NUMERICAL SIMULATION OF VOWEL QUALITY SYSTEMS: THE ROLE OF PERCEPTUAL CONTRAST

JOHAN LILJENCRANTS and BJÖRN LINDBLOM

Royal Institute of Technology, Stockholm

A numerical model is developed in order to establish the extent to which the principle of maximal perceptual contrast can be used in phonological theory to explain the phonetic structure of vowel systems. Preliminary results obtained with the model indicate that perceptual contrast appears to play an important role as a determinant of such systems. Therefore, it is likely that this principle (along with other factors) should be included among the variables in an explanatory phonological theory. However, the incorporation of numerically stated conditions on phonological structure appears to presuppose a formalism different from that which has developed within current descriptions of phonology. Some refinements and extensions of the present framework are suggested. It is proposed that predictions of phonological facts be derived as consequences of the structure of the mechanisms available for human speech communication and the optimization of their use. Such an extension would constitute a theory that would be different from traditional 'Saussurean' linguistics in several respects; e.g., it would be quantitative, and deliberately substance-based. The research reported represents a preliminary attempt to apply such a program.*

1. THE ROLE OF PERCEPTUAL CONTRAST. In this section we shall describe an attempt to predict the phonetic structure of vowel systems, based on a numerical interpretation of the principle of maximal contrast. This principle has a long tradition in linguistics; see, e.g., Jakobson 1941, Martinet 1955 (quoting Passy 1890:227), and de Groot (1931:121). It is discussed by Moulton 1962, and more recently by Wang (1968a:34), who applies it in an interpretation of formant frequency data. It is not, however, mentioned by Chomsky & Halle 1968 in their attempt to develop a theoretical account of phonological systems. Our aim is to evaluate, albeit in a preliminary fashion, the role that perceptual contrast plays in vowel quality systems, and to try to assess the explanatory value of this principle in relation to other possible factors. Seeking a quantitative formulation of the problem, we have made certain assumptions about the acoustic space universally available for vowels, the perceptual representation of this space, and the quantitative interpretation of the notion of contrast.

1.1. THE VOWEL SPACE AND THE QUANTIFICATION OF CONTRAST. A vowel sound can be characterized acoustically in terms of its formant pattern, i.e. its lowest three formant frequencies. However, only a subset of all logically possible combinations of those formant frequencies is associated with human vowels. Thus a formant pattern of F_1 = 100 Hz, F_2 = 200 Hz, and F_3 = 300 Hz would be a reasonable description of a machine-made 'vowel sound', but an absurd statement about a natural vowel. An articulatory model of speech production, if designed so as to reflect the natural degrees of freedom of the vocal tract, would provide a basis for defining this subset. The definition of the human

* This work was supported by the National Institutes of Health under Research Grant NS 04003-08.

vowel space adopted in the present study is based on the framework of Lindblom & Sundberg 1969, 1971. This model (henceforth, the LS model) offers a procedure for deriving an acoustic output characteristic of vowels, based on specifications of the position of the jaw, the state of the lip muscles, the shape and position of the tongue body, and the position of the larynx. Data on these parameters are used to calculate vocal tract shape and the cross-sectional areas along the tract. The latter in turn serve as input information in the derivation of the final output, the formant frequency pattern. The vocal tract dimensions are closely similar to those of a typical male speaker. The built-in articulatory constraints of the model delimit the range of vowel sounds that can be generated, and accordingly imply a hypothesis about the vowel qualities that we should expect to be possible in human speech. In other words, this is a hypothesis about a linguistic UNIVERSAL: the acoustic space available for vowels. The vowels that can be produced with the LS model are located in a three-dimensional acoustic space whose dimensions are the first three formant frequencies; its properties are discussed in Lindblom & Sundberg 1969. An approximate representation is shown in Figures 1a–b. Fig. 1a depicts the space in terms of the first and the second formants, and Fig. 1b in terms of the third and the second formants.

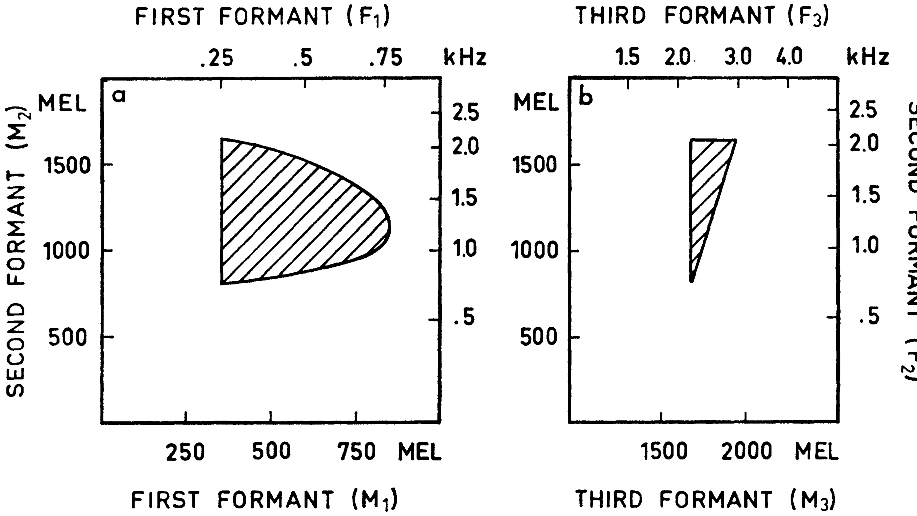

FIGURE 1a FIGURE 1b

By the transformation of the linear frequency scales into mel scales, the space assumes a shape that is more satisfactory from an auditory point of view. All calculations to be reported below were carried out using this 'perceptual' representation. As a working definition of contrast between two arbitrary vowels, we selected a measure depending on the linear distance in mel units between the points representing those vowels. The criterion used to maximize intervocalic distances, or perceptual contrast, is given by

$$\sum_{1}^{m} 1/r_i^2 \rightarrow \text{minimized}$$

where r refers to the distance between the ith pair of vowels, and the number of pairs per system is $m = n(n - 1)/2$ where n is equal to the number of vowels in the system.

This formula is adopted from physics, where it is applicable to the computation of forces in potential fields. To illustrate, let us consider the analogy of two particles with an equal electrical charge. They will repel each other with a force that is inversely proportional to the square of their distance. If we now place these particles in a limited space within which they can move freely, then the particles will move away from each other because of the force of repulsion. Eventually they will hit the boundary of the space, and then possibly move along the boundary, if their mutual distance can be increased that way. Finally, an equilibrium is reached where their distance cannot be increased any more. Characteristic of this state is the fact that the mutual energy has reached a minimum. If other particles are introduced into the space, the whole set will move to new positions, always fulfilling the very general equilibrium criterion, that of minimal energy. For this analogy, we must of course assume that the energy released when the particles move apart is dissipated in some way; otherwise the system would exhibit perpetual oscillations, like the molecules in a gas.

Other physical phenomena exhibit a force–distance dependence of the same type. In gravitation, the force will have the wrong direction for use as an analogy here. Another example is magnetic force. It is interesting to note that the energy minimization that we have performed computationally in this study can also be made with a simple physical experiment. For this purpose the repulsion between similarly oriented magnets may be used. A set of small permanent bar magnets is attached to cork floats in such a way that the assemblies can float on a water surface with the magnets having a vertical polar axis. A two-dimensional boundary of suitable shape may be constructed at the water surface, using e.g. a non-magnetic metal wire. The system will quickly settle to an equilibrium, and the energy released will be dissipated in the water by the viscous friction. This experiment is very illustrative and easy to perform, and it can be recommended for qualitative studies of the minimum-energy model.

To make computations somewhat easier, r is interpreted as a two-dimensional distance. All calculations are made within the hypotenuse plane of the space (see Fig. 1). This representation of the space can be said to correspond to that in the M_1–M_2 plane (Fig. 1a), corrected with respect to the third formant. After this correction is introduced, the properties of the vowels can be given in terms of their mel coordinates: M'_2 (second formant after third-formant correction) and M_1 (first formant). The curved contours enclosing the vowels are

$$(1)\quad M_2' = 1150 + (M'_{20} - 1150)\sqrt{(850 - M_1)/(850 - 350)}$$

$$350 \leq M_1 \leq 850$$

where $M'_{20} = 1700$ for the upper part of the contour and $M'_{20} = 800$ for the lower part. For $M_1 = 350$, M'_2 varies between these values.

1.2. COMPUTER IMPLEMENTATION. The data obtained so far in this study are the output of a computer program. The input to the program is the number of points (n) to be placed in the vowel plane as defined above. As an arbitrary initial condition, the points are evenly placed on a circle of radius 100 mel, with its center at $M_1 = 600$ and $M'_2 = 1200$ mel. Then the total energy measure is computed according to the formula

$$(2)\ \mathrm{E} = \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} 1/r_{ij}^2 \qquad \text{where } r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2.$$

Each term in the sum will thus represent the 'force' between points number i and j, and the double sum will cover all the mutual forces between the n points. The ith point is represented by its coördinates in the M_1/M'_2 plane which are denoted by x_i and y_i, respectively.

The task for the program is now to modify the coördinates until a minimum value of E is found. The elementary procedure to modify one of the points is as follows. First the point is moved a certain distance, and checked for being still inside the boundary; if this is the case, a new value of E is computed. This is repeated for a number of directions, usually six, out from the original location of the point. Then the optimum direction is selected, and the point is repeatedly adjusted in this direction until either the boundary is hit or E no longer decreases. Then another direction search is made, and so on until a minimum for this point is established. This whole procedure is repeated for all the points. After this has been done once, it is started from the beginning one or more times until E does not decrease any more. Usually the procedure is performed a first time with the adjustments of the points in 100 mel steps, and a second time with 10 mel steps.

It can be easily understood that the computation time for the iterations grows rapidly with an increasing number of points n. For simplicity, the procedure was written in the interpretive FOCAL programming language, which is not very fast in execution. For these reasons the computations were very time-consuming. In general, the computer was left overnight to compute the solutions for $n = 3$ to $n = 12$.

For a more thorough investigation of possible solutions, the iteration procedure will have to be improved. Special problems arise when the boundary is reached, since the pilot searches always tend to go outside the boundary, and the resulting effect is very similar to the mechanical friction effect that will be obvious if the magnet experiment is performed. Probably a more satisfactory operation can be obtained if the boundary is not 'hard' but in itself repels the points also at a distance. This would be reflected in the E formula as terms containing the distance between the points and the boundary.

1.3. RESULTS. Formant frequency values have been obtained for the vowels of systems ranging in size from three through twelve vowels. The formant frequency data are represented in Figure 2, which shows the horseshoe-shaped vowel area in terms of linear frequency scales for the second formant (ordinate) and first formant (abscissa). The horizontal and vertical lines correspond to divisions of F_1 at every 200 Hz from 200 to 800 Hz and of F_2 at every 500 Hz from 500 to 2500 Hz. In the top row the rightmost figure contains all the data points shown in the other diagrams. We shall discuss these results in terms of a 'broad' transcription. Symbols have been assigned to points along the contour

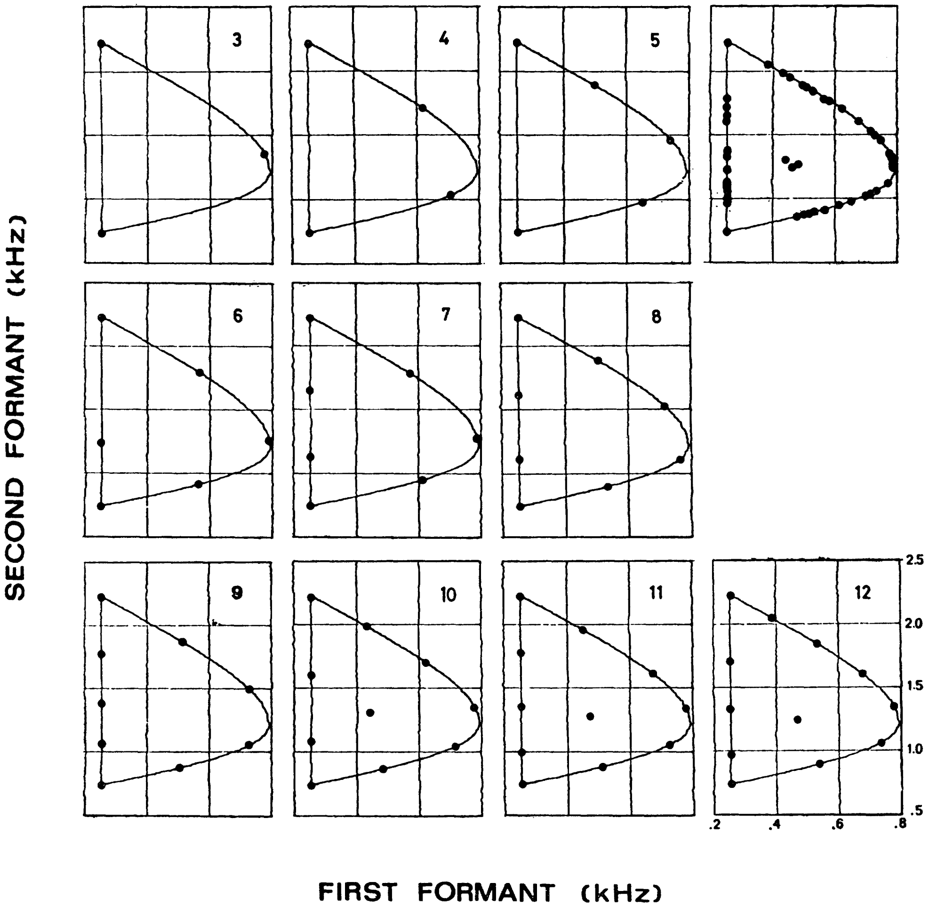

FIGURE 2

of the space, since this is primarily where the data points tend to occur. Formant patterns in Table 1 (cf. also Figure 3) were transcribed as indicated at the top of each column.

Points were selected at approximately equal distances in F_1 for vowels differing in opening, as well as in F_2 for the close series [u ɨ ʉ ü i]. The labeling in Table 1 and Figure 3 is not incompatible with the auditory values that might be assigned to synthetic realizations of these formant patterns. The occurrence of a

	i	e	ɛ	æ	a	ɑ	ɔ	o	u	ɨ	ʉ	ü	ə
F_1	250	400	550	700	800	675	550	400	250	250	250	250	450
F_2	2225	2000	1775	1600	1250	1000	925	825	750	1100	1475	1850	1200
F_3	3000	2935	2800	2700	2500	—	—	—	—	—	2650	2850	—

TABLE 1

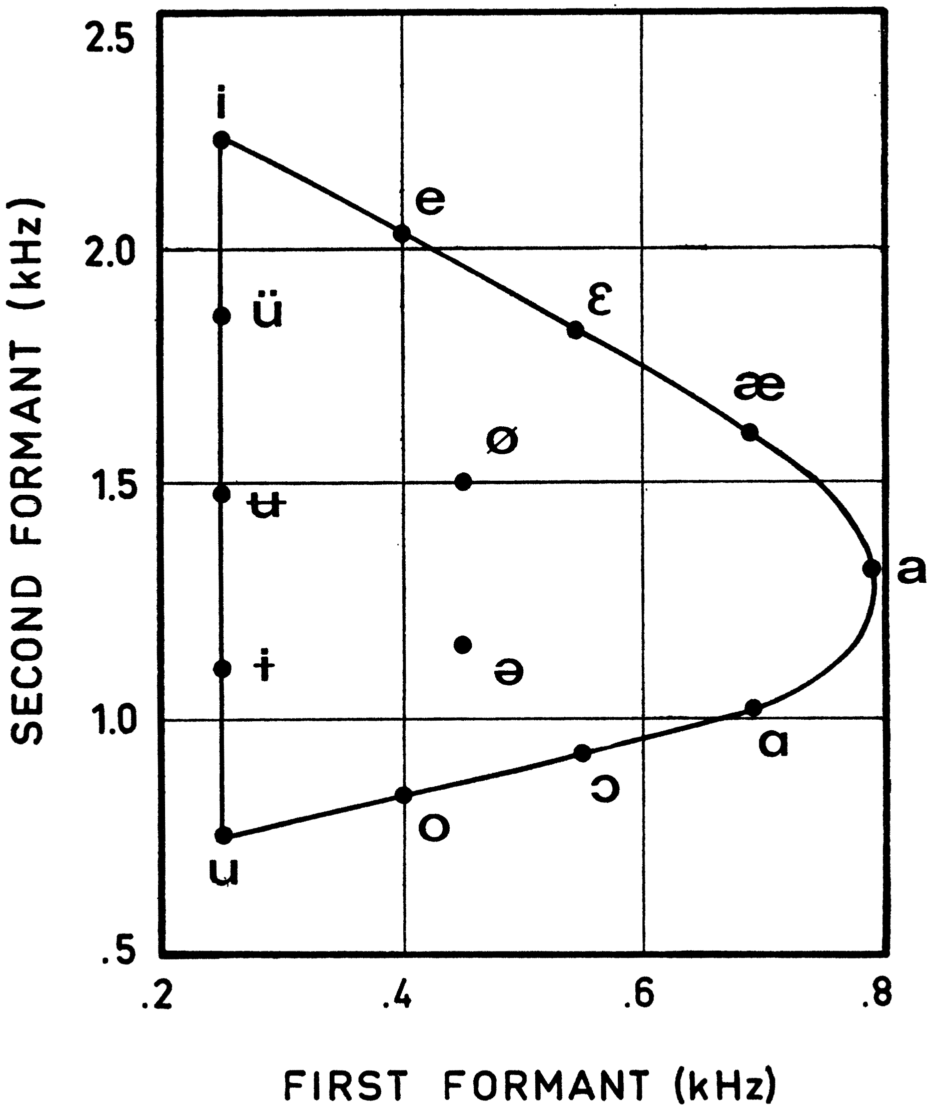

FIGURE 3

given symbol in the predicted systems below should be interpreted as indicating that the formant pattern in question was closer to the point corresponding to that symbol than to any other point. Since F_3 is known to be an important determinant of vowel quality in the case of front vowels, it has also been included in the table. Owing to the computational constraints imposed, its values are somewhat higher than normal. This is a second-order effect, however. In spite of the broadness of our transcription, the predictions are more narrowly specified than the data available in the literature; this should be borne in mind when evaluating the predictions. The data have been taken principally from three sources: Trubetzkoy 1929, Hockett 1955, and Sedlak 1969. These authors have based their vowel-system typologies on phonemic analyses, and often fail to comment on fine phonetic details.

(3) THREE-VOWEL SYSTEMS:

Predicted:

i		u
	a	

Observed:

(a)

i		u
	a	

(b)

i		
	o	
a		

(c)

ə
e
a

3a, Trubetzkoy reports, occurs 'im Lak (Zentral-Dagestan), im Neupersischen und im Arabischen'. In footnotes, he refers to the fact that all these languages show allophonic variation of quality in certain consonantal environments. Hockett exemplifies 3a with 'Arunta, Cree, Eskimo (most dialects, perhaps not all), some Arabic dialects (including Iraqi), Salishan (except Coeur d'Alene ...), Muskogean (except Creek-Seminole), Ojibwa, Kechua before the introduction of Spanish loans, Totonac, Lak and Wishram'. Sedlak's list contains some additional examples: Bella Coola (Salish), Hiligaynon (Austronesian), Gugu-Yalanji, Nyangumata, Western Desert (all Australian). In the latter two, the inventory of vowel segments is doubled by a length distinction.

3b is exemplified once in Sedlak's survey, where it is attributed to Mikasuki (Muskogean). There is also a length distinction.

3c, a 'linear' vowel system, is described by Trubetzkoy (1958:87). He attributes such systems to Caucasian languages, viz. Abkhaz and Adyge, and with some hesitation to Ubykh. The phonetic realizations of these vowels exhibit rich consonant-determined variation.

The investigations made show 16 examples of 3a, three examples of 3c, and a single example of 3b.

(4) FOUR-VOWEL SYSTEMS:

Predicted:[1]

i	u
ε	a

Observed:

(a)

i	ə
ä	a

(b)

i	o
e	a

(b′)

i		
e		o
	a	

(c)

i		u
e		
	a	

(d)

i		u
		o
	a	

(e)

i	ɨ	u
	a	

(f)

i		u
	ə	
	a	

(g)

i	ï
e	o

4a is attributed by Trubetzkoy to Rutul. He states that /ə/ → [u] after *w* and rounded consonants, which makes the similarity to the predicted system somewhat greater. Compare also the 'open' character of the predicted [ε] vowel with Trubetzkoy's choice of symbol for this member of the system.

[1] The ε symbol stands for a quality closer to [æ] than to [e].

4b is the arrangement used by Hockett to describe Rutul: 'the high back vowel is sometimes rounded, sometimes not, depending on environment.' In this connection he also mentions Fox and Shawnee, which exhibit an unrounded/rounded alternation in the realization of the /a/ vowel. 4b is further claimed to occur in 'Apachean, Campa, Chatino nasal vowels, many dialects of Nahuatl (not Tetelcingo), Creek-Seminole (/e/ rare and limited to occurrence before /y/), and Wichita'. Sedlak's version of 4b′ is used to describe Campa, Klamath and Nahuatl (which both have a length distinction), Apache and Seneca (where there are distinctions of length and nasality).

4c is mentioned only by Sedlak; it is supposed to be found in Ignaciano, Tacana, Shawnee (cf. Hockett under 4b), and Wichita (cf. Hockett), which has a long–short distinction.

4d approximates Hockett's analysis (p. 86) of 'Cebuano Bisayan, and perhaps some dialects of Tagalog'. These languages are said to have [i] and [e] which 'are both heard as allophones of a single vowel'. Sedlak lists Kalinga and Crow, the latter with long–short oppositions.

4e covers Sedlak's Binokid, Bontoc, Ilocano, and Ivatan, and Hockett's Amahuaca. Sedlak points out that phonetically Ivatan may differ somewhat from the others with respect to /ɨ/. Hockett puts forward the possibility that, from a phonetic point of view, Amahuaca should be placed in 4f.

4f includes Sedlak's Squamish and Pashto (with a special system of phonemically long vowels), and Hockett's Ilocano and Dibabaon.

4g is Sedlak's Arapaho.

There is clearly some inconsistency among the authors quoted with respect to phonetic interpretation. If we disregard all the unclear cases, there is still a majority of examples for 4b. The predicted values appear to be closest to 4c and possibly 4a.

In their discussion of marking conventions for vowels, Chomsky & Halle (1968:410) predict that an optimal four-vowel system will consist of /i a u/ + either /æ/, /e/, or /o/. Cases 4c and 4d provide partial evidence for this prediction. They further state that 'it does not seem implausible that there is an optimal four-vowel system—namely /a i u æ/' (cf. our predicted 4-vowel system).

(5) Five-vowel systems:

Predicted:		i		u
		ɛ		ɑ
			æ/a	
Observed:	(a)	i		u
		e		o
			a	
	(b)	i	ü	u
		e		a
	(c)	i		o
			ə	
		e		a

(d)	i	ü	ï	u
		a		
(e)	i	ɨ		
		ə	o	
		a		

The predicted qualities resemble those of 5a most closely. There are differences with respect to the realizations of the 'back mid' vowel and the 'low' vowel. However, a narrow phonetic representation of the /o/ would probably be [ɔ] in most cases. The predicted 'low' vowel denoted by *æ/a* is a 'front' variety of [a] approaching cardinal vowel number four.

5a is given by Sedlak for 27 languages, with another 27 that display various patterns of length, nasalization, and retroflexion oppositions in addition to the five basic vowels. His list can be supplemented with examples from the other two sources used, the total number of cases being about 65: Ainu, Aniwa, Arabela, Benabena, Buhid, Gahuku, Ganda, Greek (Cypriot and literary Dhimotiki), Hawaiian, Ifugao, Japanese, Kamano, Maori, Movima, Mundari, Negrito, Neo-Aramaic, Pocomchi, Serbo-Croatian, Siane, Spanish, Tolai, Tzeltal, Wiyot, Yaqui; (with length distinction) Aguatec, Bedauye, Chitimacha, Huasteco, Khasi, Kolami, Kota, Lushai, Malayalam, Piro, Tojolabal, Tonkawa, Wik-Munkan, Yuma, Zuni; (with additional distinctions) Chinantec, Saho, Slave; Badaga; Jamaican Creole; Burmese, Dakota, Fanti, Ioway-Oto; Karok; Chatino; Arabic (Syrian).

5b is given by Trubetzkoy for Tabassaran and Kyuri. Hockett adds Huichol which has '/ɨ/ instead of /ü/'.

5c is given by Hockett for Potawatomi.

5d is given by Sedlak for Chacobo.

5e is given by Sedlak for Manobo.

(6) SIX-VOWEL SYSTEMS:

Predicted:

i	ʉ	u
ɛ		ɔ
	a	

Observed:

(a)

i	ü	u
e		o
	a	

(b)

i	ɨ	u
e		o
	a	

(c)

i		u
e		o
ɛ		
	a	

(d)

i		u
	ə	
e		o
	a	

(e)

i		u
e		o
		ɔ
	a	

(f)

i	ü		r	u
		ə		
		a		

6a is given by Sedlak for Basque and Taki-Taki, and by Trubetzkoy for Ukrainian and for late Classical and early Medieval Greek.

6b is given by Hockett for Cashibo, Bulgarian, Mazahua, Otomi nasal vowels, Mixteco, Sierra Popoluca, and Zoque; by Sedlak for Gilyak, Itonama, Kalagan, Miwok, Mongol (Dagor), Sangir, Zoque, and Siona (with nasalized oppositions); and by Trubetzkoy for Malay.

6c is given by Hockett for Chipewyan, Dargwa, Menomini, Persian, Ukrainian, and Yuchi; by Sedlak for Bats, Galla, Persian, Zapotec, and for Chipewyan (with additional length and nasalization distinctions); and by Trubetzkoy for Dargwa. Trubetzkoy (1958:90) also attributes a pattern to Uzbek (cf. 6e) which is similar to 6c. He represents the lowest two vowels as æ and ɔ.

6d is given by Sedlak for Angami, Asmat, Chontal, Fore, Gaddang, Kanite; Atsugewi, Yurok, and Vogul have additional systems.

6e is given by Sedlak for Uzbek.

6f is given by Hockett for Mandarin Chinese.

The predicted system seems closest to 6b, the [ʉ] symbol standing for a vowel quality intermediate between [ɨ] and [ü], as these symbols were defined initially in this section. The second best match is 6a. The largest number of examples was found for 6b, 6c, and 6d. Again the assignment of a given language to different types by the authors points to the need for more reliable phonological as well as phonetic data.

(7) SEVEN-VOWEL SYSTEMS:

Predicted:

i	ü/ʉ	ɨ	u
ɛ			ɔ
	a		

Observed:

(a)	i		u
	e		o
	ε		ɔ
		a	
(b)	i	ü	u
	e	ö	o
		a	
(c)	i	ɨ	u
	e	ə	o
		a	
(dα)	i	ɨ	u
	e		o
	ε		a
(dβ)	i	ɨ	u
	e		o
	æ	a	
(dβ′)	i		u
	e	ə	o
	æ	a	
(dγ)	i	ь	u
	e		o
	ä		a
(dγ′)	i	y	u
	e	æ	o
		a	
(e)	i		u
	e	ə	o
	æ		a
(e′)	i		u
	e	ə	o
	æ		ä
(f)	i		u
	e	ə	o
		a	ɔ
(g)	i		u
	e ö		ə o
		a	
(h)	i *v̈*		*v* u
	e		o
		ɑ	

7a is given by Hockett for 'Italian, late Latin (?), some dialects of Portuguese (oral vowels), Tunica stressed vowels, Loma, Kiowa, Tetelcingo Nahuatl, and the oral syllables of Bariba, Senadi, and Supide'; by Sedlak for Italian, Portuguese, Susu, and Tunica (basic system also in Ibibio, Tswana, Yoruba, Bengali, Loko, Ewe, Baule, Gbeya, Senufo-Senadi, Slovene, Sango, Wolof); and by Trubetzkoy (1929:116, 1958:92, 110) for Burmese, Italian, and Ostyak.

7b is given by Hockett for 'German, Dutch (some dialects and over-all pattern for standard Dutch), Hungarian (some dialects), French (some dialects, particularly in the south), Zyryan'; by Sedlak for German, Hungarian (with length contrasts), and French (minimal system); and by Trubetzkoy for Zyryan (1929:117) and Mongol (1958:104).

7c is given by Hockett for Terena, Maidu, Rumanian; and by Sedlak for Ossetic, Rumanian, Kashmiri (with *ɔ* for *ə*), and Rawang (with *ü* for *ɨ*).

7dα is given by Hockett for 'some Bulgarian dialects and perhaps in Votyak'. He continues: 'The other version of this (with /ü/ instead of /ɨ/) is not attested.'

7dβ is given by Sedlak for Tamil, and 7dβ′ for Sinhalese.

7dγ is given by Trubetzkoy for 'einiger archaischen neubulgarischer Dialekte'; and 7dγ′ for Votyak (cf. quotation from Hockett above).

7e is given by Hockett, and 7e′ by Sedlak for Lifu.

7f is given by Sedlak for Tagabili.

7g is given by Sedlak for Iraqw.

7h is given by Trubetzkoy (1958:111): 'im zentral-chinesischen Dialekt Siang-tang (Provinz Honang) ... *v*, *v̈* sind die charakteristischen "summenden" (gingivalen) Vokale, die in vielen chinesischen Dialekten vorkommen.' (These vowels are [+coronal] in the terminology of Chomsky & Halle.)

No observed system exhibits both an [ɨ] and an [ʉ] or [ü], that is, three 'high' vowels as predicted. System 7a appears to be more frequent than other systems in the sources consulted.

(8) EIGHT-VOWEL SYSTEMS:

Predicted:

i		ü/ʉ		ɨ	u
	ɛ			ɔ	
		æ	a/ɑ		

Observed:

(a)	i		ü		ɨ	u
	e		ö		a	o
(a′)	i		ü		ɨ	u
		e		ö	o	
				a		
(b)	i			ü		u
	e			ö		o
		æ			a	
(c)	i		ü		(u̇)	u
	e		ö		ə	o
					a	

(d)	i		ɨ		u
	e		ə		o
		a		ɔ	
(e)	i		ɨ		u
	e				o
	æ		a		ɔ
(f)	i				u
	e				o
	ɛ				ɔ
	æ				a
(g)	i				u
	e				o
	ɛ		ʌ		ɔ
			a		

8a is given by Hockett for Turkish; by Sedlak for Balkar, Chuvash, Tartar (with *ĭ* instead of *ɨ*), and Kirghiz (with *ɨ*); and by Trubetzkoy 'im Osmantürkischen und in vielen anderen Türksprachen' (1929:119) (his transcription is *o a ö ä* for the [−high] vowels and *u y ü i* for the [+high] vowels.)

8a′ is the same system, redrawn to facilitate comparison with the predicted pattern.

8b is given by Hockett for 'Finnish and some Hungarian dialects'; by Sedlak for Albanian (long vowels, with *ɛ* instead of *æ*); and by Trubetzkoy for Finnish (1929:117) and Northern Albanian (1958:111).

8c is given by Sedlak for Cheremis (with *ü*) and Khalkha (with *u̇*). No comment is made to clarify the phonetic difference between these two symbols. However, it is stated that 'Khalkha ... differs from Cheremis ... in that the *u̇* is usually high central rounded in its phonetic manifestations, although its function in the vowel harmony system is analogous to Cheremis *ü* and thus the same distinctive features are relevant ...'

8d is given by Hockett for 'Cuitlateco and perhaps some dialects of Korean ... If Korean dialects have this pattern, the low vowels are front /æ/ and back /a/, rather than unrounded /a/ and rounded /ɔ/ as in Cuitlateco.'

8e is given by Sedlak for Cayuvava.

8f is given by Hockett for Polish dialects. The system shown contrasts front with back vowels, but other types are said to occur, substituting *a* and *ɒ* for *æ* and *a*, producing two series contrasting in terms of rounding. Trubetzkoy also gives 8f for Polish dialects (1958:100). Sedlak lists a configuration similar to 8f for Ibo; but he doubts that this is correct, since other authors (Stewart 1967, Chomsky & Halle 1968) have claimed that a dimension of advanced tongue root or covered–non-covered may be involved.

8g is given by Hockett for Portuguese dialects.

The predicted system has features also found in the data. For instance, four 'close' vowels are observed in 8a. There are systems which exhibit a contrast between [ɛ]- and [ɔ]-like vowels and lack other 'mid' vowels such as [ö], e.g. 8e and 8f. The contrast between the predicted 'low' vowels also occurs in the examples given; cf. 8b, 8d, and 8f. However, there is no system which completely agrees with the predicted pattern. In spite of discrepancies ([æ] instead of [ö]), 8a seems to come closest to the predicted configuration.

(9) NINE-VOWEL SYSTEMS:

Predicted:

i	ü	ʉ	ɨ	u
	ɛ		ɔ	
		æ	ɑ	

Observed:

(a)

i	ɨ	u
e	ə	o
æ	a	ɔ

(b)

i	ü	u
e	ö	o
æ	a	ɔ

(c)

i	ï	u
e	ë	o
ɛ	a	ɔ

(dα)

i	ü	u
e	ö	o
ɛ	ɔ̈	ɔ

(dβ)

i	ü	u
e	ö	
ɛ	ɔ̈	ɔ
	a	

(dγ)

i	y	u
e	ø	o
æ		å
	ɑ	

(e)

i		ï	u
e	ö	ə	o
		a	

(f)

i	ü		u
e		ë	
ɛ	ɔ̈		ɔ
		a	

(gα)

i	ü		u
e	ö	ë	o
ɛ/æ		ȧ/ä	

(gβ)

i	ü	u
o	ö	o
ä	õ	a

(h)

i	ü	ï	u
e	ö	ë	o
	a		

(i)

i		y	ʉ	u
	e		ø	o
		ɛ	(œ)	
	(æ)			ɑ

9a is given by Hockett for Trukese, Thai, Temoayan and Mazahua Otomi, and English; and by Sedlak for Kannada (with ï instead of ɨ), Banda-Linda, English, and Karen (with *ɛ/æ* instead of *æ*).

9b is given by Sedlak for Scottish (Barra), with length contrasts and nasal subsystems; and also by Trubetzkoy (1958:111).

9c is given by Sedlak for Shoshone.

9dα is given by Hockett for Estonian (cf. 9g below).

9dβ is given by Sedlak for Icelandic.

9dγ is given by Trubetzkoy for Norwegian (1958:103).

9e is given by Sedlak for Korean.

9f is given by Sedlak for Khakass.

9gα is given by Sedlak for Bashkir, Estonian, and Tatar.

9gβ is given by Trubetzkoy for Estonian (1929:117). Note that all the quoted authors agree in attributing a nine-vowel system to Estonian; but the impreciseness of their phonetic specifications makes comparisons difficult. Apparently Hockett's [ɔ̈] corresponds to Trubetzkoy's [õ] and to Sedlak's [ë] (cf. Lehiste 1960).

9h is given by Sedlak for Akha.

9i occurs in Swedish, for the set of phonologically long vowels. Quality differences are observed between this set and the corresponding short ones. Moreover, pre-*r* allophones of /ɛ/ and /ø/ are found, viz. [æ] and an open variety of [œ].

We conclude that the predicted set comes closest to 9i. However, the lack of complete agreement between the facts and the model suggests that an [ø]-type vowel be substituted for the [ɨ] of the prediction, and that somewhat 'closer' varieties of the two non-high front vowels and the mid back vowel be obtained.

The consequences of treating the major allophones of the long Swedish vowels as an eleven-vowel system rather than as a set of nine can be seen below.

(10) TEN-VOWEL SYSTEMS:

Predicted:

i		ü	ʉ	ɨ	u
	e				
				o/ɔ	
	ɛ/æ				
		a	ɑ		

Observed: (a)

i		ü	ɨ	u
e		ö	ə	o
	æ		a	

(b)

i	ü	u
e	ö	o
ɛ	œ	ɔ
	a	

(c)

i		y		u
e		ø		o
ɛ				ɔ
	a		ɑ	

10a is given by Hockett for 'Koibal, Karagin, and possibly some dialects of Korean'; and by Trubetzkoy for Koibal and Karagin (with *y*, *ü*, *œ*, *ä* instead of Hockett's *ü*, *ɨ*, *ə*, *æ*).

10b is given by Hockett for the Lyster dialect of Norwegian; and by Sedlak for Marathi (with *ɨ*, *ə*, ʌ instead of *ü*, *ö*, and *œ*), Vietnamese (with *ï*, *ë*, ʌ instead of *ü*, *ö*, and *œ*), and Norwegian as spoken in Wisconsin (*ɔ̈* for *œ*).

10c is given by Trubetzkoy for French (1958:111).

The five 'close' vowels of the model are best matched by 10a with four 'close' vowels. Note that the predicted system shows four 'heights' in the 'front' series, but three in the 'back' series. This asymmetry does not appear in the data. 10b and 10c have only one vowel between [i] and [u]. 10c agrees with the model with respect to the front unrounded and low vowels, but deviates in having [ø] and an additional back vowel instead of the [ʉ] and [ɨ] of the model.

(11) ELEVEN-VOWEL SYSTEMS:

Predicted:

i		ü	ʉ		ɨ	u
	e					
		ə				
					ɔ	
		ɛ/æ				
		a		ɑ		

Observed: (a)

i	ü	u
e	ö	o
ɛ	ɔ̈	ɔ
	ȧ	ä

(b)

i	ɨ	u
e	ᵻ	o
ɛ	ə	ɔ
æ	a	

(c)

```
i   y   ɯ  u
e   ɵ   ɜ  o
  æ       å
        ɑ
```

11a is given by Sedlak for French (maximal system).

11b is given by Sedlak for Nicobarese (with length and nasalization contrasts).

11c is given by Trubetzkoy for 'das Vokalsystem des Ostjak-Samojedischen (heute "Sölkupischen" ...)' (1958:95).

(12) Twelve-vowel systems:

Predicted:

```
i   ü   ʉ  ɨ  u
  e
        ə  ɔ
    ɛ
      æ a ɑ
```

Observed: (a)

```
i   ü      ɨ  u
e     é ȯ    o
ɛ   ɔ̈      a  ɔ
```

(b)

```
i    ü  ï   u
e    ö  ë   o
ɛ     a ä   ɔ
```

12a is given by Sedlak for Tibetan.

12b is given by Sedlak for Akha.

Here 12b shows some resemblance to the model.

1.4. Summary of results. The empirical systems that approach the calculated results most closely are 3a, 4c, 5a, 6b, 7b, 8a, 8b, 9i, 10a, 11c, 12b. The extent to which these 'best matches' agree or differ from the predicted patterns is demonstrated in Figure 4. This figure is identical with Fig. 2 except for the additional data points (open circles), which were obtained by translating the symbols of the best matches into the formant frequency patterns listed in Table 1. The translation of symbols not included in this table was made as follows: *y* = *ü*; *ï* and *ɯ* = *ɨ*; for Trubetzkoy, *ɑ* has been interpreted as *a*, *å* as *ɑ*, *ø* and *ɵ* as *ö*, *ë and ɜ* as *ə*. The symbol *ö* is represented as $F_1 = 450$, $F_2 = 1500$. Furthermore, *e* and *o* have been assigned values corresponding to *ɛ* and *ɔ* in the cases for which it is clear that only three 'heights' are involved. Sedlak's use of the symbol *a* in 11a indicates that it stands for an [ɑ]-like quality; but an alternative interpretation as [a] may also be possible, in which case Sedlak's *a* will have to be taken to represent a front variety and his *ɔ* a more [ɑ]-like quality. In Fig. 4, parentheses around circles imply that these data points may be interchanged to give alternative best matches. Thus Turkish 8a or Finnish 8b may be selected in the case of eight-vowel systems. The parentheses in the twelve-vowel plot in Fig. 4 were introduced to reflect the alternative interpretations possible.

Summarizing the observations that can be made from the comparisons in Fig. 4, we note that approximately correct results are obtained in the case of three-, four-, five- and six-vowel systems. For the larger systems various errors occur, but in general not more than one per system:

Three-vowel system: No major discrepancies.

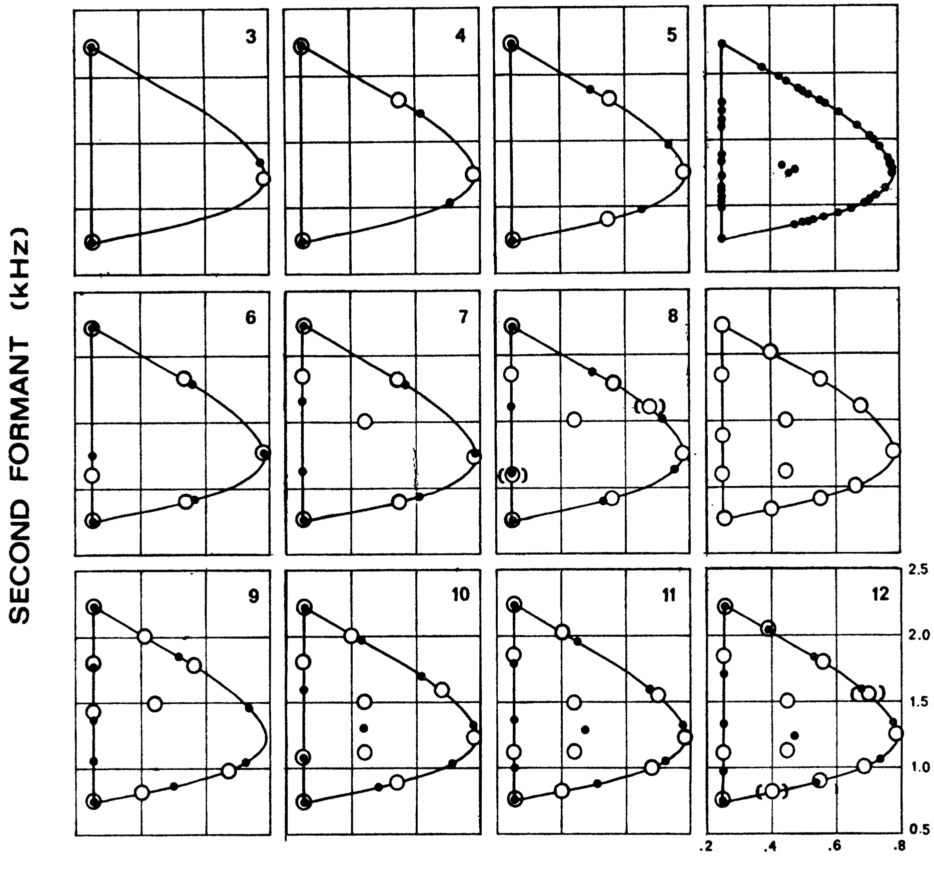

FIGURE 4

FOUR-VOWEL SYSTEM: The predicted 'low' vowel is [ɑ]-like rather than [a]-like.

FIVE-VOWEL SYSTEM: 'Open' variety of [ɔ] and 'front' variety of [a] are predicted.

SIX-VOWEL SYSTEM: Fairly close agreement.

SEVEN-VOWEL SYSTEM: The predicted system has [ɨ] instead of [ø].

EIGHT-VOWEL SYSTEM: The predicted system has [ɨ] instead of [ø] (Finnish), or [æ] instead of [ø] (Turkish).

NINE-VOWEL SYSTEM: The predicted system has [ɨ] instead of [ø] and [æ ɛ ɔ] instead of [ɛ e o]. Note, however, that *ɛ → æ*, *e → ɛ*, *o → ɔ* are among the phonological rules of Swedish, whose system is used for comparison in this case.

TEN-VOWEL SYSTEM: The predicted system has [ɑ] instead of [ø].

ELEVEN-VOWEL SYSTEM: The predicted system has [ʉ] instead of [ø].

TWELVE-VOWEL SYSTEM: The predicted system has [ʉ] instead of [ø].

It is clear that the major deficiencies of the model are its inability to generate an [ø]-like vowel and its predilection for 'close, central' vowels such as [ɨ] and [ʉ]. By and large, however, we conclude that, although based on simplifying assumptions about the universal space for vowels, its perceptual representation and the quantification of contrast, our predictions are not too far from the facts. The model produces about nine clear errors in a comparison involving 75 vowel qualities. As a preliminary conclusion, it seems justified to infer that CONTRAST plays an important role as a determinant of the structure of vowel systems. It is likely that it should be included among the variables in an explanatory phonological theory.

2. CONTRAST AND SOME OTHER PHONETIC BOUNDARY CONDITIONS. Below we shall analyse the results obtained. We shall discuss the quality of the data against which the predictions have been evaluated, and consider various refinements that might be introduced into the model in future research. Such improvements have to do with practical and computational as well as theoretical and conceptual aspects. On the one hand, it is necessary to comment on the prediction of a single rather than several systems, and on the approximations chosen to represent the notions of contrast and vowel space. Here factors such as predictability and perception enter. On the other hand, the integration of articulatory variables, e.g. 'ease' of articulation, 'co-articulability' etc., and 'syntagmatic' aspects beside 'paradigmatic' ones, must be considered.

2.1. THE DATA. The predictions are formulated in terms of numerical data on formant frequencies. The raw data on the structure of vowel systems are available in the form of broad phonemic transcriptions. To evaluate the predictions, we have defined a procedure for translating between these formats. To some linguists this may appear perfectly justified, since, after all, interest ought to be focused primarily on 'structural relations' rather than 'phonetic substance'. Others may raise objections, since 'low-level' phonetic quality differences, e.g. those between German [ü] and Swedish [y], Japanese [u] and Twi [u] etc., must necessarily be regarded as facts which also require an explanation. Ideally one would like to satisfy the adherents of both views. Making the comparisons in terms of formant frequencies is complicated by the fact that reliable data are available only for a small number of languages. Furthermore, the data that have been published cannot be used until they have been carefully normalized with respect to non-linguistic, speaker-dependent factors such as vocal tract size. However, the theoretical basis of such normalization remains to be developed.[2]

2.2. THE QUESTION OF UNIQUENESS. The computer program was written so

[2] The normalization of formant frequency data has been investigated by Fant 1966, who found that inter-sex differences in formant frequencies for Swedish and American English could be accounted for in terms of typical differences in vocal tract anatomy. So far it has not been shown whether inter-sex variations in vocal tract size are qualitatively similar to those that would be observed, say, in groups of male speakers. Since, for technical reasons, acoustic phonetic data on vowel systems usually pertain to groups of male speakers, normalization procedures must be based on intra-sex rather than inter-sex observations.

as to produce a single system, namely that for which the criterion of maximal contrast is fulfilled. However, when the algorithm has found such a system, there is no guarantee that this solution represents the only possibility, nor in fact that for which the degree of contrast is maximized in an absolute sense. It merely represents that system which exhibits greater over-all contrast than the other systems examined during the computations. The systems that the program selects for testing are thus highly dependent on the manner in which the search procedure is set up. Thus there may be several sets of vowels that meet the criterion as well as the ones listed above in §1.4. Finding all such sets has been considered relevant only when preliminary calculations produced positive results. Since the present findings are indeed encouraging, the question of uniqueness is an obvious topic for future work on the model. Thus the reason for predicting a single system, rather than several, has to do with a desire to keep computations within practical limits. It is not a conceptual defect of the theory.

2.3. THE INTERPRETATION OF CONTRAST. The appeal of the principle of maximal contrast is no doubt based on the belief that vowels can serve as more efficient carriers of differences in meaning as they become more dissimilar, and the risk of confusing them decreases. In the preceding presentation we have equated 'vowels' with 'vowel phonemes', not with underlying phonological segments, or with phonetic segments in the sense of generative phonology. We are aware that the approach selected is a simplification. It does not appear unrealistic to suppose that, in real speech, any allophone of a vowel phoneme, whether it functions as the major representative or as a combinatory variant, runs the risk of being assigned to an incorrect category by a listener. Consequently, it might be proposed that, to improve the predictions, we consider phonetic rather than phonemic segments as the relevant empirical data. This will offer no improvement, however, since in most languages the context-dependent selection of allophones and various distributional constraints, e.g. vowel harmony, tend to make the predictability of the vowels vary non-uniformly. Since the extent to which it would be motivated to differentiate the perceptual quality of a given segment ought to be related to the predictability of that segment, it is clear that the assignment of MAXIMAL contrast to all allophones of a system is a-priori wrong. Nevertheless it appears that a more satisfactory theory would make its predictions at the phonetic level of allophonic segments on the basis of both contrast and contextual restrictions of occurrence.

2.4. THE REPRESENTATION OF THE VOWEL SPACE. It is unlikely that the LS model defines an acoustic vowel space that is greatly in error, since the formant patterns which it can generate are in close agreement with those of basic vowel qualities. Another source of error which must also be regarded as minor is the stylization of the space introduced in the computations (Fig. 1). The fact that a two-dimensional rather than a three-dimensional representation was used should also be borne in mind. More important is the present use of the mel scale. A mel scale transformation is commonly carried out on acoustic phonetic data as a first step toward an auditory representation. However, much more research is needed to establish the true nature of auditory patterns and the perceptual space for speech sounds.

2.5. ARTICULATORY FACTORS. The preceding remarks draw attention to a tacit assumption underlying the present calculations. We have assumed that the vowel space is a 'homogeneous' one. Metaphorically, the vowel magnets float around, as it were, in a homogeneous medium. As a consequence, a given distance *r* represents the same contribution to the 'magnetic field' wherever we observe it within the space. We have argued above that a given perceptual distance does not necessarily represent equal 'communicative efficiency' wherever we observe it, owing to the constraints that phonological structure imposes on segment distribution. It seems reasonable to suppose that a vowel system which has been optimized with respect to communicative efficiency consists of vowels that are not only 'easy to hear' but also 'easy to say'. Consequently a further improvement of the present theory might be obtained if we found a way of quantifying and incorporating 'ease of articulation'. The LS model offers some possibilities in that the notion of NEUTRAL SHAPE of lips and tongue can be interpreted in a numerical and physiologically realistic manner; for some attempts to explore this feature of the model, see Lindblom & Sundberg 1971, where a quantitative treatment of the principle of 'least effort', or articulatory synergy, is used to explain why [i], [u] are likely to be pronounced as 'close' vowels whereas [a], [ɑ] must be 'open' vowels in all languages. The acoustic space will shrink considerably if neutral tongue shape is assigned to all vowels, especially with regard to the second formant. As a result vowels will differ chiefly in terms of the first formant and degree of opening. Although equilibrium positions of 'vowel magnets' introduced into this reduced space have not been calculated, it is clear that a three-vowel system would very closely resemble the so-called 'linear vowel systems' with a high central vowel, a mid central vowel, and a low [a]-like vowel. We conclude that the neglect of physiological aspects of vowel production may be another source of discrepancies between facts and predictions.

A hypothesis proposed to throw light on the processes that determine how languages select their inventories of speech sounds is the 'plateau' theory of Stevens. In a series of articles (1968, 1969, MS), he draws attention to the fact that the relation between articulatory and acoustic parameters is highly non-linear and non-monotonic. These circumstances can give rise to 'plateaus', i.e. regions with well-defined acoustic attributes. These plateaus are stable and insensitive to articulatory errors and imprecision. Stevens (MS) argues that language 'seeks out these regions, as it were, and from them assembles an inventory of phonetic elements that are used to form the code for communication by language'. This hypothesis is an example of how an articulatory condition may be involved in the delimitation of sound shapes suitable for linguistic communication. Further research is needed to determine the relative roles that should be assigned to criteria such as ACOUSTIC STABILITY and PERCEPTUAL CONTRAST in phonological theory. Whatever the relative role of 'stability' criteria, Stevens' work is a pioneering attempt to provide a quantitative phonetic basis for the prediction of phonological facts. As such it exemplifies a novel, substance-oriented approach to the study of linguistic form (cf. the discussion below).

2.6. Syntagmatic aspects. Implicit in our discussion so far has been the assumption that an improved version of the present model would result if we could succeed in quantifying the notions of 'easy to say' and 'easy to hear' in all their various forms. The principles of perceptual differentiation and reduction of articulatory energy expenditure should be assumed to operate 'syntagmatically' as well as 'paradigmatically'. It might, for instance, be proposed that, as a consequence of syntagmatic conditions on the optimization of syllable structure, consonant and vowel segments will be favored that maximize perceptual distances not only within each such class of segments separately, but also within the total set. In other words, consonants and vowels are chosen so that the extent of formant transitions in arbitrary CV combinations is maximized. Adherents of such a view might argue that large spectral changes tend to facilitate perceptual processing and recognition more efficiently than small changes. Preliminary computations have been undertaken within the present framework to study the effect of various consonant systems on the shape of the vowel system. A full account of these calculations falls outside the scope of the present article; nevertheless, it is of some interest to note that the effect of a given consonant system is found to be more pronounced for small vowel systems, and manifests itself by displacing vowels toward [+low] regions in the vowel space, i.e. regions with high F_1 values, as in [a ɑ æ]. These observations bring to mind the phonological systems of Caucasian languages such as Kabardian, whose underlying vowel segments are confined to /ə/ and /ɑ/ but whose consonant system is extremely rich (Halle 1970).

In summary, the purpose of the preceding remarks has been to make clear that boundary conditions other than perceptual differentiation are likely to play a role in phonology. Although nothing has been said about the diachronic and sociolinguistic origins of phonological structure (the treatment of these aspects falls outside the scope of the present limited study), we believe that, to be fully explanatory, a theory of vowel systems, or of phonological systems in general, cannot avoid reference also to the circumstances under which such systems arise.

3. The relevance of phonetics to the study of language.[3] In phonology the efforts of many linguists are directed toward the construction of a theory that predicts and explains various aspects of phonological structure. Ideally, such a theory should explain why certain systems, rules, segment classes, and phonotactic patterns are more probable than others, and give an account of the historical changes that such aspects of language structure actually did and might undergo. The formulation of these goals appears to have become particularly explicit with the advent of generative phonology (Halle 1962, 1964; Chomsky & Halle 1965, 1968). Recently a great deal of research has been undertaken in the spirit of this program. Generally it has been based on a distinctive-feature framework and the notational devices of generative phonology. In the final chapter of *The sound pattern of English*, Chomsky & Halle attempt to weigh the success that the theory of generative phonology has had in attaining the de-

[3] For a fuller discussion of this topic and some preliminary applications of the research program sketched in this section, see Lindblom, ms.

sired goals. Their appraisal is (1968:400) that 'The entire discussion of phonology in this book suffers from a fundamental theoretical inadequacy ... The problem is that our approach to features, to rules, and to evaluation has been overly formal ... In particular, we have not made use of the fact that the features have intrinsic content ...' To remedy the situation, they propose a theory of marking that purports to refer to intrinsicality in a more satisfactory manner.

Several problems have emerged in the course of testing the generative framework against various data. Let us briefly review some of them (cf. Ladefoged 1967, 1971a; Fromkin 1968; Wang 1968b). Difficulties arise, for instance, in connection with the marking conventions, the choice of the correct set of features, the non-orthogonality of features, the use of variables, and the choice between binary or *n*-ary dimensions. Thus marking conventions have not yet been sufficiently developed to predict and explain the structure of vowel and consonant systems (Chomsky & Halle 1968:420–1). It can be objected that, in spite of their alleged function, marking conventions fail to reflect the 'intrinsic content' of features to a sufficient degree, and appear to mirror known facts in an observational rather than explanatory manner. Furthermore, the correct set of features remains to be established. For example, there seem to be differing opinions (cf. Wang 1968b) concerning the representation of tongue height or opening, which would seem from a phonetic point of view a highly non-controversial dimension in comparison with features such as covered, advanced tongue root, tense–lax etc. Are there three or four contrastive degrees of opening? (Cf. Kiparsky 1968, Wang 1968b.) Fant 1971 illustrates the difficulties involved in making a unique choice among alternative distinctive-feature frameworks for vowels that are compatible both with phonetic and phonological facts. Furthermore, arguments have been presented in favor of multi-valued or *n*-ary rather than binary features (Ladefoged 1967, 1971b; Bruce 1970).

Again, variable notation that has been used to define natural classes appears to be too rich, and needs to be constrained (Zwicky 1970). As Chomsky & Halle (1968:400) point out, the simplicity criterion (symbol counting) assigns the same complexity to [αback, αround] as to [αlow, αround], in spite of the greater 'naturalness' of the former class of segments. The authors further state (401) that 'it would be a mistake to try to eliminate the less natural classes by a sharper definition of "naturalness" that makes use only of formal properties of features and feature specifications, for it is actually the content of the features and not the form of the definition that decides these questions of naturalness'. A similar problem arises in a system using n binary features which would in principle be capable of handling 2^n segments. Since feature dimensions are interdependent—e.g., no vowels can be [+lateral] or [+obstruent], for obvious phonetic reasons—such systems must be supplemented with conventions that reduce the capacity of the system in a rather Procrustean fashion. For example, [+high] vowels must not also be [+low]. In the case of vowels, Wang 1968b suggests the use of the features [palatal] and [velar]. This represents an improvement on the Chomsky–Halle system in which [covered] is tentatively introduced as the distinctive feature of the Swedish and Norwegian /ʉ/ vowels. In Wang's framework /ʉ/ is [−palatal, −velar]. Note, however, that it becomes necessary to rule out [+palatal, +velar] by a special statement, as in the case of the features of tongue height. Can phonetic theory be applied to resolve some of these difficulties?

Phonetic research on speech has traditionally departed from the models of language structure postulated by the linguist. Evidence for this can be found in the numerous studies devoted to finding the phonetic correlates of phonemes, syllables, distinctive features, tones, accents etc., and to studying the physical mechanisms underlying these and various other linguistic units. During the past decades many important contributions have been made that have significantly advanced our understanding of the speech signal and the manner in which it 'realizes' the underlying linguistic organization of utterances. Further progress will no doubt continue to be made. On the other hand, we must concede that very little research has been consciously undertaken in the fields of phonetics and speech communication to elucidate how language rather than speech is structured. And it appears legitimate to question whether phonetics, pursued principally as a science of 'realizations', could ever offer linguists help in reaching the goals set up in explanatory phonology, i.e. in accounting for systems, rules, natural classes, features, syllable structures, historical changes etc. According to this picture, the role of the phonetician has been that of passive 'consumer' of linguistic 'products' such as the phoneme, distinctive features etc. As long as phonetics is used in linguistics chiefly to interpret linguistic form in terms of phonetic substance and for assigning physical shape to the output of grammars, there is every reason to give a pessimistic answer to the question raised at the end of the previous paragraph.

Suppose, however, that we let the horse and cart change places. Rather than accept it as a-priori, we attempt to derive linguistic form as a consequence of various substance-based principles pertaining to the use of spoken language and its biological, sociological, and communicative aspects. We would then proceed by asking the following questions:

(a) What are the mechanisms available for human speech communication?

(b) How can the use of these mechanisms be constrained and optimized with respect to various psychological conditions and communicative efficiency?

(c) Given a system of hypotheses about the speech mechanisms and constraints on its use, how should we expect these mechanisms to be used in communication by language; or equivalently, what predictions about the organization of linguistic structure can be derived?

A common question these days is: What is the psychological reality of a given phonological construct or process? The above approach is exemplified by changing the question into: What constructs and processes would we derive as consequences of the speech mechanisms and the general conditions on speech communication? This may strike the reader as involving insurmountable difficulties, but is in our opinion a necessary, complementary formulation. Answering (a) consists in the construction of quantitative models that reflect how the central nervous system, motor and sensory systems delimit the class of universal speech signals that can be generated by the speaker and processed in the listener. Here we recognize the search for the 'phonetic capabilities of man' (Chomsky

& Halle 1968:299), and the development of a numerical phonetic theory of speech production and speech perception.

Under (b) belong hypotheses about perceptual differentiation and physiological energy minimization ('maximal perceptual effect produced with minimal articulatory means'). But we can also foresee that maturational constraints involved in language acquisition and speech development, and imposed not only on production and perception but also on learning, memory, serial ordering, and cognitive mechanisms, will play a role in the selection of speech signal 'realizations' that the physical mechanisms studied under (a) make available. Criteria anchored in production, perception, learning, memory etc. and their development will thus determine a hierarchy among possible uses of these mechanisms, thereby introducing discreteness and a scale of naturalness. Will highly ranked predictions bear a resemblance to the facts of language? Will this model, placed in the context of language acquisition and sociolinguistic variables (Weinreich et al. 1968), permit the identification of sources of instability underlying the mechanisms of language change? Such a research program might lead to a more productive application of phonetic theory to phonology. Possibly, formulated in general terms, the above remarks may in principle be valid for other aspects of language as well.

Clearly the present proposal for a substance-based theory represents a break with the Saussurean tradition in the study of linguistic form. This break is a necessary consequence, however, of the requirement that theories of language be predictive and explanatory. It is foreshadowed in the critical remarks about the lack of 'intrinsicalness' in the definition of features (Chomsky & Halle 1968, Fromkin 1968), in the 'overly formal' approaches so far explored (Chomsky & Halle 1968), and in the requests for phonetic explanation and realism in accounts of phonological structure and historical change (Wang 1970, Vennemann 1972). It is in line with the criticisms and objections that many phoneticians have raised in the course of the recent development of phonology (Ladefoged 1971a, Fant 1969). The proposed approach is consequently at variance with the following recommendation by Chomsky (1968:12), which seems to represent a common conception of the competence–performance distinction, and clearly reveals the Saussurean heritage:

> It seems to me that the most hopeful approach today is to describe the phenomena of language and of mental activity as accurately as possible, to try to develop an abstract theoretical apparatus that will as far as possible account for these phenomena and reveal the principles of their organization and functioning, without attempting, for the present, to relate the postulated mental structures and processes to any physiological mechanisms or to interpret mental function in terms of 'physical cause'.

In this quotation we find an example of the view discussed above, viz. that physics and physiology are to be used in an interpretative fashion, rather than as part of the predictive theoretical machinery. Although Chomsky's remarks may be partly appropriate and realistic for practical, experimental purposes, theoretical considerations force us to re-word the last part of the sentence as follows: '... and functioning, and to attempt, as far as this is possible at present, to relate the postulated mental structures and processes to physiological mechanisms and physical causes.' This wording is an important emendation, since (although difficult to live up to) it is a safeguard against artefacts. On the other hand, the knowledge sought under (a) and (b) could be identified with some of the properties that the generative school would no doubt accept unconditionally as prerequisites for language acquisition. The phonetic mechanisms, as well as the directions for optimized use, must clearly 'be assumed to be available to the child learning a language as an a-priori, innate endowment' (Chomsky & Halle 1968:4).

REFERENCES

Bruce, G. 1970. Diphthongization in the Malmö dialect. Working Papers, Phonetics Laboratory, Lund University, 3.1–19.

Chomsky, N. 1968. Language and mind. New York: Harcourt, Brace & World.

——, and M. Halle. 1965. Some controversial questions in phonological theory. Journal of Linguistics 1.97–138.

——, ——. 1968. The sound pattern of English. New York: Harper & Row.

Fant, G. 1966. A note on vocal tract size factors and non-uniform F-pattern scalings. STL-QPSR (KTH, Stockholm) 1/1966.22–30.

——. 1969. Distinctive features and phonetic dimensions. STL-QPSR (KTH, Stockholm) 2–3/1969.1–18. (Now published in Applications of Linguistics, ed. by G. E. Perren and J. L. M. Trim, Cambridge University Press, 1971.)

——. 1971. Notes on the Swedish vowel system. Form & substance, phonetic and linguistic papers presented to Eli Fischer-Jørgensen, ed. by L. L. Hammerich et al., 259–68. Copenhagen: Akademisk Forlag.

Fromkin, V. 1968. The what and why of intrinsicalness. Working Papers in Phonetics, UCLA, 10.156–68.

Groot, A. W. de 1931. Phonologie und Phonetik als Funktionswissenschaften. (TCLP 4.) Prague.

Halle, M. 1962. Phonology in a generative grammar. Word 18.54–72.

——. 1964. On the bases of phonology. The structure of language, ed. by J. Katz and J. Fodor, 324–33. Englewood Cliffs: Prentice-Hall.

——. 1970. Is Kabardian a vowel-less language? Foundations of Language 6.95–103.

Hockett, C. F. 1955. A manual of phonology. (IJAL, Memoir 11.) Baltimore.

Jakobson, R. 1941. Kindersprache, Aphasie und allgemeine Lautgesetze. Uppsala. [Reprinted, Selected writings I, 328–401. The Hague: Mouton, 1962.]

Kiparsky, P. 1968. Linguistic universals and linguistic change. Universals in linguistic theory, ed. by E. Bach and R. Harms, 170–202. New York: Holt, Rinehart & Winston.

Ladefoged, P. 1967. Linguistic phonetics. Working Papers in Phonetics, UCLA, 6.

——. 1971a. The limits of phonology. Form & substance, phonetic and linguistic papers presented to Eli Fischer-Jørgensen, ed. by L. L. Hammerich et al., 47–56. Copenhagen: Akademisk Forlag.

——. 1971b. An alternative set of vowel shift rules. Working Papers in Phonetics, UCLA, 17.25–8.

Lehiste, I. 1960. An acoustic-phonetic study of internal open juncture. Phonetica 5 (Suppl.), 1–54.

Lindblom, B. E. F. ms. Phonetics and the description of language. Proc. 7th Intl. Cong. of Phonetic Sciences, Montreal. To be published by Mouton.

——, and J. Sundberg. 1969. A quantitative model of vowel production and the distinctive features of Swedish vowels. STL-QPSR (KTH, Stockholm) 1/1969.14–30.

——, ——. 1971. Acoustical consequences of lip, tongue, jaw and larynx movement. JASA 50.1166–79.

Martinet, A. 1955. Économie des changements phonétiques. Berne: Francke.

Moulton, W. G. 1962. Dialect geography and the concept of phonological space. Word 18.23–32.

PASSY, P. 1890. Études sur les changements phonétiques et leurs caractères généraux. Paris.

SEDLAK, P. 1969. Typological considerations of vowel quality systems. Working Papers on Language Universals, Stanford University, 1.1–40.

STEVENS, K. N. 1968. Acoustic correlates of place of articulation for stop and fricative consonants. Quarterly Progress Report, MIT Research Laboratory of Electronics, 89.199–205.

——. 1969. Pharyngeal consonants. Quarterly Progress Report, MIT Research Laboratory of Electronics, 93.207–19.

——. MS. The quantal nature of speech: evidence from articulatory-acoustic data. To appear in Human communication: a unified view, ed. by E. E. David Jr. and P. B. Denes.

STEWART, J. M. 1967. Tongue root position in Akan vowel harmony. Phonetica 16.185–204.

TRUBETZKOY, N. S. 1929. Zur allgemeinen Theorie der phonologischen Vokalsysteme. TCLP 7.39–67.

——. 1958. Grundzüge der Phonologie. Göttingen: Vandenhoeck & Ruprecht.

VENNEMANN, T. 1972. Phonetic detail in assimilation: problems in Germanic phonology. Lg. 49.863–92.

WANG, W. S-Y. 1968a. The basis of speech. Project on Linguistic Analysis Reports (2d series) 4, Berkeley.

——. 1968b. Vowel features, paired variables, and the English vowel shift. Lg. 44.695–708.

——. 1970. Project DOC: its methodological basis. J. Am. Orient. Soc. 90.57–66.

WEINREICH, U.; W. LABOV; and M. I. HERZOG. 1968. Empirical foundations for a theory of language change. Directions for historical linguistics, ed. by W. P. Lehmann and Y. Malkiel, 95–188. Austin: University of Texas Press.

ZWICKY, A. M. 1970. Greek-letter variables and the Sanskrit *ruki*-class. Linguistic Inquiry 1.549–54.

[Received 11 June 1971]