if it is less than 3 Bark, the vowel is classified as [+front]. In Syrdal and Gopal's model, Bark differences of two types are used to encode two different levels of representation: an "intermediate auditory" analysis is encoded by the specific Bark differences, and a phonetic feature representation is based on critical (< or >3) Bark differences.

The main difference between Syrdal and Gopal's (1986) view of vowel perception and the one assumed in the present study is that the 3-Bark rule operates at different stages of analysis in the two approaches. In Syrdal and Gopal's model, the 3-Bark rule comes into play at the point of assigning phonetic feature values to vowels. In other words, vowel-formant values are treated as "given" and the 3-Bark rule is used to assign phonetic features to the vowel. In the view of vowel perception adopted by the present author, on the other hand, the 3-Bark rule comes into play at the stage of formant extraction. The center-of-gravity effect as seen from this viewpoint is analogous to a frequency analysis into critical bands (Scharf, 1970; Zwicker, 1961) but with a larger filter bandwidth (approximately 3 Bark). Thus, whereas Syrdal and Gopal start their consideration of vowel normalization with the actual vowel formants (the actual resonances of the vocal tract) as given and then look at how vowel normalization might be accomplished from this representation of vowels, I am including a stage of processing in which the perceptual estimation of vowel-formant peaks may be influenced by the 3-Bark rule.

Unless we posit that subjects make use of specific Bark-level differences (as opposed to phonetic feature specifications), the shift in identification observed in the present study in the subject's responses to the [ɪ]-[e] continuum cannot be accounted for by Syrdal and Gopal's (1986) model. This is because openness is determined solely by the values of F0 and F1 (which were identical for the two "hid"- "head" continua), whereas F3 manipulation would exclusively affect the determination of frontness. The manipulation of F3 frequency in the "hid"- "head" continua was not enough to shift the vowels from [+front]

to [+back], so basing identification responses solely on the assignment of phonetic features, Syrdal and Gopal's model predicts no difference between the two front-vowel continua. If, however, subjects are able to make use of specific Bark-level differences, we must posit that a shift in the values relevant for the front-back distinction (distance between F3 and F2) influences subjects' judgments along a high-low continuum. Although this difference between the model and the data may not be irreconcilable, I would suggest that the approach taken in the present study, in which F3 plays a role in the perceptual estimation of F2, provides a more straightforward explanation of the [ɪ]-[e] data. In addition, note that if specific Bark-level differences can be used by subjects in making vowel classifications, there is nothing in Syrdal and Gopal's model to prevent an F3-F2 difference of more than 3 Bark from playing a role in the perceptual classification of back vowels. The data reported here seem to indicate that there is, in the human perceptual system, some limitation of the degree to which the relationship between F3 and F2 can influence the perceptual identities of vowels. The suggestion that the center-of-gravity effect applies in the process of formant-peak extraction provides a principled explanation of the shift of identification observed here in the front-vowel series *and* the lack of shift in the back-vowel series.

The model of vowel normalization proposed by Bladon, Henton, and Pickering (1984) includes a frequency-smearing function in the auditory processing of speech that is comparable to the application of the center-of-gravity effect in the approach I have chosen. Bladon et al.'s model of normalization involves three steps. First, a peripheral auditory spectral analysis is performed, which results in a spectrum with frequency resolution in critical bands (Bark). In the next stage, a process of frequency masking (Schroeder, Atal, & Hall, 1979) is assumed to apply. The result of this stage is a smoothed frequency spectrum in which adjacent peaks have been integrated. It is at this stage that the integration of F3 and F2 would occur. The final stage in Bladon et al.'s model is a speaker-normalization process in which the auditory spectrum is shifted down by 1 Bark if the F0 of the utterance is high enough to indicate that the speaker is female. Vowel identification is taken in this model to be a process of template matching, rather than one involving formant extraction (Bladon, 1982).

The main difference between Bladon et al.'s (1984) model and the approach that I have assumed has to do with how vowels are identified (template matching vs. formant extraction), not with how higher formant normalization might be accounted for. In both approaches, F3 and F2 are integrated during an auditory stage of processing and the higher formants are not explicitly used in speaker-normalization processes at the time of vowel identification. It is possible that the center-of-gravity approach I have employed and the frequency-masking approach used by Bladon et al. make different predictions, specifically about the frequency separation between formants

that is possible before shifts in vowel quality are to be expected. In general, however, the two approaches are quite comparable.

The issue of whether vowel perception is accomplished by spectral template matching or by formant-frequency extraction is beyond the scope of this paper (see Syrdal & Gopal, 1986, p. 1093, and Bladon, 1982, for a discussion of some of the issues involved). One point that Syrdal and Gopal have made which bears repeating is that spectral templates encode a certain amount of information that has nothing to do with vowel identification. Vowel spectra include information about formant amplitudes, formant bandwidths, source spectral tilt, overall vowel amplitude, and so forth. None of these properties are distinctive characteristics of particular vowels, so vowel templates must be flexible enough to ignore these bits of information while at the same time retaining a high degree of sensitivity to the locations of spectral peaks. How this differs substantially from vowel identification based on formant values is not clear.

To summarize the present experiments, then, the data reported here suggest that the higher formant normalization effect is the result of a process in auditory perception and nothing more. In the experiments reported in this paper, it was found that when F3 and F2 were within 3 Bark of each other, there was a shift in vowel-identification functions when F3 was shifted. When F3 and F2 were not within 3 Bark of each other, however, a shift in F3 did not produce the expected shift in vowel identification. This work extends that of Carlson et al. (1975) in that the detectability of F3 in the back-vowel continuum was measured, so one possible explanation of the lack of an effect of higher formants in the perception of back vowels was addressed.

Although these results do not rule out the possibility that hearers use higher formants in combination with other phonetic cues in a cognitive-level process of adjusting to a new speaker, they do seem to rule out the possibility that higher formants alone are used in a cognitive-level process of speaker normalization. In this regard, note that it has been reported (Fujisaki & Kawashima, 1968; Slawson, 1968) that in continua for which no higher formant-normalization effect was observed, there seemed to be an increased perceptual normalization (over the shift in identification observed when F0 alone was varied) when the higher formants covaried with F0. It may be that this type of covariant cue is a triggering environment for a cognitive process of speaker-based normalization that utilizes higher formant values in an estimate of the speaker's vocal-tract size.

Note also that these data do not rule out the possibility that higher formants might be used in cognitive-level normalization processes when no F0 is present in the signal (i.e., in noise-excited vowels). Fujisaki and Kawashima (1968) assumed that the use of noise-excited tokens increased the higher formant-normalization effect because

the amplitudes of the higher formants had been increased (in fact, they found that when the spectrum of the noise source was given a 12-dB per octave tilt, the effect was reduced). The data reported here, however, give rise to the possibility that the use of higher formants in a range-normalization process is triggered more by the absence of F0 than by the relative amplitude of the higher formants (although, of course, they must be audible in order to play a role in perception).

## REFERENCES

- BLADON, R. (1982). Arguments against formants in the auditory representation of speech. In R. Carlson & B. Granström (Eds.), *The representation of speech in the peripheral auditory system* (pp. 95-102). Amsterdam: Elsevier.
- BLADON, R., HENTON, C., & PICKERING, J. (1984). Towards an auditory theory of speaker normalization. *Language & Communication*, 4, 59-69.
- CARLSON, R., FANT, G., & GRANSTRÖM, B. (1975). Two-formant models, pitch and vowel perception. In G. Fant & M. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 55-82). London: Academic Press.
- CHISTOVICH, L., SHEIKIN, R., & LUBLINSKAJA, V. (1979). "Centres of gravity" and spectral peaks as the determinants of vowel quality. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 143-158). London: Academic Press.
- FLANAGAN, J. (1957). Difference limen for formant amplitude. *Journal of Speech & Hearing Disorders*, 22, 205-212.
- FUJISAKI, H., & KAWASHIMA, T. (1968). The roles of pitch and higher formants in the perception of vowels. *IEEE Transactions on Audio Electroacoustics*, AU-16, 73-77.
- GERSTMAN, L. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio Electroacoustics*, AU-16, 78-80.
- JOHNSON, K. (1988). *Processes of speaker normalization in vowel perception*. Unpublished doctoral dissertation, Ohio State University, Columbus, OH.
- KARNICKAYA, E., MUSHNIKOV, V., SLEPOKUROVA, N., & ZHUKOV, S. (1975). Auditory processing of steady-state vowels. In G. Fant & M. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 37-54). London: Academic Press.
- KLATT, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67, 971-995.
- NEAREY, T. (1987). Static, dynamic and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 81, S16.
- PETERSON, G., & BARNEY, H. (1952). Control methods used in a study of the identification of vowels. *Journal of the Acoustical Society of America*, 24, 175-184.
- SCHARF, B. (1970). Critical bands. In J. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 1, pp. 157-202). New York: Academic Press.
- SCHROEDER, M. R., ATAL, B. S., & HALL, J. L. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 217-229). London: Academic Press.
- SLAWSON, A. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of the Acoustical Society of America*, 43, 87-101.
- SYRDAL, A., & GOPAL, H. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79, 1086-1100.
- TRAUNMÜLLER, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America*, 69, 1465-1475.
- ZWICKER, E. (1961). Subdivision of the audible frequency range into critical bands (frequenzgruppen). *Journal of the Acoustical Society of America*, 33, 248.

NOTES

1. Whether it is more accurate to describe speaker normalization as an estimation of vocal-tract length or as an estimation of formant ranges (as in Gerstman, 1968) cannot be addressed here.

2. The possibility that speaker normalization may occur during vowel identification is not precluded by this explanation of higher formant nor-

malization. See Johnson (1988) for a review of and contribution to the evidence that F0 is used in just such an adjustment-to-talker normalization process.

(Manuscript received September 21, 1988;  
revision accepted for publication February 16, 1989.)

## Announcement

**30th Annual Meeting of the Psychonomic Society  
Atlanta, Georgia  
November 17-19, 1989**

The 30th Annual Meeting of the Psychonomic Society will be held in Atlanta, Georgia, November 17-19, 1989. The meetings will begin Friday morning and continue until Sunday at noon. The headquarters hotel will be the Hyatt Regency.

The program and hotel reservation cards will be mailed to members and associates in September. A copy of the program will be published in the November issue of the *Bulletin of the Psychonomic Society*.

For further information, please contact the secretary-treasurer of the Society: Michael E. Rashotte, Department of Psychology, Florida State University, Tallahassee, FL 32306-1051 (Telephone: 904-644-2040; BITNET: Rashot@FSU).