THE HYPERSPACE EFFECT: PHONETIC TARGETS ARE HYPERARTICULATED

KEITH JOHNSON
University of Alabama at Birmingham

EDWARD FLEMMING RICHARD WRIGHT
UCLA

A commonly made, but rarely defended, assumption is that phonetic reduction processes apply to hyperarticulated phonetic targets. Results from experiments reported in this paper support this assumption. In various experimental conditions, listeners adjusted the input parameters of a speech synthesizer until the vowels it produced sounded like the vowels found in a set of example words. A preliminary study indicated that the method of adjustment is a feasible tool for studying vowel systems. Interestingly, listeners in the study chose vowels that were systematically different from those measured in productions of the set of example words: high vowels were higher, low vowels were lower, front vowels were farther front, and back vowels were farther back. We hypothesized that this extreme vowel space corresponds to phonetic targets that are hyperarticulated: HYPERSPACE. This hypothesis was tested in the two main experiments. The first experiment controlled for possible effects of instructions and phonetic training on the listeners' choices. In the second experiment, we improved the naturalness and distinctiveness of the synthetic vowels. The results indicate that the extreme vowels chosen by the listeners were consistent with those produced in hyperarticulated speech; moreover, the hyperspace effect is robust across experimental conditions. These results validate the hypothesis that phonetic targets are hyperarticulated, and are consistent with a two-stage model of phonetic implementation: at the first stage distinctive features are mapped to hyperarticulated phonetic targets, and at the second stage these phonetic targets are reduced.*

INTRODUCTION

1.1. It is a commonplace observation that one sentence can be produced in many different ways, even by a single speaker. For example, the sentence *Did you eat yet*? can be realized in ways that are approximately transcribed in 1. These range from the most distinct, clear production, often called HYPERARTICULATED, to a very reduced utterance, a style often characterized as 'casual speech'.

(1) a. [dɪd ju iʔtʰ jɛʔt]
b. [dɪd ju iʔt jɛʔt]
c. [dɪdʒuiʔjɛʔ]
d. [dʒuiʔjɛʔ]
e. [dʒiʔjɛʔ]

* This work was supported by an NIH training grant (T32 DC00029) to Peter Ladefoged and Patricia Keating, and an NIH FIRST award (R29 DC01645-01) to Keith Johnson. We would like to thank Bob Port for his encouraging words in the early stages of this work, Peter Ladefoged for his enthusiastic interest, Doug Whalen for suggesting Experiment 2, and Mary Beckman and Jim Flege for comments on an earlier version of the paper. Our colleagues at the Indiana University Speech Research Laboratory, the Indiana University Department of Linguistics, the UCLA Phonetics Laboratory, and the Acoustical Society of America gave us many helpful comments. Two anonymous reviewers helped us see things we hadn't seen before.

Some of the variation in utterances of a sentence may be introduced by optional categorical phonological rules, but much of the variability is along continuous parameters, e.g. durations, and so cannot be accounted for in terms of categorical rules. Thus, much of the variation must be the result of phonetic realization.

Most approaches to phonetic realization do not directly address the issue of casual speech. Two that do are Lindblom's 1990 'Hyper- and Hypoarticulation' (H&H) theory of phonetic variation and Browman & Goldstein's 1986, 1990 articulatory phonology. Lindblom 1990 conceptualizes speech production as a feedback system in which the input is the goal and the actual production is the output. The extent to which the output matches the input goal depends on the gain, or amplification, of the feedback loop. The gain of the feedback loop is thus analogous to effort. The salient feature of this model is that the input goal is the most distinctive, hyperarticulated speech, since this is the signal that the output approximates as the gain is maximized. Reduction processes are not conceived of as altering the goals, but result from expending less effort and thus falling short of the goal.

In fact it appears that this general approach is widely assumed, since almost all discussions of the clear speech/casual speech continuum are cast in terms of 'undershoot' (failure to achieve target), 'reduction', and related concepts. Of course, this is not the only possible account of these phenomena. We could postulate the existence of both reduction and hyperarticulation processes, in which case the canonical phonetic representation would be of an intermediate level, perhaps akin to citation forms. We can probably reject the third logical possibility, which is the hypothesis that there are only hyperarticulation processes, since the most reduced forms of words are so indistinct that it would be difficult to derive the clear distinctions that exist between them in hyperarticulated speech if they were the starting point. This is essentially the same type of consideration that led Jakobson & Halle (1956) and Hockett (1955), among others,[1] to state that the most clearly articulated speech is most relevant to phonological analysis because it contains the most information.

Browman & Goldstein's 1990 analysis of casual speech in the framework of articulatory phonology is consistent with the hypothesis that there are both reduction and hyperarticulation processes. In articulatory phonology, lexical items are represented as a series of coördinated articulatory gestures (the 'gestural score'), each of which specifies the formation of a linguistically significant constriction of the vocal tract, e.g. labial closure or velum lowering. They propose that casual speech variants of canonical lexical gestural representations are produced by increasing temporal overlap between gestures and reducing gestural magnitudes spatially and temporally. Although only reduction

[1] Daniel Jones' cardinal vowels come to mind because they 'have by definition tongue-positions as remote as possible from "neutral" position' (1960:31–39). However, these extreme vowel qualities were intended to be used as part of a descriptive and pedagogical system. Some of the cardinal vowels have extreme qualities because Jones wanted the system to cover the range of physically producible vowel sounds, not because he assumed that hyperarticulated speech is more basic than normal speech.

processes are discussed in Browman & Goldstein 1990, there is no claim that the lexical representation is the most hyperarticulated form of a word, and it seems that hyperarticulation processes could be expressed in their model as decreasing overlap and increasing gestural magnitudes.

The experiments reported in this paper explored phonetic targets by means of a method that reveals listeners' perceptual expectations of speech sounds. The results support the traditional view, more recently expressed in Lindblom 1990, that phonetic targets are hyperarticulated. Section 1.2 describes a fundamental methodological problem in making crosslinguistic acoustic/phonetic comparisons (the normalization problem) and an experimental technique which solves the problem, the METHOD OF ADJUSTMENT. In the method of adjustment (henceforth MOA), listeners adjust one or more parameters of a speech synthesizer until the machine correctly pronounces a speech sound (Scholes 1967, 1968, Nooteboom 1973, Ganong & Zatorre 1980, Samuel 1982, Johnson 1989). A preliminary test of the MOA with vowels is described in §2; that test showed that listeners using the MOA were consistent from trial to trial and with each other. Additionally, the preliminary experiment found an interesting mismatch between measured vowel production and the perceptual results: the perceptual vowel space was expanded relative to the production space. A replication and test of this result is reported in §3. That experiment showed that the perceptual vowel space expansion did not depend on the phonetic sophistication of the listeners, or on the instructions given them. The experiment also showed that the perceptual vowel space expansion is similar to the expansion seen when one compares very carefully or clearly produced (hyperarticulated) speech with normal citation productions of the same words. We therefore call the perceptual vowel space expansion the HYPERSPACE EFFECT. Section 4 reports another test of the hyperspace effect that showed that it is not the result of the lack of natural acoustic cues in the synthetic stimuli. In §5 we argue that the experimental results show that phonetic targets are hyperarticulated, and we briefly outline the implications of this conclusion for the theory of phonetic realization.

1.2. PHONETIC COMPARISON OF VOWEL SYSTEMS AND THE METHOD OF ADJUSTMENT. While the most common experimental techniques used in speech perception research focus on the boundaries between phonetic categories, the MOA provides information about linguistic/phonetic targets for speech sounds. Repp & Liberman (1987) discuss the need for data on the internal structure of phonetic categories, as opposed to category boundaries, and assume that the MOA provides information about the 'prototypes' of phonetic categories (see also Samuel 1982 concerning this assumption). However, they note that, 'until recently, no one had used methods designed to identify prototypes' (p. 90), and that 'the application of such methods has so far failed to yield entirely satisfactory results'. Since 1987 speech perception researchers have begun to focus on prototypes with greater success (Johnson 1989, Kuhl 1991, Miller & Volaitis 1989).

The MOA is useful for crosslinguistic comparisons of vowel systems because personal differences in vocal tract anatomy (vocal tract size, the ratio of oral cavity size to pharynx cavity size, palate doming, lip shape, etc.) give rise to

acoustic differences among speakers (Ladefoged & Broadbent 1957). Research since the late 1940s has shown that the speech signal varies quite considerably from speaker to speaker even within the same language and dialect (Joos 1948, Peterson & Barney 1952), and thus crosslinguistic acoustic/phonetic comparisons of vowels confound personal and linguistic differences. Any crosslinguistic acoustic comparison of vowel systems includes some unknown amount of personal variation.

One way to compensate for personal variation in making crosslinguistic comparisons is to scale the measured formant values by a factor corresponding to vocal tract size. The scale factor may be derived from the range of observed formant values for a particular speaker (Gerstman 1968), the mean of the observed formant values (Lobanov 1971), the mean of the log-transforms of the observed formant values (Nearey 1977), or a function of the speaker's fundamental frequency (F0) (Miller 1989, Syrdal & Gopal 1986). Disner (1980:257) suggested that it is valid to use the formant mean or range 'so long as the data are drawn from a single language or dialect, such that the same set of vowel phonemes is shared by all speakers'. Therefore, the methods suggested in Gerstman 1968, Lobanov 1971, and Nearey 1977 are not valid for making crosslinguistic comparisons when the languages have different numbers of vowels or vowels of differing quality. Methods of formant normalization which use the speaker's F0 are also flawed, because they rely on the observed APPROXIMATE correlation of F0 with vocal tract length, and are also valid only for comparisons of vowels produced in similar prosodic contexts.

Another class of normalization techniques uses a crosslinguistic formant average as the normalizing factor. For instance, Disner 1980 used PARAFAC (a three-mode factor analysis technique; Harshman 1970) to compare the vowels of English, German, Danish, Swedish, and Dutch. The PARAFAC procedure relates measurements from individuals to the overall mean of the data set and finds speaker constants which scale the individual's vowel space to the overall vowel space. In later work, Disner (1986) compared the vowels of various languages using analysis of variance models which included factors for vowel, speaker, and language. In both of these approaches the differences between languages are tested as deviations from crosslinguistic means.

Two limitations are inherent in this class of normalization techniques. First, as with range normalization and average value normalization, only comparable vowel qualities can be tested crosslinguistically, which means that it is not possible to compare whole vowel systems when they contain unequal numbers of vowels or vowels of differing quality. Second, the method assumes that the average individual deviation from the overall mean is not correlated with language. To see the problem with this assumption, consider an extreme example. If all of the data for language X are taken from recordings of female speakers while all of the data from language Y are taken from recordings of male speakers, an analysis of variance would show quite large differences in formant values as a function of language, even though speaker is included as a factor in the model. Thus, although the statistical techniques proposed by Disner may provide a better solution to the normalization problem (for crosslinguistic comparisons) than those provided by other approaches, the problem is not solved

by using crosslinguistic formant averages as normalizing factors, because the results still depend on the comparability of the groups of speakers who represent each language (see Behne 1989).

The MOA offers a way of making crosslinguistic phonetic comparisons that circumvents these difficulties. By using a single synthetic voice, the MOA makes it possible to ascertain the listener's perceptual expectations for vowel sounds in a particular language FOR THAT VOICE. Thus, the linguistic and personal aspects of the speech signal are disentangled.

The studies reported here started with a basic test of the MOA using speakers of Southern California English. In particular, we wanted to know how much reliable information about a language's vowel space could be generated by means of the technique. Beyond this basic test, we found in the preliminary test that listeners chose vowels that did not match those produced by ANY speaker in normal speech. The discrepancy between listeners' choices in the MOA and speakers' productions is interesting because it was systematic. The perceptual vowel space was expanded relative to the production space: high vowels were higher, low vowels lower, front vowels more front, and back vowels more back. We will present data that indicate that the MOA vowel space corresponds to the vowel space seen in hyperarticulated speech.

PRELIMINARY STUDY

2. A preliminary study was designed to give a first indication of the feasibility of the MOA for crosslinguistic phonetic research. We sought answers to two questions: (1) can listeners do the task in a reasonable amount of time and with reasonably low within- and between-listener variability? And (2) will the task provide interpretable data for crosslinguistic comparisons?

2.1. SUBJECTS. Ten female and four male university students served as volunteer subjects. They had self-reported normal speech and hearing and had recently completed a one-quarter course in phonetic transcription. This pool of subjects represented fairly diverse linguistic backgrounds: four were monolingual English-speaking Southern Californians, six were English-dominant Southern Californians, one was a native of Maryland, two were native speakers of Serbo-Croatian, and one was a native speaker of Spanish. We will present data collected from the ten Southern Californians.

2.2. MATERIALS. A software formant synthesizer (Klatt & Klatt 1990) was used to produce 330 steady-state isolated vowel stimuli by varying the two lowest vocal tract resonances (the first formant, F1, and second formant, F2) independently over a large range of values. There were fifteen possible values of F1 (from 250 Hz to 900 Hz) and twenty-two possible values of F2 (from 800 Hz to 2800 Hz). F1 corresponds to vowel height, while F2 corresponds to vowel frontness; this set of stimuli covered the entire range measured in the speech of adult male speakers. The intervals between successive stimuli along both dimensions were about four tenths of an auditory critical band (0.37 Bark; see Scharf 1970 for more information about auditory frequency resolution). These intervals are slightly larger than the just-noticeable-differences for vowel formants reported in Flanagan 1957.

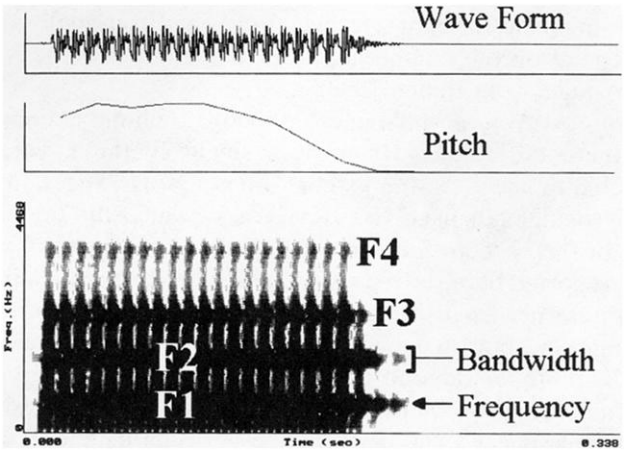

FIGURE 1. Wave form, pitch trace, and spectrogram of one of the synthetic vowels showing some of the acoustic parameters that were controlled in the synthesizer.

Figure 1 shows a spectrogram of one of the stimuli with the two lowest formants (F1 and F2) labeled. As indicated in the figure, the stimuli were composed of several acoustic parameters in addition to F1 and F2. These additional parameters contributed to the naturalness of the synthesized vowels and either were the same for all of the stimuli or were estimated from F1 and F2 by formulas.

All of the stimuli were 250 milliseconds (ms) long, and they all had the same pitch contour (see Fig. 1, which shows the output of a pitch tracker applied to the stimulus). The fundamental frequency of voicing (F0) was steady at 120 Hz over the first half, and fell gradually over the last half of the vowel to a final value of 105 Hz. Because the F0 was low, the synthetic stimuli sounded as if they had been produced by a male speaker. The frequency of the third formant (F3) was estimated by regression formulas (published in Nearey 1989) that take into account the empirical relationship between F3 and various combinations of F1 and F2 in English. The fourth formant was constrained to be at least 300 Hz higher than F3 and no lower than 3500 Hz. The bandwidths of the three lowest formants (B1, B2, B3) were estimated using a strategy that was similar to that used to estimate F3. Bandwidth is the width of the formant resonance in Hz (see Fig. 1); it must be controlled in a synthesizer to get natural-sounding speech. The empirical relationships between the bandwidth values and the values of F1, F2, and F3 suggested in Klatt 1980 for synthetic vowels were expressed by linear formulas which were then used to calculate the bandwidths in the stimuli. The regression formulas are given in 2–4. (The statistic r^2 is a value between 0 and 1, where 1 means that the formula will predict the bandwidth perfectly and 0 means that there is no relationship between the bandwidth and the formants.)

$$(2)\quad B1 = 29.27 + 0.061*F1 - 0.027*F2 + 0.02*F3 \qquad r^2 = 0.605$$
$$(3)\quad B2 = -120.22 - 0.116*F1 + 0.107*F3 \qquad r^2 = 0.497$$
$$(4)\quad B3 = -432.1 + 0.053*F1 + 0.142*F2 + 0.151*F3 \qquad r^2 = 0.595$$

As the r^2 values suggest, these formulas provided only a rough fit to the bandwidth values suggested by Klatt, and extreme formant values resulted in unnatural bandwidths. The bandwidth of F4 was constant for all of the stimuli at a typical value (200 Hz). To increase the naturalness of the stimuli further, the 'natural' voice-source in the synthesizer was used and other parameters relating to the mode of vocal cord vibration changed over the time-course of the syllable, simulating changes seen in naturally produced syllables (see Klatt & Klatt 1990).

In addition to these synthetic stimuli, we compiled a list of common English words illustrating the vowels of English for use in both the production and perception parts of the experiment. The list, with IPA symbols for the vowels, was: *heed* [i], *hid* [ɪ], *aid* [eɪ], *head* [ɛ], *had* [æ], *HUD* (acronym for the Federal Housing and Urban Development agency) [ʌ], *odd* [ɑ], *awed* [ɔ], *owed* [oʊ], *hood* [ʊ], *who'd* [u]. Note that this list differs minimally from the list of words used by Peterson & Barney (1952) and subsequent researchers. We changed this classic list by substituting some words with initial glottal stop (like *aid*) because these words are more common than the corresponding words with /h/ and because glottal stop and /h/ have similarly small effects on the formants of the following vowel in speech production.

2.3. PROCEDURE. The experiment involved two tasks. First, the subjects were asked to read ten repetitions of the list of English words (in the carrier phrase *say* _____ *again*). The order was randomized separately each time through the list. The subjects were seated in a sound booth and recordings were made using high-quality equipment (Sennheiser microphone, Symetrix SX202 preamplifier, and Tascam 122 cassette recorder). Formant values from these recorded utterances were measured using CSpeech (Paul Milenkovic) from an LPC spectrum which was calculated from a point early in the vowel (between 1/3 and 1/2 of the way through) as determined in a digital wave-form display.

Second, subjects served as listeners in the MOA task using the same list of words as visual stimuli. This part of the experiment was run on-line by an IBM PC-AT. The listener saw a word at the top of a CRT screen, and a two-dimensional grid. Each square in the grid corresponded to one of the vowel sounds in the F1, F2 matrix; the listener used a mouse to select a particular square and clicked a mouse button to hear the synthetic vowel associated with that square.[2] The synthetic vowel sounds were stored on disk and were played out through a Data Translation DT2801A digital-to-analog converter. The sampling rate was 10 kHz and the signal was low-pass filtered at 4.2 kHz before

[2] In the region of the acoustic vowel space where F1 and F2 are close to each other, the acoustic vowel space has a corner cut off (F1 was always at least 250 Hz below F2), while in the visual display this is not true. This complication was handled by filling in the corner of the visual display with copies of nearby tokens. If the listener were to choose the square that would correspond to an F1 of 1000 Hz and an F2 of 1200 Hz, for example, a token with the same F1 and the next higher F2 value would be presented. Thus, the vertical dimension of the display always corresponded to different F1 values, but, in one corner of the visual display, changes in the horizontal dimension of the grid did not result in changes of F2.

being amplified (BGW Systems, Model 85) and presented diotically over headphones (Sony MDR-V4). (For more details concerning the setup, see Johnson & Teheranizadeh 1992.) The listener's task was to find the location in the grid associated with a synthetic vowel that sounded like the vowel in the visually presented word. After choosing the F1 and F2 values for the vowel of a particular word, the listeners were asked to rate their choice on a scale from one to ten. The rating data were used to eliminate mistakes, primarily accidental terminations of trials. This task was repeated 10 times for each of 11 words in the list.

Because we wanted to collect several adjustment trials for each of several vowels and we did not want the listeners to rely simply on visual cues in making their judgments, we changed the orientation of the acoustic vowel space on the screen randomly from trial to trial. On 50% of the trials, therefore, the high F1 stimuli were at the top of the screen, and on the other 50% of the trials they were at the bottom of the screen. Similarly, the relationship of F2 to the horizontal dimension of the grid also changed from trial to trial.

2.4. RESULTS AND DISCUSSION. Table 1 shows the standard deviations in the MOA task for each word in the responses from the ten native Californians. The first two columns show the overall standard deviations (calculated across all responses) of F1 and F2, while the next two columns show the average within-listener standard deviations of F1 and F2 (calculated for each listener separately and then averaged). The ratios shown in Table 1 indicate how much of the variability in the data is due to listeners' inconsistency from trial to trial (however small) and how much is due to differences among the listeners. The average ratio of within-listener to total standard deviation for F1 is 0.83. The ratio of within-listener to total standard deviation for F2 is 0.73. Because the

TABLE 1. Comparison of total variability and average within-listener variability in the preliminary experiment. The first two columns show the overall standard deviations (in Hz) of F1 and F2 in the MOA trials. The middle two columns show the average within-listener standard deviations (in Hz) for the same data. The last two columns show the ratio of average within-listener to total variability.

	TOTAL		WITHIN-LISTENER		RATIOS	
WORD	SDF1	SDF2	SDF1	SDF2	F1	F2
heed	19.5	164.6	16.2	99.9	.83	.61
hid	29.7	158.4	26.3	124.9	.89	.79
aid	55.9	229.6	37.1	151.2	.66	.66
head	49.7	173.0	41.7	114.6	.84	.66
had	61.4	175.5	56.4	126.3	.92	.72
odd	56.5	56.6	50.0	52.2	.89	.92
awed	55.7	42.4	48.8	35.8	.88	.84
HUD	48.9	98.5	41.1	70.6	.84	.72
owed	35.5	57.6	31.5	46.0	.89	.80
hood	36.2	96.9	35.7	85.8	.99	.89
who'd	67.9	112.4	41.9	86.2	.61	.77
AVERAGE	47.0	124.1	38.8	90.3	.83	.73

average within-listener standard deviation is about as large as the total standard deviation, these data show that most of the variability in the MOA task occurred within the responses of individual listeners rather than appearing as between-listener variability in the formant values chosen. Note also that standard deviations tend to be higher for higher formant values (SDF2 is higher than SDF1). This reflects the fact that the equal auditory step-sizes in the stimulus set resulted in larger acoustic differences between the stimuli as the frequency increased. One surprising observation to be noted in Table 1 is that /i/ had a smaller ratio of within-listener to total variation in F2 than did the other vowels, and /u/ had a smaller ratio in F1. Listeners were internally consistent in their choices for these vowels, but there was relatively more discrepancy among listeners than with most of the other vowels. Also note that ratios were low for both of the formants for *aid*. This was probably due to the fact that the synthetic stimuli were steady-state vowels while the target vowel is diphthongal. The vowel in *owed* is less diphthongal in this dialect than in other dialects of English. Although there are some interesting patterns in the variability found in this preliminary study, the most important finding is that the variability is low—average standard deviations of about 50 Hz for F1 and 100 Hz for F2.

The vowels in *odd* and *awed* were merged in the acoustic measurements of the vowels produced by eight native Southern Californian males in the citation readings of Experiment 1 (shown in Figure 2a).[3] Note also that the vowels in *aid* and *hid* had very similar formant values.

As in the measured formant values, Fig. 2b shows that the native Southern California listeners in the preliminary experiment merged the vowels in *odd* and *awed* in the MOA. In addition to the merger of the low nonfront vowels, the data indicate that the listeners kept the vowels of *aid* and *hid* more distinct in the MOA than they are in production. This tendency was noted in an earlier MOA study of vowels (Johnson 1989). As suggested in that earlier report, when potential cues such as formant movement are not available, listeners in the MOA may exaggerate an existing small spectral difference in order to maintain a linguistic distinction. It is also possible that the production data in Fig. 2a do not accurately represent the spectral properties of /e^{I}/ because of our choice of measurement location. This explanation of the discrepancy between the production and perception result seems rather unlikely, however, because we made the acoustic measurements early in the diphthong. Therefore, we would expect if anything to see an even lower F1 and a higher F2 for /e^{I}/ if we were to measure later in the vowel. If we were to make the formant measurements later in the vowel, then, we would expect to see even more discrepancy between the production and perception data, not less.

One of the listeners in the preliminary experiment was older than the other listeners and was born in New York City. Figure 3a shows that he maintained the distinction between *odd* and *awed* in production. In other respects his

[3] These production data from Experiment 1 are used for comparison here because there were too few male subjects in the preliminary experiment and because it is necessary to compare the MOA results with male formant values because the synthetic stimuli had a male voice.

acoustic vowel space is similar to the average vowel space shown in Fig. 2a. Fig. 3b shows that this subject also selected different formant values for the vowels in *odd* and *awed* in the perception experiment (and the difference between /e^{1}/ and /ɪ/ was also quite expanded in the perception space). Keep in mind that, although the comparison between production vowel spaces (Figs.

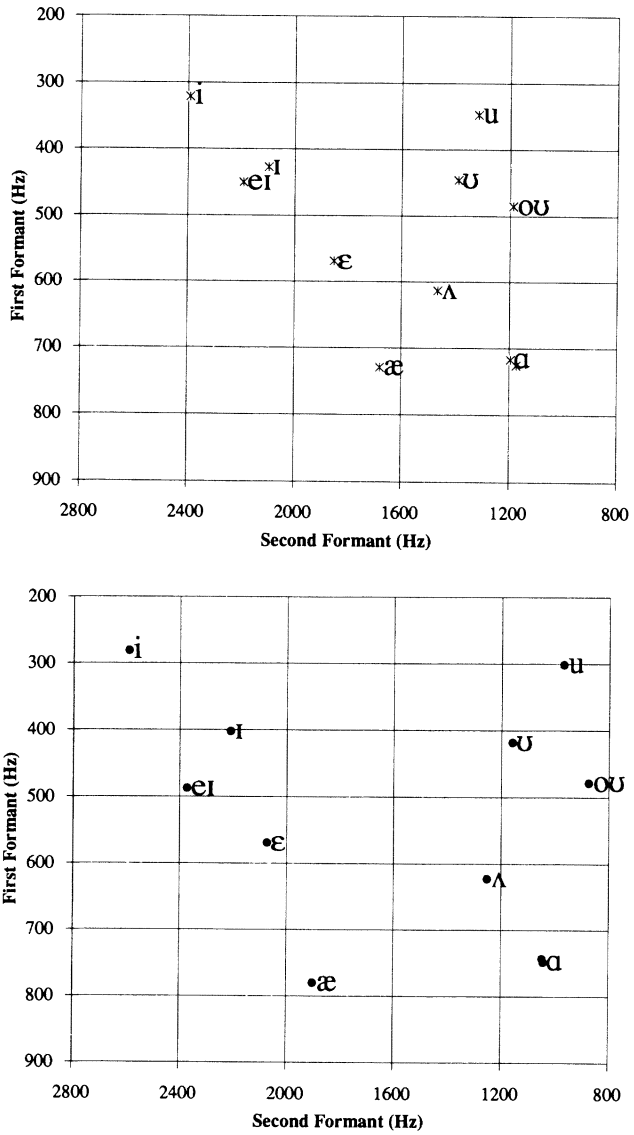

FIGURE 2. (a) Average measured formant values of vowels produced by the eight male native Southern California English speakers from Experiment 1. These vowels were produced in the 'citation' reading condition. (b) Average MOA results for the native Southern California English speakers in the preliminary experiment.

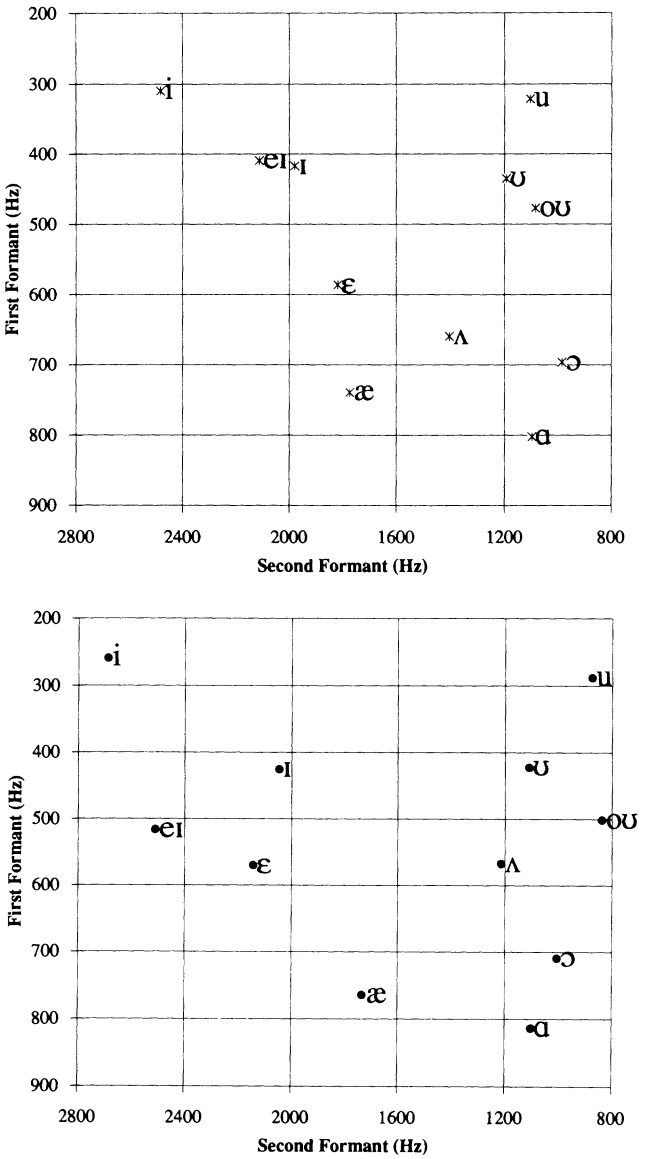

FIGURE 3. (a) Average measured formant values of vowels produced by one male speaker in the preliminary experiment. This speaker maintained a distinction between the vowels in *awed* and *odd*. (b) Average MOA results obtained from the same speaker.

2a and 3a) gives us about the same picture of the difference between this speaker and the others, the perception vowel spaces (Figs. 2b and 3b) allow for a better comparison: because the listeners are telling us their expectations for the vowel space of a single synthetic speaker, speaker and dialect information are not confounded.

EXPERIMENT 1: INSTRUCTION SET

3. The preliminary data suggested that the MOA may indeed be a useful tool in the study of vowel systems. In the remainder of this paper we will focus on a methodological puzzle and its significance both for the use of the MOA in studying vowels and for theories of phonetic realization.

The puzzle is illustrated by a comparison of the average perception vowel space from the preliminary experiment (Fig. 2b) with the average citation production space from the 8 male speakers in Experiment 1 (Fig. 2a). This comparison shows that the vowel space chosen in the MOA was expanded relative to the production vowel space. In other words, listeners' expectations for vowels produced by a male synthetic voice were quite different from the vowels actually produced by male speakers in citation speech. This is a conundrum if we assume that listeners' perceptual expectations are based on experience.

There were a couple of aspects of the preliminary experiment that made us doubt the generality of this discrepancy between production and perception vowel spaces. The listeners were not phonetically naive; they had completed an undergraduate course in phonetics and thus knew the cardinal vowel system. They may therefore have been inclined to select extreme cardinal vowel qualities in the MOA task where naive speakers might choose formant values more similar to those found in production. Additionally, we suspected that the instructions given to the listeners might have biased them toward extreme vowel qualities: we asked the listeners to find the 'best' vowel sound for each word. After the fact we realized that this instruction could have been interpreted to mean, 'Find the most distinct example of the vowel'.

Experiment 1 was designed to investigate these issues by using (1) naive listeners and (2) a careful manipulation of the instruction set. One group of listeners was instructed to find the best example of the vowel in each word; another group of listeners was instructed to find the vowel sound which most closely matched their own pronunciation of the vowel in each word.

3.1. SUBJECTS. Ten female and eight male university students were recruited through the student newspaper and paid a small sum for their participation. They were monolingual English speakers who reported normal speech and hearing ability and who had attended high school in Southern California. The subjects were divided into two groups as described below, with five females and four males in each group.

3.2. MATERIALS. As in the preliminary experiment, 330 steady-state isolated vowel stimuli with fifteen possible values of F1 and twenty-two possible values of F2 were synthesized using a software formant synthesizer (Klatt & Klatt 1990). The formant ranges in Experiment 1 were larger than those used in the preliminary experiment, because the listeners in the preliminary experiment chose formant values which were more extreme than we had anticipated. F1 ranged from 250 Hz to 1000 Hz in increments along an auditory frequency scale (0.42 Bark), while F2 ranged from 800 Hz to 2900 Hz, again in equal auditory frequency increments (0.39 Bark).

3.3. PROCEDURE. Experimental sessions in Experiment 1 were very much like sessions in the preliminary experiment. After the subjects had completed the perception part of the experiment, however, they were asked to read the word list a second time. In this second reading of the words we elicited hyperarticulated or clear-speech versions of the words by saying 'What?' or 'Huh?' after each sentence, prompting the speaker to read each sentence again more clearly. This procedure was explained to the speakers prior to starting the tape recorder. We will call the first reading the CITATION reading and the second the HYPERARTICULATED reading. One other procedural difference in Experiment 1 concerned the instructions given to the listeners in the MOA task. We asked one group of listeners (5 female, 4 male) to find the best examples of the vowels and another group of listeners (5 female, 4 male) to find the vowel sound which most closely matched their own pronunciation of the vowel in each word. We will call the first instruction set the BEST condition and the second set the AS YOU SAY IT condition.

Finally, the recordings were analyzed using CSL (Kay Elemetrics) rather than CSpeech. In analyzing these productions we chose measurement points from spectrographic displays and identified an early steady-state portion of the vowel as the point representing the acoustic vowel 'target'.

3.4. RESULTS AND DISCUSSION. The perception results from Experiment 1 are shown in Figure 4a. In separate repeated-measures analyses of variance there were no reliable effects of instruction set on the choices made by the listeners on F1 or F2 for any of the vowels. The most robust difference as a function of instruction set was for the F1 of *awed,* which tended to be greater in the AS YOU SAY IT group than in the BEST group (806 Hz versus 758 Hz, respectively), but this difference was only marginally reliable ($F[1,16] = 3.19$, $p = 0.093$). No other differences between groups proved to be statistically reliable.

Although the vowel spaces chosen were not affected by instruction set, the ratings given to the synthetic stimuli (shown in Table 2) were. Listeners in the BEST condition were consistently more critical of the stimuli than were the listeners in the AS YOU SAY IT condition. The statistical comparison of these conditions was complicated by ceiling effects in the rating data, but the trend is clear. The result is that the instruction set manipulation had an effect on ratings, but it did not have an effect on the formant values chosen in the MOA.

In addition, a comparison (shown in Fig. 4b) of the perceptual vowel spaces of naive listeners from Experiment 1 and phonetically trained listeners from the preliminary experiment averaged over instruction conditions suggests that phonetic training had no effect on the results of the MOA task. It is not valid to attempt a statistical comparison of the data shown in Fig. 4b, because there were several small changes in the method (in particular, the range of possible F1/F2 combinations was expanded in Experiment 1). Still, the differences appear to be of the same magnitude as the nonsignificant effects of manipulating the listeners' instructions (Fig. 4a) and are certainly nothing like the mismatch between measured values from citation forms and the MOA results (Fig. 2a versus Fig. 2b).

Thus, the perceptual vowel space resulting from the MOA task is robust. Speakers of the same dialect give the same answers regardless of the specifics of the instructions they are given or their previous training in phonetics. This robust pattern of performance raises an interesting question—namely, what in the experience of the listener underlies the difference between the MOA vowel

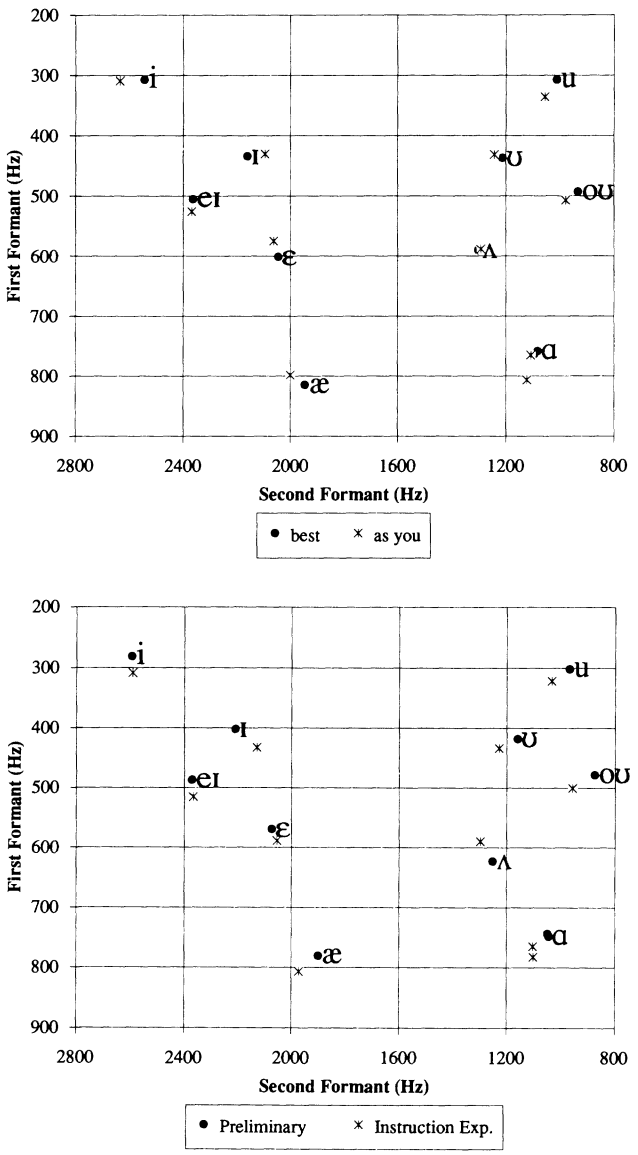

FIGURE 4. (a) Average MOA results of the instruction set experiment. Filled circles show the average responses of the BEST group, and stars show the average responses of the AS YOU SAY IT group. (b) Comparison of the MOA results from the preliminary experiment averaged over listeners (filled circles), and the MOA results of Experiment 1 averaged over listeners and instructions (stars).

TABLE 2. Average rating values given to synthetic vowels in the perception part of Experiment 1. Rating values (on a scale from 1 to 10) were averaged without including vowels rated 1 because the listeners were asked to use 1 to indicate that they had accidentally terminated a trial early. Data are presented by vowel and by listening condition. Starred items in the AS YOU SAY IT column were significantly higher than average ratings of the same vowel in the BEST condition.

WORD	BEST	AS YOU SAY IT
heed	8.9	9.7
hid	8.6	9.0
aid	8.7	9.5
head	8.9	9.4
had	8.6	9.4
odd	9.0	9.7
awed	8.8	9.9*
HUD	8.8	9.5
owed	8.1	9.8**
hood	8.2	9.4**
who'd	8.6	9.2
AVERAGE	8.7	9.5

*$p < 0.1$ **$p < 0.05$

space and the acoustic vowel space found in normal productions of the same words?

We hypothesized that the perceptual vowel space which is found in the MOA task reflects hyperarticulated versions of the vowels rather than the vowel qualities found in less carefully produced speech. We will call this the HYPERSPACE HYPOTHESIS. With this hypothesis in mind, we asked the speakers to read the word list in a hyperarticulated style. As has been found before (Picheny et al. 1986, Moon & Lindblom 1989), the hyperarticulated versions of the vowels generally had more extreme vowel formants than did the less carefully produced vowels (see Figure 5a). This is just the sort of vowel space expansion that we saw in comparing the MOA results with citation readings of the words.

A comparison of the average vowel formants in hyperarticulated productions and the average formant values chosen in the MOA (shown in Fig. 5b) suggests that the hyperspace hypothesis is on the right track. Further, when we looked at the hyperarticulated vowel spaces for individual speakers, we found that all of the formant values chosen in the perception task were represented in the productions of at least one speaker. The listeners chose vowels comparable to hyperarticulated productions rather than normal citation productions.[4]

[4] One referee remarks concerning the AS YOU SAY IT condition, 'Since the stimuli represented male productions, female listeners were matching some kind of idealized abstract values in the vowel space, rather than sounds that their own vocal tracts/larynges would have been capable of producing. The hyperspace is more abstract than the paper makes it seem'. We agree that the vowel space derived in the MOA is abstract, but would caution against the impression that this 'ideal' space is in any way tied to the male voice. In another experiment not reported here we found the hyperspace effect in the responses of male listeners to a synthetic female voice.

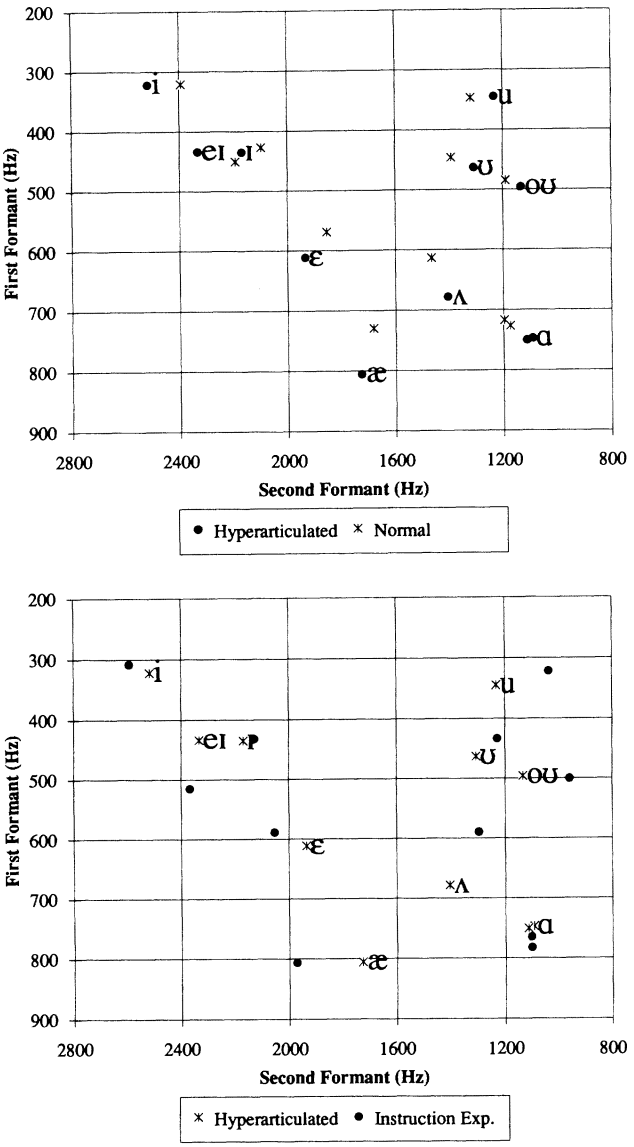

FIGURE 5. (a) Average measured formant values of vowels produced by the eight male speakers in Experiment 1. Filled circles show the average values in the hyperarticulated condition and stars show the average values in the citation-form condition. (b) Comparison of hyperarticulated productions (8 male speakers from Experiment 1) with MOA results. Stars show the average measured formant values from hyperarticulated versions of the vowels and filled circles show the MOA results of Experiment 1 averaged over listeners and instructions.

Experiment 2: Intrinsic F0 and duration

4. One factor that may have had an effect on the MOA results both in the preliminary experiment and in Experiment 1 is that the stimuli were impoverished relative to natural speech. While vowels in English differ in intrinsic pitch, duration, and formant trajectories (Peterson & Barney 1952, Peterson & Lehiste 1960, Lehiste & Peterson 1961), the stimuli that we used in the MOA task did not vary along these dimensions. This aspect of the stimuli complicates any interpretation of the MOA results, because listeners may have attempted to compensate for a loss in the overall distinctiveness of the synthetic vowel stimuli (resulting from the absence of redundant cues) by increasing distinctiveness in the vowel space. Therefore, the hyperspace effect may have been an artifact of the stimuli. We tested this possibility in a second experiment.

In Experiment 2, patterns of intrinsic vowel F0 and duration (in American English) were modeled in the synthetic stimuli. We varied F0 and duration based on the F1 and F2 of the vowel in a way that is similar to their observed variation in English. Thus, some portion of the redundant information which was missing from the stimuli used in the preliminary experiment and in Experiment 1 was present in these stimuli. If the hyperspace effect occurred in those earlier experiments because of the lack of redundant information in the stimuli, we should find a reduction (but probably not a total elimination) of the effect in Experiment 2.

4.1. Subjects. Three (two males and one female) native speakers of Southern Californian English volunteered for the experiment. The listeners reported normal speech and hearing abilities and had completed two introductory phonetics courses. Because we found no differences in performance of the MOA as a function of phonetic training between the preliminary experiment and Experiment 1, these listeners were taken to be representative of Southern California English.

4.2. Materials. As in the earlier experiments, 330 isolated steady-state vowels were synthesized. The formants and bandwidths were the same as those in the stimuli synthesized for Experiment 1; however, F0 and duration varied from stimulus to stimulus, rather than being fixed as they were in the earlier experiments.

The method used to derive F0 and duration values for the stimuli was analogous to the method used in the earlier sets to derive bandwidth values by rule (formulas 2–4 above). Average F0 and formant values for male speakers from Peterson & Barney's 1952 study of American English vowels were entered into regression analysis in which F0 was predicted by F1 and F2. As a result of our use of the regression formula, shown in 5, to calculate F0 values for the synthetic stimuli, F0 ranged from 110 Hz to 142 Hz. As in the earlier experiments, F0 was steady over the first half of the vowel and then fell gradually to about 85% of its original value over the last half.

$$(5)\quad \text{F0 (in Hz)} = 153.44 - 0.035*\text{F1} - 0.00275*\text{F2} \quad r^2 = 0.939$$

Similarly, average duration measurements from Peterson & Lehiste 1960 and

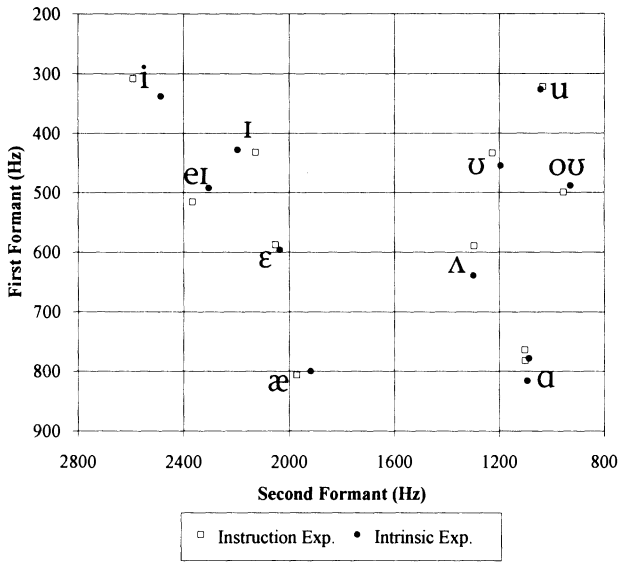

FIGURE 6. Average MOA results from Experiment 2 (filled circles) compared with the MOA results of Experiment 1, averaged over listeners and instructions (open squares).

formant measurements from Peterson & Barney 1952 for tense vowels in English were analyzed, and a regression formula (6) was calculated for duration as a function of F1 and F2.[5] The resulting durations ranged from 210 ms to 305 ms for the range of F1, F2 combinations in the vowel array. The duration equation is problematic for English, because lax vowels have much shorter durations than their tense counterparts, even though their formant values are comparable. Thus, the duration formula captures only vowel duration variation which is correlated with F1 and F2 variation, and not the duration differences between tense and lax vowels.

(6) Dur (in ms) = 191.754 + 0.121*F1 − 0.00347*F2 $r^2 = 0.792$

4.3. PROCEDURE. In contrast to the earlier experiments, no production data were collected in Experiment 2. The MOA task was conducted using the same equipment and software as in the earlier experiments. The listeners were instructed to find vowels that sounded like the ones they produced in the words (the AS YOU SAY IT condition of Experiment 1). Each of the eleven English words used in the earlier experiments was presented in random order 7 times.

4.4. RESULTS AND DISCUSSION. Figure 6 shows the results of Experiment 2 compared with the average results of Experiment 1. This figure indicates that there were only very minor differences between the vowel formants chosen when F0 and duration varied as a function of vowel quality and when they did not. In particular, the vowel space did not uniformly shrink when intrinsic F0

[5] We assume that the pattern of intrinsic vowel durations for vowels produced in isolation would be similar to those reported in Peterson & Lehiste 1960 for vowels in CVC context.

and duration were modeled in the synthetic stimuli. These results show that the expanded vowel space that was found in the preliminary experiment and in Experiment 1 was not an artifact of the stimuli. If the absence of redundant information such as intrinsic F0 differences or duration differences between vowels had caused an expansion of the vowel space, we would have expected some contraction of the vowel space in this experiment. This result did not occur; the hyperspace effect persisted.

Conclusion

5.1. Why do listeners choose a hyperspace in the MOA task? One possible answer is that the result is an experimental artifact. Although we tested several possible artifacts—instruction set, phonetic training, and lack of redundant cues—there may still be some aspect of the stimuli or of the task itself that biases listeners toward a hyperspace. For example, if we were to present synthesized versions of whole words rather than isolated vowels, listeners might be inclined to choose less extreme vowel qualities (Lindblom & Studdert-Kennedy 1967); but since measured formant values from /hVd/ contexts are very similar to those found in isolated vowel productions (Peterson & Barney 1952), it seems unlikely that the hyperspace effect occurred because the stimuli were isolated vowels.

Another possible explanation of the effect as an artifact of the experiment is that the formality of the test situation may have biased listeners toward a hyperspace in the MOA task. Against this explanation we can note that, while it is difficult to elicit casual speech in experimental situations, it is also quite difficult to elicit hyperarticulated speech. We found that speakers, when simply instructed to speak clearly, would initially produce quite hyperarticulated speech, but after only a few utterances they would revert back to the same style of speaking that they used in the citation reading. This is why we adopted a special procedure to elicit hyperarticulated speech in Experiment 1. It is not clear why hyperarticulation would be the listener's response, but not the speaker's, to the formality of an experimental situation.

Instead of viewing the hyperspace effect as an experimental artifact, we propose to interpret it in light of the nature of the perceptual task. The MOA is a PHONETIC task. The results are measurable along continuous dimensions like F1 and F2. Additionally, the MOA gets at phonetic TARGETS (Repp & Liberman 1987, Samuel 1982). Listeners tell us how a synthesizer is supposed to pronounce a sound by reference to a representation of that sound in memory. Thus, because the task is designed to reveal the nature of phonetic targets, the most straightforward interpretation of the hyperspace result is that PHONETIC TARGETS ARE HYPERARTICULATED.

5.2. Long before the discovery of the hyperspace effect, many phonologists believed that hyperarticulated (clear-speech) variants of sounds have a special status in phonological analysis. Jakobson & Halle expressed this view and offered what we will call the 'information argument' when they said, 'The slurred fashion of pronunciation is but an abbreviated derivative from the explicit clear-speech form which carries the highest amount of information ...

When analyzing the pattern of phonemes and distinctive features composing them, one must resort to the fullest, optimal code at the command of the given speakers' (1956:6). The argument is that clear-speech forms must be basic because reduced speech can be derived from clear speech, but not vice versa. Instead of this view we could assume that there are hyperarticulation processes (fortitions) that produce hyperarticulated forms from less elaborated representations. In adopting this sort of view, Donegan & Stampe (1979:142) state that fortitions 'apply in situations and styles where perceptibility is highly valued'. However, fortitions can only make the identity of a word clearer if the properties they introduce were already inherent in the representation (just as a telescope can increase the perceptibility of stars only if the stars were there all along). Donegan & Stampe's characterization (with which we agree) of fortitions as perceptually motivated makes sense only if the 'information' is in the targets, not in fortition processes. Therefore, fortitions are more accurately seen as descriptions of the pronunciation of phonetic targets in the absence of lenitions, and hence it is apparently the case that for every lenition there is an equal and opposite fortition.

A second argument for the phonetic-targets-are-hyperarticulated hypothesis comes from Hockett's discussion of 'clarity-norm' speech. He argued that 'clarity-norm' speech (i.e. hyperarticulated speech) avoids 'phonologic complications which seem to be of no morphophonemic relevance' (1955:221). For instance, in normal speech the morphological association between *hat* [hæt] and *hatter* [hæɾɚ] is obscured by flapping, but it is apparent in hyperarticulated speech ([hæt^{h}] : [hæt^{h}ɚ]). Hockett seems to have considered this to be only a practical issue; he was agnostic about whether phonologists should describe hyperarticulated speech or normal speech, but he noted that morphological analysis requires the former. Still, it wouldn't be surprising if speaker/hearers also made implicit use of the morphophonemic relevance of hyperarticulated speech in language acquisition. Therefore, the 'morphophonemic argument' for hyperarticulated phonetic targets is this: the language learner must 'undo' some reduction processes in order to uncover the morphological structure of some forms. The 'undone' versions of the forms are available to the child as hyperarticulated speech. There is thus an impetus for the child to attend to and remember hyperarticulated forms for the sake of uncovering morphological structure (although even in clear speech some relationships remain opaque—e.g. *divine : divinity*).

Recent x-ray studies of speech movements are consistent with the conclusion that phonetic targets are hyperarticulated. Sproat & Fujimura (1990:23) found that '/l/s in English all have both a dorsal and an apical gestural component', and that the articulatory realization of this ubiquitous gestural composition differed depending on the location of /l/ in a syllable and the strength of any following boundary. They found that a wide variety of surface forms could all be related to an invariant gestural description using phonetic implementation rules that introduced 'continuous variation sensitive to a number of factors' (23). This is a theme in the latest articulatory research. Forms which appear to be quite different, and which would be transcribed differently, can be mod-

elled as articulations composed of the same gestures which blend with each other or cover each another, or which are undershot to varying degrees (see Browman & Goldstein 1990). This is an extension of the information argument. The form that has the most phonetic material realized—the gestures unhidden, unblended, with targets reached—reveals information which may be hidden in other productions. What is important with regard to the phonetic-targets-are-hyperarticulated hypothesis is that the articulatory gestures are not deleted, only prevented from making an acoustic appearance.

5.3. The theory of phonetic realization must account for the wide range of realizations of the same utterance that a single speaker produces in differing situations. Some of the variation may be introduced by optional phonological rules, but much of it is continuous and very low-level in nature, and must surely be the result of phonetic implementation. The details of a phonetic implementation model consistent with our experimental results have not been fully worked out, but an outline is clear. This type of model, represented in 7, includes (a) a mapping from categorical representations to parametric representations that correspond to hyperarticulated speech, and (b) phonetic reduction processes. The first stage is a categorical-to-parametric mapping of distinctive features to phonetic parameters like vocal-tract shapes or formant values. The second stage is a parametric-to-parametric mapping that may do much to account for the wide range of realizations of the same utterance that a speaker can produce.

(7) Categorical representations
(a) |
Hyperarticulated targets
(b) |
Normal speech

An alternative model of phonetic realization has phonetic implementation rules that are context-sensitive, i.e., the categorical-to-parametric mapping itself is variable. For instance, the feature [+high] might be realized as various F1 targets, depending on the degree of effort the speaker is willing to expend, because the mapping from features to parameters is a function both of the distinctive feature and of various parameters of the performance context. A conceptual difficulty with this model is that the different realizations of a feature all have equal status as phonetic realizations of that feature (this is also a problem for Keating's 1988 'window' model of coarticulation). Thus, a reduced schwa-like version of /i/ is just as good an example of a high vowel as is a hyperarticulated, maximally distinct /i/. As shown in the MOA task, this runs counter to the intuitions of naive listeners.

One implication of assuming that phonetic targets are hyperarticulated is that the search for the phonetic correlates of distinctive features (Stevens & Keyser 1989) should focus on hyperarticulated speech. If phonetic targets are hyperarticulated, the phonetics/phonology interface is not a mapping between phonological representations and normal speech, but between phonological representations and very carefully articulated speech. Thus, the search for

acoustic and articulatory correlates of phonological units, in order to be successful, should focus on carefully produced speech (e.g. Potter et al. 1947).

Additionally, if phonetic realization includes a parametric-to-parametric mapping, then it is reasonable to expect certain differences in the nature of phonetic processes as compared with phonological rules. For instance, there is much evidence showing that phonetic processes, unlike phonological rewrite rules, are gradient, producing variants that are sometimes hard to capture in phonetic transcriptions (see 1 above).

REFERENCES

BEHNE, DAWN M. 1989. A comparison of the first and second formants of vowels common to English and French. Research on Speech Perception Progress Report 15.269–82. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.

BROWMAN, CATHE P., and LOUIS GOLDSTEIN. 1986. Towards an articulatory phonology. Phonology 3.219–52.

——, ——. 1990. Gestural specification using dynamically-defined articulatory structures. Journal of Phonetics 18.299–320.

DISNER, SANDRA F. 1980. Evaluation of vowel normalization procedures. Journal of the Acoustical Society of America 67.253–61.

——. 1986. On describing vowel quality. Experimental phonology, ed. by John J. Ohala and Jeri J. Jaeger, 69–79. Orlando: Academic Press.

DONEGAN, PATRICIA JANE, and DAVID STAMPE. 1979. The study of natural phonology. Current approaches to phonological theory, ed. by Daniel A. Dinnsen, 126–73. Bloomington, IN: University of Indiana Press.

FLANAGAN, JAMES. 1957. Estimates of the maximum precision necessary in quantizing certain 'dimensions' of vowel sounds. Journal of the Acoustical Society of America 29.533–4.

GANONG, WILLIAM F., and ROBERT J. ZATORRE. 1980. Measuring phoneme boundaries four ways. Journal of the Acoustical Society of America 68.431–9.

GERSTMAN, LOUIS H. 1968. Classification of self-normalized vowels. IEEE Transactions on Audio and Electroacoustics AU-16.78–80.

HARSHMAN, RICHARD. 1970. Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis. Los Angeles, CA: UCLA Working Papers in Phonetics 16.

HOCKETT, CHARLES. 1955. Manual of phonology. (International Journal of American Linguistics, Memoir 11.) Baltimore: Waverly Press.

JAKOBSON, ROMAN, and MORRIS HALLE. 1956. Fundamentals of language. 'S-Gravenhage: Mouton.

JOHNSON, KEITH. 1989. On the perceptual representation of vowel categories. Research on Speech Perception Progress Report 15.343–58. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.

——, and HENRY TEHERANIZADEH. 1992. Facilities for speech perception research at the UCLA phonetics lab. Los Angeles, CA: UCLA Working Papers in Phonetics 81.123–39.

JONES, DANIEL. 1960. An outline of English phonetics. 9th edn. Cambridge: Cambridge University Press.

JOOS, MARTIN. 1948. Acoustic phonetics. (Linguistic Society of America, Language Monograph 23.) Baltimore: Waverly Press.

KEATING, PATRICIA A. 1988. The window model of coarticulation: Articulatory evidence. Los Angeles, CA: UCLA Working Papers in Phonetics 69.3–29.

KLATT, DENNIS. 1980. Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America 67.971–95.

——, and LAURA KLATT. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America 87.820–57.
KUHL, PATRICIA. 1991. Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. Perception & Psychophysics 50.93–107.
LADEFOGED, PETER, and DONALD BROADBENT. 1957. Information conveyed by vowels. Journal of the Acoustical Society of America 29.98–104.
LEHISTE, ILSE, and GORDON E. PETERSON. 1961. Transitions, glides and diphthongs. Journal of the Acoustical Society of America 33.268–77.
LINDBLOM, BJÖRN. 1990. Explaining phonetic variation: A sketch of the H&H theory. Speech production and speech modeling, ed. by William J. Hardcastle and A. Marchal, 403–39. Dordrecht: Kluwer.
——, and MICHAEL STUDDERT-KENNEDY. 1967. On the role of formant transitions in vowel recognition. Journal of the Acoustical Society of America 42.830–43.
LOBANOV, B. M. 1971. Classification of Russian vowels spoken by different speakers. Journal of the Acoustical Society of America 49.606–8.
MILLER, JAMES D. 1989. Auditory-perceptual interpretation of the vowel. Journal of the Acoustical Society of America 85.2114–34.
MILLER, JOANNE L., and LYDIA E. VOLAITIS. 1989. Effect of speaking rate on the perceptual structure of a phonetic category. Perception & Psychophysics 46.505–12.
MOON, S. J., and BJÖRN LINDBLOM. 1989. Formant undershoot in clear and citation-form speech: A second progress report. Speech Transmission Laboratory–Quarterly Progress Status Report 1.121–23.
NEAREY, TERRANCE M. 1977. Phonetic feature systems for vowels. Storrs: University of Connecticut dissertation.
——. 1989. Static, dynamic and relational properties in vowel perception. Journal of the Acoustical Society of America 85.2088–113.
NOOTEBOOM, SIG G. 1973. The perceptual reality of some prosodic durations. Journal of Phonetics 1.25–46.
PETERSON, GORDON E., and H. L. BARNEY. 1952. Control methods used in a study of vowels. Journal of the Acoustical Society of America 24.175–84.
PETERSON, GORDON E., and ILSE LEHISTE. 1960. Duration of syllable nuclei in English. Journal of the Acoustical Society of America 32.693–703.
PICHENY, M. A.; N. I. DURLACH; and LOUIS D. BRAIDA. 1986. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. Journal of Speech & Hearing Research 29.434–46.
POTTER, RALPH K; GEORGE A. KOPP; and HARRIET GREEN. 1947. Visible speech. New York: Dover.
REPP, BRUNO H., and ALVIN M. LIBERMAN. 1987. Phonetic category boundaries are flexible. Categorical perception, ed. by Steven N. Harnad, 89–112. New York: Cambridge University Press.
SAMUEL, ARTHUR. 1982. Phonetic prototypes. Perception & Psychophysics 31.307–14.
SCHARF, B. 1970. Critical band. Foundations of modern auditory theory, vol. 1, ed. by J. V. Tobias, 159–202. New York: Academic Press.
SCHOLES, ROBERT J. 1967. Phoneme categorization of synthetic vocalic stimuli by speakers of Japanese, Spanish, Persian, and American English. Language and Speech 10.46–68.
——. 1968. Perceptual categorization of synthetic vowels as a tool in dialectology and typology. Language and Speech 11.194–207.
SPROAT, RICHARD, and OSAMU FUJIMURA. 1990. On the nature of allophonic variation and the phonology/phonetics mapping: The case of /l/ in English. Murray Hill, NJ: AT&T Bell Labs, MS.
STEVENS, KENNETH N., and SAMUEL JAY KEYSER. 1989. Primary features and their enhancement in consonants. Lg. 65.81–106.

SYRDAL, ANN, and H. S. GOPAL. 1986. A perceptual model of vowel recognition based on the auditory representations of American English vowels. Journal of the Acoustical Society of America 79.1086—100.

Keith Johnson
Department of Biocommunication
University of Alabama at Birmingham
Birmingham, AL 35294-0019

[Received 18 November 1992;
accepted 24 January 1993.]