UNIVERSALS OF VOWEL SYSTEMS: THE CASE OF CENTRALIZED VOWELS

Jean-Marie Hombert, Linguistics,
University of California, Santa Barbara, USA 93106

This paper attempts to explain why centralized vowels (i.e. vowels which are not located on the periphery of the vowel space) are relatively less common than peripheral vowels.

1. Surveys of phonemic systems, phonetic universals and "exotic" languages.

If one is interested in discovering phonetic universals some of the most fruitful places to search for potential universals are large scale surveys of phonetic and phonemic inventories. Despite the criticism leveled against these surveys it is our belief that such surveys are useful in that asymmetries or systematic gaps in these inventories may reveal in their explanation universal phonetic processes. Once such a potential universal or universal tendency has been uncovered each language exhibiting this process should be reexamined through careful study of available sources, consideration of possible reinterpretations of the data, and when possible, accurate phonetic data should be obtained.

Until very recently the bulk of available phonetic data, especially perceptual data, has come from a handful of languages. Due to the availability of phonetic equipment and presence of research groups located in the countries where these languages are spoken available phonetic data has been largely limited to Danish, Dutch, English, French, German, Japanese and Swedish. It is clear that if we are to understand universal phonetic processes, our data base must be extended to include more "exotic" languages.

Most perceptual data has been gathered from experiments conducted under laboratory conditions using linguistically sophisticated subjects. Obviously if we are to gather similar data from languages spoken in areas remote from laboratory facilities, it is necessary to design techniques of data gathering suitable for use in the field with linguistically naive subjects. In Section 3 one such design will be discussed.

2. The case of centralized vowels.

It is clear from surveys of vowel systems that centralized vowels are less commonly found than peripheral ones. In the case of languages which do have centralized vowels it is not rare that different sources will vary in the treatment of such vowels by

attributing to a given vowel different phonetic qualities. These variations suggest that either these vowels are more prone to historical change or are more difficult to identify correctly by the investigator. It appears, then, from these surveys that non-peripheral vowels, that is, vowels which in acoustic terms have a second formant of approximately 1200-1700 Hz, are rare and that they are more subject to change than peripheral vowels.

In Section 3 we will use data from a perceptual experiment carried out on the Grassfield Bantu languages of Cameroon. Because of space constraints in this paper, we will use only data from one speaker of the Fe?fe? language[1] to suggest possible explanations for the rarity as well as instability of non-peripheral vowels.

3. Experimental paradigm

Fe?fe? contains eight long vowels in open syllables. These vowels are [i, e, a, ɒ, ɔ, u, ʉ, ə]. A word list consisting of eight meaningful Fe?fe? words contrasting these eight vowels was elicited from native Fe?fe? speakers. The Fe?fe? speakers were asked to read these eight words which were listed five times each, in random order. After the repetition of each word, the final sound of the word, that is the vowel, was repeated once. Both the vowels of the meaningful words and the vowels in isolation were subsequently analyzed.

Subjects were then asked to listen to 53 synthetic vowel stimuli, each presented five times in random order. After the presentation of each stimulus the subjects were instructed to point out which Fe?fe? word in the eight-word list that they had previously read, contained the same "final sound", i.e. vowel, as the stimulus. Subjects had the option to claim that some of the stimuli did not sound like any of the eight Fe?fe? words. The 53 synthetic stimuli were selected to maximally cover the vowel space; F1 was varied between 250 Hz-750 Hz, F2 between 650 Hz-2350 Hz and F3 between 2300 Hz-3100 Hz. This task was designed so that native speakers would divide the vowel space according to their own vowel systems.[2]

4. Results

The results of the acoustic analysis and of the perceptual experiment for one Fe?fe? speaker are presented in Figure 1 and Figure 2 respectively. Since F3 values are not relevant for the point that we want to make here the data are presented in an F1 x F2 space. Each vowel indicated in Figure 1 is the average of five measurements. The spectra were computed 100 msec. after vowel onset using LPC analysis. The phonetic symbols appearing in Figure 2 indicate that at least four times out of five this stimulus was identified by the Fe?fe? speaker as the same vowel.

We will consider the two vowels [a] and [ə]. Two unexpected results emerge from the data:

1. When comparing acoustic and perceptual data it is not surprising to find that the stimulus with F1 at 750 Hz and F2 at 1250 Hz is identified as the vowel [a] since a vowel with such a formant structure could have been produced by a Fe?fe? speaker with a larger vocal tract size than the speaker considered here. What is surprising, though, is that the stimulus with the formant structure F1 at 750 Hz and F2 at 850 Hz was also identified as [a]. These results are even more surprising when one considers that the intermediate stimulus (750 Hz - 1050 Hz) was identified as [ɒ]. It is likely that in the case of the stimulus with F1 at 750 Hz and F2 at 850 Hz the two formant peaks were perceived as one formant peak, that is as F1. One thing remains to be explained: in the acoustic data, the Fe?fe? vowel [a] has a peak around 1600 Hz but the stimuli with F1 at 750 Hz and F2 at 850 Hz does not have a peak in this frequency region. Let us just say for the moment that the saliency of the peak at 1600 Hz seems to be perceptually secondary.

2. Two stimuli (F1 at 350 Hz, F2 at 1500 Hz and F1 at 450 Hz, F2 at 1500 Hz) are identified as [ə], which is what we would expect considering the location of [ə] in Figure 1. However the identification of the stimulus with F1 at 450 Hz and F2 at 650 Hz with [ə] comes as a surprise. Notice that F1 and F2 are also close to each other for this last stimulus, which could have lead to the perception of them as one peak corresponding to the first formant. But notice also that this stimulus does not have a peak around 1500 Hz. As in the case of the vowel [a] it appears that the perceptual saliency of the peak around 1500 Hz did not play a major role in the identification of the [ə].

---

(1) For more data and a more complete description of the experimental paradigm, see Hombert (in preparation).

(2) It should be noticed that this method does not allow study of diphthongs since all stimuli used have steady state formant frequencies.

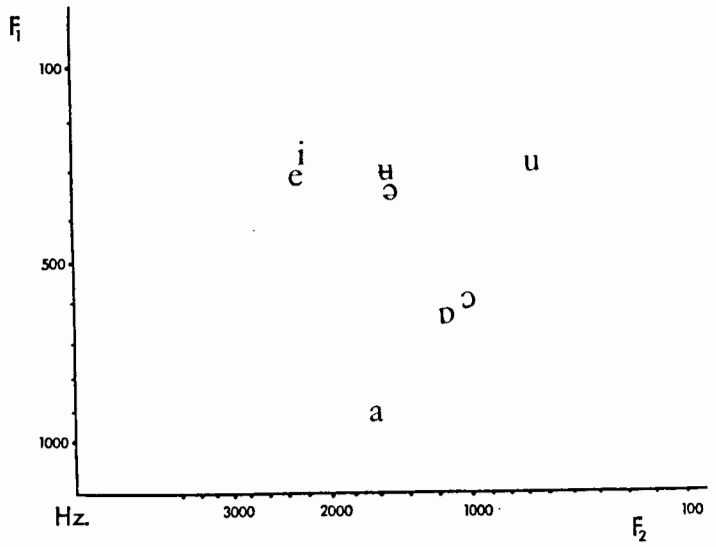

Figure 1. Acoustic data: the Fe?fe? vowel system, (one speaker, average of five measurements).

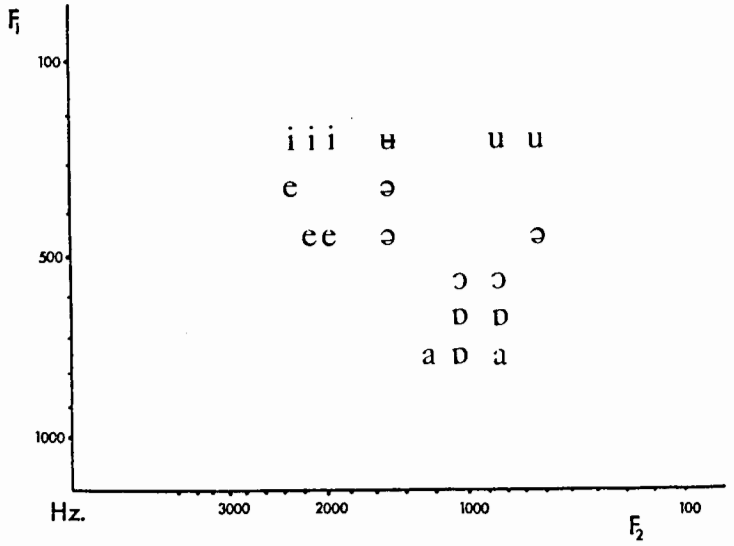

Figure 2. Perceptual data: only stimuli for which the Fe?fe? subject gave at least four out of five identical responses are presented on this graph.

5. Discussion[3]

Two possible explanations to account for the lack of saliency of formant peaks around 1500 Hz are being explored now.

1. Spectrum-based representation of vowels.

Our results would be compatible with a mechanism of vowel perception which looks for certain amounts of energy within frequency regions rather than formant peaks. In the cases which we discussed in the previous section, the unexpected vowel identification happened with stimuli which had their first and second formants very close to each other. In such cases the closeness of the first two peaks leads to an increase in amplitude of the spectrum. This increased amplitude may have created sufficient energy in the 1500 Hz region to lead to these "perceptual mistakes".

2. Place vs. periodicity mechanisms.

Pitch is processed by different mechanisms depending upon its frequency region. The boundary between these two mechanisms (place vs. periodicity) is not well defined. It is possible that for some subjects a defective overlap between these two mechanisms in the 1500 Hz region could create the perceptual mistakes presented in Section 4.

6. Implications

The explanation generally provided to account for the relative scarcity of non-peripheral vowels is based on the principle of maximum perceptual distance presented by Liljencrants and Lindblom (1972). Our results suggest a different explanation - non-peripheral vowels are avoided because one of their components (F2) is located in a relatively less salient perceptual zone. If this is the case we can now understand why processes leading to vowel centralization (vowel nasalization, rounding of front vowels, unrounding of back vowels) are relatively uncommon.

Finally we should point out that "perceptual mistakes" such as the ones reported in Section 4 were found in approximately one out of five subjects, with the "mistake" being consistently made by the one subject. These results would be consistent with a theory of sound change which claims that sound changes are initiated by a minority of speakers.

---

(3) The reason why previous experiments on vowel perception did not uncover this problem may be due to the nature of the experimental paradigm as well as the range of stimuli used in this experiment.

(vii) { a ⟶ ɔ/w __ ; ø ⟶ œ/ __ r } (was, swan, quarrel; Middle English). ([ø:va] vs [œ:ra]; Swedish).

(viii) { x ⟶ { ç / +front V __ ; x / elsewhere } ; /h/ realizations of Japanese cf. (v) above. } ([lɪçt] vs [axt]; German).

(ix) { n ⟶ m / b __ ; /n/ realizations of Swedish cf. (iii) above. } ([ha:bm] (haben); German).

(x) { r ⟶ r̥ / [-voic, -son] __ ; ʒ ⟶ ʒ̊ / __ [-voic] ; k ⟶ k̊ / __ [+voic] } ((try, cry, pry; English). ([nɛʒfɔ̃dy], [sak̊dɔʀ] French)

The above examples of pro- and regressive assimilations suggest that assimilation be hypothetically described as a reduction of articulatory distance in articulatory space. Do they imply a syntagmatic pronounceability condition, favoring a reduction of the physiological equivalent of a power constraint, mechanical work (force x distance)/time (a LESS EFFORT principle)? Can at least some phonological facts be interpreted as cases of contrast-preserving articulatory simplifications? What is their behavioral origin?

3. Speech - a Physiological Pianissimo.

3.1 The question also arises whether spoken language underexploits the degrees of freedom that in principle the anatomy and physiology of speech production make available. Seen against the full range of capabilities, speech gestures, like many other skilled movements, appear to be physiologically "streamlined" both as regards muscle recruitment and force levels (cf. jaw closure as a speech gesture and in mastication, speech breathing vs respiration in general, articulatory gestures vs swallowing etc.). Extreme displacements of articulatory organs do not occur (PIKE 1943, 150) although such configurations are available and yield acoustically equivalent results (evidence from non-speech: body-arm, eye-head coordination; and from speech: lip/tongue-mandible and tongue blade-tongue body coordination (LINDBLOM et al 1974)). Do we in these circumstances see the operation of an economy of effort principle? A principle that we should invoke to explain how and why speech and non-speech sounds differ and to account for certain phonological regularities as well as the instances of hypo-articulation (reductions, ellipses, co-articulations etc.) in spontaneous speech. "Todays's allophonic variation leads to tomorrow's sound change..." OHALA (1979).

3.2 Pronounceability and Syllable Structure.

FIG. 1 shows average measurements of jaw positions for Swedish apical consonants in the environment [a'Cɑ:]. The production of these consonants permits a variable influence of the open jaw positions of the vowels. Thus the dimension of jaw opening reveals one aspect of their "willingness" to coarticulate. It is of considerable interest to see that this measure correlates well with their universally favored position in initial and final phonotactic structures (ELERT 1970). If the present observations are generalized, they imply that the phonetic structure of clusters can be explained at least in part with reference to ease of co-articulation (ELERT 1970, BRODDA 1972).

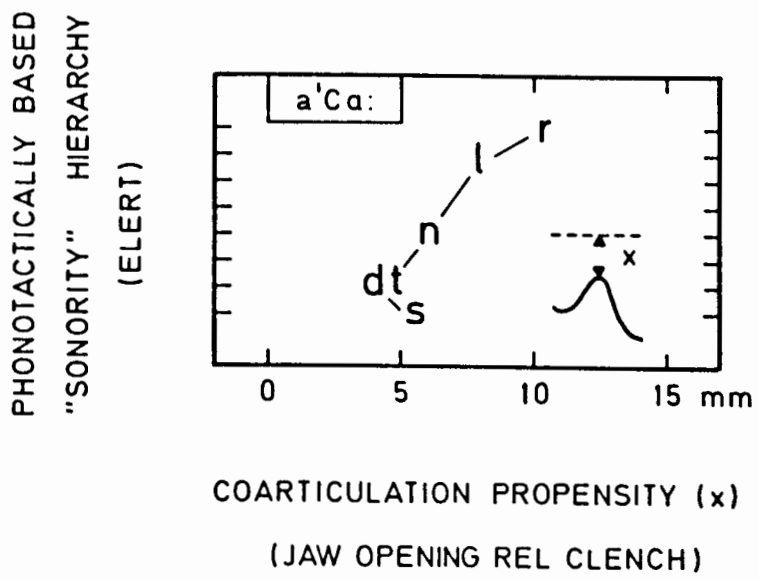

FIG. 1

4. The Distinctiveness "Conspiracy".

4.1 Language structure exhibits redundancy at all levels.

4.2 Speech generation is an output-oriented process: The reference input to the speech control system is specified in terms of a desired output. The dimensions of the target specifications are sensory, primarily auditory. Evidence supporting the primacy of auditory targeting comes from work on compensatory articulation, speech development and the psychological reality of phonological structure (LINDBLOM et al to appear, LINELL 1974).

4.3 Speech understanding is an active (top-down or conceptually driven) process. (Cf. the demonstrations of context-sensitive processing, resistance to signal degradation, phonemic restoration, verbal transformation etc.)

4.4 The speech system may possess specialized mechanisms that contribute towards enhancing the distinctiveness of stimulus cues. Examples of such hypothetical mechanisms are "feature detectors" in speech perception. Specialization of speech production has been suggested in the case of the phylogenetic development of the human supralaryngeal vocal tract whose shape LIEBERMAN (1973) interprets as a primarily speech-related adaptation increasing the acoustic space available for speech sounds.

4.5 Phonetic targets are selected so as to retain acoustic stability in the face of articulatory imprecision (STEVENS 1972).

The properties listed in 4.1 through 4.3, do they have a common origin in a basic principle of language design viz., the DISTINCTIVENESS CONDITION: different meanings sound different? The preservation of meaning across encoding and decoding seems to be favored by redundancy, output-oriented and active processing (rather than by lack of redundancy, exclusively input-oriented encoding and purely passive decoding strategies). Thus the question arises whether these at first seemingly unrelated attributes form an evolutionary "conspiracy". Do they constitute three different ways of coping with a physical signal which is inevitably going to be noisy, variable and ambiguous? 4.4 and 4.5 could offer related advantages. What is the behavioral origin of the distinctiveness condition?

5. Speech Development.

5.1 Imperfect learning: Can perceptual similarity and articulatory reinterpretation serve as a source of phonological innovation (cf. JONASSON (1971))? Many sound substitutions in children's speech appear compatible with this interpretation: θ→f, ɪ→w cf. 2.1. The child is a cognitive and phonetic bottle-neck through which language must pass. Does the process of acquisition leave its imprints on language structure?

5.2 Selection of the fittest: A speech community may use in free variation several realizations of a given form. The set of fricatives may contain /f, s, ʃ, ç/ and /h/ with the /ʃ/ produced as [ʃ] and [ʂ] (cf. Swedish). The distinctiveness principle favors [ʃ] which contrasts better with [ç] than [ʂ]. The lower confusion risk of the pair [ʃ] / [ç] promotes its reception and learning by the child. There is in this case thus a behavioral rather than teleological motivation for the distinctiveness condition. If sound patterns show evidence of perceptual differentiation, is communicative "selection of the fittest" among several competing forms one of the evolutionary mechanisms? Selection occasionally occurs from a rich variety of hypo- as well as hyper-articulated forms (STAMPE 1972). Is hyperarticulation another behavioral source of distinctiveness?

6. Non-Phonetic Origins of Sound Patterns: Social Biasing.

Selection of speech forms is influenced not only by production and perception factors. Phonological contrasts vary as a function of social variables (prestige, age, class, sex, style etc.). Does the interaction of the sometimes conflicting requirements of social and phonetic factors account for the fact that there is no evidence (GREENBERG 1959) that language change leads to more efficient linguistic systems? Is local rather than global phonetic evaluation of systems (KIPARSKY 1975) another reason why languages do not seem to be converging toward a single optimum equilibrium?

The emergence of a phonological system can be simulated on the basis of current models of production and perception. FIG. 2 shows some computational results obtained by an application of

$$\sum_{i=2}^{n} \sum_{j=1}^{i-1} T_{ij}(t) \cdot L_{ij}(t) \cdot S_{ij}(t) < \text{CONSTANT} \qquad (1)$$

where n is the size of a universal inventory of segments, T_{ij} represents a (time-varying) talker-dependent measure of evaluation for a given contrast (pronounceability condition), L_{ij}

refers to a listener-dependent evaluation (distinctiveness condition), and S_{ij} reflects the balance between social and phonetic factors. FIG. 2 illustrates the

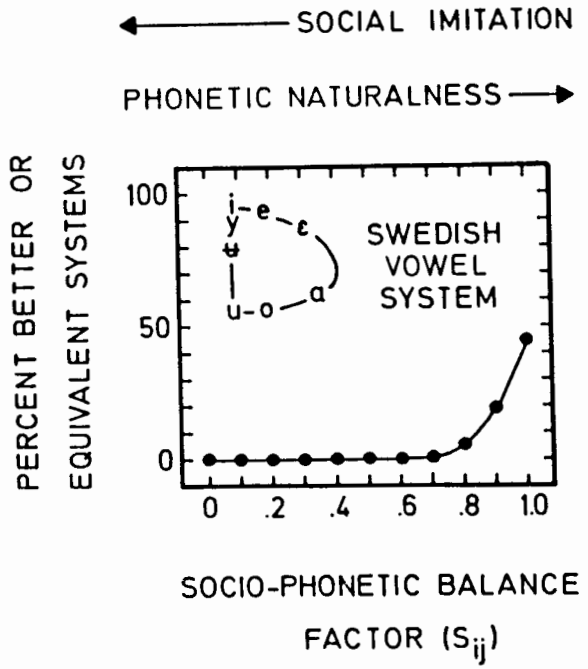

FIG. 2

interaction between the criteria of distinctiveness and social imitation in deriving the Swedish vowel system from a larger set of universal vowel types (represented in terms of canonical auditory patterns). The socio-phonetic balance varies from zero ("social imitation" dominates) to unity (natural phonetic factors, T and L, dominate). It is applied to the contrasts of Swedish with the values shown. For non-Swedish contrasts S=1. Apparently there are many systems (out of a total of 92378) that meet our present criterion of distinctiveness equally well or better. If we had reason to believe that the role of natural phonetic factors in the genesis of the Swedish vowels was correctly and exhaustively reflected in our calculations we would conclude that social factors are quite important in their development. We don't. A great deal of work on phonetic naturalness remains to be done before any safe conclusions can be drawn.

However, we believe that the approach will be useful in studying phonological contrasts particularly in child language and cross-linguistically.

7. A "Darwinian" Theory of Phonological Universals.

Suppose that we answer all the questions of the preceding discussion in the affirmative. We accept as our null hypotheses the assumptions that learnability, pronounceability and perceptibility conditions can account for differences between speech and non-speech sounds, that discreteness reflects the operation of memory, learning and decoding mechanisms, that sound changes are influenced by social variables and shaped by demands for perceptual efficiency and convenience of production, and that the origin of such demands is prosaically behavioral rather than mysteriously teleological. Such acceptance boils down to the idea that phonological structure arises both phylogenetically and ontogenetically by "natural selection" of sound patterns from sources of phonetic variation. Language structure emerges in response to the biological and social conditions of language use. Natural selection is based on the communicative (perceptual as well as social) value of contrasts and "phonetic variation" is defined with respect to possible segment, possible sequence and their possible variation. According to this "Darwinian" theory, phonological universals will be explained with reference to how speech is acquired, produced and understood, or rather in terms of our models of these processes.

This conclusion may seem uncontroversial. However, a truly quantitative and explanatory theory along these lines is not likely to appear until we learn to recognize its full intellectual, educational and administrative implications for how linguistics should be done. Language is the way it is partly because of our brains, ears, mouths as well as our minds. In this sense linguistics is a natural science. Phonetics can contribute by formulating its behavioral explanans principles.

8. References.

BRODDA, B. (1973): "Naturlig Fonotax", unpubl. manuscript, Stockholm University.

CHAFE, W.L. (1970): Meaning and Structure of Language, Chicago and London: The University of Chicago Press.

ELERT, C.C. (1970): Ljud och Ord i Svenskan, Stockholm: Almqvist & Wiksell.

GREENBERG, J.H. (1959): "Language and Evolution", in MEGGERS, B.J. (ed.): Evolution and Anthropology: A Centennial Appraisal, pp. 61-75.

JONASSON, J. (1972): "Perceptual Factors in Phonology", in RIGAULT, A. & CHARBONNEAU, R. (eds.): Proceedings in the Seventh International Congress of Phonetic Sciences, pp. 1127-1131, The Hague: Mouton.

KIPARSKY, P. (1972): "Explanation in Phonology", in PETERS, S. (ed.): Goals of Linguistic Theory, pp. 189-227.

KIPARSKY, P. (1975): "Comments on the Role of Phonology in Language", in KAVANAGH, J.F. and CUTTING, J.E. (eds.): The Role of Speech in Language, pp. 271-280.

LIEBERMAN, P. (1973): "On the Evolution of Language: A Unified View", Cognition 2 (1), pp. 59-94.

LINDBLOM, B., PAULI, S. and SUNDBERG, J. (1974): "Modeling Coarticulation in Apical Stops", in FANT, G.: Speech Communication, vol. 1, pp. 87-94, Almqvist & Wiksell Int.

LINDBLOM, B., LUBKER, J. and GAY, T. (in press):"Formant Frequencies of Some Fixed-Mandible Vowels and a Model of Speech Motor Programming by Predictive Simulation", J. Phonetics.

LINELL, P. (1974): "Problems of Psychological Reality in Generative Phonology: A Critical Assessment", Reports from Uppsala University Department of Linguistics nr 4.

MANDELBROT, B. (1954): "Structure Formelle des Langues et Communication", Word 10, pp. 1-27.

MILLER, G.A. (1956): "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", Psychological Review 63, pp. 81-97.

OHALA, J.J. (1979): "The Contribution of Acoustic Phonetics to Phonology", to be published in LINDBLOM, B. and ÖHMAN, S. (eds.): Frontiers of Speech Communication Research, London: Academic Press.

PIKE, K.L. (1943): Phonetics, Ann Arbor: The University of Michigan Press.

STAMPE, D. (1972): "On the Natural History of Diphthongs", Papers from the 8th Regional Meeting, Chicago Linguistic Society, pp. 578-590.

STEVENS, K.N. (1972): "The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data", in DAVID, E.E. and DENES, P.B. (eds.): Human Communication: A Unified View, New York: McGraw Hill.