European Conference on
Speech Technology
Edinburgh, Scotland, UK
September 1987

10.21437/ECST.1987-221

A KNOWLEDGE-BASED SYSTEM FOR VOICELESS PLOSIVE DECODING

Charles Barrera*.

ABSTRACT

A local learning system for acoustic phonetic decoding of French voiceless plosives is described. A rule-base is worked out from both a labelled acoustic phonetic data-base and a statistical one.

Locally, rules using the concept of fuzziness are arrived at through:
- yielding automatically (based on statistics) restriction boundaries for each µPFi membership function that is associated to a (parameter P, function F) pair,
- working out a synthetic, arborescent AND/OR rule...

Globally, an automaton is defined for each phoneme whose states refer to a part of the rule-base.

Operating within a decoding procedure, this system allows to define rules and to test the various parameter/function pairing combinations.

INTRODUCTION

Nowadays we have, available, an acoustic phonetic data-base (ref 1) that is set up as follows: the vocal signal, being processed through a filter bank, yields a spectrum that undergoes an automated pre-segmentation into homogeneous infra-phonemic units. In turn, these units become specified, in terms of acoustic, phonetic, syllabic properties, through labels that are applied by a skilled phonetician, as so many tags (macro-class, phonemic code, acoustic phonetic realization modality (fricative, nasal,...) contextual attributes, acoustic phases, syllabic code (ref 2). This latter procedure is the keystone of the whole system: upon it, depends the creation of references; thus allowing to adapt any further use of the data-base to a wide variety of units; i.e., syllable, phoneme, infra-phoneme, etc..

Initial results have been extracted (ref 3), through various types of statistical analysis (discriminant, main-component analyses, automatic classification), and stocked within a statistical data-base. We can show it is possible to go further, and simultaneously make use of both the data-bases described above, in order to practice both semi-automatic learning and acoustic phonetic decoding.

1. LOCAL LEARNING (Fig. 1)

1.1 Decoding Plosives We currently concentrate on French voiceless plosives. We are aware that, during occlusion proper or within the area immediately following the ensuing burst, depending upon whether the location of the articulation target is labial, dental or velar, discrimination cues (ref 4, 5) display different behaviors —both spectral and temporal (V.O.T.). However, all three categories do share an energy curve and a spectrum that are both definitely non-continuous and characteristic. Thus, while labelling, we proceeded according to three separate acoustic phases (E implosion, T partial or total occlusion, O coda with or without burst, release with or without aspiration, friction). Learning begins with this observation, and operates locally as it creates per-phoneme and per-acoustic phase rules.

* Laboratoire C.E.R.F.I.A., UA 824-CNRS GRECO, 118 Route de Narbonne,
31062 Toulouse cedex, France

1.2 Setting up rules The user has, available, both a set of parameters (acoustic cues: acute/grave, closed/open,... , formants, prosodic parameters, etc.) and evaluative functions (min., max., mean, derivative,...). Thus, for the occlusive part, the "minimum" of "energy" can be retained; whereas, for the burst, the "derivative" of the cue "Acute/Grave" and/or of "energy" will be preferred.

The parameters are organized into a hierarchy by means of a statistical analysis; namely, their relative importance is determined through a main-component analysis (MCA). Rules are then set up in two stages:

-automated creation —based on elementary statistics— of membership-functions (ref 6) µPFi, that correspond to a (parameter, function) pair. Both the support and the core of the fuzzy function are calculated; using mean, standard deviation, minimum and maximum over several files.

-arborescent and/or human- or expert-controlled organization of µPFi functions. A decoding rule, therefore, emerges as an agregate of fuzzy objectives. Each such objective is weighed according to its relative importance by means of the MCA coefficients.

However, membership-functions amount to standards that may be altered according to context. For instance, within a fricative context, the cue "acute" should be de-emphasized. Within the system, this translates either as an alteration of the µPFi functions (applying + or - linguistic operators, ...), or as an increasing/decreasing criterion weight along the rule-tree. Learning does not proceed blindly: on the contrary, it is assisted by knowledge issuing from the expertise phase, involved in labelling.

1.3 Evaluation The parameters, the rule refers to, are computed over spectral infra-phonemic units and stored within a temporary work-base. The setting in correspondence of a parametric value with a fuzzy form µPFi, yields a belonging-value. An overall score is then atributed to the complete rule, which combines the various fuzzy criteria together, thanks to min. (AND) and max. (OR) operators. Since weight-coefficients are issued from statistics, they impart a more objective character to the various combinations. The decision concerning phoneme-phase is secured through a simple comparison with a set threshold.

For each phase, rules are independently tested over the already labelled spectrum; and this allows to evaluate their quality. Through comparing with the score, obtained for previous rules, the definition of selected rules can be refined; therefore, retaining those that perform best. Similarly, it is possible to consider other parameters, to complete statistical analyses and to test the rules over a larger number of corpuses.

2. GLOBAL LEARNING AND DECODING (Fig. 2)

Once rules have thus been defined, it is then a matter of setting them into a temporal arrangement, in the form of an automaton (ref 7) that represents the associated phoneme. This automaton is made up by several states (the various acoustic phases) along with the transition between them. Both a beginning- and an end-state are fictitiously slapped on. The sequence should be capable of taking into account the various positions the plosive may assume within the phrase —initial, between two vowels, final. For the time being, clusters whose detection is more difficult are not taken up. If located between two vowels, a plosive will successively go through the phases E, T and Q. However, if it is located in the initial position, direct access to phase T is allowed. As well, in the case of an imperceptible occlusion, direct access to the coda Q is possible. Similarly, it is possible to exit from the automaton at the transition T-F, whenever the burst is weak, etc... . Moreover, since an acoustic phase can consist of a string of infra-phonemic units, it is also possible to loop iteratively over each state; although there is a duration rule that steps in to stop this. A coda state Q can be dealt with as a whole or exploded into sub-automata; taking into account finer phenomena in the vicinity of the burst — e.g., micro-cue detection (ref 8). Furthermore, the states are in one-to-one correspondence with the rules, described above. A transition between two successive states is triggered either by excessive duration in one state or by the local score achieved by a rule that is too weak.

CONCLUSION

The system herein described brings together several knowledge sources(ref 9): labelling expertise, acoustic phonetic data and statistical results. It affords a structuring of parameters, thanks to local rules that are linked up in the form of an automaton. Altering rules is made easier through both the modular and the convivial character of the system. Indeed, the package is written in Turbo-Pascal, and it is equipped with powerful graphic interfaces that include pop-up menus and windows for the acquisition of both automata and rules.

This system is akin to an expert system, since it displays the same components; namely, a fact-base (acoustic and phonetic data-base), a rule-base (here, erected in strata) and an inference generating device (here, a rule applicator that works both locally —when looking for relevant rules- and globally —when linking up the temporal phases of a phoneme. This system is devised for voiceless plosives, but it can easily be adapted to other phonemes.

ACKNOWLEDGMENT

I wish to thank J.F. Malet of CSU Sacramento for translating the original French text and for his kind suggestions.

REFERENCES

1. J. Caelen, N. Vigouroux, "Producing and Organizing Phonetic Knowledge from Acoustic Facts in Multi-Level Data-Information",IEEE-ICASSP 2,1209 (1986).

2. C. Barrera, J. Caelen, G. Caelen-Haumont, J.F. Malet, N. Vigouroux, "Towards an Automatic Labelling System", 11th. ICPS-Tallin (1987).

3. G. Caelen-Haumont, N. Vigouroux, "Les indices de distribution spectrale: étude comparative au travers de deux analyses discriminantes monolocuteur et interlocuteur", Speech Com. 2-3, 133 (1983).

4. A. Lahiri, L. Gewirth, S. Blumstein, "A Reconsideration of Acoustic Invariance for Place of Articulation in Diffuse Stop Consonant: Evidence from a Cross-language Study", J.A.S.A. 76-2, 391 (1984).

5. S. Blumstein, E. Isaacs, J. Mertus, "The Role of the Gross Spectral Shape to Place of Articulation in Initial Stop Consonants, J.A.S.A. 68-3, 836 (1980).

6. R. De Mori, Computer Models of Speech Using Fuzzy Algorithms (Plenum Press, NY, 1983) 113.

7. R. Gubrynowicz, K.Marasek, W.W.Wiezlak, "Reconnaissance de mots isolés par la méthode descriptive de traits phonétiques", XV J.E.P., GALF-CNRS, 235 (1986).

8. C. Barrera, J. Caelen, "Micro-indices dans les occlusives sourdes", XV J.E.P., GALF-CNRS, 311 (1986).

9. N. Vigouroux, "Apport Mutuel de savoir entre une base de connaissance et un expert phonéticien", AFCET-RFIA, 297 (1986).

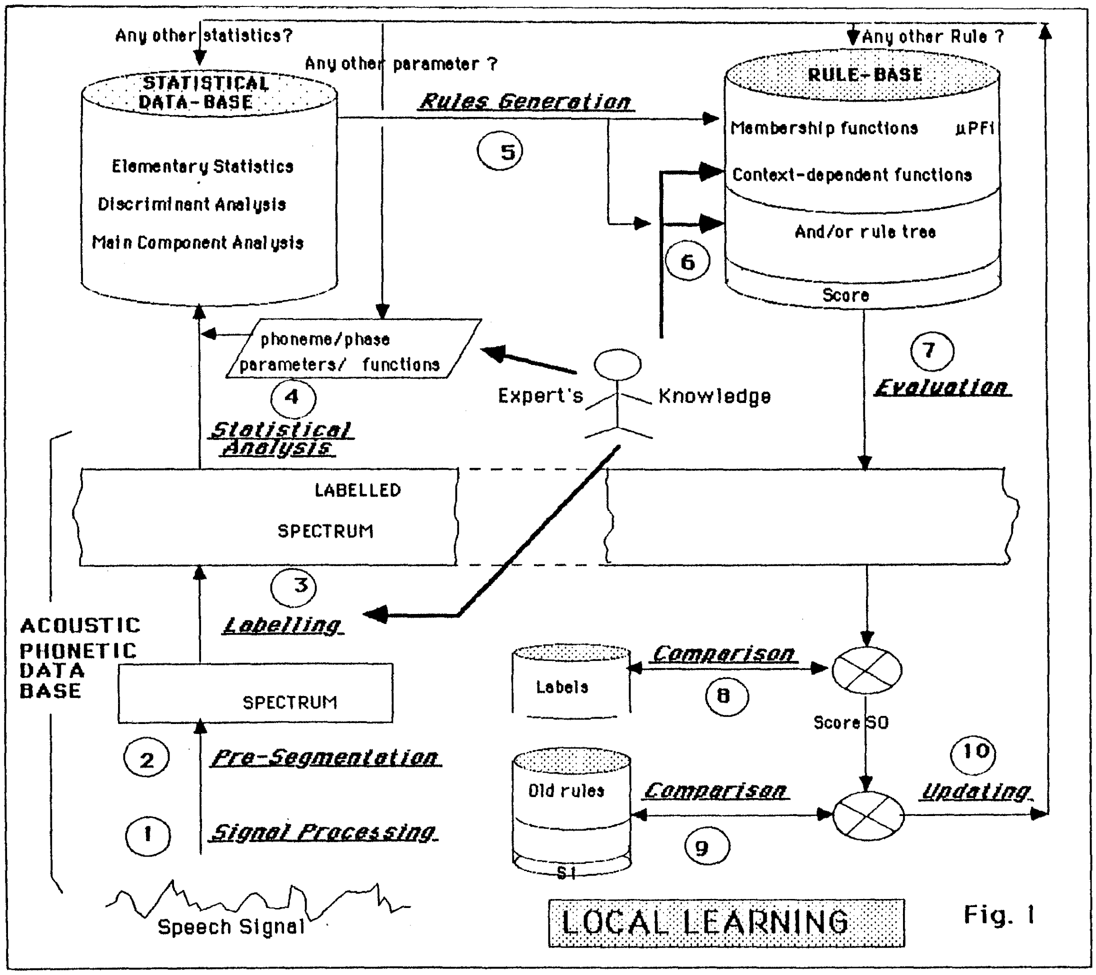
Any other statistics?
Any other parameter ?
Any other Rule ?
STATISTICAL DATA-BASE
Elementary Statistics
Discriminant Analysis
Main Component Analysis
Rules Generation
5
RULE-BASE
Membership functions
µPFi
Context-dependent functions
And/or rule tree
6
Score
phoneme/phase parameters/ functions
4
Expert's
Knowledge
7
Evaluation
Statistical Analysis
LABELLED
SPECTRUM
3
Labelling
ACOUSTIC PHONETIC DATA BASE
Labels
Comparison
8
Score S0
SPECTRUM
2
Pre-Segmentation
Old rules
Comparison
10
Updating
1
Signal Processing
9
S1
Speech Signal
LOCAL LEARNING

Fig. 1

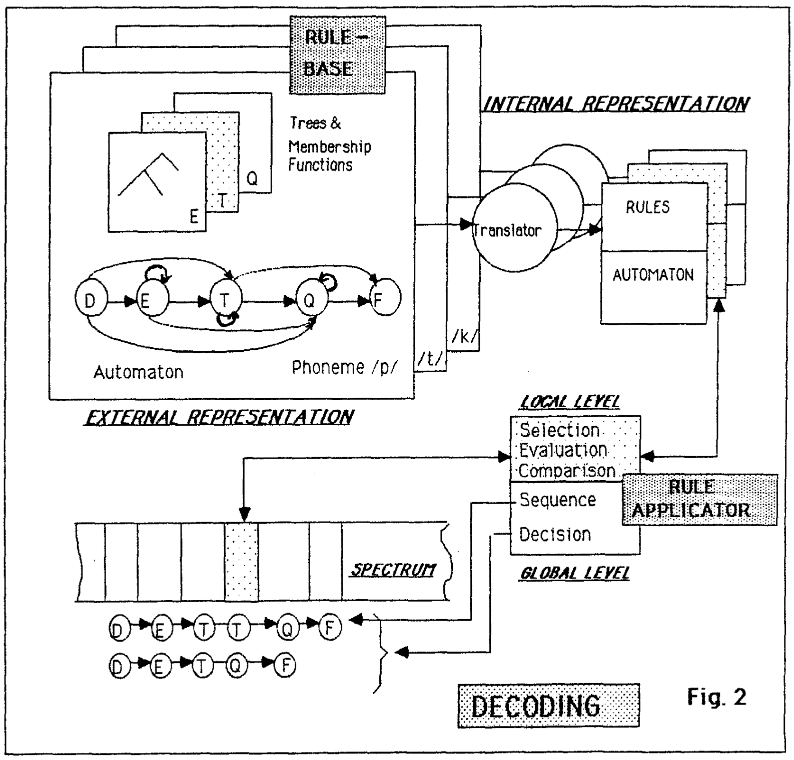
RULE-BASE
Trees &
Membership
Functions
Q
T
E
INTERNAL REPRESENTATION
Translator
RULES
AUTOMATON
D
E
T
Q
F
Automaton
Phoneme /p/
/t/
/k/
EXTERNAL REPRESENTATION
LOCAL LEVEL
Selection
Evaluation
Comparison
Sequence
Decision
RULE APPLICATOR
SPECTRUM
GLOBAL LEVEL
DECODING

Fig. 2