A STUDY ON DISTINCTIVE FEATURES AND FEATURE HIERARCHIES THROUGH "PHONEME ENVIRONMENT CLUSTERING" (PEC)

M. Dantsuji* and S. Sagayama**

*Faculty of Letters, Kansai University, Suita, Osaka, JAPAN
**Dept. of Speech Processing, ATR Interpreting Telephony Research Laboratories, Kyoto, JAPAN

ABSTRACT

The present study concerns the study of distinctive features by means of "Phoneme Environment Clustering" (PEC). The PEC algorithm, originally developed for automatic speech recognition, selects the optimal set of allophones and estimates missing contexts automatically. We have examined approximately 2,000 segments from 216 phonemically balanced words uttered by a male informant of Japanese using PEC. The results show that the feature [sonorant] is separated from others in the earliest stages of the process of the tree structure and coincide with the feature hierarchies proposed in the field of current non-linear phonology.

1. INTRODUCTION

In this paper we would like to describe an attempt to reconsider the hierarchical structure of distinctive features by means of a "Phoneme Environment Clustering" (PEC). Early generative phonologists adopted the distinctive features of the Jakobsonian framework [6]. Later, they revised the distinctive features in many respects. In the framework of SPE, distinctive features are mainly described from an articulatory point of view [1], and the same inclination has been maintained among current approaches. This, however, does not mean that acoustic and auditory aspects have lesser importance, but rather that it was difficult to make an exact and precise description of the acoustic characteristics of distinctive features at that time.

With respect to the hierarchy of distinctive features, several kinds of feature hierarchies have been proposed.

We would like to introduce another kind of hierarchy based on the acoustic distance. PEC, which was originally developed for automatic speech recognition, is one such experiment and attempts the establishment of the feature hierarchy.

2. THE CONCEPT OF PHONEME ENVIRONMENT CLUSTERING (PEC)

We can consider a number of possible factors, which may affect the sound patterns of a given language, such as a preceding phoneme, a phoneme before a preceding phoneme, a center phoneme (the current phoneme itself), a succeeding phoneme, a phoneme after a succeeding phoneme, speakers, pitch frequency, power, speaking rate, stress position, phoneme position in the utterance, background noise, emotion and so forth. The combination of these factors makes an abstract space which is called the environment space E. Each allophone is assumed to be a point e in the space E. On the other hand, each allophone is observed as an acoustic pattern which can be assumed to be a point v in a vector space (V), after some normalization of pattern durations as well.

If we have a set of phonetically labeled acoustic segments, each is a point e in the environment space E as well as a point v in the pattern space V. Denoting a mapping function from the space E to V by $\varphi : E \longrightarrow V$, the acoustic pattern of each allophone $v=\varphi(e)$ varies from sample to sample and has a certain spread in the space V. This spread is measured by some distortion measure, such as an averaged Euclidean distance from the centroid, and denoted by d(v). The image in a subspace Ei of the phoneme environment space E through the mapping function is also a subspace $Vi = \varphi(Ei)$ in the vector space V. Its spread in V is denoted by d(Vi).

The aim of the phoneme environment clustering is to find the optimal set of n subspaces $\{Ei\}_{i=0}^{n-1}$ to cover all variations of acoustic segments. It is defined as the minimization of the total distortion defined by:

$$D = \sum_{i=1}^{n} d(\varphi(Ei))$$

where

$$E = E1 \cup E2 \cup E3 \cup \ldots \cup En$$
$$\text{and } Ei \cap Ej = 0 (i \neq j)$$

That is, PEC aims to find an optimal division of the phoneme environment space to minimize the total sum of the distortions of images of environment subspaces. This formulation means a sort of piecewise approximation of a mapping function such that, if an arbitrary phoneme environment is given, its pattern is predicted with a minimum error. Since it is not easy to obtain the real minimum, the solution to the above problem is approximated by successive splitting of the environment subspaces, which has significant advantages such as, the clustering algorithm is simple, all produced subspaces are convex, the splitting process derives a binary decision tree, and so forth.

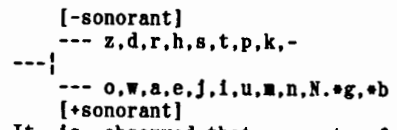
3. EXPERIMENTS ON PEC AND DISTINCTIVE FEATURES

As has been mentioned above, the process of successively splitting subspaces forms a tree structure which is interpreted as a similar grouping of phonemes and the phoneme environment. The concept of PEC can be applied as well to the distinctive features, which are components of phonemes. For Example, Fant (1973) stated that "the phonetic value of a distinctive feature can be regarded as a vector in a multidimensional signal space. The variability due to context shall be expressible by rules which define how the feature vector is changed when the conditioning elements are varied" [5]. Therefore, distinctive features may be extracted to some extent using the PEC procedure.

We have examined how sets of phonemes are divided into allophones in the process of PEC. Experiments were carried out under the following condition.

1)Informant and texts: Approximately 2,000 segments out of 216 phonemically balanced words for one male adult. 2)Acoustic parameters: cepstrum, delta-cepstrum, log-power, delta-log-power. 3)Dimension: 34. 4)Regression window: 90 ms triangular. 5)Window length:30 ms. 6)Window shift: 10 ms. 7)Sampling frequency: 12kHz. 8)Environment factors: 5. 9)Distance measure: weighted Euclidean distance.

The results indicate that allophones depending on phonetic environment are extracted at lower nodes. Phonemes as sets of allophones appropriately correspond to upper nodes which bind the lower nodes of allophones. Still upper nodes tie several phonemes into bundles and these bundles correspond to natural classes. Following diagrams represent parts of the tree structure which was formed through the process of successive splitting using PEC.

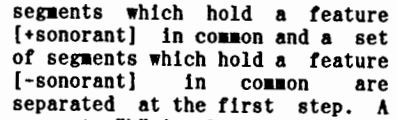


It is observed that a set of segments which hold a feature [+sonorant] in common and a set of segments which hold a feature [-sonorant] in common are separated at the first step. A segment "h" is classified as a member of segments having [-sonorant] in this analysis. In the case of Japanese, the phoneme /h/ occurs as allophones [ç], [ɸ], [x], and [ɦ] in addition to [h], and this phoneme is not usually classified as a glide. Therefore, there is no problem in classifying this segment as [-sonorant].

With respect to /r/, this segment is an approximant (semi-vowel) in the case of English, and this would be classified as [+sonorant]. In the case of Japanese, however, this segment has quite a number of allophones and free variations. For example, /r/ is often represented as a kind of plosive at word initial positions, and as a flap at word-medial positions. It is assumed that this segment is accordingly classified as [-sonorant] in this instance.

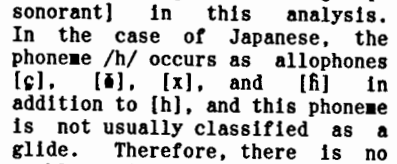
Attaching an asterisk (*) to g and d implies a special case. These segments are originally voiced plosives and should be classified as [-sonorant]. At the stage of labeling preconditioned the phoneme environment clustering, transition portions of formants were not included in vowels but included in voiced plosives. Therefore, some properties of vowels, which should be classified as [+sonorant], are assigned to these segments in this analysis. Furthermore, /g/ and /b/ seldom occur as voiced plosives [g] and [b]. Rather, they occur as voiced fricatives [ɣ] and [β] or velar nasal [ŋ] called "bidakuon". These are also assumed to be factors.

In the next step, the segments that have features [-high, -consonantal] in common were separated from [+sonorant].

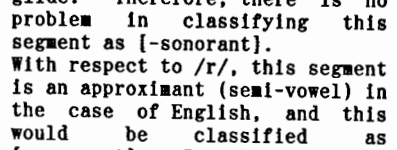


In this analysis /w/ is classified as [-high], although it is classified as [+high] in the case of English. In the case of English, [w] is produced with a constriction between the upper and lower lips and the back of the tongue and soft palate as well, and is a so-called voiced labial-velar approximant. On the other hand, in the case of Japanese, the degree of raising the back of the tongue is lower even at the word initial position, and it is pointed out that is still lower at the word medial position. Therefore, the informant of this analysis reflects such properties of Japanese, and /w/ was classified as [-high].

The group which holds [-high, -consonantal] is subdivided into a group which has a feature [+round], viz./o/ and /w/, and a group which has a feature [-round], viz. /a/ and /e/.



The segments that have a feature [-round] in common are still subdivided into individual phonemes of /a/ and /e/ by a feature [+/- low]. The low vowel /a/ and the non-low vowel /e/ are separated by this feature.



Other groups of segments are also subdivided into individual phonemes in a similar way.

4. DISCUSSION AND CONCLUSION

Recently, there is a tendency to revise not only partial problems but also the total framework of feature systems in many ways. One of the main concerns among them is setting up a hierarchy structure or groupings for the feature arrangement. Until now, several kinds of feature hierarchies or groupings of features have been proposed. For example, in a Jakobsonian framework, Fant (1973) discussed a feature hierarchy depending on the economy of description [5]. From the automatic recognition study, Dantsuji (1989) proposed a feature hierarchy making use of auditory distance [3]. In a generative phonology framework, for example, Clements (1985) discussed feature hierarchy geometrically organized from a phonological point of view considering articulatory aspects. and Sagey (1986) elaborated this feature hierarchy from phonetic and physiological facts [2,9]. These phonetic and physiological facts mean that speech sounds are produced with the movement and action of a physiologically limited number of articulators, as was pointed out by Maddieson and Ladefoged (1989), etc. [7]. Movable articulators are lips, tongue tip, tongue blade, tongue dorsum, tongue root, soft palate, larynx and so forth. Therefore, as terminal features [high]. [back] and [low] have, for example, relevance to the movement of the dorsum of the tongue, they are dominated by a non-terminal node dorsal. As labial, coronal and dorsal are related to the place of articulation, these nodes are dominated by a higher node place. Furthermore, the place node and soft palate node are dominated by a still higher node, the supralaryngeal. However, major class features such as [sonorant] and [consonantal] are directly dominated by a root node which is the highest position of the hierarchy, or situated as special features that constitute the root node.

On the other hand, the analysis by PEC establishes another type of feature hierarchy which reflects the acoustic distance. Features such as [sonorant] and [consonantal] are extracted at quite early steps in this experiment. For example, [sonorant] is extracted at the first step of the clustering. These matters indicate that the acoustic distance between segment groups corresponding to the feature [+sonorant] and [-sonorant] is considerably great. Therefore, this confirms the view that the feature [sonorant] is placed at a higher position of the feature hierarchy, as proposed in current literature of non-linear phonology based on articulatory and physical facts.

5. REFERENCES

[1]CHOMSKY,N. AND M.HALLE (1968), "The Sound Pattern of English", New York: Harper and Row.
[2]CLEMENTS,G.N.(1985), "The Geometry of Phonological Features",Phonology Year Book 2,225-253.
[3]DANTSUJI,M.(1989),"A Tentative Approach to the Acoustic Feature Model", Revue de Phonétique Appliquée, #91,92,93, 147-159.
[4]DANTSUJI,M. AND S.SAGAYAMA(1989),"A Study on Acoustic Aspects of Phoneme Environment Clustering and Distinctive Features", IEICE Technical Report SP 89-79, 25-32, (in Japanese).
[5]FANT,G.(1973), "Speech Sounds and Features", Cambridge, MA, The MIT Press.
[6]JAKOBSON,R, C.G.M.FANT AND M.HALLE (1952), "Preliminaries to Speech Analysis: The Distinctive Features and their Correlate". (1969, Cambridge, MA, The MIT Press)
[7]MADDIESON, I. & P. LADEFOGED (1989),"Multiply articulated segments and the feature hierarchy", UCLA Working Papers in Phonetics, 72, 116-138.
[8]SAGAYAMA,S.(1989), "Phoneme Environment Clustering for Speech Recognition", ICASSP-89, 397-400.
[9]SAGEY,E.(1986),"The Representation of Features and Relations in Non-linear Phonology", Diss., MIT.