675

Speech Communication 8 (1989) 1–15
North-Holland

TRAVERSING UPPER VOWEL SPACE: A SMOOTH OR A BUMPY RIDE?

Bruno H. REPP
Haskins Laboratories, New Haven, CT 06511–6695, U.S.A.

Received 5 November 1987
Revised 21 April 1988

Abstract. It has been found in vowel imitation studies that subjects' responses to steady-state isolated vowels from an acoustic continuum exhibit categorical tendencies: Some adjacent vowels are responded to more similarly than others, and the distribution of formant frequencies in the total set of responses is decidedly nonuniform. The present study investigated whether these tendencies originate in perception or in articulation. Subjects were asked to produce continuous glides from /i/ to /æ/ and from /æ/ to /i/ at three different speeds, as well as discrete vowels along the same continuum, without any auditory models. The results show that categorical tendencies of the kind observed previously in imitation tasks emerge as soon as vowel glides are produced at a rate that requires conscious control over the articulatory trajectory. The categorical tendencies thus seem to originate in the speech production system, although on-line perceptual guidance through auditory feedback cannot be ruled out at this stage.

Zusammenfassung. Studien über die Wiederholung von Vokalen haben gezeigt, dass die Reaktion von Informanten auf gedehnte Vokale, welche innerhalb eines akustischen Kontinuums selektiert wurden, gewisse kategorielle Tendenzen aufweisen. Für gewisse benachbarte Vokale gleichen sich die Wiederholungen weit mehr als für andere, und die Verteilung der Formantenfrequenzen innerhalb der Gesamtheit der Antworten ist ungleichmässig. Die hier vorgestellte Studie untersucht inwieweit diese Tendenzen auf die Artikulationsebene oder auf die Perzeptivebene zurückzuführen sind. Die Sprecher produzierten, ohne auditives Modell, Gleitlaute von [i] nach [æ] und von [æ] nach [i] mit drei verschiedenen Geschwindigkeiten, und ebenfalls diskrete Vokale innerhalb des gleichen Kontinuums. Die Resultate zeigen, dass ähnliche kategorielle Tendenzen sich einstellen, sobald die Gleitlaute mit einer Geschwindigkeit produziert werden, welche eine bewusste Kontrolle der Artikulationskurve verlangt. Es erscheint somit als ob diese kategoriellen Tendenzen ihren Ursprung innerhalb der Sprachproduktion finden, obwohl eine schritthaltende perzeptive Rückwirkung im Moment nicht ausgeschlossen werden kann.

Résumé. Lors d'expériences d'imitation de voyelles, il a été montré que les réponses des sujets à des voyelles soutenues sélectionnées parmi un continuum acoustique faisaient preuve de tendances catégorielles. Les imitations de certaines voyelles adjacentes étaient plus proches que d'autres, et la distribution des fréquences formantiques parmi l'ensemble des réponses est clairement non-uniforme. Notre travail vérifie si ces tendances prennent naissance au plan articulatoire ou au plan perceptif. Des sujets ont produit des voyelles glissantes allant de [i] vers [æ] et de [æ] vers [i] à trois vitesses différentes, ainsi que des voyelles discrètes sur le même continuum, en l'absence de modèles auditifs. Les résultats montrent que des tendances catégorielles de même nature que celles observeés antérieurement dans des tâches d'imitation apparaissent dès que les glides sont produits à une vitesse qui réclame un contrôle conscient de la trajectoire articulative. Les tendances catégorielles semblent ainsi prendre naissance au sein du système de production de la parole, bien qu'une guidance perceptive en temps réel au travers de la rétroaction auditive ne puisse être totalement exclue au stade actuel.

Keywords. Vowel imitation, categorical tendencies, articulatory control, perceptual guidance.

1. Introduction

In a now classic study, Chistovich et al. (1966) found that a (single, experienced) subject's vocal imitations of isolated, synthetic vowels from an /i/–/a/ continuum followed a categorical pattern: Some vowels adjacent in formant space were responded to similarly, while other such neighbors received very different responses. The number of clusters of stimuli receiving acoustically similar re-

sponses exceeded the number of phonemic categories in the subject's native language (Russian), so it was hypothesized that the clusters reflected a subphonemic, articulatory categorization of the input. Another piece of evidence suggesting that the clusters did not reflect phonemic categorization was that imitation reaction times did not vary across the continuum, while phonemic classification latencies varied in accordance with the relative ambiguity of the stimuli.

Kent (1973) attempted to replicate these findings with four phonetically trained, English-speaking subjects. He used both /i/–/æ/ and /u/–/i/ continua, expecting to find stronger categorical responses with the former, since it crossed several phoneme categories (/ɪ/,/ɛ/, perhaps /e^{i}/) whereas the latter did not include any English vowels other than the endpoints. Kent's results showed considerable individual variability and categorical tendencies that were weaker than those observed by Chistovich et al. Such tendencies were present on both continua, though perhaps somewhat more systematic along the /i/–/æ/ continuum. Schouten (1977), in a related study using a large number of stimuli ranging from /i/ to /æ/ with speakers of several languages, also found some evidence for categorical tendencies in imitation.

In two recent experiments, Repp and Williams (1985, 1987) followed up on Kent's research. The first experiment used Kent's stimuli and the two authors as subjects. The results confirmed Kent's findings of categorical tendencies and individual differences in their pattern and extent. The second experiment replicated the first with similar vowel continua whose stimuli had been selected from the two subjects' responses in the first experiment. Even when imitating vowels produced by themselves, the two subjects evinced systematic deviations from the targets in their imitation responses. The pattern of these deviations, or categorical tendencies, was similar to that observed in the first experiment with synthetic stimuli. It was concluded that the categorical tendencies were not an artifact of using synthetic stimulus materials.[1]

With one relatively trivial explanation ruled out, several possible explanations of the categorical tendencies remain. One is that their origin is in perception. Categorical tendencies in vowel discrimination have been found in several studies using high-uncertainty paradigms (Healy and Repp, 1982; Pisoni, 1973; Repp et al., 1979). Even though such findings have often been attributed to strategies based on covert phonemic labeling, they could also derive from a discontinuous precategorical coding of the stimuli. Speech perception theorists generally agree that speech sounds must be represented internally in some auditory or articulatory code before they are categorized. It is not clear, however, why that code itself should have quasi-categorical properties. Such properties would have to originate in multiple nonlinearities in auditory processing that lead to a distorted auditory vowel space. Although at least one such possible nonlinearity, the 3.5 Bark formant separation "boundary", has been proposed (Syrdal and Gopal, 1986), it seems unlikely that there would be several such boundaries to account for three or four "steps" in imitation responses to stimuli from a vowel continuum. Moreover, Macmillan et al. (1987) have shown that there are no discontinuities in discrimination along an /i/–/ɪ/–/ɛ/ continuum when a low-uncertainty paradigm is used, which more directly taps into low-level auditory processes.

One alternative to a perceptual explanation is production-based. That is, even though the auditory representation of the speech input may be continuous, its conversion into a motor output (i.e., an imitation) may give rise to discontinuities. These discontinuities could arise from constraints inherent in the articulatory system, which make some vocal tract configurations easier to achieve than others, and/or from articulatory habits acquired in learning a native language. Although it may be argued that such discontinuities, if they exist, will in turn affect the nature of perceptual representations and phonemic categories,

[1] As to the strength of the categorical effects in imitation, it became clear in our 1987 study that the /i/–/æ/ continuum used (based on Kent (1973) and indirectly on Peterson and Barney (1952)) did not pass through all of the subjects' relevant phonemic categories for isolated vowels. This may be one reason why categorical effects did not emerge as strongly as in the study by Chistovich et al. (1966), whose continuum was modelled directly on the subject's production of isolated vowels.

so that perception and production are hopelessly intertwined, a theoretical separation of these levels may be of some heuristic value.

There is a third possible explanation, which attributes the categorical effects to the conversion of vocal tract shapes into acoustic output. Stevens' (1972) well-known "quantal" theory of vowel production postulates discontinuities in that conversion, though probably not enough of them to account for all trends in imitation. Nevertheless, they may account for some of them, and if so, this would be important to know. In addition, Stevens' theory makes predictions about the acoustic variability of vowel productions in different regions of the vocal tract, although the empirical evidence bearing on these predictions is somewhat conflicting (Perkell and Nelson, 1985; Pisoni, 1980).

The present study attempted to assess the latter two, production-based accounts. Since every vowel to be imitated must first be heard and perceived, it is very difficult to think of an imitation experiment that might rule out a perceptual explanation of categorical tendencies. It is possible to produce vowels, however, according to an internally generated plan, without a perceptual model. Of course, this is what happens all the time in speech production. In normal speaking, however, as well as in most speech production experiments in the laboratory, the talker's task is to produce discrete utterances from his or her language inventory. While the articulation of vowels, in particular, may change substantially across different contexts, for a fixed context a talker usually has only one optimal target – his or her language norm. In the case of isolated vowels, the targets would be the vocal tract shapes and sounds corresponding to /i, ɪ, e, ɛ, æ/, etc. Talkers are rarely required to produce something "between" the categories of their language, except in certain imitation tasks, where perceptual models are provided. The present study is novel in that it required subjects to produce vowels along an /i/–/æ/ continuum *without* any perceptual models.

To facilitate this unusual task, two steps were taken. First, the vowels were always to be produced in order, going from one endpoint of the continuum to the other. Thus the successive targets along the continuum were self-generated rather than designated by some visual-spatial code. Second, the experiment proceeded in five stages, starting with the production of continuous glides from one vowel endpoint to the other, first at a fast, then at a slower, and finally at a very slow speed; this was followed with a slow glide with interrupted phonation, and only then did the subjects attempt to produce discrete vowels along the continuum. This progression of tasks not only provided practice for the subjects and enabled them to map out the articulatory space between the continuum endpoints, but it also included two important independent variables: Fast versus slow rate, and continuity versus discreteness. If there are discontinuities in the conversion from vocal tract shape to the acoustic output, they should manifest themselves independently of these variables. That is, they should be evident even in fast, diphthong-like glides. If discontinuities arise in motor control, however, they should depend on the relative speed and continuity of the movement.

Since the task is so unusual, there is no theory of speech production that predicts how subjects generate utterances of the kinds described. It is clear, however, that the likelihood of "phonemic guidance" increases from stage 1 to stage 5. When rapidly gliding from one vowel to another, even across a wide distance in vowel space, subjects probably need only set up motor instructions specifying the endpoints and let the articulators find a natural path between them. At the slower speeds, however, deliberate control over the articulatory path is required, and as the articulators proceed through their tortuous journey, articulatory constraints or habits may modulate their local velocity or cause deviations from the trajectory. An appropriate analogy would be a steel ball rolling through a field of magnets at different speeds: At a fast speed, the ball may roll in a straight path, but at a slow speed it may depart from it in the direction of the magnetic forces. Discontinuous phonation may be an additional factor that invokes articulatory habits and thus increases any discontinuities. Finally, the production of discrete vowels should exhibit such tendencies most clearly. The question, then, was at what stage in the game any acoustic discontinuities would emerge.

Although the search for discontinuities in the acoustic signal was the main purpose of this study, the novel task offered the opportunity of making additional observations relevant to questions of articulatory control in general. Thus it was of interest whether subjects would follow precisely the same articulatory path regardless of speed and regardless of direction (from /i/ to /æ/ vs. from /æ/ to /i/), and how their continuous productions would relate to their discontinuous ones, including the production of phonemic vowel "prototypes" in isolation. Because of the laboriousness of the data analysis, only a small sample of subjects was tested that, however, represented a wide range of age and experience. This was intended both to increase the generality of any consistent findings and perhaps also to provide clues to possible effects of phonetic training and experience.

2. Methods

2.1. Subjects

Four subjects, three men and one girl, participated in this study. Subject DW was a graduate student who had been a subject and collaborator in the earlier vowel imitation studies (Repp and Williams, 1985, 1987). He is a native speaker of American English and does not speak any other language fluently. He has had some phonetic training. Subject BR was the author, who is a native speaker of German but fluent in English, which he has spoken almost exclusively for many years, though with a noticeable foreign accent. He has not had any formal phonetic training and, though very experienced as a listener in speech perception tasks, has rarely served as a subject in speech production experiments. Subject AA was a senior phonetician with much experience in the controlled production of speech sounds. He is a native speaker of American English but has command of several other languages. Subject MR was a child, the author's daughter, who was 9.6 years old at the time of testing. She is a native speaker of American English and has had little exposure to other languages, although both parents are nonnative speakers with different foreign accents.

Subjects DW and BR were included because they had served in the earlier experiments and also because it was important to have subjects who "knew what they were doing", since categorical tendencies in production were to be *avoided* as much as possible. Subject AA was included as a presumable expert in articulatory control, even though the task was rather novel to him also.[2] Subject MR was included because of her ready availability and eagerness to participate. While it may seem that the task should be even more difficult for a child, this need not necessarily be so. When it comes to avoiding categorical tendencies, someone with more articulatory flexibility and fewer ingrained habits may have an advantage. The possibility of exploring this issue was considered worth the risk of encountering problems in obtaining accurate formant frequency estimates from the child's voice.

2.2. Recording procedures

All recording was done in a sound-insulated booth using a Sennheiser microphone and an Otari MX5050 tape recorder located in an adjacent booth. The subject sat in front of the microphone, facing a Tektronix oscilloscope placed on the same table about two feet to the right. The variable speed of the oscilloscope beam moving cyclically across the screen was used to pace the subjects' utterances. Subject MR, the child, was coached through the session by the author who was with her in the booth.

In *stage 1* the task was to produce rapid glides from /i/ to /æ/. The oscilloscope beam was set to sweep across the screen in 0.5 s (i.e., at a rate of 2 Hz). The subject was instructed to time the onset of each utterance with the emergence of any beam at the left edge of the screen and to terminate the utterance when the beam reached the right edge. After some practice, 10 vowel glides were recorded, at intervals of at least 1.5 s (three intervening sweeps).

In *stage 2*, the oscilloscope beam was set to

[2] AA initially misunderstood the experiment as being a test of his expertise in producing fine *categorical* distinctions; once this became clear from the data analysis, the experimental session was repeated. Only the results of this second session are reported.

sweep across the screen in 2 s. The instructions were the same, and 10 productions were recorded after some practice, separated by at least 2 s (one intervening sweep).

In *stage 3*, the oscilloscope beam was set to sweep across the screen in 5 s. The instructions remained the same. Ten productions were recorded after some practice, separated by at least 5 s (one intervening sweep).

In *stage 4*, the oscilloscope sweep speed was kept at 5 s, but the beam was set to pulse on and off as it traversed the screen, at a rate of approximately 2.5/s. Thus there were about 12 pulses during one sweep, with equal duty cycles. The task was to move slowly from /i/ to /æ/, as in stage 3, but to phonate only when the beam was on (i.e., to produce short vowels along the vowel glide trajectory in synchrony with the light pulses). After some practice, 10 productions were recorded, at least 5 s apart.

In *stage 5*, the beam was set back to continuous and to the 2 s speed, and subjects were instructed to produce as many individual vowels as they could between /i/ and /æ/. It was stressed that each vowel should be closer to /æ/ than its predecessor and that the mouth should be *closed* after each production. Subjects were told to start with /i/ and to stop when they reached /æ/. After some practice, 10 vowel series were recorded, with at least 2 s (one intervening sweep) between individual vowels.

Following a short break, the five tasks were repeated going from /æ/ to /i/. At the end of the experiment, each subject produced 10 repetitions of five "prototypical" isolated vowels. The adult subjects read at their own pace from a list containing the phonetic symbols /i, ɪ, e, ɛ, æ/ 10 times in random order. For the child, the author pointed in the same random order to the vowel letters on a poster showing the words BEET, BIT, BAIT, BET, BAT. The task of producing the vowel sounds in isolation was readily understood by the child.

Although each subject did his or her best to follow the instructions, there were some deviations, mostly in stage 4, which will be mentioned below. On the whole, the experiment was stressful from a phonatory viewpoint, and it provoked much vocal fry and clearing of the throat between utterances.

2.3. Acoustic analysis

All utterances were digitized at a sampling rate of 10 kHz, low-pass filtered at 5 kHz, and stored in computer files. Subsequently, they were subjected to LPC analysis (ILS package, Version 4.0, distributed by Signal Technology, Inc.) using 14 coefficients and a 20 ms Hamming window moving in 10 ms steps across each file. Formant frequency estimates for the vowel glides (stages 1–4) were obtained with the root-solving method. This turned out to be crucial because the alternative peak-picking routine contained in ILS was found to generate artifactual steps in the formant tracks, due to its fixed-size Fourier transform algorithm. That peak-picking routine, however, was used for the discrete vowel productions (stage 5 and prototypes) because of its greater speed. The errors introduced in the analysis of relatively steady-state vowels seem to be minimal (cf. Repp and Williams, 1987: fn. 2).

Further analysis used software written for that purpose by the author. The result of the LPC analysis of each utterance was an array containing estimates of the frequencies of the first four formants (*F*1–*F*4) and overall amplitude (in dB) at 10 ms intervals. A well-known problem, however, is that there are often missing or extra formant peaks in LPC estimates, so the frequency estimates needed to be assigned to their correct slots. A Fortran program was written for this purpose and modified during analysis as required. The program featured an amplitude threshold to exclude silence preceding and following an utterance (the gaps in the discontinuous stage 4 glides were left intact), estimates of formant starting frequencies to "catch" the right formants at the beginning, reasonable upper and lower limits for *F*1–*F*3, and computation of running means of each formant (usually over the five preceding frames, although this "memory span" was a variable) against which the formant estimates in each frame were compared. The program rearranged the frequency estimates in each frame as required and set extra or out-of-range values to zero. For the discrete vowel productions of stage 5 and the prototypes, no starting frequencies were specified; instead, the program computed preliminary formant means for each vowel and then

made a second pass through the array comparing all values against these means and rearranging or zeroing out any values that deviated by more than one standard deviation before recomputing the means and standard deviations.

To be able to compare the formant trajectories for utterances of different duration (stages 1–4), the cleaned-up formant estimates of each production were time-normalized by linear expansion or compression into 100 time frames. These time-normalized formant tracks were displayed on a screen and each formant of each token was examined for abnormalities or unexpected discontinuities. These were examined and traced back to unusual constellations of LPC formant frequency estimates or insufficiencies in the formant tracking program, and adjustments were made until all such errors were eliminated. In a small number of cases, this required direct intervention with the LPC data, usually by zeroing out formant values that seemed unreasonable and/or misled the formant tracker. Once all tokens of an utterance type had been cleaned up, average formant values and standard deviations across the 10 tokens were computed for each of the 100 time frames, which then could be displayed as formant tracks over time or as paths in two-dimensional formant space. In averaging across stage 4 tokens, which posed the most difficult analysis problem because of the silent gaps in them, a minimum of 5 nonzero formant values was required for an average to be computed. Since the discontinuous vowels of these productions were not precisely aligned in time, despite the subjects' efforts, this criterion had the effect that averages were computed only for regions of considerable energy overlap across tokens.

For the discrete vowels of stage 5 and the prototypes, average formant frequencies were computed for each vowel, assuming it to be essentially steady-state. The within-vowel standard deviations were inspected to detect errors, which were then cleared up. Grand average formant frequencies and standard deviations were computed over the 10 tokens of each prototype. These vowel formant frequencies could be displayed as points (with or without standard deviations) in two-dimensional formant space.

3. Results and discussion

3.1. Vowel glides at three rates

Table 1 shows that actual average durations and standard deviations of the vowel glides intended to be 500, 2000, and 5000 ms long. As can be seen, they were generally in the right ballpark; maximum accuracy was not essential. The only systematic deviations occurred at the fast rate, where all subjects produced utterances that were 100–300 ms too long. The reason for this overshoot is unknown, as the optic timing signal was accurate.

The time-normalized average formant tracks ($F1$ below, $F2$ and $F3$ above) are shown in Figs. 1 (from /i/ to /æ/) and 2 (from /æ/ to /i/). Rows represent the three rates of production, columns the individual subjects. The faint dots accompanying each formant track represent plus/minus one standard deviation over the 10 repetitions.

Two measurement problems should be pointed out right away. For AA, measurement of $F3$ was difficult or impossible at the /i/-end of the con-

Table 1
Average durations and standard deviations (in parentheses) of continuous vowel glides (in ms)

Direction	Rate	Subjects			
		DW	BR	AA	MR
/i/–/æ/	Fast	762 (34)	702 (26)	716 (99)	798 (27)
	Medium	2176 (92)	1991 (61)	2511 (80)	2073 (101)
	Slow	5210 (287)	5009 (211)	5360 (224)	4684 (464)
/æ/–/i/	Fast	797 (87)	602 (39)	582 (43)	769 (30)
	Medium	2104 (152)	2041 (128)	2204 (91)	2376 (152)
	Slow	4982 (377)	4679 (223)	5033 (317)	5387 (307)

tinua. (Note also the unusual convergence of *F*2 and *F*3 for this speaker, especially in the middle of the /i/–/æ/ glides.) For MR, the child, the *F*1 tracks in /i/–/æ/ glides started consistently with a relatively high plateau, which may represent the first harmonic of the fundamental rather than *F*1 itself; note, however, that *F*2 shows a corresponding (sloping) plateau, and that *F*1 in /æ/–/i/ glides showed a more typical pattern. While the child's *F*1 tracks should perhaps be regarded with caution, her *F*2 and *F*3 generally proved measurable, even though they exhibited considerable variability in some conditions.

A general feature of these formant plots is that the speakers did not move from one endpoint vowel to the other at a constant rate. This is especially evident at the fast rate, where the transitional movement typically occupied less than half of the total time, leading to sigmoid-shaped formant tracks. The increased variability in the transitional region reflects the fact that the transition occurred at somewhat different times in different repetitions; the average slope was in fact steeper and less variable than is suggested by the figure. In the slower rate conditions the formant changes tend to be more gradual, but they still deviate from linearity. These global deviations reflect general movement strategies, not categorical tendencies.

Categorical tendencies would show up as steps (local plateaus) in the formant tracks, perhaps in all three formants simultaneously. It is difficult to discern such steps in Figs. 1 and 2. The clearest examples occur in DW's productions at the slow rate. Several other trends can be seen if the dotted (standard deviation) lines are examined closely. However, one reason why steps are difficult to see is that they may have occurred at different times in different repetitions, so that they were blurred by the averaging over tokens. One way to get around this problem is to examine individual utterances (tokens), as was in fact done in the course of data analysis. A more efficient way is to consider the *distributions of formant frequencies* for all tokens combined. This representation of the data also makes comparison with the earlier vowel imitation data possible, which were treated in the same way (Repp and Williams, 1985, 1987). If a number of tokens show plateaus at the same frequencies, then there should be peaks at these frequencies in the distribution. If there are no plateaus, the distribution should be even. More precisely, since the subjects tended to dwell on the endpoint vowels, especially at the faster rates, the absence of plateaus should be manifested in a relatively smooth valley between two major peaks at the ends of the distribution.

These distributions are shown in Fig. 3. The upper panels represent *F*1 distributions and the lower panels represent *F*2 distributions; *F*3, which is less informative, has been omitted for simplicity. In each graph, the distributions for the three rates of production are superimposed. Each curve represents the envelope of a 38-bin histogram comprising the relevant frequency range, which is different for each subject. Each histogram is based on close to 1000 frequency values. The highest peaks are cut off in the figure, as their precise height is not of interest here. Note that the *F*1 and *F*2 graphs are not aligned with each other, and while /i/ is on the left and /æ/ on the right in the *F*1 plots, their positions are reversed in the *F*2 plots.

Consider subject DW first, who had shown fairly strong categorical tendencies in the earlier imitation studies. His fast productions are quite smooth, and no major peaks are evident between the endpoint peaks, if the twin *F*2 peak for /æ/ in /i/–/æ/ glides is ignored. At the medium rate, however, additional peaks emerge on the /i/–/æ/ continuum, a striking narrow peak in *F*1 and a broader peak in the *F*2 distribution. The /æ/–/i/ medium-rate productions remain fairly smooth. At the slow rate, there are multiple peaks in the /i/–/æ/ *F*1 and *F*2 distributions, as well as in the /æ/–/i/ *F*1 distribution. This subject, therefore, shows clear evidence for discontinuities in slow-rate vowel glides.

Subject BR shows less pronounced peaks, on the whole, than does DW. Especially along the /i/–/æ/ continuum, the distributions are fairly even, except for an unexpected extra *F*1 peak close to /i/ at the fast rate. Along the /æ/–/i/ continuum, however, clear *F*1 peaks emerge at the medium and slow rates, as well as an *F*2 peak at the medium rate. Thus this subject, too, shows some evidence of plateaus in the formant tracks.

Subject AA was an experienced senior phone-

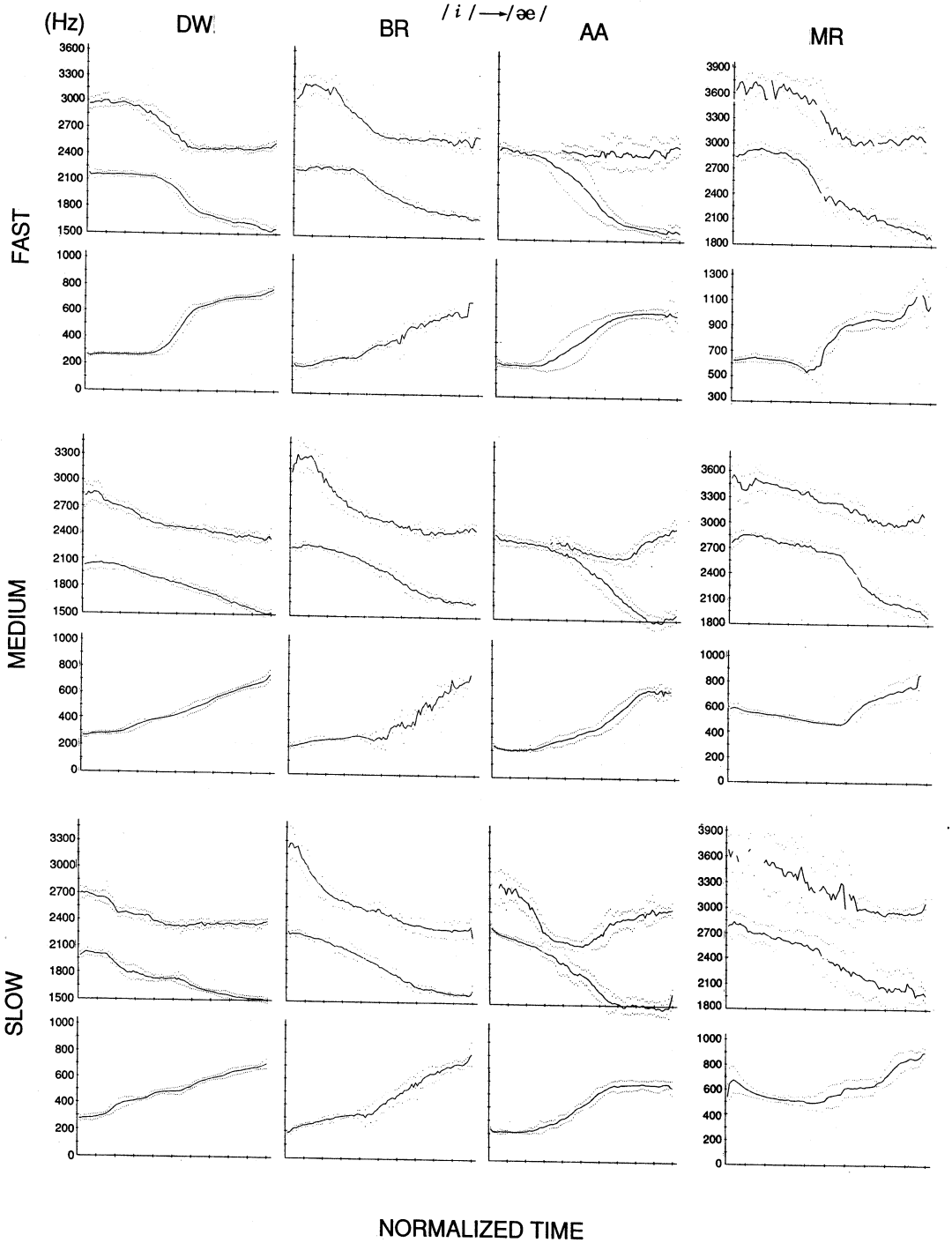

Fig. 1. Time-normalized formant trajectories ($F1$ below, $F2$ and $F3$ above) of /i/–/æ/ vowel glides produced at three different speeds. The faint dotted lines represent plus/minus one standard deviation.

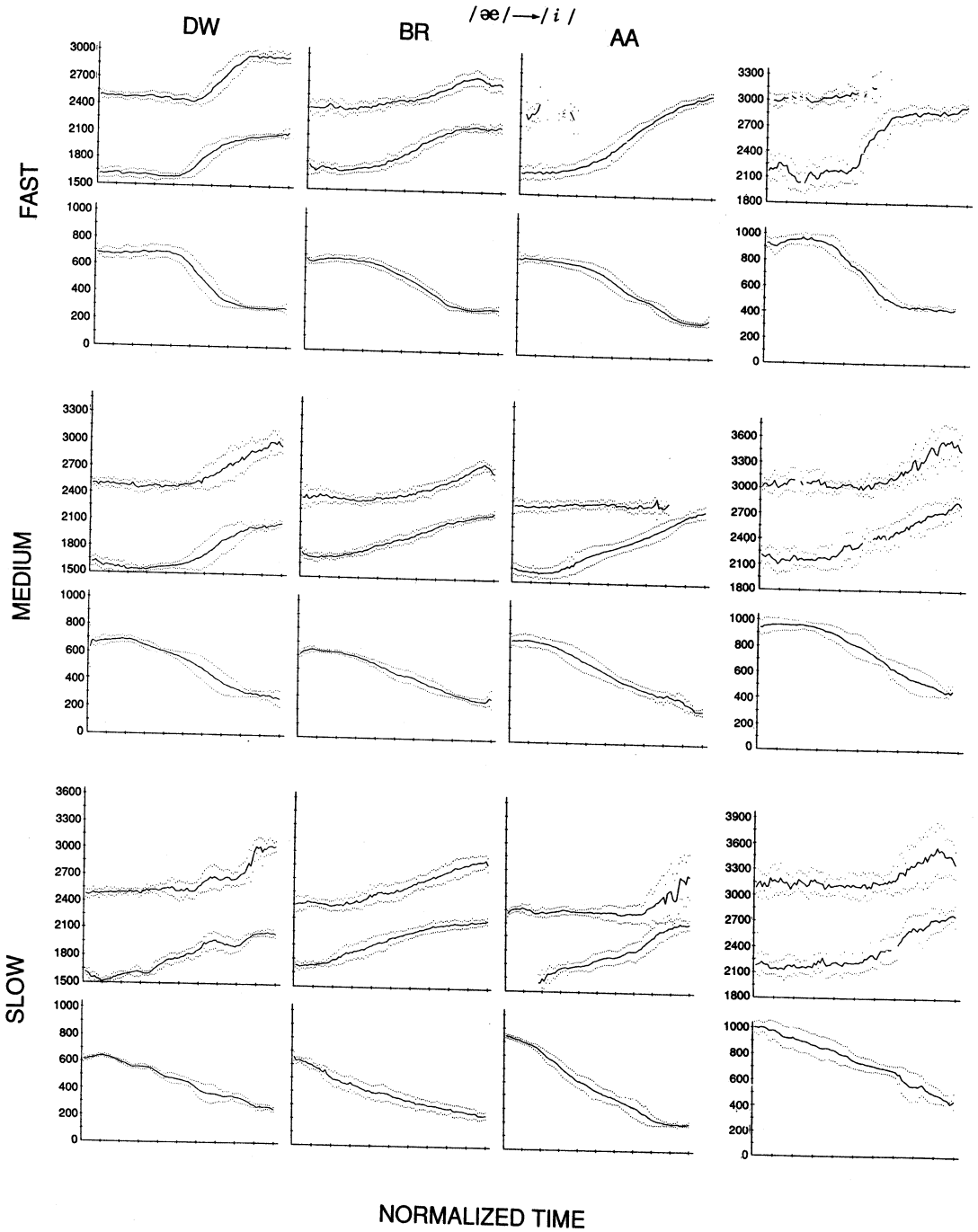

Fig. 2. Formant trajectories of /æ/–/i/ vowel glides.

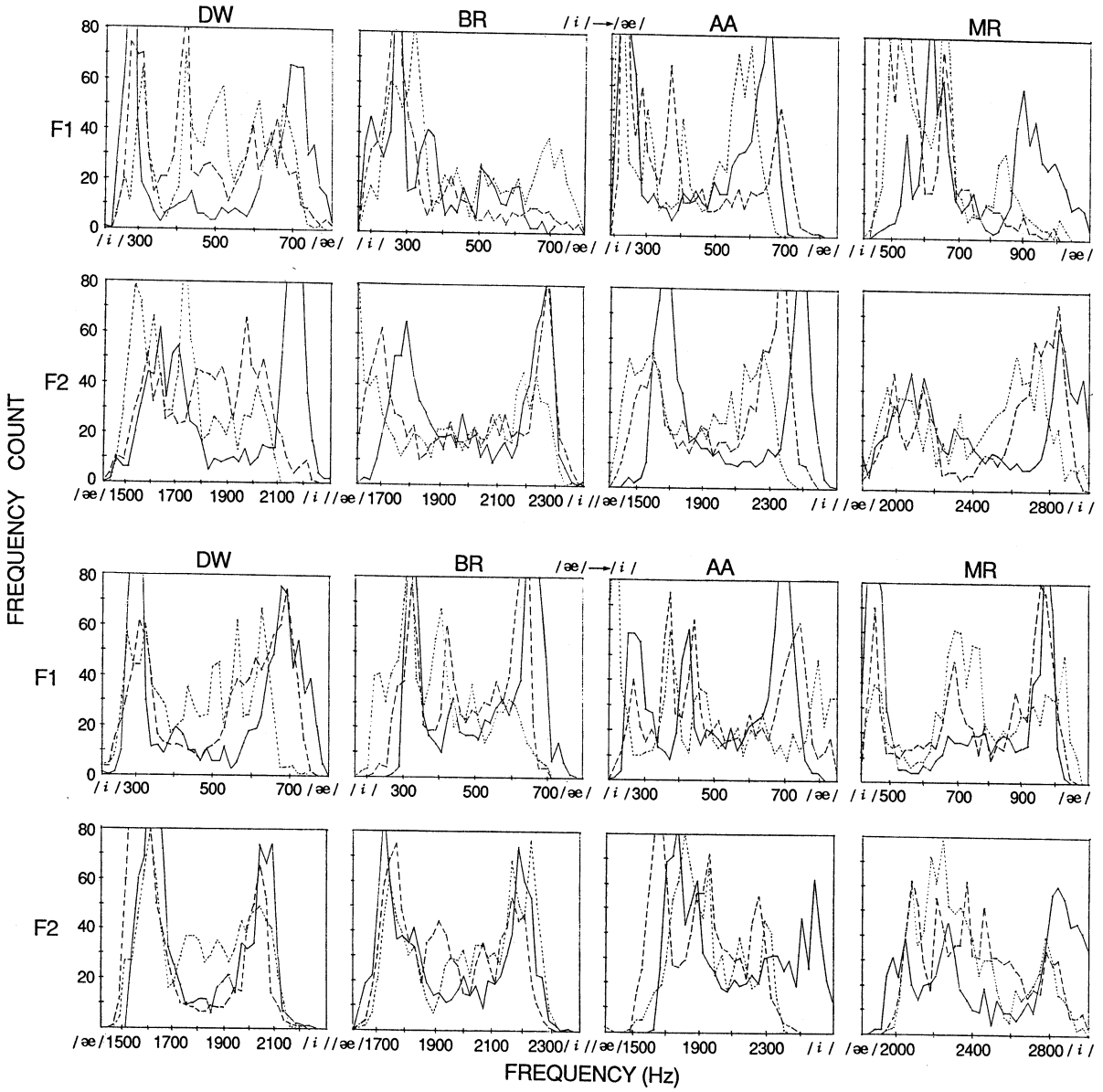

Fig. 3. *F*1 and *F*2 frequency histograms for vowel glides produced at three rates (fast: solid line; medium: dashed line; slow: dotted line).

tician. If anyone could avoid discontinuities in vowel glides, it was thought it would be he. While his fast /i/–/æ/ glides are very smooth, there is a striking extra *F*1 peak in his fast /æ/–/i/ glides and an emerging peak in the *F*2 distribution as well. At the medium speed, all distributions except *F*2 for /i/–/æ/ glides show multiple peaks, and the same is true at the slow speed. Thus this speaker likewise shows strong evidence of plateaus in the formant tracks, in one instance even at the fast speed.

Finally, we turn to MR, the child. She already shows some discontinuities at the fast rate, which become more pronounced at the medium and

slow rates. The *F*1 distributions for the /æ/–/i/ glides provide a nice example of an emerging mid-frequency peak as a function of rate.

Table 2 presents an informal count of extra peaks (i.e., not including the endpoint peaks). It may be seen that the number of peaks increased as rate slowed down; that there were more *F*1 peaks, on the average, than *F*2 peaks; that there was not much difference between /i/–/æ/ and /æ/–/i/ glides in this respect; and that the child tended to show *more* peaks than the adults. This last result suggests that the child, rather than being more flexible in her phonetic skills, had more trouble avoiding phonemic influences (if that is what caused the peaks).

Figure 3 contains additional information about influences of glide direction and rate on the distribution of formant frequencies. Peaks at different rates often did not coincide. The peaks representing the endpoint vowels, in particular, tended to shift with rate. There was some consistency among the four speakers, especially with regard to *F*2 in /i/–/æ/ glides: the *F*2 peaks for both endpoint vowels shifted toward lower frequencies as rate decreased. These effects, and especially effects of direction of movement, are easier to see in plots of formant tracks in *F*1–*F*2 space. Such plots also reveal whether or not plateaus in *F*1 and *F*2 occur at the same time. If they do, they would not change the path through *F*1–*F*2 space but only reduce the velocity of movement along it. If they occur at different times, however, deviations from the path should be evident. Such deviations are predicted by the analogy of "magnetic attraction" from prototypes.

Table 2
Number of extra peaks observed in the histograms of Figure 3. (Half a count is made for small or doubtful peaks.)

Direction	Rate		Subjects				
			DW	BR	AA	MR	Sum
/i/–/æ/	Fast	F1	0	2	0	1.5	3.5
		F2	0.5	0	0	0.5	1
	Medium	F1	1.5	0	1.5	2	5
		F2	1	0	0	1	2
	Slow	F1	3	1.5	2	1	7.5
		F2	1.5	0	0	0.5	2
/æ/–/i/	Fast	F1	0	0.5	1	0	1.5
		F2	0	0	1	1	2
	Medium	F1	0	1	2	1	4
		F2	0	1.5	1	3	5.5
	Slow	F1	2.5	1.5	2	1.5	7.5
		F2	0.5	1	1	1	3.5
Sum			11.5	9	11.5	14	

Figure 4 shows these paths in *F*1–*F*2 space, together with the locations and standard deviations of the prototype vowels, as produced by the subjects at the end of the experiment. The most striking effect was the shift in the trajectory as a function of rate. Vowel glides produced at the slower rates were more centralized (i.e., had lower *F*1 and *F*2 frequencies) than those produced at the fast rate. In some instances, most notably for subject AA, there was also a difference between the medium and slow rates. The starting points and endpoints, especially the latter, varied with rate in the same way. Steps were observed in some of the trajectories, usually together with slowed velocity (visible at the slow and medium rates as concentrations of dots or dashes). In other instances, velocity changes seemed to exist without accompanying deviations, indicating coincidence of *F*1 and *F*2 plateaus. Disappointingly, however, these discontinuities bore no clear relationship to the prototypes, nor did the trajectories themselves. In some instances (e.g., subject AA at the slow rate) the trajectory was quite far from the prototype locations. Thus there is little support for the "magnetic attraction" theory here.

The trajectories were quite linear in a number of cases, particularly at the fast rate of production. Some striking deviations from linearity occurred in the /i/–/æ/ glides of subjects BR and MR in the form of a "knee" or bulge in the /ɪ/ region. It is suspected that this deviation was caused by a /j/-like constriction in moving from /i/ to /æ/; indeed, /ijæ/ seems a more natural utterance than /æji/. The articulatory strategy of these two subjects (father and daughter!) may have been to lower the tongue body (mainly affecting *F*2) before lowering the jaw. Subjects DW and AA, on the other hand, may have moved their jaw throughout, with or without additional tongue movement. It may be noted here that it is possible to produce a reasonable /i/–/æ/ continuum with either tongue or jaw movement alone, with the

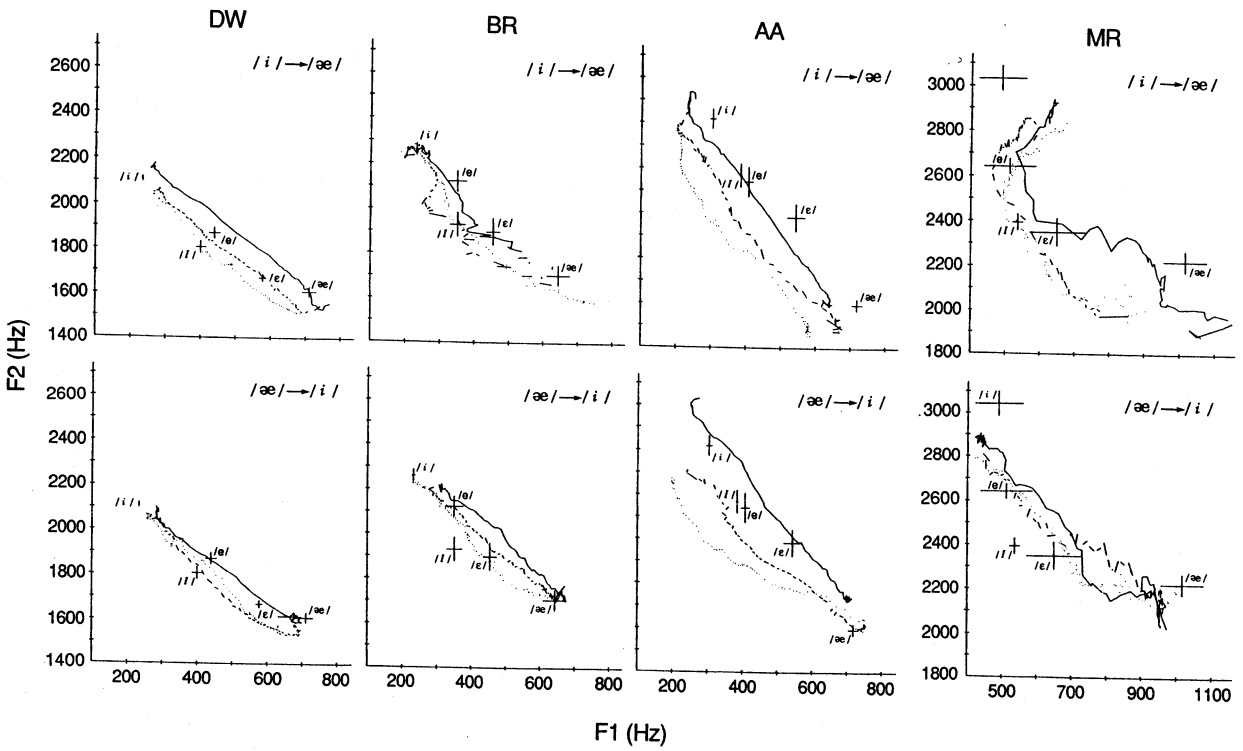

Fig. 4. Trajectories of vowel glides at three rates (fast: solid line; medium: dashed line; slow: dotted line) in *F*1–*F*2 space. The spacing of the dashes or dots reflects the velocity of the glide. The crosses represent the prototypical vowel productions plus/minus one standard deviation along each dimension.

other articulator fixed at an intermediate position. (To what extent each of the two articulators contribute to the observed trajectories is an interesting question for further research.) Differences between the /i/—/æ/ and /æ/–/i/ trajectories also occurred for subjects DW and AA, but they were less striking and not consistent with each other.

3.2. *Interrupted vowel glides*

These data can be dealt with very briefly. Inspection of formant frequency histograms suggested that whatever categorical tendencies were already present in the continuous vowel glides were not noticeably enhanced in the interrupted productions. Moreover, there were fewer data points available, some measurement problems at the edges of the phonated segments, and an irregularity in the productions of subject DW, who did not really stop phonating between vowel segments; therefore, these data were not very tidy. Compared to the long (5 s) continuous glides, the interrupted productions of equal duration exhibited higher *F*2 values for subjects DW, MR (/i/–/æ/ direction only), and especially AS, but lower *F*2 values for BR on the /i/–/æ/ continuum. All subjects except BR exhibited an "approach effect" in that changes between successive vowels were initially large and then grew smaller as the endpoint of the continuum was approached. Only subject DW showed some discontinuities in the *F*1–*F*2 trajectories that could be related to his prototype categories, but perhaps this was caused by his quasi-continuous phonation.

3.3. *Discrete vowel productions*

The discrete vowel productions were analyzed in the same way as the prototype productions, that is, by computing a single set of average for-

mant frequencies for each vowel. The data from this condition thus consisted of points in $F1$–$F2$ space.

Before inspecting these point swarms, the number of vowel productions per talker is of interest. The instructions were to produce as many discrete vowels as possible between the specified endpoints, and to stop when the second target was reached. The average numbers of vowels produced by each talker, and the ranges across the 10 repetitions, are shown in Table 3. BR was less adept at this task than were DW and AA, but the most striking difference is between the child, MR, and the three adults. Even though the instructions were clearly understood by MR, she was never able to produce more than six discrete vowels along either continuum.

Figure 5 shows the discrete vowel productions as points in $F1$–$F2$ space, together with the prototype productions (crosses). These plots show that a wide variety of vowel qualities was produced by the subjects. In general, vowels were more widely spaced apart at the beginning of the continuum and closer together towards the end (the "approach effect" again). There was relatively little prototype-related clustering of responses; possible exceptions are /i/ in BR's productions and /ɛ/ in AA's productions. AA produced a number of vowels with unusually low $F2$. MR's strategy appeared to be to make a big jump following the initial target and then to vary only one formant at a time, depending on the continuum ($F1$ on the /i/–/æ/ continuum, $F2$ on the /æ/–/i/ continuum). To the extent that variation in

Table 3
Average number (range in parentheses) of vowels produced in the discrete vowel production task

Direction	Subjects			
	DW	BR	AA	MR
/i/–/æ/	11.7 (10–14)	8.7 (7–11)	10.6 (9–13)	4.7 (4–6)
/æ/–/i/	14.3 (11–18)	8.0 (7–10)	13.0 (12–14)	4.8 (4–6)

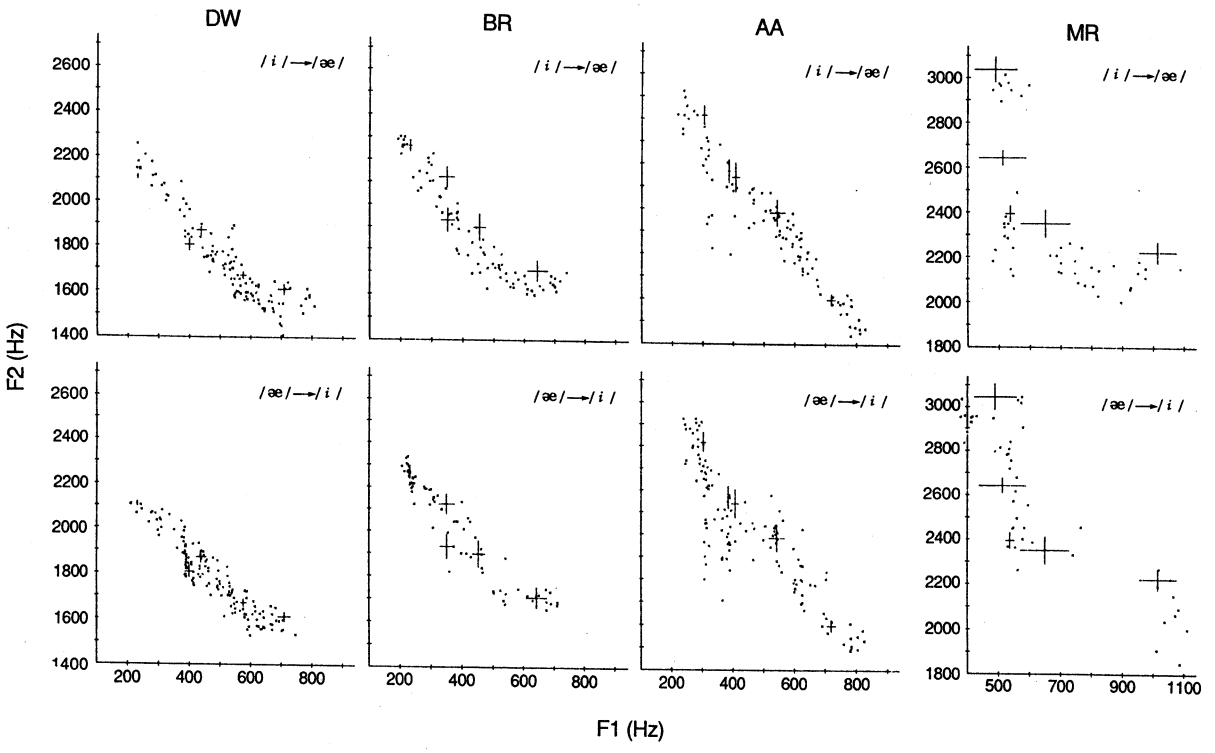

Fig. 5. Scatter plots of discrete vowel productions in $F1$–$F2$ space. Also shown are the prototypical vowel productions as crosses; see Fig. 4 for their labels.

these formant frequencies is primarily associated with jaw and tongue position, respectively (Lindblom and Sundberg, 1971), it may be inferred that the child preferred to vary only one articulator at a time. If so, this may imply a certain difficulty in interarticulator coordination in this novel task.

Because of the relatively small number of data points, no formant frequency histograms were derived from these data. Nevertheless, it is evident from several of the plots that certain *F*1 frequencies occurred more often than others. Note, for example, the "column" extending upwards from the /ɪ/ prototype in DW's /æ/–/i/ continuum data, the /i/ column in BR's corresponding plot, and the vertical structure in AA's data. Unless these tendencies are due to some as yet unrevealed artifact of analysis, they suggest, in agreement with the slow glide productions, that there are certain preferred *F*1 frequencies, i.e., jaw positions, in isolated vowel production. Moreover, the data are not inconsistent with the proposition that these jaw positions are related to the prototypical vowels, with deviations from the prototypes being caused mainly by tongue position.

3.4. Prototypical vowels

As to the prototype productions themselves, some individual differences are noteworthy. For all speakers, the vowels /ɪ/, /ɛ/, and /æ/ fell approximately along a straight line in *F*1–*F*2 space, with /i/ lying above that line. This is predicted by Peterson and Barney's (1952) classical norms. The child's /i/ productions, however, had an unusually high *F*2 relative to her other vowels, and in fact her vowel glides never came near the /i/ prototype. The formant frequencies of the four vowels named are compatible with the Peterson and Barney data, taking into account differences in vocal tract size. Subject AA, however, commanded an unusually wide range of formant frequencies for a male, comparable to that of the average female speaker. Individual differences were notable with respect to the fifth vowel, /e/, which (being a diphthong in standard American English) was not included in the Peterson and Barney norms. Subjects DW and AA, both native speakers, made only a minimal distinction between (non-diphthongized) /e/ and /ɪ/, which suggests that for them the distinction between these two vowels rests primarily on the presence or absence of formant movements. Subject BR, however, who is a native speaker of German, a language that has a monophthongal /e/, clearly distinguished that vowel from /ɪ/ by raising its *F*2. Interestingly, subject MR, BR's daughter but a native speaker of English, did the same, suggesting some possible influence of exposure to her father's (and mother's) foreign accent.

4. Conclusions

The present results suggest that categorical tendencies – steps in formant tracks and peaks in formant frequency distributions – similar to those observed in vowel imitation tasks emerge even in continuous glides between two vowel targets, provided the rate of the glide is sufficiently slow. At a fast rate approaching that of a diphthong in natural speech, the articulatory trajectory is presumably of a ballistic nature and requires no active control during its execution. In general, the acoustic output was smooth also, suggesting that there were no "quantal" steps in the articulatory-to-acoustic transform. At slower speeds, however, the talker must exert continuous control over the steady progress of articulatory movement, which apparently leads to hesitations and deviations.

These hesitations and deviations appear to be similar in nature to the categorical tendencies observed during the production of discrete vowels, and both may be caused by "attraction" to articulatory prototypes. These relationships are not straightforward, however. Discontinuities occurred at different formant frequencies depending on the rate of vowel glides, and at yet different frequencies in discrete vowel productions. If they are caused by influences of phonemic reference points, then it must be the case that either the reference system is flexible and shifts with rate and manner of production or, perhaps more plausibly, that influences of a fixed reference system become manifest at different points in a movement path, depending on the dynamics of the task.

Since categorical tendencies emerged in vocalic productions that were not guided by an auditory-perceptual model, it may be concluded that these tendencies originate not in the perceptual system but in speech production – that is, in the construction of articulatory plans and/or in the execution of articulatory movements. Yet it is possible that perceptual guidance was exerted over the utterances by means of an auditory feedback loop, especially in the slow vowel glides. Indeed, the different location of formant peaks in the different rate conditions might be attributed to a constant lag in processing the auditory feedback input. Such an explanation seems less plausible in the case of discrete vowel productions, however, which represent single discrete articulatory targets. Perception may have been involved, however, in computing the next target based on the preceding production. To rule out perceptual mediation completely, future studies will have to disturb or eliminate auditory feedback.

The present subjects included an experienced phonetician and a child. Although no firm conclusions can be derived from such individual observations, the data suggest that phonetic experience did not decrease categorical tendencies, whereas phonetic inexperience enhanced them. In addition, there was a suggestion that the child found it difficult to control two articulators at the same time, and preferred to vary either jaw height or tongue position. This issue warrants further research, as it may bear on the acquisition of coordinative structures in articulation.

Acknowledgments

This research was supported by NICHD Grant HS-01994 and BRSG Grant RR-05596 to Haskins Laboratories. I am grateful to Kevin Munhall for comments on an earlier draft of the manuscript.

References

L.A. Chistovich, G. Fant, A. de Serpa-Leitão and P. Tjernlund (1966), "Mimicking and perception of synthetic vowels", *Quarterly Progress and Status Report* (Royal Technical University, Speech Transmission Laboratory, Stockholm), No. 2, pp. 1–18.

A.F. Healy and B.H. Repp (1982), "Context independence and phonetic mediation in categorical perception", *J. Experimental Psychology: Human Perception and Performance*, Vol. 8, pp. 68–80.

R.D. Kent (1973), "The imitation of synthetic vowels and some implications for speech memory", *Phonetica*, Vol. 28, pp. 1–25.

B.E.F. Lindblom and J.E.F. Sundberg (1971), "Acoustical consequences of lip, tongue, jaw, and larynx movement", *J. Acoust. Soc. Am.*, Vol. 50, 1166–1179.

N.A. Macmillan, L.D. Braida and R.F. Goldberg (1987), "Central and peripheral processes in the perception of speech and nonspeech sounds", in *The Psychophysics of Speech Perception*, ed. by M.E.H. Schouten (Martinus Nijhoff, Dordrecht), pp. 28–45.

J.S. Perkell and W.L. Nelson (1985), "Variability in production of the vowels /i/ and /a/", *J. Acoust. Soc. Am.*, Vol. 77, pp. 1889–1895.

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

D.B. Pisoni (1973), "Auditory and phonetic memory codes in the discrimination of consonants and vowels", *Perception & Psychophysics*, Vol. 13, pp. 253–260.

D.B. Pisoni (1980), "Variability of vowel formant frequencies and the quantal theory of speech: A first report", *Phonetica*, Vol. 37, pp. 285–305.

B.H. Repp and D.R. Williams (1985), "Categorical trends in vowel imitation: Preliminary observations from a replication experiment", *Speech Communication*, Vol. 4, pp. 105–120.

B.H. Repp and D.R. Williams (1987), "Categorical tendencies in imitating self-produced isolated vowels", *Speech Communication*, Vol. 6, pp. 1–14.

B.H. Repp, A.F. Healy and R.G. Crowder (1979), "Categories and context in the perception of isolated, steady-state vowels", *J. Experimental Psychology: Human Perception and Performance*, Vol. 5, pp. 129–145.

M.E.H. Schouten (1988), "Imitation of synthetic vowels by bilinguals", *J. Phonetics*, Vol. 5, pp. 273–283.

K.N. Stevens (1972), "The quantal nature of speech: Evidence from articulatory-acoustic data", in *Human Communication: A Unified view*, ed. by E.E. David and P.B. Denes (McGraw-Hill, New York), pp. 51–66.

A.K. Syrdal and H.S. Gopal (1986), "A perceptual model of vowel recognition based on the auditory representation of American English vowels", *J. Acoust. Soc. Am.*, Vol. 79, 1086–1100.