Individual differences in vowel production

Keith Johnson,[a] Peter Ladefoged, and Mona Lindau
Department of Linguistics, UCLA, Los Angeles, California 90024-1543

(Received 26 August 1991; accepted for publication 8 April 1993)

It is often assumed that a relatively small set of articulatory features are universally used in language sound systems. This paper presents a study which tests this assumption. The data are x-ray microbeam pellet trajectories during the production of the vowels of American English by five speakers. Speakers were consistent with themselves from one production of a word to the next, but the articulatory patterns manifested by this homogeneous group were speaker specific. Striking individual differences were found in speaking rate, the production of the tense/lax distinction of English, and in global patterns of articulation. In terms of a task-dynamic model of speech production, these differences suggested that the speakers used different gestural target and stiffness values, and employed different patterns of interarticulator coordination to produce the vowels of American English. The data thus suggest that, at some level of speech motor control, speech production tasks are specified in terms of acoustic output rather than spatiotemporal targets or gestures.

PACS numbers: 43.70.Aj, 43.70.Bk, 43.70.Hs

INTRODUCTION

Most studies of speech production find some differences between speakers. In this paper, we report an x-ray microbeam study of such individual differences in the production of American English vowels.

Individual differences in the production of American English have been noted for many years (we will consider some examples which, in our view, cannot be accounted for in terms of regional or social variation).[1] For instance, Kenyon (1924, p. 54) remarked that there was disagreement among American phoneticians concerning the pronunciation of /s/, some describing an articulation in which the tongue tip touches the back of the lower teeth and some describing a production with the tongue tip higher in the mouth behind the upper teeth. This observation was confirmed much later by Borden and Gay (1979, p. 22) who described the different articulations as *tongue tip up* and *tongue tip down* stating that "these two variant articulations of /s/ are common." Similarly, Ladefoged (1982, p. 7) notes that the dental fricatives /θ/ and /ð/ are pronounced with the tongue tip protruding below the upper teeth by some Californian speakers of American English, while for others the tongue tip remains behind the upper teeth. It has also been noted many times (e.g., Heffner, 1950, p. 148; Ladefoged, 1982, p. 78) that American English speakers produce "r" in two different ways. Some speakers keep the tongue tip down and the body of the tongue is raised in a *bunched* /ɪ/, while others raise the tongue tip toward the post-alveolar region of the roof of the mouth in a *retroflex* /ɪ/. Finally, we have also found that our students report a variety of pronunciations of initial "l," some having a lateral opening on one side or the other of the tongue, and some producing "l" with a lateral opening on both sides.

As these examples illustrate, there is much anecdotal evidence of individual differences in the production of American English. However, detailed instrumental studies are rare. As mentioned above, Borden and Gay (1979) studied individual differences in the production of /s/. Also, Raphael and Bell-Berti (1975) reported EMG data on the tense/lax distinction in American English vowels (e.g., [i] vs [ɪ], [eᶦ] vs [ɛ], and [u] vs [ʊ], see also Bell-Berti *et al.*, 1978). Ladefoged *et al.* (1972) controlled for several factors which might reasonably be expected to underlie individual differences. They reported differences in the position of the jaw at the vowel midpoint in tense/lax vowel pairs, and differences in tongue height among front vowels. The speakers were all male Midwesterners, and to insure that the speakers all spoke the same dialect, the recordings were evaluated by a panel of linguists. To control for contextual effects on vowel production, the speakers were asked to read a set of /hVd/ words in the carrier phrase, "say__again."

The present study has its roots in this earlier work. Ladefoged *et al.* noticed that some of their speakers showed nearly identical jaw positions for members of tense/lax pairs, and others did not. They argued that these individual differences in vowel production show that speakers aim for auditory targets rather than articulatory targets, because in producing the "same" vowels individuals achieve different articulatory postures. However, only one repetition of each vowel per speaker was analyzed, so one cannot be certain that the observed differences do not reflect within-speaker variability rather than true individual differences.

The study of individual differences in speech production bears on a fundamental issue in linguistic phonetics. Phoneticians assume that a relatively small set of phonetic categories (*phonetic features*) describing the place and manner of oral articulation, voicing, nasality, etc. are *universally* used in language sound systems (Pike, 1943; Ja-

---

[a]Now at: Department of Biocommunication, University of Alabama at Birmingham, UAB Station, Birmingham, AL 35294-0019.

 0001-4966/93/94(2)/701/14/$6.00

kobson *et al.*, 1952; Chomsky and Halle, 1968; Ladefoged, 1971; Clements, 1985). This assumption has been remarkably useful for many years. For example, the International Phonetic Alphabet has proven to be a valuable tool in both language description and teaching as well as clinical speech pathology. Also, analyzing the sounds of the world's languages in terms of phonetic features has led to the discovery of many important cross-linguistic generalizations (Maddieson, 1984). Therefore, it is important to remember that statements like "vowels tend to be nasalized before nasal consonants" assume the cross-linguistic validity and universality of terms like "vowel," "consonant," "nasal," etc.

Phonetic features are usually defined in terms of speech *articulation* (Chomsky and Halle, 1968; Clements, 1985; but see Jakobson *et al.*, 1952) because in most cases it is simpler to define features by reference to articulation rather than acoustics (we will discuss exceptions to this generalization below). For example, "nasal" can be easily defined by reference to the opening of the velopharyngeal port, but an acoustic definition (spectral zero, nasal murmur, relative amplitude) is more complicated and given the acoustic theory of speech production (Fant, 1960) the acoustic characteristics can be derived from the articulatory definition. Place of articulation in stop consonants, to take another example, is similar because while it can be easily defined in terms of the activities of the articulators, an acoustic definition involves seemingly unrelated acoustic attributes such as the spectral characteristics of the release burst and the transitions of the vowel formants immediately before and after the consonantal occlusion.

The assumptions (1) that the same set of features is used in all languages, and (2) that those features are defined by reference to articulation, specify a hypothesis about the phonetic aspect of language. We will call this the *universal articulatory phonetics hypothesis*. In the present study, following Ladefoged *et al.* (1972), we tested the universal articulatory phonetics hypothesis by examining productions of a set of test words read in a carrier phrase by a dialectally homogeneous group of speakers. Previous tests of the hypothesis have compared "similar" sounds across languages. Reviews of research taking this approach (Ladefoged, 1980; Keating, 1985) suggest that there are indeed problems with the hypothesis, but it should be noted that "similar" sounds across languages may not provide the best test of the hypothesis because the set of phonetic features or the number of possible combinations of features may be larger than previously assumed. For instance, the fact that phoneticians have transcribed measurably different sounds in both French and English as /p^h/, may simply mean that the feature(s) which distinguish them across the languages have not been noticed before, or that the two languages have different combinations of features in the sounds [see for example Stevens and Keyser (1989) for a discussion of "enhancement" features]. Thus since cross-linguistic tests of the universal articulatory phonetics hypothesis are equivocal, the study of individual differences among speakers of a particular language may be a better way to test the hypothesis.

Phonetic similarity within language is an approachable topic of research because language is a shared system. Each speaker of a language communicates using words and sounds which are the "same" (at some level) as those used by every other speaker. The study of individual differences can help us identify the level at which the behaviors of different individuals are the "same." Because linguistic theory aims to describe the shared aspect of language, it is important to identify the idiosyncratic aspects of linguistic behavior. However, it would be a gross oversimplification to assume that the universal articulatory phonetics hypothesis predicts that all of the speakers of a language will show exactly the same patterns of articulation, but according to the hypothesis we *do* expect to be able to relate the variation we find among individuals to factors like vocal tract geometry or preferred speaking rate. In other words, although the hypothesis does not necessarily predict that every detail of articulation will be identical for all speakers of a language, it does predict that the observed variation will be lawful. Therefore, it is not enough to simply report that there are individual differences among a homogeneous group of speakers, we must also consider ways in which such variability is lawful, because this variability *must* be made to square with fact that language is a shared system.

Individual differences are also important in speech motor control theory because motor control theorists seek to define hierarchical motor control structures in which the number of free parameters, the degrees of freedom, is reduced (Sherrington, 1906; Weiss, 1941; Bernstein, 1967; Gelfand *et al.*, 1971; Greene, 1972; Easton, 1972; Turvey, 1977; Fowler *et al.*, 1980; Abbs *et al.*, 1984; Turvey, 1990). As in linguistic phonetics, the goal is to define a domain, or coordinate system, in which underlying similarities between events can be captured, and thus, as in linguistic phonetics, the study of individual differences is important. What is more, motor control theory provides a framework for investigating the lawfulness of individual differences in speech production. In order to outline the relevance of individual differences in speech motor control theory and the relevance of motor control theory for the universal phonetics hypothesis we will review a detailed task-dynamic model speech production (Saltzman and Munhall, 1989) considering the ways in which speakers might differ according to the model.

Saltzman and Munhall (1989) posit two "functionally distinct but interacting levels of coordination" in speech production. At the *interarticulator level*, gestural units "serve to organize the articulators ... into functional groups or ensembles of joints and muscles (i.e., synergies) that can accomplish particular gestural goals" (pp. 336–337). Gestures are defined as spring-mass dynamical equations (with target and stiffness parameters) which create and release vocal tract constrictions (we disregard here a possibly crucial problem with "constrictionless" vowels like [æ] in English). A projection maps the gestural coordinate system, defined in terms of vocal tract constrictions, to a coordinate system defined in terms of articulators.

One of the most important properties of Saltzman and Munhall's task dynamic model is that invariant control structures produce variable movements. Thus the task dynamic model achieves flexibility in articulatory performance while maintaining invariance in organization. In view of the topic of this paper, the logical next question is, "Can a task dynamic model account for variation between speakers as well?" Just as an invariant system of control may underlie variable movements in an individual's speech, so individual differences in production may reflect the responses of a shared system of motor control to the unique constraints imposed by differing vocal tract geometries. In terms of the universal phonetics hypothesis, this might offer a way to define universal phonetic features in terms of articulatory constants which when realized by different individuals exhibit lawful variation.

This paper presents a preliminary investigation of the lawfulness of individual variation in the production of vowels in American English. The analysis is limited in scope, but we hope to identify some issues for future research.

I. METHOD

The data were collected at the X-ray Microbeam facility at the University of Wisconsin (Fujimura *et al.*, 1973; Kiritani *et al.*, 1975; Abbs *et al.*, 1988). Some aspects of these data have been reported previously (Lindau and Ladefoged, 1989, 1990; Johnson, 1991).

A. Subjects

Five speakers (2 males and 3 females) of midwestern American English served as speakers in the experiment. They were paid a small sum for their participation and were recruited by the staff at Wisconsin from the university community. The subjects were unaware of the specific purposes of the experiment, and reported no history of speech or hearing deficiencies and had no dental fillings. We evaluated the dialect homogeneity of the speakers by having them read several sets of dialect diagnostic words (e.g., "merry," "Mary," and "marry," "cot" and "caught"). One of the male speakers (RP) did not distinguish between [ɑ] and [ɔ], so he did not read the [ɔ] words.

B. Materials

The speakers read sentences containing the vowels /i ɪ eᶦ ɛ æ ɑ ɔ ʌ oᵘ ʊ u/ in three different contexts: (1) between alveolar stops /d__t/ or /d__d/, (2) between bilabial stops /b__b/, and (3) between alveolar fricatives /s__s/. These contexts were chosen so as to cause the vowels to be between consonants that make very different demands on the position of the jaw. Not all of these sequences were real words in English and so the subjects were instructed in the pronunciation of the non-English sequences by pointing out words which rhyme with the test sequences. For instance, "suss" [sʌs] rhymes with "fuss" in this dialect. In an attempt to balance the demands of using actual words with the desire for a factorial experimental design, some of the words had CVC structure and some had CV structure with the following word in the carrier phrase supplying the final C of the sequence. For example, instead of being asked to read, "Say sace (/seᶦs/) serenely," the subjects read, "Say say serenely." A full list of the sentences is presented in Table I.

TABLE I. List of materials.

(1) alveolar stops Say dee to me.	(2) bilabial stops Say bee between.	(3) alveolar fricatives Say see serenely.
did	bib	sis
day	bay	say
dead	beb	cess
dad	bab	sass
Dodd	bob	soss
daw	baw	saw
doe	boe	sew
dood	···	soos
do	boo	sue
dud	bub	suss

C. Procedure

Small (2.5 mm) gold pellets were glued to the speakers' lips, teeth, and tongue along the midline of the vocal tract (Fig. 1). The locations of two further pellets were tracked to indicate head movement. These pellets were glued to the bridge of the nose and to the border of the upper incisor and gums. One pellet was glued to the border between the lower incisors and the gums, thus indicating the location of the jaw. Two pellets were glued to the vermilion ridges of the upper and lower lips, and four pellets were placed at intervals of approximately 15 mm on the protruded tongue, with the first pellet being about 8–10 mm behind the tongue tip. Figure 1 shows the locations of the pellets at the midpoint of the vowel averaged across all vowels and speakers. When the tongue was not protruded, the tongue pellets were about 10 mm from each other. As the talkers read the experimental materials, the movements of the pellets were tracked by a computer controlled x-ray

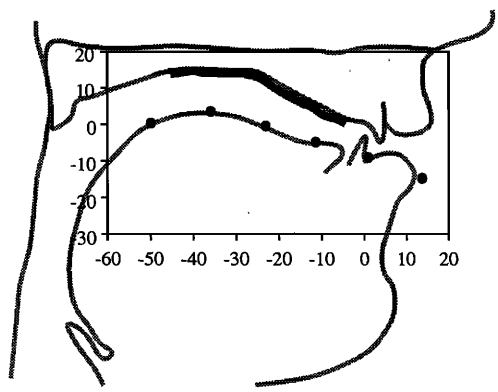

FIG. 1. Placement of the pellets on the surfaces of the vocal tract. This figure shows the average locations (across speakers, consonants and vowels) of the tongue, jaw, and lower lip pellets in a maxillary coordinate system originating at the upper tooth. Units of measure shown on the axes are millimeters. The dark trace on the palate is a smoothed average of palate trace measured for three of the speakers. The lighter sagittal outline of the vocal tract was taken from an earlier x-ray study and is presented merely as an illustration of the pellet placements. Later figures show the pellet locations without the sagittal illustration.

TABLE II. Number of tokens classified by speaker and consonant environment.

Speaker	Cons	#	Reps	Missing tokens	Extra tokens		# of utterances
DM	bVb	10	6	6 reps of [ɔ], [eᴵ] [i], [oᵁ], and [u]			30
	dVd	11	6	1 rep of [ɔ]			65
	sVs	11	3				33
						total	128
AM	bVb	10	3	3 reps of [i]			27
	dVd	11	3		3 reps of [ʊ]		36
	sVs	11	3				33
						total	96
RP	bVb	9	6	1 rep of [ɛ], [eᴵ]			52
	dVd	10	6	3 reps of [i]			57
	sVs	10	6	3 reps of [ʌ], [ʊ], [u]			51
						total	160
MB	bVb	10	3		3 reps of [i]		33
	dVd	11	3	3 reps of [ɑ], [æ], [ɛ], [eᴵ], [ɪ], and [i]			15
	sVs	11	3				33
						total	81
AO	bVb	10	6				60
	dVd	11	3				33
	sVs	11	6				66
						total	159
						Grand total	624

system (Abbs *et al.*, 1988). A small beam of x-ray tracked each pellet, and the locations (in both the vertical and horizontal dimensions) of the pellets were recorded at intervals of 10 ms (for tongue, lower lip, nose, and upper tooth) or 20 ms (for jaw and upper lip). Accuracy of the measurements was on the order of fractions of a millimeter. Additionally, the speech wave form was simultaneously sampled at a rate of 10 kHz.

The speakers read the list of materials in the order shown in Table I. Each sentence was repeated three times before moving on to the next sentence in the list. Table II shows the number of utterances recorded and measured for each speaker. The aim was to have each speaker read through the whole set of materials twice. However, recording x-ray microbeam data is not without its problems. Due to various instrumental failures and other errors, our aim was not achieved. We have valid data for one reading of the material by speakers AM and MB, except for the first 5 vowels of the /dVd/ list for speaker MB. Thus with this exception, we have three repetitions of each vowel in each consonant context for these two speakers. For speaker DM there are two valid readings of the /bVb/ list and the /dVd/ list, and one of the /sVs/ list, so the data include three repetitions of the vowels in the /s/ environment and six repetitions of the vowels in the other environments. The valid data for speaker AO are for the vowels three times in /dVd/ contexts and six times in the other contexts. The carrier phrase for the /bVb/ list as read by speaker DM was "Say__to me," so measurements of the vowels in open syllables could not be made because the consonant context was not symmetric (see below for more details on the measurements). As mentioned above, speaker RP did not read words containing [ɔ]. Because the data set was not fully balanced, the statistical analyses were based on unequal numbers of observations in the various conditions.

After the data had been collected, the nose and upper incisor pellet traces were used to correct for head movements, rendering the other pellet traces in terms of movement relative to the speaker's occlusal plane rather than absolute movement.

Five events were located in the trajectory of the primary consonant articulator in each CVC sequence (Fig. 2). Thus for CVC sequences involving /d/ and /s/ the events were located in the trajectory of tongue tip movement, and for sequences involving /b/ the events were located in the trajectory of the lower lip (see the Appendix for more details). The five events were: (1) the point of maximum absolute displacement toward consonant closure during the initial consonant, (2) the point of maximum speed from consonant closure to vowel opening, (3) the point of maximum vowel opening, (4) the point of maximum speed from vowel opening to the final consonant closure, and (5) the point of maximum absolute displacement toward consonant closure during the final consonant. We will be focusing in this paper on the locations of the pellets (in the maxillary coordinate system illustrated in Fig. 1) at the point of maximum vowel opening. This is the point during the vowel which shows the least influence from the neighboring consonants, and hence the smallest degree of overlap between C and V gestures.

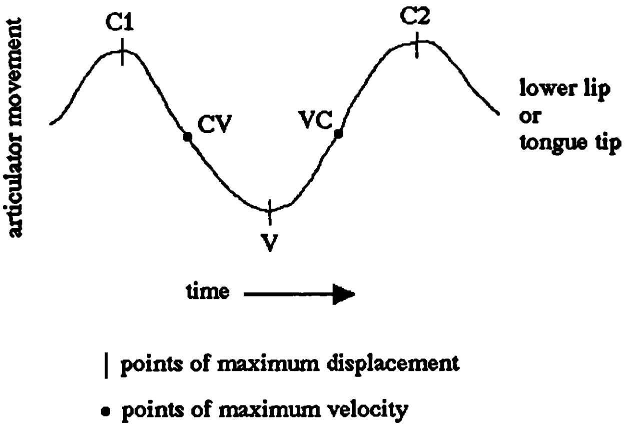

FIG. 2. Schematic representation of the articulatory landmarks at which pellet locations were measured. Represented are the moments of maximum displacement of the consonant articulator (either tongue tip or lower lip depending on the consonant environment) during the initial consonant (C1), medial vowel (V), and final consonant (C2) as well as the moments of peak articulator speed between the initial consonant and the medial vowel (CV) and between the medial vowel and the final consonant (VC). The articulatory landmarks were measured from two-dimensional data, but for simplicity are shown here in one dimension. See the Appendix for more details.

II. RESULTS

A. Within-speaker consistency

We will describe the results of four analyses of the data. First we investigated within-speaker consistency in the positions of the pellets at maximum vowel opening and across time. We will present the results of three correlational analyses which indicate that the speakers produced consistent movements when articulating the vowels.

The first analysis compared the locations of all six pellets at the midpoint of the vowel across different repetitions of the test materials. Pearson product-moment correlation coefficients were calculated within speakers and consonant contexts, comparing each reading of the list with every other reading. Thus if the speaker read the /dVd/ list six times, there was a 6×6 matrix of repetition correlations for /dVd/ for that speaker. The lowest correlation in this analysis was $r=0.982$.

As would be expected given the physiology of the vocal tract, the locations of some of the pellets were correlated across vowels. For instance, the vertical locations of the lower lip and the jaw were correlated with each other ($r=0.8$) as were the jaw and tongue tip vertical locations ($r=0.79$). Because these interpellet correlations may have inflated the correlations reported above, another analysis was conducted, focusing on consistency in the location of the tongue dorsum pellet in repetitions of the materials. This pellet was used because the tongue dorsum horizontal and vertical locations were not correlated with each other (average $r=0.18$) and showed substantial variation as a function of vowel quality (on the order of 10–15 mm). The lowest correlation across repetitions of the list of vowels within consonant contexts and speakers was $r=0.973$.

Because data concerning the locations of the pellets at a single point in time can only provide inferences about movements, the third analysis investigated movement consistency more directly by calculating repetition correlations across time. The locations of the tongue dorsum pellet were examined at three points in time (C1, CV, and V). The smallest correlation across repetitions of vowels within consonant contexts and speakers was $r=0.975$. A similar analysis of tongue tip movement found slightly lower correlations than this. Most of the tongue tip repetition correlations were high but in three cases they fell below $r=0.9$.

Taken together, these analyses suggest that the speakers in this study were remarkably consistent in their productions. They achieved the same posture at the vowel midpoints in each repetition of the list of materials and the consonant opening gestures were also very consistent across repetitions. Previous researchers have reported less consistency. Russell (1928) found substantial variability in vowel production. However, Paramenter and Treviño (1932) found that speakers produce very similar vocal tract configurations in repetitions of the same utterance. They also reported that inconsistencies from one production of a word to the next such as those reported by Russell (1928) may arise from articulatory adaptations accompanying changes in posture. Abbs *et al.* (1984) reviewed several studies which found that movement and EMG patterns were not consistent across utterances, but rather showed motor equivalence. That is, the activities of individual muscles or articulators varied from utterance to utterance, but the combined effect of several movement components (such as jaw and lower lip movement in the production of [ɑ]) was to produce a consistent vocal tract area function (Abbs and Netsell, 1973). Articulatory variation resulting from postural changes such as those reported by Paramenter and Treviño (1932) were not expected in the present data, but articulatory trading relations (Abbs *et al.*, 1984) probably occurred. The correlation analyses suggest that the magnitude of such articulatory trading must have been small relative to the differences between vowels.

B. Speaking rate

Second, we investigated individual differences in speaking rate. One way that speakers may vary is that they may choose different characteristic rates of speaking, which may influence the position of the articulators in vowel production (Lindblom, 1963). Therefore, even though the speakers were instructed to use a normal speaking rate, the durations of the test vowels from the initial consonant closure (C1) to the final consonant closure (C2) may have been dependent on the rate. Figure 3 shows these duration measurements averaged across vowels and consonant contexts for each speaker. The figure suggests, and an analysis of variance confirms, that vowel duration differed from one speaker to the next [$F(4,611)=62.6$, $p<0.01$].

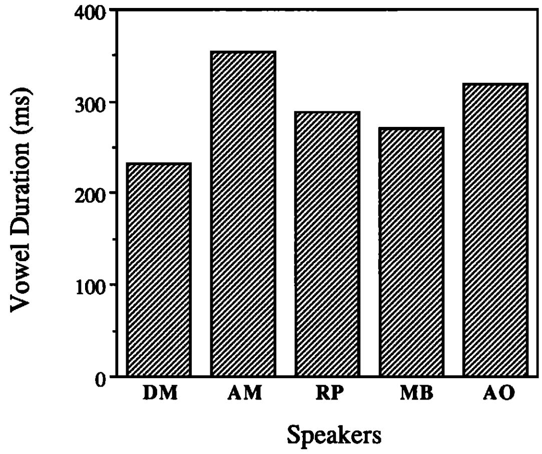

FIG. 3. Vowel duration (averaged across vowels and consonant contexts) by speaker.

C. Tense versus lax vowels

Third, we investigated differences in the production of the tense/lax distinction. Ladefoged *et al.* (1972) investigated some ways in which speakers vary in producing the distinction between tense and lax vowels (see also Raphael and Bell-Berti, 1975; and Bell-Berti *et al.*, 1978 concerning individual differences in EMG patterns which are consistent with the different patterns reported by Ladefoged *et al.* and the current data). Table III shows the results of statistical analyses of jaw height (mandible pellet *y* dimension) and tongue height (tongue body 2 pellet *y* dimension) in tense/lax vowel pairs in each of the consonant contexts. Note that our tongue height measure refers to the absolute position of the tongue in the mouth, rather than the position of the tongue relative to the jaw. By considering the locations of both the jaw and the tongue, as we are doing, we get an indication of the independent component of tongue movement, while retaining information about the vocal tract constriction. Each row in Table III shows the results of a separate analysis of variance, which had as factors vowel (a tense vowel versus a lax vowel), and speaker. The vowel main effects show whether there was an effect of vowel averaged across speakers, while the speaker main effects show whether there are effects due to physiological differences among the speakers such as the relative size or shape of the vocal tract, which is present regardless of the tenseness or laxness of the vowel. The interactions between speaker and vowel effects are most important for this discussion, because they indicate that the speakers produced the distinction between the tense and lax vowels in different ways.

As the table shows there were a number of reliable interactions. We will describe in some detail the data from the /dVd/ context for four speakers (speaker MB's data for these vowels in /dVd/ contexts were incomplete), which illustrate the type of differences documented in Table III. Figure 4a shows jaw height plotted against tongue height at the midpoint of the vowels in "dee," "did," "day," "dead," "do," and "dood" as produced by speaker DM. Note that because we are focusing on differences in patterns of articulation rather than absolute differences between the speakers, the panels in Fig. 4 have different scales. The vowel in "dee," in the top right corner of panel (a), was produced with relatively high jaw and tongue positions. The lax vowel in "did" (directly below it) had the same jaw height ($t[10]=0.1$, $p=0.92$) but was separated from the tense vowel in "deed" by a lower tongue

TABLE III. Results of statistical tests of jaw height and tongue height in tense/lax vowel pairs. Separate analyses of variance of the jaw (mandible pellet, *y* dimension) and tongue (tongue body 2 pellet, *y* position) were calculated for three tense/lax pairs in three consonant contexts. Each ANOVA had factors, subject, vowel, and the subject by vowel interaction. Subject and the subject by vowel interaction were treated as random factors, thus the Type III sums of squares were used to calculate *F* values. *F* values for these factors are shown in bold in this table (the degrees of freedom of the statistic are shown in parentheses), and each row shows the results of a separate ANOVA (*=$p<0.05$).

			Vowel	Subject	Vowel ×subject
/dVd/					
	Jaw	[i] vs [ɪ]	**0.11** (1,25)	**25.67*** (3,25)	**1.86** (3,25)
		[eˈ] vs [ɛ]	**0.13** (1,28)	**32.5*** (3,28)	**14.98*** (3,28)
		[u] vs [ʊ]	**16.67*** (1,35)	**8.98*** (4,35)	**3.33*** (4,35)
	Tongue	[i] vs [ɪ]	**509.4*** (1,25)	**837.4*** (3,25)	**13.8*** (3,25)
		[eˈ] vs [ɛ]	**496.9*** (1,28)	**398.4*** (3,28)	**12.7*** (3,28)
		[u] vs [ʊ]	**271.9*** (1,35)	**275.0*** (4,35)	**2.42** (4,35)
/sVs/					
	Jaw	[i] vs [ɪ]	**9.39*** (1,32)	**39.40*** (4,32)	**3.35*** (4,32)
		[eˈ] vs [ɛ]	**11.54*** (1,32)	**49.72*** (4,32)	**17.7*** (4,32)
		[u] vs [ʊ]	**24.04*** (1,26)	**48.92*** (4,26)	**1.52** (4,26)
	Tongue	[i] vs [ɪ]	**1169.5*** (1,32)	**542.8*** (4,32)	**12.25*** (4,32)
		[eˈ] vs [ɛ]	**562.6*** (1,32)	**282.2*** (4,32)	**13.1*** (4,32)
		[u] vs [ʊ]	**1265.3*** (1,26)	**732.5*** (4,26)	**9.4*** (4,26)
/bVb/					
	Jaw	[i] vs [ɪ]	**26.5*** (1,34)	**21.7*** (4,34)	**0.28** (2,34)
		[eˈ] vs [ɛ]	**7.4*** (1,31)	**31.1*** (4,31)	**1.16** (3,31)
	Tongue	[i] vs [ɪ]	**67.9*** (1,34)	**76.3*** (4,34)	**0.001** (2,34)
		[eˈ] vs [ɛ]	**709.2*** (1,31)	**159.2*** (4,31)	**5.17*** (3,31)

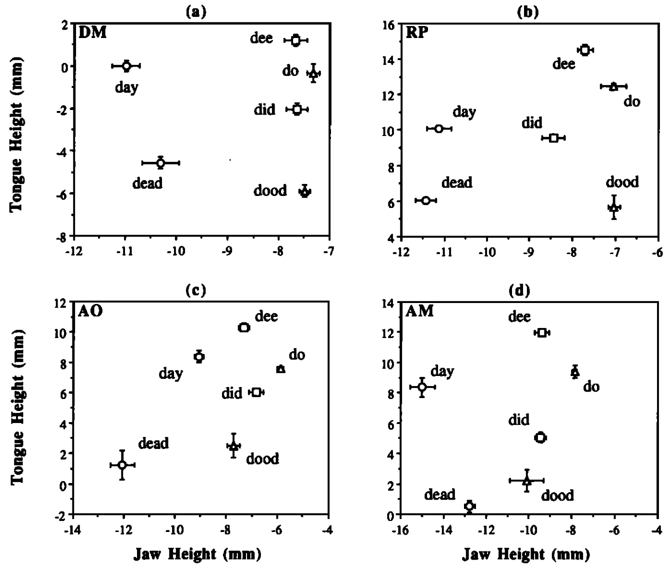

FIG. 4. The relationship of jaw height and tongue height in tense and lax vowels for four speakers in the /dVd/ environment. Error bars are standard deviation.

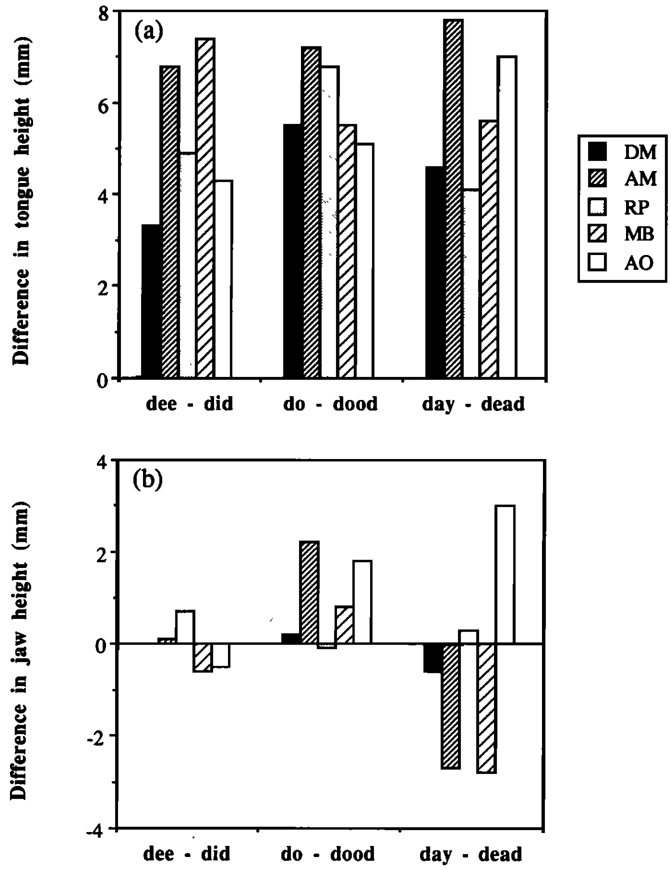

FIG. 5. (a) Difference in tongue height between tense and lax vowels. (b) Difference in jaw height between tense and lax vowels. Data are from the /sVs/ context for speaker MB and the /dVd/ context for the others. Positive values indicate a higher tongue or jaw position in the tense member of each pair.

height ($t[10]=-8.6$, $p<0.01$). The vowels in the "day/dead" pair were separated in exactly the same way (different tongue heights: $t[10]=-11.6$, $p<0.01$, same jaw height: $t[10]=1.5$, $p=0.16$), as were the vowels in "do/dood" (tongue height: $t[10]=-10.6$, $p<0.01$, jaw height: $t[10]=-1$, $p=0.34$). Also, note that jaw height was correlated with linguistic vowel height (i.e., the [-high] vowels /eᶦ/ and /ɛ/ had a lower jaw position than did the others which are linguistically [+high]). Thus this speaker's strategy appears to involve varying jaw height to separate linguistically high and mid vowels, and varying the tongue height independently of the jaw to separate the tense and lax members of each vowel pair. Speaker RP [Fig. 4(b)] had the same pattern (except that the test of jaw height contrasting "dee" and "did" was marginal: $t[7]=-2.4$, $p=0.0501$).

Speaker AO [Fig. 4(c)] used the same mechanism for separating the vowels of the "dee/did" pair as the previous speaker's (jaw: $t[4]=1.4$, $p=0.24$, tongue: $t[4]=-15$, $p<0.01$). However, for the vowels in the other two vowel pairs, "day/dead" and "do/dood," she used tongue height and jaw height in a different way. The tense and lax vowels in "day/dead," and in "do/dood" differed both in terms of tongue height ([eᶦ] vs [ɛ]: $t[4]=-6.9$, $p<0.01$, [u] vs [ʊ]: $t[4]=-6.5$, $p<0.01$) and in terms of jaw height ([eᶦ] vs [ɛ]: $t[4]=-5.96$, $p<0.01$, [u] vs [ʊ]: $t[4]=-6.9$, $p<0.01$), with the height of the tongue correlated with the height of the jaw in the distinction between [eᶦ] and [ɛ] and between [u] and [ʊ]. Thus this speaker patterned with speakers DM and RP for [i]/[ɪ] and in the other vowel pairs she produced the distinction between tense and lax vowels with the same correlated jaw and tongue locations found in the distinction between high and mid vowels.

Speaker AM [Fig. 4(d)] exhibited a more complex pattern. The tense and lax vowels in "dee/did" were produced with different tongue positions ($t[4]=-19.1$, $p<0.01$) and without a change in jaw position ($t[4]=-0.1$, $p=0.9$). This pattern was thus found in all four speakers for this vowel pair. The vowels in the back "do/dood" pair were separated in the same way as for speaker AO. The vowel in "do" was produced with both a higher tongue position ($t[7]=-6.6$, $p<0.01$) and a higher jaw position ($t[7]=-2.8$, $p<0.05$). However, the separation of the vowels in the "day/dead" pair involved a third type of strategy. As before, the tongue was lower in "dead" than in "day" ($t[4]=-10.6$, $p<0.01$), but the jaw positions were reversed from those of the other speakers. The lower tongue position of the lax vowel in "dead" was combined with a HIGHER jaw position ($t[4]=3.72$, $p<0.05$). As with the other speakers, this speaker produced the [+high] vowels with a higher jaw position than that in the [−high] vowels.

So, for these four speakers, linguistic vowel height was uniformly associated with the relative height of the jaw. The tense/lax distinction was, however, manifested in different ways. The results are summarized in Fig. 5, which includes data from speaker MB's productions of the vowels in /sVs/ contexts. Figure 5(a) shows that each speaker produced different tongue heights for the tense/lax vowel pairs, but the magnitude of these differences were not the same across speakers (as the vowel X speaker interactions in Table III suggest). For example, speakers AM and MB had larger tongue height differences for the [i]/[ɪ] pair than did speakers DM, RP, and AO.

Figure 5(b) illustrates the differences in jaw height between the vowels in the tense/lax pairs. The very small differences in jaw height for the vowels in the [i]/[ɪ] pair show that all speakers differentiated these two vowels by varying tongue height within a constant jaw position. But speakers AM and AO had a higher jaw in "do" than in "dood" while the other speakers showed no change in jaw position in these words (hence the vowel X speaker interaction in Table III for the jaw for [u]/[ʊ] in /dVd/ context). In the [eᶦ]/[ɛ] pair we find a third pattern for speakers AM and MB, where the jaw height differences did not correlate with the tongue height differences, that is, the jaw was lower for the tense vowel [eᶦ] than for the lax vowel [ɛ].

In summary, while linguistic vowel height was uniformly associated with different jaw positions (compare [i]/[ɪ] with [eᶦ]/[ɛ] in Fig. 4), the tense/lax distinction in American English was produced by varying tongue height within a fixed jaw position, or by coordinating the tongue and the jaw in the same manner as for varying linguistic vowel height, or even by varying the tongue and jaw in opposite directions. Earlier researchers (Ladefoged *et al.*, 1972; Bell-Berti *et al.*, 1978) have suggested that speakers differ in their strategies for producing tense and lax vowels. The present data show a slightly more complicated picture, because not only did different speakers produce the tense/lax distinction in different ways, but the same speaker sometimes patterned differently for different tense/lax pairs.

The fact that some speakers showed correlated jaw and tongue positions for some tense/lax pairs while others did not, suggests that, in terms of Saltzman and Munhall's (1989) model, the mapping from vowel gestures to articulators differed across this group of speakers. Therefore, the fact that the jaw is involved in making the tongue height distinction between "do" and "dood" for speakers AM and AO (with average tongue height differences of 7.2 and 5.1 mm, respectively) but not for speakers DM and RP (with average tongue height differences of 5.5 and 6.8 mm, respectively), suggests that the jaw is more closely tied to the tongue body gesture for AM and AO than it is in the same gesture produced by DM and RP. Similar differences in interarticulatory coordination may also underlie the curious pattern of differences found in the production of "day" versus "dead."

D. Global patterns

Finally, we investigated individual differences in overall patterns of vowel articulation. A factor analysis was implemented to investigate broad patterns of movement and posture which distinguish the vowels of English. The gestures in a task-dynamic model have been described as "cohesive patterns of movement" (Browman and Goldstein, 1989, p. 202), so factor analysis was used to discover cohesive articulatory patterns in vowel production.

Harshman *et al.* (1977) introduced the PARAFAC model of multidimensional scaling into the study of speech articulation. This method of data analysis reduces complex patterns of data into a few basic components or factors. Using PARAFAC, Harshman *et al.* found that tongue postures during American English vowels (the x-ray traces from Ladefoged *et al.*, 1972) could be decomposed into two basic tongue shapes; each vowel derived by a unique combination of the basic shapes. The PARAFAC procedure also estimates speaker-specific parameters to account for some individual differences in articulation. The assumption of the model is that each speaker uses the same patterns of articulation to distinguish different vowels, and thus the speaker-specific parameters in the model are simply scale factors to account for vocal tract size differences (see Jackson, 1988a,b). The data presented in the previous section suggest that this assumption of the PARAFAC model is not true for data which include both tongue and jaw position measurements (although for tongue shape alone the assumption may still hold). Rather, the data on individual differences in the production of the tense/lax distinction suggest that speakers differ from each other not only in terms of vocal tract size, but also in terms of interarticulator coordination. Therefore, we performed both a pooled analysis, in which speaker differences were disregarded, and separate factor analyses of each speaker's articulations. First, we will describe the results of the pooled analysis and then we will discuss the results of the separate analyses.

The horizontal and vertical locations of all six pellets were examined at three points in time for each target vowel. The times were the maximum displacement of the consonant articulator (lower lip or tongue tip) during the vowel, and the speed maximums in the transitions to and from the vowel (see the Appendix). Thus there were 36 raw variables per observation. We used canonical discriminant analysis (Kshirsagar, 1972; SAS Institute, 1988) to reduce these data to two factors along which the vowels, pooled over speakers and consonant contexts, could be best discriminated. In this pooled analysis and in 3 of the 5 individual analyses the third factor accounted for an appreciable amount of variance. However, because two of the speakers did not appear to use the third factor, and because the major interest is in making between-speaker comparisons, the discussion of the results will be limited to the first two factors. It should be noted that unlike PARAFAC or other types of multidimensional scaling, canonical discriminant analysis gives the same solution regardless of the number of factors one cares to examine. In this regard it is like principle component analysis. We will consider two aspects of the analysis: (1) the patterning of the vowels in the factor space, and (2) the articulatory patterns which correspond to the factors. The vowel space formed by the vowel scores on the first two factors of the pooled analysis (shown in Fig. 6) resembles the traditional vowel space, and can be roughly described as having "high/low" and "front/back" dimensions. The vowels /i/ and /u/ have large positive values on the first factor while /ɑ/ and /ɔ/ are at the opposite extreme. The second factor separates back vowels such as /o/ and /u/ from front vowels such as /æ/ and /i/.

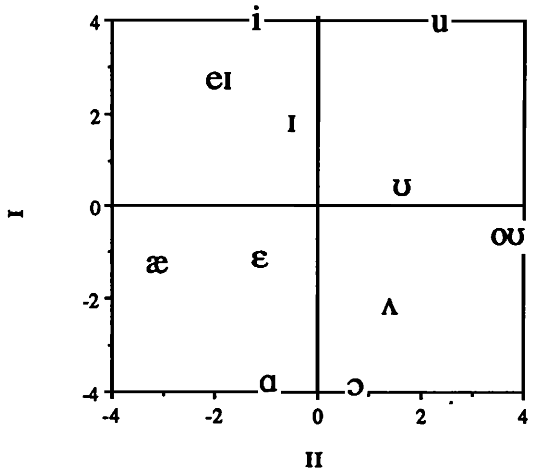

FIG. 6. The derived vowel space (vowel weightings on the factors) from a canonical discriminant analysis of data pooled across speakers and consonant environments. The vertical axis (I) shows vowel weightings on the first factor, and the horizontal axis (II) shows vowel weightings on the second factor.

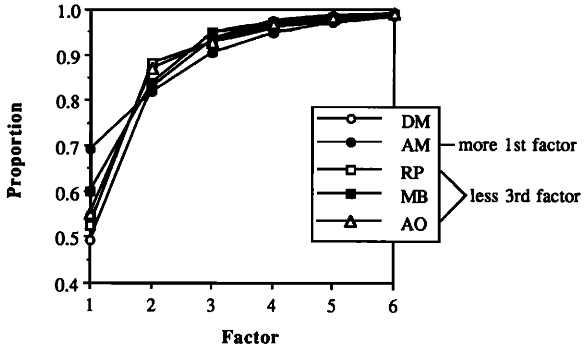

FIG. 8. Cumulative variance accounted for in each of the individual analyses. Variance accounted for is shown on the vertical axis and factor number is shown on the horizontal axis.

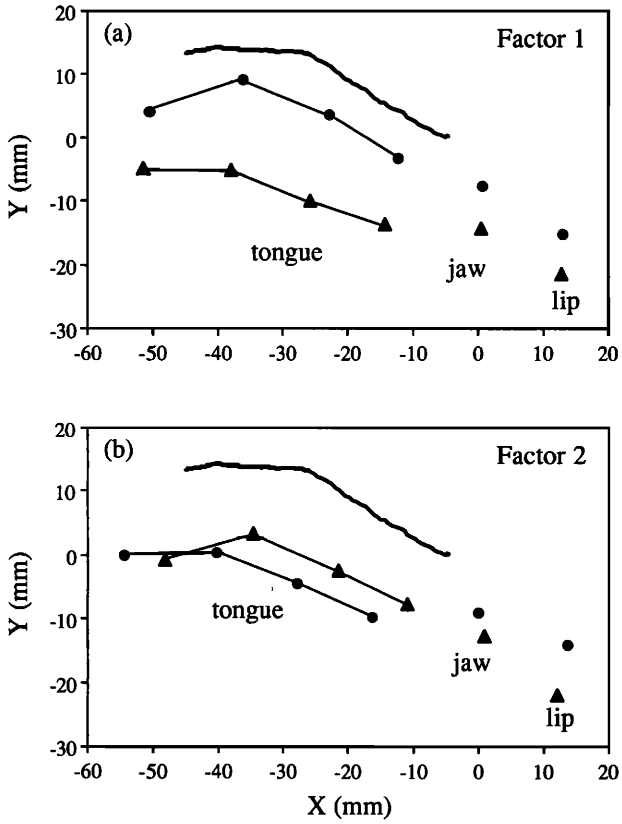

FIG. 7. Factor loadings from the pooled analysis. (a) Extreme positive (filled circles) and negative (filled triangles) values of factor 1 with factor 2 held at 0. (b) Extreme positive and negative values of factor 2 with factor 1 held at 0. The position of the palate is illustrated in both panels by the dark line near the top of each graph which is an average of palate traces measured for three speakers. See Fig. 2 for an illustration of the pellet placement.

The articulatory patterns which correspond to these factors are shown in Fig. 7. The pellet locations shown in this figure were calculated by setting one factor to zero and the other factor to the most extreme positive (filled circles) or negative (filled triangles) value observed in the factor space (see Johnson, 1991 for a more detailed description of how these were derived). For example, in panel (a) the second factor was fixed at zero and the first factor was set to 4 (see Fig. 6) to calculate the values marked by circles, and to -4 to calculate the values marked by triangles. Thus the derived tongue shapes shown in Fig. 7 do not correspond to any actual vowel, but rather to abstract correlations among the variables which capture basic articulatory patterns used in distinguishing the vowels. These abstract articulatory patterns can be interpreted in traditional phonetic terms. The first factor [Fig. 7(a)] corresponds to the traditional high/low distinction. It involves vertical displacement of the tongue and virtually no horizontal displacement. The second factor [Fig. 7(b)] corresponds to the front/back and rounded/unrounded distinctions (which in American English are correlated because only back vowels may be rounded). It involves horizontal displacement of the tongue and also lip rounding when the tongue is back.

So, this analysis of the vowel production data pooled across speakers is relatively straightforward and generally conforms with the conventional linguistic description of English vowels. However, as noted in the previous sections, there were some interesting between-speaker differences in these data. These individual differences were explored by performing separate factor analyses for each speaker.

Some individual differences are apparent in the plots of cumulative variance accounted for by each factor in the separate analyses (Fig. 8). The first factor, which primarily reflects vowel height, accounted for relatively more variance in speaker AM's productions, while factor 3 (which primarily distinguished tense and lax vowels with a change in tongue shape) accounted for very little variance in the vowels produced by speakers RP and AO.

As in the pooled analysis, each separate analysis resulted in a derived 2-factor vowel space, and these vowel spaces were similar to each other as the correlation matrices in Table IV show. This table shows the Pearson

TABLE IV. Between subject correlations of vowel weighting matrices.

	Factor 1					Factor 2				
	DM	AM	RP	MB	AO	DM	AM	RP	MB	AO
DM	···	0.85	0.89	0.92	0.92	···	0.95	0.84	0.87	0.93
AM		···	0.70	0.90	0.74		···	0.67	0.90	0.79
RP			···	0.73	0.93			···	0.60	0.93
MB				···	0.84				···	0.77
AO					···					···

product-moment correlations of the individual vowel spaces formed by the first two factors for each speaker. The correlations were generally high (which justifies further comparisons of the results), but the matrices also reveal some differences among the speakers. The vowel spaces for speakers RP and AO were very similar, as were the spaces derived from the vowels produced by speakers AM and MB. The vowel space for speaker DM correlated fairly well with all of the other speakers.

To illustrate the individual differences suggested by the plots of cumulative variance (Fig. 8) and vowel space correlations (Table IV), we will present the results of the individual analyses of two representative speakers, AM and AO, with summary data for the other speakers.

The difference between high and low vowels (factor 1) for speaker AM involved a change in the shape of the tongue, as may be seen in Fig. 9(a), which shows that the jaw was relatively stationary and the body of the tongue was curved forward and up for high vowels, but was flatter for low vowels. On the other hand, the difference between high and low vowels for speaker AO involved a greater difference in jaw height and lip height, but only a small change in tongue shape [Fig. 9(b)]. The pattern for all 5 speakers can be seen in Fig. 10 which shows the differences between jaw height (mandible pellet, y dimension) and tongue height (tongue body pellet 2, y dimension) in the articulatory patterns captured in factor 1 for each speaker. For speakers RP and AO factor 1 includes a large jaw component which is not present in the analysis of speakers DM, AM, and MB. Kent and Moll (1972) found a similar difference between the two speakers in their study. In the analysis of the tense/lax distinction reported above, speaker AO tended to distinguish between tense and lax vowels by changing both the locations of the tongue and the jaw, while AM and MB showed more independence of jaw and tongue in distinguishing tense and lax vowels, especially [eɪ] and [ɛ]. It is also interesting that the speakers are different in this way because RP and AO's productions did not require a third factor to distinguish tense and lax vowels (Fig. 8) while a third factor was needed for the other speakers.

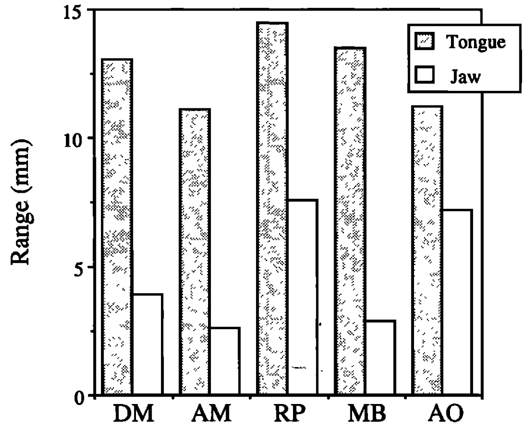

FIG. 10. Range of tongue and jaw displacement encoded in the factor loadings for factor 1 from the separate analyses.

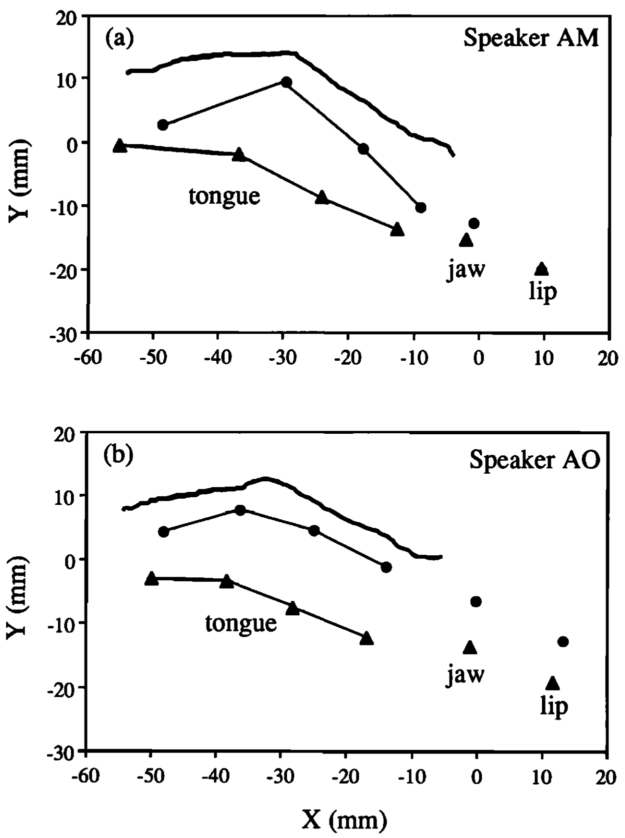

FIG. 9. Factor loadings for the first factor from the separate analyses of speakers AM and AO. The heavy dark lines show smoothed palate traces for these individuals.

As with jaw/tongue coordination in tense/lax vowels discussed in the previous section, these data suggest that the speakers differ in the articulators which they recruit to carry out a particular gesture. In Saltzman and Munhall's (1989) framework, this is a difference in interarticulator coordination. Notice also in Fig. 9 that the tongue body constriction during high vowels (those with a positive loading on factor 1, the filled circles) is more narrowly constrained to a small area on the roof of the mouth for speaker AM, while the constriction is more distributed for speaker AO. Perhaps the targets are the same, but the difference in vocal tract area function which results from the changes in interarticulator coordination indicates that different means of raising the tongue body to the roof of the mouth may produce different area functions.

Figure 11 shows the differences between speakers AM and AO for factor 2, the front–back parameter. For all of the speakers this factor involved lip rounding and horizontal movement of the tongue body, but there was a difference in the vertical position of the tongue. Speaker AM, shows very little change in tongue position associated with this second factor. The tongue moves back and up in the mouth, but the front of the tongue doesn't change position very much at all. So, although AM appears to have a correlation between velar constriction and lip rounding (filled circles), this person showed no change in the position of the front part of the tongue. On the other hand, speaker AO showed a marked difference in the position of the front part of the tongue depending on whether the lips were rounded. Like AM, lip rounding appears to be associated with a velar constriction (filled circles, in figure 11b), but unlike AM there was greater constriction at the front of the mouth when the lips were unrounded (filled triangles). Thus, both AM and AO appear to produce a constriction at the velum when the lips are rounded, but AO has a constriction in the front of the mouth when the lips are spread while AM shows no such change. Figure 12 shows the location of the tongue body 2 pellet (the third pellet from the tip of the tongue) as realized in the articulatory patterns captured by factor 2 for all speakers. Speakers RP and AO again patterned together with tongue lowering correlated with lip rounding and tongue retraction, while speakers AM and MB showed tongue raising with lip rounding and tongue retraction, and speaker DM showed very little change in the vertical dimension. In terms of Saltzman and Munhall's (1989) model, these individual differences involve a difference in gestural targets because RP and AO (unlike the other speakers) show a large difference in tongue location associated with the second factor in the factor analysis.

III. DISCUSSION

A. Explaining individual differences

The results showed within-speaker consistency and between-speaker variability in the production of the vowels of American English. In this section the individual differences reported above will be considered relative to potential artifacts, and to individual characteristics in vocal tract anatomy.

First, the different articulatory patterns that were observed might have occurred because of between-speaker differences in *pellet placement*. The pellets were placed on the protruded tongue at intervals of about 15 mm with the front-most pellet about 8–10 mm from the tongue tip. So, the relative locations of the pellets on any particular speaker's tongue depended on the size of the individual's tongue which was probably correlated with gender. Therefore, if between-speaker differences in the relative locations of the pellets are the source of the individual differences in the articulatory patterns seen in this study, the different articulatory patterns may be correlated with gender. In one case this seems to be true. Speakers AM and MB distinguished [eᴵ] and [ɛ] in the same way and they were both women. However, the third female speaker (AO) did not pattern with speakers AM and MB. Furthermore, the two speakers whose data were used to illustrate the range of between-speaker variability in the individual factor analyses (speakers AM and AO) were both women. So, to the extent that variations in relative pellet placement can be attributed to differences in vocal tract size that may relate to gender, the observed results are not likely to be due to relative pellet placement.

A second possible source of the different articulatory patterns is *speaking rate*. If a person is speaking relatively slowly and carefully there would probably be an accompanying articulatory reorganization as compared with faster more casual speech. In particular, it seems reasonable to suppose that speakers might move the mandible more in careful speech than in rapid speech. The duration data reported in Sec. II A indicate that the speakers did not speak at the same rate. However, the speakers who patterned together in the articulatory analyses do not pattern together in terms of speaking rate. The two slowest speakers (AM and AO) had quite different articulatory patterns, while the fastest speaker (DM) had a pattern which was between the extremes represented by speakers AM and AO. Therefore, speaking rate differences do not account for the other individual differences that were observed.

The third possible explanation of the individual differences reported here is that the speakers may not have been speaking the same *dialect*. There is a sound change in progress in the Northern Midwestern cities of the United States, which Labov (1972, 1991) calls the Northern Cities Shift. In this sound change, the /æ/ in words like "had" is being raised and acquiring an offglide, so it is pronounced [ɛᵊ] or [eᵊ]. Also, /ɛ/ as in "head" is lowered slightly, and /ɑ/ as in "hod" is being fronted to [a] or [æ]. Now, if these changes in vowel production were present in the speech of some of our speakers but not others, the individual differences that were observed may just reflect dialectal variations and not different articulatory strategies for producing phonetically identical vowels. There are two reasons to believe that the individual differences found in this study were not the result of dialect differences. First, before participating in the experiment the speakers were asked to read a list of words which vary as a function of dialect, including "merry," "Mary," and "marry," and "caught" and "cot." There was a small trace of the Northern Cities Shift in their speech, namely, all of the speakers lowered /ɛ/ and lengthened /æ/. Speaker RP did not distinguish between [ɑ] and [ɔ], whereas the other speakers did, but this difference is not correlated with the observed individual differences. In the factor analysis, RP and AO patterned together, despite the apparent dialect difference. Second, the vowel productions which had been tracked by the x-ray microbeam system were transcribed; the transcriptions indicated no differences in the pronunciations of the test materials. So, there is nothing to suggest that the different articulatory strategies exhibited by these speakers are due to dialect differences.

Although the observed articulatory differences cannot be accounted for in terms of pellet placement, speaking rate or dialect differences, it may still be possible to offer a principled account of these differences as lawful articula-

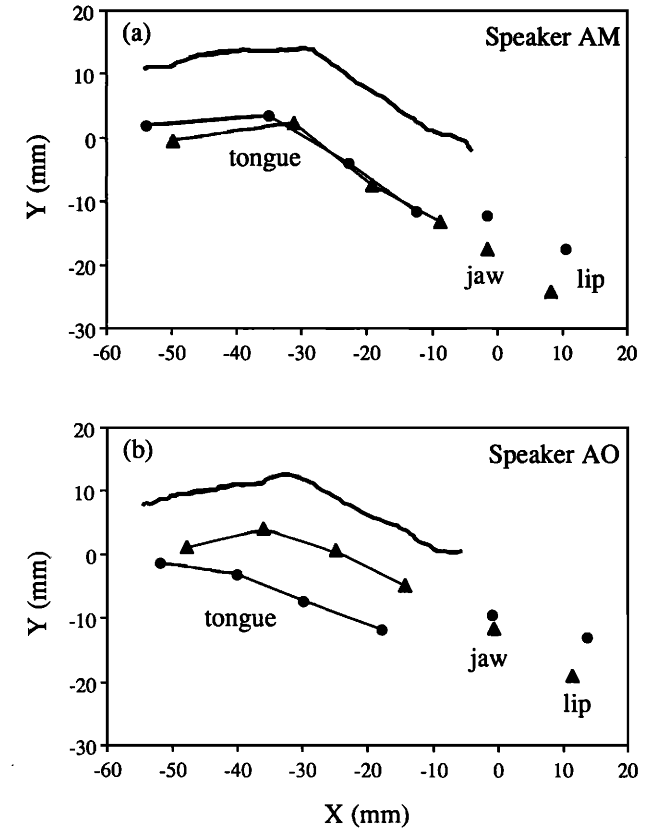

FIG. 11. Factor loadings for the second factor from the separate analyses of speakers AM and AO. The heavy dark lines show smoothed palate traces for these individuals.

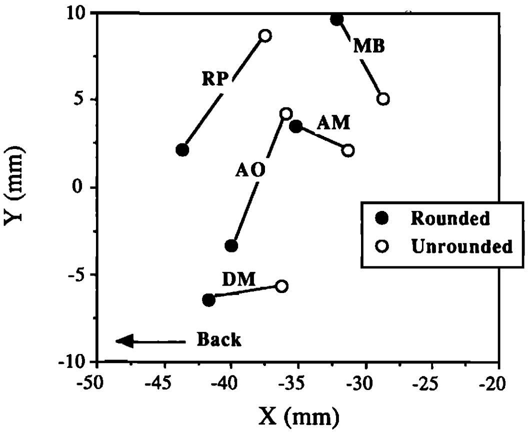

FIG. 12. The locations of tongue body pellet 2, for the two extremes of factor 2 (rounded and unrounded lip positions) from the separate analyses.

tory variability (as predicted by the universal articulatory phonetics hypothesis). First, individual differences in vocal tract anatomy may require that speakers adopt different articulatory strategies to accomplish functionally equivalent output signals. In particular, the degree of *palate doming* may have an impact on articulatory organization. The hypothesis is that if a person has a deeply vaulted palate it may be necessary to move the tongue more than if the palate is relatively shallow. Thus speakers with shallowy vaulted palates may move the jaw and tongue together while speakers with deeply vaulted palates may be inclined to use more independent tongue movement. Using the data reported by Ladefoged *et al.* (1972), correlations were calculated between the range of jaw positions observed at the midpoints of the vowels with an index of palate doming for each of their speakers. The index of palate doming was calculated by taking the ratio of the front length and the depth of the palate vault. The results showed that the range of jaw positions and the index of palate doming were inversely correlated ($r^2=0.76$) suggesting that those speakers who had shallowy domed palates showed a greater range of jaw positions at vowel midpoint than those who had deeply domed palates. Thus differences in vocal tract geometry were, for these speakers, correlated with differences in vowel articulation. However, when we repeated this analysis with the x-ray microbeam speakers, the relationship between palate doming and jaw movement did not hold. Evidently, if there is a relationship between the shape of the palate and articulatory organization it is more complex than this simple analysis could reveal. Still, this is an interesting possibility that should be investigated further.

A second possible explanation of the articulatory differences between speakers has to do with *dental occlusion* (see Angle, 1907 and Pound, 1977). Edwards and Harris (1990) found that two speakers with class II occlusions (usually characterized by a retrusion of the mandible and protrusion of the maxilla) showed more jaw translation during speech than did another speaker who had a normal occlusion. Subtelny, Mestre and Subtelny (1964) found that speakers with class II occlusions who were judged by listeners to have "normal speech" positioned the tongue further back in the mouth than did speakers with normal occlusions. While it is not obvious that dental occlusion necessarily influenced vowel production in this study, this type of anatomical data (which is not available for these speakers) might have provided some insight on the individual differences found here.

Finally, there is a possibility that *articulatory strategy* is not directly motivated by the physical requirements of speaking. Clearly, anatomical differences may require different patterns of articulation, but it is also the case that for any particular speech sound a variety of articulatory gestures may be functionally equivalent. "Normal speech seems to exploit no more than a fraction of the degrees of freedom that are in principle available for articulation" (Lindblom, 1983, p. 224). Therefore, the articulatory strategy a person adopts may be only partly determined by anatomy. Consider, by way of analogy, handwriting. The factors which determine an individual's unique style of handwriting may include some anatomical factors, but surely individual differences are also a reflection of the person's own unique habits within the bounds of communicative function and social convention. Perhaps there is a similar range of possible patterns of articulatory organization for speech production defined by functional utility and social convention and that individuals randomly adopt articulatory strategies from this set of possibilities.

B. Summary and implications

The individual differences which we found are interesting because they were of several different types. Speakers differed in terms of characteristic speaking rate, which in Saltzman and Munhall's (1989) task dynamic model is associated with gestural stiffness. There were also differences in how speakers produced tense/lax vowel distinctions which appeared to reflect variations in interarticulatory coordination in the model. Speakers differed in how they produced vowel height distinctions, which could also be due to differences in interarticulatory coordination. Finally, there were differences in how speakers maintained the distinction between front and back vowels, which seemed to indicate a variation in gestural targets.

The patterns of variation could not be unequivocally explained on the basis of vocal tract size, speaking rate, or dialect. For instance, it was predicted that those speakers who didn't move their jaws very much in producing tense/lax or height distinctions might have been speaking more quickly than the other speakers. If this had been the case it could have been argued that there were subtle differences in interarticulatory coordination at different speaking rates, and thus the individual differences would have been lawful. This explanation, like all of the others we could think of, was untenable. We are left with the conclusion that speakers of the same language and dialect may use different articulatory plans; that the universal articulatory phonetics hypothesis is wrong.

What then is it that all speakers of a language share, if they don't share the articulatory plans used to pronounce the words of their language? Consider two of the individual differences mentioned in the Introduction. In producing "r" sounds in English, it doesn't matter whether the speaker uses retroflexion or tongue bunching so long as another speaker of the language can recognize the sound. Therefore, the "r" sounds of English are often described by reference to their auditory characteristics, using a term such as *rhotic*, which says nothing about the articulations involved. Similarly, "s" sounds, which may be pronounced using at least two different articulatory strategies, are often described as *sibilant*, another auditory feature. Following Ladefoged *et al.* (1972), individual differences such as these (as well as the differences in vowel production reported above) may be interpreted as indirect evidence that the acoustic product of speaking is the crucial determinant of the organization of speech articulation. It seems clear that to some extent speech motor movements are not only goal directed, but also that the goal is an auditory one, leading us to believe in the validity of an *auditory theory of speech production*.

ACKNOWLEDGMENTS

Research supported by NIH Grant T32 DC00029 and R01 DC00330. We appreciate the help of the staff at the microbeam facility at Madison. Comments from Maureen Stone, Anders Lofqvist, and one anonymous reviewer were very helpful and thought provoking, as were discussions with members of the UCLA Phonetics Lab. Rob Hagiwara and Jenny Ladefoged helped with the illustrations.

APPENDIX

This appendix describes the algorithm and heuristics used by a computer program in making the x-ray microbeam measurements reported in this paper. The main difference between the technique reported here and the measurement techniques followed by other researchers is that for both measurements of displacement and speed we considered movements in the horizontal and vertical dimensions.

The program stepped backward in time from the acoustic onset of the vowel (which had been previously identified by Lindau and Ladefoged, 1990) to the point at which the articulator was furthest forward and front (for both tongue tip and lower lip). This strategy is based upon the fact that the jaw, as a hinge, rotates horizontally as well as vertically, and upon the assumption that movement toward consonant closure for the alveolar and bilabial consonants involves both horizontal and vertical components. After choosing a location for the initial consonant, the vowel maximal opening was identified as the point at which the articulator was maximally back and low. The location of the final consonant closure was measured using the same method which had been used in finding the location of the initial consonant. The process of finding the locations of the maximal consonant closures was repeated until the initial and final locations of the articulator were as close (in space) as possible. The tongue tip or lower lip did not always follow a smooth path toward closure, and although the program had to consider a peak as the end of the movement, it was not always clear how to terminate the process of looking for an extreme in the direction toward consonant closure. The heuristic adopted by the program was that the best choices for the consonant maxima were the peaks which resulted in the smallest spatial separation between the locations for the initial and final closures (which in these utterances were phonologically identical). After finding maxima for consonants and vowels the points of maximum speed were measured. The measurement of speed included movement in both the horizontal and vertical dimensions, hence "speed" rather than "velocity." This measurement for both the opening and closing gesture was a simple first difference. Since the sampling rate for the articulators was constant, the Euclidean distance between one sample and the next is proportional to the speed of the articulator during that period of time.

The measurements of consonant and vowel maxima produced by the program were evaluated and obvious failures were remeasured by hand. The total number of such failures was less than 10%. To further evaluate the success of the program, we compared the values obtained for the location of the vowel midpoint with similar measurements which had been previously made by hand (Lindau and Ladefoged, 1990). Although, the earlier measurements had been made just using the vertical dimension of the articulator, there was no significant difference between the locations of vowel midpoint determined by the two methods.

[1]We have tried to choose examples which illustrate individual differences as opposed to dialectal differences. Unfortunately, while many authors attend to regional variation, social variation is often overlooked.

Abbs, J. H., Gracco, V. L., and Cole, K. J. (**1984**). "Control of multimovement coordination: Sensorimotor mechanisms in speech motor programming," J. Motor Behav. **16**, 195–231.

Abbs, J. H., Nadler, R. D., and Fujimura, O. (**1988**). "X-ray microbeams track the shape of speech," Soma, 29–34 (January).

Abbs, J. H., and Netsell, R. (**1973**). "A dynamic analysis of two-dimensional muscle force contributions to lower lip movement," J Acoust. Soc. Am. **53**, 295.

Angle, E. H. (**1907**). *Malocclusion of the Teeth* (S. S. White Dental Manufacturing Co., Philadelphia).

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., and Sawusch, J. R. (**1978**). "Some relationships between articulation and perception," Haskins Lab.: Status Rep. Speech Res. **SR-55/56**, 21–32.

Bernstein, N. (**1967**). *The Coordination and Regulation of Movements* (Pergamon, London).

Borden, G. J., and Gay, T. (**1979**). "Temporal aspects of articulatory movements for /s/-stop clusters," Phonetica **36**, 21–31.

Browman, C. P., and Goldstein, L. (**1989**). "Articulatory gestures as phonological units," Phonology **6**, 201–251.

Chomsky, N., and Halle, M. (**1968**). *The Sound Pattern of English* (Harper & Row, New York).

Clements, G. N. (**1985**). "The geometry of phonological features," Phonol. Yearbook **2**, 225–252.

Easton, T. A. (**1972**). "On the normal use of reflexes," Am. Sci. **60**, 591–599.

Edwards, J., and Harris, K. S. (**1990**). "Rotation and translation of the jaw during speech," J. Speech Hearing Res. **33**, 550–562.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fowler, C. A., Rubin, P., Remez, R. E., and Turvey, M. T. (**1980**). "Implications for speech production of a general theory of action," in *Language Production*, edited by Butterworth (Academic, New York), pp. 373–420.

Fujimura, O., Kiritani, S., and Ishida, H. (**1973**). "Computer controlled radiography for observation of movements of articulatory and other human organs," Comput. Biol. Med. **3**, 371–384.

Gelfand, I. M., Gurfinkel, V. S. Tsetlin, M. L., and Shik, M. L. (**1971**). "Some problems in the analysis of movements," in *Models of the Structural-functional Organization of Certain Biological Systems*, edited by I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, and M. L. Tsetlin, translated by Carol R. Beard (MIT, Cambridge, MA), pp. 329–345.

Greene, P. H. (**1972**). "Problems of organization of motor systems," in *Progress in Theoretical Biology: 2*, edited by R. Rosen and F. M. Snell (Academic, New York), pp. 303–338.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693–707.

Heffner, R.-M. S. (**1950**). *General Phonetics* (The University of Wisconsin, Madison).

Jackson, M. T. T. (**1988a**). "Analysis of tongue positions: Language-specific and cross-linguistic models," J. Acoust. Soc. Am. **84**, 124–143.

Jackson, M. T. T. (**1988b**). "Phonetic theory and cross-linguistic variation in vowel articulation," UCLA Working Papers in Phon. **71**, 1–184.

Jakobson, R., Fant, G., and Halle, M. (**1952**). *Preliminaries to Speech Analysis* (MIT, Cambridge).

Johnson, K. (**1991**). "Dynamic aspects of English vowels in /bVb/ sequences," UCLA Working Papers Phon. **80**, 99–120.

Keating, P. A. (**1985**). "Universal phonetics and the organization of grammars," in *Phonetic Linguistics* edited by V. A. Fromkin (Academic, New York), pp. 115–132.

Kent, R. D., and Moll, K. L. (**1972**). "Tongue body articulation during vowel and diphthong gestures," Folia Phoniatrica. **24**, 278–300.

Kenyon, J. S. (**1924**). *American Pronunciation: A Text-book of Phonetics for Students of English* (George Wahr, Ann Arbor).

Kiritani, S., Itoh, K., and Fujimura, O. (**1975**). "Tongue-pellet tracking by a computer-controlled x-ray microbeam system," J. Acoust. Soc. Am. **57**, 1516–1520.

Kshirsagar, A. M. (**1972**). *Multivariate Analysis* (Marcel Dekker, New York).

Labov, W. (**1972**). *Sociolinguistic Patterns* (Univ. of Pennsylvania, Philadelphia).

Labov, W. (**1991**). "The three dialects of English," in *New Ways of Analyzing Sound Change*, edited by P. Eckert (Academic, New York).

Ladefoged, P. (**1971**). *Preliminaries to Linguistic Phonetics* (University of Chicago, Chicago).

Ladefoged, P. (**1980**). "What are linguistic sounds made of," Language **56**, 485–502.

Ladefoged, P. (**1982**). *A Course in Phonetics* (Harcourt Brace Jovanovich, New York), 2nd Ed.

Ladefoged, P., DeClerk, J., Lindau, M., and Papcun, G. (**1972**). "An auditory-motor theory of speech production," UCLA Working Papers Phon. **22**, 48–75.

Lindau, M., and Ladefoged, P. (**1989**). "Methodological studies using an x-ray microbeam system," UCLA Working Papers Phon. **72**, 83–90.

Lindau, M., and Ladefoged, P. (**1990**). "Interarticulatory relationships in vowel production," UCLA Working Papers in Phon. **74**, 115–123.

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

Lindblom, B. (**1983**). "Economy of speech gestures," in *The Production of Speech*, edited by P. MacNeilage (Academic, New York).

Maddieson, I. (**1984**). *Patterns of Sound* (Cambridge U.P., Cambridge).

Paramenter, C. E., and Treviño, S. N. (**1932**). "Vowel positions as shown by x-ray," Q. J. Speech. **18**, 351–369.

Pike, K. (**1943**). *Phonetics* (University of Michigan, Ann Arbor).

Pound, E. (**1977**). "Let *S* be your guide," J. Prosth. Dent. **38**, 482–487.

Raphael, L. J., and Bell-Berti, F. (**1975**). "Tongue musculature and the feature of tension in English vowels," Phonetica **32**, 61–73.

Russell, G. O. (**1928**). *The Vowel: Its Physiological Mechanism as Shown by X-ray* (McGrath, College Park, MD).

Saltzman, E. L., and Munhall, K. G. (**1989**). "A dynamical approach to gestural patterning in speech production," Ecolog. Psychol. **1**, 333–382.

SAS Institute (**1988**). *SAS User's Guide: Statistics, 1988 Edition* (SAS Institute, Inc, Cary, NC).

Sherrington, C. S. (**1906**). *The Integrative Action of the Nervous System* (Yale U.P., New Haven, CT).

Stevens, K. N., and Keyser, S. J. (**1989**). "Primary features and their enhancement in consonants," Language **65**, 81–106.

Subtelny, J. D., Mestre, J. C., and Subtelny, J. D. (**1964**). "Comparative study of normal and defective articulation of /s/ as related to malocclusion and deglutition," J. Speech Hear. Disord. **29**, 269.

Turvey, M. T. (**1977**). "Preliminaries to a theory of action with reference to vision," in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, edited by R. Shaw and J. Bransford (Erlbaum, Hillsdale, NJ), pp. 211–265.

Turvey, M. T. (**1990**). "Coordination," Am. Psychol. **45**, 938–953.

Weiss, P. (**1941**). "Self-differentiation of the basic pattern of coordination," Compar. Psychol. Monogr. **17**, 21–96.