A DATA BASE TO EXTRACT PROSODIC KNOWLEDGE WITH A NUMERICAL AND SYMBOLICAL LEARNING SYSTEM

J.J. Schneider and H. Méloni
Laboratoire d'Informatique d'Avignon, France

ABSTRACT

In this paper we present a representation language to conceive prosodic data bases to extract prosodic knowledge in explicit rules form with a learning system working with numerical and symbolical information.

The multiplicity of parameters at work (acoustic, linguistic, syntactic, semantic, pragmatic, ...) make the realisation of a prosodic labelling complex. In the first time, we present the common basis principles of the prosodic models, and we define a representation language suited to this models. In the second part of this paper we present, in succinct way, the general methodology and tools to extract the prosodic knowledge. And, to illustrate our aims, we propose a prosodic labelling of the utterance to realise a prosodic data base on which we use the previous tools.

PROSODIC LABELLING - METHODOLOGY

All prosodic models, in sprite of their differences, have two common basis principles : (a) the melodic, rhythmic and intensity continuum could be segmented into discreet units, (b) the great congruence between the organisation of the utterance and prosodic organisation of the message [1][2][4].

THE INSTANCE LANGUAGE

For each unit, we combine a description of the prosodic parameters with a description of the organisation of the utterance. To realise this language we define the following descriptions.

An Elementary Description

The smaller description unit, called elementary description, is a pair (attribute value). Attribute is a symbolic label whose meaning is known by the expert, and value is an homogeneous data list which characterise the elementary description.

Example 1 : elementary descriptions

```
(PointA (+1.1 -3.05 +5.2))
(ValeurMax (128))
(Couleur ("bleu foncé"))
```

A simple data list don't exactly characterise an elementary description. For each one [3], we combine a type, a nature and an environment defined as following.

The type is the set of possible values. We define two types, the first one is numerical to describe crude data, and the second is symbolical to describe abstract concepts.

To describe the data behaviours we define three natures. The atomic descriptor's values are independent each other, the linear descriptor's values belong to a completely arranged set, and the structured descriptor's values belong to a generalisation tree.

The environment contain several information, one to describe the data presentation after organisation, and one to describe the list of mathematics tools needed.

The Description of a Parameter

For a best organisation, we define a parameter to group some elementary descriptions together.

```
(Energie
  (AlterationE (+1.3))
  (ContourE (71 66 65)))
(FrequenceFondamentale
  (ContourLocalF0 (167 120 137))
  (ContourGlobalF0 (135 120 116))))
```

The Description of an Instance

Two elements compose an instance, the description of the prosodic parameters, and the description of the organisation of the utterance. Each element of an instance is a list of descriptions of parameters.

```
( ( (FrequenceFondamentale
       (ContourLocalF0 (167 120 137))
       (ContourGlobalF0 (135 120 116)))
    (Duree
       (AlterationD (-11.4)))          )
  ( (StructureGrammaticale
       (Mot ("verbe"))
       (Groupe ("groupe verbal")))     ) )
```

The Data Base

An instances succession compose the data base, in which all instances must have the same structure.

DATA ORGANISATION - KNOWLEDGE ACQUISITION

In the first time, we regroup all instances in which the organisation of the utterance is exactly the same. This step could be supervised by an expert according to the clustering technical used. We obtain a representation more concise and structured of the data. To express this result we propose a set of possible representations.

The next step determine, with a generalisation on the set of the descriptions of the organisation utterance, a characterisation of the phenomena observed on the corpus used.

PROSODIC LABELLING

The Corpus

The corpus chosen to validate our approach had been conceived by V. Aubergé and G. Bailly within the context of the MULTIDIF 1991 contract.

The Description Unit

In this paper the description unit chosen is the vowel. It could be easy extended to any other unit, like syllable or word, according to the phenomena looked for.

The Prosodic Parameters

We propose for the prosodic parameters (fundamental frequency, duration and intensity) a set of marks and a representation of them.

Fundamental Frequency

To underscore the local and global phenomena we define two elementary descriptions, the first one called local fundamental frequency, and the second one called global fundamental frequency. For each description we propose three points defined in the *Figure 1* and *Figure 2*.

Figure 1: local fundamental frequency

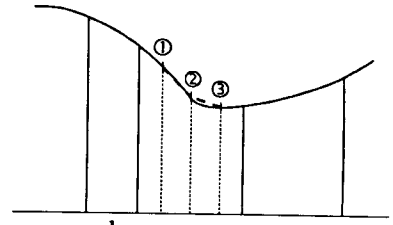
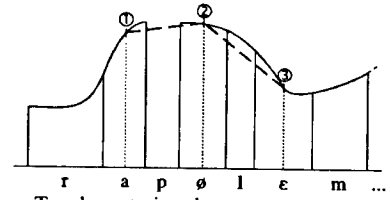


Figure 2: global fundamental frequency



To characterise these movements we use the variation rate between the three points, in a height values scale [4]. The following example is a description of the prosodic parameter fundamental frequency.

```
(FrequenceFondamentale
  (ContourLocalF0 (-3 -1))
  (ContourGlobalF0 (+1 -3)))
```

Duration

For each vowel we could establish the intrinsic duration [5]. We calculate with (1) an alteration rate between the measured and the theoretical duration.

$$\%D(V) = \frac{d(V)}{d_{th}(V)} - 1 \qquad (1)$$

with %D(V) duration alteration rate
d(V) measured duration
$d_{th}(V)$ theoretical duration

We don't calculate the intrinsic duration of each vowel but we calculate an average duration. If vowels are in varied linguistic contexts and if the corpus is big enough, this value is satisfactory.

$$d_{th}(V) = \frac{1}{N}\sum_{i=1}^{N} d_i(V)$$

with $d_{th}(V)$ theoretical duration
N total number of apparitions
$d_i(V)$ i^{th} measured duration

To express this alteration rate we use a symbolical elementary description defined in the *Table 1*.

Table 1 : relations between numerical and symbolical values.

[... -10%]	[-10% 10%]	[10% ...]
"réduite"	"normale"	"allongée"

The following example is a description of the prosodic parameter duration.

```
(Duree
  (AlterationD ("réduite")))
```

Intensity

To describe the prosodic parameter intensity we use two elementary descriptions. The first one, like duration, is an alteration rate between the theoretical intensity and the measured intensity of the vowel.

$$\%E(V) = \frac{e(V)}{e_{th}(V)} - 1 \qquad e_{th}(V) = \frac{1}{N}\sum_{i=1}^{N} e_i(V)$$

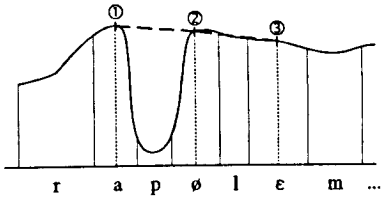
To express this alteration rate we use a symbolical elementary description defined in *Table 2*.

Table 2 : relations between numerical and symbolical values.

[... -5%]	[-5% 5%]	[5% ...]
"basse"	"normale"	"élevée"

The second elementary description, like fundamental frequency, underscore the relative intensity evolution between the previous vowel, the current vowel and the follow vowel in the sentence. We propose three points defined in the *Figure 3*.

Figure 3: intensity movement



To characterise these movements we use the variation rate between the three points, in a height values scale. The following example is a description of the prosodic parameter intensity.

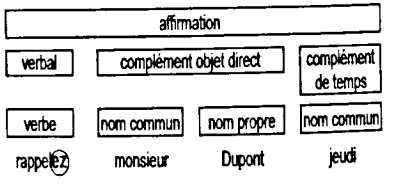
```
(Energie
  (ContourE (-1 -2))
  (AlterationE ("normale")))
```

The Utterance Organisation Parameters

The Syntax

In French the syntax is the most influent utterance organisation parameter.

Figure 4 : syntactic tree representation



From the syntactic tree (see *Figure 4*), we define two parameters. The first one describe the grammar nature of each node of this tree. The second describe the position in the tree of each constituent in the upper constituent. The following example is a description of the grammar structure and the positions structure.

```
(StructureGrammaticale
  (Mot ("verbe"))
  (Groupe ("groupe verbal")))
(StructurePositions
  (Voyelle (3 3))
  (Mot (1 1))
  (Groupe (1 3)))
```

The Phonetic Context

To underscore the influence of the phonetic context on the prosodic parameters we propose three parameters. Two for the phonetic description of the adjacent consonants and one for the phonetic description of the current vowel. For each parameter we use two elementary descriptions, one for the nasality and one for the opening (for a vowel), and one for the voicing and one for the articulatory mode (for a consonant). The following example is a description of the phonetic context.

```
(ConsonnePrec
  (ModeCP ("vocalique"))
  (VoisementCP ("voisée")))
(ConsonneSuiv
  (ModeCS ("occlusive"))
  (VoisementCS ("non voisée")))
(Voyelle
  (NasaliteV ("orale"))
  (OuvertureV ("ouverte")))
```

The Pause

To underscore the influence of the pause on the prosodic parameters we propose two parameters. The first one describe the distance to the next pause, expressed in milliseconds, and the second describe the duration of this pause expressed milliseconds to. The following example is a description of the pause.

```
(Pause
  (DistancePause (430))
  (DureePause (90)))
```

RESULTS - CONCLUSION

We have made tests on a corpus labelled in a voluntary simple way and the results obtained are very interesting.

With this language we can describe the general phenomena found in the prosodic models. The prosodic data base presented could be extended to any other parameters like semantic, pragmatic, ...

The originality of our approach is the system only know the structure of an instance. The segmentation unit and the description of the prosodic parameters and the organisation of the utterance parameters could be different according to the corpus and the phenomena looked for. In the same way, clustering and generalisation tools are used like resources, and they could be different.

In conclusion, the most interesting aspect of the environment presented is his opening, to the phenomena descriptions, to the description unit and to the techniques.

REFERENCES

[1] J.J. Schneider P. Robineau and H. Méloni "*Un système d'Apprentissage Symbolique et Numérique pour la Formalisation de connaissances Prosodiques*" XX^{ème} Journées d'Etude sur la Parole - Trégastel 1994

[2] L. Mortamer F. Emerard and L. Miclet "*Attempting automatic prosodic knowledge acquisition using a database*" Workshop on Speech Synthesis - Autrans 1989

[3] R.S. Michalski "*A Theory and Methodology of Induction Learning*" Machine Learning : An Artificial Intelligence Approach vol. 1 - 1983

[4] G. Caelen-Haumont "*Stratégie des locuteurs en réponse à des consignes de lecture d'un texte : analyse des interactions pragmatiques et des paramètres prosodiques*" Thèse de Doctorat d'état 1991

[5] Y. Nishinuma S. Barber and D. Hirst "*Estimation de la durée intrinsèque des voyelles*" XII^{ème} Journées d'Etude sur la Parole - Montréal 1981