Phonetica

Editor: K. Kohler, Kiel
Publishers: S. Karger, Basel
Separatum (Printed in Switzerland)

Phonetica *39:* 1–22 (1982)

Individual Variation in the Perception of Vowels: Implications for a Perception-Production Link

Robert Allen Fox

Department of Linguistics, Ohio State University, Columbus, Ohio, USA

Abstract: A scaling experiment using dyadic comparisons of vowel pairs was performed to determine if significant perceptual variation existed on an individual level and whether such variation could be correlated with differences among subjects in their production of vowels. Subjects were required to scale six separate sets of vowel stimuli. Each stimulus set consisted of the 36 possible dyads from the vowel set /i, ɪ, ɛ, æ, a, ʌ, o, ʊ, u/. Analysis of the resulting data indicated that significant individual perceptual variation was present which could be seen as a function of perceptual structure differences between the subjects and not merely error variance. In addition, such variations were related to aspects of each subject's acoustic vowel space – indicators of tongue height and advancement during vowel articulation. These results were interpreted as suggesting a link between perception and production.

A theoretical position held by many speech researchers is that a nontrivial link exists between the processes of perception and production in the human cognitive system. This position has been incorporated into theories of speech perception [Liberman et al, 1967; Stevens, 1972] and production [Ladefoged et al., 1972] as well as in the design of speech therapy programs [Van Riper, 1963; McReynolds et al., 1975]. However, although direct empirical support for this hypothesized link has been offered by a number of experimental studies, such results have often been prone to alternative interpretations or have been unreplicated in subsequent experiments.

A number of child language studies have reported a positive correlation between speech sound discrimination and articulation ability – especially between poor discrimination and misarticulation [Kronvall and Diehl, 1954; Cohen and Diehl, 1963; Madison and Fucci, 1971; Monnin and Huntington, 1974; McReynolds et al., 1975]. But studies reporting contradictory results are equally numerous

[POWERS, 1957; FARQUHAR, 1961, DICKSON, 1962; AUNGST and FRICK, 1964; HAGGARD et al., 1971; MARQUARD and SAXMAN, 1972]. WEINER [1967] claimed that such inconclusive results were all attributable to a failure to control for the effects of age and severity of the defect and noted that studies consistently found a positive relationship between articulation and speech sound discrimination below the age of 9 years. STITT and HUNTINGTON [1969] suggested that even this upper age limit was only an artifact of the testing procedure and reported findings of a significant relationship between articulation proficiency and speech signal discrimination and identification abilities in college-age subjects. However, WALDMAN et al. [1978] in a study comparing speech production and perception in children with articulation disorders, controlling for both age and sex of the subjects found no significant relationship between articulation and discrimination when performance was tested on the same stimulus items.

Several studies by COOPER [1974; COOPER and LAURITSEN, 1974; COOPER and NAGER, 1975] involving perceptuo-motor adaptation have shown that the voice onset time (VOT) values in subjects' production of syllable-initial voiceless stops decreased immediately following repeated exposure to an adapting speech sound. For example, COOPER [1974] demonstrated that subjects' VOTs in production of [p^hi] were significantly shorter following perceptual adaptation to [p^hi] but not to [i]. These adaptation effects on production are analogous to selective adaptation effects on perception of VOT [EIMAS and CORBIT, 1973] and suggest that similar mechanisms are involved in both perception and production. However, these results have recently been brought into question by SUMMERFIELD et al. [1980], who were unable to replicate COOPER's basic perceptuo-motor adaptation effect using a slightly modified version of his experimental procedure. In addition, BAILEY and HAGGARD [1978] found no correlation between a subject's average VOT and his perception of VOTs as phonemically voiced or voiceless which might be expected if such a link exists.

Other evidence supporting such a link is found in BELL-BERTI et al. [1979]. Their data indicated that speakers could be divided into different populations in terms of production strategies for the articulation of the English vowels /i, ɪ, e, ɛ/. Subjects were divided into two separate groups on the basis of EMG potentials recorded from the genioglossus muscle during vowel production. These recordings showed two different patterns of muscular contractions suggesting that one group of speakers utilized a production strategy involving tongue height to differentiate the tense/lax vowel pairs while the second group utilized tongue tension. Perceptual measures were obtained from the subjects using vowel identification tests requiring subjects to label the members of a seven-step synthetic vowel continuum (/i/ to /ɪ/) under both equal-probability and anchoring conditions. They found that the magnitude of the boundary shift in the anchoring condition (toward the more frequently occurring stimulus) was greater for subjects using the tongue height strategy. They hypothesized that the perceptual differences reflect different articulatory strategies and thus suggest a common mechanism between perception and production, but cautioned that this hypothesis was not the only one possible (e.g., perceptual reliance by subjects upon durational differences among the vowels would not correspond in any direct manner to tongue height or tension strategies during articulation).

As the preceding studies have indicated, empirical support for a common mechanism between perception and production can be found, but not without conflicting results or interpretations. To establish firmly the reality of such a mechanism, additional support must be found and it is to this end that the present study is directed.

If perception and production are mediated by a common process or otherwise linked in some fashion, it is reasonable to assume that when subjects differ in terms of vowel articulation then, ceteris paribus, they should exhibit differences in terms of vowel perception. Similar assumptions are the basis of most studies discussed above including BELL-BERTI et al. [1979], BAILEY and HAGGARD [1978], and WALDMAN et al. [1978] and involve comparison of appropriate perception and production measures. A method for obtaining a useful measure of individual vowel perception is the technique of multidimensional scaling using programs such as INDSCAL or PARAFAC, which not only extract the features used in perception, but also indicate individual reliance upon those extracted features [CARROLL and CHANG, 1970; HARSHMAN, 1970]. Methods of obtaining direct measurements of vowel articulation are more problematic. However, comparison of an individual subject's perceptual performance with acoustic measurements of his vowel space should be closely equivalent to a comparison of his perceptual performance with actual articulatory parameters such as tongue height and tongue advancement since the acoustic and articulatory measures are so highly correlated [LINDAU, 1978]. Indeed, relevant acoustic measurements of subjects' vowel perception could represent a very accurate reflection of vowel production processes, since speakers may organize their articulation of vowel quality in acoustic terms rather than articulatory terms [LADEFOGED et al., 1972]. The experiment to be discussed here was designed to determine (1) whether significant variation in vowel perception existed on a strictly individual level and (2) if significant variation did exist, to what extent the variations could be predicted on the basis of acoustic vowel space differences (seen as representative of production differences) among the subjects. Support for a perception-production link will only obtain if both significant variation is found and a systematic relationship between the perceptual differences and the acoustic space differences can be established.

Method

The perceptual data were obtained using a perceptual scaling paradigm (dyadic comparison) which produced estimates of the perceptual distances among a set of vowel stimuli. Subjects (corresponding to the Listener variable) were required to scale six different sets of vowel stimuli on 3 successive days. Each vowel set was produced by a different speaker and corresponds to the Speaker variable. This design allows us to discover whether an individual subject's perceptual responses are significantly (and systematically) different from one another over a series of different vowel sets. If perceptual data were obtained from the subjects while only scaling a single stimulus set, then even if statistical analysis demonstrated significant variation among subjects, we could not be sure to what extent the variation reflected chance error differences (or differences among subjects only with regard to the particular set of stimuli used) or actual systematic perceptual differences among the subjects. With several different stimulus sets (which were systematically varied in terms of acoustic characteristics) one can better determine the degree of individual variation as a function of the underlying perceptual system.

Stimuli

The stimuli consisted of nine different American English vowels: /i, ɪ, ɛ, æ, a, ʌ, ʊ, u/ appearing in the phonetic environment [h___d]. These vowels include most of the relatively monophthongal vowels in American English and are representative of a variety of articulatory and acoustic distinctions (e.g. front/back, high/low, round/nonround, closed/open). There were 36 different vowel dyads (excluding vowels paired with themselves and neglecting order). Six different vowel sets were recorded, each set produced by a different speaker. Each vowel set consisted of four trials, each trial containing all 36 vowel pairs plus false pairs at the beginning and end to avoid scaling 'noise' related to first or last pair responses. The vowel dyads were presented in one pseudorandom order in trials 1 and 3 (random order except for the requirement that no vowel appeared twice in consecutive pairs) and in the reverse order in trials 2 and 4. The vowels within a given dyad appeared in the sequence AB in trials 1 and 2 and BA in trials 3 and 4. Different pseudorandom orders were used for each different speaker.

The groups of speakers producing the vowel sets included 3 men and 3 women. Within each of the sex subgroups there was a speaker with a relatively low F_0, one with a relatively medium F_0, and one with a relatively high F_0. The sex differentiation in the stimulus sets produces a systematic variation among the separate speaker sets in terms of relative formant height – women tend to have shorter vocal tracts than do men and thus exhibit higher formant frequencies. Requiring each sex subgroup to have speakers of high, medium, and low F_0 allows one to have vowel sets which vary relatively independently with regard to each of these aspects of speaker quality. Spectrograms were made of 20 different tokens of each vowel in each of the six speaker sets. Measurements obtained from these spectrograms included F_0, F_1, F_2, F_3 and duration. Formant frequency values were measured at the midpoint of the vowel in each token. Mean values of F_0, F_1, F_2, F_3, F_2–F_1, and duration of the nine vowels occurring in each of the six speaker sets appear in table I.

The vowels were produced by each speaker under the direction of the experimenter who monitored all vowel stimuli produced. Vowel dyads to be recorded were presented on separate index cards. The speakers heard pairs of low-volume timing pulses 300 ms in length during the recording session. They were instructed to produce the vowel dyads in concert with these pulses and to make each token the same duration as the pulses. There was a 500-ms period of silence between the timing pulse for the first token in a dyad and the pulse for the next token. This was to ensure that the speaker paused between the two tokens, reducing possible coarticulatory effects. In addition speakers were specifically instructed to make such a pause. This procedure was designed to ensure relatively consistent vowel durations and interstimulus intervals across all stimulus sets. The stimuli

Table I. Mean acoustic measures of stimulus vowels (± SD)

Vowel	F_0	F_1	F_2	F_2–F_1	F_3	Duration ms
Speaker 1 (male)						
[i]	176 ± 6.8	308 ± 25	2,087 ± 45	1,779	2,779 ± 41	288 ± 22.8
[ɪ]	172 ± 5.0	395 ± 25	1,920 ± 21	1,525	2,664 ± 55	229 ± 20.6
[ɛ]	175 ± 9.3	437 ± 24	1,803 ± 49	1,366	2,634 ± 47	250 ± 17.2
[æ]	171 ± 2.1	661 ± 17	1,700 ± 53	1,039	2,634 ± 40	287 ± 24.4
[a]	172 ± 4.6	728 ± 20	1,191 ± 30	463	2,629 ± 73	296 ± 14.5
[o]	172 ± 4.6	413 ± 46	1,166 ± 28	753	2,548 ± 97	296 ± 18.1
[ʌ]	172 ± 4.9	674 ± 40	1,255 ± 40	581	2,543 ± 98	241 ± 33.4
[ʊ]	174 ± 6.1	419 ± 22	1,226 ± 57	807	2,557 ± 44	229 ± 21.6
[u]	177 ± 7.0	337 ± 20	1,209 ± 85	872	2,314 ± 83	274 ± 19.3
Speaker 2 (female)						
[i]	216 ± 9.6	251 ± 40	2,267 ± 98	2,016	3,000 ± 140	290 ± 26.4
[ɪ]	213 ± 6.1	268 ± 14	1,983 ± 89	1,715	2,608 ± 91	261 ± 28.2
[ɛ]	210 ± 12.2	630 ± 23	1,742 ± 86	1,112	2,469 ± 68	266 ± 33.8
[æ]	212 ± 8.4	753 ± 41	1,783 ± 91	1,030	2,582 ± 89	317 ± 36.6
[a]	208 ± 8.4	774 ± 35	1,283 ± 85	509	2,522 ± 86	315 ± 41.1
[o]	214 ± 9.3	416 ± 58	993 ± 39	577	2,287 ± 97	322 ± 33.0
[ʌ]	211 ± 11.7	648 ± 34	1,174 ± 41	526	2,444 ± 83	277 ± 47.7
[ʊ]	208 ± 8.4	308 ± 61	1,079 ± 98	771	2,290 ± 63	256 ± 34.3
[u]	217 ± 9.1	265 ± 41	876 ± 68	611	2,232 ± 73	285 ± 31.0

Table I. (continued)

Vowel	F_0	F_1	F_2	F_2–F_1	F_3	Duration ms
Speaker 3 (female)						
[i]	305 ± 14.3	278 ± 38	2,440 ± 26	2,162	3,061 ± 79	326 ± 19.
[ɪ]	303 ± 4.9	416 ± 49	2,174 ± 25	1,758	2,800 ± 55	337 ± 21.
[ɛ]	299 ± 13.0	707 ± 38	1,931 ± 98	1,224	2,840 ± 136	333 ± 20.
[æ]	299 ± 13.7	821 ± 48	1,734 ± 49	913	2,799 ± 93	339 ± 18.
[a]	325 ± 17.6	777 ± 32	1,201 ± 32	424	2,843 ± 139	357 ± 20.
[o]	300 ± 12.4	484 ± 39	913 ± 41	429	2,652 ± 96	361 ± 21.
[ʌ]	298 ± 14.1	733 ± 46	1,347 ± 61	614	2,776 ± 50	322 ± 16.
[ʊ]	302 ± 18.5	446 ± 20	1,033 ± 71	587	2,658 ± 36	310 ± 8.8
[u]	297 ± 15.6	309 ± 21	822 ± 82	513	2,603 ± 107	347 ± 17.
Speaker 4 (male)						
[i]	144 ± 9.9	322 ± 66	2,287 ± 63	1,965	2,867 ± 119	209 ± 17.
[ɪ]	144 ± 7.0	379 ± 29	2,130 ± 82	1,751	2,620 ± 93	140 ± 5.
[ɛ]	146 ± 10.3	451 ± 54	1,974 ± 48	1,523	2,579 ± 82	159 ± 10.
[æ]	146 ± 10.3	503 ± 67	1,965 ± 51	1,457	2,530 ± 55	238 ± 8.
[a]	143 ± 7.0	814 ± 63	1,109 ± 54	295	2,557 ± 130	232 ± 19.
[o]	147 ± 5.3	487 ± 79	939 ± 65	452	2,278 ± 148	229 ± 17.
[ʌ]	143 ± 7.2	571 ± 59	1,119 ± 68	548	2,452 ± 107	218 ± 6.
[ʊ]	142 ± 9.9	423 ± 65	967 ± 65	544	2,252 ± 62	221 ± 4.
[u]	147 ± 11.8	311 ± 68	872 ± 53	561	2,204 ± 63	218 ± 18.
Speaker 5 (male)						
[i]	121 ± 9.3	247 ± 47	1,869 ± 58	1,622	2,382 ± 58	312 ± 16
[ɪ]	118 ± 7.0	282 ± 39	1,713 ± 62	1,431	2,209 ± 78	285 ± 28
[ɛ]	118 ± 7.0	339 ± 40	1,561 ± 46	1,222	2,148 ± 42	294 ± 23
[æ]	118 ± 9.9	424 ± 45	1,522 ± 44	1,098	2,151 ± 97	356 ± 25
[a]	119 ± 10.3	587 ± 50	991 ± 59	404	2,160 ± 87	335 ± 32
[o]	119 ± 6.7	364 ± 46	808 ± 45	444	2,047 ± 47	341 ± 24
[ʌ]	119 ± 6.8	492 ± 53	1,061 ± 46	569	2,113 ± 28	295 ± 37
[ʊ]	119 ± 9.8	348 ± 35	922 ± 55	574	2,026 ± 51	295 ± 24
[u]	119 ± 6.8	283 ± 33	905 ± 69	622	1,974 ± 68	312 ± 26
Speaker 6 (female)						
[i]	327 ± 11.8	369 ± 33	2,662 ± 63	2,293	3,371 ± 107	348 ± 2
[ɪ]	326 ± 19.4	440 ± 55	2,383 ± 68	1,943	2,984 ± 103	306 ± 9
[ɛ]	445 ± 17.7	441 ± 52	2,227 ± 69	1,786	2,905 ± 76	317 ± 21
[æ]	322 ± 16.0	935 ± 36	1,718 ± 78	783	2,711 ± 89	381 ± 1
[a]	319 ± 9.9	1,027 ± 43	1,192 ± 65	165	2,471 ± 83	388 ± 2
[o]	306 ± 26.0	491 ± 33	1,087 ± 56	596	2,516 ± 68	363 ± 2
[ʌ]	362 ± 9.9	922 ± 80	1,357 ± 60	435	2,653 ± 106	301 ± 2
[ʊ]	316 ± 16.4	505 ± 32	1,200 ± 56	695	2,558 ± 94	296 ± 1
[u]	318 ± 11.0	426 ± 26	1,139 ± 31	713	2,581 ± 70	325 ± 1

All formant frequency values are given in Hertz. F_2–F_1 values were calculated from mean F_2 and

Mean values are based on a sample size of 20.

were recorded using a Tandberg series 9100X tape recorder with a realistic 'Highball' microphone. The speaker sat in an IAC sound-attenuated room and care was taken to achieve similar sound levels across all stimulus sets. The final tapes were produced by careful tape editing of the recorded stimuli. This editing process made sure that all vowels were of the proper phonetic quality (as evaluated by the experimenter), in the correct presentation sequence, and at a uniform loudness level. The vowel pairs appeared on the stimulus tatpe at 6-second intervals. Each pair was preceded by a short attention tone to alert the subject.

Experimental Task

The scaling task required subjects to listen to two sequentially presented vowels and to judge the similarity or dissimilarity of the vowel pair on a 9-point scale. Listeners were instructed to use the entire 9-point range. The subjects read typed instructions[1] describing the experimental task to ensure that identical instructions were given to all subjects. Listeners were given a set of 10 vowel pairs to scale as introduction to the task. Following the introductory set the first experimental stimulus set was presented. The listeners heard the stimuli as reproduced by a Tandberg 9100X series tape recorder through Nova Pro earphones while sitting in an IAC sound-attenuated room.

The experiment required listeners to scale the six separate vowel sts in three 1-hour sessions. Listeners scaled two separate vowel sets, a woman's and a man's at each session. In two sessions listeners scaled a man's vowel set first, in the other session they scaled a woman's set first. An introductory set of 10 vowel pairs (different for each session) was presented to listeners before each of the three experimental sessions.

Care was taken to avoid subject fatigue. Each listener was required to rest between speaker set presentations at each session. The rest break was designed not only to avoid taxing the subjects, but to maintain a distinct separation between all six speaker sets scaled.

---

[1] The following is the instruction set given to subjects:

This experiment is designed to obtain similarity judgments from listeners on sets of word pairs from various speakers. There are sets of word pairs from 6 different speakers. Each listener will listen to two sets of word pairs during each 1-hour session.

Each particular set of word pairs from a speaker consists of all possible pairs of the following words: heed, hid, head, had, hod, who'd, hood, hud, and hoed. Note that these words differ only in terms of the medial vowel sound in each. You will hear a list of these words at the beginning of the tape.

For each word pair you hear you are to make a judgment as to how similar or different-the vowel sounds in the two words are. Your judgment should not be based on considerations of any slight pitch, loudness, or durational differences between the vowels. Judgments will be done on a 9-point scale:

S . . . . D

For vowel sounds that are very similar, in your opinion, please circle a dot near 'S' (similar), for example:

S ⊙ . . . . D

For vowel sounds that are very different, please circle a dot near 'D' (different):

S . . . . ⊙ D

Please circle only the dots. We are interested in your initial reactions so please mark down your first judgments only. Try to use the entire 9-point scale.

One word pair will be presented every 6 s. Preceding each word pair is a short attention tone to alert you that a pair is to follow immediately.

To acquaint you with the task there will be a short 10 word pair practice set at the beginning of the tape. ARE THERE ANY QUESTIONS?

Listeners were not told the object of the experiment until after the end of all three sessions. Listeners were not instructed on what basis they were to make their similarity judgments, but the typed instructions made it clear that they were to make similarity judgments of vowel quality and not other attributes such as pitch, loudness, or duration.

Listeners

16 subjects, all nonlinguistics, took part in the scaling experiment and were evenly divided into men and women. Both speakers and listeners were chosen so that they were from the same general dialect. Each listener was paid $ 10.00 for his participation in the three-session experiment. At the end of the third and final session the subjects were required to read a list of the words they had heard in the experiment (i.e., heed, hid, head, had, hod, who'd, hood, hoed, hud). Eight examples of each token were recorded and were used in the ensuing acoustic measurements. Spectrograms were made of these words and measured in a manner identical to that for the Speaker vowels. Mean values of F_0, F_1, F_2, F_2–F_1, F_3 and duration for each of the listener's nine vowels appear in table II. These measurements are representative of each listener's acoustic vowel space and will be used in looking for significant correspondence between articulatory differences among subjects and perceptual differences among subjects should the latter prove to be significant.

Table II. Mean acoustic measures of listener's vowels

Vowel	F_0	F_1	F_2	F_2–F_1	F_3	Duration ms
Listener 1 (female)						
[i]	256 ± 14.7	406 ± 41	3,102 ± 50	2,696	3,536 ± 74	305 ± 8.
[ɪ]	259 ± 10.9	511 ± 36	2,666 ± 57	2,155	3,348 ± 76	290 ± 25.
[ɛ]	236 ± 2.6	1,011 ± 42	2,457 ± 88	1,446	3,391 ± 163	296 ± 28.
[æ]	236 ± 3.9	1,043 ± 70	2,282 ± 49	1,239	3,283 ± 116	324 ± 17.
[a]	232 ± 14.4	1,211 ± 43	1,371 ± 53	160	3,000 ± 109	348 ± 13.
[o]	249 ± 10.9	493 ± 36	1,261 ± 48	768	3,261 ± 156	343 ± 9.
[ʌ]	249 ± 4.1	968 ± 19.1	1,641 ± 36	673	3,163 ± 83	292 ± 15.
[ʊ]	249 ± 10.8	550 ± 41	1,319 ± 41	769	3,014 ± 108	259 ± 16.
[u]	245 ± 5.8	420 ± 21	1,174 ± 36	754	2,870 ± 101	313 ± 10.
Listener 2 (female)						
[i]	206 ± 12.4	290 ± 21	2,406 ± 22	2,116	2,449 ± 26	348 ± 23.
[ɪ]	197 ± 10.7	377 ± 21	2,232 ± 41	1,855	2,724 ± 81	288 ± 7.
[ɛ]	188 ± 10.9	696 ± 30	1,794 ± 55	1,098	2,748 ± 94	311 ± 14.
[æ]	193 ± 6.2	754 ± 41	1,855 ± 41	1,101	2,898 ± 82	352 ± 15.
[a]	194 ± 5.7	891 ± 37	1,587 ± 37	696	2,717 ± 147	354 ± 47.
[o]	194 ± 6.6	413 ± 22	1,033 ± 64	620	2,533 ± 56	361 ± 14.
[ʌ]	190 ± 11.4	739 ± 36	1,381 ± 56	648	2,750 ± 64	332 ± 15.
[ʊ]	206 ± 6.6	376 ± 54	1,261 ± 71	885	2,522 ± 36	264 ± 11.
[u]	203 ± 6.6	337 ± 19	1,152 ± 37	815	2,381 ± 35	317 ± 15.

able II. (continued)

owel	F^0	F_1	F_2	F_2–F_1	F_3	Duration ms
istener 3 (male)						
]	144 ± 2.6	365 ± 59	2,757 ± 45	2,392	3,304 ± 59	295 ± 21.9
]	144 ± 3.2	409 ± 23	2,174 ± 39	1,768	2,765 ± 59	249 ± 20,8
ɛ]	141 ± 3.3	539 ± 56	2,044 ± 50	1,505	2,687 ± 36	255 ± 18.7
æ]	144 ± 2.8	642 ± 38	1,696 ± 52	1,054	2,316 ± 89	305 ± 24.4
a]	138 ± 3.4	685 ± 33	1,142 ± 61	457	2,282 ± 100	303 ± 14.5
o]	153 ± 4.1	478 ± 8.7	1,104 ± 44	626	2,145 ± 46	287 ± 7.8
ʌ]	136 ± 3.7	609 ± 13	1,522 ± 68	913	2,333 ± 23	251 ± 20.4
ʊ]	144 ± 3.3	478 ± 21	1,218 ± 34	740	2,174 ± 69	229 ± 15.3
u]	151 ± 7.5	392 ± 36	1,104 ± 73	712	2,226 ± 49	303 ± 25.3
Listener 4 (male)						
[i]	144 ± 2.0	276 ± 41	2,087 ± 72	1,811	2,609 ± 94	290 ± 6.3
[ɪ]	140 ± 6.1	348 ± 13	1,855 ± 41	1,507	2,580 ± 21	253 ± 18.8
[ɛ]	141 ± 5.6	457 ± 22	1,717 ± 22	1,260	2,555 ± 56	282 ± 13.2
[æ]	136 ± 6.2	638 ± 21	1,623 ± 21	985	2,493 ± 37	338 ± 3.6
[a]	134 ± 5.6	597 ± 37	968 ± 20	371	2,533 ± 55	330 ± 18.9
[o]	148 ± 5.7	403 ± 36	970 ± 42	567	2,478 ± 44	328 ± 24.1
[ʌ]	141 ± 10.8	598 ± 55	1,207 ± 65	609	2,500 ± 72	291 ± 8.4
[ʊ]	141 ± 1.6	348 ± 62	913 ± 42	565	2,406 ± 54	278 ± 17.3
[u]	141 ± 5.6	326 ± 22	935 ± 39	609	2,402 ± 36	301 ± 21.9
Listener 5 (male)						
[i]	128 ± 6.6	229 ± 21	2,044 ± 36	1,815	2,859 ± 74	300 ± 12.9
[ɪ]	128 ± 6.7	229 ± 42	1,859 ± 121	1,630	2,478 ± 113	248 ± 9.9
[ɛ]	123 ± 7.6	392 ± 23	1,594 ± 50	1,202	2,348 ± 87	261 ± 5.7
[æ]	128 ± 6.6	587 ± 70	1,500 ± 51	913	2,316 ± 67	370 ± 9.8
[a]	118 ± 3.2	767 ± 24	1,028 ± 33	261	2,174 ± 96	351 ± 4.5
[o]	122 ± 3.9	333 ± 25	797 ± 26	464	1,935 ± 90	362 ± 9.9
[ʌ]	118 ± 4.1	347 ± 50	1,065 ± 52	718	2,087 ± 69	292 ± 8.1
[ʊ]	131 ± 3.7	322 ± 27	971 ± 66	649	1,854 ± 78	254 ± 15.8
[u]	125 ± 4.6	239 ± 21	848 ± 74	609	1,979 ± 150	317 ± 9.8
Listener 6 (female)						
[i]	254 ± 6.2	319 ± 54	2,521 ± 43	2,202	3,072 ± 55	270 ± 13.6
[ɪ]	240 ± 5.7	435 ± 17	2,294 ± 93	1,859	2,935 ± 109	284 ± 6.3
[ɛ]	241 ± 6.2	594 ± 21	2,087 ± 71	1,493	2,913 ± 62	290 ± 16.7
[æ]	227 ± 6.2	696 ± 18	1,913 ± 107	1,217	2,870 ± 62	307 ± 15.6
[a]	249 ± 9.3	935 ± 42	1,326 ± 86	391	2,859 ± 112	312 ± 12.4
[o]	400 ± 10.9	468 ± 36	1,089 ± 36	621	2,794 ± 79	306 ± 28.1
[ʌ]	266 ± 28.9	707 ± 19	1,522 ± 36	815	2,782 ± 137	283 ± 19.1
[ʊ]	254 ± 12.4	536 ± 21	1,218 ± 37	682	2,652 ± 62	283 ± 4.6
[u]	253 ± 5.7	425 ± 47	1,055 ± 102	630	2,586 ± 38	269 ± 18.9

Table II (continued)

Vowel	F^0	F_1	F_2	F_2–F_1	F_3	Duratio ms
Listener 7 (female)						
[i]	271 ± 6.1	304 ± 20	2,913 ± 25	2,609	3,630 ± 90	225 ±
[ɪ]	167 ± 5.9	362 ± 31	2,505 ± 28	2,143	3,130 ± 81	173 ±
[ɛ]	158 ± 6.1	478 ± 30	2,203 ± 40	1,725	2,971 ± 83	196 ±
[æ]	158 ± 10.3	717 ± 38	2,326 ± 66	1,609	3,000 ± 77	256 ±
[a]	162 ± 12.0	985 ± 47	1,362 ± 80	377	2,950 ± 91	234 ±
[o]	161 ± 7.1	794 ± 42	967 ± 37	173	2,783 ± 83	233 ±
[ʌ]	168 ± 6.9	739 ± 25	1,250 ± 72	511	3,022 ± 101	168 ±
[ʊ]	171 ± 6.8	435 ± 30	1,072 ± 80	637	2,768 ± 75	168 ±
[u]	167 ± 10.1	261 ± 21	924 ± 30	663	2,681 ± 60	200 ±
Listener 9 (male)						
[i]	125 ± 5.9	290 ± 22	1,855 ± 44	1,565	2,435 ± 174	175 ±
[ɪ]	118 ± 3.9	333 ± 21	1,739 ± 55	1,406	2,347 ± 60	157 ±
[ɛ]	112 ± 6.5	391 ± 37	1,696 ± 39	1,305	2,314 ± 43	187 ±
[æ]	118 ± 3.3	463 ± 41	1,594 ± 49	1,131	2,304 ± 60	188 ±
[a]	112 ± 4.1	630 ± 33	1,087 ± 39	457	2,304 ± 69	200 ±
[o]	112 ± 3.7	348 ± 28	957 ± 39	609	1,978 ± 45	225 ±
[ʌ]	112 ± 2.9	461 ± 21	1,319 ± 21	858	2,232 ± 30	166 ±
[ʊ]	112 ± 3.5	370 ± 19	1,261 ± 39	891	2,043 ± 47	162 ±
[u]	112 ± 2.6	283 ± 24	1,152 ± 22	869	1,783 ± 68	189 ±
Listener 10 (male)						
[i]	203 ± 6.3	348 ± 12	2,087 ± 38	1,739	2,544 ± 45	385 ±
[ɪ]	188 ± 6.1	350 ± 31	1,968 ± 47	1,620	2,565 ± 62	282 ±
[ɛ]	190 ± 6.5	417 ± 37	1,903 ± 19	1,486	2,511 ± 48	315 ±
[æ]	194 ± 10.7	468 ± 37	1,859 ± 23	1,391	2,489 ± 62	296 ±
[a]	184 ± 9.2	917 ± 35	1,082 ± 71	165	2,382 ± 68	343 ±
[o]	194 ± 6.4	413 ± 22	859 ± 23	446	2,434 ± 62	351 ±
[ʌ]	190 ± 6.5	609 ± 62	1,218 ± 78	609	2,456 ± 86	304 ±
[ʊ]	193 ± 12.3	420 ± 21	1,028 ± 41	608	2,392 ± 44	280 ±
[u]	190 ± 6.1	348 ± 31	957 ± 31	609	2,276 ± 63	322 ±
Listener 11 (female)						
[i]	197 ± 10.2	290 ± 20	2,900	2,610	3,504 ± 70	271 ±
[ɪ]	207 ± 9.1	383 ± 15	2,365	1,985	3,043 ± 80	258 ±
[ɛ]	206 ± 6.5	435 ± 36	2,035	1,600	2,948 ± 108	285 ±
[æ]	184 ± 6.6	609 ± 62	1,870	1,261	2,957 ± 91	323 ±
[a]	184 ± 8.0	587 ± 39	1,174	587	2,991 ± 84	345 ±
[o]	197 ± 2.7	470 ± 41	1,152	602	2,939 ± 51	320 ±
[ʌ]	197 ± 6.1	493 ± 39	1,276	785	2,913 ± 62	252 ±
[ʊ]	210 ± 10.5	452 ± 31	1,104	652	2,696 ± 23	280 ±
[u]	219 ± 5.3	377 ± 28	985	608	2,722 ± 50	285 ±

All formant frequency values are given in Hertz. F_2–F_1 values were calculated from mean F_2 and Mean values were based on a sample size of 8. The acoustic data from listener 8 do not appear because recorded material was destroyed due to equipment malfunction (he was not included in the regression lyses). The mean F_3 measurements are not as accurate as are the other formant measures for the high vowels, since the amplitude of this formant for many tokens was so low as to not allow spectrographic anal

Results

Raw Scaling Results

In order that the perceptual distance data be as free of random error as possible (the individual perceptual variation of interest may be very small and easily masked by random variation) and to ensure that each subject's data represent reliable, and replicable, scaling judgments, consistency statistics were done on the obtained perceptual data. Correlations were obtained among the four trials in each of the six stimulus sets from each of the 16 subjects (i.e., correlations were obtained between trials 1 and 2, 1 and 3, ..., 3 and 4). A listener's responses were required to exhibit a correlation coefficient of at least 0.44 between every trial within each of the six sets of stimulus vowels scaled. The 0.44 value ensures a level of significance for a two-tailed test of at least $p < 0.01$.

Data from 11 of the 16 subjects demonstrated the critical consistency and were used in all the analyses. For the INDSCAL analyses these raw data were collapsed into perceptual distance matrices (6 stimulus sets by 11 listeners meant 66 separate distance matrices) and symmetricized. Each distance matrix was the result of pooling data across the four separate trials of a particular speaker stimulus set. Each perceptual distance matrix is a vowel by vowel matrix (9×9) with cell values corresponding to the estimated perceptual distance between the row vowel and the column vowel.

Analysis of Variance

Two separate analyses of variances (ANOVA) were done before the similarity judgments were collapsed into perceptual distance matrices to determine whether the responses demonstrated significant variance with regard to either Listener or Speaker. The two separate three-way ANOVAs done were (1) Speaker by Stimulus pair by Trial and (2) Listener by Stimulus pair by Trial.

The first ANOVA indicated a significant main effect due to Stimulus pair ($F = 23.09$, $p < 0.001$) which contributed to the largest proportion of the variance, but neither of the other two main effects (Trial and Speaker) were significant. Though several of the interaction effects were marginally significant, none accounted for more than 3% of the variance. The second ANOVA of course again indicated a significant main effect due to Stimulus pair and no main effect due to Trial. However, the main effect due to Listener was significant ($F = 24.0$, $p < 0.001$) as was the Listener by Stimulus pair interaction ($F = 9.28$, $p < 0.001$). Other interactions were either statistically nonsignificant or trivial in terms of strength or influence. These ANOVA results thus suggest (as did HIERCLUS analyses not discussed here) that subjects (corresponding to the Listener variable) differed perceptually across a number of different stimulus sets.

Given this fact, our concern is directed to the second question: Can such perceptual variation among listeners be related with variation in some aspect of the acoustic space of their own vowels? To answer this question it is necessary to compare some measure(s) of each listener's acoustic space with some measure(s) of the perceptual responses of each listener. The acoustic measures to be used in this comparison will be those collected from each listener during the scaling experiment. The perceptual measures will be obtained through the use of multidimensional scaling.

INDSCAL Analysis

One of the more useful and important methods of analyzing perceptual distance data is multidimensional scaling. This procedure views similarity judgments among a set of objects as indicating the perceptual distance between those objects. A multidimensional scaling analysis accounts for this set of perceptual distances by optimally modeling the underlying perceptual space, that is, the stimuli are seen as objects in an n-dimensional space and the distance between the objects in this space reflect the experimentally obtained perceptual distances. Following appropriate rotation of the coordinate axes of this space, these dimensions are seen to represent the perceptual features utilized by the subjects in making their similarity judgments [see CARRO.L and WISH, 1974, or KRUSKAL and WISH, 1978, for a more detailed discussion of these techniques].

INDSCAL [CARROLL and CHANG, 1970], the multidimensional scaling procedure used here, additionally considers differences in perceptual structures across subjects and determines for each individual subject the salience or importance of each perceptual dimension. The model assumes that when subjects differ in judgments of similarity within the same set of stimuli they do so in terms of the relative importance which they assign to each of the relevant perceptual dimensions. As GANDOUR and HARSHMAN [1978] note, INDSCAL (and HARSHMAN'S PARAFAC) represents individual variations in judgments of similarity as systematic variations in the use of a common set of perceptual dimensions. One advantage which is obtained from consideration of individual differences is that the orientation of the axes of the resulting solution is fixed and cannot be rotated without worsening the overall fit of the solution to the perceptual data. Since the orientation of the axes is fixed, the claim that the dimensions have some sort of psychological reality is thereby strengthened.

In terms of the present study, four-way INDSCAL [CARROLL and ARABIE, 1980] was used which generated (1) a set of n-dimensional

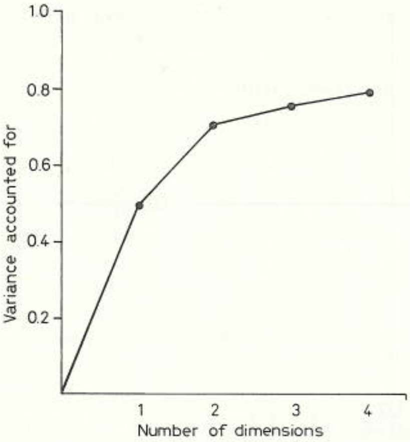

Fig. 1. Proportion of variance accounted for by INDSCAL solutions in one to four dimensions.

coordinates for each stimulus vowel, (2) a salience for each dimension for each Speaker condition, and (3) a salience for each dimension for each individual Listener. Since these Listener saliences indicate the relative importance which each individual subject places upon particular perceptual dimensions, they will serve as suitable perceptual measures against which to compare the acoustic measures of each listener.

INDSCAL Results

INDSCAL solutions were obtained in 1–5 dimensions. Each different dimensional solution was done several times at different random initial configurations. A 'fit' curve plotting the variance accounted for by the number of dimensions in the solution is shown in figure 1. Choice of the correct dimensionality is, of course, an important decision which involves several different criteria: First, one tries to find an 'elbow' in the fit curve which should appear at the correct dimensionality. An elbow would indicate that addition of one more dimension accounts for substantially less of the variance than did addition of the last dimension and is perhaps accounting for noise in the data. Second, INDSCAL theoretically produces a unique solution up to and including the 'correct' number of dimensions and at higher dimensionalities produces multiple solutions. If multiple solutions are found at n-dimensions which account for equal amounts of the variance (and thus do not reflect local minima), it suggests that n—1 dimensions are

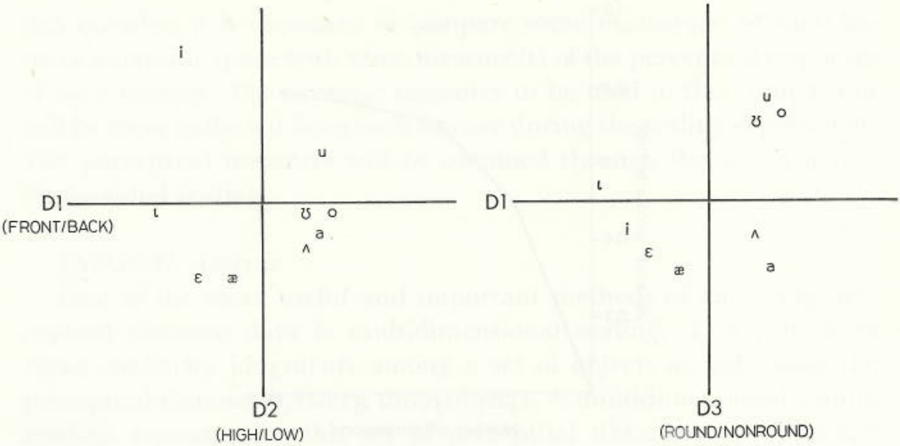

Fig. 2. Two different planes of the three-dimension INDSCAL solution.

required. Finally, although interpretation is subjective, one relies upon how 'interpretable' the extracted dimensions are. One would question the validity of a solution if the interpretation of the results radically differed from results of previous studies or corresponded to no known linguistic attribute of the stimuli. This last criterion is particularly important for dimensions accounting for very little variance.

The three-dimensional solution was selected as the correct solution in this study. Although the fit curve shows essentially no elbow at three dimensions, there were multiple solutions at four dimensions and each of the three dimensions extracted were easily interpretable (addition of a fourth dimension also caused a degeneration in the interpretability of the other three dimensions). Figure 2 shows two different planes of the INDSCAL solution. In terms of traditional linguistic categories, the three dimensions can be labeled front/back, high/low, and round/nonround, though clearly there is skewing in the empirically derived dimensions from what might be expected theoretically. For example, [i] was perceived as somehow very different from all the other vowels. Additionally, on the high/low dimension (D2) [a] is perceived as higher than [æ] or [ɛ]. In terms of relating these dimensions to acoustic characteristics of the stimulus vowels, correlations were obtained between the coordinate values for all nine vowels on each dimension and mean F_0, F_1, F_2, F_2–F_1, F_3 and duration values for the stimulus vowels. These correlations are shown in table III.

Table III. Correlations between mean acoustic measures and vowel coordinates on perceptual dimensions

Acoustic measures	Perceptual dimensions D1	D2	D3
F_0	0.385	0.371	0.347
F_1	0.324	0.634*	0.757*
F_2	0.978*	0.340	0.420
F_2–F_1	0.960*	0.463	0.136
F_3	0.823*	0.498	0.494
Duration	0.192	0.014	0.462

* Significant at 0.05 level or beyond, one-tailed.

D1 was best correlated with F_2 and F_2–F_1 while D2 and D3 were both better correlated with F_1 than with any other acoustic measure. The discovery of significant correlations between the front/back dimension (D1) and F_2, on the one hand, and the high/low dimension (D2) and F_1 on the other hand, is a reflection of the well-known fact that the first two formants are the most important determinants of phonetic quality and corroborates the results of numerous other perceptual studies [e.g., SINGH and WOODS, 1970; SHEPARD, 1972; FOX, 1981] and supports the acoustic interpretation of these two vowel features as suggested by LADEFOGED [1975]. But why should F_1 be significantly correlated with the rounding dimension (D3)? Note that the vowel feature rounding, unlike height or frontness, is not considered to be as strongly linked to any *single* acoustic parameter. For example, in English, degree of rounding increases as a function of height (and therefore F_1), but only in the back vowels (i.e., those with a low F_2–F_1). The corresponding perceptual dimension found in TERBEEK's [1977] study which was labeled 'rounding' was reasonably predicted by a weighted sum of *both* F_1 and F_2 ($r=0.93$). The significant correlation found here between F_1 and D3 indicates that the best single acoustic predictor of this dimension was F_1.

The measures of interest obtained via INDSCAL analysis in the present study, however, are not the vowel coordinates per se, but rather the saliences derived for each listener on each perceptual dimension. These weights indicate the relative importance of each perceptual dimension for each individual subject and thus are an indication of perceptual differences among individual subjects. The complete INDSCAL solution, including vowel coordinates, Speaker weights, and Listener saliences appear in table IV.

Table IV. Complete specification of the four-way three-dimensional INDSCAL solution

	Dimension 1	Dimension 2	Dimension 3
[i]	−0.463	−0.830	0.068
[ɪ]	−0.523	0.031	−0.106
[ɛ]	−0.329	0.330	0.217
[æ]	−0.130	0.322	0.435
[a]	0.301	0.097	0.410
[o]	0.331	0.049	−0.423
[ʌ]	0.216	0.178	0.226
[ʊ]	0.254	0.054	−0.357
[u]	0.316	−0.231	−0.471
Speaker saliences			
Set 1	0.629	0.658	0.641
Set 2	0.619	0.628	0.675
Set 3	0.623	0.655	0.674
Set 4	0.650	0.641	0.615
Set 5	0.663	0.615	0.592
Set 6	0.646	0.633	0.622
Listener saliences			
Listener 1	0.292	0.224	0.348
Listener 2	0.279	0.359	0.308
Listener 3	0.244	0.391	0.316
Listener 4	0.321	0.207	0.273
Listener 5	0.308	0.233	0.328
Listener 6	0.337	0.187	0.303
Listener 7	0.291	0.326	0.313
Listener 8	0.310	0.283	0.285
Listener 9	0.297	0.341	0.271
Listener 10	0.321	0.273	0.300
Listener 11	0.302	0.397	0.259

Regression Analysis

The hypothesis being evaluated is that a systematic relationship can be seen between a listener's perception of vowels (done on the basis of certain perceptual features) and his production of those vowels (seen in terms of the appropriate acoustic measures of his vowel space). This evaluation will be done using stepwise multiple linear regression; in particular, for each dimension the Listener saliences will be regressed against various measures of his acoustic vowel space. Of interest here will be the predictive abilities of corner vowel (/i, u, a/) measures as opposed to noncorner vowel (/æ, ʌ, o/) measures. The corner vowels are good indicators of the range and position of any talker's acoustic space and, indeed, some have suggested that talker normalization (done

Table V. Best prediction of Listener saliences using one, two or three variables in regression equation

Vowel Set	Variable name	r_m	Beta	F
Dimension 1				
Corner vowels				
	F_2–F_1u	0.496	−0.496	2.617
	F_2–F_1u F_2–F_1i	0.648	−0.536 −0.418	2.527
	F_2–F_1u F_2–F_1i F_3a	0.937	−0.543 −1.237 1.063	14.485***

by the perceptual system to eliminate speaker quality variation among different talkers) is done in terms of reference to these corner vowel measures. Any set of noncorner vowels will indicate something of the acoustic space of a speaker, but will not give as clear an estimate of the acoustic space of the speakers as do the corner vowels. Since we view the acoustic measures (particularly F_1, F_2, and F_2–F_1) of a subject's vowel space as directly reflecting variables such as tongue height and advancement, then if a direct link exists between vowel production and perception some systematic relationship should be evident between a subject's corner vowels and his perceptual measures – at least more than would exist between a subject's noncorner vowels and those same perceptual measures. The discovery of such a relationship would suggest a perception-production link, though clearly would not allow precise determination of how such a link manifests itself.

The goal of the multiple regression, then, was to determine how well acoustic measurements on a speaker's vowels would predict the salience of the three INDSCAL dimensions in his perceptual judgments of another speaker's vowels. Discussion of the multiple regression results to follow includes a summary of numerous analyses regressing the Listener saliences from each dimension against different sets of acoustic measures from two different sets of vowels: /i, u, a/ and /æ, ʌ, o/. The acoustic measures from each of these vowels used in these analyses included F_1, F_2, F_2–F_1, F_3, F_0, and duration. Table V summarizes the results from the regression analyses. Shown are the best predictors of Listener saliences at the one-, two- and three-variable levels using both corner and noncorner vowel acoustic measures[2].

---

[2] Since the number of items to be predicted was relatively small, discussion of results using more than three variables would be meaningless.

Table V. (continued)

Vowel Set	Variable name	r^m	Beta	F
Noncorner vowels (with F_0)				
	F_0	0.312	0.312	0.861
	F_0	0.571	0.660	1.688
	F_1æ		−0.591	
	F_0	0.683	0.690	1.748
	F_1æ		−0.699	
	F_2–F_1ʌ		−0.387	
Noncorner vowels (without F_0)				
	F_2–F_1ʌ	0.285	−0.285	0.702
	F_2–F_1ʌ	0.459	−0.466	0.934
	F_1o		−0.404	
	F_2–F_1ʌ	0.565	−0.285	0.939
	F_1o		−0.814	
	F_2–F_1æ		0.559	
Dimension 2				
Corner vowels				
	F_2–F_1a	0.660	0.660	6.179*
	F_2–F_1a	0.708	0.727	3.525
	F_1i		0.274	
	F_2–F_1a	0.843	0.970	4.919*
	F_1i		0.875	
	F_1u		−0.724	
Noncorner vowels				
	Dur. ʌ	0.371	−0.371	1.279
	Dur. ʌ	0.436	−0.338	0.823
	F_2–F_1ʌ		0.232	
	Dur. ʌ	0.470	−0.220	0.566
	F_2–F_1ʌ		0.360	
	F_2–F_1æ		0.238	
Dimension 3				
Corner vowels				
	F_1a	0.784	0.784	12.736***
	F_1a	0.873	0.989	11.171***
	F_3a		−0.435	
	F_1a	0.945	0.911	16.610***
	F_3a		−0.829	
	F_2i		0.566	
Noncorner vowels				
	F_1æ	0.638	0.638	5.495*
	F_1æ	0.792	1.109	5.900*
	F_2o		−0.665	
	F_1æ	0.888	1.027	7.46**
	F_2o		−1.259	
	F_2ʌ		0.767	

* $p < 0.05$; ** $p < 0.025$; *** $p < 0.01$.

For the corner vowels, the best predictors of D1 saliences were the F_2–F_1 measures of /u/ and /i/ at the two-variable level. This is quite interesting since D1 was interpreted as reflecting F_2–F_1 of the stimulus vowels. Regression analyses limited to acoustic measures other than F_2–F_1 indicated F_2u and F_2i among the best predictors – note that the correlation between D1 and F_2 of the stimulus vowels was second only to F_2–F_1. Analyses not including F_2–F_1 afforded a much worse prediction of the D1 saliences (e.g., use of F_1 values yielded a multiple correlation (r_m) of 0.487, N.S., at the three-variable level).

The best prediction at the three-variable level included F_3a. Predictions of D1 saliences using noncorner vowel measures were significantly worse. Without including F_0 the noncorner vowel measures were only able to yield $r_m = 0.565$ (NS); even with F_0 r_m at the three-variable level was only 0.683.

The best prediction of D2 saliences at the three-variable level included F_2–F_1a, F_1i, and F_1u ($r_m = 0.843$, $F = 4.919$, $p < 0.05$) in the regression equation. Again, one should note that D2, though slightly skewed, was interpreted as reflecting F_1 of the stimulus vowels. Regression analyses done after eliminating F_1 measures from the pool of variables from which the regression program could choose showed no significant prediction of the saliences at the three-variable level. For the noncorner vowel measures, the best three-variable prediction using Dur. ʌ, F_2–F_1, ʌ, and F_2–F_1æ yielded $r_m = 0.470$ (NS).

Examination of the D3 regression results again demonstrates that prediction of Listener saliences is better using corner vowel measures than noncorner vowel measures, though the differences in predictions are less dramatic in this case. The best single predictor for all vowels was F_1a (${}_m = 0.784$, $p < 0.01$) and for the noncorner vowels F_1æ ($r_m = 0.638$, $p < 0.05$). Note also that D3 also reflected F_1 of the stimulus vowels (even better than did D2).

Discussion

The experimental results described in this paper suggest a link between perception and production. Our data indicated consistent perceptual structure differences among listeners in judging vowel quality similarities across a number of different vowel stimulus sets (each stimulus set varying substantially in terms of speaker quality). In addition, the data have shown a high degree of correspondence between perceptual differences and articulatory differences (seen in terms of various acoustic measures) among the subjects. We draw this conclusion from the fact that acoustic measures of the corner vowels from listener consistently afforded better prediction of Listener saliences than did the noncorner vowel measures and the fact that the best predictors of these saliences on a particular perceptual dimension tended to be those related to the acoustic interpretation of that dimension. There was no reason to expect, a priori, that this would be the case if no direct relationship between perception and production existed. One explanation of how such a link might manifest itself might include reference to a prototype model of speech perception [Repp, 1977; Oden and Massaro, 1978] and would propose that the ideal internalized perceptual prototype of a particular vowel (defined in terms of features relating to important acoustic characteristics of that vowel) may be identical to the idealized acoustic/articulatory target at which a speaker aims during vowel production.

Obviously, these results do not, in and of themselves, constitute undeniable proof of the existence of such a link; they may be prone to different interpretations or may reflect some fortuitous perceptual data. In terms of the latter concern, clearly these findings must be replicable – this is a universal requirement in any experimental science. In terms of the first concern, one should note that most, if not all experimental results supporting a perception-production .link have been prone to different interpretations. However, as Liberman and Studdert-Kennedy [1978] remarked, 'no single piece of evidence is, by itself, wholly convincing; it is only the pattern that tells.' As the bits of evidence for a fairly concrete perception-production link mount from a variety of experimental sources, the more likely it is that such a link has a psychologically real existence.

References

Aungst, L.F.; Frick, J.V.: Auditory discrimination ability and consistency of articulation of /r/. J. Speech Hear. Res. *29:* 76–85 (1964).

Bailey, P.J.; Haggard, M.P.: Perception and production: some correlations on voicing of an initial stop. Lang. Speech *21:* 189–195 (1978).

Bell-Berti, F.; Raphael, L.J.; Pisoni, D.B.; Sawusch, J.R.: Some relationships between articulation and perception. Phonetica *36:* 373–383 (1979).

CARROLL, J.D.; ARABIE, P.: Multidimensional scaling. A. Rev. Psychol. *31:* 607–649 (1980).

CARROLL, J.D.; CHANG, J.J.: Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckert-Young decomposition. Psychometrika *35:* 283–319 (1970).

CARROLL, J.D.; WISH, M.: Multidimensional perceptual models and measurement methods; in CARTERETTE, FRIEDMAN Handbook of perception, vol. II: Psychophysical judgment and measurement (Academic Press, New York 1974).

COHEN, J.H.; DIEHL, C.F.: Relation of speech-sound discrimination ability to articulation-type defects. J. Speech Hear. Disorders *28:* 187–190 (1963).

COOPER, W.E.: Adaptation of phonetic feature analyzers for place of articulation. J. acoust. Soc. Am. *56:* 617–627 (1974).

COOPER, W.E.; LAURITSEN, M.R.: Feature processing in the perception and production of speech. Nature, Lond. *252:* 121–123 (1974).

COOPER, W.E.; NAGER, R.M.: Perceptuo-motor adaptation to speech: an analysis of bisyllabic utterances and a neural model. J. acoust. Soc. Am. *58:* 256–265 (1975).

DICKSON, S.: Differences between children who spontaneously outgrow and children who retain functional articulation error. J. Speech Hear. Res. *5:* 263–271 (1962).

EIMAS, P.D.; CORBIT, J.D.: Selective adaptation of linguistic feature detectors. Cognitive Psychol. *4:* 99–109 (1973).

FARQUHAR, M.S.: Prognostic value of imitative and auditory discrimination tests. J. Speech Hear. Disorders *26:* 342–347 (1961).

FOX, R.A.: Perceptual structure of English monophthongs and diphthongs. 17th Regional Meet. Chicago Ling. Soc., Chicago 1981.

GANDOUR, J.T.; HARSHMAN, R.A.: Crosslanguage differences in tone perception: a multidimensional scaling investigation. Lang. Speech *21:* 1–33 (1978).

HAGGARD, M.P.; CORRIGALL, J.M.; LEGG, A.G.: Perceptual factors in articulatory defects. Folia phoniat. *23:* 33–40 (1971).

HARSHMAN, R.A.: Foundations of the PARAFAC procedure: models and conditions for an 'explanatory' multidimensional factor analysis. Working Papers Phonet., No. 16 (University of California, Los Angeles 1970).

KRONVALL, E.L.; DIEHL, C.F.: The relationship of auditory discrimination to articulatory defects of children with no known organic impairment. J. Speech Hear. Disorders *19:* 335–338 (1954).

KRUSKAL, J.B.; WISH, M.: Multidimensional scaling (Sage Publications, London 1978).

LADEFOGED, P.: A course in phonetics (Harcourt, Brace, Jovanovich, New York 1975).

LADEFOGED, P.; DECLERK, J.; LINDAU, M.; PAPÇUN, G.: An auditory-motor theory of speech production. Working Papers Phonet., No. 22, pp. 48–75 (University of California, Los Angeles 1972).

LIBERMAN, A.M.; COOPER, F.S.; SHANKWEILER, D.P.; STUDDERT-KENNEDY, M.: The perception of the speech code. Psychol. Rev. *74:* 431–461 (1967).

LIBERMAN, A.M.; STUDDERT-KENNEDY, M.: Phonetic perception; in HELD, LIEBOWITZ. TEUBER Handbook of sensory physiology, vol. VIII: Perception (Springer, Berlin 1978).

LINDAU, M.: Vowel features. Language *54*: 541–563 (1978).

MADISON, C.L.; FUCCI, D.J.: Speech-sound discrimination and tactile-kinesthetic discrimination in reference to speech production. Percept. Mot. Skills *33:* 831–838 (1971).

MARQUARDT, T.P.; SAXMAN, J.H.: Language comprehension and auditory discrimination in articulation deficient kindergarten children. J. Speech Hear. Res. *15:* 382–389 (1972).

MCREYNOLDS, L.; KOHN, J.; WILLIAMS, G.: Articulatory-defective children's discrimination of their production errors. J. Speech Hear. Disorders *40:* 327–328 (1975).

MONNIN, L.M.; HUNTINGTON, D.A.: Relationship of articulatory defects to speech-sound identification. J. Speech Hear. Res. *17:* 352–366 (1974).

Oden, G.C.; Massaro, D.W.: Integration of featural information in speech perception. Psychol. Rev. *85:* 172–191 (1978).

Powers, M.H.: Functional disorders of articulation-symptomatology and etiology; in Travis Handbook of speech pathology (Appleton Century Crofts, New York 1957).

Repp, B.H.: Dichotic competition of speech sounds: the role of acoustic stimulus structure. J. exp. Psychol. hum. Perception Performance *3:* 37–50 (1977).

Shepard, R.N.: Psychological representation of speech sounds; in David; Denes: Human communication: a unified view (McGraw-Hill, New York, 1972).

Singh, S.; Woods, G.: Perceptual structure of 12 American vowels. J. acoust. Soc. Am. *49:* 1861–1866 (1970).

Stevens, K.N.: Segments, features, and analysis by synthesis; in Kavanagh, Mattingly The relationships between speech and reading (MIT Press, Cambridge ,Mass. 1972).

Stitt, C.; Huntington, D.: Some relationships among articulation, auditory abilities, and certain other variables. J. Speech Hear. Res. *12:* 576–593 (1969).

Summerfield, Q.; Bailey, P.J.; Erickson, D.: A note on perceptuo-motor adaptation of speech. J. Phonet. *8:* 491–499 (1980).

Terbeek, D.: A cross-language multidimensional scaling study of vowel perception. Working Papers Phonet., No. 37, pp. 1–271 (University of California, Los Angeles, 1977).

Van Riper, C.: Speech correction: principles and methods (New York 1963).

Waldman, F.R.; Singh, S.; Hayden, M.E.: A comparison of speech-sound production and discrimination in children with functional articulation disorders. Lang. Speech *21:* 205–220 (1978).

Weiner, P.: Auditory descrimination and articulation. J. Speech Hear. Disorders *32:* 19–38 (1967).

Received: May 26, 1981; accepted: September 23, 1981

Dr. Robert Allen Fox, Cunz Hall of Languages, Department of Linguistics, The Ohio State University, 1841 Millikin Road, *Columbus*, Ohio 43210 (USA)