

MICRO * : UN SYSTEME MULTI-AGENTS POUR LA COMPREHENSION DE LA PAROLE

Bertrand CAILLAUD, Jean-Yves ANTOINE, Jean CAELEN, Geneviève CAELEN-HAUMONT

Institut de la Communication Parlée
46, avenue Félix Viallet
38031 GRENOBLE Cedex1 - FRANCE

Tel: +33-76.57.48.50
Fax: +33-76.57.47.10

caillaud@icp.grenet.fr, antoine@icp.grenet.fr, jcaelen@icp.grenet.fr, gcaelen@icp.grenet.fr

Résumé

Cet article propose une architecture de système de compréhension de la parole inspirée par des modèles cognitifs permettant d'approcher les capacités d'adaptation de l'être humain. Les résultats acquis en neuropsychologie et psycholinguistique nous ont conduit à doter notre système d'une architecture distribuée constituée d'un réseau d'agents. De plus, compte tenu de la latéralisation des processus cérébraux dans les activités langagières, nous avons choisi d'ajouter au traitement analytique de la parole une seconde voie rapide reposant sur l'analyse prosodique du signal et dont la finalité est de fournir des points d'ancrages pour les processus analytiques. Après avoir sommairement rappelé les théories cognitives ayant inspiré notre recherche, nous décrivons l'architecture globale du système, pour enfin nous attarder sur deux groupes d'agents particuliers : les groupes de bas niveau d'abstraction des voies analytiques et holistiques.

Mots-clés : système cognitif, compréhension de la parole, Intelligence Artificielle distribuée, système multi-agents, analyse holistique.

Abstract

This paper focuses on the description of MICRO, a speech understanding system largely inspired by cognitive models. The main aim of this project is to improve the adaptative abilities of the system by means of this cognitive modelization. Following cognitive modularist theories, MICRO was thus given a multi-agent structure. Furthermore, we took into account the functional difference between analytic and holistic processing that has been characterized in human cognition. Consequently, a fast prosodic way of analysis, devoted to the definition of anchorage points, was added to the classical analytic strategy (from acoustic-phonetic decoding to linguistic analysis).

We first present a brief overview of cognitive theories that represent the background of the project. The global structure of MICRO system is then described. Two particular groups of agents, corresponding to the low-level processings of both analytic and holistic analyses are finally described.

Keywords : speech understanding, cognition, distributed artificial intelligence, multi-agents system, holistic analysis.

1. INTRODUCTION

La finalité du projet MICRO consiste à mettre en œuvre un système de compréhension automatique de la parole fondé sur les avancées récentes en sciences cognitives. Notons que la plupart des systèmes actuels sont construits sur un principe hiérarchique découpant les fonctions de classification, de filtrage et d'interprétation, en blocs indépendants. Ainsi en est-il des systèmes à bases de connaissances hétérogènes mais aussi des systèmes statistiques ou des systèmes auto-organiseurs. Il est bon de constater qu'en dépit de leurs bonnes performances d'ensemble, ces systèmes restent incapables d'affronter les situations inattendues ou de s'adapter à des conditions ou locuteurs nouveaux, facultés pour lesquelles l'être humain est particulièrement efficace. Il peut paraître raisonnable de ce point de vue de s'inspirer de théories récentes en sciences cognitives pour définir l'architecture d'un système ayant des capacités identiques. L'hypothèse modulariste fournit un cadre possible de simulation sous forme distribuée à l'aide d'agents "réactifs" et "cognitifs" bien connus en IAD (Intelligence Artificielle Distribuée). Dans une telle architecture multi-agents, la maîtrise des processus de contrôle revêt une importance primordiale. Les neurosciences et la psychologie cognitive n'apportant que peu d'indications sur un tel contrôle, nous avons tenté de définir des stratégies de contrôle par une première phase de simulation par un expert humain.

Après avoir sommairement rappelé les théories cognitives ayant inspiré notre recherche, nous décrivons l'architecture globale du système, le propos de cet article étant centré sur les apports de la simulation pour le contrôle. C'est pourquoi nous la présenterons en détail ainsi que son analyse et les conclusions qui en découlent.

2. APPROCHE COGNITIVE

A partir de l'étude des avancées récentes en sciences cognitives, trois grandes caractéristiques de la cognition ont retenu notre attention: la modularité, l'interactivité et la co-existence de deux voies d'analyses (analytique et globale).

2.1. MODULARITE

De nombreux travaux en psychologie cognitive présentent le cerveau comme un ensemble de modules coopérant entre eux tout en opérant

chacun dans leur propre domaine de compétence [Minsky 85][Fodor 83][Marr 82]. De même, la neuropsychologie progressant vers une description de plus en plus précise de l'anatomie fonctionnelle du cerveau humain accrédite elle aussi la théorie modulariste tout en révisant certains aspects. Actuellement, de nombreux chercheurs présentent sous forme distribuée la structure fonctionnelle des activités cognitives impliquées dans la vision [Kosslyn 80] ou la parole [Morton 84][Altman 90:12-14]. La question de l'indépendance de ces modules n'est pas tranchée. Par exemple, Fodor et Minsky s'opposent sur ce point. Néanmoins, l'expérience acquise à travers le système à tableau noir DIRA [Caelen 90] a montré les limites adaptatives d'un modèle simplement modulaire où les experts n'ont pas la maîtrise de leur contrôle. En conséquence, MICRO est doté d'une structure distribuée s'appuyant sur des techniques multi-agents.

2.2. INTERACTIVITE

La compréhension du langage naturel par l'humain semble être un processus hautement intégré [Mehler 90] dans lequel il semble bien difficile de démêler des niveaux de traitement, comme par exemple sémantique et syntaxe qui s'enchevêtrent intimement. Il est clair que stratégie ascendante et stratégie descendante sont des expressions vides de sens en cognition humaine. Ainsi, la psycholinguistique nous propose actuellement des modèles de reconnaissance de la parole essentiellement interactifs [Mc Clelland et Elman 86][Marshall-Wilson 87]. S'il est nécessaire en simulation par ordinateur de définir des hiérarchies qualitatives de traitements, cela nous entraîne à supposer un processus de diffusion des informations et des mécanismes de parallélisme pour les activités de computation des agents. Nous définissons donc sous le terme d'*interactivité* la production et la prise en compte simultanées des informations ascendantes (stratégie *bottom-up*) et descendantes (stratégie *top-down*). L'interactivité, vue comme une conséquence de la modularité, implique que tous les agents peuvent opérer en parallèle en échangeant leurs informations par des processus diffusants.

2.3. DEUX VOIES D'ANALYSE

Sperry a montré dès les années soixante une différenciation fonctionnelle entre l'hémisphère droit du cerveau, chargé des processus cognitifs

non verbaux, et l'hémisphère gauche siège de l'analyse du langage [Sperry 61]. La réalité est certes moins manichéiste, et la controverse sur cette latéralisation est loin d'être close. Néanmoins, les travaux en neuropsychologie récents s'accordent sur un point: la co-existence d'une voie d'analyse détaillée (analytique) et d'une voie de traitement plus global (holistique), sans présumer de leur localisation physiologique [Gardner 81]. Ainsi la parole dans ses aspects verbal et acoustique subirait de tels traitements différenciés. Ayant par ailleurs observé empiriquement la nécessité d'une double analyse dans le système DIRA — pour éviter les "retours-arrière" induits par un raisonnement séquentiel en profondeur d'abord et incompatibles avec la gestion temps-réel du signal parole — nous proposons de prendre en compte ces deux voies de la manière suivante :

- une voie analytique parcourue par les processus de décodage acoustico-phonétique et linguistique,
- une voie de pré-marquage global reposant sur l'analyse de la prosodie du signal de parole, pour orienter les moments de l'analyse fine effectuée par la première voie.

3. DESCRIPTION DE *MICRO*

Le système *MICRO* est développé dans un environnement générique dédié à la conception de systèmes experts de type multi-agents [Baujard 90] : il s'agit de l'environnement MAPS (Multi-Agent Problem Solver). L'architecture de *MICRO* ne peut se comprendre sans une description sommaire de cet environnement.

3.1. MAPS

La conception de MAPS [Baujard 90] repose sur la distinction entre deux classes de connaissances :

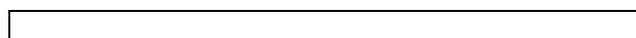
- *les connaissances descriptives* sur les éléments du problème (données, hypothèses, résultats),
- *les connaissances opératoires et décisionnelles* (règles et meta-règles), définissant des opérations et un mécanisme de raisonnement sur les éléments du problème.

MAPS associe un type d'agent à chaque classe de connaissance, respectivement les KS (Knowledge Server) et les KP (Knowledge Processor). Les agents communiquent par envoi de messages et réagissent ainsi à des événements extérieurs. Les

agents KS communiquent usuellement avec des agents KP (soumission de données et problèmes), et réciproquement les agents KP avec des agents KS (transmission et demande d'informations). MAPS permet désormais une exécution parallèle des processus attachés aux agents; deux modes de synchronisation (synchrone et asynchrone) entre les agents ont été introduits pour permettre la définition de différents comportements (agents indépendants, coordonnés ou concurrents). Les communications en mode synchrone correspondent à des relations client / serveur, tandis qu'en mode asynchrone, les communications sont de type *boîte aux lettres*. Enfin, MAPS permet désormais la création d'architectures multi-couches par l'introduction de meta-agents, chargés du contrôle respectif des tâches et des connaissances.

3.2 ARCHITECTURE GLOBALE DU SYSTEME *MICRO*

Le système *MICRO*, décrit en figure 1, constitue un réseau d'agents MAPS (voir [Caelen 92] pour une description détaillée). On remarque la structure en deux couches du système, où des groupes d'agents correspondant à des niveaux de traitement de tâches sont supervisés par des agents (Meta-KP) chargés du contrôle des tâches. On distingue les deux voies de traitement holistique et analytique symboliquement placées à droite et à gauche. Le système comprend en outre des groupes d'agents simulant les entrées (modèle du système auditif) et les sorties (modèle d'interface de communication avec les autres "aires" de traitement, notamment la pragmatique). Les agents sont rassemblés en groupes correspondant aux deux niveaux d'articulation de la langue : linguistique d'une part, acoustico-phonétique d'autre part.



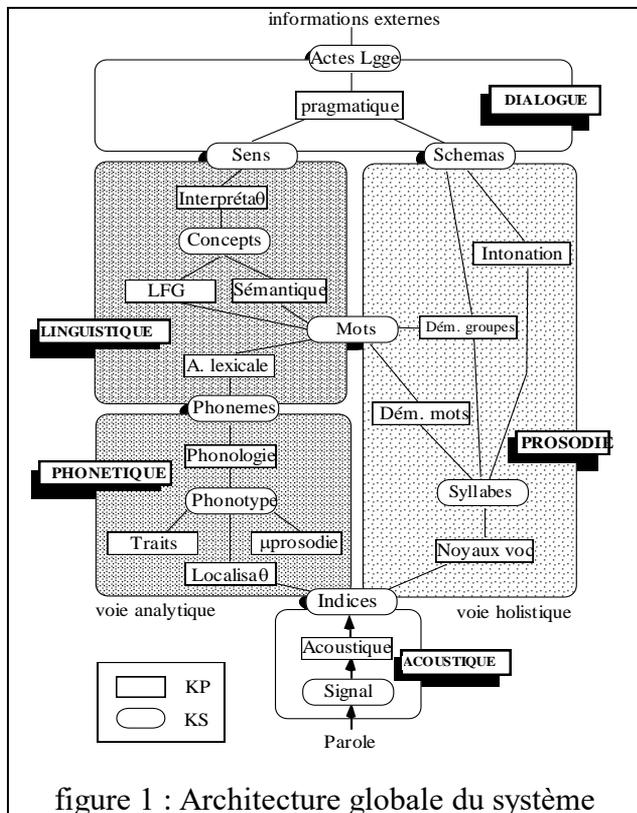


figure 1 : Architecture globale du système

Groupe perception

Ce groupe correspond aux traitements du système auditif humain. Il est simulé à l'aide d'un modèle mathématique d'oreille [Caelen 79] comportant plusieurs étages de traitement.

Groupe acoustico-phonétique (analytique)

Ce groupe rassemble les traitements bas-niveau de la voie analytique. Les agents qui le composent doivent produire des hypothèses phonétiques à partir d'une représentation en traits de la parole (fournie par les agents du groupe perceptif) et des contraintes phonétiques fournies par le groupe linguistique. Ce groupe d'agents sera étudié plus complètement au paragraphe 5.2.

Groupe linguistique (analytique)

Ce groupe d'agents à la charge de produire la représentation conceptuelle de la phrase, en intégrant les hypothèses phonétiques du groupe acoustico-phonétique, les contraintes de haut-niveau produites par le groupe dialogue et les informations prosodiques fournies par la voie holistique.

L'analyse linguistique est mise en oeuvre à l'aide d'un mécanisme de coopération entre les agents syntaxique (grammaire LFG) et sémantique qui travaillent concurremment [Antoine 93]. Comme le montre la figure 2, cette coopération est réalisée par l'intermédiaire d'une fusion de leur productions concurrentes, qu'elles soient ascendantes (structures hypothétiques de la phrase) ou descendantes (contraintes sur l'analyse lexicale).

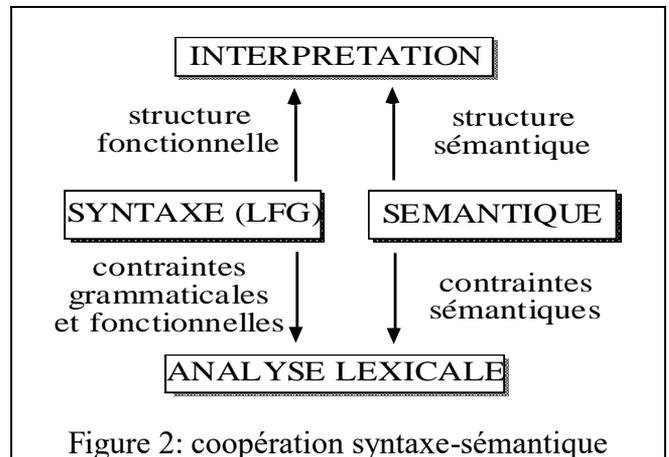


Figure 2: coopération syntaxe-sémantique

Cette équivalence des analyses syntaxique et sémantique est attestée à l'heure actuelle par de nombreux travaux présentant l'analyse linguistique comme un processus interactif et intégré [Rastier 87,p. 12]. Ils rompent ainsi avec la tripartition classique plaidant pour une succession ascendante des traitements syntaxique, sémantique et finalement pragmatique [Morris 46].

Groupe prosodie (holistique)

Ce groupe constitue l'ensemble de la voie holistique. Il effectue des traitements rapides basés sur l'analyse prosodique du signal de parole, afin de fournir des informations (démarcation de groupes, intonation, etc.) qui serviront de points d'ancrages aux groupes linguistique et pragmatique pour leurs analyses combinatoires. Nous décrirons ultérieurement ce groupe plus en détail.

Groupe dialogue

Ce groupe d'agents confronte les structures de reconnaissance produites par les niveaux inférieurs avec les informations non-auditives. Il répercute également les contraintes pragmatiques vers les niveaux linguistiques. Ainsi, il constitue essentiellement l'interface entre le système de compréhension et les niveaux pragmatiques.

3.3 UNIFICATION

Les agents KP de chaque groupe construisent, dans leur domaine de compétence, des solutions partielles qui sont regroupées dans un (ou plusieurs) agents KS. Le problème consiste donc à faire *émerger* les solutions globales du groupe par confrontation de ses diverses solutions partielles puis à les diffuser vers les autres groupes qui ont fait de même pendant le même temps. On peut alors aisément imaginer un processus global de stabilisation dont l'état émergent produirait la solution finale. La question fondamentale reste néanmoins de savoir si la convergence d'un tel système doit être contrôlée ou non. Dans le cas d'une réponse négative il ne serait guère possible de simuler un tel processus bouclé sur une machine séquentielle et avec des agents qui échangent des messages de nature symbolique. C'est pourquoi, dans un premier temps nous avons opté pour une simulation avec un contrôle en deux passes :

- un processus d'unification dans chaque groupe (émergence des solutions partielles),
- la diffusion, avec un léger retard temporel, de ses solutions, de ces solutions partielles aux autres groupes permet à ces derniers d'intégrer cette nouvelle donnée comme contrainte sur leur solution partielle pour le prochain segment de parole.

Ainsi, c'est la prise en compte en parallèle d'informations élaborées par divers processus de nature différente qui autorise un comportement réellement adaptatif [Lautrey 90], faculté primordiale en compréhension de la parole. En effet, l'existence de points de vue différents est essentielle pour la remise en cause d'hypothèses éventuellement erronées.

Nous allons maintenant décrire la réalisation des groupes prosodique et acoustico-phonétique.

4. GROUPE PROSODIQUE

Le groupe d'agents prosodie représenté figure 3 est actuellement en cours d'élaboration en

particulier en ce qui concerne les traitements de bas niveau. L'objectif global de ce groupe d'agents est de fournir rapidement des informations qui doivent permettre de synchroniser les traitements du groupe linguistique sur des îlots de confiance (KP démarcation de mots et de groupe) et de fournir des informations pour les actes de dialogue (KP intonation). Toutes ces informations sont calculées sur les noyaux vocaliques (en général centres des voyelles) avant d'être catégorisées.

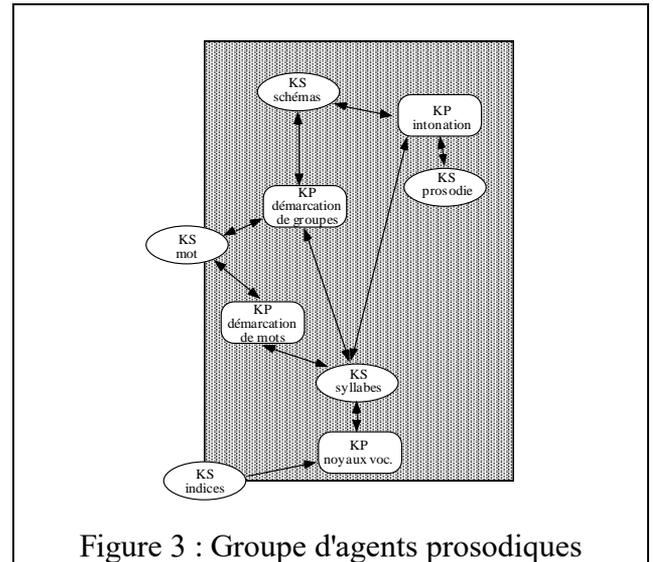


Figure 3 : Groupe d'agents prosodiques

4.1. KP NOYAUX VOCALIQUES.

4.1.1. Description

Cet agent doit d'une part détecter les noyaux vocaliques à partir des courbes d'énergie et de fréquence fondamentale fournies par les agents de bas niveau, et d'autre part mettre à la disposition des agents de plus haut niveau un ensemble de paramètres utiles pour le repérage des groupes prosodiques. Ce sont :

- l'énergie moyenne sur chaque noyau,
- la fréquence fondamentale moyenne,
- la durée de chaque noyau,
- l'accélération du débit d'un noyau au suivant,
- la durée de pause éventuelle précédant le noyau.

La détection des noyaux vocaliques est un traitement clef puisque tous les autres agents s'appuient sur cette information pour fournir la leur. Des contraintes de rapidité de calcul sont également imposées pour que le rôle de marquage dévolu à la voie holistique dont font partie ces agents, soit correctement rempli avant l'analyse fine. Afin d'obtenir un traitement suffisamment rapide, la détection des noyaux vocaliques se fait uniquement par une recherche

de formes spécifiques (pics et plateaux) sur la courbe d'énergie, ce paramètre étant l'un des premiers fournis par les agents de bas niveau.

Ces contraintes n'empêchent néanmoins pas une bonne robustesse, comme le montre l'étude présentée au paragraphe suivant..

4.1.2. Robustesse de détection

Deux analyses de robustesse ont été effectuées. La première a été obtenue sur un corpus lu extrait de BDSON [Serignat 92]. Le tableau suivant montre que dans ce cas, la robustesse est très satisfaisante:

Nombre de noyaux non détectés	Nombre de noyaux détectés mais faux	Nombre de noyaux détectés mais doubles	Nombre total de noyaux
11	25	58	944
1.1 %	2.6 %	6.1 %	

Notons que dans beaucoup de cas les noyaux doubles (2 voyelles dans le même noyau détecté) ne pénalisent pas les traitements ultérieurs.

La seconde analyse a été effectuée sur un corpus de dialogue enregistré dans un environnement bruyé. La encore, les résultats sont satisfaisants:

Nombre de noyaux non détectés	Nombre de noyaux détectés mais faux	Nombre de noyaux détectés mais doubles	Nombre total de noyaux
3	8	10	192
1,6 %	4,2 %	5.2 %	

On peut donc en conclure à une bonne robustesse d'ensemble de l'agent responsable de la détection des noyaux vocaliques.

4.2. KP DEMARCATIION DE MOTS

Cet agent est chargé de placer des marqueurs de début et fin de mots ; s'il le peut, il doit aussi qualifier les mots marqués (en termes de mot lexical ou mot gramatical). Cette tâche de démarcation est réalisée par l'application d'un ensemble de règles prosodiques issues des travaux réalisés dans ce domaine par [Nasri 89] à la suite de l'expertise fournie par [Caelen-Haumont 93].

Exemple de règle :

SI $Fo(n) = MAX_Fo$
 ET $D(n-2) \leq D(n-1) < D(n)$
 ET $D(n) \geq (D(n-1) \times 1.5)$
 ET $D(n) = 4$

ALORS FML(n)

où $Fo(n)$ = valeur de la fréquence fondamentale quantifiée en niveaux sur le noyaux vocalique n° ,
 et $D(n)$ = durée quantifiée du noyau vocalique n° . FML(n) marque de Fin de Mot Lexical posée sur le noyau n° .

Nous travaillons actuellement sur un affinement de ces règles par l'étude d'une expertise faite par un expert humain, en particulier pour les adapter à un environnement de dialogue. Ces règles font aussi l'objet d'un projet de test et de validation sur des corpus importants (SIDOC [Munteanu 92]).

4.3. KP DEMARCATIION DE GROUPES

Cet agent doit déterminer les début et fin de groupe syntaxique à partir des mots démarqués par le KP démarcation de mots et des informations fournies par le KP noyaux vocaliques. Il doit aussi déterminer le type des groupes repérés : groupe majeur ou mineur dans la phrase. Il utilise des règles similaires à celle du KP démarcation de mots et qui font l'objet de la même étude. Il n'est pas complètement mis en œuvre à l'heure actuelle.

5. SIMULATION DU GROUPE ACOUSTICO-PHONETIQUE

5.1. PROBLEMATIQUE

Dans une architecture multi-agents, la maîtrise des processus de contrôle est de la première importance. Ce contrôle apparaît nécessaire à la fois pour gouverner les stratégies d'adaptation contextuelle des agents et pour réguler les stratégies locales d'unification. Pour caractériser ce contrôle, nous devons améliorer nos connaissances sur l'adaptation des traitements des agents ainsi que sur les stratégies d'unification. Pour ce faire, nous avons choisi de simuler le travail des agents du groupe acoustico-phonétique par une série d'expertises humaines. L'expert dispose pour cela d'interfaces représentant l'état de chaque agent KP.

5.2. DESCRIPTION DU GROUPE ACOUSTICO-PHONÉTIQUE SIMULÉ

L'expérience acquise avec le système DIRA [Caelen 90] a permis une détermination aisée de la structure et des algorithmes de reconnaissance de ce groupe, même si la prise en compte de certains phénomènes contextuels reste imparfaite (figure 4). Nous avons adjoint un agent spécialisé dans l'unification des solutions partielles opérant sur une sorte de mémoire commune locale.

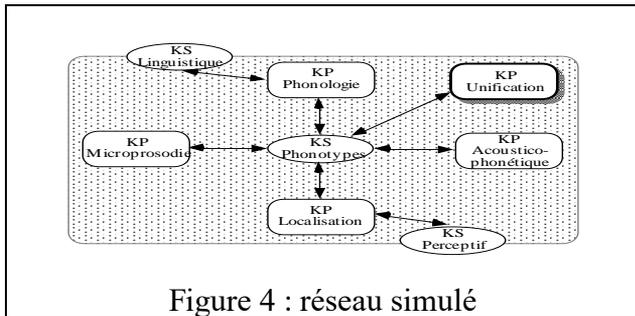


Figure 4 : réseau simulé

Deux autres mémoires communes servent de relais avec les deux autres groupes voisins (non représentés sur la figure). Il ne restait alors plus qu'à simuler diverses stratégies d'unification dans le but de retenir la meilleure.

Nous allons maintenant décrire les rôles des différents agents composant le groupe.

KP Localisation

Cet agent détermine les localisations temporelles des phonotypes¹ du signal à partir d'une sur-segmentation du signal de parole effectuée par le groupe perception (les agents du groupe acoustico-phonétique travaillent tous sur ces phonotypes).

KP Acoustico-phonétique

Cet agent classe les phonotypes à l'aide des traits acoustico-phonétiques suivants :

- *Macro-classe* : trait de catégorisation acoustico-phonétique large. Les macro-classes sont : occlusif, fricatif, vocalique, consonantique, silence,
- *Traits* : l'expert utilise quatre traits de caractérisation. Ils apportent des informations acoustiques complémentaires aux macro-classes, permettant une restriction éventuelle de l'espace de recherche. Ce sont aigu-grave, fermé-ouvert, doux-strident, bémolisé-diésé, écarté-compact.

¹phonotype: c'est la réalisation acoustique d'un phonème.

KP Microprosodie

Cet agent caractérise les phonotypes à l'aide des traits micro-prosodiques suivants :

- *Voisement* : trait de catégorisation binaire (voisé / non voisé) sur le voisement du phonotype.
- *Energie* : trait de catégorisation micro-prosodique décrivant la forme de la courbe d'énergie. Il est analogue, sur son domaine de compétence, au trait "Macro-classe",
- *Durée intrinsèque* : trait de caractérisation qualifiant la durée du phonotype par rapport à une durée étalon (moyenne sur plusieurs locuteurs).

KP Phonologie

Contrairement aux KP précédents, cet agent a une production descendante vis à vis du groupe: il produit des hypothèses phonotypiques à partir de la description phonémique des mots candidats fournis par le groupe linguistique. L'unification, réalisée par un KP spécifique, s'opèrera entre ses hypothèses partielles et celles des autres agents fonctionnant plutôt en production ascendante.

KS Phonotypes

Il mémorise l'ensemble des informations produites par les agents KP du groupe acoustico-phonétique.

KS Perception et Acoustico-Phonétique

Ce deux agents simulent respectivement les informations venant du groupe perception et du groupe linguistique.

KP Unification

Cet agent est responsable de la fusion des hypothèses ascendantes et descendantes proposées par les différents KP du groupe. Les agents Acoustico-Phonétique et Microprosodie travaillent sur des domaines de compétence disjoints. Par contre, les hypothèses phonémiques produites par le KP Phonologie englobent ces deux domaines (voir tableau). Le processus d'unification doit donc posséder des mécanismes d'union et d'intersection d'hypothèses

Hypothèses ascendantes		Hypothèses descendantes	
KP Acoustico-Phonétique	Macro-classe	Macro-classe	KP Phonologie
	Traits 1 ... 4	Traits 1 ... 4	
KP Micro Prosodie	Voisement	Voisement	
	Energie	Energie	
	Durée intrinsèque	Durée intrinsèque	

Traits calculés sur le signal	Traits typiques aux phonèmes
-------------------------------------	------------------------------------

5.3. INTERFACES DE SIMULATION

Les interfaces de simulation présentent les informations et les outils utiles à l'expert pour simuler les agents KP. Chaque interface présente:

- une zone d'information regroupant l'ensemble des informations fournies par les KS voisins,
- une zone de décision dans laquelle l'expert humain simule les traitements du KP,
- une zone d'hypothèses produites qui regroupe les hypothèses résultant de ses décisions.

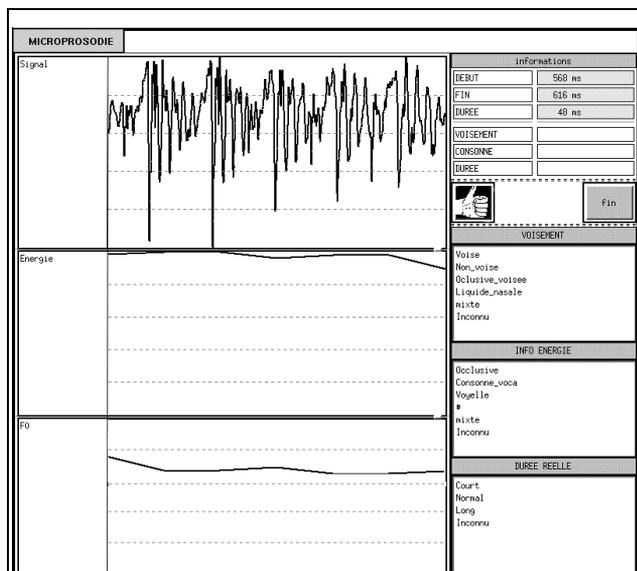


figure 5 : interface du KP Microprosodie

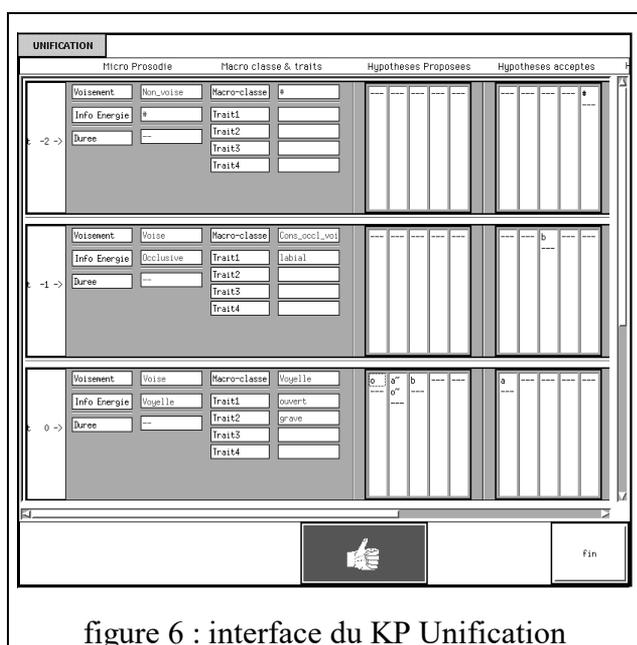


figure 6 : interface du KP Unification

5.4. RESULTATS D'EXPERTISE

Dans un premier temps, l'expert a travaillé sans contraintes. Nous voulions dégager une expertise sur le fonctionnement des agents KP ainsi que sur la stratégie d'unification la plus naturelle. Rien n'assure cependant que la meilleure stratégie pour un système automatique corresponde à celle employée par l'expert humain. C'est pourquoi nous avons ensuite imposé à l'expert diverses stratégies d'unification pour comparer leurs performances. L'expertise a porté sur un corpus extrait de BDSONS [Serignat 92].

5.4.1. Expertise libre

Utilisation et robustesse des traits

Pour chaque trait, nous avons comptabilisé le nombre de fois où l'expert a pu (ou a estimé nécessaire de) qualifier ce dernier par un attribut :

Nom du trait ¹	Utilisation
Macro-classe	100 %
Trait 1	75,4 %
Trait 2	36,8 %
Trait 3	3,5 %
Trait 4	0,0 %
Voisement	96,5 %
Energie	86,0 %

On peut tirer les conclusions suivantes:

- Deux traits semblent suffire à l'expert pour sa prise de décision, les suivants n'étant pas toujours très robustes,
- Les traits "Macro-classe", "Voisement" et "Energie" sont des paramètres très utilisés car robustes. Pour l'expert, les traits "Voisement" et "Energie" n'apportent souvent qu'une confirmation (forte redondance) à l'information portée par "Macro-classe".

Confiance accordée aux traits

Pour caractériser la stratégie d'unification employée par l'expert humain, il était intéressant de s'intéresser au taux de confiance que ce dernier accorde à chaque trait au moment de l'unification des hypothèses. Pour cela, nous avons calculé trois taux de confiance:

- *Confiance forte (OK+)*: lors de l'unification, l'expert a conservé uniquement les hypothèses

¹ le trait "Durée intrinsèque" n'a pas été analysé, la trace ne fournissant pas une connaissance suffisante sur son utilisation. Un enregistrement audio de l'expert est nécessaire pour permettre une telle analyse.

produites par le KP phonologie qui étaient compatibles avec le trait considéré,

- *Confiance moyenne (OK-)*: en sus des hypothèses précédentes, l'expert a conservé celles qui ne contredisaient pas ce trait,
- *Confiance nulle (NO)*: l'expert a conservé au moins une hypothèse entrant en contradiction avec le trait considéré.

Enfin, on calcule un taux de confiance "OK Tot" correspondant aux cas de confiances forte et moyenne. Le tableau suivant présente, en pourcentages, les résultats d'analyse obtenus:

Nom du trait	OK+	OK-	OK Tot	NO
Macro-classe	78,9	3,5	82,4	17,6(1)
Trait 1	26,8	22,0	48,8	51,2
Trait 2	23,8	23,8	47,6	52,4
Voisement	92,7	-	92,7	7,3
Energie	83,7	-	83,7	16,3 (2)

(1) rencontré en cas de conflit avec le KP Phonologie, ou avec le trait "Energie" (2) conflit avec "Macro-Classe"

On remarque que:

- l'expert accorde une grande confiance aux traits ("Macro-classe", "Energie" et "Voisement"), à l'exception des cas où ces derniers se contredisent mutuellement. Si de plus, aucun candidat du KP Phonologie n'est compatible avec ces traits, toutes les hypothèses sont conservées. Dorénavant ces traits sont appelés *traits principaux*.
- l'expert accorde une confiance limitée aux autres traits sauf peut-être dans des situations prototypiques simples. Dorénavant ces traits sont appelés *traits secondaires*.

5.4.2. Stratégie d'unification de l'expert

L'analyse ci-dessus nous donne une idée assez fidèle de la stratégie d'unification employée par l'expert : généralement, celui-ci préfère adopter une stratégie prudente qui consiste à garder un maximum d'hypothèses, en éliminant progressivement les hypothèses les plus mauvaises ou redondantes. La stratégie générale d'unification correspond à un filtrage des hypothèses par les traits "Macro-classe", "Voisement" et "Energie". Dans de rares cas contextuellement caractérisés, l'expert a choisi une stratégie différente. Ce sont ces cas qui nous intéressent maintenant.

5.4.3. Expertise contrainte

Nous avons demandé à l'expert de reprendre ses expertises d'unification en lui imposant tour à tour diverses stratégies systématisées afin de les comparer. Trois stratégies ont ainsi été définies: en premier lieu, une stratégie normale, proche de celle employée par l'expert humain. En second lieu, deux stratégies stricte et large qui sont respectivement plus et moins contraignantes que la première stratégie.

Les trois stratégies se distinguent dans le contexte d'analyse le plus fréquent, où l'on rencontre un forte compatibilité entre les traits. On définit alors les trois stratégies comme suit:

- *stratégie normale* : l'expert garde une hypothèse compatible avec tous les traits principaux et deux traits secondaires,
- *stratégie large* : l'expert garde une hypothèse compatible avec tous les traits principaux,
- *stratégie stricte* : l'expert garde une hypothèse compatible avec tous les traits.

Les autres contextes d'analyse correspondent à des cas particuliers. Les stratégies sont alors identiques. En premier lieu, s'il y a incompatibilité entre les traits eux-mêmes, on détend les contraintes sur les traits secondaires jusqu'à atteindre un point de compatibilité faible.

Enfin, si aucune hypothèse phonémique n'est compatible avec les traits, on conserve alors toutes les hypothèses. Ainsi, la confiance ultime est-elle accordée en tout dernier recours aux hypothèses descendantes. Malheureusement, il se peut que celles-ci peuvent avoir été auparavant orientées dans une mauvaise direction par une production ascendante défectueuse.

Analyse qualitative des stratégies

Nous avons caractérisé la qualité des trois stratégies à l'aide de deux critères: la robustesse des hypothèses produites et l'explosion combinatoire générée.

- *Robustesse* : nous avons observé la robustesse des stratégies au regard du groupe. Celle-ci est caractérisée par la qualité des propositions au sortir du groupe, c'est-à-dire les hypothèses lexicales filtrées à partir des unités phonétiques d'unification¹. Le tableau suivant présente les taux de robustesse correspondant à chaque stratégie : celui-ci est bien entendu

¹ Dans le cas de la simulation, la qualité des hypothèses lexicales reflète totalement celle du groupe acoustico-phonétique, car le groupe linguistique ne fournit ici aucune contrainte sur celles-ci.

inversement proportionnel au caractère contraignant de chaque stratégie:

Stratégie	Taux de robustesse
Stricte	62,5 %
Normale	75,0 %
Large	87,5 %

- *Explosion combinatoire*: ce second critère a été évalué à partir du nombre d'hypothèses lexicales conservées à chaque fin d'analyse. Cet indice est d'importance puisqu'un grand nombre d'hypothèses nécessitera des efforts de désambiguïsation soutenus de la part du groupe linguistique. Il en résultera un accroissement des risques d'erreurs et de temps de calcul. Les résultats sont les suivants:

Stratégie	Production multi-hypothèses	Nb hypothèses finales (1)
Stricte	12,5 %	1,2
Normale	37,5 %	1,4
Large	62,5 %	1,9

(1) En moyenne. Le facteur de branchement à chacun des 3 premiers phonèmes des unités lexicales du corpus est de 5.

Là encore, on retrouve des résultats attendus: une stratégie large est source d'une plus grande explosion combinatoire...

- *Robustesse Pondérée*: les deux critères précédents qualifient totalement la stratégie mais sont antinomiques. Il nous a donc semblé nécessaire de les regrouper en un seul: il s'agit de la robustesse pondérée, pour laquelle on pondère le critère de robustesse par le nombre d'hypothèses finales produites. On obtient alors le résultat suivant:

Stratégie	Robustesse pondérée
Stricte	54,2 %
Normale	64,2 %
Large	63,1 %

Les écarts de performances se sont réduits, montrant par là qu'on ne peut préférer une stratégie à une autre dans toutes les situations. En outre, leurs performances globales restent insuffisantes (notons au passage que la stratégie stricte reste sensiblement la plus faible, ce qui explique

certainement les réticences de l'expert humain à l'employer en simulation libre).

5.4.4. Conclusion: stratégie adaptative

Il n'existe pas de stratégie unique ayant des performances acceptables. La seule solution semble être une stratégie variable, suivant le contexte, choisie parmi les trois stratégies précédentes qui tour à tour sont plus ou moins performantes. En examinant de près les cas où ces performances sont les meilleures, nous avons ainsi déterminé quatre ensembles de paramètres pour l'adaptation contextuelle de la stratégie. Ce sont :

- les paramètres d'environnement acoustique,
- la concordance informative entre les traits,
- le type de phonotype candidat,
- la trace du comportement des agents KP.

En jouant sur ces paramètres il est possible d'obtenir des critères de commande de la stratégie d'unification et des règles de méta-stratégie de groupe. Cette méta-stratégie règle le comportement des agents ou conditionne le processus d'unification. Nous allons maintenant l'esquisser.

Paramètres d'environnement acoustique

Les paramètres d'environnement acoustique sont le bruit d'enregistrement ainsi que la position de l'unité phonétique reconnue dans le mot (début parole, début mot, fin mot).

Ces deux paramètres sont pertinents en ce qui concerne la netteté supposée du phonotype. Dans le bruit comme en fin de mot ceux-ci sont moins nets donc plus difficiles à identifier. Le taux de confiance des traits obtenus sera plus faible et la décision plus prudente.

Concordance des traits

La concordance des traits est un effet renforçateur qui peut être utilisé pour contraindre plus strictement la stratégie d'unification (et à l'opposé la non-concordance pour l'élargir). Ainsi l'analyse statistique des expertises montre que l'accord des traits "Macro-classe" et "Energie" donne une robustesse de 96,9%.

Type du phonotype candidat

La robustesse des méthodes de décodage acoustico-phonétique varie fortement avec le type d'unité phonétique attendu. Typiquement, les fricatives sont très bien identifiées tandis que les consonnes liquides posent toujours problème. La méta-stratégie d'adaptation est donc :

Si X est un phonotype candidat et si X présente des traits de:

- *Fricative*: alors stratégie_stricte (le trait strident est très robuste dans ce cas).
- *Occlusive ou Voyelle*: alors stratégie_normale (confiance moyenne dans les traits).
- *Autre* alors stratégie_large

5.4.4. Observation du comportement des agents

A un niveau plus élevé de contrôle, on peut utiliser la trace du comportement de l'agent d'unification pour analyser a posteriori la qualité des traits obtenus par tel ou tel agent. Par exemple, si l'information de voisement contenue dans "Occlusive_Sonore" est souvent écartée en phase d'unification (choix de la contrainte "Occlusive"), on relâche la contrainte sur cet attribut. Une fois mis en oeuvre, ce processus conduira à doter chaque agent d'une capacité d'auto-jugement, caractéristique essentielle à la définition d'agents réellement cognitifs.

CONCLUSION

Nous avons présenté une approche cognitive pour un système multi-agents de compréhension automatique de la parole. L'architecture distribuée de *MICRO* permet une modélisation de l'hypothèse modulariste ainsi que la co-existence de deux types d'analyses cognitives (analytique et holistique). Ces caractéristiques permettent d'espérer une amélioration de la robustesse de l'analyse, mais posent de nombreux problèmes de contrôle. C'est la raison pour laquelle nous avons procédé à une simulation d'une partie du système. La principale conclusion de cette simulation réside dans la nécessité d'adopter une stratégie adaptative, orientée par le contexte d'exécution et la trace du fonctionnement des agents. Quatre grands critères ont été dégagés, qui déterminent cette stratégie.

BIBLIOGRAPHIE

J.Y. Antoine, B. Caillaud, J. Caelen, 1993, "Syntax-semantics cooperation in MICRO, a multi-agent speech understanding system", pp. 2139-2142, EUROSPEECH'93, Berlin.

- G.T.M. Alman, 1990, "Introduction" in G. Altman (Ed.) "Cognitive models of speech processing", MIT Press, Cambridge, Mass.
0. Baujard, C. Garbay, 1990, "A programming environment for distributed expert system design", Expert System Applications, *ExperSys*, pp. 27-32.
- J. Caelen, 1979, "Un modèle d'oreille. Analyse de la parole. Reconnaissance phonémique, Thèse d'Etat, Toulouse.
- J. Caelen et al., 1990, Architecture et fonctionnement du système DIRA. De l'acoustique aux niveaux linguistiques", *Revue de traitement du signal*, vol. 7, no 4, pp. 345-366.
- J. Caelen, C. Garbay, 1992, "Modularité pour la reconnaissance de la parole", ECCOS'92.
- G. Caelen-Haumont, 93, "Cognitive processes and prosodic encoding. Speaker's adaptation to discourse conditions", *Communication and cognition - Artificial Intelligence*, 10:4.
- J. Fodor, 1983, "The Modularity of Mind", MIT Press, Cambridge MA.
- R.W. Gardner, 81, "The analysis of unanalysed perceptions", in M.Kubovy, J.R. Pomerantz (eds.), "Perception and organisation", Lawrence Erlbaum, Hillsdale NJ, pp. 119-139.
- S.M. Kosslyn, 1980, "Image and Mind", Harvard University Press, Cambridge MA.
- J. Lautrey, 1990, "Esquisse d'un modèle pluraliste du développement cognitif", in Reuchlin et al (eds), "Cognition, l'individuel et l'universel", PUF, Paris.
- J.L. McClelland et J.L. Elman, 1986, "The TRACE model of speech perception", *Cognitive Psychology*, 18, pp. 1-86.
- D. Marr, 1982, "Vision: a Computational investigation into the human representation and processing of visual information", Freeman, San Fransisco.
- W.D.Marslen-Wilson, 87, "Functional parallelism in spoken word recognition", *Science*, 25, pp. 71-102.
- J. Mehler et E. Dupoux, 1990, "Naître humain", Odile Jacob, Paris.

- M. Minsky, 1985, "The society of Mind", Simon & Schuster, New-York.
- C. Morris, 1946, "Signs, language and behavior, Prentice Hall, New-York.
- J. Morton, 1984, "Brain-based and non-brain-based models of language", in D. Caplan, L. Lecours, A. Smith (eds), "Biological Perspectives in Language", MIT Press, Cambridge MA, pp. 40-64.
- P. Munteanu, 92, "Modification d'un questionnaire de connaissances sur la parole", Rapport de stage ingénieur de l'Institut polytechnique de Bucarest-Roumanie, ICP/INPG, Grenoble.
- M.K. Nasri, G. Caelen-Haumont et J. Caelen, 1989, "Using Prosodic Rules in Continuous Speech Recognition", Proc. of ICASSP-IEEE, Glasgow, Vol. 1, pp. 671-674.
- F. Rastier, 1987, "Sémantique et Intelligence Artificielle: Présentation", *Langages*, 87, pp. 5-19.
- J.F. Sérignat, J.M. Dolmazon, M. De Calmes, G. Pérennou, 1992, "BDSONS - BDLEX : Bases de données vocales et lexicales", Rencontres Recherches en Informatique - Industrie, Toulouse, PP 88-89.
- R.W. Sperry, 1961, "Cerebral Organization and Behavior", *Science*, 133, pp. 1749-1757.