UNIVERSALS OF VOWEL SYSTEMS:
THE CASE OF CENTRALIZED VOLWELS*

Jean-Marie Hombert
University of California
Santa Barbara

*Contribution to symposium on Phonetic Universals in Phonological Systems and their Explanation, Ninth International Congress of Phonetic Sciences, Copenhagen, 6-11 August 1979.

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 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 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

---

(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.

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 in 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].

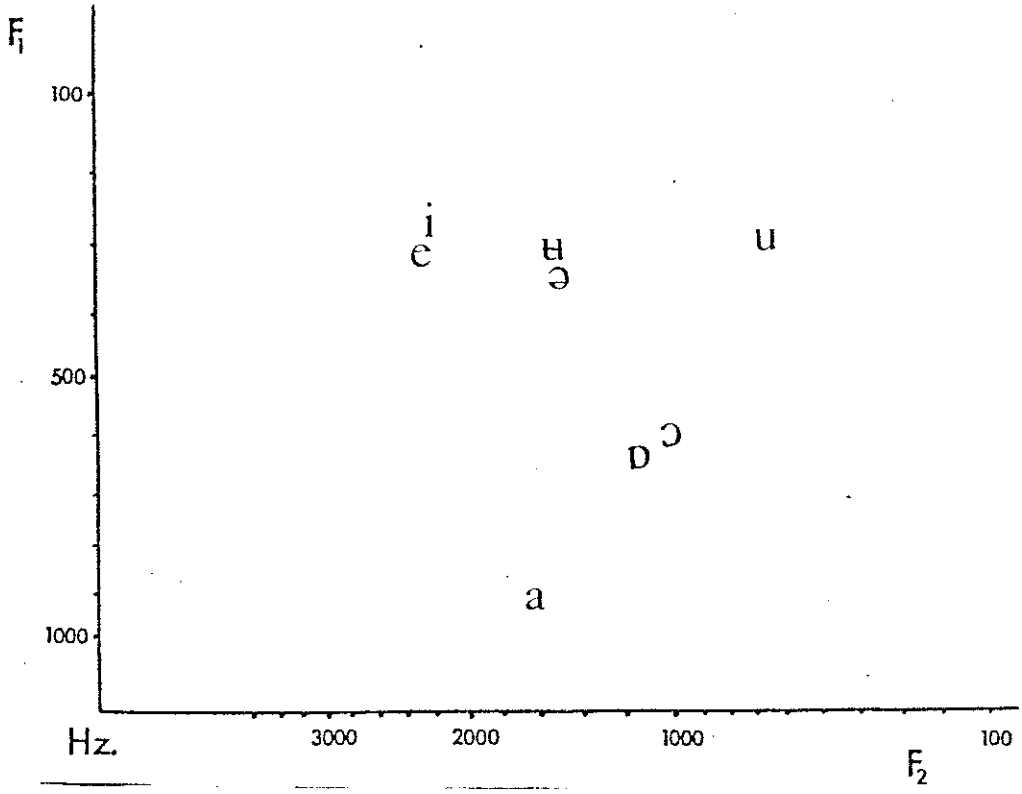

Figure 1. Acoustic data: the Fe?fe? vowel system, (one speaker, average of five measurements).

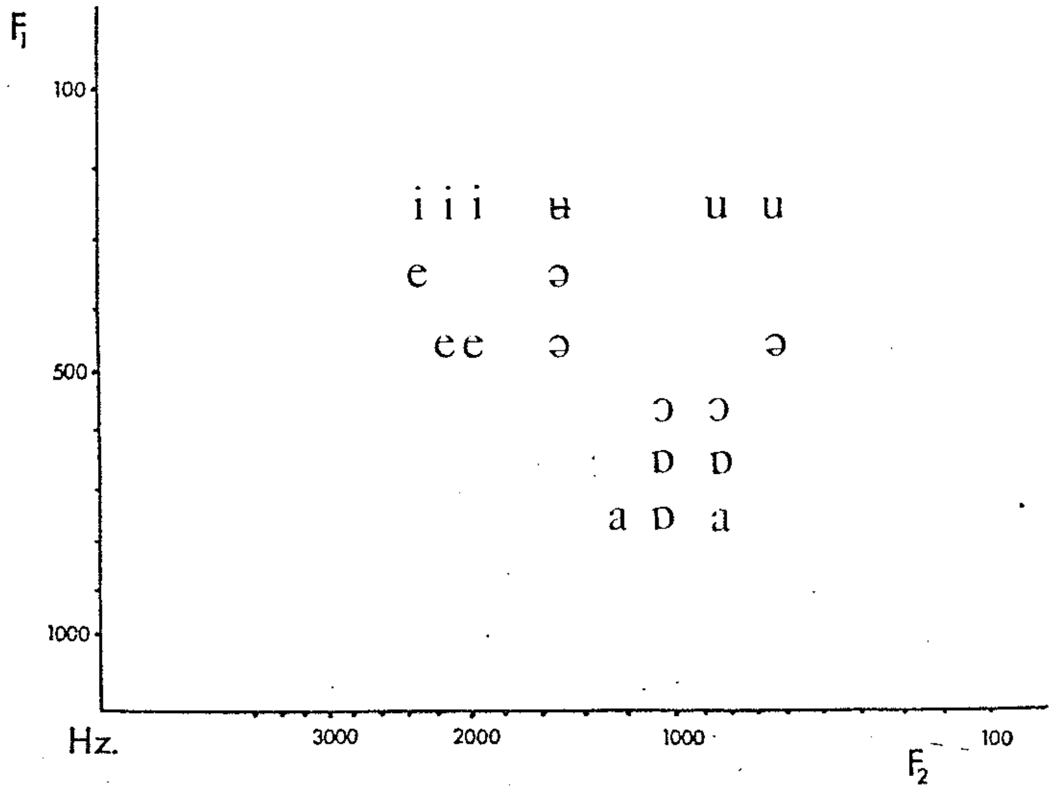

Figure 2. Perceptual data: only stimuli for which the Fe?fe? subject gave at least four out of five identical responses are presented on this graph.

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.

7. Conclusion

In summary we have shown that interesting sound patterns uncovered by comparison of phonemic inventories can be the basis for field investigation and can stimulate new explanations. In particular we have suggested that the relative scarcity of non-peripheral vowels may be due to the lack of perceptual saliency of a frequency region around 1500 Hz.

Acknowledgements

We would like to thank the members of the Bantu Grassfields Working Group who provided us with phonemic analyses for the languages they were familiar with. The Fe?fe? phonemic analysis was provided by Larry Hyman. The data collection was supported by an NSF Grant made to the Bantu Grassfields Working Group (USC) and the data analysis by an NSF Grant made to the UCLA phonetics laboratory.

References

Hombert, J. M. (in preparation): "Acoustic and Perceptual Data from Ten Grassfields Bantu Vowel Systems".

Liljencrants, J. and B. Lindblom (1972): "Numerical simulation of vowel quality systems: the role of perceptual contrast", Language, 48.4, 839-862.