ACADÉMIE D'AIX-MARSEILLE
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE

THÈSE

Présentée à l'Université d'Avignon et des Pays de Vaucluse
pour obtenir le diplôme de DOCTORAT

SPÉCIALITÉ : Informatique

Traitement de la prosodie en reconnaissance automatique de la parole

par

Philippe Langlais

Soutenue le 11 octobre 1995 — devant le Jury composé de :

MM	Gérard Chollet	Examinateur
	René Collier	Examinateur
	Albert Di Cristo	Rapporteur
	Marc El-Bèze	Président
	Joseph-Jean Mariani	Rapporteur
	Henri Méloni	Directeur de Thèse

À mes parents
À Patricia

ABSTRACT

In this study, we investigate the treatment of prosody in automatic speech recognition systems. The major steps of a classical prosodic approach (parametrization, microprosodic and perceptive corrections, application of suprasegmental rules) are discussed, thus introducing the choices made for each of them.

In the first part of this dissertation, the segmental variations are analyzed in detail. A selection of the phenomena which have been most investigated in the past is proposed. Each is then studied on isolated words corpora for two specific purposes : firstly, to evaluate if the automatic parameters extraction techniques used here allow the use of microvariations to improve an acoustico-phonetic decoding process ; secondly, to measure the pertinence of automatic microprosodic corrections. The study shows that only few phenomena can be significantly observed by means of these techniques. The reliable phenomena have been successfully integrated in a lexical access system.

The second part presents the major difficulties of prosodic analysis carried out by experts and attempts to explain why statistical methods are more widely applied. An automatic correlative system has been elaborated ; firstly, it upholds an assistance to prosodic analysis (providing vizualization and query tools), and secondly, it gives a predictive function of the linguistical structure of the message to decode. Two applications of this system are proposed ; a first one for the recognition of decimal numbers (our system is able to locate the word "virgule"in an unknown number, only by means of prosodic information) and a second one for the recognition of read isolated sentences. The results obtained fully validate the approach we proposed.

Pour l'intérêt qu'ils ont spontanément témoigné à mes activités de recherche, je tiens à exprimer mes remerciements et ma sincère reconnaissance aux membres du jury :

Henri Méloni, professeur à l'Université d'Avignon et des Pays de Vaucluse, qui a dirigé mes recherches et m'a encouragé de ses précieux commentaires, conseils et critiques ;

Albert Di Cristo, professeur à l'Institut de Phonétique d'Aix-en-Provence, qui m'a fait l'honneur d'examiner ce travail à la lumière de ses compétences en prosodie ;

Joseph-Jean Mariani, directeur du Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, qui malgré ses très nombreuses responsabilités a pris le temps d'étudier minutieusement ce mémoire ;

Gérard Chollet, directeur de recherche CNRS à l'École Nationale Supérieure des Télécommunications, qui m'a chaleureusement accueilli au sein de l'équipe parole de l'IDIAP et qui n'a eu de cesse de me voir mener à bien ce travail ;

Marc El-Bèze, professeur à l'Université d'Avignon et des Pays de Vaucluse, qui a toujours fait preuve de gentillesse et de disponibilité ;

René Collier, professeur à l'Institute for Perception Research, spécialiste en prosodie, qui a eu l'amabilité de faire également partie de ce jury.

Table des matières

Liste des figures

Liste des tableaux

Chapitre 1

Position du problème

Ce premier chapitre a pour objectif d'introduire le travail présenté dans ce mémoire, en prenant soin d'expliquer les choix méthodologiques retenus à la lumière de la littérature concernée.

1.1 Définition de la prosodie

Durant plusieurs années consacrées à l'étude de la prosodie, j'ai souvent dû répondre à la même question lorsque quelqu'un était suffisamment curieux pour me demander le sujet de ma thèse :

> "La prosodie... C'est quoi ?"

Une première réponse simple mais peu courtoise aurait été de renvoyer mon interlocuteur à un dictionnaire, ce qui ne lui aurait pas rendu grand service tant les définitions que l'on peut y trouver sont confuses :

> PROSODIE [*ρzdi*] n.f. — 1573 "bonne prononciation" 1562 ; gr. *prosôdia* "accent, quantité, dans la prononciation". Caractères quantitatifs (durée) et mélodiques des sons en tant qu'ils interviennent dans la poésie (→ **métrique, versification, mètre, pied**) ; règle concernant ces caractères. "En apprenant la prosodie d'une langue, on entre plus intimement dans l'esprit de la nation qui la parle" (Staël). Règles concernant les rapports de quantité, d'intensité, entre les temps de la mesure et les syllabes des paroles, dans la musique vocale. Ling. Étude de l'accent et de la durée des phonèmes.

J'ai alors très vite adopté un comportement consistant à ne répondre que de manière succincte (par une formule ressemblant à : "la prosodie ? Et bien c'est en quelque sorte la musique de la parole") proposant ensuite aux personnes non satisfaites de cette réponse des explications plus fournies...
Le terme de prosodie est souvent utilisé pour regrouper différents phénomènes liés de la communication orale (comme l'accent, le rythme, l'intonation, les tons). La conception

des éléments prosodiques a évolué avec le nombre des études, de telle façon que le simple rôle de véhicule des expressions et des sentiments qu'on leur accordait avant les années 70 a fait place à un statut linguistique incontestable (au moins pour un sous-ensemble de ces éléments). Il convient de décomposer les faits prosodiques en trois classes bien distinctes[1] [38] :

- le niveau spontané qui permet la manifestation des réactions instinctives (douleur, joie, émotion, état psycho-physiologique, *etc.*),
- le niveau des formes prosodiques expressives employées de façon intentionnelle,
- et le niveau référentiel où les unités prosodiques assument des fonctions morphologiques, syntaxiques et informatives spécifiques à chaque langue.

On sent intuitivement, que la prosodie ne peut être définie de manière rigoureuse que par une spécification des éléments qui la composent ainsi qu'un ensemble de règles permettant de décrire leurs interactions possibles avec les structures des différents niveaux linguistiques étudiés. On évalue aisément l'ampleur de la tâche aussi existe-t-il deux conceptions traditionnelles de la prosodie :

- On la définit généralement comme l'ensemble des phénomènes suprasegmentaux de la parole c'est-à-dire relevant de domaines plus larges que l'unité linguistique distinctive minimale — le phonème — parmi lesquels la syllabe, le mot, les clauses, les phrases, les paragraphes, *etc.* Lehiste ajoutait à juste titre que les traits suprasegmentaux sont établis par comparaison d'items en séquence, alors que les traits segmentaux sont identifiables par la seule étude du segment lui-même [83].
- Une autre conception répandue est de considérer la prosodie comme l'étude des corrélats phonétiques de fréquence fondamentale, de durée et d'intensité.

La nature du travail qui est présenté ici peut fort bien s'accommoder de ces "définitions". Il faut cependant reconnaître qu'un recours à une présentation plus formalisée de la prosodie nous a été salutaire à plus d'un titre ; aussi allons-nous rappeler brièvement quelques points développés par Di Cristo [39] qui ont fortement influencé notre conception de la prosodie.
Si l'on considère le rythme comme relevant de l'étude distributionnelle des accents, et que l'on fait abstraction des langues dites tonales (pour la majorité asiatiques et africaines), alors l'accentuation et l'intonation sont les deux principaux éléments (ou structures) prosodiques.
La figure 1.1 précise les typologies et les fonctions associées des différents accents en référence à diverses terminologies couramment employées. Les fonctions de l'accent dépendent de l'appartenance de la langue considérée à l'un des deux groupes suivants :

[1]La première catégorie qui ne relève pas du domaine de la langue, ne fera l'objet d'aucune investigation dans la suite de cet exposé.

les langues à accent fixe où la place de l'accent est prévisible (comme le français, le tchèque, *etc.*) et les langues à accent libre (comme l'anglais, l'italien, l'allemand, *etc.*) où la place de l'accent — *a priori* imprévisible — est déterminée à partir de critères morphologiques et/ou sémantiques. Pour la langue française — seule considérée dans notre travail — l'accent non-emphatique assume une fonction démarcative (l'accent tombe sur la dernière syllabe du syntagme ou l'avant-dernière dans le cas d'un [ə] terminal), et génératrice d'intonèmes (un intonème étant un fait intonatif à valeur fonctionnelle). L'accent emphatique qui se distingue par son caractère facultatif remplit également une fonction démarcative (en affectant généralement la première syllabe des mots) et peut accessoirement avoir une fonction contrastive. Ce type d'accent ne faisant pas l'objet d'investigations poussées dans la suite de cet exposé, nous emploierons dès maintenant le terme d'accent pour désigner l'accent non-emphatique. La substance acoustique et perceptive de l'accent est à caractère pluriparamétrique (la hiérarchie des divers paramètres étant dépendante de la langue considérée [103]). Ainsi comme le précise Rossi [134] avec des notations légèrement différentes pour une conception similaire, l'*accent lexical* ou interne qui a pour domaines la syllabe et l'unité accentuelle est caractérisé principalement en français par l'indice de durée [47] ; l'*ictus mélodique* (ou accent secondaire pour Di Cristo) a pour domaine la syllabe, il est caractérisé par une proéminence tonale et assume des fonctions essentiellement rythmiques ; et l'*accent de focalisation* (ou accent intellectuel) dont les domaines sont la syllabe et le mot, il se caractérise par des proéminences du fondamental et de l'énergie.

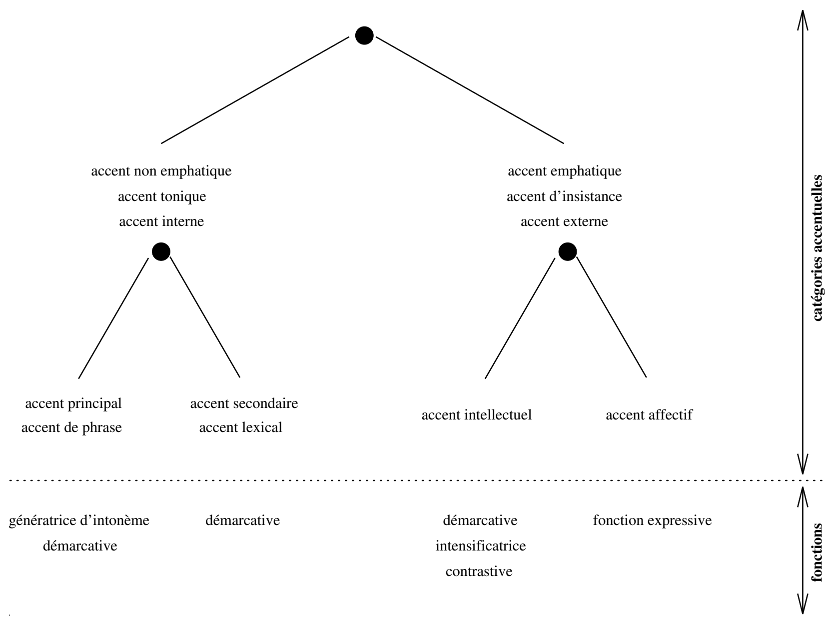

Figure 1.1: Catégories et fonctions accentuelles du français selon Di Cristo. Les dénominations catégorielles attachées à un même nœud sont équivalentes et reprennent les principales terminologies.

Le rôle de l'intonation est d'expliquer le comportement d'un locuteur en fonction d'une situation et des personnes à qui s'adresse le message. Les fonctions de l'intonation recensées par Di Cristo [39] sont au nombre de douze, témoignant — si besoin était — du rôle important de la prosodie et principalement de l'intonation dans la communication linguistique.

Fonction intégrative : l'intonation permet d'accorder à une suite non grammaticalement correcte (ex: Max travaille d'une façon[2]) le statut d'énoncé intono-syntaxiquement bien formé.

Fonction modale : de nombreuses études [74, 111] et [126, pp. 149–177] ont mis en relief l'importance de la configuration intonative pour la détermination de la modalité de la phrase (affirmative, impérative, interrogative, *etc.*). La figure 1.3 présente la courbe du fondamental de trois réalisations de la même phrase avec des modalités différentes.

Fonction informative : la participation de l'intonation à la structuration du message (sémantique) a été abondamment étudiée, notamment pour l'identification du thème (fonction *thématique*) et le rhème (fonction rhématique) [120, 93, 126, 133], mais également dans le cas d'une emphase non expressive ou à valeur fonctionnelle comme dans l'exemple : "*C'est* celle-ci *que je veux*" (fonction *contrastive*). La figure 1.4 propose un exemple de cette fonction.

Fonction identificatrice : l'intonation participe également à la structuration de l'énoncé en délimitant les frontières de certains constituants syntaxiques [129, 120, 93, 31, 16, 21] (voir figure 1.5).

Fonction hiérarchisante : certains auteurs s'accordent à dire que la structure intonative a pour rôle de proposer une hiérarchie des constituants syntaxiques [126, pp. 223,233], [39].

Fonction de désambiguïsation : cette fonction illustrée par l'exemple classique : *la belle ferme le voile*[3] est d'autant plus intéressante qu'elle est souvent la seule à pouvoir trancher sur la structure profonde de la phrase [68, 113, 123, 167]. Les figures 1.6 et 1.7 proposent des exemples de réalisations de phrases ambiguës.

Fonction attitudinelle : l'intonation véhicule des informations sur l'attitude (doute, surprise, *etc.*) du locuteur envers le message énoncé.

Fonction quantificatrice : qui définit une notion d'intensité des différentes modalités expressives.

Fonction d'emphatisation : l'intonation permet finalement au locuteur de mettre en relief une portion d'un énoncé (voir une illustration en figure 1.2).

---

[2]Exemple proposé par Gross [60].
[3]Exemple proposé par Malmberg [88].

Comme pour l'accentuation, Rossi [134] rappelle l'aspect pluriparamétrique de l'intonation et spécifie que le domaine de l'intonation est la phrase et ses constituants. La table 1.1 résume les intonèmes (morphèmes de l'intonation) qu'il distingue (et que l'on retrouve habituellement dans la littérature) ainsi que leur caractérisation paramétrique.

intonèmes	Désignation	Fonction	*f0*	durée	loudness
CT	continuatif majeur	hiérarchisation pragmatique et syntaxique	+	+	+
ct	continuatif mineur	démarcatif	+	+	
CA	vocatif non terminal	marqueur du topique			
PAR	parenthétique	marqueur d'apposition	stable		
CC	conclusif majeur	marqueur terminal et rhématique	−	+	−
cc	conclusif mineur	marque la fin d'une parenthèse droite ; marqueur de disjonction entre les syntagmes internes d'une parenthèse			

Table 1.1: Intonèmes répertoriés par Rossi avec leurs caractérisations paramétriques.

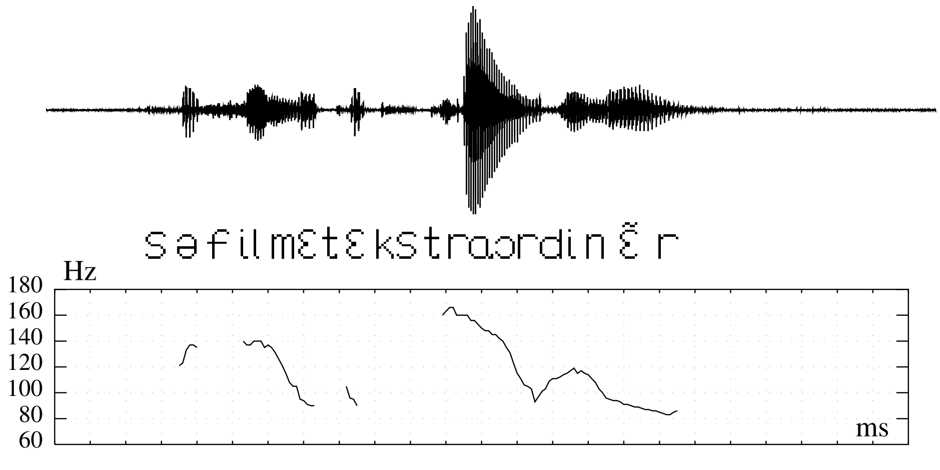

Figure 1.2: Illustration de la fonction d'emphase assurée par la prosodie pour une réalisation de la phrase : "Ce film est ex**tra**ordinaire" : l'emphase est ici marquée par des valeurs maximales de l'intensité et de la fréquence fondamentale sur la voyelle [*a*] du mot *extraordinaire*.