UNIVERSITÉ DE FRANCHE-COMTÉ

ÉCOLE DOCTORALE «LANGAGES, ESPACES, TEMPS, SOCIÉTES»

Thèse en vue de l'obtention du titre de docteur en
Sciences du langage

Option : Traitement Automatique des Langues Naturelles

DÉCOMPOSITION ADAPTATIVE DU SIGNAL DE PAROLE APPLIQUÉE AU CAS DE L'ARABE STANDARD ET DIALECTAL

Présentée et soutenue publiquement par

Christian GUILLEMINOT

Le 19 décembre 2008

Sous la direction de Monsieur Henri Madec

Membres du jury :

Mohamed Embarki, MCF HDR, Université Paul Valéry, Montpellier III
Henri Madec, MCF HDR, université de Franche-Comté
Jean–Noël Pernin, Professeur, université de Franche-Comté
Mohamed Yéou, MCF HDR, université Chouaïb Doukkali d'El Jadida

UNIVERSITÉ DE FRANCHE-COMTÉ

ÉCOLE DOCTORALE «LANGAGES, ESPACES, TEMPS, SOCIÉTES»

Thèse en vue de l'obtention du titre de docteur en
Sciences du langage

Option : Traitement Automatique des Langues Naturelles

DÉCOMPOSITION ADAPTATIVE DU SIGNAL DE PAROLE APPLIQUÉE AU CAS DE L'ARABE STANDARD ET DIALECTAL

Présentée et soutenue publiquement par

Christian GUILLEMINOT

Le 19 décembre 2008

Sous la direction de Monsieur Henri Madec

Membres du jury :

Mohamed Embarki, MCF HDR, Université Paul Valéry, Montpellier III
Henri Madec, MCF HDR, université de Franche-Comté
Jean–Noël Pernin, Professeur, université de Franche-Comté
Mohamed Yéou, MCF HDR, université Chouaïb Doukkali d'El Jadida

Remerciements

Toute ma gratitude va à Monsieur Henri Madec dont le soutien, l'humour et la confiance sans faille m'ont permis de mener à bien ce travail.

Je tiens également à remercier chaleureusement,

Mohamed Embarki, pour son amitié, ses conseils et ses critiques sans concession ;

Madame le Professeur Sylviane Cardey, Directeur du Centre Tesnière pour son accueil, ses encouragements et son soutien qui furent déterminants ;

Monsieur le Professeur Peter Greenfield, pour son attention et ses sympatiques encouragement à aborder le langage C, ce que je me refusais alors et qui me fut très utile ensuite ;

Christelle Dodane dont l'amitié et le soutien tant moral que scientifique ne m'ont jamais fait défaut ;

Anwuli Echenim pour ses compétences et son aide à propos de MIXMOD ;

Fabien Brachère pour sa disponibilité et la réécriture d'un passage de son logiciel Guimauve ;

Rémy Gribonval et l'équipe MPTK qui sont toujours prêts à répondre à une question, à expliquer un concept ;

et toutes celles et tous ceux que je ne peux pas citer ici, en particulier les artisans du formidable mouvement du logiciel libre qui mettent leurs efforts, leurs découvertes et leurs œuvres, gracieusement au service de la communaté scientifique internationale.

À ma mère qui m'a si souvent répété
qu'il fallait travailler à l'école.
C'était il y a longtemps...
À mon fils Carlos et à mon épouse Tania
que j'ai trop délaissés durant ce travail
ainsi qu'à Lidia Ester dont l'accueil a toujours eu
la chaleur exubérante de Cuba.

Mots clefs et abréviations

– a Mots clefs

analogique	atome	atomique (décomposition)	classification
cluster	clustering	compression	constructale
coarticulation	complexité	dialecte	distance
formant	FFT	gaussienne	information
locus (équation de)	Matching Pursuit	pharyngalisée	non-pharyngalisée
numérique	signal	similarité	similitude

– b Abréviations

AD, arabe dialectal	ASC, arabe standard contemporain
V, voyelle	C consonne
V_1, première voyelle de VCV	V_2 deuxième voyelle de VCV
V_{2onset}, début de V_2	V_{2mid}, milieu de V_2
F1, premier formant	F2, F3, F4, formants 2, 3 et 4
H0 fréquence fondamentale	Hn, harmonique de rang n
TFD transformée de Fourier discrète	FFT, transformée de Fourier rapide
IA, intelligence artificielle	MP, Matching Pursuit
MPTK, Matching Pursuit Tool Kit	Fig, figure
s, écart type	

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Introduction

Dans le présent travail, nous introduisons en phonétique la décomposition granulaire ou atomique appelée encore Matching Pursuit pour l'étude de séquences de parole en arabe et nous validons nos résultats par des références phonétiques incontestables issues d'études effectuées selon les modalités classiques de la discipline.

Depuis 1940 environ, l'analyse spectrale permet d'observer l'évolution des paramètres du signal sonore dans les dimensions – temps – fréquence – amplitude. L'outil privilégié est le sonagraphe, un analyseur à balayage qui applique un filtre passe-bande (réglable) balayant la largeur de la bande passante du signal dans une fenêtre qui se déplace sur l'axe temporel. D'autres appareils[1] destiné à d'autres types de détections comme ceux de J–P. Rousselot[2], le kymographe (1847), le palatographe, le glottographe, etc. ont été utilisés, mais le sonagraphe constitue depuis son invention l'instrument de base du laboratoire de phonétique. En effet, une fois les détails physiologiques de la production des sons d'une langue connus, il est relativement facile de les relier à l'image sonagraphique.

Le sonagraphe est une machine dédiée jusque dans les années 1980, époque ou il commence à être remplacé par des programmes informatiques. La plupart des études

[1] http ://web2.bium.univ-paris5.fr/livanc/?p=2&cote=53034x01&do=pages
[2] Principes de Phonétique expérimentale, Paris, 1897-1908

en phonétique s'appuient aujourd'hui encore largement sur des mesures faites avec cet instrument.

Pour nous, l'informatisation du sonagraphe à certes apporté un réel confort de travail par rapport aux mesures effectuées sur les sonagrammes papier, mais de nombreux points restent insatisfaisants comme la fiabilité des mesures automatiques des formants et du fondamental et nous acceptons difficilement l'élévation de l'écran d'ordinateur, même avec les hautes résolutions actelles, au rang d'appareil de mesure. Quant à la pose des bornes limitant les segments à analyser, son automatisation ne relève pas de ce type d'appareil et il n'existe que des aides peu efficaces pour effectuer ce travail. Ces considérations nous ont toujours tellement rebuté que nous avons définitivement opté pour les mesures manuelles : l'expérience montre qu'il y a plus de travail à faire pour détecter puis corriger les nombreuses erreurs rencontrées dans les mesures que de faire ces dernières directement.

Dès l'origine, le sonagraphe n'a pas manqué de critiques de la part de ses créateurs même, comme Denis Gabor. Ces critiques se positionnaient au niveau mathématiques (utilisation des séries de Fourier), de la théorie de l'information et de la théorie quantique. Pour Gabor tout signal ondulatoire devait relever de la théorie quantique, les séries de Fourier n'étant qu'un cas limite donc extrêmement simplifié de la réalité.

Or, depuis 1940, l'analyse du signal et des formes en général à fait de grands progrès et notre vision des phénomènes ondulatoires s'est considérablement modifiée. Pendant longtemps seul le paradoxe de la lumière considérée comme un phénomène ondulatoire ou corpusculaire, était vulgarisé mais depuis des théories et des réalisations tendent à généraliser la dualité onde–corpuscule. Il y a peu, l'atome a pu être vu en laboratoire sous une forme ondulatoire et l'onde sonore peut être décomposée en corpuscules ou atomes. Les récents développements en acoustique concernant les lentilles acoustiques et les miroirs à retournement temporel nous imposent de considérer l'onde acoustique sous un angle nouveau.

La théorie de l'information propose depuis près d'un siècle des solutions en évolution

constante pour l'analyse des données. L'ubiguïté de l'information et sa nature polymorphe ont entraîné la naissance de l'informatique théorique. L'informatique théorique dont nous trouvons les bases dans les travaux d'Alan Thuring [115], nous donne des théorèmes puissants pour comparer, classer, des objets malgré leur apparent manque de relation. Les propriétés de l'information et les théorèmes qui en découlent sont mis en oeuvre dans les techniques de séquençage moléculaire, en astrophysique et partout ou il y a de grandes bases de données avec des organisations sous-jacentes cachées, à analyser.

Participant à des recherches en phonétique depuis longtemps, nous cherchions une nouvelle façon de représenter un signal évoluant sans cesse comme celui de la parole. Sur le sonagamme, les formants sont perçus par beaucoup comme des entités continues sur l'espace ou ils apparaissent alors qu'il n'en est rien puique que le sonagraphe ouvre des fenêtres temporelles et fréquentielles de largeur fixe sur le signal. Cette technique ne peut prétendre rendre compte de tous les phénomènes relatifs à l'acoustique de la parole, en particulier de ceux dont la durée ou la variation fréquentielle est inférieure à la dimension de la fenêtre.

Des découpages différents du plan temps–fréquence ont été proposés pour tenter de remédier aux carences de l'analyse sonagraphique. C'est le cas notamment avec les ondelettes où l'on associe un découpage temporel variant avec la hauteur du son des composantes spectrale du signal associé à des familles de petites ondes dites ondelettes (ou waveletts) qui sont les ondes élémentaires susceptitble de reconstituer le signal par sommation. Cette approche intéressante nous a semblé toutefois peu adaptée à notre problématique car elle implique la définition d'un quadrillage déterminé du plan temps fréquence et d'une famille d'ondellettes. Face à cette situation, et tout en surveillant l'évolution de cette technique, nous avons exploré d'autres voies.

L'approche par ondelettes est à la base d'une évolution fondamentale dans l'analyse temps–fréquence des signaux. En effet si l'on considère que le plan temps–fréquence peut être découpé en fonction des besoins locaux et l'ondelette peut être calculée en fonction des mêmes besoins, nous approchons d'une solution à notre problème. C'est l'agorithme Matching Pursuit que nous décrirons plus loin qui va nous donner la solution pour une

décomposition fine en temps et en fréquence du signal.

Reste des questions fondamentales en suspend dont celle relative à la nature ondulatoire ou corpusculaire du son. L'expérience montre que l'on peut parfaitement traiter les « atomes » issus de Matching Pursuit comme des corpuscules. Ces atomes sont réversibles et peuvent se montrer sous leur aspect ondulatoire lors de la synthèse musicale notamment et comme les atomes et particules constituant la matière, les atomes sonores sont assimilables à des grains d'énergie.

Les considérations et résultats précédent vont avoir des conséquences importantes au niveau pratique car si nous pouvons décomposer une séquence sonore en une séquence d'atomes, nous allons pouvoir traiter les sons de la même manière que des molécules chimiques. En particulier nous pourrons regarder du côté du séquençage biologique pour lesquel de grands progrès ont été faits. Il y a de nombreux projets en cours tel le séquençage de l'ADN, de l'ARN, des mollécules complexes, etc. mais aussi de la surveillance dans des conditions extrêmes avec le projet SETI[3] ou les travaux de détection des dysfonctionnements organiques en médecine. Le principe reste toujours le même, il s'agit de mettre l'objet à analyser sous la forme d'une suite caractères et de faire des comparaisons.

Le classement des objets dans des ensembles cohérents étant une activité scientifique fondamentale, notre l'hypothèse est que nous allons pouvoir utiliser les techniques de décomposition atomique pour classer des productions phonétiques, distinguer des locuteurs et des langues. D'après nos lectures et des correspondances entretenues avec des chercheurs, la voie que nous explorerons ici semble abandonnées depuis quelques années en phonétique. Le logiciel Guimauve que nous utilisons pour la décomposition atomique a été développé à l'Observatoire du Pic du Midi pour étudier les variations de la trajectoire d'un satellite à partir des photographies envoyées par l'engin. Son développement

[3]SETI vise à trouver dans le rayonnemenent cosmique des structures organisées en provenance d'une vie extra-terrestre. Au-delà d'un projet aux apparences ésotériques, il y une recherche scientifique et technologique de reconnaissance des formes dans les signaux bruités qui peut apporter beaucup à notre recherche.

a cessé à partir de 2002, tandis que d'autres projets basés sur MP prenaient la relève.

Nous pensons que les phonéticiens ont été bloqués par le manque d'outils d'analyse de résultats car si Matching Pursuit présente en lui-même un défit informatique, l'analyse des séquences obtenues en est un plus grand encore.

En effet, MP consiste à trouver le plus grand atome pouvant « entrer » dans un signal donné sachant que le logiciel devra déterminer la fréquence, l'intensité et la durée de cet atome. On cherche l'atome qui va bien en variant ses caractéristiques jusqu'à l'obtention d'un résultat satisfaisant Ceci nécessite une très grande quantité d'itérations qui sont des boucles d'essais, donc un temps machine important. Mp est un algorithme dit glouton et la réponse à ce problème se trouve dans la recherche d'algorithmes parcimonieux. Seule l'informatique actuelle permet de mettre en œuvre un tel algorithme

Ensuite il faut comparer les séquences d'atomes obtenues. Cette comparaison est une opération subtile appelée étude de similarité ou les objets à comparer présentent des traits communs mais impossible de quantifier précisément. Nous ne pouvons pas utiliser la superposition, les homothéties ni les échelles. C'est là qu'interviennent des applications qui dépendent directement de la théorie de l'information puis de l'informatique théorique dont les premiers résultats pratiques sous forme de logiciels accessibles à tous datent seulement de 2003.

Chapitre 2

Généralités

2.1 L'espace arabophone

Nous travaillons sur un corpus de différents dialectes arabes et nous donnons ici quelques indications sur cette langue en discutant quelques des idées qui nous passionnent. L'arabisation est une tentative d'unification linguistique relativement récente puisque qu'elle date, si l'on s'en tient au calendrier de l'Hégire, de l'an 622 du calendrier grégorien ou du 23^e siècle du calendrier imazigh, zone maghrébine dominée aujourd'hui par les arabophones. Avant cette époque, il est question de langues pré-islamiques ou de proto-arabe. En réalité l'arabe existe avant 622 parmi les langues pré-islamiques qui sont constituées des nombreux dialectes des sociétés polythéistes de l'époque. La Mecque était depuis longtemps un important centre commercial et culturel ou les différentes tribus se mêlaient. Les dialectes pré-islamiques étaient indifféremment utilisés pour le commerce, des joutes oratoires et poétiques et les diverses activités sociales de ce lieu cosmopolite. C'est l'un de ces dialectes qui aurait émergé et se serait imposé et répandu, fédérant l'ensemble des tribus sous une seule langue, une seule culture.

Des chercheurs ont des idées sur la tribu dont la langue a été à l'origine de l'arabe du Coran mais c'est toujours un sujet de discussion et nous avons relevé deux idées intéres-

santes. La première pose que c'est la langue comprise par le plus grand nombre aurait été choisie, par exemple la langue la plus usité dans les relation inter-communautaires, et la deuxième qu'elle a aurait été construite à partir des dialectes de l'époque de façon à être comprise par tous. Cette inter-compréhension était en effet indispensable pour réussir le projet coranique et l'arabisation fut effectivement une réussite. Toutefois, les expériences de contruction de langue n'int jamais bien réussi.

Dans une conférence, Judith Rosenhouse [89] fait une liste non exhaustive des chercheurs qui ont jalonné l'histoire de la langue arabe depuis le VII^{e} siècle de notre ère. Nous notons que dès le début de son existence officielle, l'arabe a ses spécialistes connaissant parfaitement la grammaire, la phonétique et la phonologie qu'ils ne distiguent pas complètement.

2.1.1 Évolution et modernité

La résistance au fait dialectale

Langue révélée, l'arabe n'admettrait aucune variante, ce qui n'est évidemment pas le cas. Admettre qu'il y ait des variétés multiples c'est admettre qu'il n'y a pas unité, mais au contraire divergence, buts et visions différentes du monde, ce qui est impensable dans le cadre d'un univers monothéiste centralisé. Or, la nécessité faisant force de loi, il a bien fallu adapter l'idiome original à la grande variabilité des conditions écologiques, sociologiques, économiques des régions où il s'est implanté, puis avec l'évolution des sociétés humaines aux objets technologiques, scientifiques et culturels nouveaux.

L'évolution a eu lieu presque toujours à l'insu des puristes de la langue et des locuteurs. Parmi les raisons des évolutions, il y a celle de l'expansion sur des lieux occupés par des populations indigènes parlant leurs propres langues. Ces populations en intégrant la langue arabe à leurs parlers lui ont conféré un accent étranger, conséquence de leur propres habitudes langagières. En effet, la première chose que l'on perçoit d'une langue, avant d'en comprendre le sens, c'est sa musique, autrement dit son organisation pro-

sodique. Chaque langue possède une organisation accentuelle, rythmique et mélodique spécifique, particulièrement évidente lorsque qu'un locuteur transpose la musique de sa langue maternelle dans une langue étrangère. Ce phénomène que l'on nomme l'accent étranger nous révèle que toutes les langues ne chantent pas sur le même air

L'expansion rapide de l'espace couvert par l'arabe au sud de la Méditerranée et vers l'Orient a impliqué un grand nombre de langues-substrat, ce qui laisse entendre qu'il doit y avoir, rien que pour cette raison, de nombreuses variétés dialectales d'arabe. Bien que les grammairiens arabes anciens aient fait de grand efforts pour décrire la prononciation phonétique du « bon arabe » cette expansion rapide des territoires conquis n'a pas permis de former en nombre et en qualité suffisante les enseignants nécessaires. La nature souvent désertique du terrain avec d'immenses distances et des relations difficiles donc réduites entre les communautés a eu pour conséquence une dilution des savoirs centraux dans des savoirs locaux.

2.1.2 Pourquoi y-t-il de nombreuses langues plutôt qu'une seule ?

Cette question n'est pas triviale. Elle est du même ordre que celle qui demande pourquoi il y a quelque chose plutôt que rien. Dans ce mémoire, nous travaillons sur des variétés d'une langue : l'arabe. Nous admettons donc l'existence d'une langue originelle ou langue mère, laquelle aurait divergé selon les lieux et selon des lois qui nous restent obscures. La variabilité des langues, tout comme la variabilité biologique, suit des lois dont nous ne pouvons qu'observer les effets. Bien sûr il est possible d'exclure certaines variations tant pour les langues que pour les espèces vivantes, mais nous ne pouvons pas dire quand, ni comment, une variation va entrer dans un registre linguistique ou vivant donné.

L'idée de la langue unique originelle est un mythe ancien que des linguistes comme Merritt Ruhlen [92] ont tenté de replacer dans le registre des probabilités scientifiques avec une argumentation qui paraît cohérente. Cette hypothèse ne manque pas d'arguments et si l'on part de l'idée que l'humanité est née sur le territoire restreint d'une

petite tribu, elle paraît évidente.

Dans la préface de la traduction française du livre de Ruhlen, l'anthropologue et généticien André Langaney [93] note qu'il y a une forte corrélation entre l'appartenance génétique et linguistique. De plus la proximité génétique des 6 milliards d'êtres humains actuels implique une origine biologique commune. Or Langaney avance que des indices laissent entendre que l'humanité a failli disparaître au cours d'une glaciation qui se serait produite entre 30 000 et 100 000 ans avant notre ère. La population aurait été réduite à quelques dizaines de milliers d'individus, soit juste le nombre d'individus nécessaire pour empêcher l'extinction de l'espèce.

L'homogénéité de l'espèce humaine, confirmée par la génétique, s'expliquerait par le « goulet d'étranglement » que constitue cet accident climatique où seule une population regroupée de plus de 10 000 individus a pu maintenir les liens indispensables lui permettant de survivre. Ce regroupement imposait une grande solidarité donc la construction de liens forts avec pour conséquence une langue unique. Cette langue a pu se forger sur une très longue période au regard du temps humain, ce qui a pu permettre d'en faire une langue maternelle, même si des variantes ont existées. Cet événement aurait donc entraîné la fusion-disparition des hypothétiques langues humaines pré-existantes tandis que les autres langues, dont celles possibles d'autres espèces d'hominidés vivant antérieurement à la glaciation disparaissaient avec ces espèces [49] et [50].

La théorie de la langue originelle unique est fortement critiquée par des chercheurs comme Metoz et al. [73]. Metoz montre qu'un modèle probabiliste trouve des cognats à 100% dans toutes les langues et arrive donc au mêmes résultats que Ruhlen. Dans un article, Anne Szulmajster-Celnikier [108] fait un bilan des critiques portées à la théorie de Ruhlen, tant au plan linguistique que génétique.

2.1.3 Théorie de l'aire ancestrale

C'est la théorie de l'apparition de l'humanité sur un territoire limité. Dans sa préface au livre de Ruhlen, Langaney tempère un peu les critiques qui pourraient être adressée à l'auteur en posant que Rulhen peut avoir raison au plan de l'évolution même si son argumentation est insuffisament étayée au plan linguistique. En effet si l'accident glaciaire à bien eu lieu, la survie de petits groupes humains dans des conditions extrêmes est difficilement acceptable. L'hypothèse retenue par Laganey est celle d'un groupe humain important concentré géographiquement pratiquant des échanges intenses pour survivre et forgeant par nécessité une langue unique.

Une autre hypothèse propose l'existence de foyers multiples auquel cas il faut aussi que, en chaque lieu, les populations aient été suffisantes pour survivre. Ces deux théories reprennent celle de l'aire ancestrale et proposent que l'humanité ait connu au moins deux situations de ce type au cours de son existence.

2.1.4 De l'unicité à la multiplication

Les langues sont des systèmes vivants soumis aux principes de l'évolution, donc aux pressions de l'environnement. Tout les chercheurs ne seront pas d'accord avec cette assertion, pourtant les formes qui perdurent sont celles qui sont les mieux adaptées au contexte où elles s'épanouissent. En ce qui concerne l'être humain et semble-t-il toutes les formes de vies complexes, l'environnement social et culturel prend le pas sur les nécessités écologiques primaires.

L'évolution darwinienne des espèces est un processus lourd modifiant en profondeur des éléments biologiques stables. Elle rencontre une forte résistance au niveau de ce matériel biologique, même quand l'adaptation de l'organisme au milieu est indispensable. Les langues étant soumises aux lois de l'information ont plus de facilité à évoluer que les organismes vivants. Et c'est bien ce que nous pouvons observer chaque jour en comparant des environnements relativement stables comme l'Europe et des milieux nettement plus

instables comme les pays où l'urbanisation rapide modifie profondément les modes vie et les parlers, y compris ceux des populations d'accueil [72].

2.2 Thermodynamique et diversification des langues

Nous avons déjà soutenu une théorie néo-darwinnienne de l'évolution des langues. Nous tenterons dans ce qui suit d'affirmer une vision thermodynamique de cette évolution. La thermodynamique est en effet au coeur de toute sortes d'échanges, dont l'information.

2.2.1 La théorie constructale

En 1996 Adrian Bejan [5] écrivait : « Pour qu'un système fini puisse persister dans le temps, il doit évoluer de manière à offrir un accès facilité aux flux qui le traversent ». La loi constructale est le principe qui génère la forme la plus parfaite par évolution adaptative, ce qui donne (en réalité) la forme la moins imparfaite possible, celle qui est adaptée à une situation et à un instant donné. Les langues sont aussi soumises aux lois de l'univers, dont la thermodynamique fait partie aussi étrange qu'elle puisse paraître dans ces principes. Dans sa théorie constructale Adrian Bejan démontre que toute forme efficace, c'est-à-dire remplissant la fonction attendue est parfaite au sens matérialiste du terme, c'est-à-dire physique, thermodynamique, mécanique et nous ajouterons informationnel et social. Seulement il y a une quantité non dénombrable de formes efficaces, donc « parfaites » possibles.

La théorie constructale nous permet de poser que toute langue, pour un locuteur et à un moment donné, est parfaite et nous justifions par la-même la variabilité des formes observées y compris dans une même famille et pouvons mettre leur étude sur un même plan : il n'y a pas de langue meilleure ou supérieure à une autre.

Au cours d'une génération au XX^{e} siècle, le français a perdu l'usage d'un grand voca-

bulaire adapté à des techniques et modes de vie disparus dans ce temps très court. Nous pouvons citer l'agriculture avec la disparition de la paysannerie, les transports, de nombreuses industries ainsi que tout un pan de l'artisanat, son outillage et ses techniques qui ont soit disparu, soit se sont radicalement transformé. D'autres pratiques avec d'autres mots sont venus combler le vide laissé par les disparus : il y a eu adaptation.

Une langue imparfaite ne peut pas exister en l'état

Nous posons que toute langue qui ne serait pas parfaite ne peut exister en l'état. Dans les faits nous observons bien que les langues ne remplissant pas ou plus correctement leur rôle sont amenées à changer ou à disparaître. Nous pouvons donner l'exemple du latin qui a divergé vers le français, l'espagnol, l'italien, le portugais et leurs variantes, les parlers qui évoluent très vite dans les régions de forte expansion urbaine afin d'adapter un langage souvent d'origine rurale à la réalité de la ville ou les parlers d'entreprises qui, isolés de toute forme sociale étendue (ils sont réduits à quelques individus), n'ont plus qu'un lointain rapport avec l'anglais et ne survivent, comme le latin d'église que par un effet de communauté d'intérêt.

Réflexions à partir de la théorie constructale

Dans sa thèse, « L'auto-organisation de la parole » Pierre-Yves Oudeyer [81] utilise l'IA pour montrer qu'un langage entre des machines peut émerger et s'auto-organiser. Dans leur présentation et discussion de ce travail, Jean-Paul Baquiast et Christophe Jacquemin[1] écrivent « *on peut s'étonner précisément que Pierre-Yves Oudeyer ne fasse pas allusion à cette théorie* [la théorie constructale] *et aux applications qui en sont données dans différentes disciplines, allant de la recherche fondamentale à l'ingénierie quotidienne. Il nous semble que ses propres hypothèses et celles d'Adrian Bejan se complètent fort bien.* »

[1] http ://www.admiroutes.asso.fr/larevue/2003/50/pyo.htm, consulté le 10/09/2008

La théorie constructale est un moyen indépendant de la linguistique qui nous permet de comprendre pourquoi il n'y a pas et il n'y aura jamais de langue unique, universelle. Elle justifie à elle seule l'intérêt de l'étude des variantes linguistiques et phonétiques ainsi que l'importance de la dialectologie. L'observation de l'évolution phonétique à travers la théorie constructale et en particulier celle des dialectes devrait apporter de nombreuses réponses à des questions qui restent ouvertes.

2.3 La phonétique quantique

Á une réunion du groupe ATALA[2], une équipe internationale de chercheurs a présenté des éléments de « phonétique quantique ».

Cette théorie accorde un statut égal aux dimensions articulatoire, acoustique et auditive de la langue parlée contrairement aux théories de Jakobson, Chomsky, Halle et de quelques autres. Elle établit des rapports entre les structures abstraites posées par la phonologie (notamment les phonèmes) et leur expression phonétique. Depuis quelques années, Stevens [100] a développé une nouvelle démarche dans le cadre de la théorie quantique des traits, en remarquant que les rapports non monotoniques entre les modifications du conduit vocal et les paramètres acoustiques qui leur sont associés favorisent la catégorisation des sons :

- il y a des non-linéarités entre les paramètres acoustiques et les réponses auditives qui leur sont associées qui, par leur analogie, favorisent la catégorisation ;
- l'inventaire de sons de chaque langue est fait de traits distinctifs définis par les régions de stabilité acoustique ou auditive ainsi dégagées.

La théorie quantique tente d'intégrer des modèles phonologiques basés sur des modèles sonores détaillés pour chaque langue. Elle vise des applications médicales, l'orthophonie, l'informatique, l'enseignement des langues et l'ingénierie. Une des conclusions des recherches dans cette voie est que pour progresser en phonétique, il est nécessaire

[2]http ://www.atala.org/Recherches-actuelles-en-phonetique

de faire des analyses acoustiques de plus en plus fines. C'est justement ce que nous nous proposons de faire avec la décomposition atomique des sons de la parole.

D'autres approches tentent catégoriser les sons avec une meilleure précision que les méthodes habituelles. Notons la théorie des régions et modes distinctifs [78] ou la théorie de dispersion / focalisation [95].

Chapitre 3

Méthodologie

Notre travail porte sur l'organisation matérielle des formes du signal sonore supportant l'information véhiculée par la parole , ce que l'on peut nommer « information structure ». C'est à partir de cette information que nous allons tenter de classer, rapprocher ou dissocier des parlers par différentes méthodes. Notre critique de la méthode la plus généralement employée va nous permettre au su de la théorie de l'information de justifier la recherche d'une approche nouvelle pour étudier le signal de parole.

3.1 Méthodes d'analyse de la parole

L'analyse de la parole peut se faire à partir de l'observation directe, d'un modèle de production en partant des sources de bruit qui produisent les phonation-articulation (conduit vocalique, observation des mouvements) ou sans a priori sur l'origine du signal par écoute et/ou enregistrement dans l'environnement du locuteur. Nous utiliserons la technique sans apriori pour nos études en remarquant qu'un signal sonore peut être enregistré et reproduit d'une façon quasi-parfaite par les systèmes électroacoustiques actuels qui sont sans aucun rapports avec les organes vocaux humains. Il est devenu quasiment impossible de distinguer la production humaine de la production électroacoustique.

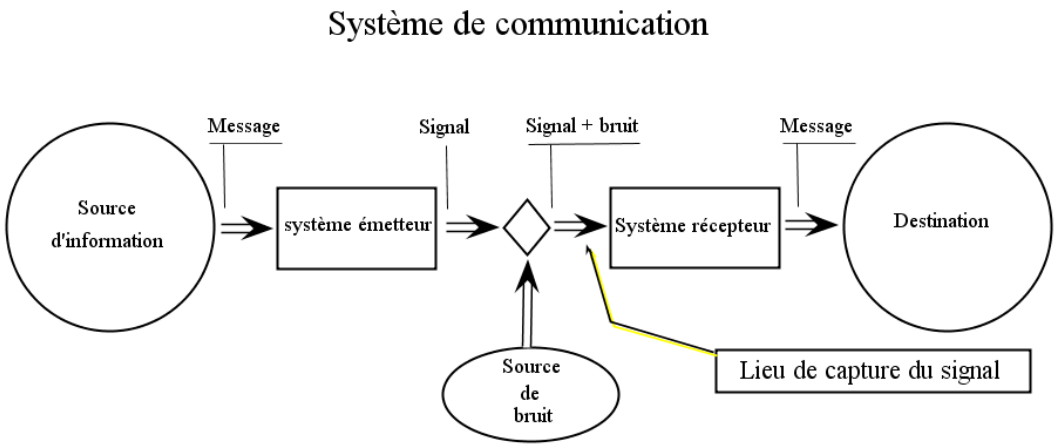

FIG. 3.1 – *le système de communication d'après Claude Shannon adapté*

Le schéma 3.1 primitif mais désormais classique de la communication, nous permet de préciser le lieu de saisie de l'information que nous traiterons [97]. Nous pouvons observer sur ce schéma que le signal capté ne correspond pas à un signal perçu par l'émetteur ni à celui perçu par le récepteur. En effet :

a l'émetteur perçoit sa production selon deux voies principales, par conduction interne (os, oreille interne) et externe, par plusieurs chemins qui vont de l'ouverture des lèvres aux oreilles externes selon les conditions environnementales ;

b le récepteur, lui, perçoit un son modifié par la distance à l'émetteur selon les lois de l'acoustique, par les résonances et réverbérations du lieu d'échange et par les bruits ambiants. En outre il est influencé notamment par ce qu'il voit, attitude du locuteur, mouvement des lèvres (lecture labiale) et autres facteurs ambiants ;

c le schéma original ne tient pas compte des boucles de rétroaction entre l'émetteur et lui-même, l'émetteur et le récepteur et le récepteur et l'émetteur. Il est plutôt destiné à la communication entre machines mais sa simplicité convient parfaitement à notre propos.

C'est N. Wiener [120] qui en 1950 ajoutera les boucles de rétroaction au schéma de Shannon. Notre approche n'étudie que le signal acoustique et ne tient pas compte des

éléments de communication qui passent par d'autres canaux que le son comme les attitudes, gestes, etc. Ces éléments font partie de l'étude de la communication multimodale [99] que nous ne pouvons pas ignorer, mais qui n'intervient pas directement dans notre problématique.

3.1.1 Étude de référence pour valider notre projet

Pour l'étalonnage de la méthode instrumentale basée sur Matching Pursuit que nous nous proposons de décrire au chapitre 3, nous nous appuierons sur des travaux effectués avec des méthodes classiques dans le cadre de plusieurs publications. Ces travaux de référence reposent sur l'étude de la coarticulation en ASC et en AD dans les situations de consonnes pharygalisées vs non-pharygalisées et notamment sur des mesures de la résistance coarticulatoire à l'aide de l'équation de locus. L'exercice consiste à distinguer les traits phonétiques propres à chaque région et les différences entre ASC et AD et de vérifier la concordance des résultats obtenus avec chaque méthode.

Cette première étude utilise les méthodes standards de la phonétique et a fait l'objet de validation par des comités de lectures. Quand à la deuxième, elle est tout à fait originale dans la mesure où nous n'avons pas trouvé d'équivalent dans les publications que nous avons pu consulter. Il faut noter que la plupart des informations et des logiciels utilisés sont très récents et ont été développés à d'autres fins que la phonétique. Avant d'entrer dans le vif du sujet nous tenons à préciser quelques points et la façon dont nous comprenons les idées sous-jacentes.

3.2 Signal et information

Nous nommons « signal » toute forme contenant de l'information [6]. Un signal est donc l'élément matériel qui transmet l'information.

3.2.1 L'information

L'information est une forme au sens physique qui, pour un être humain, appartenant à une culture donnée, induit des représentations, des idées. Une même information peut être présentée de multiples façons en s'appuyant sur une grande variété de supports matériels et en particulier en tant que savoir dans un système nerveux. La parole est un « signal sonore » qui transmet de l'information sous forme d'onde sonore.

L'information n'a pas de masse, mais elle a besoin de l'énergie et de la matière comme support pour exister. Elle est polymorphe car elle peut utiliser indifféremment tout support matériel y compris biologique et adopter des formes diverses pour un même signifiant (textes, idéogrammes, langues, fichiers informatiques, onde électrique, lumineuses, structures nerveuses, sonores, etc.) Un signal sonore n'est pas un objet, n'a pas de masse, mais il est porteur d'énergie.

Nous pouvons encore définir l'information (ou néguentropie) comme le facteur d'organisation qui s'oppose à la tendance naturelle au désordre et au chaos ou entropie. L'exemple classique est celui d'un organisme vivant qui ne peut rester vivant (organisé) que par les informations qui lient ses différentes parties. Toute rupture du flux d'information (nerveuse, chimique, etc.) entraîne la dégradation d'une partie ou de l'ensemble et la dégradation d'une partie de l'ensemble entraîne souvent celle du tout : pour l'être vivant, information, matière et énergie sont indissociables. Cette constatation va nous fournir les moyens d'analyse de nos données.

Notre approche de la théorie de l'information s'est faite à travers deux auteurs principaux, Henri Laborit et Paul Watzlawick et l'école de Palo Alto. Tous deux traitent de la transmission de l'information, Laborit [58] au niveau de la biologie et du système nerveux et Paul Watzlawick au niveau psychologique. Watzlawick pose avec son équipe les bases axiomatiques de la communication (il est impossible de pas communiquer, etc., [119]) et distingue deux types d'information : l'information analogique qui est copie ou imitation de la forme du phénomène à transmettre, suggestion de l'idée de l'objet et l'information numérique qui étiquette ou donne des mesures de l'objet à identifier en

établissant une relation bijective entre un ensemble de réalités ou supposé tel et un ensemble de symboles ou de nombres. Il montre aussi l'importance vitale de la rétroaction dans les processus de communication.

3.2.2 Information analogique et numérique

Pour l'onde sonore, l'information analogique peut être inscrite dans la forme du sillon du cylindre de cire du phonographe d'Edison ou du champ magnétique des bandes du même nom. L'information numérique elle, est le mot de vocabulaire, le numéro de série ou la suite de bits qui donne, dans le cas de nos enregistrements, une valeur numérique et une seule de la pression atmosphérique aux instants $t_1, t_2 \ldots t_n$, où t_n est l'instant de la mesure ou de l'échantillonnage du signal.

Si l'information analogique et sujette à la déformation et à l'usure du support, l'information numérique est beaucoup plus robuste, permettant une reconstruction de l'original à partir d'un support partiellement abîmé, des corrections calculées et une possibilité de duplication en chaîne et sans perte des documents. Seule la qualité du support en limite la pérennité.

3.2.3 Information structure et information circulante

Pour Henri Laborit [58] existent l'information-structure qui stratifie le monde en niveaux d'organisation, par exemple les niveaux (...particules ⇒ atomes ⇒ molécules ⇒ cellules ⇒ organes ⇒ organismes ⇒ sociétés...), niveaux d'organisation classés ici par ordre de complexité croissante ou négantropie, et l'information-circulante qui permet le passage d'un niveau à l'autre dans les deux sens. Il y a similarité/complémentarité entre les visions de Laborit et de Watzlawick et cela concerne directement notre champ de recherche.

Nous observons que la parole se situe au niveau de l'information circulante puisque

sa fonction est justement de traverser les divers niveaux d'organisation de l'expérience humaine. Toutefois, c'est son information structure que nous étudions ici.

La parole est un « signal sonore » qui transmet de l'information avec des ondes sonores, elle utilise l'énergie biologique du locuteur sous forme d'air comprimé par les poumons qui est ensuite modulé par les diverses cavités de l'organe phonatoire. Elle se présente sous forme de séquences sonores structurées et articulées entre elles qui ont besoin du temps 1) pour s'exprimer et 2) pour être décryptées.

3.2.4 Les apports de la théorie de l'information

La théorie de l'information fourni un certain nombre de théorèmes utiles à notre projet. D'une part la démarche scientifique consiste à réduire la complexité de l'information à traiter sous une forme accessible à notre compréhension et d'autre part la démarche historique constante de l'espèce humaine a été de transformer et de transmettre l'information sur des supports de plus en plus petits, de masse de plus en faible, utilisant donc de moins en moins d'énergie : des tables de pierre aux faisceaux de lumière cohérente, d'une multitude d'événements et d'objets divers vers une formulation synthétique. Les forces de la nature ne sont plus que quatre aujourd'hui et la physique cherche à réduire encore ce nombre tandis que les objets que nous cherchons à classer deviennent de plus en plus abstraits. C'est désormais l'information et ses formes qui font l'objet de traitements ce qui est le cas pour nous avec l'étude des signaux sonores de la parole.

3.2.5 Complexité et organisation sous-jacente

La complexité d'un objet peut être définie par la proportion d'information, matière et énergie qu'il contient. Le niveau de complexité d'un système est déterminé par ses constituants, ceux-ci étant des systèmes ayant leurs propres niveaux d'organisation, chaque niveau étant défini par le type d'énergie et de structure qui y est mis en œuvre. Ces considérations nous permettent de dire que les sociétés humaines qui sont des orga-

nismes vivants ne peuvent exister sans la parole qui est le vecteur informationnel qui lie les êtres qui les composent. La parole est une caractéristique fondamentale des sociétés humaines et non de l'être humain, qui ne peut l'exercer que dans le cadre d'une communauté et la multiplicité des groupes humains a pour conséquence naturelle la multiplicité des langues et des dialectes.

3.2.6 La rétroaction

La rétroaction est une fonction dynamique fondamentale du monde vivant et du calcul analogique. Elle est présente à tous les niveaux de l'activité vivante et donc dans toute situation de communication. Si l'effecteur est l'organe ou organisme agissant, le principe consiste à renvoyer une partie de l'information présente en sortie, sur son entrée. Dans une situation de communication, l'effecteur est l'émetteur du message et le retour se fait de plusieurs façon dans le cas de la communication orale : il y a ce que le locuteur ressent et entend de lui-même et ce qui lui est renvoyé par l'interlocuteur.

La rétroaction permet trois type de comportement selon que le signal renvoyé en entrée s'additionne ou se retranche du signal d'excitation : la stabilité en sortie quelque soient les conditions en entrée, c'est un régulateur ; une sortie proportionnelle à l'entrée, c'est un amplificateur ; une sortie en tout ou rien, il y a là une rétroaction positive qui est celle que l'on trouve dans les situations explosives qui peuvent être émotionnelles, chimiques, mécaniques, ou situationnelles. La rétroaction positive est utilisé dans les circuits logiques qui ne donnent en sortie que 0 ou 1, (tout ou rien).

La notion de rétroaction à été le principe de base qui a permis le développement de la cybernétique, rendu possible la compréhension du mouvement dans le monde vivant et la réalisation des premières machines autocontrôlées zoomorphes (tortues) ainsi que le développement de la théorie de l'information. Le principe de rétroaction est fondamental pour la réalisation de systèmes (effecteurs) régulés ou actifs, la stabilité des systèmes biologiques et des mécanismes régulés (sortie constante, comme la température du corps ou d'un habitat) en dépendent, de même que l'action motrice (gestes) qui se

voit régulée par ce moyen. Par exemple le ralentissement de l'influx nerveux provoque des tremblements lors des gestes en raison de la rétroaction trop tardive qu'il provoque (situations de fatigue, maladies nerveuses, médicaments, drogues, etc.)

Dans l'étude de la communication multimodale, la rétroaction devient un des éléments essentiels de la relation inter-locuteurs.

3.3 Les systèmes numériques

Le monde numérique pour qui toute sortie résulte d'un calcul n'a pas besoin de la rétroaction. Depuis l'avènement de l'ère numérique, les systèmes de traitement de l'information sont dotés de dispositifs permettant de corriger les erreurs par calcul, ce qui a permis de comprendre que le monde vivant et en particulier les humains corrigeaient aussi leurs erreurs par calcul de telle façon qu'une information incomplète ou dégradée puisse être intégrée correctement. C'est un phénomène qui existe lors de l'apprentissage d'une lange, qu'elle soit maternelle ou autre. Nous pensons que ces points sont importants et même indispensables à notre réflexion en phonétique au niveau de toute réflexion sur la transmission de la langue parlée et les messages messages qu'elle véhicule.

3.3.1 L'intelligence artificielle (IA)

Nous nous intéressons depuis longtemps à ce qui se rapporte à cette science qui tente d'imiter le vivant. Les systèmes analogiques sont désormais complétés par des systèmes numériques qui permettent de simuler l'intelligence du vivant par un assouplissement et une multiplication des possibilités de mémorisation et l'apprentissage. Le développement des micro puis nano-techniques et des réseaux apporte chaque jour son lot d'inovations. Micro-robots, implants non invasifs, réseaux de toutes sortes permettent de construire des systèmes capables de communiquer et d'évoluer entre eux et d'inter-réagir avec le vivant en communicant, en apprenant et en se transformant eux-mêmes. La construc-

tion de ces systèmes permet, en limitant strictement les conditions de départ, d'étudier l'évolution d'un phénomène ou, inversement, de rechercher les conditions nécessaires à sa réalisation. Ces dispositifs sont utilisés par exemple pour étudier l'évolution d'un langage entre machines ou dans un ensemble mixte, vivant-mécanique. Les résultats permettent au spécialiste prudent de mieux comprendre certains phénomènes cachés du monde vivant et à la robotique de progresser.

L'apprentissage automatisé

Un apprentissage peut être automatisé par programmation et s'adapter à des conditions extérieures. Si l'objectif peut-être atteint par une suite d'opérations simples, nous disposons à coup sûr d'un algorithme efficace et d'un langage adaptés à cette tâche. Ainsi, il est relativement simple d'apprendre à une machine à calculer la durée d'une série de sons, de faire une moyenne, d'avertir l'autorité compétente à la suite d'un événement particulier ou d'une combinaison d'événements analysés à travers un réseau bayésien.

Cependant, dans de nombreux cas, nous sommes dans l'incapacité d'expliquer une situation ou sa complexité est tellement grande que nous n'avons pas d'autre choix que de réaliser un nombre très important d'essais successifs : c'est que l'on appelle la résolution de problème par force brute. Cette technique permet par exemple de « casser » le code de documents chiffrés, de reconnaître un texte ou un mot parmi une multitude de documents ou de bruits. Le temps requis est souvent un frein à ce type de méthode. C'est la raison d'être d'algorithmes comme la Méthode de Monte Carlo qui utilise le hasard le plus parfait possible pour résoudre des problèmes difficiles comme le calcul de fonctions non intégrables ou sans solution classique. L'utilisation d'un hasard le plus parfait possible évite alors de perdre du temps à tirer plusieurs fois le même « jeton ». De même nous verrons que la compression de fichiers la plus parfaite possible offre une solution élégante pour leur comparaison.

L'apprentissage artificiel prend tout son sens dans la résolution de problèmes avec des

données multiples pour lesquels l'ordinateur va pouvoir apprendre une tâche en faisant le bilan de milliers d'exercices : pour être efficace, le programme doit être conçu pour retenir la configuration la plus favorable en fin d'exercice. Ces techniques peuvent être supervisées ou non par un expert humain.

Les réseaux de calcul

Les structures les plus communément dédiées à l'apprentissage se présentent sous formes de réseaux. La plus célèbre concerne les réseaux neuronaux, terme qui paraît magique par sa similitude supposée avec notre cerveau. Pourtant ces réseaux sont très différents de celui-ci, par leur simplicité d'abord et parce les réseaux neuronaux sont des structures plutôt rigides, avec des connexions en anneau, en treillis, etc. mais ils sont très efficaces dans leur spécialité. Ils demandent malheureusement un apprentissage souvent long et des essais multiples et fastidieux qui sont à refaire de nombreuses fois pour affiner les résultats, travail qui doit-être recommencé à chaque modification de la configuration du réseau ou des besoins. Nous avons rejeté ce type d'approche à cause de ces contraintes et parce qu'elle a largement été explorée.

Les Réseaux de Markov Cachés sont plus souples que les réseaux neuronaux et très utilisés en reconnaissance de la parole. Nous avons également étudié les possibilités offertes par les réseaux bayésiens qui sont entièrement probabilistes et ne nécessitent pas d'à priori.

D'autres systèmes tentent d'imiter le monde vivant avec une programmation leur permettant de s'adapter aux conditions environnementale et d'évoluer. Notre attention s'est portée sur les automates évolutionnaires qui permettent de résoudre des problèmes insolubles autrement. Ils permettent par exemple d'étudier l'évolution d'un génome artificiel sur des millions de générations au rythme de 1 ms par génération pour l'automate contre 20 mn pour une bactérie réelle. Ce qui nous intéresse est le fait qu'ils ont été utilisés pour la reconnaissance de la parole, pour l'étude de l'évolution de langages artificiels [81] ainsi que dans des expériences mêlant humains et machines en réseau.

Nous vivons d'ailleurs dès maintenant une expérience en temps réel où des automates capables d'adaptation, doués de facultés de production et de reconnaissance de la parole, ayant la possibilité de répondre à des demandes concrètes, s'insèrent sans heurt apparent dans le tissus vivant de nos sociétés.

3.4 Analyse des sons

3.4.1 Nature de l'objet étudié

L'onde sonore est une pression qui se déplace dans tout matériau solide, liquide ou gazeux. Il n'y a pas son dans le vide. La célérité dans les gaz dépend de la masse volumique (ou molaire) du gaz et de sa température ou $v = (\frac{P}{\rho})^{\frac{1}{2}}$ avec P = pression et ρ = masse volumique. Il y déplacement de l'onde de pression et non de gaz. Si l'on se donne l'image hypothétique d'un gaz où les molécules ont une position bien définie, ces molécules oscilleraient autour de cette position tel que la résultante des forces en ce lieu soit nulle.

Un son est caractérisé par :

1. sa hauteur ou fréquence fondamentale F_0. On dit qu'un son est périodique quand l'onde se reproduit semblable à elle-même dans le temps. La période (T) est la durée nécessaire pour effectuer un cycle, elle est donc l'inverse de la fréquence $T = \frac{1}{F_0}$. Quand un son n'est pas périodique, on dit qu'il est apérioque, il s'agit d'un bruit dont la fréquence varie constamment ou d'une impulsion (Dirac). La variation de fréquence d'un signal apériodique peut suivre une loi définie : elle est alors prévisible (chirp). Si elle ne suit pas de loi simple à décrire, elle est imprévisible. C'est le cas des bruits blancs ou des bruits de consonnes comme [s] ;
2. sa composition harmonique ou spectrale, qui donne le timbre ou couleur du son. Joseph Fourier a montré qu'un signal périodique de forme complexe pouvait être décomposé en une somme de signaux périodiques simples dont la fréquence est

multiple de la fréquence fondamentale qui est la fréquence plus basse contenue dans le signal ;

3. son intensité, c'est-à-dire l'énergie qu'il contient. L'air étant un gaz presque parfait, son élasticité n'engendre pratiquement pas de pertes, mais l'intensité sonore étant une pression, si l'énergie peut être considérée comme constante, l'aire sur laquelle elle s'applique croit rapidement avec la distance (onde sphérique, puis onde plane) de sorte que l'intensité du son décroit très vite avec celle-ci avant de se stabiliser.
4. sa durée dans le temps ;
5. son évolution temporelle. Il y a des signaux stationnaires et non-stationnaires.

Le corpus que nous analysons ici a premièrement été traité par la transformée rapide de Fourier (FFT) discrète à l'aide du logiciel gratuit PRAAT utilisé par de nombreux chercheurs. PRAAT permet la segmentation de la production sonore sur plusieurs niveaux et permet de décomposer celle-ci par transformée rapide de Fourier. Un oscillogramme des sons est disponible avec diverses fonctions comme la détection des passages à zéro du signal. Ce logiciel offre des automatismes et permet des analyses statistiques. En outre, il accepte des scripts dans un langage spécifique.

Malheureusement nous avons souvent mis en échec la détection automatique de la fréquence fondamentale et celle des formants dont l'importance est essentielle à la plupart de nos travaux. Face au nombre d'erreurs générées par le logiciel et au temps qu'il aurait fallut pour les corriger, nous avons préféré effectuer toutes nos mesures manuellement : dans notre cas il fallait plus de temps pour corriger que de faire les mesures manuellement. Utiliser une méthode qui donne de toute évidence un niveau d'erreur important pour un travail d'étalonage relève de la plus grande négligence même si l'utilisation de statistiques peut en réduire les effets.

L'imprécision de l'analyse par formants n'est pas étonnante car nous avons souvent été mis devant des choix difficiles lors de nos mesures mais contrairement à la machine, il nous a toujours été possible de modérer ou d'éliminer la valeur improbable au profit d'une autre plus vraisemblable. Cette situation nous à conforté dans l'idée de chercher

une alternative adaptée au genre d'étude que nous faisions.

3.5 Stationnarité et non stationnarité du signal de parole

La théorie des signaux définit deux types de signaux, les signaux stationnaires et les signaux non-stationnaires. Un signal stationnaire est stable dans le temps contrairement signaux non stationnaires dont la fréquence, la forme, l'intensité, la durée sont variables. Les signaux stationnaires sont périodiques, c'est-à-dire leur forme se reproduit semblable à elle-même dans un intervalle de temps constant appelé période et leur durée est infinie comme la raie de longueur d'onde 21cm de l'hydrogène ou l'onde sinusoïdale d'un générateur de signaux horaires. Leur description présente une complexité minimum : $y = asin(\omega t)$ pour un son pur avec y, l'intensité du signal à la constante a près, ω la vitesse angulaire, t la valeur du temps.

Les signaux non stationnaires, au contraire, ne peuvent être définis ni par une fréquence ni par leur forme, ni par leur intensité. Ils sont essentiellement variables dans le temps. Il existe des classes de signaux non stationnaire décrits simplement mais la plupart ne peuvent être approchés que par des expressions complexes et décrits avec précision que par eux-mêmes.

3.5.1 Voyelles et consonnes (phonèmes)

Il est commun de distinguer deux types principaux de sons de la parole, les voyelles et les consonnes. Les voyelles sont présentées généralement comme des sons périodiques tandis que les consonnes sont des sons dits apériodiques, des bruits.

En fait, la réalité est moins simple que cela et la périodicité des voyelles n'est qu'une vision réductrice de leur évolution temporelle. En premier lieu elles ont une durée limitée à quelques périodes dans le temps ce qui élimine d'emblée le caractère de stabilité tempo-

rel. Lors de leurs courte existence elles évoluent sur plusieurs cycles durant lesquels leur fréquence et leur énergie varient plus ou moins rapidement. Enfin elle connaissent souvent de fortes perturbations au contact des sons environnant. La seule chose que nous pouvons affirmer est que la voyelle est relativement simple à décrire par une décomposition harmonique approchée. La voyelle fait habituellement l'objet de descriptions simplifiées.

Le cas des consonnes n'est pas moins compliqué. Leur spectre n'est pas quantifiable en terme d'harmoniques puisqu'il varie en permanence et il existe des consonnes voisées, ce qui signifie qu'elles sont modulées par une onde pseudo-périodique ce qui en fait des sortes de voyelles bruitées. Enfin il y a dans la parole et plus particulièrement dans certaines langues, des impulsions, bruits très brefs qui doivent être traitée comme appartenant à une distribution de Dirac.

Consonnes et voyelles n'existent pas ou peu isolément. Elles sont influencées par leur environnement sonore et certains traits d'un son de parole peuvent traverser plusieurs sons voisins, se retrouver dans le sons précédent (anticipation), le suivant (persistance), les deux à la fois et même traverser plusieurs phonèmes. Au niveau des liaisons entre phonèmes interviennent des micro et macro-facteurs mélodiques où coarticulatoires qui sont importants dans la caractérisation des parlers.

Malgré ces imbrications complexes, la parole est une suite d'événements sous forme de chaîne de formes que nous savons reconnaître, donc nommer et classer. Plusieurs idées venant d'horizons différents suggèrent l'existence de possibilités de mécanisation de certaines analyses comme c'est le cas en biologie pour la description de l'ADN et de l'ARN, la recherche des virus et des bactéries, la chimie des molécules complexes et autres.

En conclusion de cette rapide et incomplète description nous pouvons affirmer que les langues sont des signaux totalement non stationnaires puisque même ceux qui sembleraient l'être (les voyelles) varient en permanence. C'est sur cette base que nous posons notre critique de l'usage de la théorie de Fourier et de ses dérivées comme la FFT et

tentons de proposer une nouvelle approche de l'onde sonore, l'approche temps/fréquence.

3.5.2 Le timbre d'un son

Le timbre ou couleur du son dépend de sa composition harmonique ou spectre. En phonétique on nomme formant (F1, F2, F3...) la réunion d'un groupe d'harmoniques présentant une intensité plus grande que ceux environnants. La distinction de ces groupes est importante car elle permet de caractériser le timbre du son et partant, de le nommer [voyelle V, consonne C] et de déterminer sa provenance : femme, homme, enfant, machine, instrument de musique, etc. Toutefois la définition du formant en tant que groupe d'harmoniques renforcés pose la question de savoir à partir de quel niveau d'intensité un harmonique doit être considéré comme pouvant être rattaché au groupe. L'autre question est de savoir si l'onde élémentaire incorporée automatiquement dans le formant appartient bien au signal de parole étudié.

Stabilité de timbre

Il est généralement admis que la voyelle est stable en fréquence et en timbre au point milieu de sa durée sur le deuxième formants. La plupart des mesures se font ainsi et les valeurs atteintes en ce point distinguent les voyelles entre elles. La définition précédente nous permet de poser que la stabilité du timbre d'un son est l'intervalle de temps où le rapport entre les formants F1, F2, F3, etc. est constant.

Dans la réalité cet intervalle de stabilité est soumis à de légères variations qui sont très facile à détecter par nos instruments. Se pose alors la question de savoir à partir de quel taux d'instabilité l'oreille perçoit une variation de timbre ? En effet, l'oreille ne peut détecter un changement que si l'excitation dépasse une certaine variation ou seuil différentiel. Selon Dodane [24], il serait plus approprié de définir la stabilité de timbre de la voyelle par exemple comme l'intervalle de temps où les rapports entre les trois premiers formants évoluent en dessous d'un certain seuil de perception. Au-delà de ce

seuil, l'auditeur perçoit un changement de timbre. Si nous sommes capables de délimiter la zone de stabilité de timbre, il est facile de délimiter la durée des autres sous-segments, c'est-à-dire des intervalles de temps correspondant à la transition initiale « tête » et la transition finale ou « queue » de la voyelle.

Notons que du fait des lois de décroissance de l'énergie acoustique dans l'atmosphère, les harmoniques les plus faibles disparaissent rapidement avec la distance à la source de telle sorte que le timbre d'un son varie avec cette distance. Le timbre du son perçu par l'interlocuteur varie avec sa distance au locuteur.

Chapitre 4

Décomposition du signal

La décomposition des signaux non-stationnaires en éléments plus simples et un domaine complexe qui relève des mathématiques supérieures et de certains domaines des statistiques et des probabilités. Depuis les années 1940, c'est la décomposition en série de Fourier à l'aide du sonagraphe qui est à la base des analyses harmoniques de la parole. C'est ce type d'analyse qui permet de suivre pas-à-pas l'évolution du signal sonore dans ses dimensions temporelles, fréquentielles et énergétique. Toutefois pour en arriver là, un certain nombre d'artifices doivent être utilisés pour adapter le principe de la décomposition harmonique de Fourier à l'objet d'étude réel.

4.0.3 L'enregistrement du signal acoustique

Le signal acoustique est un phénomène fugitif qu'il faut mémoriser pour pouvoir en étudier et comprendre l'évolution. Notre système nerveux pratique cette mémorisation de façon complexe tandis que les premiers enregistreurs mécaniques furent des systèmes analogiques ou l'onde sonore était enregistrée sous forme d'une forme matérielle analogue comme un sillon gravé dans la cire ou une matière plastique. Puis il y a eu les enregistreurs magnétiques ou l'information analogique était inscrite sous forme d'aimantation orientée de particules magnétiques noyées dans une colle couchée sur une

bande en matière plastique souple. L'enregistrement et la lecture se fait par induction magnétique. Pour mémoire, le premier magnétophone a utilisé un simple fil de fer pour enregistrer les sons, mais la continuité magnétique et électrique de ce matériau exigeait des longueurs extrêmement grandes de fil pour séparer les informations et des vitesses de défilement en rapport : les bobines faisaient environ deux mètre de diamètre pour quelques secondes d'enregistrement...

Les systèmes d'enregistrement/reproduction sonore analogiques ne permettent pas une reproduction fidèle du contenu informationnel, mais leurs défauts « naturels » les font toujours apprécier de certains mélomanes. À chaque utilisation, copie ou transfert, il y a usure du support donc modification des formes, dégradation et perte d'information. Ces pertes sont irréversibles car il n'est pas possible de connaître la loi permettant de retrouver l'original : toute tentative de correction abouti à un résultat pire que l'acceptation de l'évolution normale de ces supports.

C'est pourquoi dès que la technologie numérique a été rendue possible par l'évolution de l'électronique, le domaine du traitement numérique de l'information n'a cessé de s'étendre. L'avantage est, cette fois, la possibilité d'effectuer des calculs directement sur les données, de reproduire et de transmettre l'information sans perte avec une grande robustesse car des corrections calculées d'erreurs sont possibles. La robustesse tient au fait qu'il est possible de reconstituer un papier ou un livre très abîmés sur de simples indices et probabilités que nous sommes capables d'appliquer nous-mêmes, naturellement. En informatique, il est possible de reconstituer des fichiers inscrits sur un disque dur malgré l'effacement qui n'a en fait que diminué drastiquement l'intensité du champ magnétique des signaux inscrits : ces signaux étant des 0 ou des 1 cela facilite grandement l'opération. Pour un texte, un fragment de mot, de phrase permet de reconstituer le mot ou la phrase. La théorie de l'information montrait enfin sa puissance plus d'un demi siècle après avoir été formulée par Shanon, Thüring, Gabor, Heisenberg, etc.

4.1 Les systèmes de reconnaissance de la parole

Une idée naïve est de regarder du côté des techniques de reconnaissance de la parole. Nous avons analysé trois systèmes libres, donc utilisables gratuitement et au code ouvert.

Le premier, SIROCCO[1] fait partie du projet METISS de l'IRISA-INRIA. Ce système utilise des réseaux de neurones et une base TIMIT [38] anglophone regroupant dix enregistrements de 630 locuteurs. NTIMIT est la version destinée à la téléphonie de TIMIT. La reconnaissance repose sur la comparaison des éléments sonores de parole en entrée avec la base de données en utilisant des algorithmes accélérant le processus. On imagine déjà un procédé de reconnaissance que nous nommons par « force brute » par analogie à certaines techniques de décryptage des messages codés.

Dans le même style, nous avons testé NICO [2] qui est une boîte de composants informatiques destinés à la production de systèmes de reconnaissance vocale. Nous n'avons pas analysé le code de ces deux premiers systèmes car le troisième projet étudié, qui est directement utilisable et leur est très proche, est d'un accès moins technique (il n'y a pas de programmation à faire) et fonctionne parfaitement : il s'agit des programmes Sphinx 3 et 4 de l'université de Carnegie Mellon[3], basé sur les Modèles de Markov Cachés et également une base TIMIT.

4.1.1 Analyse du système Sphinx

Sphinx 4 (figure 4.1) est programmé en Java™ et ne présente pas de difficulté particulière pour être installé sur toute plateforme supportant Java. Les principales composantes de Sphinx-4 sont le frontal, le linguiste, et le décodeur. Une application interagit avec Sphinx 4 via le système Recognizer.

[1] Ce projet semble avoir disparu et son site est inacessible. Recherché le 10/09/2008
[2] http ://www.speech.kth.se/NICO/index.html
[3] http ://www.cmu.edu/index.shtml

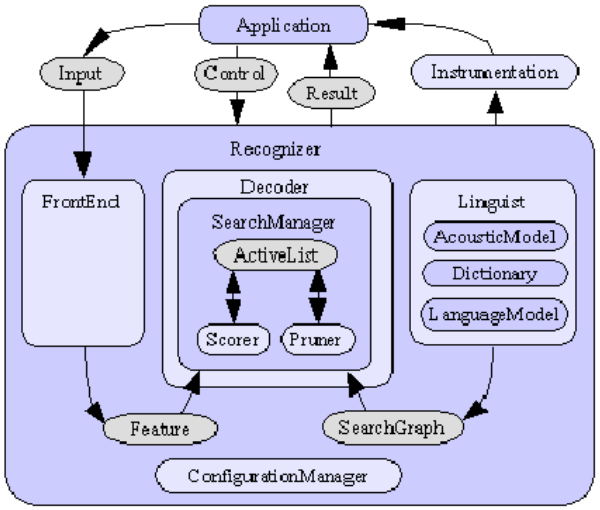

FIG. 4.1 – *Schéma synoptique de Sphinx 4, système de reconnaissance de la parole de l'université de Carnegie Mellon*

Schéma commenté de Sphinx 4

Audio : ce sont les données audio à décoder mais le système peut accepter des données spectrales ou cepstrales.

FrontEnd : c'est un DSP (processeur spécialisé, un logiciel dans le cas présent) qui effectue le traitement numérique du signal entrant ;

Features : il est utilisé pour estimer les paramètres acoustiques ;

Linguist : base de connaissances linguistiques du système, utilisé par le décodeur pour déterminer les mots ou phrases prononcées. Le linguiste produit un modèle graphique de la structure sur laquelle la recherche s'effectue en utilisant des algorithmes différents ;

Acoustic Model (modèle acoustique) : il contient une représentation (souvent statistiques) des sons, créés en utilisant beaucoup de données acoustiques ;

Dictionary – dictionnaire : il détermine le mot qui a été prononcé selon une probabilité ;

LanguageModel – Modèle de langue : contient une représentation probabiliste des

occurrences de formes, de mots ou d'expressions possibles dans la langue cible ;

SearchGraph : contient toutes les séquences possibles de phonèmes, diphones, triphones ou plus basées sur le Modèle de langue ;

Décodeur : il traite les informations reçues du FrontEnd, les analyse, les compare avec la base de données pour envoyer un résultat à l'application ;

Gestionnaire de Recherche : il effectue la recherche en utilisant certains algorithmes comme la recherche à droite, à gauche, en profondeur, par rapport aux meilleurs résultats antérieurs, etc. ;

ActiveListe : la liste des jetons représentant tous les états de la recherche graphique qui sont actifs dans le cadre actuel ;

Marqueur : partition de l'actuel cadre caractéristique actif à partir de l'ActiveList ;

Pruner : élague la liste active selon certaines stratégies ;

Résultat : dépouillement des résultats qui contient habituellement les N-meilleurs résultats ;

ConfigurationManager : charges les données de configuration de Sphinx à partir d'un fichier basé sur XML et gère le volet du cycle de vie des objets.

4.1.2 Discussion

L'analyse du système montre que Sphinx est un programme d'une grande complexité qui nécessite des éléments aux fonctions fort éloignées les unes des autres comme la base mécanique, les informations acoustiques, phonétiques et linguistiques implémentées. Plusieurs dictionnaires sont nécessaires et plusieurs modèles doivent être créés et comparés avec des patrons qui doivent être eux-mêmes créés à la volée.

Pour fonctionner, Sphinx a besoin de modèles acoustiques et d'un modèle de langue. Les modèles acoustiques sont basés sur une représentation statistique de séquences sonores contenues dans TIMIT ou NTIMIT. Ces bases doivent être crées pour chaque

langue [94] aussi n'avons nous pu faire des tests qu'avec un petit nombre de mots contenus dans la base réduite fournie avec le logiciel. Malgré notre mauvaise prononciation de l'anglais, les résultats sont assez bons.

Très gênant pour la recherche en phonétique, est le fait que la reconnaissance de la parole ne passe pas vraiment par la reconnaissance des sons. Ce sont des indices acoustiques qui sont pris en compte (modèles statistiques sonores) et la probabilité qu'ils soient bien à leur place dans une chaîne de sonore définie par le modèle de langue. Pour cela il n'y a pas besoin de précision temporelle et plusieurs mesures peuvent être faites sur un intervalle donné, la plus probable étant retenue. Ainsi si un [a] est détecté, ce peut-être en n'importe quel point de son évolution temporelle et si plusieurs points convenables sont trouvés sur un intervalle déterminé, la probabilité que ce soit un [a] croit. Le son est alors défini comme [a] avec une bonne probalitilité dans l'arborescence de recherche et il suffit considérer sa position par rapport à la suite de phonèmes environnants et de soumettre le résultat à un dictionnaire électronique phonétique mais de même style que les correcteur automatique de vocabulaire des traitement de texte et enfin de croiser le résultat avec le modèle de langue.

En conséquence les algorithmes utilisés en reconnaissance de la parole ne sont pas une voie exploitable pour l'étude phonétique ou l'on cherche à étudier avec précision la catégorie du son étudié, ses frontières temporelles et son timbre, etc. Les systèmes de reconnaissance vocales sont la conséquence des recherches en phonétiques, phonologie et linguistique générale et non l'inverse.

L'impression donné par des logiciels comme Sphinx tient dans leur réussite à reconnaître la parole. Même si la grande complexité de ce type de programme et de ses annexes se paye par un taux d'erreurs qui peut parfois être important, mais la plupart du temps ce n'est pas trop gênant, car les erreurs sont fréquentes et « naturelles » dans une situation de communication réelle.

4.2 Décomposition du signal sonore

L'étude précédente montre que nous avons procédé à une fouille dans les connaissances et les techniques actuelles en quête d'une hypothétique réponse à nos besoins. Nous constatons que la reconnaissance de la parole n'utilise pas de technique particulière susceptible de nous aider dans l'analyse du signal sonore. En fait il n'est pas possible de travailler directement sur le signal, ni sur les fichiers obtenus par numérisation de la pression sonore, car nous ne percevons pas ce signal, ni des bits ou des formants.

4.2.1 Analyse de l'information phonétique

Rappelons que la numérisation consiste à effectuer des mesures régulières dans le temps échantillonnage, ces mesures étant écrites dans un fichier au format binaire. L'échantillonnage doit se faire à une fréquence aussi précise et stable que possible à l'aide d'une horloge électronique pilotée par quartz ou mieux. Un tel système génère une quantité très grande d'informations binaires, de l'ordre 64.10^3 à 1536.10^3 bits/seconde, selon les systèmes actuels. Cette suite de bits ne présente pas d'information exploitable autrement qu'en la parcourant pour reproduire les sons originels grâce à la réversibilité des fonctions qui lui ont donné naissance, c'est-à-dire essentiellement la fréquence d'échantillonnage et le nombre de bits par échantillons.

La suite de 0 et de 1 obtenue ne laisse apparaître aucune structure particulière ni une quelconque relation avec le phénomène qui lui a donné naissance : le fichier ne peut être décrit que par lui-même ou alors au prix d'une simplification qui interdit irrémédiablement de retrouver l'original. C'est pourquoi, afin de réduire la dimension des fichiers sonores numérisés, les algorithmes utilisé généralement détruisent une partie de l'information présente. C'est le cas des compressions MP3™ ou Ogg[4] qui altèrent définitivement les fichiers sonores : il est impossible de reconstruire le fichier original à partir de ces dernier.

[4]Ogg est un format de compression à sources ouvertes

Il existe pourtant des algorithmes permettant de compresser des fichiers musicaux, mais ils ne semblent efficace que sur la musique. C'est le cas de Flac[5], logiciel Open source, qui permet de réduire le volume d'un morceau de musique jusqu'à 50% sans pertes ; son utilisation sur nos séquences de parole ne montre pas de compression significative et de toute façon cette voie est une impasse épistémologique.

En 1948, Claude Shannon a posé les bases de la théorie de l'information avec son article « A Mathematical Theory of Communications » [97]. Au niveau de l'acquisition des données, il démontre le théorème qui porte son nom et qui définit la fréquence minimale d'échantillonnage nécessaire pour numériser un phénomène périodique ou pseudo-périodique.

Théorème de Shanon : la fréquence d'échantillonnage d'un signal doit être au minimum le double de la fréquence harmonique la plus haute contenue dans le signal à échantillonner.

L'oreille humaine (jeune et en bon état) perçoit les fréquences qui vont de 20 à 20 000 Hz à environ 3dB près (fréquence de coupure), il faut donc échantillonner au minimum à 40 000 Hz pour pouvoir espérer reproduire une fréquence de 20 kHz. Pour un échantillonage à 40 kHz il y a un sérieux problème de phase entre le point d'échantillonnage et le signal à la fréquence maximale désirée puisque si l'échantillonnage tombe au passage à zéro, l'onde à 20 kHz sera toujours d'intensité zéro, sinon et si la mesure se fait à l'intensité maximum I_{max} de l'harmonique, celui-ci aura une forme rectangulaire d'intensité I_{max}, forme qui pose de sérieux problèmes de filtrage. Aussi pour limiter les effets indésirables la fréquence d'échantillonnage est choisie un peu plus grande que le double de la fréquence de coupure désirée, soit par exemple 44 100 Hz pour le disque numérique standard. Le filtre passe-bas appliqué au signal décodé est réglé sur une fréquence de coupure de 20 kHz.

[5]Flac : http ://flac.sourceforge.net/

Précision de la forme du signal

Un signal échantillonné se présente sous la forme d'une suite d'échelons au pas de l'échantillonnage. Le nombre de bits de mesure défini nombre de paliers possibles donc la précision du lissage entre les valeurs mesurées et le signal réel. Il est à l'origine d'un « bruit numérique » spécifique à la reproduction des enregistrement numériques et agit sur le timbre en modifiant le contenu harmonique. Dans certains cas l'échantillonnage est choisi délibérément bas (8 bits par exemple) avec sur-échantillonnage à la reproduction. Le sur-échantillonnage consiste à calculer la moyenne entre deux échantillons successifs au moment de la reproduction et d'intercaler cette valeur calculée entre les deux valeurs réelles. Cette technique permet une reproduction à une fréquence d'horloge double de la fréquence d'échantillonnage originale avec des échelons de durée divisée par deux, donc plus faciles à filtrer.

Qualité des fichiers sonores

Dans tous les cas la précision du signal et son rapport avec l'original est défini arbitrairement par les contraintes techniques qui sont imposées dès l'enregistrement, notament la fréquence d'échantillonnage et le nombre de bits par mesure. Quand nous nous référerons par la suite à une réversibilité parfaite d'un processus de décomposition, c'est d'une réversibilité vers ce que produit le fichier informatique original dont nous traiterons. En fait nous devrions dire que nos traitements n'ont pas altéré le produit du fichier original et lui seul car nous ne pouvons pas, par réversibilité retrouver directement les échantillons du fichier numérisé.

4.2.2 La décomposition harmonique de Fourier

La première formulation décrivant mathématiquement une onde stationnaire a été l'oeuvre de Joseph Fourier (1768-1830). Cette formulation a précédé de près d'un siècle ses applications pratiques. Au printemps 1927, Heisenberg énonçait son principe d'indé-

termination en physique quantique et Gabor le reformulait pour l'appliquer à l'acoustique où l'analyse des signaux non-stationnaires posait des problèmes théoriques et techniques ardus. C'est de ce dernier travail qu'est issu le sonagraphe.

Ces théories et les critiques les accompagnant contenaient également les germes de nouvelles formes d'analyse. Ces nouvelles formes d'analyse demandaient toutefois un pas technologique très important. À l'époque, la machine de Turing n'était qu'une entité abstraite, sans équivalent sous la forme d'ordinateur comme aujourd'hui, et les théorèmes concernant l'information de purs exercices théoriques. Il a fallut attendre les années 1980 pour que la puissance des ordinateurs et leur diffusion de masse puissent les faire entrer dans tous les laboratoires et toucher les chercheurs concernés en leur permettant de mettre en oeuvre des algorithmes inutilisable auparavant.

La première transformation a été de revoir le sonagraphe type « Key Elemetric »^{TM}, les oscilloscopes et oscillographes et de les écrire sous forme de programmes informatiques beaucoup plus souples que les machines dédiées antérieures. Nous pensons que les progrès faits sur sur le sonagraphe ont malheureusement bloqué les avancées qui accompagnaient, ailleurs, l'évolution des théories et des techniques de traitement du signal.

D'autres algorithmes ont vu le jour mais n'ont pas ou peu été utilisés dans le domaine de la parole. Ce sont des algorithmes dit « gloutons », c'est-à-dire consommant beaucoup de puissance machine dans des boucles récursives. Ils ont été jugé inutilisables dans le cadre de l'étude de la parole à une époque (1980-90) ou les ordinateurs n'était effectivement pas encore assez puissants pour les mettre en oeuvre facilement. Depuis ils sont restés dans l'ombre avec un préjugé défavorable ; c'est le cas de Matching Pursuit de Mallat [69]. Mais compte tenu de la loi de Moore[6], il nous semble qu'une révision de ces jugements doit être faite aujourd'hui. La loi de Moore s'est avérée relativement juste jusqu'à aujourd'hui : elle prévoit en effet un doublement du nombre de transistors dans les circuits intégrés environ tous les deux ans, donc une progression régulière de la puissance des micro-processeurs (de l'ordre de 20% sur cette période).

[6]Loi présentée en français en 1965 dans « Electronics Magazine », revue aujourd'hui disparue.

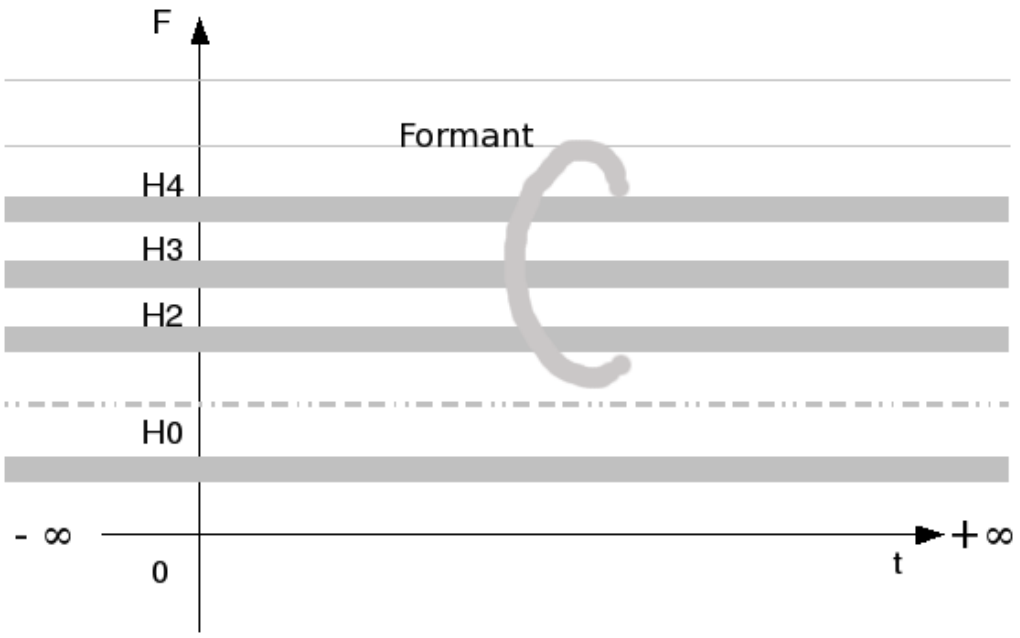

FIG. 4.2 – *exemple d'un son continu de fréquence fondamentale H0 avec trois harmoniques H2, H3, H4 que l'on peut rassembler pour constituer un formant.*

La transformée de Fourier continue (1807)

Nous traiterons en premier lieu des séries de Fourier et de leurs évolutions et utilisations.

Définition : toute fonction périodique continue, de forme quelconque, est décomposable en une série infinie de fonctions (ou harmoniques) de forme sinusoïdale et de fréquences multiples de la fréquence la plus basse contenue dans le signal, appelée fréquence fondamentale (c'est la note en musique).

La figure 4.2 donne l'exemple d'une représentation de type sonagramme pour un son continu. Le fondamental H0 et les harmoniques H2, H3, H4 ont des coéfficients de Fourier non nuls. L'hamonique 1 et tous les harmoniques de rang supérieur à 4 ont des coefficient de Fourier nuls.

Si T est la période du signal, c'est-à-dire le temps que la forme élémentaire le constituant met à redevenir semblable à elle-même et F sa fréquence, on a la relation $T = \frac{1}{F}$. Si A et B sont les coefficients de Fourier, l'intensité instantanée du signal en fonction du temps est donnée par l'équation :

$$i(t) = \sum_{n=0}^{n=+\infty} A_n \cos n\omega t + B_n \sin n\omega t \quad avec\ \omega = \frac{2\pi}{T} = 2\pi F$$

Si la série de Fourier à l'ordre n de f est la fonction obtenue en sommant les harmoniques successifs, on a :

$$f_n(t) = \sum_{k=-n}^{k=n} C_k e^{i\omega t} \quad avec\ \omega = \frac{2\pi}{T} = 2\pi F$$

et

$$f(t) = \sum_{-\infty}^{+\infty} C_k e^{i\omega t} avec\ \omega = \frac{2\pi}{T} = 2\pi F$$

Or il n'y a pas de signal stationnaire dans la nature autres que certaines approximations en physique comme la longueur d'onde spécifique d'un élément ou certains signaux produits par les astres. C'est pourquoi des alternatives aux séries de Fourier ont été cherchées. La transformée de Fourier en est une. Elle est définie par :

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt$$

Mathématiciens et physiciens ne s'accordent en général pas sur la définition de la transformée de Fourier et son écriture peut varier, de même que les représentations générées.

4.2.3 Conséquences

Les séries de Fourier permettent de décomposer des ondes complexes en leurs composantes les plus simples, c'est-à-dire sinusoïdales. Les séries de Fourier ne s'appliquent

en toute rigueur qu'aux ondes périodiques, c'est-à-dire strictement stables en forme et en fréquence, c'est-à-dire de durée infinie en raison des perturbations qui seraient engendrées s'il y avait un début et une fin du signal. Elles sont parfaites pour la détection des éléments chimiques en astrophysique puisque leur durée se compte en milliards d'années. Pour l'analyse de signaux non-stationnaires, elles ne permettent pas d'avoir une précision simultanément en temps et en fréquence (principe de Heisenberg).

4.3 L'ère numérique

L'arrivé des traitements numériques de l'information a balayé toutes les technologies antérieures, ou presque. Le passage très rapide des machines electro-mécaniques aux machines électroniques puis l'invention du microprocesseur ont rendu possible des travaux impensables sans eux. Nous avons assisté à ce profond bouleversement, tentant de l'accompagner malgré nos faibles moyens. Il nous a donc fallu remettre en cause de nombreuses fois nos savoirs et l'outil informatique nous amène désormais à traiter les fichiers numérisés par des transformées discrètes. La transformée de Fourier discrète (TFD) est un outil mathématique de traitement du signal numérique, qui est l'équivalent discret de la transformée de Fourier continue utilisée pour le traitement du signal analogique. Sa définition mathématique pour un signal s de N échantillons est la suivante :

$$S(k) = \sum_{n=0}^{N-1} s(n) \cdot e^{-2i\pi k \frac{n}{N}}$$

Des algorithmes de calcul rapides ont été cherchés et nous utilisons désormais la FFT (Fast Fourier Transform) ou transformée de Fourier rapide (TFR). C'est une transformée discrète qui s'écrit :

$$F_j = \sum_{n=0}^{N-1} f_n e^{\frac{-i2\pi j}{N}} \quad tel\ que\ j \in \{0, ...N-1\}$$

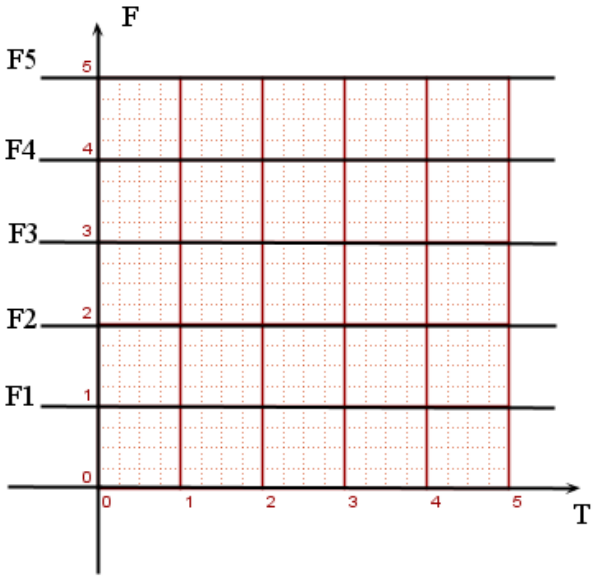

FIG. 4.3 – *Découpage du plan temps fréquence de Gabor, utilisé dans le sonagraphe.*

La FFT optimise les calculs sur une fenêtre temporelle qui est nécessairement une puissance d'un nombre d'échantillons k. On prend généralement k=2 et la fenêtre mesure généralement entre 2^7 et 2^{12} échantillons. Les zones de la fenêtre qui ne contiennent pas de signal sont simplement remplies de zéros et elle admet un certain nombre d'aproximations. La vitesse de calcul peut être 100 fois plus rapide que celle nécessaire à la TFD. Cet algorithme est couramment utilisé pour l'analyse numérique du signal du signal de parole ; il est implémenté notamment dans les analyseurs de spectre comme Praat, SFS et bien d'autres.

4.3.1 Le sonagraphe

Les premiers sonagraphes étaient des machines mécaniques traitant le signal par un filtre passe bande analogique qui balayait la largeur de la bande passante du signal. Deux types de filtres peuvent être utilisés, l'un à bande étroite qui permet de produire une image de tous les harmoniques présents dans le signal, l'autre, à bande large, qui

montre les groupes d'harmoniques ou formants.

Le sonagraphe repose sur les séries de Fourier et les idées de Gabor qui à permis l'élaboration de l'appareil que nous connaissons ; le plan temps–fréquence utilisé est celui de Gabor, (fig : 4.3).

Aujourd'hui les éléments analogiques ont été entièrement remplacés par des fonctions numériques, de l'enregistrement des sons au traitement par des transformées FFT : le sonagraphe est devenu un programme informatique proposé gratuitement par plusieurs auteurs qui s'il simplifie beaucoup d'opérations dans l'étude la parole, maintient des principes et donne des résultats équivalent à ceux des machines analogiques antérieures.

Nous avons dit plus haut que la FFT était un filtre passe-bande dont on pouvait faire varier la largeur. Nous utilisons le plus souvent un filtrage large qui donne des représentations ou empreintes très parlantes du signal de parole (sonagramme). L'importance de la largeur de bande du filtre fait perdre toute précision en fréquence, mais conserve la précision temporelle comme me prévoit le principe d'Heisenberg.

L'avantage incontournable de la FFT et du sonagraphe se tient dans des décennies d'expérimentation et une expertise mondiale immense : les autres approches peinent à pénétrer le milieu des phonéticiens, surtout celles qui semblent ésotériques comme la décomposition atomique du son.

Les inconvénients de l'analyse sonagraphique

La figure 4.4 donne un exemple d'échec de la détection automatique du deuxième formant de [a] qui est celui qui nous intéresse. L'examen visuel des deux premiers formants de V_2 ne justifie pas, à priori, une telle déviation du suivi formantique d'autant plus qu'il est correct pour V_1

Le nombre important d'erreurs de ce type nous a fait renoncer à toute détection automatique...

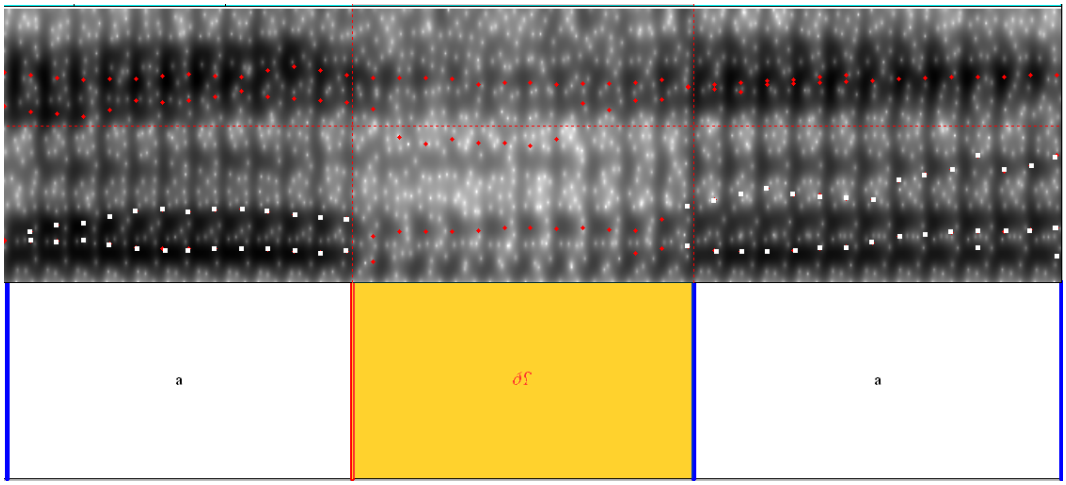

FIG. 4.4 – *La détection automatique échoue sur le deuxième formant de* V_2

4.3.2 Autres approches

L'analyse du signal faisant l'objet de recherches intensives, de nombreuses méthodes ont été proposées. Citons l'analyse cesptrale [2], les analyses temps-fréquence avec les distributions de Wigner-Ville et les transformées en ondelettes. Toutes ces techniques peuvent être approchées par les livres ou mieux pour nous, par les programmes disponibles sur internet.

Les ondelettes

Face aux critiques concernant la TFD qui découpe le plan temps/fréquence en bandes et au principe d'incertitude d'Heisenberg sur l'impossibilité de connaître à la fois le temps et la fréquence, Gabor a proposé d'améliorer la précision par un quadrillage régulier de la carte temps/fréquence (fig : 4.3), chaque case pouvant contenir une forme ou sorte de corpuscule appelé atome ou gaborette. La critique porte cette fois sur ce quadrillage qui traite avec la même résolution les basses et les hautes fréquences alors que les hautes fréquences contiennent un nombre de formes plus grand que les basses. C'est pourquoi un premier perfectionnement est venu corriger ce problème avec les ondelettes

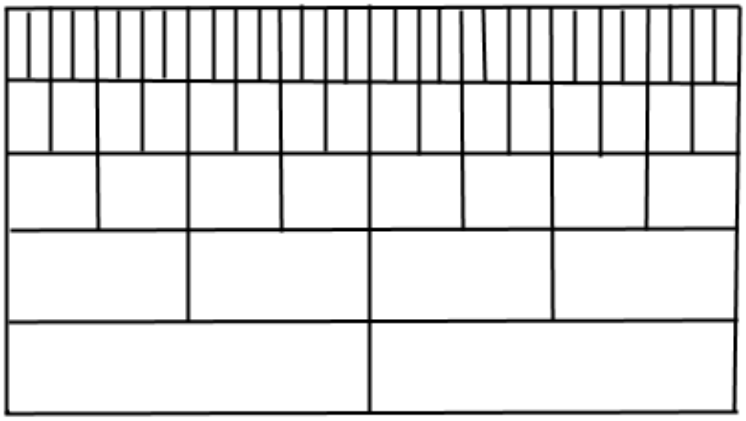

FIG. 4.5 – *Le plan temps fréquence pour la description du signal par les ondelettes*

(ou wavelets),.

En établissant un quadrillage variant en fonction de la fréquence (figure 4.5), plus large pour les basses et plus étroit pour les hautes fréquences, le tout associé avec une famille de petites ondes adaptées, on obtient une plus grande précision dans la description du signal, même si la représentation demande plus de calculs, qu'elle est moins parcimonieuse que précédemment. Toutefois, en consultant la littérature au sujet des ondelettes, nous observons que les chercheurs ont créé une grande variété de familles d'ondelettes adaptées à chaque cas spécifique. Cette démarche va à l'encontre de la nôtre qui est d'obtenir un système le plus simple et général possible capable de traiter la variabilité du signal de parole. Nous remarquons aussi qu'à ce point d'évolution de notre démarche toutes les méthodes étudiées proposent un découpage rigide du plan temps/fréquence.

Un pas supplémentaire devait être fait avec l'idée d'un algorithme suffisamment souple, mais requérant des machines puissantes, appelé Matching Pursuit ou décomposition adaptative des ondes. On peut en effet imaginer avoir un découpage du plan temps/fréquence adapté à la situation locale, de même que des atomes appartenant à plusieurs familles et, eux aussi, adaptables aux conditions locales. Nous obtenons alors les avantages des ondelettes sans en avoir les inconvénients, la complexité supplémentaire

étant gérée par le programme informatique effectuant les calculs.

4.4 Matching Pursuit ou décomposition atomique du signal

Avec Matching Poursuit, nous entrons dans une nouvelle génération de systèmes d'analyse du signal. C'est une méthode temps-fréquence qui devrait être considérée comme un choix intéressant pour l'analyse d'un signal aussi complexe que celui de la parole, lequel révèle des caractéristiques rythmiques et transitoires nombreuses et variées. MP, ou Poursuite Assortie, ou encore Analyse Granulaire Adaptative, offre des avantages par rapport aux autres méthodes d'analyse : sa haute résolution, son adaptativité locale à des structures transitoires et de plus les représentations graphiques obtenues sont compréhensibles.

Malgré ces caractéristiques intéressantes, son application généralisée est limitée par la complexité de l'algorithme qui réclame de grandes puissances de calcul et un grand nombre de boucles. MP peut exiger du temps de calcul : il s'agit d'un ensemble d'algorithmes récursifs dits « gloutons » pour la raison précitée, ce qui entraîne une intense recherche en informatique pour les rendre plus rapides. On dit que l'on recherche des algorithmes parcimonieux. Un bref parcours du site METISS de l'IRISA[7] montre que la parcimonie est au cœur des préoccupations des chercheurs en informatique travaillant sur MP et d'autres algorithmes gloutons.

L'algorithme Matching Pursuit est un outil d'analyse de signaux extrêmement flexible dont les principales propriétés théoriques montrent un fort potentiel pour le codage, le débruitage, la séparation de sources et l'extraction de descripteurs pour la reconnaissance de formes. Matching Poursuit généralise les transformées de Fourier et transformées en ondelettes en s'affranchissant de leurs limitations. Pour décomposer les signaux, il s'appuie sur des dictionnaires et des algorithmes adaptatifs parcimonieux.

[7] http ://www.irisa.fr/activites/equipes/metiss

Ces propriétés exceptionnelles nous ont incité à nous pencher sur les différents programmes proposant cet algorithme. Nous avons testé les possibilités de MP sur le site de P. Durkas[8] qui présente des l'analyses d'électroencéphalogrammes, puis avec Lastwave[9], avec MPTK[10] et enfin Guimauve[11]. MP est un puissant système d'analyse des signaux non-stationnaires dont les applications ne sont limitées que par l'imagination de ses utilisateurs potentiels.

4.4.1 La dualité onde-corpuscule

L'idée de présenter l'onde sonore sous forme de corpuscule choque encore beaucoup de chercheurs, d'ingénieurs et de techniciens (section 4.8, page 81). Rappelons que depuis longtemps la physique est confrontée à la dualité ondulatoire et corpusculaire de la lumière et que récemment, un élément aussi gros à l'échelle des particules élémentaires que l'atome, a pu être vu en tant qu'onde en laboratoire.

MP décompose une onde en corpuscules sous forme d'atomes. Ces atomes sont des « grains énergétiques » caractérisés par une fréquence (en physique on dirait leur longueur d'onde) et une position dans l'espace temps-fréquence. Les atomes les plus couramment utilisés sont les atomes de Gabor qui sont une sinusoïde modulée par une gaussienne (fig : 4.6) mais toute autre forme peut convenir si elle est adaptée au signal à traiter. Ces atomes de base sont placés dans un dictionnaire de formes qui peut être très simple. Un dictionnaire peut contenir des atomes gaussiens, des atomes de Dirac qui permettent de rendre compte des impulsions contenues dans le signal et d'autres encore.

[8] http ://brain.fuw.edu.pl/d̃urka/
[9] http ://www.cmap.polytechnique.fr/b̃acry/LastWave/
[10] http ://gforge.inria.fr/projects/mptk/
[11] http ://webast.ast.obs-mip.fr/people/fbracher/

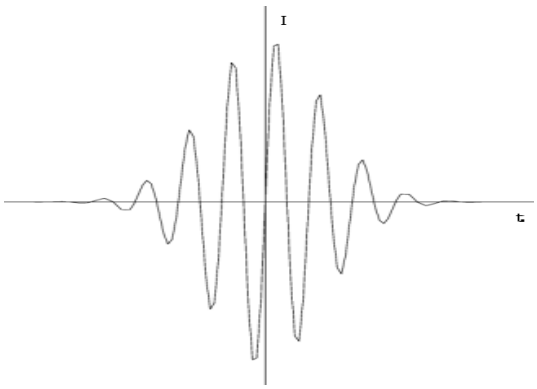

FIG. 4.6 – *fonction ou atome de Gabor simplifié,* $y = 4sin(12x)e^{-x^2}$

Le dictionnaire d'atome

Les atomes les plus usités sont les atomes de Gabor car ils offrent une localisation temps-fréquence optimale. La fonction de Gabor généralisée peut s'écrire ainsi :

$$g_{\gamma(t)} = K(\gamma)e^{-\pi(\frac{(t-u)}{n})^2}\sin(2\pi\frac{\omega}{N}(t-u)+\phi)$$

ou N est la taille du signal pour lequel le dictionnaire est construit et $\dot{K}$ est tel que $\|g_\gamma\| = 1$. Les paramètres du dictionnaire de fonctions temps-fréquence des atomes sont données par $\gamma = \{u, \omega, s, \phi\}$. La durée du signal (N points), propose des zones où les paramètres des fonctions de Gabor peuvent être raisonnablement monté à un moment donné. Toutefois, aucune plages d'échantillonnage n'est a priori définie, et nous sommes confrontés à un espace continu tridimensionnel qui se traduit par une infinité de taille de dictionnaires possible. Par conséquent, dans la pratique, nous utiliserons des sous-ensembles de l'éventuel dictionnaire de fonctions.

Dans le dictionnaire mis en oeuvre initialement par Mallat et Zhang [69], les paramètres des atomes sont choisis parmi des séquences dyadiques d'entiers. Leur prélèvement est gouverné par un paramètre supplémentaire – octave j, (entier). s correspond à la largeur de l'atome dans le temps, c'est la dérivé de la séquence dyadique

$s = 2^j, 0 \leq j \leq L$ pour un signal de taille $N = 2^L$. Les paramètres $\dot{u}$ et ω, qui correspondent à la position de l'atome dans le temps et à la fréquence donnée, sont échantillonnés pour chaque octave j, avec un intervalle $\dot{s} = 2^j$, ou, si un sur-échantillonnage est introduit, avec l'intervalle 2^{j-se}.

Pour un signal donné, on cherche dans un dictionnaire de formes celle qui représente le maximum d'énergie du signal, on la soustrait du signal et on écrit cette forme dans un livre. On répète l'opération sur le signal résiduel jusqu'à un point choisi en fonction de la profondeur de décomposition voulu. À la fin nous disposons d'un livre contenant les atomes extraits et un résidu qui est toujours une onde sonore.

En faisant la somme des atomes du livre on peut recomposer le signal approché. En faisant la somme des atomes avec le résidu, on reconstruit exactement le signal original (même qualité). Contrairement aux autres méthodes de décomposition ou de compression des signaux, MP est parfaitement réversible, ce qui est important pour le traitement de l'information comme nous le verrons plus loin.

MP produit des livres où les atomes trouvés sont répertoriés, avec leurs durée, leur fréquence, leur énergie et leur position dans le temps, ce qui permet de créer la carte temps $\times$ fréquence du signal considéré. Cette carte peut être générée en deux ou trois dimensions ce qui donne une représentation très similaire aux images très médiatisées de la physique des particules.

Le mécanisme du fonctionnement de MP se trouve décrit dans les lignes qui suivent. Étant donné une série de fonctions contenues dans un dictionnaires $D = \{g1, g2, ..., gn\}$ tel que $\|g_i\| = 1$, nous pouvons définir une approximation comme une expansion optimale M-, en minimisant l'erreur ε d'une approximation du signal $f(t)$ par M ondes :

$$\epsilon = |f(t) - \sum_{i=1}^{M} \omega_i g_{\gamma i}(t)|$$

où γi tel que $i = 1, \ldots, M$, représente les indices des fonctions choisies $g_{\gamma i}$.

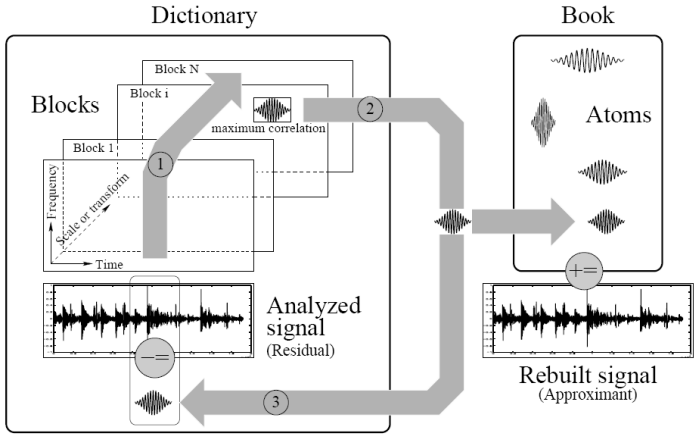

FIG. 4.7 – *synoptique de l'algorithme Mathing Pursuit.*

Trouver une telle approximation optimale $NP-$ est un problème difficile à résoudre. La solution existe par le biais d'une procédure itérative, telle que l'algorithme de recherche MP. Dans la première étape de MP, on choisit l'onde $g_{\gamma 0}$. qui correspond le mieux au signal $f(t)$. Dans chacune des étapes suivantes, on obtient une onde $g_{\gamma n}$ correspondant au mieux au signal résiduel après avoir soustrait les résultats de l'itérations précédente :

$$\begin{cases} R^0 f = f \\ R^n f =< R^n f, g_{\gamma n} > g_{\gamma n} + R^{n+1} f \\ g_{\gamma i} = argmax \in D \mid R^n f, g_{\gamma n} > | \end{cases}$$

L'orthogonalité de $R^{n-1} f$ et $g_{\gamma n}$ à chaque étape implique la conservation de l'énergie.

Pour un dictionnaire complet la procédure converge vers :

$$f = \sum_{n=0}^{\infty} < R^n f, g_{\gamma n} > g_{\gamma n}$$

De cette équation nous pouvons dériver une distribution temps–fréquence de l'énergie du signal :

$$Ef(t,\omega) = \sum_{n=0}^{\infty} |< R^n f, g_{\gamma n} >|^2 \, W g_{\gamma n}(t,\omega)$$

Par souci de parcimonie, il est nécessaire d'éviter autant que possible de faire deux fois la même opération [54]. Par exemple, dans les blocs, la mise à jour de la transformée temps-fréquence est exécutée seulement le long de la partie du signal qui a été modifiée en soustrayant un atome à la passe précédente. Les résultats sont stockés en tant que vecteurs une seule dimension. Après chaque mise à jour de cette transformée, le lieu de corrélation maximum peut être n'importe où le long du vecteur aussi pour éviter de parcourir chaque fois toute la longueur pour une recherche de maximum de corrélation, une structure arborescente garde une trace des maxima locaux. Seules les parties grisées, modifiées par l'opération précédente, sont examinées (Fig : 4.8).

4.5 Résultats obtenus avec Matching Pursuit

Des travaux montrent qu'il est possible de reconnaître des formes avec MP, et dans le projet METISS, un volet important est dédié à la séparation aveugle de sources mélangées dans un signal monophonique voire polyphonique [61].

Nous avons testé les logiciels Lastwave et MPTK, programmés par le même groupe de chercheurs. Il existe aussi des boîtes à outils libres pour Matlab™ [3] ou Octave qui est un logiciel libre mais nous n'avons pas réussi à compiler le visualisateur Octaviz qui nous promettait des graphiques de qualité optimale. Finalement nous avons découvert les possibilités de Guimauve de Fabien Brachère du Laboratoire d'astrophysique de Toulouse. Guimauve ne conserve que le module MP de Lastwave et offre les fonctions essentielles que nous souhaitions expérimenter, c'est-à-dire une sortie graphique en deux et trois dimensions et une sortie des caractéristiques des atomes sous forme de textes.

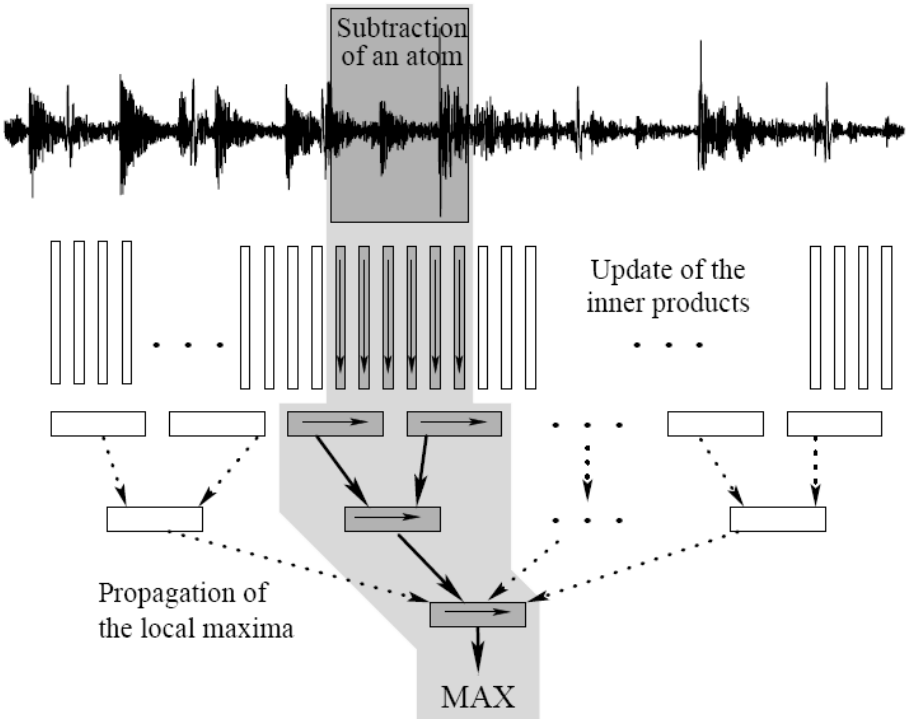

FIG. 4.8 – *schéma montrant l'algorithme de recherche de MP*

Pour cette expérience, nous avons pris trois occurrences de chaque séquence VCV des fichiers étudiés dans le cadre de la coarticulation et nous leurs avons appliqué une décomposition atomique avec Guimauve. Nous obtenons alors des fichiers contenant les atomes classés dans l'ordre inverse de leur contenu énergétique. Ensuite nous avons cherché une méthode de classification capable de discriminer les séquences pharingalisées vs pharyngalisées. Pour simplifier ce travail d'approche, nous avons appliqué la méthode à un seul locuteur de chacune des quatre grandes régions arabophones choisies.

Le processus est alors le suivant :

1. extractions des séquences VCV au format binaire (.wav) à partir de Praat ;
2. conversion des données binaires au format texte :
3. le fichier .asc résultant est alors traité par Guimauve.

Vu le nombre de fichiers à traiter et pour éviter les erreurs nous avons automatisé tout ce qui pouvait entrer dans des boucles en Perl, ce qui fait que seul l'envoi des fichiers vers Guimauve a été fait un par un. Nous montrons ci-dessous le résultat d'une décomposition sous forme de carte temps/fréquence et quelques possibilités de MP concernant la manipulation des atomes. MP permet en effet d'éliminer ou d'ajouter des atomes au livre du signal, ce qui fait dire aux auteurs de MP, que ce livre est indépendant du signal. Parmi les possibilités intéressantes de MP, il y a l'élimination du ou des bruits, comme un bruit continu ou des impulsions parasites (Dirac).

4.5.1 Les représentations graphiques

La figure 4.9 donne la carte complète des 200 atomes calculés sur une séquence as$^{\varsigma}$a. Il est possible de sélectionner des atomes pour construire un signal à partir de ceux-ci (figure 4.10) ou de les retirer (figure 4.11) du signal pour éliminer des bruits par exemple. Ces possibilités n'ont pas encore été exploitées en phonétique à notre connaissance.

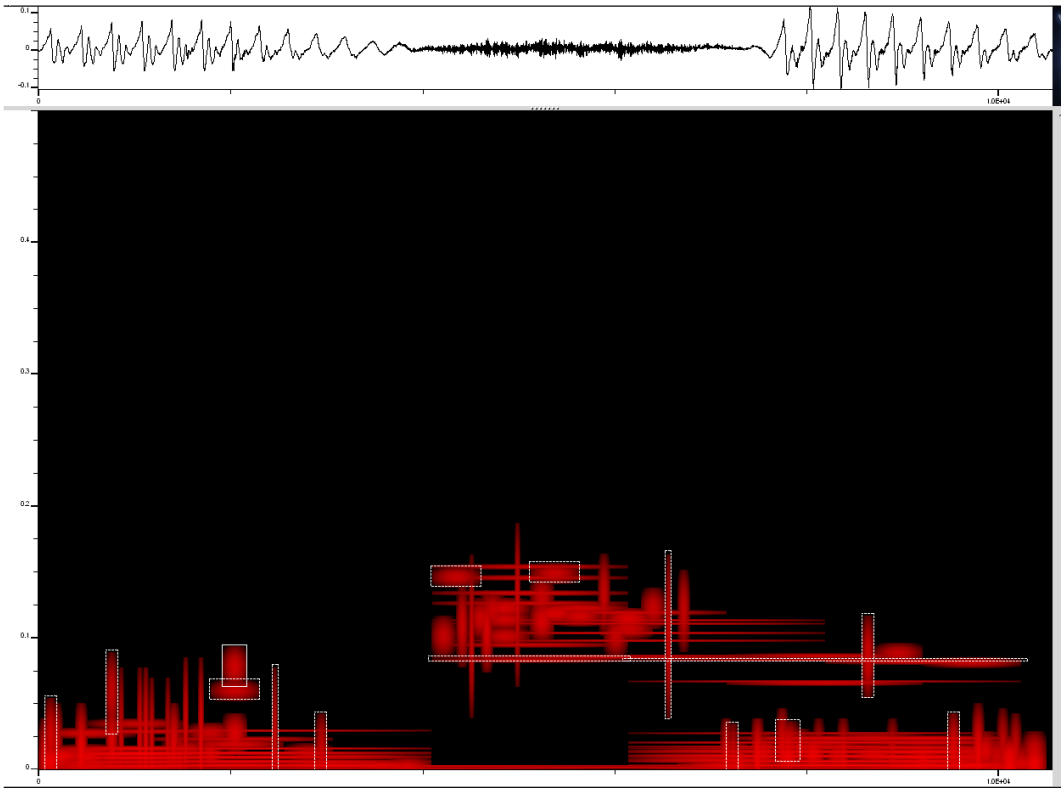

FIG. 4.9 – *carte temps-fréquence des 200 atomes extrait d'une séquence* [as$^{\text{ʕ}}$a]

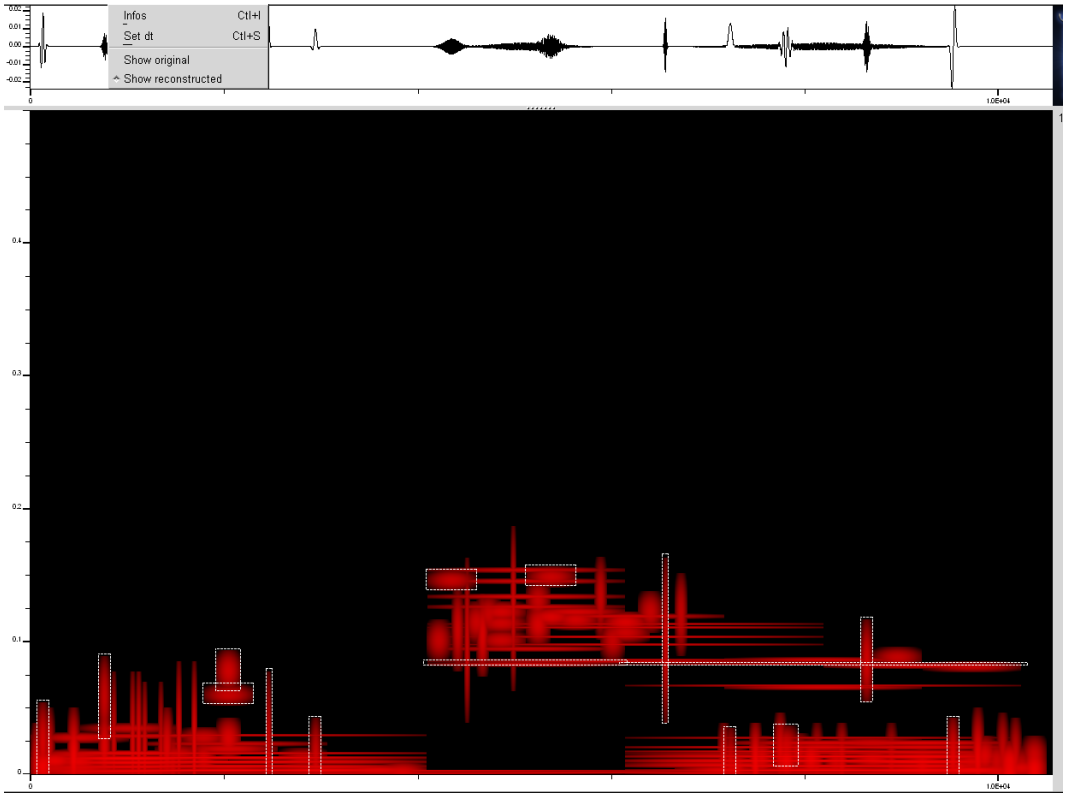

FIG. 4.10 – *sélection de quelques atomes sur le même signal* [as$^{\int}$a]

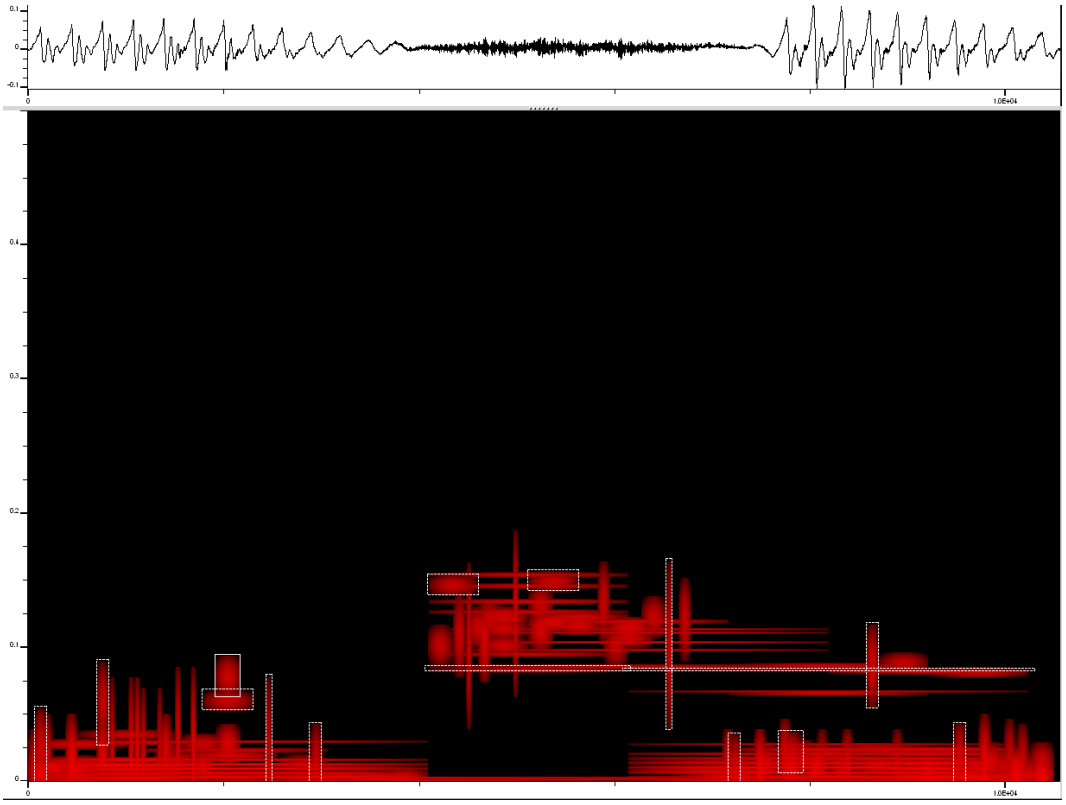

FIG. 4.11 – *soustraction de quelques atomes du signal*[as$^{\text{f}}$a]

4.5.2 Les données littérales

La décomposition avec Guimauve nous donne des fichiers d'atomes sous forme d'un texte formaté facile à lire et à comprendre. Exemple :

```
atom #:1, coeff2: 0.286017, octave:8, time: 9600.000000, frequency: 0.007812
atom #:2, coeff2: 0.273861, octave:8, time: 9344.000000, frequency: 0.009766
atom #:3, coeff2: 0.176226, octave:7, time: 9920.000000, frequency: 0.011719
atom #:4, coeff2: 0.168513, octave:8, time: 2048.000000, frequency: 0.009766
atom #:5, coeff2: 0.156791, octave:8, time: 10240.000000, frequency: 0.011719
atom #:6, coeff2: 0.153241, octave:8, time: 2304.000000, frequency: 0.007812
atom #:7, coeff2: 0.144899, octave:7, time: 9024.000000, frequency: 0.007812
atom #:8, coeff2: 0.128123, octave:7, time: 2880.000000, frequency: 0.007812
atom #:9, coeff2: 0.121010, octave:8, time: 2560.000000, frequency: 0.005859
```

Avec :

atom # : le numéro de l'atome ;

coeff2 : l'énergie de l'atome (le premier atome est le plus énergétique) ;

octave : l'octave de l'atome ;

time : la position temporelle de l'atome ;

frequency : la féquence de l'atome ;

Nous avons modifé ces textes avec des progammes en Perl [96], en supprimant les étiquettes qui sont inutiles puisque nous savons à quoi correspond chaque colonne, et certaines colonnes en fonction des expériences effectuées.

4.5.3 Représentations temps-fréquence

Guimauve permet des représentations en deux ou trois dimensions, mais les fichiers d'atomes peuvent recevoir d'autres traitements. Pour chaque séquence VCV, nous avons choisi une décompossition en 200 atomes. Ce choix nous a donné de bons résultats et nous l'avons conservé, mais rien ne prouve qu'une autre valeur ne serait pas plus optimale. Parallèlement nous avons établi une discussion avec l'auteur du logiciel Guimauve qui nous a fourni des idées fondamentales sur l'usage de MP et proposé une amélioration du code de son logiciel afin d'obtenir en sortien un format de fichier directement exploitable.

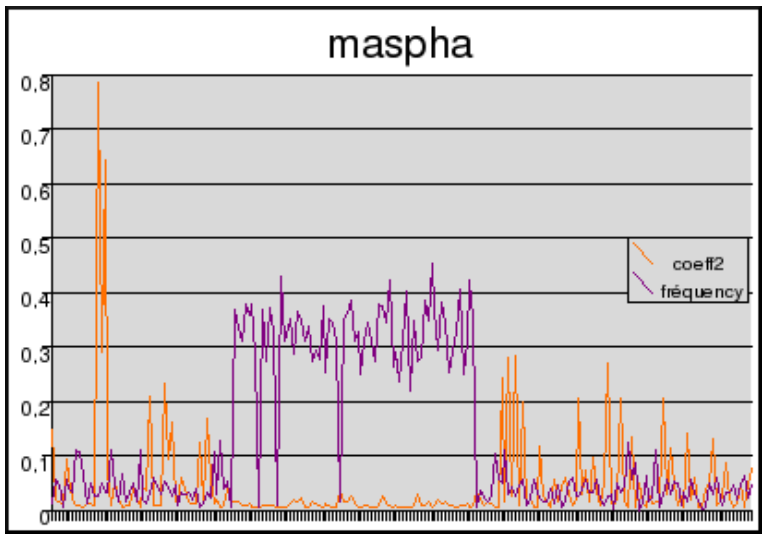

FIG. 4.12 – *Évolution de l'énergie et de la fréquence des atomes de* ad$^{\varsigma}$a.

4.6 Présentation des résultats

Nous ne connaissions pas de méthode capable d'exploiter les listes d'atomes extraites de chaque séquence VCV et de les comparer entre elles. Nous ne connaissions des systèmes comme Weka [121] ou R [113] que par des tests superficiels et des exemples donnés dans des articles. Nous avons également étudié les possibilités de la logique floue [45] mais cette théorie mathématique, qui connaît quelques applications pratiques en électronique pour la conduite de machines, ne nous a pas apporté de solution, du moins pour l'instant.

4.6.1 Présentation des données temps-fréquence avec un tableur

Curieusement pour nous, c'est avec un simple tableur que nous avons commencé à récolter des informations intéressantes. Le graphique 4.12 montre que la fréquence des voyelles, courbe foncée, est basse, tandis que la consonne au centre contient des fréquences élevées. En ce qui concerne l'énergie des atomes, celle-ci est élevée pour les voyelles, représentée par des pics importants, alors qu'au niveau de la consonne, leur énergie est faible. Cette observation est conforme à la théorie, valide les potentialités

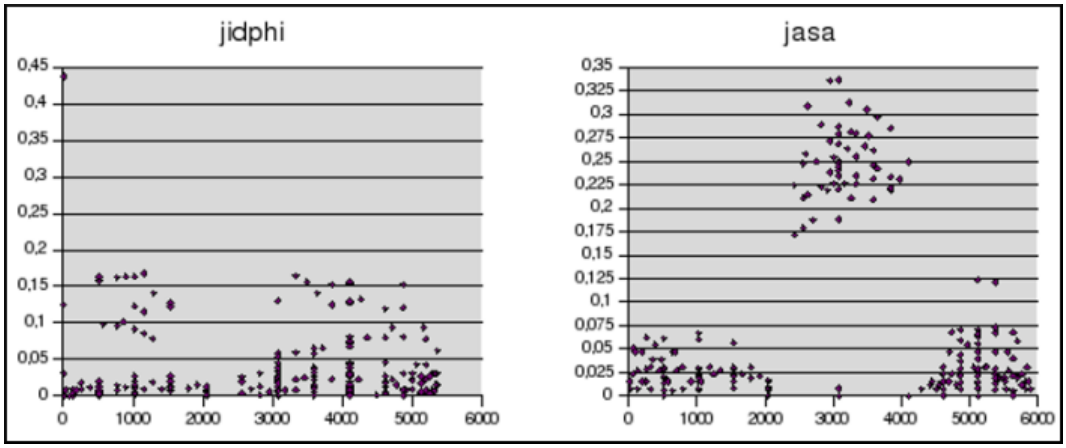

FIG. 4.13 – *Positionnement temps-fréquence des atomes avec un tableur*

MP comme outil d'analyse et nous donne une représentation nouvelle de la séquence VCV.

Autre point particulièrement intéressant, nous obtenons une vision inédite de la transition VC et CV, qui si nous n'en avons pas encore fait une analyse poussée devrait nous donner des indications fort précieuses sur ce qui se passe à ce niveau. En changeant simplement de représentation, nous avons obtenons le graphique par points suivant avec idphi pour [ad$^{\varsigma}$a] et [asa].

Seules les valeurs temps et fréquences sont prises en compte : nous observons donc uniquement les positions des centres des atomes. Pour aller plus loin, il faudrait pour cela un logiciel capable de prendre en compte l'intensité de ces atomes, de distinguer les différentes zones de regroupement appelées généralement clusters ou de les caractériser de façon concise, par une régression ou toute autre manière permettant de donner une signature aux nuages de points obtenus. Une autre direction que nous n'avons pas approfondie est celle des centres de gravité. Nous avons préféré nous orienter vers des logiciels de fouille de données ; chaque séquence VCV est composée de 200 atomes qui devraient pouvoir être réunis dans trois groupes ou clusters ce qui est cas d'école banal en calcul de probabilités.

4.6.2 Analyse en clusters avec Weka

Les graphiques 4.12, 4.13 montrent clairement les trois zones V, C et V. En remplaçant le graphique par traits par un graphique par points on obtient une représentation beaucoup plus parlante, mais les tableurs ne contiennent pas à notre connaissance de fonction permettant de regrouper ces points en familles ou clusters. Il y a des outils spécialisés pour cela et nous avons testé Weka.

Weka est un outil de fouille de données ou « Data mining » qui permet le classement, le tri, le rassemblement (clustering) des données. La fouille des données à été utilisée pour la cassifiction de variétés de langues [118]. Weka offre de nombreux algorithmes et il est très ergonomique, permettant de passer instantanément d'une vision d'un ensmble de résultats à une autre. Il permet un fonctionnement en ligne de commande, donc l'exécution de scripts qui permettent de réaliser des opérations complexes. Pour notre problématique nous avons appliqué la fonction de clustering KM (K-moyennes) sur les 200 atomes de nos fichiers. Weka propose un graphique qui montre trois nuages de points bien séparés pour une séquence [asa] ainsi que le résultat détaillé de ses calculs.

```
=== Run information ===
Scheme:       weka.clusterers.SimpleKMeans -N 3 -S 10
Relation:     j-a-s-a-2.gui.asc
Instances:    200
Attributes:   2
              time
              frequency
Test mode:    evaluate on training data

=== Model and evaluation on training set ===

kMeans
======
Number of iterations: 4
Within cluster sum of squared errors: 3.44884590874933
Cluster centroids:

Cluster 0
Mean/Mode:  2854.1099    0.2559
Std Devs:    527.8304    0.0398
Cluster 1
```

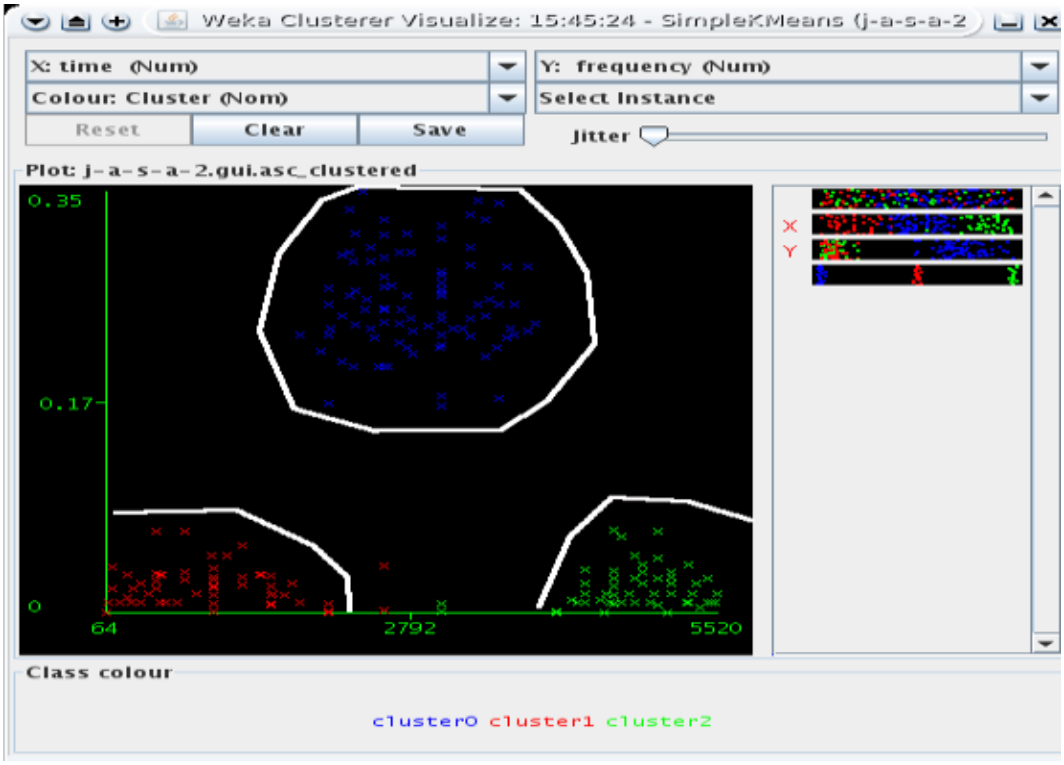

FIG. 4.14 – *Clustering avec Weka, ici les atomes sont correctement regroupés.*

```
Mean/Mode:  1018.3448    0.0209
Std Devs:    624.6978    0.0154
Cluster 2
Mean/Mode:  4779.1373    0.0198
Std Devs:    518.126     0.0168

Clustered Instances

0       91 ( 46%)
1       58 ( 29%)
2       51 ( 26%)
```

Un grand nombre de tests sur nos données montre que Weka détermine automatiquement souvent plus de trois clusters. Il est alors nécessaire de le forcer à rassembler les fréquences en trois clusters (analyse supervisée). Dans le rapport précédent, le résultat numérique est correct, mais lorsque Weka rassemble des points chevauchant les trois zones, les clusters sont inutilisables pour une analyse. La méthode EM donne les mêmes résultats que la méthode des K-moyennes employée ci-dessus. EM est un algorithme différent de KM qui permet aussi de former des groupes.

En conclusion, Weka permet d'obtenir le graphe des nuages de points plus facilement qu'un tableur car il n'y a pas besoin de d'élaguer le fichier de données, mais il manque de fiabilité quand il s'agit de former correctement des clusters. Ce manque fiabilité nous a poussé a stopper notre recherche avec Weka, bien que nous ayons quelques idées sur les voies à explorer avec ce logiciel.

4.6.3 Autres logiciels testés

FindFraud et Baldr sont des logiciels destinés au départ à découvrir la fraude aux examens rendus sous forme de fichiers informatiques. À la différence de Weka qui traite un fichier de données à la fois, nous avons ici affaire à des systèmes qui cherchent la similitude entre des ensembles de fichiers en se basant sur la compression de ceux-ci. Les questions de compression et de similitude ressortissent directement de la théorie de

l'information présentée plus haut dans ce chapitre. Le compresseur doit être sans pertes ce qui exclu l'usage des logiciels de compression de musique ou d'images. La similitude mathématique est une notion délicate qui ne sera pas traitée ici, si tant est que nous en ayons la compétence.

Les résultats obtenus avec FindFraud et Baldr ne sont pas significatifs car tous deux trouvent que les fichiers issus de Guimauve sont tous différents deux à deux. Pour FindFraud, l'échelle d'appréciation de la similitude va de 0 à 1 avec égalité = 0 et totalement différent = 1. Sur notre corpus, FindFraud donne des valeurs allant de 0,6 à 0,95 avec impossibilité de faire des regroupement. De même Baldr, qui fournit un graphique et une table des correspondances, ne détecte pas de similitude autre qu'entre deux fichiers égaux de notre répertoire. En conséquence, nous avons écarté provisoirement ces deux programmes.

4.7 Discussion

La complexité est au cœur de tout travail scientifique et la recherche de méthodes pour la réduire fait partie des tâches essentielles du chercheur. La décomposition en séries de Fourier est une des premières méthode utilisée pour approcher et comprendre un signal périodique complexe mais stable.

Pour résoudre la question des signaux de durée finie, et variant beaucoup dans le temps il a été imaginé d'effectuer le calcul de la transformée dans une fenêtre de durée choisie en fonction du signal à analyser. Nous devons cette avancée à Denis Gabor pour qui, si la formulation mathématique de Fourier est parfaite, son application pratique est extrêmement limitée parce qu'elle repose sur des signaux sinusoïdaux de durée infinie et qu'il n'y a pas de signaux de durée infinie et de forme constante dans la nature. En particulier, même dans le cas abstrait d'un signal de forme et de fréquence constantes sur une durée donnée, la théorie de Fourier ne peut exprimer ce qui se passe lors de l'établissement et de l'extinction du signal, deux instants qui peuvent avoir des aspects

et des effets très variées sur l'évolution de l'onde, donc sur sa perception s'il s'agit d'une onde sonore. Les autres passages transitoires ne peuvent pas non plus être décrits, or nous avons montré l'importance des transitions coarticulatoires dans nos analyses phonétiques précédentes.

4.8 Retour sur les principes de base

Le principe d'Heisenberg stipule qu'il est impossible d'avoir en même temps une bonne résolution en temps et en fréquence. Or dans l'évolution du signal de parole, nous avons des segments variant plus ou moins lentement et d'autres variant rapidement. L'analyse de Fourier sous forme de FFT n'est qu'un compromis, même s'il semble faire l'objet d'un consensus presque général au sein de la communauté des chercheurs en phonétique. Les phonèmes sont des éléments très courts, qui doivent s'établir à partir d'un élément phonétique précédent puis rapidement se fondre dans le suivant, le fondamental s'il existe ainsi que le contenu harmonique varient constamment pour des ajustements coarticulatoires. La question de non-stationnarité des signaux n'est pas spécifique à la phonétique et des alternatives ont été cherchées dans d'autres domaines pour pouvoir rendre compte et décrire les signaux non stationnaires qui sont majoritairement présents dans la nature. Pour tenter de mieux rendre compte des phénomènes temps/fréquence, des techniques nouvelles ont été développées à partir des années 1980, époque qui marque les débuts de l'industrialisation de l'information,. Ces travaux reposent sur ceux des pionniers de la mécanique quantique du début du XX^{e} siècle.

4.8.1 La notion de quantum

Selon certaines sources, lors d'une conférence en 1925 sur la physique quantique, N. Wiener aurait utilisé comme référence la musique pour expliquer la notion de quantum et énoncé que la précision dans le domaine temporel entraîne des imprécisions dans le domaine fréquentiel et vice-versa. Apparaît alors une notion de quantum sonore dès lors

que l'on connaît la taille minimum d'une particule sonore en fonction de sa fréquence. D'où l'idée, similaire à celle de la mécanique quantique qui décrit la matière, de pouvoir décrire les sons en termes de taille et de nombre de grains ou atomes.

Dès 1946, D. Gabor, prix Nobel de physique, avait fabriqué une machine utilisant un système de grains pour reproduire des sons. Ses théories sont dans l'article « Theory of communication » (1946) et « Acoustical Quanta and the Theory of Hearing » (1947). Pour lui, le problème fondamental posés avec l'analyse de Fourier et que l'utilisation d'ondes sinusoïdales implique une durée infinie du signal. Bien qu'à l'origine du découpage temporel utilisé dans le sonagraphe, il en conclue que cette théorie ne convient absolument pas à la description des sons à fréquences variables.

Pour Gabor il faut utiliser les théories de la physique quantique pour étudier les signaux sonores. Il effectue des études sur les seuils de discrimination et utilise les résultats des expériences menées par Buerck, Kotowski, Lichte, Shower et Biddulph. Dans une série d'expériences destinées à déterminer la durée nécessaire pour reconnaître la hauteur d'un son à diverses fréquences, il montre que :

1. les fréquences entre 500Hz et 1000Hz doivent être jouées au minimum pendant 10 ms avant d'être perçues comme des hauteurs ;
2. si une fréquence change, il faut deux fois plus de temps pour s'en apercevoir qu'il n'en a fallu pour entendre la première fréquence. Cette durée de reconnaissance du changement varie avec la fréquence. À basses fréquences, nous discernons plus vite ;
3. le discernement d'un changement d'amplitude prend un minimum de 21 ms. Gabor émet alors une théorie mathématique sur les surfaces de seuil de perception qu'il utilise pour développer les premières machines capables de changer la hauteur ou la durée de documents sonores.

Abraham Moles sera le premier à dire que le nombre de sensations que reçoivent nos organes psychophysiologiques est quantifiable [76]. Il énonce qu'« *Un message est un groupe fini, ordonné, d'éléments de perception puisés dans un répertoire et assemblés en une structure. Les éléments de ce répertoire sont définis par les propriétés du récepteur.* »

Pour pouvoir définir et déchiffrer un message visuel ou sonore il est nécessaire de prendre en considération les caractéristiques psychophysiologiques de ce récepteur. En étudiant le pouvoir de l'ouïe à résoudre de petites différences de fréquence et d'amplitude, il trouve qu'il y a environ 340 000 éléments audibles. Moles défini alors « l'atome sonore » comme une cellule tridimensionnelle ayant pour côtés le seuil différentiel de fréquence $\frac{\Delta F}{F}$, le seuil différentiel d'intensité $\frac{\Delta I}{I}$ et le seuil différentiel de durée $\frac{\Delta t}{t}$ in [4].

Nous ne pouvons pas négliger de citer ici les travaux de Iannis Xenakis en musique car celui-ci a été, dans le sillage des recherches précédentes, le premier musicien à utiliser les grains ou atomes sonores en composition musicale. Nous citerons aussi la thèse de Rocha Iturbide, M. [87] qui donne de nombreuses et précieuses informations tant théoriques que pratiques sur ce sujet.

4.8.2 En résumé

Une intense activité à eu lieu dans le sillage de la mécanique quantique. Moles et Gabor arrivent à des formulations équivalentes du quantum sonore, Moles allant jusqu'à donner une mesure du nombre de « quanta de sensation » reçu et Iannis Xenakis créé de nouvelles formes de composition musicale basées sur les atomes de Gabor, parfois appelé aussi synthèse granulaire. Il est possible de trouver des créations musicales et logicielles basées sur ces théories à l'IRCAM[12] au Centre Pompidou à Paris.

Cette nouvelle conception du signal sonore nous impose de nouvelles formes de pensée et d'analyse. Ceci ne va pas sans effets remarquables et nous observons parallèlement à l'évolution de la recherche, la réalisation de systèmes acoustiques inédits reposant sur la théorie quantique qui permettent de considérer et d'utiliser la diffraction sonore de façon équivalente à la diffraction de corpuscules quantiques, la réalisation de miroirs acoustiques à retournement temporels[13], le masquage des objets et bien d'autres techniques encore dans les laboratoires d'acoustique.

[12] http ://freesoftware.ircam.fr/

[13] Tout l'U, 138, Université de Franche-Comté, pp. 15

Il faudra donc attendre la fin du XX^{e} et le début du XXI^{e} pour voir le triomphe des idées des chercheurs cités ci-dessus, et leur mise en application dans de nombreux domaines de la recherche, de l'industrie et même de la vie courante.

Chapitre 5

Traitement des données

Ce que nous nommons « analyse phonétique standard » repose sur des mesures de formants obtenus par FFT avec le logiciel Praat [11] en ce qui concerne la première partie de ce chapitre. L'étude de la coarticulation des consonnes pharyngalisées *vs* non-pharyngalisées et l'application de l'équation de locus nous donnent un moyen reconnu pour traiter les mesures effectuées sur les formants afin d'en déduire un certain nombre d'informations concernat nos locuteurs [1].

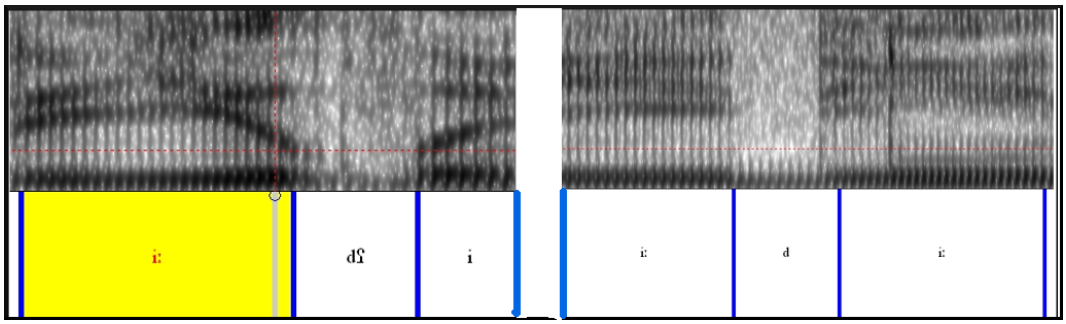

FIG. 5.1 – *effets de la coarticulation avec à gauche* [id$^{\varsigma}$i] *et à droite* idi

La coarticulation comme signe distinctif des dialectes

Dès les premiers balbutiements de l'enfant, la coarticulation se traduit par une expansion des traits propres à un segment phonétique sur les segments adjacents. Chez l'adulte elle a fait l'objet d'investigations au niveau de plusieurs articulateurs mais nous ne l'abordons ici qu'au niveau acoustique. Ses effets acoustiques ont été observés dans des séquences syllabiques simples de type VC et CV mais aussi dans des séquences complexes de type VCV, CVC, VCCV ou CVVC.

L'exploration de la coordination des gestes articulatoires dans la production de la parole montre l'existence de phénomènes de chevauchement, coproduction ou coarticulation [48], ce qui est aussi visible lors de l'observation du signal. La parole n'est donc pas une suite de sons de caractéristiques indépendantes, elle est constituée de « gestes » qui s'organisent entre eux selon des lois qui nous échappent souvent et que l'on nomme la coarticulation.

Les premiers synthétiseurs de parole fonctionnaient sur la base des quelques phonèmes pouvant entrer dans leur faible mémoire. Ils produisaient une parole hachée qui est resté la caractéristique des voix de robots, parce qu'il n'y avait pas de transitions entre les différents sons. Aucune langue humaine ne pouvant s'accomoder d'une telle technique aussi primitive, les synthétiseurs ont rapidement pris en compte la liaison entre phonème : les diphones, triphones et même au-delà ont été utilisés.

De nombreux chercheurs ont tenté d'expliquer le phénomène de coarticulation. Celui-ci peut-être vu comme un processus d'accommodation raccordant des sons contigus [17] ; [47], ou comme le résultat d'une altération des propriétés du segment par un ajustement articulatoire des deux segments adjacents. Dans ce cas, la transition entre les deux segments est réduite, mais il est aussi possible d'envisager l'action d'un seul segment dont les traces se retrouvent dans l'autre [8]. Les segments de parole ne sont donc pas constitués de gestes articulatoires isolés, mais exercent des influences multiples et réciproques les uns sur les autres.

Les caractéristiques remarquables de la coarticulation

Nous observons que les influences d'un segment sur les segments adjacents sont observables et que souvent un segment agit sur des segments éloignés [29].

Au niveau de l'analyse phonétique, la coarticulation est un processus complexe. Elle dépend de contraintes articulatoires, perceptives et phonologique du système linguistique considéré. Elle dépend de l'inventaire consonantique et vocalique du système considéré [65] ; [70] ; [66] ; [41], elle dépend de propriétés prosodiques comme l'accent [33].

Les situations de coarticulation

Nous distinguerons les effets anticipatoires qui indiquent une programmation de l'acte de parole et les effets rémanents qui signent une inertie des articulateurs. Mais en fait tous les articulateurs, procèdent à des ajustements avant l'arrivée de la cible articulatoire de même qu'ils peuvent procéder à des ajustements après.

Il existe un phénomène de résistance à la coarticulation et certains phonèmes montrent plus de résistance à la coarticulation que d'autres : les consonnes dentales ou alvéolaires et les voyelles palatales résistent plus que les consonnes labiales et vélaires et les voyelles vélaires.

Les stratégies d'économie articulatoire

Nous plaçons sous ce terme l'ensemble des ajustements opérés par les différents articulateurs dont l'action a des impacts spécifiques sur le résultat acoustique qu'est le signal de parole. Les déplacements de la langue, la protrusion labiale ou l'ouverture vélo-pharyngée donnent lieu à davantage de modifications spectrales que les mouvements des cordes vocales ou ceux de la mâchoire. Nous avons étudié la coarticulation d'un point de vue acoustique dans la séquence CV sur les différentes transitions de F2 de la consonne plosive décrites comme dépendantes du contexte vocalique adjacent,

transitions qui caractérisent différentes occurrences allophoniques de la consonne [22].

La coarticulation n'est pas seulement la résultante de limitations des articulateurs ou de stratégies d'économie articulatoire. Elle a des liens forts avec les représentations mentales [17] ; [47] ; [98] ; et les travaux convergent tous vers une même conclusion paçant ce processus complexe aux niveaux phonétique et phonologique [17] ; [47].

5.1 L'équation de locus

Le concept de locus a ouvert la voie à une série de recherches en acoustique, notamment en appliquant l'équation de locus définie par Lindblom. Pour Delattre et al., 1955, l'équation de locus permet de caractériser les aspects coarticulatoires d'un segment CV [64].

L'équation de locus est une régression linéaire de F_{2onset} (fréquence de F2 au début de la voyelle) sur F_{2mid} (fréquence de F2 au milieu de la voyelle) de plusieurs voyelles devant la même consonne – (où k et c sont la pente et l'ordonnée de la fonction de l'intersection de la droite de régression avec l'axe y).

5.1.1 Les résultats attendus de l'équation de locus

La pente de la droite de régression est caractéristique de la résistance de la consonne aux effets de la voyelle :

1. une pente plate est le signe d'un minimum de coarticulation entre les deux segments, F_{2onset} étant dans ce cas insensible à la nature de la voyelle qui suit, quelle que soit la cible fréquentielle vocalique à atteindre, ce qui traduit par la même occasion une résistance coarticulatoire maximale de l'articulation de la consonne aux effets de la voyelle.
2. une pente forte indique un effet de coarticulation maximum entre les deux segments. F_{2onset} et F_{2mid} ont la même fréquence, quelle que soit la cible à atteindre, ce

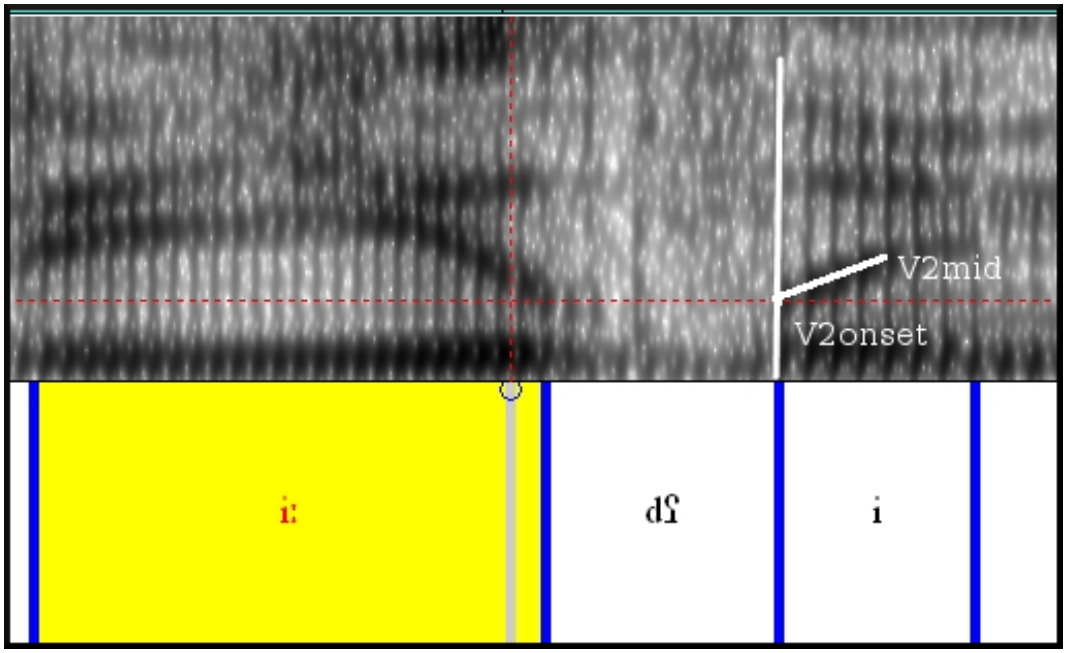

FIG. 5.2 – *mesures effectuées pour appliquer l'équation de locus*

qui traduit une résistance coarticulatoire minimale de l'articulation de la consonne.

Cette régression linéaire a été utilisée comme indicatrice du degré de coarticulation entre la consonne et la voyelle, [102] ; [103] ; [107] ; [56] ; [57], Fowler [32]. Cependant les interprétations de la coarticulation CV ont été très variables et Sussman et Krull ont observé la covariation de l'équation de locus avec le lieu d'articulation de la consonne, la valeur de la pente va décroissant pour les trois plosives : [g] > [b] > [d]. Ces auteurs ont trouvé que la consonne vélaire a une pente à peine plus forte que la consonne labiale, l'intersection-y est plus faible pour cette dernière ; la consonne dentale présente une valeur d'intersection-y élevée mais une pente plus plate.

Fiabilité de l'équation de locus dans l'étude de la coarticulation

L'équation de locus s'est avérée stable dans de nombreuses situations, indépendamment des conditions expérimentales. La hiérarchie qui apparaît en situation de discours

normal n'est pas bouleversée par l'utilisation de bite blocks [103], la pente de l'équation de locus en situation de discours spontané est plus élevée, révélant une coarticulation plus forte en passant du discours de laboratoire au discours spontané [57]. Pour la variation du débit de parole, les résultats montrent un maintien des distinctions entre les trois consonnes [g] > [b] > [d] par-delà les différences de débit, rapide *vs* lent, la hiérarchie de lieu d'articulation [g] > [b] > [d] se maintient à travers les changements de style de parole, lecture *vs* discours spontané chez 22 locuteurs anglo-américains [102]. Les valeurs de pente sont plus élevées en discours spontané qu'en lecture.

Comparaisons inter-langues

1. il a été montré que la gradation de la pente de l'équation de locus en fonction du lieu d'articulation était identique en thaï, en ourdou et en arabe égyptien du Caire, avec toutefois des différences relatives à l'inventaire phonologique de la langue [105] L'ourdou possède le contraste entre alvéolaires et rétroflexes et présente des valeurs de pente légèrement plus élevées pour les dernières comparativement aux premières. La variété arabe qui possède le contraste de pharyngalisation présente des valeurs de pente plus basses pour les consonnes dentales pharyngalisées que leurs correspondantes non-pharyngalisées.
2. en comparant les 3 plosives aspirées à leurs correspondantes voisées en anglais américain et en farsi, Modarresi et al. [75] aboutissent à des résultats qui non seulement confirment l'indication du lieu d'articulation via la pente de l'équation de locus vélaire ⇒ labial ⇒ dental, mais aussi obtiennent des distinctions de voisement : les consonnes voisées présentent une pente plus forte que leur correspondantes aspirées, excepté pour la paire vélaire [k^h] *vs* [g].
3. Molis et al. [77], relèvent une similitude entre l'anglais et le suédois, et la singularité du français en raison de l'aspiration présente dans les deux premières langues et de son absence en français.
4. une étude [29] montre que l'équation de locus permet de distinguer des variétés d'arabe.

5. une distinction des consonnes pharyngalisé vs non pharyngalisé en arabe peut être faite d'une façon fiable [28].

Critiques de l'équation de locus

Tabain et al. obtiennent des résultats différents en anglais australien [112]. Selon leurs mesures, la valeur de la pente est plus élevée pour une occlusive ou fricative non voisée [t k θ s ʃ] que pour une correspondante voisée [d g ð z ʒ]. Ils précisent toutefois que les mesures pour les occlusives prises dans le burst inversent la hiérarchie, la pente étant plus élevée pour une consonne voisée que pour une non voisée. Malgré ces différences, même dans ce cas l'équation de locus indique les lieux d'articulation. Fowler pense que l'équation de locus indique le mode d'articulation, car les différences sont significatives entre consonnes plosives et fricatives alvéolaires [32].

D'autres chercheurs contestent la validité de l'équation de locus comme indicateur du lieu d'articulation.

5.2 Classification des groupes d'atomes produits par Matching Pursuit

Les résultats obtenus par MP ne peuvent être traités par la statistique ordinaire ; nous avons besoin de nouveaux algorithmes pour le traitement de ces données. Nous avons cherché longuement une méthode simple à mettre en oeuvre pour effectuer les comparaisons qui nous intéressaient. L'inefficacité des méthodes présentées dans les livres de probabilité et de statistiques nous a poussé à chercher s'il existait de nouveaux algorithmes susceptibles de satisfaire nos exigences.

Ayant assisté à l'explosion de projets nouveaux et de grande envergure comme le séquençage de l'ADN, le projet SETI, au développement de systèmes automatiques de surveillance notamment celui des malades au niveau de l'électroencéphalogramme ou

du rythme cardiaque, d'investigations dans de nouveaux domaines de la chimie des molécules complexes, nous avons étudié les méthodes utilisées.

Nous n'avons pas manqué de faire le rapprochement entre les séquences atomiques produites par MP et celles étudiées dans le cadre de certains des domaines précités. Comme il n'y a pas de méthode applicable à une séquence ou à un signal particulier, sauf à simplifier ponctuellement l'algorithme utilisé, les techniques utilisées peuvent être généralisées à tous les domaines ou une classification-comparaison d'objets complexe et semblable est possible. Ceci n'est pas envisageable, pensons-nous, avec les séries de Fourier et leurs variantes associées à la statistique.

Retour sur le signal numérisé

Les fichiers obtenus par numérisation d'un signal sont une suite de bits ne présentant pas de redondances apparentes et partant ils ont une complexité telle qu'une tentative de simplification peut facilement arriver à un résultat encore plus volumineux que l'original. On dit qu'il présentent une entropie maximale et ils sont hors d'atteinte de notre compréhension.

Les fichiers numérisés n'ont pas de rapport avec l'information de parole qui nous concerne, ils ne sont que l'enregistrement d'une suite de valeurs de la pressions atmosphérique au temps de l'échantillonnage. Il est donc nécessaire d'en extraire des structures de niveaux d'organisation supérieures pour pouvoir les relier à notre sujet. C'est le rôle de la FFT dans la production de formants ou de MP dans celle de groupes d'atomes. Or l'équation de locus est une régression linéaire qui ne peut pas être appliquée aux fichiers d'atomes de MP.

Pour traiter la décomposition atomique nous avons les possibilités offertes par la logique floue [45]. Nous espérions tenir là un moyen simple et solide permettant de comparer nos fichiers d'atomes qui présentent une grande variabilité apparente. Cette voie, connue depuis longtemps, ne nous a toutefois pas convaincu et il nous a fallu nous

plonger dans la théorie de l'information pour trouver une technique capable de répondre à nos besoins.

5.3 Complexité et similarité.

Nous avons découvert des travaux récents, en mathématiques appliquées et en informatique, basés sur ceux de Church [14], Turing [115], Kolmogorov [52], lesquels avaient en leur temps donné des définition de l'information, une définition de sa complexité ainsi qu'une mesure de celle-ci. Nous n'ignorons pas les critiques et discussion épistémologiques sur ces sujets, mais les définitions et théorèmes que nous utilisons conviennent bien à notre propos.

5.3.1 La complexité de Kolmogorov

Si nous décrivons un ensemble d'éléments sous la forme d'une suite de caractères, la complexité de Kolmogorov sera la quantité d'information K permettant de décrire cette suite. K permet de mesurer le contenu incompressible de l'information contenu dans la suite, son entropie, nous définissons la complexité d'un désordre sans règles. La complexité de Kolmogorov est une quantité mesurable, mais non démontrable : en effet si à une suite donnée nous pouvons affecter une complexité K_1, rien n'indique qu'un autre algorithme ne nous permettra pas d'obtenir une valeur $K_2 < K_1$

La complexité de Kolmogorov devient rapidement grande dans le cas d'un fichier numérique représentant un signal sonore non stationnaire car il faut un grand nombre de règles pour définir ce fichier. Cela revient la plupart du temps à créer un fichier descriptif plus grand que celui à réduire. En conclusion le fichier numérisé d'un signal non stationnaire ne peut être représenté exactement que par lui-même, il est incompressible sans perte d'information[1].

[1]Rappelons qu'il y a quand même des méthodes pour compresser sans perte des fichiers sonres, mais

Un exemple pourra être donné par la complexité d'un signal stationnaire qui est faible : sa description toute entière tient dans la formule représentant une série de Fourrier.

Mesure de la complexité

Nous pouvons mesurer la complexité en mesurant le temps de calcul nécessaire au plus court programme qu'il est possible d'écrire pour produire notre fichier [7]. Comme cette mesure dépend de la vitesse de la machine, il est plus juste de compter le nombre de cycles nécessaires. On parle alors de « profondeur logique » de calcul. Toutefois le plus court programme dépend également du compilateur ou de l'interpréteur utilisé ; le plus court programme est en fait déterminé à une constante près. La complexité n'est pas démontrable : il est impossible de prouver que l'on connaît et utilise le plus court programme effectuant l'opération.

Conséquence

Lorsqu'il faut un très grand nombre de règles, c'est-à-dire un grand programme pour décrire un objet, il s'agit d'une complexité qui a besoin d'un très grand nombre d'informations et nous sommes dans le cas de la complexité aléatoire. Lorsque, au contraire, la situation globale est obtenue par ensemble de règles réduit, il y a complexité organisée. C'est la complexité organisée que nous devons tenter de trouver sur les ensembles d'atomes extraits des séquences VCV étudiées.

5.4 Classification et similarité

Parmi les nombreux algorithmes de traitement des données, nous avons étudié les systèmes de fouille de données (dits Data Mining en anglais) dont Weka est un exemple

elles sont spécialisées. Flac ne compresse que la musique par exemple.

créé sur spécifications du gouvernement Néo-Zélandais, simple, gratuit et facile à mettre en oeuvre.

Toute science est basée sur la classification des éléments à étudier, que ce soit des microbes, des particules élémentaires, des éléments chimiques, des molécules, le monde vivant et les corps astronomiques. La classification permet de former des ensembles au sens mathématique qui sont des rassemblement d'objets ayant une ou plusieurs propriétés communes, c'est à dire qui deviennent comparables sous certains aspects et nous notons que les nouvelles recherches théoriques en phonétique ont bien pour but d'améliorer la catégorisation des traits acoustiques, autrement dit leur classement. C'est ce qui apparaît clairement tout au long de notre travail sur le matériau phonétique arabe, notamment au niveau de la dialectologie et des phénomènes de coarticulation. Nous allons essayer de placer cette démarche dans le cadre de la théorie de l'information afin de tenter de dégager et d'ouvrir des pistes de recherche.

La similarité pose de redoutables problèmes au mathématicien. Chacun d'entre nous a eu affaire aux figures géométriques semblables et nous avons l'illusion ou l'intuition que certains objet sont ont des propriétes communes. Mais il apparaît par exemple en chimie que des molécules de même formule ne sont pas organisées exactement de la même façon et la phrase ne présente pas toujours les mots dans le même ordre. Souvent pourtant ces molécules ou phrases présentent les mêmes propriétés ; nous pouvons dire qu'elles ont le même contenu informationnel. En réalité nous devrions préciser « dans un certain domaine » car rien ne dit que la variation d'organisation n'induit pas des effets non sensibles immédiatement et non connus par nous, mais réels. La similarité n'est ni une translation d'échelle ni une homotétie, mais la relation approchée de deux objets sur les points spécifiques, voire une relation sur un seul point particulier par exemple.

5.4.1 « Est complexe ce qu'on ne peut représenter avec concision. » (Kolmogorov)

Les questions de simplification et de concision étant au cœur du travail scientifique, Kolmogorov et Occam sont deux théoriciens importants de la simplification[2]. L'expérience montre aussi qu'à partir d'un certain niveau de complexité, les systèmes deviennent instables ce qui les rend également incompréhensibles. Mais simplification et concision ne signifient pas perte d'information ce qui a également été théorisé par Thüring dans le cadre de sa machine [115]. Aujourd'hui nous utilisons souvent des systèmes redondant pour palier les failles possibles d'un logiciel complexe et la recherche en informatique s'oriente vers la représentation concise ou parcimonieuse de l'information en partie à cause de ce risque.

La mesure de Levin

Levin raffine l'idée d'Occam et prouve qu'un objet structuré est plus probable qu'un objet aléatoire. Cette idée a permis de construire des compresseurs des fichiers avec des algorithmes comme zip, gzip, lha, etc. Les fichiers sont écrits de manière plus concise que l'original mais leur complexité interne augmente.

Si s est une suite binaire, la Mesure de Levin m(s) est la probabilité définie par $m(s) = \frac{1}{2K(s)}$ où K(s) est la complexité de Kolmogorov qui est aussi la longueur du plus petit programme capable de décrire s. Cette mesure propose un monde où les objets sont produits par des programmes ou des mécanismes assimilables à des programmes.

La définition de la simplicité par la taille du plus petit programme permettant de décrire ou de reconstituer l'information [52] a également été très féconde en informatique théorique et a trouvé aujourd'hui des applications en sciences thermodynamique, physique, chimie, statistiques, en biologie, psychologie et autres. Cette théorie a aussi plusieurs fois été utilisée pour interpréter certains problèmes d'épistémologie comme le

[2]http ://www.hutter1.de/ait.htm

principe du Rasoir d'Occam[3], l'inférence déductive, et d'autres.

Un exemple intéressant d'application des notions précédentes à la comparaison interculturelle se trouve dans le livre de Pascal Baudry, *Français et Américains, L'autre rive*[4], Annexe 3, p. 270.

5.4.2 La construction des formes

Nous revenons un instant sur la théorie constructale pour complèter ce qui précède. Développée par Adrian Bejan au MIT[5], elle nous intéresse au plus haut point car elle vise à expliquer l'origine des formes qui semblent se développer selon des algorithmes comparables dans des domaines très différents (le minéral, le vivant, le sociétal). Elle doit logiquement s'intégrer dans la problématique de l'étude, de la production, mais surtout de l'évolution des langues, en particulier de la parole dont les formes fugaces sont difficiles à noter et à observer. C'est un outil nouveau particulièrement utile, à notre sens, pour l'avenir de notre discipline.

Voyons ce qu'explique Bejan. Dans la nature, la complexité naît la plupart du temps de la combinaison de processus élémentaires. Ce sont les lois simples de la physique macroscopique, et plus particulièrement de la thermodynamique qui génèrent l'apparition des formes. Ces formes se caractérisent par une optimisation destinée à diminuer les dépenses d'énergie et de matière pour lutter au mieux contre l'entropie ambiante qui est assimilable à une perte d'information. Le moteur de cette optimisation est l'évolution compétitive pour la survie dans laquelle s'affrontent les divers éléments de la matière et de la vie. Nous pouvons l'appliquer à différents aspects formels des langues et dans notre cas de la parole.

La diversité du monde du vivant est telle qu'on pourrait l'imaginer infinie et croire que toutes les formes sont possibles. En fait, les formes du vivant dépendent de l'environ-

[3]Le principe du rasoir d'Occam pose « que les explications les plus simples sont vraisemblablement plus justes que les plus complexes. »

[4]http ://www.pbaudry.com/cyberlivre/dl_livre.php ?target=cyberlivre.pdf

[5]http ://www.mit.edu/ vu le 29/01/2008

nement, elles sont donc extrêmement limitées. En plus de devoir coller aux variations du milieu, la vie doit respecter les lois de la physique, ce qui restreint encore les variations possibles. Ainsi des biologistes on pu déterminer que la mouche drosophile est basée sur un modèle semblable à celui des mammifères actuels et que c'est une structure d'ADN similaire à la nôtre qui génère ses structures.

Si l'entropie n'est pas la loi la plus fondamentale de l'univers, puisque la vie et en particulier l'information se définissent par leurs capacités de croissance, de régénération et de reproduction (néguentropie), la seule loi universelle, est d'aller au minimum de dépense d'énergie, de suivre le « principe d'économie naturelle » dit de Fermat–Leibniz, que l'on l'appelle encore le « principe d'action extrémale », ou « principe du minimum ». Cette loi se trouve à la base de la « théorie constructale » et permet de réintroduire et de donner une explication scientifique à la question de la finalité qui est au service des besoins matériels de survie et non d'une quelconque métaphysique, car si la vie respecte ce principe, c'est après un détour permettant de s'économiser, en se mettant par exemple à courir pour manger ou pour ne pas être mangé, à parler pour augmenter son pouvoir sur la nature ou sur les autres ou à se taire pour échapper à une situation désagréable.

La théorie constructale stipule que chaque fonction tend à s'optimiser à la longue, ce qui expliquerait la complexification croissante à long terme du vivant, donc de l'information dont il est fondamentalement constitué. Ce qui n'exclue pas les régressions et simplifications localisées, brutales et passagères.

Nous remarquons toutefois qu'il peut y avoir antinomie apparente entre le « principe d'économie naturelle » et le principe de plaisir qui est aussi une des caratéristiques du vivant. La langue la plus simple, ou la plus claire, n'est forcément la plus prisée par une population et à une époque donnée.

Nous constatons que de nombreuses recherches et discussions vont dans le même sens à partir de résultats théoriques vérifiés par l'expérience. Nous notons en particulier :

a que la complexité est mesurable sous certaines conditions ;

b que la complexité masque la plupart du temps des mécanismes simples ;

c qu'il est nécessaire de simplifier l'expression décrivant les objets pour pouvoir les manipuler d'une façon abstraite.

Nous avons vu deux théories qui tentent d'expliquer le monde, l'une où les objets (la complexité) sont produits par des programmes ou des mécanismes assimilables à des programmes et l'autre où la complexité (les objets) naît la de la combinaison de processus élémentaires. Notre conjecture est que ces deux théories se relient quelque part et qu'elles donneront ensemble des outils d'analyse très puissants. C'est pourquoi nous les évoquons aussi longuement.

5.5 La compression comme mécanique simplificatrice

Il existe des compresseurs sans perte pour les fichiers d'enregistrement sonore. Flac, un compresseur Open Source, offre de bonnes performances sur les fichiers musicaux, soit en général une réduction de plus de 50 % du volume du fichier initial. Nous avons voulu le tester bien que cela soit une absurdité épistémologique. Heureusement la compression avec Flac est sans effet sur nos séquences VCV. Si quelqu'un obtenait une compression, ce serait par pur hasard et cela n'aurait par conséquent, aucune valeur méthodologique..

5.5.1 Pourquoi n'est-il pas possible de compresser le fichier numérisé ?

Nous pouvons considérer sans grande hardiesse que les échantillons contenus dans les fichiers de sons numérisés sont des particules d'information. Or la parole est plutôt une chimie des sons et personne n'a jamais pu prétendre faire de la chimie avec des particules. C'est donc à un niveau d'organisation supérieur à celui des particules que doivent s'exercer les manipulations dont nous avons besoin.

5.5.2 Le choix du niveau d'organisation où agir

Plusieurs choix peuvent être faits : nous avons analysé des impasses comme la reconnaissance vocale. Notre moyen le plus ancien et usité pour compresser et manipuler l'information phonétique est le sonagraphe qui nous donne une image claire des phonèmes et de leur évolution temporelle ; ensuite, nous réduisons encore souvent cette information sous la forme de données statistiques.

Si les formants se présentent sous forme de trajectoires formantiques en fonction du temps, Matching Pursuit propose des ensembles d'atomes qui ne peuvent plus être traités de la même façon. La solution à ce problème a été donnée par Charles Benett [7]. qui a pour la première fois défini la notion de distance informationnelle ou similarité (voir aussi [19]). Pour mesurer cette distance, nous devons rassembler les atomes dans un seul objet, objet qui est naturellement un fichier compressé car nous ne savons pas définir autrement et d'une façon concise, un ensemble d'atomes.

5.5.3 Distance informationnelle

Un fois obtenus les fichiers compressés à comparer nous devons mesurer la distance qui les sépare afin de les placer les uns par rapport aux autres. La distance informationnelle entre eux est fonction de leur similitude ou similarité. Or, la similarité est une question difficile puisqu'il s'agit de comparer les objets ayant des traits communs mais qui ne peuvent pas se normaliser selon une simple translation. La mesure numérique du contenu commun en information est obtenue en utilisant des algorithmes de compression de données : meilleurs sont les algorithmes utilisés, plus fines seront les classifications obtenues.

Cette théorie a été féconde car elle a généré des recherches en linguistique (écrit), sociologie, mais surtout en génétique où elle permet le séquençage des gènes, en chimie pour comparer des molécules longues, en médecine pour l'observation des rythmes EEG ou cardiaque et dans tous les domaines ou l'on peut décrire le phénomène étudié sous

forme d'une suite de caractères ou de mots. Or la suite d'atomes obtenue avec MP est écrite sous la forme d'une suite de symboles ou de mots. Nous entrevoyons là des possibilités d'analyses phonétiques par cette technique.

5.5.4 Un exemple de classification des langues

L'élaboration d'un arbre des différentes langues et dialectes, de leur filiation, préoccupe les linguistes et les phonéticiens comme nous l'avons vu dans l'étude des dialectes arabes. La méthode de la distance de similarité peut être utilisée avec profit. P. Vitanyi et ses collaborateurs, ont tenté de construire un arbre de classification des 52 langues indo-européennes principales. « *Partant de la traduction de la Déclaration des droits de l'homme dans chacune des 52 langues, ils ont laissé leur méthode automatique mener tout le travail d'élaboration de l'arbre. Celui obtenu est conforme, pour l'essentiel, à ce qu'obtiennent les linguistes, ce qui est assez bon puisque ces mathématiciens et informaticiens ne disposent d'aucune compétence particulière en linguistique et que c'est finalement l'algorithme de compression utilisé qui a fait le travail de repérage des similarités entre langues.* » in [20].

Reconnaissance automatique des patrons ou gestes élémentaires du signal de parole

La reconnaissance des formes et les dernières avancées dans la théorie de la similitude sont précieuses pour notre projet. Des progrès importants on été faits ces dernières années dans la séparation aveugle des sources et la reconnaissance des formes sonores. Pour pouvoir extraire des sons d'un mélange, il est nécessaire de pouvoir reconnaître les diverses formes de chacun des signaux. La technique est encore lourde et les logiciels en développement, mais c'est ceux-ci que nous souhaitons utiliser dans nos recherches futures.

La première idée qui vient est de représenter le signal en utilisant le théorème de

Fourier, or il s'avère que le signal de parole est non stationnaire et que l'on ne peut le décrire sans pertes et sans déformations importantes de cette manière. Le principe de la FFT est de découper le signal en fines tranches temporelles et de chercher sur chacune le contenu harmonique. Si le théorème de Fourier permet de représenter un signal de forme constante et de durée infinie par une formule très simple et concise, ici non seulement la complexité de l'algorithme croit mais des arrangements à la limite de l'acceptable sont fait avec la théorie : dans le cas des impulsions (signaux appartenant à une distribution de Dirac), la fenêtre temporelle de mesure est remplie de zéros en dehors du pic d'énergie.

Enfin l'on démontre qu'il est impossible d'avoir une relation temps/fréquence précise : soit on connaît le temps avec précision et la fréquence est très approximative, soit l'inverse (principe d'incertitude d'Heisenberg). Dans tous nos travaux nous avons surtout eu besoin de précision en temps et nous avons accepté une large imprécision en fréquence en utilisant une fenêtre de mesure de fréquence large, appelé aussi « filtrage large ». Cette imprécision donne lieu à de nombreux dilemmes pour fixer la valeur d'une fréquence de formant, parce que nous sommes écartelé entre un besoin intellectuel de précision et un savoir qui dit que notre mesure n'est qu'une approximation commode. La plupart des travaux en phonétique restent basés sur cette méthode d'analyse comme nous avons encore pu le constater en août 2008 à Sarrebrück, à ICPhS qui est le plus important congrès international de phonétique.

Les logiciels d'analyse par FFT écrivent leurs résultats dans des fichiers que l'on peut ensuite analyser. Mais nous avons vu que les erreurs de mesure étaient en nombre important et que cette information est mal aisé à traiter pour des questions de reconnaissance des formes qui persistent à un niveau élevé de l'analyse.

Les diverses techniques de représentation des résultats

La modélisation d'une observation est un problème récurrent dans de nombreux domaines. Il est facile d'imaginer un ensemble important de signaux parmi lesquels

certains seraient susceptibles de décrire l'observation de façon satisfaisante : on utilise alors une régression linéaire où le vecteur d'observation est considéré comme la somme des vecteurs de signaux potentiels.

Si nous supposons que seul un nombre restreint de signaux suffit à une modélisation convenable, il devient nécessaire de les identifier, ce nous nommerons représentation parcimonieuse de l'observation et l'on dit que le modèle est sous-déterminé quand le nombre de prédicteurs potentiels n est beaucoup plus important que la taille du vecteur d'observation m, mais les algorithmes de sélection de variables utilisés en statistiques ne s'appliquent que dans les cas surdéterminé où $n < m$.

Matching Pursuit permet de contourner la difficulté des modèles sous-déterminés. Cet algorithme nous a parût très intéressant et c'est avec lui que nous avons décidé de faire nos premiers pas en analyse temps-fréquence. Matching Pursuit s'affranchit des contraintes en utilisant un dictionnaire de formes d'ondes qui seront adaptées à la situation locale. La machine cherche l'atome qui a l'énergie la plus proche de l'énergie du signal. Une fois cet atome trouvé, il est classé dans un livre et soustrait du signal. Le reste de cette soustraction, qui est toujours une onde sonore dans le cas qui nous préoccupe, subi la même opération et ainsi de suite jusqu'à ce que le nombre d'itérations choisi soit atteint en fonction de l'objectif que l'on s'est assigné. Il s'agit donc d'un algorithme récursif que l'on arrête quand on estime que le livre contient suffisamment d'informations pour représenter correctement le signal en fonction de nos buts :

1. la première conséquence de la méthode est que l'on maîtrise parfaitement la profondeur de description du signal, donc sa compression ;
2. la deuxième est qu'il est possible de reconstruire exactement le signal de départ, en faisant la somme des atomes du livre et du résidu de la décomposition au moment de l'arrêt. Ce résidu est aussi un fichier de type sonore qui peut ausi faire l'objet d'études.

Matching Pursuit offre un système puissant de décomposition du signal, parfaitement réversible ce qui n'est pas le cas des autres méthodes. Il permet de retrouver par

réversibilité parfaite, le signal original issu du fichier que nous avons numérisé.

Chapitre 6

Résultats

6.1 L'équation de locus des consonnes pharyngalisées de l'arabe

Nous avons vu que l'équation de locus avait été utilisée pour vérifier des différences spécifiques entre des langues. L'étude de Sussman et al. [106] montre que les interdentales de l'ourdou ont une valeur de pente relativement plus élevée qu'en ce qui concerne les alvéolaires. L'étude de Tabain et al. [112] relèvent la même tendance dans les deux langues aborigènes d'Australie. Sussman et al. ont montré que l'équation de locus permettait de distinguer nettement les consonnes dentales de l'arabe égyptien du Caire en fonction de la pharyngalisation, les consonnes pharyngalisées présentant des valeurs de pente plus basses que les consonnes non-pharyngalisées.

Dans une étude portant sur la production en arabe moderne de dix locuteurs d'origine marocaine, Yeou [122] a montré que les valeurs de pentes sont basses pour les consonnes pharyngalisées [$t^ˤ$ $d^ˤ$ $s^ˤ$ $ð^ˤ$], respectivement 0,37, 0,31, 0,35 et 0,22, comparativement à de leurs correspondantes non-pharyngalisées [t d s ð], (0,66, 0,48, 0,56 et 0,46). En arabe moderne, le contraste de pharyngalisation se révèle donc à travers l'équation de locus.

TAB. 6.1 – *pente et coefficient de régression pour les 8 locuteurs.*

	Non pharyngalisé				pharyngalisé			
C	t	d	s	ð	$t^ˤ$	$d^ˤ$	$s^ˤ$	$ð^ˤ$
Inter-y	531	579	524	411	570	479	325	439
pente	0,750	0,662	0,752	0,741	0,473	0,540	0,649	0,487
R^2	0,949	0,884	0,848	0,903	0,883	0,868	0,885	0,839

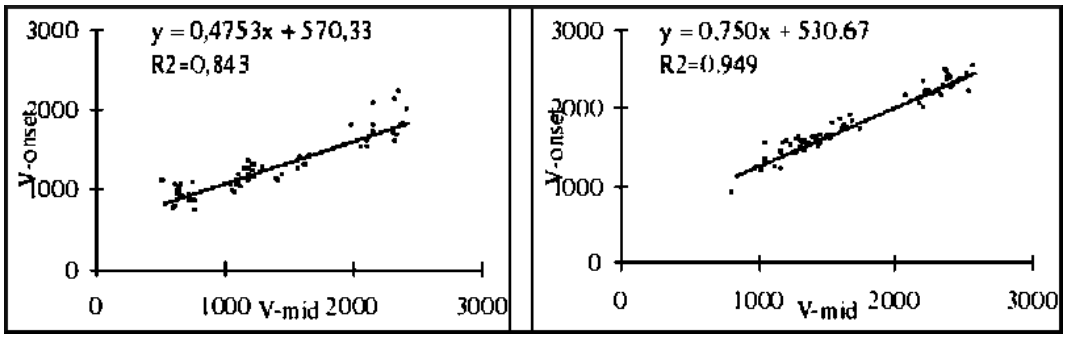

FIG. 6.1 – *droite de régression de la consonne pharyngalisée* [$t^ˤ$] *à gauche et sa correspondante non-pharyngalisée* [t] *à droite.*

6.1.1 Une étude portant sur huit locuteurs

Nous avons exploité [27] l'équation de locus pour l'étude du contraste consonantique pharyngalisé *vs* non pharyngalisé. Bien que plus élevées que dans l'étude de Yeou [122], les valeurs de pente obtenues confirment la solidité de cette méthode pour qualifier le contraste phonologique de pharyngalisation en arabe moderne (cf. tableau 6.1).

Portant sur la production en arabe moderne de locuteurs originaires de huit pays arabes différents, cette étude montre que [$t^ˤ$ $d^ˤ$ $s^ˤ$ $ð^ˤ$], ont des pentes plus basses (0,47, 0,54, 0,64 et 0,48), que celles (0,75, 0,66, 0,75 et 0,74) des consonnes non-pharyngalisées [t d s ð]. Ceci est visible sur les figures 6.1 et 6.2, qui mettent vis-à-vis la consonne pharyngalisée et sa correspondante non-pharyngalisée.

L'équation de locus permet de distinguer nettement deux groupes de consonnes, les

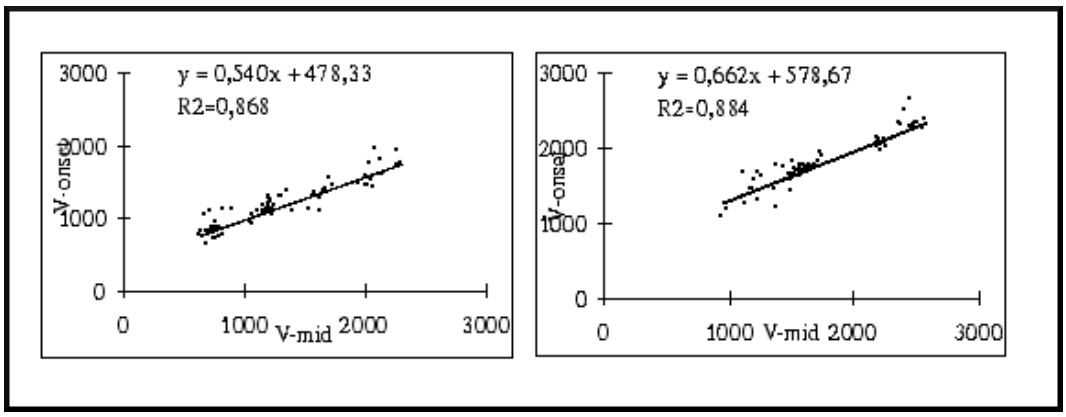

FIG. 6.2 – *droite de régression de la consonne pharyngalisée* [d^{ς}] *à gauche et la correspondante non-pharyngalisée* [d] *à droite.*

consonnes pharyngalisées présentant des valeurs de pente basses et les non-pharyngalisées qui ont des pentes plus hautes. Les valeurs que nous obtenons s'avérent plus élevées que la littérature ne l'indique et pour des valeurs données de 0,50 pour le groupe de consonnes non-pharyngalisées, nous trouvons en moyenne 0,75. L'étude de Yeou trouve pour les consonnes pharyngalisées une moyenne d'environ 0,3 tandis que nous trouvons autour de 0,57. La méthode de mesure peut influencer le résultat : la résonance de F2 dans l'explosion ou dans la friction génère des pentes faibles pour [t] et [d], respectivement 0,23 et 0,24 selon Modarresi et al. [75]. Comme la mesure est difficile dans le cas de la consonne [s], cela expliquerait les valeurs élevées trouvées par Yeou [122] pour cette consonne.

6.1.2 Équation de locus et origine dialectale

L'étude qui précède suggère que l'équation du locus peut être exploitée dans des comparaisons inter-dialectales. Tabain et Butcher [112] l'ont mis à l'épreuve avec deux langues aborigènes d'Australie très proches, le yanyuwa, le yindjibarndi, en comparaison avec l'anglais australien. Les résultats montrent une différence nette entre deux groupes. Les différences entre d'une part les deux langues aborigènes et l'anglais austra-

lien seraient dues, selon les auteurs, à l'inventaire phonologique des langues. Les langues aborigènes ont deux fois plus de lieu d'articulation consonantique dans la région dentale alvéolaire et nettement moins de voyelles que l'anglais australien. Ceci impliquerait, selon Tabain et Butcher, davantage de coarticulation dans les deux langues aborigènes et il n'y a pas, selon eux, de différences significatives entre les variétés aborigènes.

L'équation de locus a été éprouvée dans diverses situations et s'est avérée pertinente pour l'indication de la variabilité phonétique consonantique (lieu d'articulation, mode articulatoire, voisement, aspiration, pharyngalisation), du degré de coarticulation CV (consonne plus résistante ou moins résistante à la coarticulation), du type de discours (spontané, lecture, reformulation) et du débit de parole (ordinaire, lent, rapide).

6.2 L'équation de locus à l'épreuve

Compte tenu de la variabilité de plusieurs indices acoustiques d'une région du Monde arabe à l'autre, nous avons voulu éprouver la validité de l'équation de locus dans l'indication du contraste de pharyngalisation en fonction de l'origine dialectale du locuteur [29]. Le but est double : vérifier si l'équation de locus en arabe moderne varie selon l'origine dialectale des sujets ; 2) vérifier si elle révèle, chez un même locuteur, des stratégies variables de production de la séquence CV en passant de l'arabe moderne ASC (langue de scolarisation) à l'arabe dialectal AD (langue maternelle).

Ghazeli [40] a montré que la durée et le sens de la coarticulation étaient très différents chez des locuteurs tunisiens, libyens, égyptien, jordanien et irakien. Comme notre corpus ne contient que des consonnes dentales et des alvéolaires pharyngalisées [t^{ς} d^{ς} s^{ς} $ð^{\varsigma}$] et non-pharyngalisées [t d s ð], nous pouvons vérifier d'une part que l'équation de locus permet bien de faire des distinctions en fonction du concept de résistance articulatoire et d'autre part qu'elle permet de mettre en relief une possible variabilité de l'articulation des consonnes pharyngalisées en relation avec les zones dialectales étudiées. Dans notre étude, le contraste de pharyngalisation en arabe devrait nous donner des indices

spatiotemporels différents, mettant en relief des stratégies de coarticulation variant en fonction du dialecte maternel.

6.2.1 Méthodologie utilisée avec l'équation de locus

Nos variétés dialectales arabe sont les dialectes yéménite, koweïtien, jordanien et marocain en ASC et en AD. Chaque dialecte est représenté par quatre locuteurs masculins âgés de 20 à 40 ans. Les locuteurs koweïtiens et yéménites sont tous originaires de la capitale de leur pays mais la provenance régionale des locuteurs jordaniens et marocains ne nous est pas connue. Tous les locuteurs choisis pour cette étude sont nés et ont grandi dans leur pays d'origine ; ils maîtrisent tous leur langue maternelle (AD) et l'ASC.

Le corpus est constitué par une liste de 24 mots en ASC et autant en AD. Ces mots comprennent des séquences VCV symétriques iCi, uCu et aCa, apparaissant toujours en position médiane dans un mot, où C est soit pharyngalisée [tˤ dˤ sˤ ðˤ], soit non-pharyngalisée [t d s ð]. Les mots du corpus ont été insérés dans une phrase porteuse du type [qul...ljawm] (dis... aujourd'hui). Tous les locuteurs ont été enregistrés selon la même procédure.

Pour chaque variété de langue ASC et AD les séquences CV ont été segmentées et étiquetées à l'aide de PRAAT ce qui nous a donné un ensemble de 1152 élémens CV. Pour chaque séquence, la fréquence du deuxième formant (F_2) a été mesurée en deux points de la trame, au début de la résonance vocalique (F_{2onset}) puis au milieu de la voyelle (F_{2mid}) ce dernier point étant admis comme coïncidant généralement avec la partie stable de la voyelle. L'étude a porté sur un total de 4608 mesures de fréquences (1152 voyelles $\times$ 2 (AD + ASC) $\times$ 2 (mesures) = 4608).

Les résultats des 16 productions en ASC présentent les mêmes tendances observées par Sussman et al. [106] pour l'arabe égyptien, Yeou [122] et Embarki et al. (2006 [28]) pour l'arabe standard. Le contraste consonantique de pharyngalisation se traduit par des équations de locus différentes (cf. tableau 6.2).

TAB. 6.2 – *inter-y, pente, R^2 pour les 16 locuteurs en ASC.*

	Non pharyngalisé				pharyngalisé			
C	t	d	s	ð	tˤ	dˤ	sˤ	ðˤ
Inter-y	438	508	331	376	476	448	262	418
pente	0,728	0,714	0,815	0,769	0,548	0,556	0,767	0,560
R^2	0,917	0,819	0,913	0,936	0,763	0,769	0,857	0,793

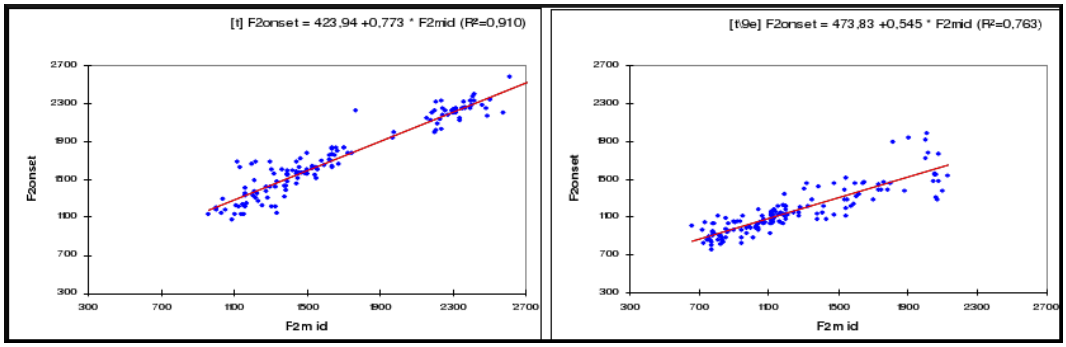

FIG. 6.3 – *droite de régression de la consonne non pharyngalisée* [t] *à gauche et sa correspondante pharyngalisée* [tˤ] *à droite.*

Les consonnes pharyngalisées présentent des pentes plus faibles que celle de leurs correspondantes non-pharyngalisées. La valeur de pente la plus basse du groupe est affectée à la consonne pharyngalisée [tˤ] (0,54), la correspondante non-pharyngalisée [t] présente une valeur (0,77) plus élevée, figure 6.3.

La comparaison des valeurs absolues en Hz de la même trame vocalique dans [tˤ] et [t] ne laisse pas apparaître de différences importantes et les calculs ANOVA à un seul facteur ne montrant pas d'effets significatifs sur les valeurs de F_{2onset}, avec F(1, 122) = 1,62, p > 0,05, ni sur celles de F_{2mid}, F(1, 122) = 1,95, p > 0,05. Pour [ð] et [ðˤ] les différences de pente sont sans ambiguïté (0,76 contre 0,55) (cf. figure 6.4) et les résultats donnés par l'ANOVA à un seul facteur sont significatifs pour F_{2onset} (F(1, 124) = 3,59, p < 0,01), et non significatifs pour F_{2mid}, $F(1, 124) = 1,57, p > 0,05$.

Pour les fricatives alvéolaires [s] et [sˤ] nous trouvons une valeur de pente proche,

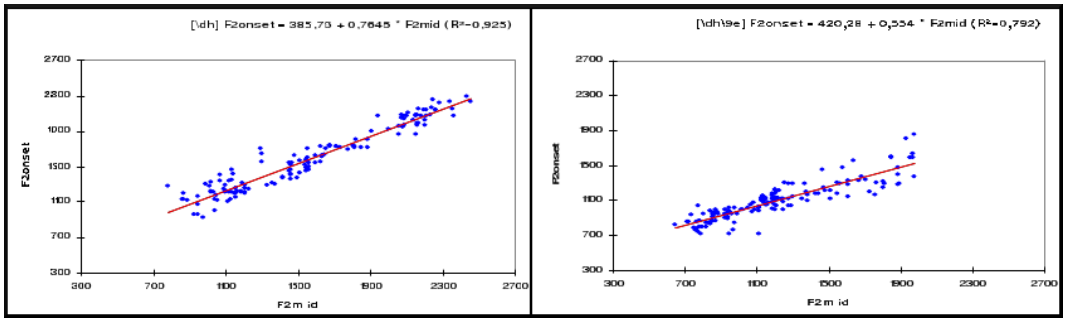

FIG. 6.4 – *équations de locus pour la consonne non-pharyngalisée* [ð] *(à gauche) et sa correspondante pharyngalisée* [ðˤ] *(à droite) (ASC).*

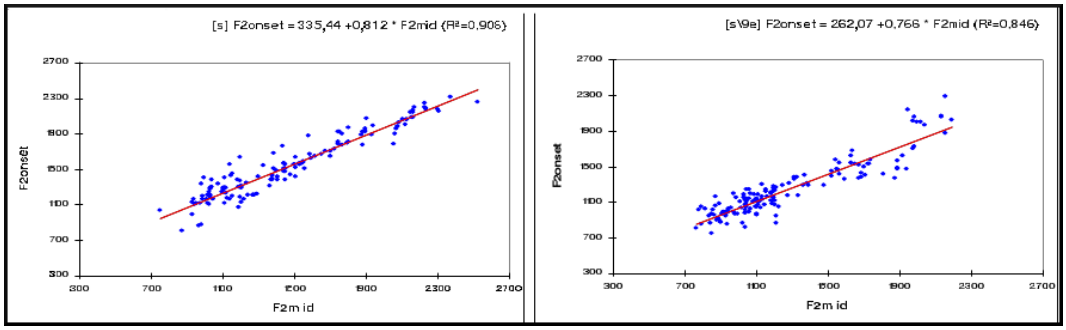

FIG. 6.5 – *droite de régression de la consonne non-pharyngalisée* [s] *(à gauche) et sa correspondante pharyngalisée* [sˤ] *(à droite) (ASC).*

respectivement 0,81 et 0,76, figure 6.5, où les différences sont significatives sur F_{2onset}, F(1, 123) = 271 ; p < 0,01 et non significatives sur celles de F_{2mid}, $F(1, 123) = 1,96; p > 0,05$.

L'opposition [d] – [dˤ] se manifeste par des valeurs de pente éloignées, (0,71 et 0,57, figure : 6.6, et nous ne trouvons des différences significatives que pour les valeurs de F_{2mid}, F(1, 126)=3,32 ; p < 0,01 avec les calculs ANOVA à un seul facteur. C'est sur le début de la voyelle que les seize locuteurs présentent en général des différences marquées

Les résultats obtenus en ASC confirment les données fourniés par la littérature. Ils soulignent la sensibilité des valeurs fréquentielles de F_{2onset} à la consonne qui précède,

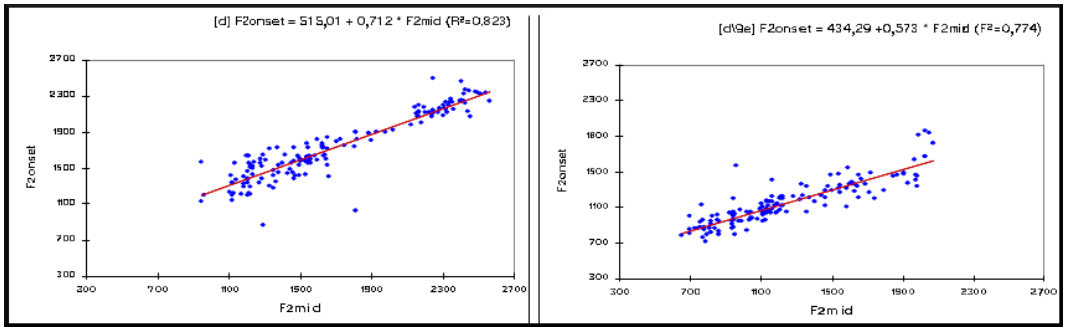

FIG. 6.6 – *droite de régression de la consonne non-pharyngalisée* [d] *(à gauche) et sa correspondante pharyngalisée* [dˤ] *(à droite) (ASC).*

qu'elle soit pharynLocusgalisée ou non-pharyngalisée.

6.2.2 L'équation de locus en arabe dialectal

Comparativement à l'ASC les résultats obtenus à partir de la production dialectale permettent de distinguer trois aspects :

1. les valeurs de pente pour la même consonne sont différentes en ASC et en AD ;
2. la hiérarchie au sein de chaque série de consonnes est différente avec : {ASC $\Rightarrow$ $t^{\varsigma} < ð^{\varsigma} < d^{\varsigma} < s^{\varsigma}$] et [d < t < ð < s] } $\not\equiv$ {AD $\Rightarrow$ [$d^{\varsigma} < ð^{\varsigma} < t^{\varsigma} < s^{\varsigma}$] et [t < d < ð < s] }. Les sifflantes [s^{ς}] et [s] sont toujours affectées de la pente la plus élevée ;
3. la distance entre consonnes pharyngalisée et non-pharyngalisée tend à diminuer en AD par rapport à l'ASC.

Ces trois aspects montrent de manière convergente que d'une part le locuteur programme et exécute la séquence syllabique CV de manière différenciée dans les deux variétés de langue et d'autre part que la coarticulation entre la consonne et la voyelle adjacente, matérialisée en utilisant l'équation de locus, témoigne de la diversité de la

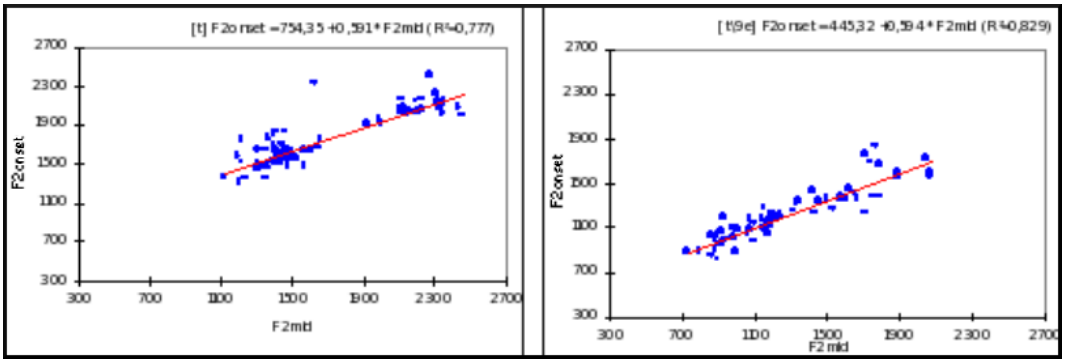

FIG. 6.7 – *droite de régression de la consonne non-pharyngalisée* [t] *à gauche et sa correspondante pharyngalisée* [tˤ] *à droite (AD).*

résistance articulatoire entre mêmes segments en passant de l'ASC à la variété dialectale : hormis [tˤ] dont la pente augmente de 0,54 à 0,59, toutes les autres consonnes sont affectées d'une valeur de pente plus basse en AD qu'en ASC :

dˤ passe de 0,54 à 0,59 (exception) ;

dˤ passe de 0,57 à 0,47 ;

sˤ passe de 0,76 à 0,66 ;

ðˤ passe de 0,55 à 0,51.

Le même phénomène se répète pour la série non-pharyngalisée :

t passe de 0,77 à 0,59 ;

d passe de 0,71 à 0,61 ;

s passe de 0,81 à 0,79 ;

ð passe de 0,76 à 0,66.

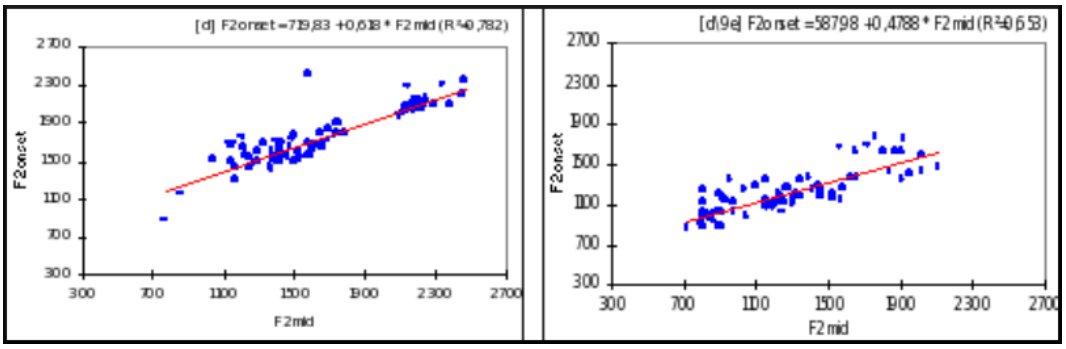

FIG. 6.8 – *droite de régression de la consonne non-pharyngalisée* [d] *à gauche et sa correspondante pharyngalisée* [dˤ] *à droite (AD).*

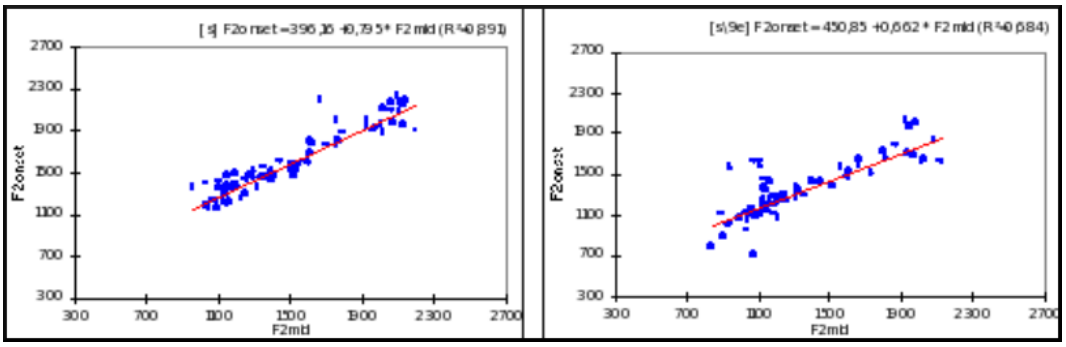

FIG. 6.9 – *droite de régression de la consonne non-pharyngalisée* [s] *à gauche et sa correspondante pharyngalisée* [sˤ] *à droite (AD).*

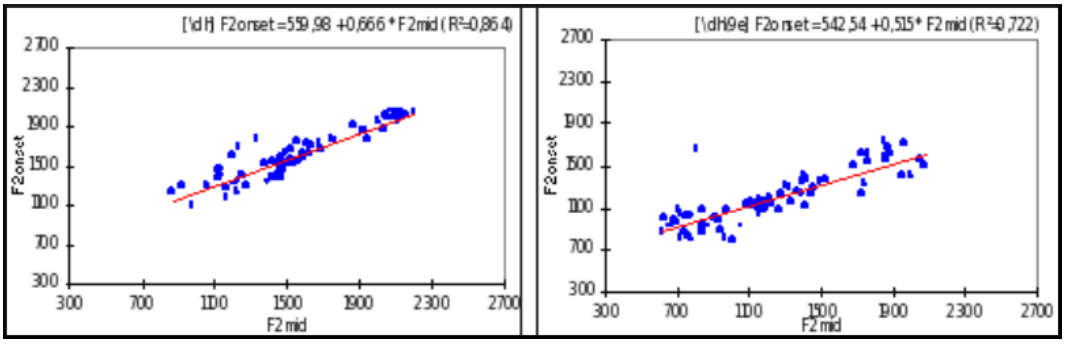

FIG. 6.10 – d*roite de régression de la consonne non-pharyngalisée* [ð] *à gauche et sa correspondante pharyngalisée* [ðˤ] *à droite (AD).*

6.2.3 Synthèse de nos résultats pour l'équation de locus

La série non-pharyngalisée résiste davantage en ASC qu'en AD aux effets de la voyelle, avec une transition très faible entre F_{2onset} et F_{2mid} tandis que la série pharyngalisée se présente avec un F_{2onset} moins sensible au timbre de la voyelle en arabe populaire qu'en ASC. La coarticulation CV en contexte pharyngalisé vs non pharyngalisé dans les deux variétés de langue présente une résistance croissante aux effets de la voyelle. Trois ordres de consonnes sur quatre l'illustrent :

1. ðˤ (POP) < ðˤ (ASC) < ð (POP) < d (ASC)
2. dˤ (POP) < dˤ (ASC) < d (POP) < d (ASC)
3. sˤ (POP) < sˤ (ASC) < s (POP) < s (ASC).

Dans cette partie nous avons montré que l'équation de locus appliquée à l'étude de l'ASC et de l'AD permet de mesurer le degré de coarticulation entre la consonne et la voyelle adjacente.

Nous avons montré qu'il était possible de distinguer sans ambiguïté deux séries de consonnes, les pharyngalisées et les non-pharyngalisées avec des valeurs dépendant de chaque variété de langue.

De ce fait nous avons montré que l'équation de locus permettait de distinguer des variétés de langues comme ASC *vs* POP.

6.3 Les résultats avec Matching Pursuit

Cette dernière partie explore les possibiltés offertes par la combinaison [Matching Pursuit, Compression, mesure de distance], aussi donnerons-nous surtout des résultats graphiques sans aller au fond des analyses possibles. Ces résultats corroborent ceux obtenus avec l'équation de locus, mais il nous semblent apporter bien plus d'éléments que celle-ci, ce qui devra faire l'objet d'investigations futures.

Conventions graphiques pour lire cette partie :

a) la première lettre correspond à l'initiale du pays. En minuscule il s'agit d'une seule occurence. En majuscule, il s'agit d'un regroupemment de plusieurs occurences : par exemples toute les consonnes pharyngalisées d'un locuteur ou d'une région ;

b) J ou j = Jordanie, K ou k = Koweit, M ou m = Maroc, Y ou y = Yémen ;

c) les consonnes non-pharyngalisées sont notées s et d ;

d) les consonnes pharyngalisées sont notées [sph ou s9e] et [dph ou d9e], (nous avons utilisé ph pour pharyngalisé ou 9e qui est une partie du code Unicode du symbole phonétique. Exemple : s9e ou sph correspondent à [$s^{\text{ʕ}}$] de code s\9e).

6.4 Méthodologie

6.4.1 Le corpus utilisé

Pour ce travail d'exploration, nous avons utilisé les productions d'un seul locuteur par région, choisi au hasard et nous avons restreint notre étude aux consonnes [s], [d], [$s^{\text{ʕ}}$] et [$d^{\text{ʕ}}$].

Nous avons utilisé le même corpus que pour l'équation de locus mais nous travaillons cette fois sur des séquences VCV traitées par MP. Nous avons obtenu d'emblée des résultats conformes à ce que nous avions trouvé avec l'équation de locus. Le type de séquence choisi ne nous favorise pas particulièrement à priori, puisqu'elle prend en compte, si cela a un sens ici, les éléments supplémentaires que sont V_1, la consonne entière et les transitions entre V et C ainsi que celles avec les éléments précédents et suivants VCV

Les séquences VCV ont été décomposées par le logiciel Guimauve avec une résolution de deux cent atomes. Lors de nos premières expériences nous avons trouvé que cela nous

donnait de bons résultats, ce que la suite n'a pas démenti, même s'il est probable qu'une optimisation soit possible.

6.5 Technologies d'analyse des ensembles d'atomes utilisées

Les fichiers d'atomes obtenus par MP imposent de nouvelles formes d'analyse prenant en compte des nuages de points en tant qu'objets indépendants. Nous nous sommes donc penché sur les théories et techniques qui permettent de traiter ce type de problèmes. La formation de clusters avec Weka ne nous a pas donné de résultats suffisamment réguliers : il y a trop souvent confusion entre les groupes d'atomes. Bien que les nuages de points soient bien formés à l'écran, il y a trop souvent des erreurs quant à leur appartenance. C'est donc avec le logiciel Complearn que nous entreprendrons l'étude des groupes de fichiers d'atomes.

6.5.1 Classification avec Complearn

CompLearn est développé par Cilibraci, R. and Vitanyi, P.[1] C'est une suite d'utilitaires simple à utiliser, issue d'une évolution en mathématiques théoriques concernant la complexité de Kolmogorov. L'approche est basée sur la compression des données et le programme peut déterminer des régularités dans des domaines complètement différents. On peut classer les styles de morceaux de musique, des écrits, identifier des compositeurs, des auteurs. Il permet de reconnaître la langue d'un texte, de découvrir des relations entre espèces vivantes et il s'est distingué dans la découverte de nouveaux virus.

Ces caractéristiques alléchantes nous ont fortement motivées. Beaucoup de domaines n'ont pas encore été explorés avec ce logiciel et nous avons décidé de le tester sur nos fichiers. La version ComplearnDemo pour les machines Microsoft^{TM}, fonctionne graphi-

[1] http ://homepages.cwi.nl/ paulv/papers/amdug.pdf

quement d'une façon spectaculaire par simple glisser déposer des fichiers à analyser dans une fenêtre.

6.5.2 Principe de Complearn

Dans un premier temps il faut décrire les objets à comparer sous forme de chaîne de caractères. Ensuite ces chaînes sont compressées avec un compresseur sans perte et Complearn analyse la distance qui existe entre les fichiers compressés. Une telle compression fait ressortir les éléments semblables dans le sens ou ils ne peuvent pas être compressés plus : ils sont réduits à leur plus simple expression. Les parties des fichiers ne pouvant être compressées gardent une forte entropie les unes par rapport aux autres et interdisent tout rapprochement. Plus les fichiers contiennent d'éléments similaires, plus ils sont proches. Pour la compression des fichiers nous avons utilisé le format .rar qui donne un bien meilleurs résultats que .zip. Il existe des formats de compression encore plus performants comme LZMA qui est un logiciel libre aux sources ouvertes, mais ayant fait nos premiers essais avec .zip puis ayant tout repris avec le format .rar, nous ne pouvions pas refaire une partie importante du travail, ni mélanger les formats d'autant que .rar donne une compression tout à fait adaptée à nos buts et qu'il est souvent utilisé avec Complearn.

Complearn construit dans un premier temps la matrice des distances entre éléments, puis un système de visualisation analyse cette matrice pour en tirer le graphe de similitude entre les fichiers.

Soit x et y deux objets à comparer mis sous forme de fichiers informatiques, C (x) est la longueur de la version compressée de x en utilisant un compresseur sans perte C. On a $C(x) = K(x)$, ou $K(x)$ est la complexité de Kolmogorov et

$$NCD(x,y) = \frac{C(xy) - min\{C(x), C(y)\}}{max\{C(x), C(y)\}}$$

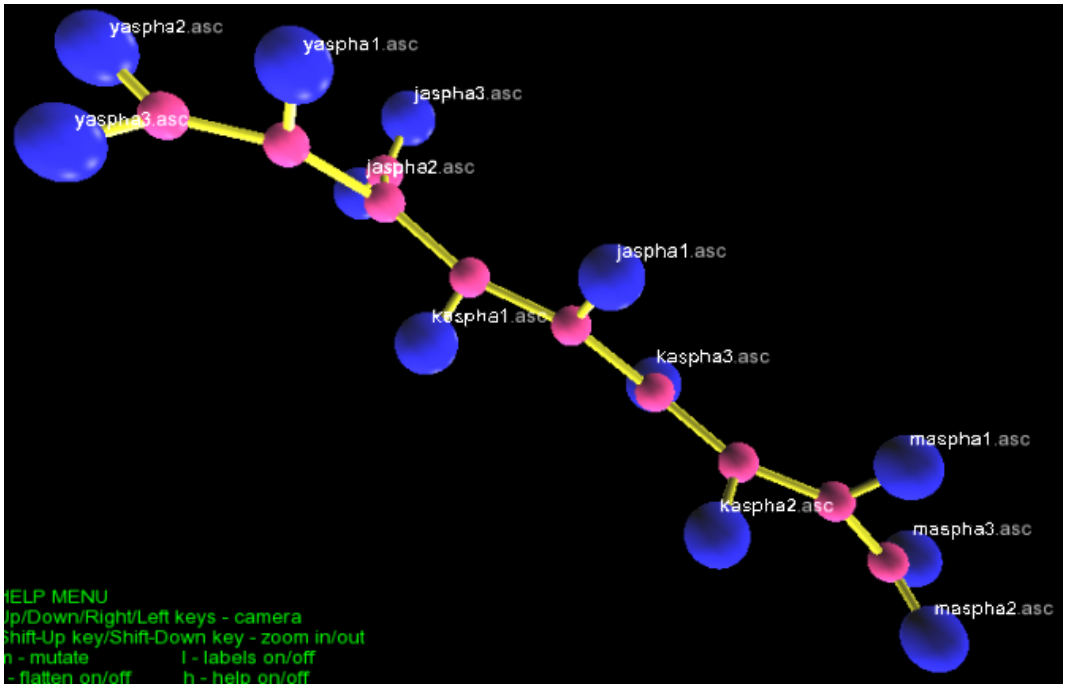

FIG. 6.11 – *arbre de regroupement pour trois occurences de* [as$^{\varsigma}$a] *par locuteur.*

ou NCD est la distance de compression normalisée entre x et y.

6.6 Les résultats obtenus

Nous ne donnerons ici que les résultats sous formes graphiques avec quelques commentaires. Une investigation plus poussée serait possible, mais rapellons qu'il s'agit ici d'un travail novateur utilisant MP et Complearn, travail qui a absorbé tous nos efforts. Il nous apparaît que les résultats obtenus sont suffisamment éloquents pour que notre méthode soit considérée comme une alternative valable à l'utilisation de la FFT et de l'équation de locus, mais elle peut l'être aussi pour d'autre études.

La figure 6.11 montre le regroupement obtenu pour les trois occurences de la séquence [as$^{\varsigma}$a] extraites des trois phrases prononcées par un locuteur de chacune des

quatre régions arabophones étudiées. Nous notons que les productions m [as$^{\varsigma}$a], y [as$^{\varsigma}$a] sont correctement classées tandis que k[as$^{\varsigma}$a] et j[as$^{\varsigma}$a] présentent une seule occurrence éloignée des deux autres. Dans ce test, Maroc et Yémen se retrouvent aux antipodes du centre tandis que Jordanie et Koweit partage ce centre.

Nous avons effectué le même exercice que précédemment avec [asa] (figure : 6.12. Nous observons sur la figure l'excellent regroupement des productions des locuteurs marocain et yéménite pour[asa] et une zone de confusion qui touche la Jordanie et le Koweit comme pour [as$^{\varsigma}$a]. Outre les qualités précédemment décrites, CompLearn fonctionne indépendamment de l'interface graphique sur les machines Unix, ce qui permet de l'inclure facilement dans une chaîne de traitement automatisée et de travailler simultanément sur un plus grand nombre de fichiers. Une meilleure lisibilité peut être attendue en sortie sous forme d'images Postscripts qu'avec l'interface que nous avons utilisé pour le prototypage présenté ici.

Sur la figure 6.13 regroupant des séquences [id$^{\varsigma}$i] le Koweit, le Maroc, le Yémen sont complètement regroupés. La Jordanie a une seule occurence éloignée des deux autres qui sont groupées.

Pour compléter cette étude nous avons regroupé consonnes pharyngalisées et non-pharyngalisées en un seul fichier par zone. Les Figures 6.14 et 6.15 présentent les résultats obtenus. Les symboles Cph représentent le regroupement de toutes les consonnes paryngalisées et C les non-paryngalisées d'un locuteur.

Seul le locuteur jordanien a une très faible distance entre pharyngalisées et non-pharyngalisées pour aCa avec prédominance des pharyngalisées. Comme sa production pharyngalisées se trouve correctement classée vis-à-vis de celles des autres locuteurs, les non-pharyngalisées se trouvent au même niveau (voir figure 6.14).

Par contre sur la figure 6.15, il y a une séparation sans ambiguïté entre iCi et iC$^{\varsigma}$i qui forment deux groupes bien distincts.

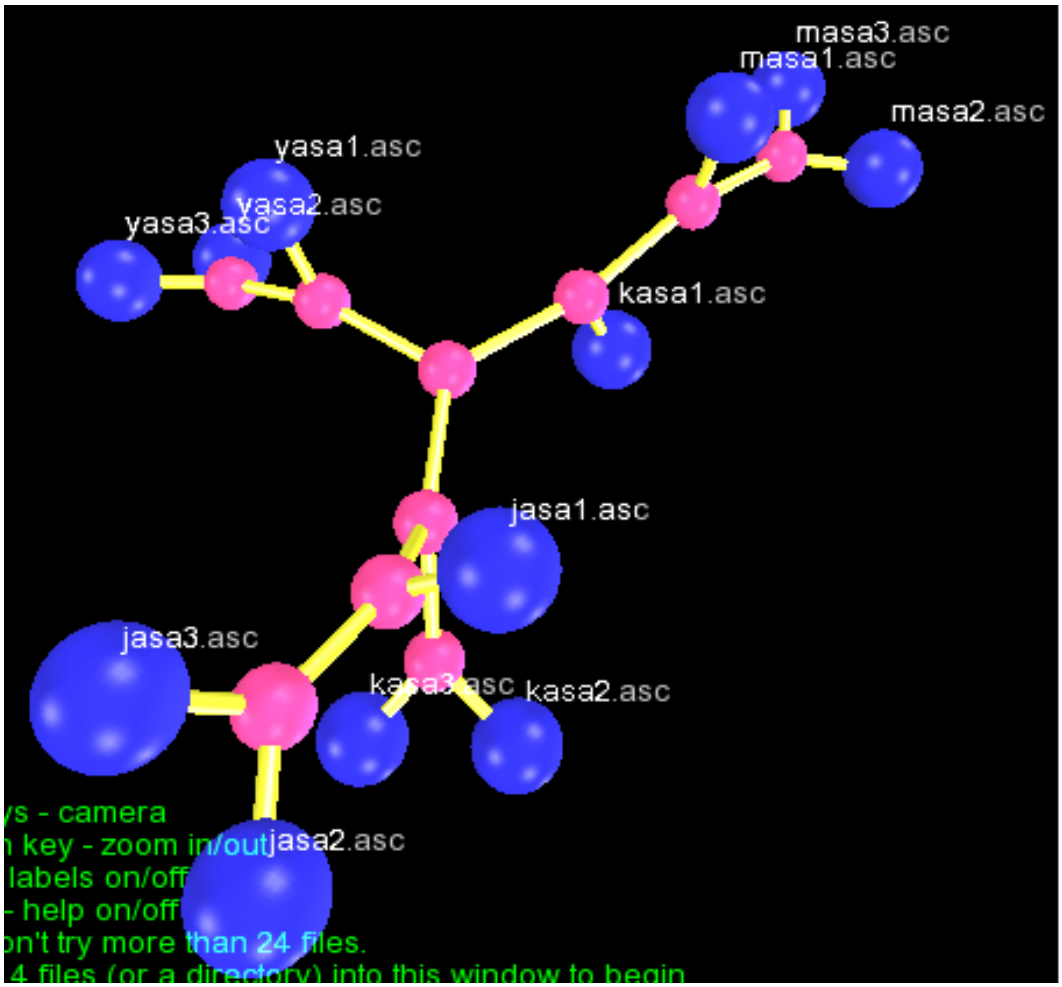

FIG. 6.12 – *arbre de regroupement avec* [asa] *pour les quatre régions arabophones.*

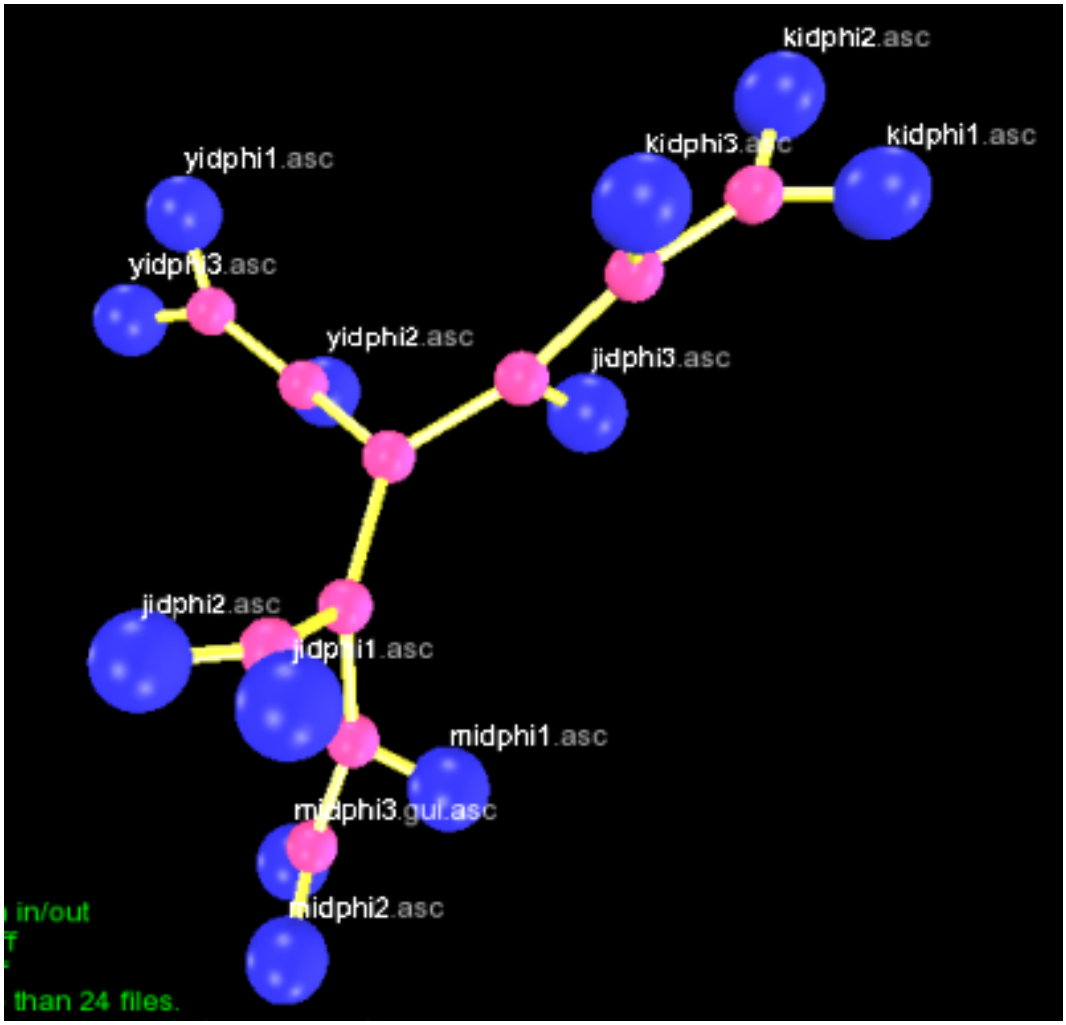

FIG. 6.13 – *le même test avec* [id$^{\mathrm{f}}$i].

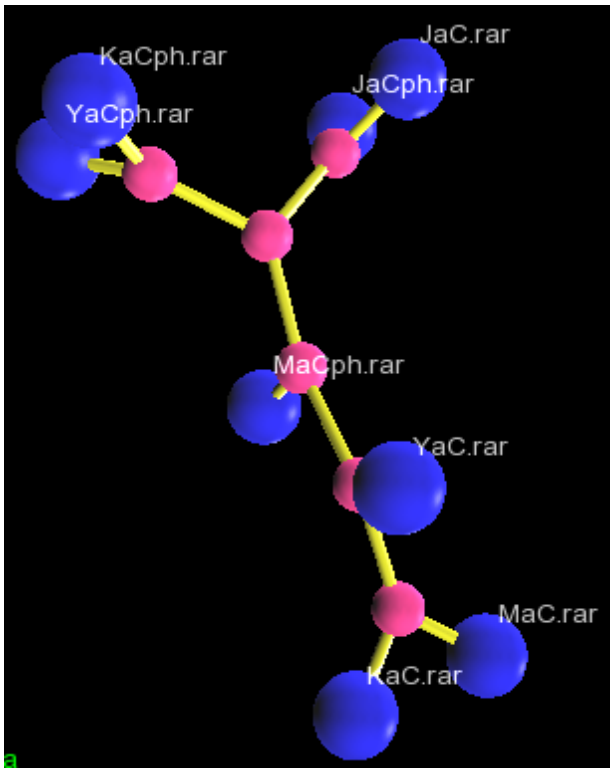

FIG. 6.14 – *carte des distances entre* aCa et aC$^{\zeta}$a.

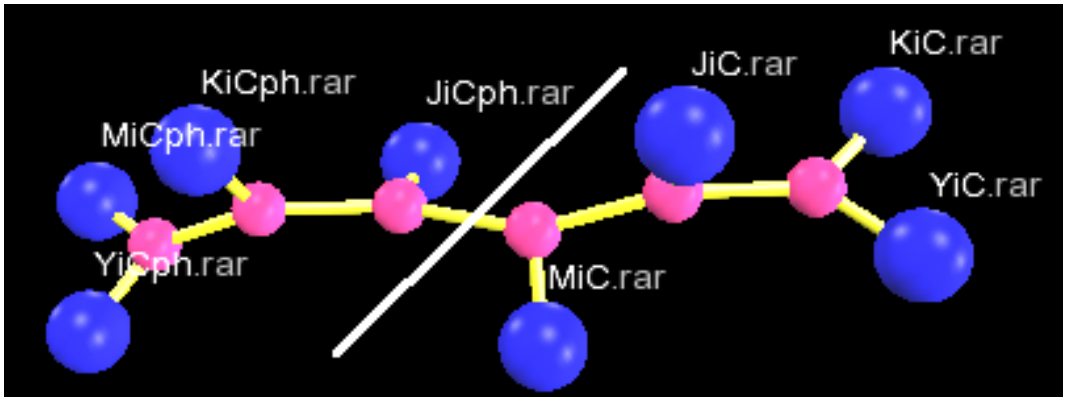

FIG. 6.15 – iCi et iC$^{\zeta}$i *forment deux groupes distincts.*

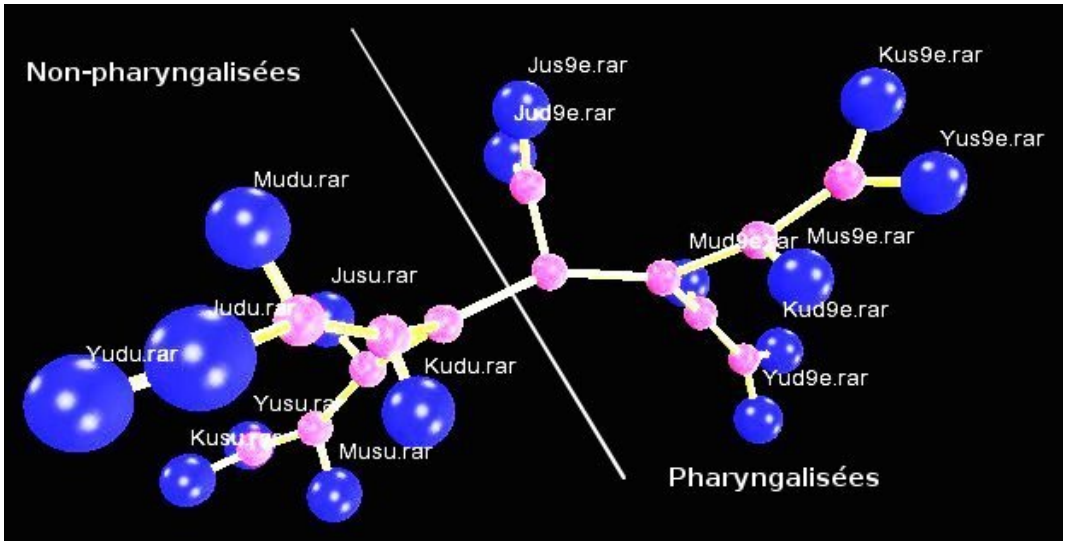

FIG. 6.16 – *sépartion des consonnes non-pharyngalisée* d *à gauche et la correspondante pharyngalisée* [d$^{\varsigma}$] *à droite en AD.*

6.7 Étude de la voyelle [u]

Les formes [uCu] nous ont posé très souvent des difficultés pour la mesure des deux premiers formants. Avec la chaîne d'analyse que nous avons mis au point, il semble que ces difficultés soient gommées, laissant entendre que notre méthode présente une grande robustesse vis-à-vis de la qualité de l'information qui lui est donnée.

6.7.1 Séparation des consonnes pharyngalisées et non-pharyngalisées

La figure 6.16 montre une expérience de tri des séquences groupées par pays contenant les consonnes non-pharyngalisées [d], [s], et les pharyngalisées [d$^{\varsigma}$] et [s$^{\varsigma}$]. Nous avons été surpris du résultat, nous attendant à une séparation moins tranchée et nous ne cacherons pas que la qualité du classement obtenu nous a laissé perplexe vu nos expériences antérieures avec les méthodes classiques.

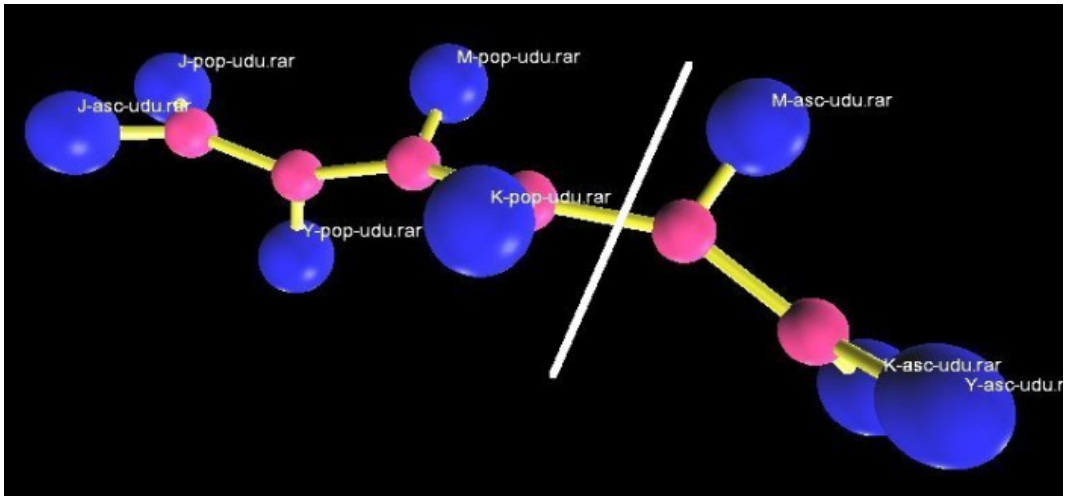

FIG. 6.17 – *séparation POP-ASC avec la consonne* [d].

6.7.2 Distinction POP-ASC avec [d]

Nous remarquons que l'ASC d'une part et le POP de l'autre sont bien séparé sauf pour la jordanie qui se trouve isolée avec se deux éléments POP et ASC dans l'espace POP.

6.7.3 Distinction POP-ASC avec [s]

Dans cet exercice nous remarquons qu'il y a l'ASC d'une part et le POP de l'autre. Seules les occurences en ASC en provenance de la Jordanie se retrouvent à l'extrémité opposée du groupe ASC avec le groupe POP. Nous notons que les classifications obtenues avec usu et udu sont, bien entendu, différentes.

6.7.4 Distinction POP-ASC avec [s] et [d]

La figure 6.19 nous donne un aperçu de la répartition du mélange des séquences [usu] et [udu] en arabe populaire et en arabe standard contemporrain. Nous observons

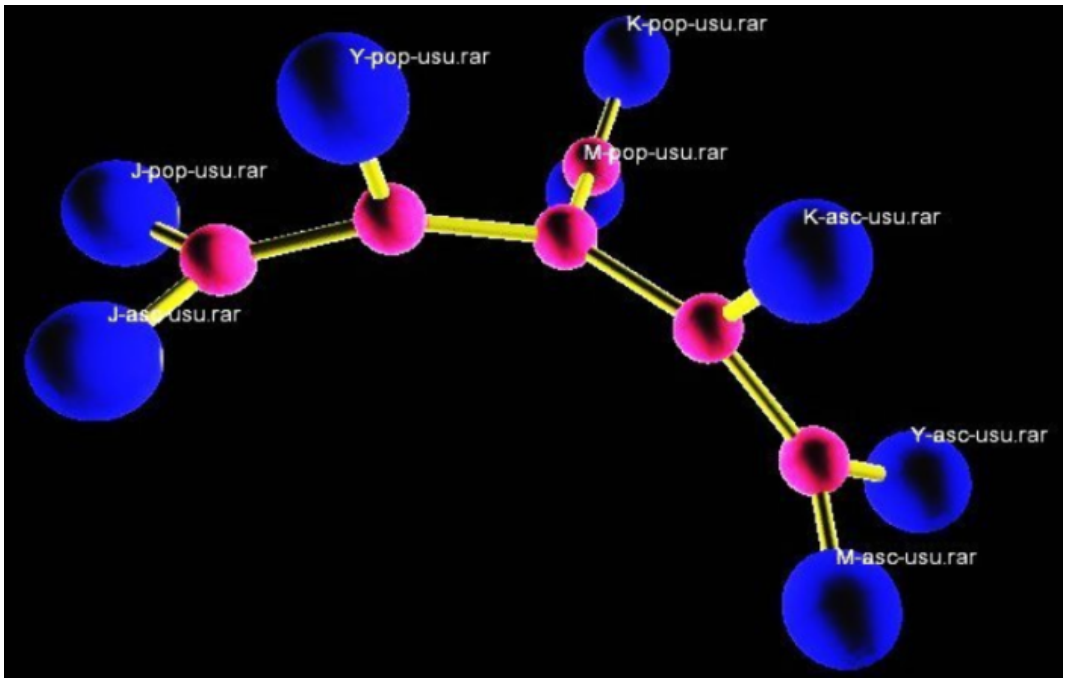

FIG. 6.18 – *séparation POP-ASC avec la consonne* [s].

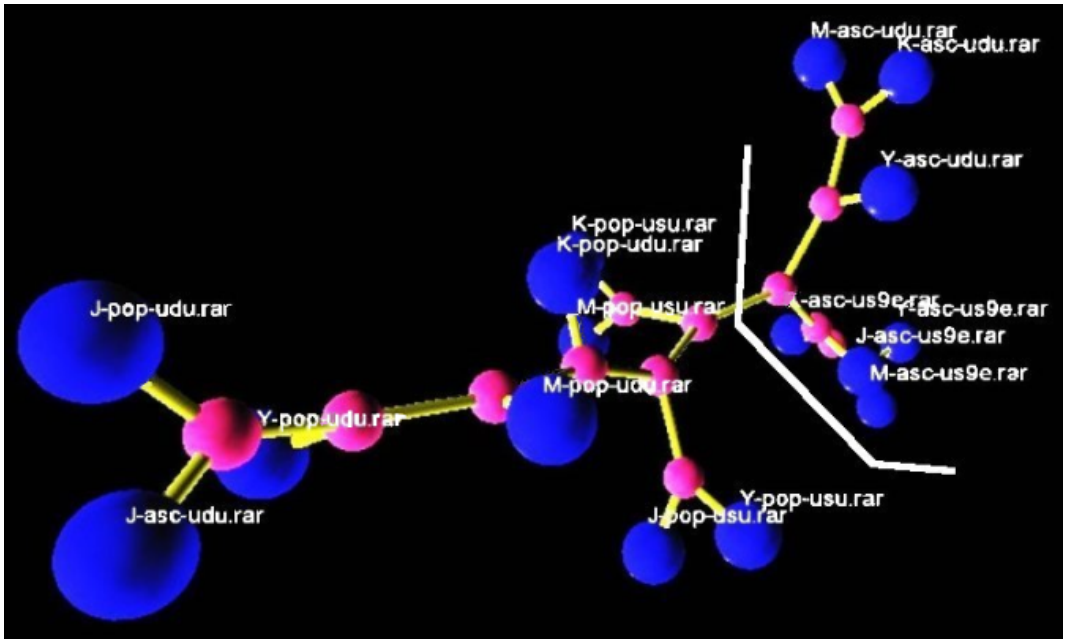

FIG. 6.19 – *séparation POP-ASC avec les consonnes* [d] et [s].

un regroupement plus serré en ASC qu'en POP. Cela paraît intuitif si l'on considère que les variétés POP sont indépendantes entre elles, reliées indirectement par la variété standard tandis que les variétés d'ASC, en contact directe, doivent être proches l'une de l'autre. Autrement dit que la variabilité de l'arabe POP doit être plus grande que celle de l'ASC et l'arabe POP doit être éloigné de l'ASC. Nous vérifions cette hypothèse ici. La mesure de distance se fait au nombre de nœuds et la ligne blanche que nous avons tracée marque la séparation très nette entre ASC et POP, sans zone de mélange.

6.8 Résultats globaux

Les résultats précédents nous donne un aperçu des classifications possible, en particulier la figure 6.19 nous donne un résulat général et détaillé des positions relatives des diverses productions. Nous avons voulu savoir s'il était intéressant de chercher à obtenir

des résultats globaux sur des regroupements encore plus généraux d'éléments VCV.

Pour chaque pays, nous avons réuni l'ensemble des séquences VCV et créé deux ensembles de séquences avec V non-pharyngalisées et V pharyngalisées. L'ensemble par pays 6.20 est noté du nom du pays, par exemple Koweit.rar, les autres ensembles sont notés par l'intiale du pays, v les trois voyelles et C les consonnes non-pharyngalisées ou Cph pharyngalisées. Chaque ensemble contenant toutes les occurences étudiées de chaque locuteur est sensé avoir une distance importante par rapport aux autres et nous l'utilisons comme attracteur pour les autres groupes.

Nous constatons en effet la constitutions de quatre pôles, Jordanie, Koweit, Maroc et Yémen. Ces pôles étirent quatres branches, une par pays, sur lesquelles viennent se greffer les ensembles séparés de pharyngalisées et de non-pharyngalisées relatifs à chaque pays. Il n'y a aucun mélange et les ensembles de pharyngalisées et de non-pharyngalisées sont proches de leurs attracteurs sur la branche corespondante.

Positionnement des zones dialectales par rapport à un attracteur

Selon le même principe que pour la figure 6.20 nous supposons que l'ensemble des productions de tous les locuteurs va nous donner un appui nous permettant de positionner l'ensemble des productions de chaque locuteur. La figure 6.21 montre que notre hypothèse est justifiée et les fichiers globaux Maroc, Yémen, Jordanie et Koweit se placent par rapport au fichier global, le Maroc se trouve proche de l'ensemble, puis vient le Yémen, la Jordanie et le Koweit se trouvent à l'opposé de l'attracteur, sur le même nœud. Notons que nous avons souvent observé que Jordanie et Koweit se trouvaient ensemble, voire mélangés

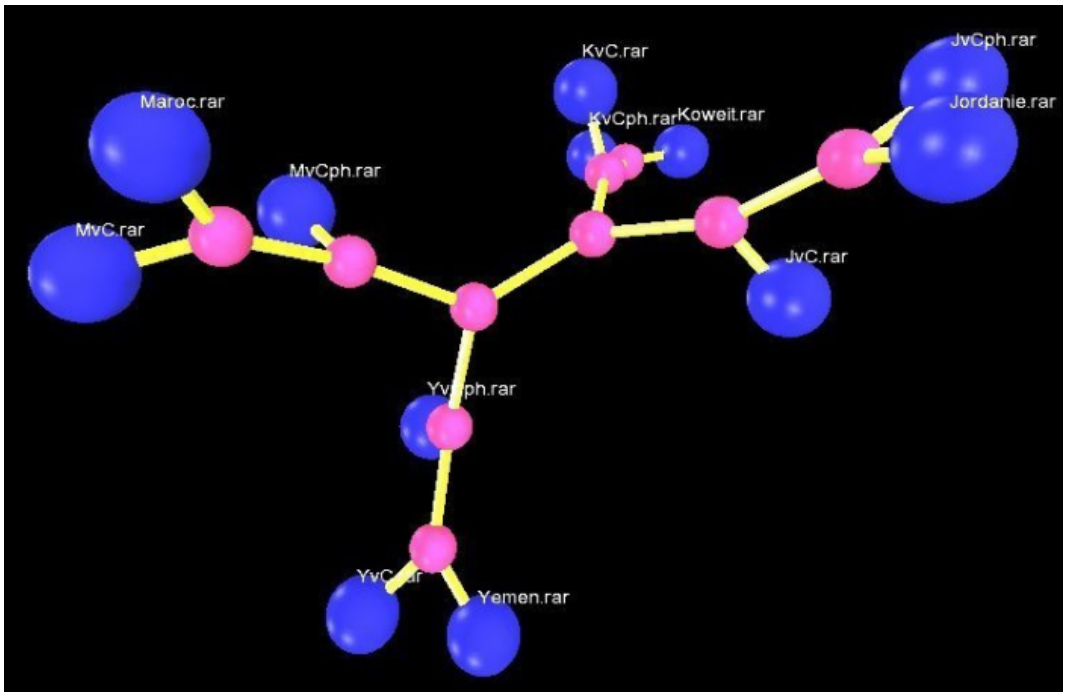

FIG. 6.20 – *placement de* C *et de* C^{ς} *par pays, par rapport à l'ensemble des productions*

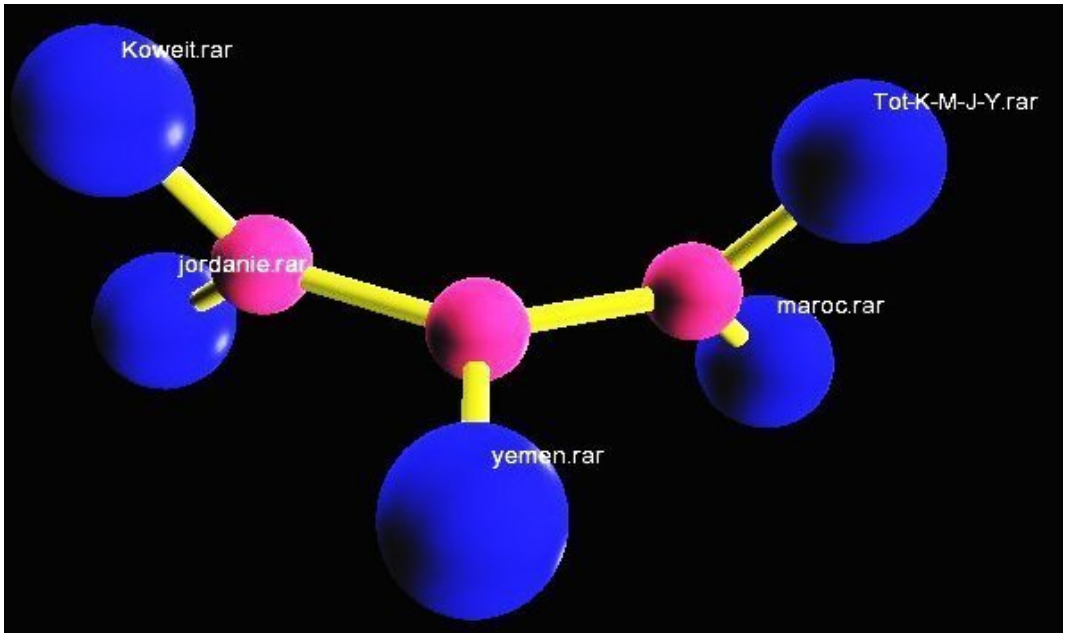

FIG. 6.21 – *Résultat global par pays*

6.9 Commentaires

Nous constatons que la décomposition atomique et l'analyse des fichiers d'atomes par le logiciel Complearn après compression sans perte, nous permet de séparer clairement les mélanges de séquences pharyngalisées *vs* non-pharyngalisées, les variétés de langue, les production intra et inter-locuteur.

Nous pouvons faire les mêmes distinctions qu'avec l'équation de locus mais bien d'autres expériences sont possibles. Une fois les données atomiques compressées, nous disposons d'une grande quantité de tests rapides et faciles à réaliser. Chaque test ne demande que des manipulations simples, ce qui évite les erreurs trop grossières.

L'obtention des fichiers sonores sur lesquels nous avons travaillé a pris un temps très court au regard de celui qui avait été nécessaire pour faire les mesures pour l'étude avec l'équation de locus.

Enfin, nous observons que l'information obtenue est beaucoup plus synthétique, plus lisible, avec notre méthode qu'avec l'équation de locus et nous ajouterons même, plus fine, car nous n'en avons pas tiré tous les enseignements, lesquels demanderaient encore beaucoup de travail.

Outre son élégance, l'outil décrit ici est extrêment efficace et versatile. Nous observons la grande qualité et la finesse des résultats de l'analyse donnée par l'association de Matching Pursuit et de Complearn dont nous espérons un jour voir se développer toute la puissance.

Chapitre 7

Conclusion

Pour ce travail nous sommes parti vers l'inconnu muni de notre seule intuition et finalement nous pouvons présenter une combinaison originale qui se résume dans l'association MPCK, (Matching Pursuit, Compression, complexité de Kolmogorov) que nous représentons selon le schéma 7.1 :

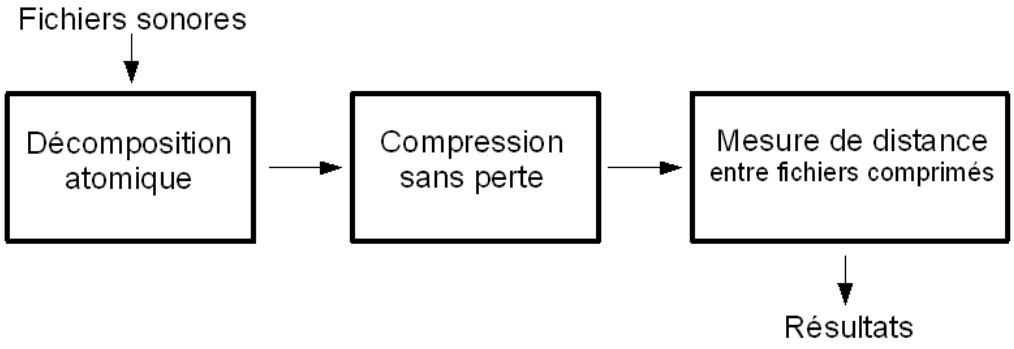

FIG. 7.1 – *schéma synoptique réprésentant MPCK*

Notre démarche nous a amené à utiliser la décomposition atomique adaptative du signal pour classer des variétés d'une même langue. Pour cela nous avons utilisé les quatre consonnes pharyngalisées [$t^ʕ$ $d^ʕ$ $s^ʕ$ $ð^ʕ$] et non–pharyngalisées [t d s ð] dans l'environnement des 3 voyelles cardinales [i a u] sous la forme VCV avec $V_1 \approx V_2$.

Dans un premier temps nous avions effectué une étude sur les mêmes phonèmes par les techniques habituelles employées en phonétique, à savoir la décomposition du signal en formants par FFT et l'emploi de l'équation de locus sur les mesures V_{2onset} et V_{2mid} de fréquence du deuxième formant de la deuxième voyelle.

L'utilisation de l'équation de locus nous a permis de mesurer le degré de coarticulation entre la consonne et la voyelle adjacente. En comparant les résultats entre l'ASC et l'AD et dans chaque variété entre les régions, nous avons pu mettre en évidence des différences caractérisées permettant de distinguer les variétés régionales et les variétés de langue.

La méthode du locus ayant donné les résultats que nous avons présenté, nous nous sommes proposé de tester MP. Malgré des jugements décourageants concernant notre théorie et notre projet, nous étions certains que l'élégance des théories mises en œuvre comme la décomposition atomique ne pouvait qu'apporter des résultats de qualité et dépasser les analyses classiques sur nombre de points. Nous pouvions espérer les améliorations significatives suivantes :

1. une présentation des résultats claire ;
2. la possibilité d'observer le signal dans ses parties les plus perturbées ;
3. une simplification des procédures, en particulier une diminution du nombre de mesures à faire tout en obtenant les mêmes informations qu'avec les méthodes classiques, sinon plus ;
4. une grande robustesse face à la viariabilité de l'objet étudié. Il semble que notre méthode élimine l'effet des coarticulations en début de V_1 et en fin de V_2 ;
5. une faible sensibilité à la précision des mesures, point qui est en rapport avec le précédent ;
6. une grande facilité pour préparer les expériences : une fois le traitement primaire du corpus effectué, les expériences doivent être facile à monter et à réaliser.
7. l'automatisation d'une grande partie des tâches qui doit être possible sans risque de résultats aléatoires.

Les résultats obtenus ne laissent pas de doute sur le fait que les tous points précédents puissent être atteints et qu'une grande partie de la chaîne d'analyse MPCK peut être automatisée. MP permet de traiter avec précision les zones perturbées du signal et l'ensemble de la chaîne d'analyse est très peu sensible au bruit : elle paraît d'une grande robutesse. Elle est également peu sensible à la précision des mesures, les extrémités de VCV ne semblant pas affecter le résultat global. Enfin nous avons pu expérimenter la facilité avec laquelle il est possible de varier de multiples combinaisons expérimentales.

S'il apparaît clairement que les résultats obtenus avec MP et Complearn sont étonnants, nous n'avons pu les obtenir que tardivement après une longue recherche en théorie de l'information. Revenant sans cesse sur les bases fondamentales, doutant de la démarche à suivre, il nous a fallut du temps et une multitude d'essais avant de trouver les bons logiciels et la bonne combinaison. Il existe d'autres programmes capables de faire les mêmes calculs, mais ils sont loin d'avoir les qualités de Guimauve et Complearn au niveau de leur utilisation dans ce cas précis.

Parmi les avantages de l'utilisation de notre système MPCK, (Matching Pursuit, Compression, complexité de Kolmogorov) nous observons que les résultats sont donnés avec beaucoup plus de finesse que dans le cas de l'équation de locus ; non seulement nous en tirons des enseignements équivalents, mais nous pouvons, de plus, évaluer une distance entre les séquences étudiées en les situant sur un arbre.

Pour lever toute ambiguïté, nous tenons à préciser ce qu'est et ce que n'est pas, pour nous, notre méthode. Commençons par ce qu'elle n'est pas :

1. ce n'est pas un système de reconnaissance de la parole ni du locuteur ;
2. elle n'est pas destinée à reconnaître les langues ;
3. elle n'est pas destinée à définir quels types de signaux est analysés mais à faire des comparaisons entre eux.

Le temps machine pris par les divers algorithmes est un obstacle à tout usage en reconnaissance de la parole en temps réel. Ensuite des éléments sans rapports entre eux

peuvent, par hasard, permettre des rapprochements. Il est nécessaire de bien définir la classe d'objets à étudier avant d'utiliser MPCK

Ce qu'est notre méthode :

1. c'est un ensemble logiciel permettant de classer des objets que l'on soupçonne avoir des structures sous-jacente proches.
2. c'est un système de classification fin ;
3. l'outil obtenu est très souple, versatile et polyvalent. Il permet de faire rapidement des comparaisons entre des séquences ondulatoire non-stationnaires plus particulièrement, là ou d'autres méthodes échouent où doivent recourir à des artifices pour réussir ;
4. c'est un outil très souple d'utilisation. Par exemple, il est possible en quelques manipulations donc en un temps très court, de comparer les productions d'un locuteur, les productions de plusieurs locuteurs, de varier les séquences à étudier, etc. ;
5. il s'agit d'un système dont la robustesse aux conditions initiales semble importante. En effet si nous distinguons bien les séquences $V_1 C V_2$ et $V_1 C^{\varsigma} V_2$, cela signifie que les éléments de coarticulation avec les phonèmes précédent V1 et suivant V2 sont minimisés ou éliminées.

L'impression déconcertante, provenant de l'usage de MPCK, tient a son aspect mystérieux de boîte noire sur laquelle nous n'avons apparemment aucune emprise et de laquelle sortent des résultats de grande qualité. En fait, à bien y regarder, nous pourrions en dire de même du sonagraphe dont le fonctionnement profond échappe à nombre de ses utilisateurs, même très qualifiés...

7.0.1 Perspectives

Nous n'avons pas jugé indispensable d'analyser l'ensemble du corpus, les résultats obtenus étant suffisament tranchés pour passer à d'autres travaux dont l'un des plus important est de mettre en chantier un outil plus pratique que la suite de programmes utilisée ici.

La version Unix de ces logiciels permet par construction leur chaînage et il est possible d'obtenir directement en fin de chaîne, une sortie graphique en très haute définition. Ce sont des questions de disponibilité de certains programmes informatiques et surtout de temps qui nous ont bloqué dans cette direction. Nous ne pouvions pas nous lancer dans un travail de programmation lourd tout en mettant au point la maquette de ce que nous souhaitions obtenir.

Nous avons décomposé les séquences VCV en 200 atomes parce qu'expérimentalement nous avons obtenu de bons résultats avec les séquences VCV étudiées. Cette question, bien que décrite théoriquement d'une façon plutôt obscure dans les manuels, doit être étudiée car nous ne savons pas encore définir la valeur convenable à utiliser en fonction du signal à analyser.

Nous n'avons pas utilisé MP sur les demi–voyelles qui ont servi à la mesure du locus alors que nous devrions nous attendre à obtenir des résultats très intéressants. C'est parce que les tests sur les séquences VCV nous ont apporté immédiatement les résultats recherchés. De nombreux tests sont encore à faire car la méthode n'a pas dévoilé toute la richesse de ses possibilités.

Un travail important consistera à évaluer les nouveaux outils du projet MPTK, mis au point à l'INRIA. Ces nouveaux programmes sont orientés vers la rapidité et l'efficacité (parcimonie), ce qui n'est pas nécessairement notre priorité, mais leur structure présente de multiples intérêts pour la réalisation d'une chaîne de traitement automatique.

Nous avons vu que MP permettait d'éliminer, d'ajouter ou de ne retenir que certains atomes du livre du signal. La richesse de ces possibilités laisse entrevoir la faisabilité

d'analyses en des points singuliers du signal, comme par exemple l'étude des transitions entre phonèmes.

Pour terminer, nous pouvons sans crainte affirmer que l'ensemble des outils employés pour ce travail est utilisable en l'état dans un confort relatif et que ceux–ci ouvrent des voies nouvelles à la recherche en phonétique tout en vérifiant et en approfondissant des travaux antérieurs par une finesse d'analyse encore jamais égalée.

Chapitre 8

Publications

A. Publications nationales et internationales (avec comité de lecture)

1. EMBARKI, M. & GUILLEMINOT, Ch. (2001), Conscience phonologique dans une situation de contact de langues, *Actes du VII^{ème} Symposium International de la Communication Sociale*, vol. II, pp. 147–151.
2. DODANE, C., GUILLEMINOT, C., (2003), « Influences de la formation musicale sur la restitution des voyelles d'une langue étrangère », 8^{e} Symposium International de Communication Sociale, Santiago de Cuba, 20–24 janvier, Actes II, pp. 1319–1324.
3. DURAND, C., GUILLEMINOT, C., (2003), « Où nous conduit l'hégémonie de la langue anglaise ? Quelques conséquences néfastes de l'utilisation de la langue anglaise en science et en technologie dans le contexte international : l'exemple français », 8^{e} Symposium International de Communication Sociale, Santiago de Cuba, 20–24 janvier, Actes II, pp.766–772
4. EMBARKI, M. & GUILLEMINOT, Ch. (2003), The moving boundaries of the first–acquired variety's phonological features: Evidence from production/perception of Moroccan Arabic's vowels, *Proceedings of 15th ICPhS*, pp. 639–642.
5. EMBARKI, M. & GUILLEMINOT, Ch. (2003), La conscience phonologique à

l'épreuve de la scolarisation : cas des langues en contact, *Actes du VIII^{ème} Symposium International de la Communication Sociale*, Vol. II, pp. 1052 –1057.

6. EMBARKI, M., GUILLEMINOT, Ch. & YEOU, M. (2006), Équation de locus comme indice de distinction consonantique pharyngalisé *vs* non pharyngalisé en arabe, *Actes des XXVIèmes JEP*, Dinard 12–16 juin, pp. 155–158.
7. EMBARKI, M., GUILLEMINOT, Ch. & BARKAT–DEFRADAS, M. (2007), Expansion nasale en arabe standard : indices acoustiques d'une coarticulation anticipatoire, *Revue Parole*, Vol. 39–40, pp. 209–234.
8. EMBARKI, M., YEOU, M., GUILLEMINOT, Ch. & AL MAQTARI, S. (2007), An acoustic study of coarticulation in Modern Standard Arabic and Dialectal Arabic: pharyngealized *vs* non–pharyngealized articulation, *Proceeds. of 16th ICPhS*, Saarbrücken, Germany, pp. 141–146.
9. EMBARKI, M., GUILLEMINOT, Ch. , YEOU, M. & AL MAQTARI, S. (2008), Voicing effects an absolute universal or language specific: New evidence from Modern Arabic and dialectal Arabic, M. Embarki (ed.), *Arabic and its Varieties: Phonetic and Prosodic Aspects*, special issue of *Languages & Linguistics*, 22 (à paraître).
10. GUILLEMINOT, Ch., YEOU, M., AL MAQTARI, S. & EMBARKI, M. (2008), Le voisement en arabe moderne, un indice de classement dialectal ? *Typologie des Parlers Arabes Modernes : Traits, Méthodes & Modèles de Classification.*
11. EMBARKI, M., GUILLEMINOT, Ch., YEOU, M. & AL MAQTARI, S. (2008), Directionality of coarticulation in Arabic VCV sequences, *Workshop international La Coarticulation : Indices, Sens et Représentation*, Montpellier, 7 décembre 2007.
12. EMBARKI, M., YEOU, M., GUILLEMINOT, Ch. & AL MAQTARI, S. (2008), Locus equation as an index of Arabic dialectal variation, *Phonetica* (sous révision).
13. EMBARKI, M., GUILLEMINOT, Ch., YEOU, M. & AL MAQTARI, S. (2008), Effets du voisement sur les obstruentes en arabe moderne, *Actes des XXVIIèmes JEP (Journées d'Etudes sur la Parole)*, Avignon, 9–13 juin, pp. 261–264.

B. Direction d'ouvrage

1. EMBARKI, M., DODANE, Ch., GUILLEMINOT, Ch. & YEOU, M. (2008), La Coarticulation : Indices, Direction et Représentation, Paris : l'Harmattan (à paraître).

C. Conférencier invité

1. EMBARKI, M., YEOU, M., GUILLEMINOT, Ch. & AL MAQTARI, S. (2007), An acoustic study of coarticulation in Modern Standard Arabic and Dialectal Arabic : pharyngealized vs non–pharyngealized articulation, J. Rosenhouse (cord.), Special session *Arabic Phonetics at the Beginning of the 3nd Millenium*, *ICPhS*, 6–10 August, Saarbrücken, Germany.

D. Conférences orales

1. EMBARKI, M. & GUILLEMINOT, Ch. (2000), Sociolinguistic representations and phonological awareness of Arabic languages spoken in Morocco, *4th AÏDA International Conference*, *Aspects of the Dialects of Arabic Today*, Marrakech 1–4 April, p. 20.
2. EMBARKI, M. & GUILLEMINOT, Ch. (2001), Conscience phonologique dans une situation de contact de langues, *VII^{ème} Symposium International de la Communication Sociale*, Santiago de Cuba, 24–27 janvier.
3. EMBARKI, M. & GUILLEMINOT, Ch. (2002), L'école publique au Maroc : un lieu de consolidation ou d'altération de la conscience phonologique ?, *5ème Conférence Internationale d'AÏDA*, Cadix, Espagne, 25–28 septembre.
4. EMBARKI, M. & GUILLEMINOT, Ch. (2003), La conscience phonologique à l'épreuve de la scolarisation : cas des langues en contact, VIII^{e} Symposium International de Communication Sociale, Santiago de Cuba, 20–24 janvier.

5. EMBARKI, M., GUILLEMINOT, Ch. & AL MAQTARI, S. (2006), Équation de locus en contexte pharyngalisé *vs* non pharyngalisé et variation régionale arabe, *7^e Conférence Internationale d'AÏDA*, Vienne, 5–9 septembre.
6. EMBARKI, M. (2007), Proximité et distance entre les dialectes arabes modernes : état de l'art et perspectives, Colloque internationale *Typologie des parlers arabes : traits, méthodes et modèles de classification*, Montpellier 14–15 mai.

E. Communications affichées

1. EMBARKI, M. & GUILLEMINOT, Ch., (2000), Linguistic awareness of monolingual and bilingual subjects: sociolinguistic and phonological representations of two languages in contact", 7th International Pragmatics Conference, Budapest, 9–14 July.
2. EMBARKI, M. & GUILLEMINOT, Ch., (2003) "The Moving Boundaries of the First–Acquired Variety's Phonological Features: Evidence From Production/ Perception of Moroccan Arabic's Vowels", Proceeds. of the 15th ICPhS, Barcelona, August 3–9.
3. EMBARKI, M., GUILLEMINOT, Ch. & YEOU, M. (2006), Équation de locus comme indice de distinction consonantique pharyngalisé vs non pharyngalisé en arabe, *XXVI^eJEP*, Dinard 12–16 juin.
4. GUILLEMINOT, Ch., YEOU, M., AL MAQTARI, S. & EMBARKI, M. (2007), Le voisement en arabe moderne, un indice de classement dialectal, Colloque internationale *Typologie des parlers arabes : traits, méthodes et modèles de classification*, Montpellier 14–15 mai.

Bibliographie

[1] J. Al-Tamini : L'équation de locus comme mesure de la coraticulation vc et cv : étude préliminaire en arabe dialectal jordanien. *XXV^e JEP*, 1, 2004.

[2] C. Alessandro : *Traitement automatique du langage parlé : Analyse, synthèse et codage de la parole*, volume 1, chapitre 1. Hermes, 2002.

[3] F. Auger, P. Flandrin, P. Gonzales et O. Lemoine : *Time-Frequency Toolbox for Use white Matlab*. CNRS, Rice University, 1996.

[4] C. Bascou : Modélisation de sons bruités par la synthèse granulaire. Rapport technique, Université Aix-Marseille II, 2004.

[5] A. Bejan : *Constructal theory of organization in nature : dendritic flows, allometric laws and flight, Design and Nature*. Brebbia, CA. and Sucharov, L. and Pascola, P., 1996.

[6] M. Benidir : *Théorie et traitement du signal*, volume 1, chapitre 1, page 1. Dunod, 2002.

[7] C.H. Bennett : Logical depth and physical complexity. in the universal turing machine : A half-century survey. *Oxford Univ. Press*, 1:227–257, 1988b.

[8] R. Bladon et A Al-Bamerni : Coarticulation resistance in english. *Journal of Phonetics*, 4:137–150, 1976.

[9] L.-J. Boë, P. Bessière et N. Vallée : When ruhlen's « mother tongue » theory meets the null hypothesis. *In XV^th International Congress of Phonetic Sciences, Barcelone*, 2003.

[10] L. J. Boë, J-L. Schwartz, R. Laboisière et N. Vallée : Integrating articulatory-acoustic constraints in the prediction of sound structures. *In Speech production seminar*, volume 4, pages 163–166, 1996.

[11] P. Boersma et D. Weenink : Praat. University of Amsterdam, June 2008.

[12] Yves Brostaux : *Introduction au système R.*

[13] M. Chen : Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22:129–159, 1970.

[14] A. Church : An unsolvable problem of elementary number theory. *Bull. Amer. Math. Soc.*, 41:332–333, 1935.

[15] R. Cilibrasi, A. Lissa Cruz et S. de Rooij : Complearn. http ://www.complearn.org/, 2008.

[16] R. Cilibrasi et P. Viteanyi : Clustering by compression. *IEEE Transactions on Information Theory*, 51:1523–1545, 2005.

[17] R Daniloff, R. Hammarberg : On defining coarticulation. *Journal of Phonetics*, 1:239–248, 1973.

[18] Kenneth de Jong et Bushra Zawaydeh : Comparing stress, lexical focus, and segmental focus : patterns of variation in arabic vowel duration. *Journal of Phonetics*, 30:53–75, 2002.

[19] J-P. Delahaye : La complexité mesurée. *Pour la Science*, 314:34–38, 2003.

[20] J.-P. Delahaye : Classer musiques, langues, images, textes et gé nomes. *Pour la Science*, 317:90–95, 2004.

[21] J.-P. Delahaye : Thé ories et thé orie de l'information. *)i(interstices*, 2008.

[22] P. C. Delattre, A. M. Liberman et F. S Cooper : Acoustic loci and transitional cues for consonants. *JASA*, 27:769–773, 1955.

[23] P. Denes : Effect of duration on the perception of voicing. *J. Acoust. Soc. Amer.*, 27:761–763, 1955.

[24] C. Dodane et C. Guilleminot : Détection de la stabilité de timbre des voyelles, vers une automatisation des tâches. *In XXIV^e JEP*, pages 101–104, 2002.

[25] Piotr Durka : *Matching pursuit and unification in EEG analysis*. Artech House Publishers, 2007.

[26] E. El Khoury : Nouvelle mèthode de segmentation et regroupement en locuteurs. *In Actes des Septièmes Rencontres Jeunes Chercheurs en Parole*, pages 76–79, ILPGA, Paris, juillet 2007.

[27] M. Embarki, C. Guilleminot et S. Al-Maqtari : équation de locus en contexte pharyngalisé vs non pharyngalisé et variation régionale arabe. *In AIDA 7*, septembre 2006.

[28] M. Embarki, C. Guilleminot et M. Yeou : équation de locus comme indice de distinction consonantique pharyngalisé vs non pharyngalisé en arabe. *In Actes des XXVIe JEP*, 2006.

[29] M. Embarki, M. Yeou, C. Guilleminot et S. Al-Maqtari : An acoustic study of coarticulation in modern standard arabic and dialectal arabic : pharyngealized vs non-pharyngealized articulation. *In ICPhS 2007*, 2007.

[30] E. Farnetani : V-c-v lingual coarticulation and its spatio-temporal domain. *In Speech production and speech modelling*, pages 93–110, 1990.

[31] J. E. Flege et R. F. Port : Cross-language phonetic interference : Arabic to english. *Language and Speech*, 24:125–146, 1981.

[32] C. A. Fowler : Invariants, specifiers, cues : An investigation of locus equations as information for place of articulation. *Perception and Psychophysics*, 55:597–610, 1994.

[33] C.A. Fowler : Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, 46:127–139, 1981a.

[34] C.A. Fowler : A relationship between coarticulation and compensatory shortening. *Phonetica*, 38:35–50, 1981b.

[35] D. Gabor : Theory of communication. *J. IEE (London)*, 93(III):429–457, November 1946.

[36] D. Gabor : Acoustical quanta and the theory of hearing. *Nature*, 159:591–594, 1947.

[37] D. Gabor : New possibilities in speech transmission. *The Journal of the Institution of Electrical Engineers*, Part III:369–387, 1947.

[38] J. S. Garofolo : Timit acoustic-phonetic continuous speech corpus. Linguistic Data Consortium, Philadelphia, 1993.

[39] S. Ghazeli : Du statut des voyelles en arabe, analyse thèorie. *In Numéro spécial « É tudes Arabes »*, pages 199–219, 1979.

[40] S. Ghazeli : La coarticulation de l'emphase en arabe. *Arabica*, 28:251–277, 1981.

[41] B. Gick, F. Campbell, S. Oh et L. Tamburri-Watt : Toward universals in the gestural organization of syllables : A cross-linguistic study of liquids. *Journal of Phonetics*, 34:49–72, 2006.

[42] G. Gonon : *Proposition d'un schéma adaptatif dans le plan temps-fréquence basé sur des critères entropiques. Application au codage audio.* Informatique, Université du Maine, 2002.

[43] M. Goodwin et M. Vetterli : Time-frequency signal models for music analysis, transformation, and synthesis. *In Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1996.

[44] Michel Grabisch : Classification et reconnaissance des formes. *In Logique Floue, Série ARAGO*, volume 14. Masson, 1994.

[45] Michel Grabish : Classification, 2000.

[46] R. Gribonval et E. Bacry : Harmonic decomposition of audio signals with matching pursuit. *IEE Transactions on Signal Processing*, VOL51:101–111, 2003.

[47] R. Hammarberg : The metaphysics of coarticulation. *Journal of Phonetics*, 4:353–363, 1976.

[48] N. Hardcastle, W. J. Hewlett : *Coarticulation. Theory, Data and Techniques.* Cambridge University Press, 1999.

[49] J.-L. Heim, L.-J. Boë et C. Abry : La parole à la portée du conduit vocal de l'homme de neandertal. nouvelles recherches, nouvelles perspectives. *CR Acad. Sc. Palevol*, 1:129–134, 2002.

[50] J.-L Heim, L.-J. Boë et P. Abry, C. Badin : Les hommes de neandertal étaient-ils handicapés du conduit vocal ? *Primatologie*, 6:219–262, 2004.

[51] W. J. Isdsardi : A phonological perspective on locus equation. *Behavorial and Brain Sciences*, 21 :2:270–271, 1998.

[52] A. N. Kolmogorov : Three approaches for defining the concept of information quantity. *Information Transmission*, vol. 1:3–11, 1965.

[53] S. Krstulovic et R. Gribonval : Mptk : Maching pursuit made tractable. *In ICASSP*, 2006.

[54] S. Krstulovic et R. Gribonval : *The Matching Pursuit Tool Kit.* INRIA, 2007.

[55] S. Krstulovic, R. Gribonval, P. Leveau et L. Daudet : A comparison of two extensions of the matching pusrsuit algorithm for the harmonic decomposition of sound. *IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, 49:259–262, 2005.

[56] D. Krull : Acoustic properties as predictors of perceptual responses : a study of swedish voiced stops. *Perilus*, 7:66–70, 1988.

[57] D. Krull : Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech. *Perilus*, 10:87–108, 1989.

[58] H. Laborit : *La nouvelle grille.* Gallimard, 1974.

[59] Ilse Lehiste : The search for phonetic correlates in estonian prosody. *In* Ilse Lehiste et Jaan Ross, éditeurs : *Estonian Prosody*, pages 11–35, 1997.

[60] Leroux : *Ondelettes et paquets d'ondelettes pour le traitement de la parole.* Thèse de doctorat, Université PARIS VI, 1994.

[61] S. Lesage, S. Krstulovic et R. Gribonval : Séparation des sources dans les cas sous-déterminé : Comparaison de deux approches basées sur des décompositions parcimonieuses. *In 2005 - GRETSI - Actes de Colloques*, 2005.

[62] P. Leveau : *Décompositions parcimonieuses structurées : application à la représentation objet de la musique. Modèles de signaux, algorithmes, applications.* Thèse de doctorat, Université Pierre Marie Curie-GET-Télécom, Paris, 2007.

[63] M. Li, X. Chen, X Li, B. Ma et P. Vitanyi : The similarity metrics. *IEEE transactions on information theory*, 50:3250–3264, 2004.

[64] B. Lindblom : On vowel reduction. Rapport technique, The Royal Institute of Technology, Speech Transmission Laboratory, 1963.

[65] B. Lindblom : Explaining phonetic variation : A sketch of the h&h theory. *In* Hardcastle W.J. March A., éditeur : *Speech Production and Speech Modelling*, pages 403–439. The Netherlands : Kluwer Academic, 1990.

[66] B. Lindblom, S. Guion, S. Hura, S.-J. Moon et R. Willerman : Is sound change adaptive ? *Rivista di Linguistica*, 7:5–36, 1995.

[67] I. Maddieson et A.-S. Abramson : Patterns of sounds by ian maddieson. *Acoustical Society of America Journal*, 82:720–721, août 1987.

[68] S. Mallat et L. Hwang, W : Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38(2):617–643, 1992.

[69] S. Mallat et Z. Zhang : Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41:3397–3415, 1993.

[70] Sharon Y. Manuel : The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *JASA*, 88(Manuel, Sharon Y):1286–1298, 1990.

[71] Robert McAllistera, James E. Flegeb et Thorsten Piskec : The influence of l1 on the acquisition of swedish quantity by native speaker of spanish, english and estonian. *Journal of Phonetics*, 30:229–258, 2002.

[72] L. Messaoudi : Etudes sociolinguistiques. Rapport technique, Université Ibn Toufail, 2003.

[73] L. Métoz, N. Vallée, I. Rousset, J.-L. Boë et P Bessière : Des formes phonétiques aux proto-formes de la langue originelle. analyse méthodologique et évaluation des limites. *In XXIV^e Journée d'étude sur la Parole*, 24-27 juin 2002.

[74] F Mitleb : Voicing effect on vowel duration is not an absolute universal. *Journal of Phonetics*, 12(1):23–27, 1984.

[75] G. Modarresi, H.M. Sussman, B. Lindblom et E. Burlingame : Locus equation encoding of stop place : revisiting the voicing/vot issue. *Journal of Phonetics*, 33:101–113, 2005.

[76] A. Moles : *Information theory and esthetic perception.* Urbana : University of illinois press., 1968.

[77] M.R. Molis, B. Lindblom, W. Castelman et Carré R : Cross-language analysis of vcv coarticulation. *JASA*, 95:2925, 1994.

[78] M. Mrayati, R. Carre et B. Guerin : Distinctive regions and modes : articulatory-acoustic-phonetic aspects. *Speech Commun.*, 9(3):231–238, 1990.

[79] M. J. Munro : Productions of english vowels by native speakers of arabic : acoustic measurements and accentedness ratings. *Language and Speech*, 36:39–66, 1993.

[80] NooteBoom : *Production and perception of Vowel Duration.* Thèse de doctorat, Univ. of Utrecht, 1972.

[81] P-Y. Oudeyer : *L'auto-organisation de la parole.* Thèse de doctorat, Paris 6, 2003.

[82] E. Paradis : *R pour les débutants.* Université Montpellier II, F-34095 Montpellier cedex 05, 2002.

[83] Peterson et Barney : Control methods used in a study of the vowels. *JASA*, 24:175–184, 1952.

[84] R. F. Port, S. Al-Ani et S. Maeda : Temporal compensation and universal phonetics. *Phonetica*, 3:235–252, 1980.

[85] P. Prandoni, M. Goodwin et M. Vetterli : Optimal time segmentation for signal modeling and compression.

[86] N. Rjaïbi-Sabhi : *Approches Historique, Phonologique et Acoustique de la Variabilité Dialectale Arabe : Caractérisation de l'Origine Géographique en Arabe Standard.* Thèse de doctorat, Université de Franche-Comté, Besançon, 1993.

[87] M. Rocha Iturbide : *Les techniques granulaires dans la synthèse sonore.* Thèse de doctorat, PARIS VIII, 1999.

[88] X Rodet et P. Depalle : Spectral envelopes and inverse fft synthesis, 1992.

[89] J. Rosenhouse : Arabic phonetics at the beginning of the third millenium. *In ICPhS XVI*, pages 131–134, 2007.

[90] S. Rossignol : *Segmentation et indexation des signaux sonores musicaux*. Thèse de doctorat, Paris 6, 2000.

[91] Cristèle Roux : Méthodes de tri pour l'extraction des signaux faibles.

[92] M. Ruhlen : *L'origine des langues*. Belin, 2000.

[93] M. Ruhlen et A. Langaney : *L'origine des langues*, page 7. Belin, 2000.

[94] H. Satori, M. Harti et N. Chenfour : Système de reconnaissance automatique de l'arabe basé sur cmu sphinx, 2007.

[95] J.L. Schwartz, L.J. Boë, N. Vallé et C. Abry : The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, 25:255–286, 1997.

[96] R.L. Schwartz et T. Phoenix : *Introduction à Perl*. O'Reilly, 2002.

[97] Claude Shannon : A mathematical theory of communication. *Bell System Technical Journal*, vol. 27:379–423, 1948.

[98] Manuel Sharon : Cross-language studies : relating language-particular coarticulation patterns to other language-particular facts. *In* W.J. Hardcastle & N. Hewlett Nigel, éditeur : *Coarticulation : Theory,Data and Techniques*, pages 179–198. Cambridge University Press, 1999.

[99] J.-L. Shwartz, P. Escudier et P. Tessier : *Reconnaissance de la parole*, volume 2, chapitre 5, pages 141–178. Hermes, 2002.

[100] K.N. Stevens : On the quantal nature of speech. *Journal of Phonetics*, 17:3–45, 1989.

[101] K.N. Stevens : *Acoustic Phonetics (Current Studies in Linguistics)*. New Ed edition, 2000.

[102] H. M. Sussman, E. Dalston et Gumbert : The effect of speaking style on a locus equation characterization of stop place of articulation. *Phonetica*, 55:204–225, 1998a.

[103] H. M. Sussman, D. Fruchter et A. Cable : Locus equations derived from compensatory articulation. *JASA*, 97 (5):3112–3124., 1995.

[104] H. M. Sussman, D. Fruchter, J. Hilbert et J. Sirosh : Linear correlates in the speech signal : The orderly output constrain. *Behavioral and Brain Sciences*, 21:241–299, 1998b.

[105] H. M. Sussman et K. A. Hoemeke : A cross-linguistic investigation of locus equation as a phonetic descriptor for place of articulation. *Journal of the Acoustical Sociéty of America*, 94 :3:1256–1268, 1993.

[106] H. M. Sussman, K. Hoemke et F. Ahmed : A cross-linguistic investigation of locus equations as relationally invariant descriptor for place of articulation. *JASA*, 94:1256–1268, 1993.

[107] H. M. Sussman, H. A. McCaffrey et S. A. Matthews : An investigation of locus equations as a source of relational invariance for stop place categorization. *JASA*, 90(3):1309–1325, 1991.

[108] Anne Szulmajster-Celnikier : La question de l'origine des langues : vaine quête du graal ? *Marges linguistiques*, 11:1–16, 2006. Collège de France.

[109] M. Tabain : Consistencies and inconsistencies between epg and locus equation data on coarticulation. *In* 5th International Conference on Spoken Language Processing, éditeur : *ICSLP-1998*, page paper 0668, 1998.

[110] M. Tabain : Coarticulation in cv syllables : a comparison of locus equation and epg data. *Journal of Phonetics*, 28:137–159, 2000.

[111] M. Tabain : Voiceless consonants and locus equations : A comparison with electropalatographic data on coarticulation. *Phonetica*, 59:20–37, 2002.

[112] M. Tabain, J.G. Breen et A.R Butcher : Vc vs. cv syllables : a comparison of aboriginal languages with english. *Journal of the International Phonetic Association*, 34:175–200, 2004.

[113] R Development Core Team : *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, cran édition, 2008. ISBN 3-900051-07-0.

[114] A. Turing : On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, 43(2):544–546, 1936.

[115] A. Turing : *Mechanical intelligence*. North-Holland Publishing Co., Amsterdam, The Netherlands, 1992.

[116] J. Véronis : Annotation automatique de corpus : panorama et état de la technique. l (ed.), c (pp. 111-129). paris : Editions hermès. [lire]. *In Ingénierie des langues*, pages 111–129. Hermès, 2000.

[117] Kees Versteegh et C. H. M. Versteegh : *The Arabic Language*. Edinburgh University Press, 1997.

[118] B. Vieru-Dimulescu, P. Boula de Mareuïl et M. Adda-Decker : Identification de 6 accents étrangers en français en utilisant des techniques de fouille de données. *In RJC Parole*, juillet 2007.

[119] P. Watzlawick, J. H. Beavin et D. Jackson : *Une logique de la communication*, chapitre 1, pages 57–65. Seuil, 2000.

[120] N. Wiener : Speech, language and learning. *J. acoust. Soc. Amer. 22*, 6:696–697, 1950c.

[121] I. H. Witten et F. Eibe : *Data Mining : Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2005.

[122] M. Yeou : Locus equation and degree of coarticulation of arabic consonants. *Phonetica*, 54:187–202, 1997.

[123] S. Zimmerman et S. Sapon : Note on vowel duration seen cross-linguistically. *Journal of the Acoustical Society of America*, 30:152–153, 1958.

Annexe A

Programmes et autres graphiques

Scripts utilisés pour formater les fichiers issus de Guimauve.

Les différents logiciels utilisés pour les expériences demandent des formatages différents et plus ou moins stricts pour les fichiers en entrée. Nous avons écrits ou adapté quelques scripts Perl (extension .pl) afin de rendre le passage d'un format à un autre plus facile et le traitement par lot. Chaque script est rendu exécutable (Unix) :

```
chmod 700  monsript.pl
```

et lancé dans le répertoire choisi :

```
./monsript.pl/chemin/monrépertoire/
```

Suppression de colonnes dans un fichier. Les données sont séparées pas une virgule ou par une espace et nous faisons la réécriture avec l'extention .csv, seule acceptée par Weka par exemple.

```
for  fichier in *asc
```

```
do
cat $fichier | awk  '{ print $4,",", $5 }' > $fichier.csv
done
```

Praat écrit la liste des séquences sélectionnées, mais ne prévoit pas de sauvegarder les fichiers sonores que l'on veut étudier. De plus Praat ne prévoit pas un nommage incrémental ou basé sur les étiquettes ce ce qui serait plus adapté à une sélection. Il faut donc renommer chaque fichier un à un. Nous n'en avions pas trouvé de script faisant la sauvegarde de l'ensemble des fichiers renommés et nous avions commencé à le faire fichier par fichier mais Cedric Gendrot nous a proposé d'en écrire un :

```
# Script Praat
# extraction automatique des éléments sonores
# sélectionnés sur une TextGrid
# par Cedric Gendrot, Chercheur, MCF Paris III
# Institut de Linguistique et de Phonétique Générales Appliquées
# 19, rue des Bernardins 75005 Paris

nb_objects = numberOfSelected()

# premier passage en revue pour identifier les fichiers
for x from 1 to nb_objects
fullName_'x'$ = selected$ (x)
type_'x'$ = extractWord$ (fullName_'x'$, "")
name_'x'$ = extractLine$ (fullName_'x'$, " ")
endfor

for x from 1 to nb_objects
fullName$ = fullName_'x'$
```

```
type$ = type_'x'$
name$ = name_'x'$
select 'fullName$'

if type$ = "TextGrid"
Write to text file... 'name$'.TextGrid
elsif type$ = "Sound"

Write to WAV file... 'name$'.wav
else
pause type de fichier non prévu ... on passe au suivant
endif
endfor
```

Transformations les fichiers sonores Guimauve n'accepte que le format ASCII avec l'extension .asc pour les fichiers de sons. Il faut donc convertir les fichiers binaires .wav classiques en fichiers texte. Nous employons un script Perl capable de traiter d'un seul coup l'ensemble d'un répertoire avec Sox pour les conversions.

Sox est un utilitaire capable de faire toutes les conversions imaginables entre les formats de fichiers sonores. Ici nous convertissons de plus la fréquence d'échantillonnage qui passe de 44100 Hz à 22050 Hz, argument [-r 22050]. Les commandes sox (programme externe à Perl et mv (move) faisant partie du Shell, sont appelées par la commande Perl system().

```
#!/usr/bin/perl
# Conversion de wav en asc
# Ouverture du rertoire en  ligne de commande de commande
opendir (DIR, $ARGV[0])  || die ("ne peut ouvrir $ARGV[0]");
```

```
@liste_rep = readdir(DIR);
closedir DIR;

# lecture du tableau contenant les noms de fichier
foreach $fichier (@liste_rep)  {
if ($fichier =~ /wav/)  {
system "sox $fichier -r 22050 $fichier.dat";
system "mv $fichier.dat  $fichier.asc";
}
}
```

Il est parfois nécessaire de supprimer les lignes d'entête aussi avons nous utilisé le scripts suivant :

```
#!/usr/bin/perl
# Conversion de wav en asc
# Ouverture du répertoire en ligne de commande de commande

opendir (DIR, $ARGV[0])  || die ("ne peut ouvrir $ARGV[0]");
@liste_rep = readdir(DIR);
closedir DIR;

# lecture du tableau contenant les noms de fichier

foreach $fichier (@liste_rep)  {
if ($fichier =~ /asc/)  {
system "sed -i  2d $fichier";
system "sed -i  1d $fichier";
}
```

```
}
```

Modification de la sortie de Guimauve

Le but est d'avoir avoir un fichier de sortie demandant moins de manipulations.

Il est toujours difficile de relire un programme informatique des années après. Malgré ces années, Fabien Brachère, auteur du logiciel Guimauve, a bien voulu se pencher à nouveau sur son programme. Il nous à proposé une modification et indiqué comment obtenir une sortie formatée selon nos besoins.

Il faut modifier la ligne 175 du fichier read_ sig.c contenu dans l'archive des sources guimauve-0.2.4.tar.gzpour n'écrire que les informations désirées dans le fichier de sortie :

Télécharger guimauve-0.2.4.tar.gz

ensuite taper : tar xvzf guimauve-0.2.4.tar.gz puis cd guimauve-0.2.4

Chercher read_ sig.c et changer la ligne 175 comme suit :

```
printf(fp,"%f %f %f %d \%d \n",word->atom[0]->timeId*book->dx,
word->atom[0]->freqId/(float)book->fftSize/book->dx,word->
atom[0]->coeff2,word->atom[0]->octave,i+1);
```

Par exemple pour obtenir uniquement le temps, la fréquence et l'intensité (coeff2), il suffit de remplacer cette ligne par :

```
printf(fp,"%f %f %f \n",word->atom[0]->timeId*book->dx,word->atom[0]->
freqId/(float)book->fftSize/book->dx,word->atom[0]->coeff2);
```

Ensuite remplacer l'ancien fichier read_ sig.c par le nouveau et compiler Guimauve. Cette modification ne vaut que pour les systèmes Unix même si Guimauve s'installe aussi

sur celui du leader du marché. Dans ce dernier cas, l'accès aux sources et la compilation est plus difficile.

Exemple de dictionnaire utilisé pour Matching Pursuit (dictionnaire de test pour MPTK par Benjamin ROY)

```
 <?xml version="1.0" encoding="ISO-8859-1"?>
<dict>
<libVersion>0.2</libVersion>
<blockproperties name="GAUSS-WINDOW">
<param name="windowtype" value="gauss"/>
<param name="windowopt" value="0"/>
</blockproperties>
<block uses="GAUSS-WINDOW">
<param name="type" value="gabor"/>
 <varparam name="fftSize">
<var>64</var>
<var>32</var>
<var>16</var>
<var>8</var>
</varparam>
<param name="windowLen" value="8"/>
<varparam name="windowShift">
<var>32</var>
<var>16</var>
<var>8</var>
</varparam>
    </block>
    <block uses="GAUSS-WINDOW">
<param name="type" value="gabor"/>
<param name="windowLen" value="256"/>
```

```
<param name="windowShift" value="64"/>
<param name="fftSize" value="256"/>
    </block>
        <block uses="GAUSS-WINDOW">
<param name="type" value="gabor"/>
<param name="windowLen" value="512"/>
<param name="windowShift" value="128"/>
<param name="fftSize" value="512"/>
    </block>
   <block uses="GAUSS-WINDOW">
<param name="type" value="gabor"/>
<param name="windowLen" value="1024"/>
<param name="windowShift" value="256"/>
<param name="fftSize" value="1024"/>
    </block>
<block uses="GAUSS-WINDOW">
<param name="type" value="harmonic"/>
<param name="fftSize" value="256"/>
<param name="windowLen" value="256"/>
<param name="windowShift" value="128"/>
<param name="f0Min" value="340"/>
<param name="f0Max" value="1000"/>
<param name="numPartials" value="5"/>
        </block>
<block uses="GAUSS-WINDOW">
<param name="type" value="harmonic"/>
<param name="fftSize" value="512"/>
<param name="windowLen" value="512"/>
<param name="windowShift" value="256"/>
<param name="f0Min" value="440"/>
```

```
<param name="f0Max" value="1000"/>
<param name="numPartials" value="10"/>
        </block>
       <blockproperties name="GAUSS-WINDOW-FOMIN" refines="GAUSS-WINDOW">
<param name="f0Min" value="440"/>
    </blockproperties>
<block uses="GAUSS-WINDOW-FOMIN">
<param name="type" value="harmonic"/>
<param name="fftSize" value="1024"/>
<param name="windowLen" value="1024"/>
<param name="windowShift" value="512"/>
<param name="f0Max" value="1000"/>
<param name="numPartials" value="10"/>
        </block>
<block uses="GAUSS-WINDOW-FOMIN">
<param name="type" value="harmonic"/>
<param name="fftSize" value="2048"/>
<param name="windowLen" value="2048"/>
<param name="windowShift" value="1024"/>
<param name="f0Max" value="1000"/>
<param name="numPartials" value="10"/>
</block>
<block>
<param name="type" value="dirac"/>
        </block>
</dict>
```

Utilisation de MIXMOD

MIXMOD est un logiciel libre de l'INRIA qui permet de traiter des problématiques

d'estimation de densités, de classification ou d'analyse discriminante. Il maintenu par Florent Langrognet du laboratoire de mathémathiques de Besançon. Il est développé par une équipe de chercheurs de l'INRIA et de plusieurs laboratoires. Des ingénieurs travaillent sur le projet dont Anwuli Echenim qui a bien voulu prendre sur ses loisirs pour nous aider à traiter nos données.

Anwuli Echenim a effectué pour nous une recherche des données les plus significatives obtenues à partir de Guimauve et écrit le progamme d'analyse correspondant. Les graphes obtenus sont des mélanges de gaussiennes. La documentation statistique de MIXMOD traite de cette question ainsi que de nombreux ouvrages de statistique/probabilités.

Pour utiliser MIXMOD, il faut :

1. installer SCILAB ;
2. installer MIXMOD ;
3. ouvrir SCILAB ;
4. exécuter la commande : « initMixmod.sci », pour cela, cliquer sur File -> Files Operations, chercher le fichier « initMixmod.sci »qui est dans le répertoire "MIXMOD", le sélectionner et cliquer sur « EXEC » : après cette étape, toutes les fonctions de mixmod sont chargées dans l'espace de travail ;
5. charger les données à étudier avec la commande « read ». Attention : il faut enlever la première ligne des fichiers texte qui contient généralement des labels, tandis qu'en paramètres la fonction « read »prend le nombre de lignes et de colonnes du fichier à lire ;
6. exécuter mixmod sur les données en suivant l'user's guide ;

Script MIXMOD/SCILAB (nouvelle version)

```
// script de Anwuli Echenin, ingenieur stagiaire au
```

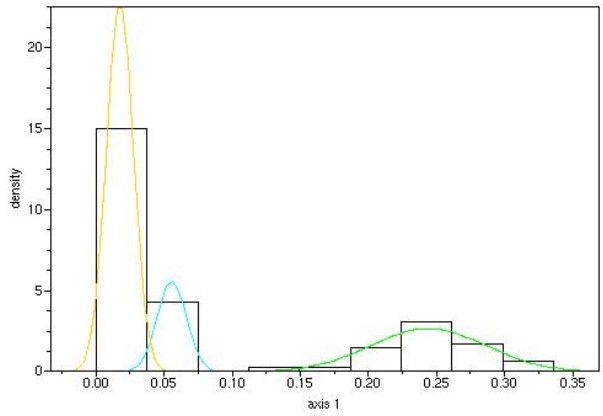

FIG. A.1 – *Représentation du mélange de gaussiennes dans le fichier d'atomes de* [asa]

```
// Laboratoire de Mathetiques  de Besancon
// export PathToMixmod=$HOME/MIXMOD
// export PATH=$PathToMixmod/BIN:$PATH

exec('/home/cege/MIXMOD/initMixmod.sci'); // chargement Fonctions
models = mixmodInputModel('allGaussianModel'); // var modes gaussiens
files=listfiles('AtomesMIX'); //var fichiers repertoire AtomesMIX
n=size(files,1); // nombre de fichiers
for i=1:n
data=read("PATH$/AtomesMIX/"+files(i),200,2); //lecture d'un fichier
data = data(:,2); // on s'intesse seulement aux frequences
output = mixmod(data,3,'model'); // exution de mixmod
mixmodView(output,list([1]),'densityComponent'); // graphiques
title(files(i)); // ajout d'un titre sur le graphique
end;
```

Exemples de graphiques obtenu

Au vu d'un ensemble assez important de résultats, nous avons estimé que l'interpré-

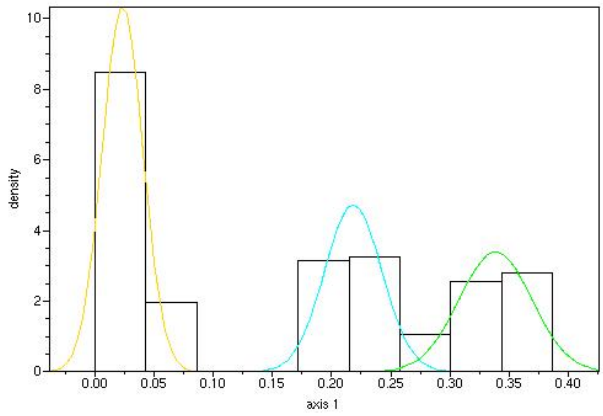

FIG. A.2 – *représentation d'une autre occurence de* [asa] *avec Mixmod*

tation de ceux-ci était trop ardue et nous découvrions Complearn. C'est pourquoi nous avons provisoirement délaissé la voie de MIXMOD.

Résumé

Le présent travail introduit en phonétique la décomposition atomique du signal, appelée aussi Matching Pursuit, traite les fichiers d'atomes par compression sans perte et enfin mesure la distance des fichiers comprimés par des algorithmes de Kolmogorov.

L'étalonnage est basé sur une première analyse classique de la coarticulation de séquences sonores VCV et CV, où V ∈ {i u a} et C ∈ {t d s ð } ∪ {t^{ς} d^{ς} s^{ς} $ð^{\varsigma}$}, extraites d'un corpus issu de quatre régions arabophones. L'équation de locus de CV vs $C^{\varsigma}V$, permet de différencier les variétés de langue.

La deuxième analyse applique un algorithme de décomposition atomique adaptative ou Matching Pursuit sur des séquences VCV et $VC^{\varsigma}V$ du même corpus. Les séquences atomiques représentant VCV et $VC^{\varsigma}V$ sont ensuite compressées sans perte et la distance entre elles est recherchée par des algorithmes de Kolmogorov. La classification des productions phonétiques et des régions arabophones obtenue est équivalente à celle de la première méthode.

Ce travail montre l'intérêt de l'introduction de Matching Pursuit en phonétique, la grande robustesse des algorithmes utilisés, et suggère d'importantes possibilités d'automatisation des processus mis en œuvre, tout en ouvrant de nouvelles directions d'investigation.

Abstract

The present work introduces in phonetics, the atomic decomposition of the signal also known as the Matching Pursuit and treats a group of atoms by compression without losses and finally measures the distance of the list of atoms compressed using the Kolmogorov's algorithms.

The calibration is based on an initial classical analysis of the co-articulation of sound sequences of VCV and CV, or V ∈ { i u a } and C ∈ {[t d s ð} ∪ { t^{ς} d^{ς} s^{ς} $ð^{\varsigma}$}) the excerpts culled from a corpus made up of four arabic speaking areas. The locus equation of CV vs C^{ς}, makes it possible to differentiate the varieties of the language.

In the second analysis, an algorithm of atomic adaptative decomposition or Matching Pursuit is applied to the sequences VCV and $VC^{\varsigma}V$ still on the same corpus. The atomic sequences representing VCV et $VC^{\varsigma}V$ are then compressed without losses and the distances between them are searched for by Kolmogorov's algorithms. The classification of phonetic recordings obtained from these arabic speaking areas is equivalent to that of the first method.

The findings of the study show how the introduction of Matching Pursuit's in phonetics works, the great robustness of the use of algorithms and suggesting important possibilities of automation of processes put in place, while opening new grounds for further investigations.