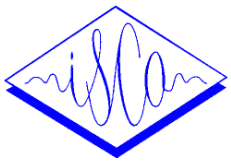

The ESCA Workshop
on Speech Synthesis
Autrans, France
September 25-28, 1990

MULTI-STYLE PROSODIC MODEL FOR FRENCH TEXT-TO-SPEECH SYNTHESIS

Valérie Pasdeloup

Université du Québec, Institut National de la Recherche Scientifique,
3, Place du Commerce, Ile-des-Soeurs, Québec, CANADA H3E 1H6

ABSTRACT

We present an attempt for model prosody, including a prosodic hierarchical organization and a series of linguistic and phonotatic rules. Unlike most other existing French speech synthesis systems where prosodic organization depends mainly on syntactic structure, we adopt an approach where the phonotatic constraints (on length of stress group, number of sequential unstressed syllables...) resulting biological and psychological prerequisites for speech activity are taken into account. In text-to-speech synthesis, our model proves effective in generating various acceptable prosodic structures for a given sentence.

1. INTRODUCTION

Most of the prosodic models developed during the last two decades are originally based upon a syntactic analysis to generate accentuation or intonation: prosody is considered to be congruent to syntax (Martin, 1975 ; Di Cristo and Rossi, 1977 ; Liberman and Prince, 1977, Di Cristo, 1981 ; Dell, 1984). Therefore, according to this hypothesis, the following two sentences which have identical syntactic structure but have a different syllable count (10 compared to 20) are associated with an equivalent stress structure (stressed syllables are coded I, unstressed syllables -).

Le chat de Bill di gère un gros pou let.
\- I - I - I - I - I

Les rhi no cé ros de Mar gue rite re gar daient de gi gan tes ques pé li cans.
\- - - - I - - - I - - I - - - I - - - I

In more recent work on prosodic models, some importance has been given to phonotatic (also called eurhythmic) phenomena such as: principles of accentual alternation (strong/weak) and syllabic balancing, constraints on the length of the stress group and on the number of consecutive unstressed syllables (Dell, 1984 ; Rossi, 1985 ; Martin, 1986). However, these phonotatic constraints are generally considered only after having derived the prosodic structure from the syntactic structure, therefore compensating for the weaknesses of these models.

In general, phonotatic principles - which result from biological and psychological constraints, according to our hypothesis - and linguistic principles are not considered to belong, in the same way, to speech activity.

Prosodic and perceptual analysis of 400 read utterances and temporal analysis of 96 utterances (reiterated with ma-ma-ma syllables) have been carried out. We present a prosodic model which considers linguistic and phonotatic constraints. This model offers also the possibility to generate various acceptable prosodic structures for a given sentence.

2. LINGUISTIC CONSTRAINTS

In spoken communication, linguistic constraints (demarcation and juncture of constituents) are partially realized in the prosodic organization. According to the type of communication, we need to consider the hierarchy and boundaries of the various linguistic structure levels.

These principal linguistic structure levels are the morphemic level, the lexical level, the syntactic level, the semantic level, the enunciative level (supply of information, presupposition and hierarchy of information (focus)), the expressive level.

The weight and nature of the linguistic constraints depend on: the linguistic structure level (e.g., syntactic, lexical), the hierarchical degree of this level (e.g., minor or major syntactic boundary), type of speech (colloquial speech, scientific lecture and reading...) and also on the speaker's regional and individual characteristics. Linguistic constraints vary in strength and can be related to different domains (syntactic, lexical units...). Some are obligatory (demarcation between subject and verbal phrases), some are optional (systematic occurrence of a secondary accent at the beginning of a word) and some are dependent on constraints of other types (phonotatic for example).

3. PHONOTATIC CONSTRAINTS

In the prosodic organization of a sentence, the phonotatic principles are associated with biological and psychological constraints of rhythm production (Fraisse, 1956, 1967, 1974). According to our hypothesis, secondary accent has a regulatory function in the stress structuration of a sentence. Thus, the stress pattern in a given utterance would be consistent with the biological and psychological standards of rhythm production.

The main phonotatic constraints are the following :

- *size of stress groups*: a stress group - i.e. in French, one or more unstressed syllables followed by a stressed one - is composed of few syllables: generally about 2 to 5, in average 3. The stress group is made up of at least one stressed syllable and of a maximum of 7 or 8 syllables. The average and maximum size of stress groups varies with the speaking rate (Fraisse, 1974) and the speaker.

- *number of consecutive unstressed syllables*: as a rule, we try to avoid more than 4 consecutive unstressed syllables (exception occur in cases of: bracketing, rapid speaking rate...). The stress clash rule is complementary to this one.

- *rule of the first accent in a sentence*: this one is realized as soon as possible, usually on one of the first stressable syllable which is often the first syllable starting with a consonant in the first word of the sentence.

- *phonotatic function of secondary accent*: secondary accent in French is located mainly on the first syllable of a word or of a group, on the antepenultimate syllable of a word, or at the boundary of a morpheme in a polymorphemic word (mélo/dramatique) (Pasdeloup, 1988). According to our hypothesis, secondary accent has a phonotatic and linguistic function (demarcation of units), whereas primary accent (placed on the last syllable of a word) has only a linguistic function. Secondary accent generally occurs so that a series of unstressed syllables never exceed the count of 4, and is distributed to an average of one accent (primary or secondary) per 3 syllables (Pasdeloup, 1990). Therefore, when two primary accents are too distant from each other (example: "L'avancée dans la technologie"), a secondary accent (example: on the syllable "tech") would allow avoiding a lengthy series of unstressed syllables (Fraisse, 1967).

- *rhythmic structure of the whole utterance*: accentual or syllabic duration alternation principles (Bruce, 1983 ; Duez et Nishinuma, 1985 ; Dell, 1984) ; recurrence of stress groups, intonative patterns and temporal sequences principles (Fraisse, 1956 et 1967 ; Fónagy, 1983 ; Wioland, 1984).

4. A MODEL FOR PROSODIC ORGANIZATION OF A SENTENCE

Based upon linguistic and phonotatic constraints described in previous sections, we propose an example illustrating the steps required to obtain various acceptable prosodic structures for the same sentence (for more details refer to Pasdeloup 1988 and 1990_b):

(1) Sentence segmentation into rhythmic sequences: major intonation groups (acoustically defined by a considerable amplitude of the pitch contour and an important final-syllable lengthening). Example :

Ph1 / Les rhinocéros de Marguerite / regardaient de gigantesques pélicans /.

(2) Segmentation into rhythmic words: minor intonation groups (medium or weak amplitude of pitch contour, negligible or no syllabic lengthening).

Ph1A / [Les rhinocéros] [de Marguerite] / [regardaient] [de gigantesques] [pélicans] /.
Ph1B / [Les rhinocéros] [de Marguerite] / [regardaient] [de gigantesques pélicans] /.

(3) Accentuation:

(3a) All potential primary and secondary accents are located (Pasdeloup, 1988, 1990_b): the rhythmic word's last stressable syllable, and the antepenultimate syllable of a lexical word of at least 3 syllables are accentuated ; the rhythmic word's first stressable syllable is accentuated.

```
      [-  -  I  -  I]  [-  -  -  I]  /  [-  -  I]  [-  -  -  -  -  -  I]
Ph1A  [Les rhi no cé ros]  [de Mar gue rite]  /  [re gar daient]  [de gi gan tesques pé li cans].
      [-  I  -  -  -]  [-  I  -  -]  /  [I  -  -]  [-  I  -  -  -  -  -]

      [-  -  I  -  I]  [-  -  -  I]  /  [-  -  I]  [-  -  -  I  -]  [-  -  I]
Ph1B  [Les rhi no cé ros]  [de Mar gue rite]  /  [re gar daient]  [de gi gan tesques]  [pé li cans].
      [-  I  -  -  -]  [-  I  -  -]     [I  -  -]  [-  I  -  -  -]  [I  -  -]
```

(3b) stress structures are generated considering the phonotatic rules stated in the preceding section:

```
      [Les rhi no cé ros]  [de Mar gue rite]  /  [re gar daient]  [de gi gan tesques pé li cans].
Ph1A1 /-  I  -  -  I  -  -  -  I  /  -  -  I  -  I  -  -  -  -  I/
Ph1A2 /-  -  I  -  I  -  -  -  I  /  -  -  I  -  I  -  -  -  -  I/
Ph1A3 /-  I  -  -  I  -  I  -  I  /  -  -  I  -  I  -  -  -  -  I/

      [Les rhi no cé ros]  [de Mar gue rite]  /  [re gar daient]  [de gi gan tesques]  [pé li cans].
Ph1B1 /-  I  -  -  I  -  -  -  I  /  -  -  I  -  -  -  I  -  -  -  I/
Ph1B2 /-  I  -  -  I  -  -  -  I  /  -  -  I  -  I  -  I  -  -  -  I/
```

5. CONCLUSION

Speech activity is the result of a compromise between constraints and possibilities which are of linguistic (ex: syntactic and lexical), and biological and psychological type. A biolinguist such as Lieberman (1984, 1986) states the existence of constant interaction between animal and human

behavior with the biological mechanisms to which they are associated: "human beings have innate biological capacities that underlie linguistic ability" and "human language is the result of a mosaic of biological mechanisms" (1986).

The proposed model takes into consideration both linguistic and phonotatic constraints and not only, as in the past, linguistic constraints. Therefore two sentences having the same syntactic structure but composed of a different number of syllables will not be given the same prosodic structure. However, further research on other aspects of phonotatic constraints is required (as an example, to what extent does the proportion of accents to the number of syllables vary with speaking rate?)

Also, this models has the advantage of generating different prosodic structures for a given sentence. It is thus possible to define in a speech synthesis system different styles or prosodic strategies. For example, a style characterized by the systematical presence of a secondary accent at the beginning of a word (when possible) can be defined. This model has been tested in text-to-speech synthesis system. First results show a significant improvement on the prosodic quality of synthesized speech.

REFERENCES

Bruce, G. (1983) On rhythmic alternation, *Working papers of Lund.*

Di Cristo, A. et Rossi, M. (1977) Propositions pour un modèle d'analyse de l'intonation, *8e J. E. P.*, Aix-en-Pce.

Di Cristo, A. (1981) L'intonation est congruente à la syntaxe: une confirmation, *L'intonation, de l'acoustique à la sémantique, Etudes Linguistiques*, **25**, Klincksieck, Paris.

Dell, F. (1984) L'accentuation dans les phrases en français, *Forme sonore du langage*, Hermann, Paris .

Duez, D. ; Nischinuma, Y. (1985) Some evidence on rhythmic patterns spoken french, *Perilus*, **3**, 1984-1985, Stockholm.

Fónagy, I. (1983) La Vive Voix, Payot, Paris.

Fraisse, P. (1956) Les structures rythmiques, *Studia Psychologica*, Publications Universitaires de Louvain.

Fraisse, P. (1967) Psychologie des rythmes humains, colloque "Les rythmes", Lyon 4 Déc. 1967, *Journal Français d'Oto-Rhino-Laryngologie*, sup. n° 7, SIMEP.

Fraisse, P. (1974) Psychologie du rythme, PUF, Paris.

Rossi, M. (1985) L'intonation et l'organisation de l'énoncé, *Phonetica*, **42.**

Liberman, H. ; Prince, A. (1977) On Stress and Linguistic Rhythm, *Linguistic Inquiry*, **8.**

Lieberman, Ph. (1984) The Biology and Evolution of Language, Harvard University Press, Cambridge.

Lieberman, Ph. (1986) A reply to Jacques Mehler's Review of "The Biology and Evolution of Language", *JASA*, **80.**

Martin, Ph. (1975) Eléments pour une théorie de l'intonation, *Rapp. Inst. Phoné. Bruxelles*, **9** (1).

Martin, Ph. (1986) Structure prosodique et structure rythmique pour la synthèse, *15e J. E. P.*, Aix-en-Pce, 27-30 Mai.

Pasdeloup, V. (1988) Essai d'analyse du système accentuel du français: distribution de l'accent secondaire, *17e J. E. P.*, Nancy, 20-23 Sept. 1988.

Pasdeloup, V. (1990_a) Les contraintes linguistiques et phonotactiques dans la structuration accentuelle d'un corpus lu, *1er Congrès d'Acoustique*, Lyon, 10-13 Avril 1990.

Pasdeloup, V. ; Modèle de règles rythmiques du français appliqué à la synthèse de la parole, Thèse de Doctorat de 3ème cycle, Université d'Aix-en-Provence, Aix-Marseille I (soutenance prévue en 1990_b).

Wioland, F. (1984) Organisation temporelle des structures rythmiques du français parlé, *Bulletin des rencontres régionales de linguistique*, Lausanne, 7-9 Juin 1984.