

Stratégies perceptuelles et identification automatique des langues

Melissa Barkat-Defradas*, Ioana Vasilescu, François Pellegrino***

** Laboratoire Dynamique du Langage (UMR CNRS 5596)
Institut des Sciences de l'Homme
14, avenue Berthelot 69363 Lyon cedex 07 – FRANCE
melissa.barkat@univ-lyon2.fr
francois.pellegrino@univ-lyon2.fr*

*** ENST- LTCI Département Traitement du Signal & de l'Image
46 rue Barrault
75634 Paris cedex 13 – FRANCE
Ioana.Vasilescu@enst.fr*

RÉSUMÉ : *L'identification Automatique des Langues (i.e., IAL) est par essence un champ de recherche pluridisciplinaire. Toutefois, la majeure partie des études menées dans ce domaine aborde cette problématique uniquement sous l'angle d'une extension multilingue des acquis de la Reconnaissance Automatique de la Parole. Cette approche n'est que modérément satisfaisante car elle délaisse les connaissances acquises dans les domaines de la linguistique et de la psycholinguistique, pourtant susceptibles de permettre une amélioration significative des systèmes automatiques. Cet article présente un état de l'art des expériences faites en identification perceptuelle des langues, puis montre comment l'interprétation des stratégies perceptives développées par des auditeurs humains pour identifier des dialectes arabes peut influencer la conception d'un système automatique.*

MOTS-CLES : *Identification Automatique des Langues, perception de la parole, dialectes arabes, systèmes vocaliques.*

ABSTRACT : *Automatic Language Identification (i.e., ALI) is clearly an area of research requiring much interdisciplinary effort. However, the major part of the research simply consider ALI as a multilingual development of the methods developed in Automatic Speech Recognition. This approach is only efficient to some extent since it ignores the knowledge gathered both by the linguistic and psycholinguistics approaches, even though they may lead to significant improvements in the results of the automatic systems. The present work provides a state of the art of the experiments led in perceptual identification of languages and then focuses on how the interpretation of the perceptual strategies developed by human listeners for the identification of Arabic dialects may be taken into account for the design of an automatic system.*

KEYWORDS : *Automatic Language Identification, Speech Perception, Arabic Dialects, Vocalic Systems.*

1. Introduction

A l'heure où l'on parle de plus en plus de communication internationale et de mondialisation, les approches courantes en Identification Automatique des Langues (désormais IAL), bien loin de mettre en exergue une utilisation exclusive de la langue anglaise, nécessite une prise en compte efficace de la diversité des langues et dialectes parlés par les usagers. L'espace européen, de plus en plus multiculturel et multiethnique, fournit ainsi un exemple très représentatif de ces nouvelles exigences et des thèmes de recherche qui en découlent. Parmi ceux-ci, l'IAL à partir de la parole, se révèle ainsi en pleine expansion. Le travail que nous présentons ici est né de la volonté de répondre à plusieurs questions, axées autour des interactions entre linguistes et chercheurs en traitement automatique des langues, et pouvant permettre d'améliorer les systèmes d'IAL ; il est centré sur l'étude des processus d'identification perceptuelle des dialectes arabes et est issu de l'observation selon laquelle les stratégies développées par les humains pour identifier des idiomes étrangers présentent une robustesse toute particulière, reconnue par les deux communautés de chercheurs suscitées. Par conséquent, le paradigme expérimental que nous avons employé vise à faire émerger une hiérarchie d'indices perceptuels discriminants potentiellement intéressants pour une modélisation dans un système d'IAL. Par ailleurs, des tests automatiques préliminaires montrent la saillance des indices ainsi mis en évidence et nous encourage dans la poursuite de ce type de démarche.

Les méthodes expérimentales en identification *perceptuelle* des langues, développées durant les deux dernières décennies, fournissent un élément de comparaison intéressant quant aux performances des systèmes automatiques, mais elles peuvent également être considérées comme un moyen de mettre en évidence d'éventuelles sources d'amélioration des systèmes en question. L'idée qui sous-tend cette approche est que l'être humain demeure l'un des « systèmes » de reconnaissance les plus performants en termes de nombre de langues traitées, d'apprentissage et de taux de réussite lors d'un test de discrimination. Cette affirmation s'appuie sur plusieurs expériences menées pour évaluer la perception humaine dans une perspective de reconnaissance linguistique, en particulier pour fournir un étalonnage pour des systèmes d'IAL. Muthusamy *et al.* (1994) ont par exemple dressé une évaluation des approches automatiques, en comparant leurs performances avec celles d'auditeurs naïfs testés à l'aide des mêmes corpus. La conclusion des auteurs est que les indices utilisés par les sujets humains afin de discriminer des langues inconnues et qui ont été partiellement décrits par ceux-ci dans une phase post test, pourrait fournir de nouvelles pistes quant à l'amélioration des performances des systèmes automatiques. Notons aussi que Pols (1997) parvient à la même conclusion lorsqu'il évalue les réactions des sujets humains face aux difficultés de reconnaissance dues à la dégradation progressive de la qualité des signaux d'entrée : la capacité de reconnaissance d'un être humain est préservée malgré toute perturbation ou distorsion du signal, qu'il s'agisse de bruit, de réverbération, ou de distorsion spectrale ou temporelle. L'être humain semble avoir la capacité de restaurer l'information acoustique manquante (Greenberg *et al.*, 2001 ; Barkat *et al.*, 2002). Toujours selon Pols (1997), la compréhension des aspects perceptuels constitue un pré-requis indispensable pour les développeurs de systèmes de reconnaissance automatique et elle doit être considérée comme une source potentielle de progrès. Enfin, avec le même objectif Lippmann (1997) dresse un bilan des travaux qui comparent les performances des systèmes modernes de reconnaissance de la parole avec celles des locuteurs humains. Ces observations mènent l'auteur à estimer que des améliorations devraient être apportées par une modélisation acoustico-phonétique de bas niveau adéquate, par l'amélioration de la robustesse face au bruit et à la variabilité des conditions de transmission, ainsi que par la modélisation plus précise de la parole. L'être humain apparaît comme capable de développer des stratégies sophistiquées pour extraire les informations les plus saillantes pour la reconnaissance et pour contrebalancer les données manquantes dans des conditions variables.

Il semble donc pertinent *a priori* de considérer les critères utilisés par l'homme pour la discrimination des langues comme potentiellement pertinents pour l'amélioration des systèmes d'IAL. Ces critères font appel à de l'information relevant de plusieurs niveaux linguistiques : segmental, supra-segmental et/ou phonotactique.

Les sections suivantes proposent un état de l'art des différents types d'expériences perceptives d'identification réalisées jusqu'à ce jour avec des sujets humains et insistent tout particulièrement sur la nature et la pertinence des indices acoustiques mis en évidence par ces méthodes.

2. Les expériences d'identification des langues en parole naturelle

Les premières expériences *en parole naturelle* ayant pour but la comparaison des résultats obtenus par des sujets humains avec ceux résultant de la reconnaissance automatique sont mentionnées par Muthusamy *et al.*

(1994) et datent des années 60. Le paradigme expérimental consiste à présenter une suite de stimuli sonores généralement de courte durée à un groupe d'auditeurs. Les stimuli sont extraits d'enregistrements acoustiques en plusieurs langues que les sujets testés ne parlent pas. On demande aux sujets d'identifier la langue d'origine de chaque stimulus. Une phase d'apprentissage leur permet de se familiariser avec le matériel acoustique et la tâche expérimentale. Le principal traitement des réponses des sujets consiste en l'analyse statistique des scores de réussite. Toutefois, Muthusamy et ses collaborateurs (1994) estiment que le nombre restreint de locuteurs et le type de matériel utilisé limitent la portée de cette première tentative. En effet, le premier corpus qui a réellement permis de mener des expériences perceptives adaptées a été OGI-TS (Muthusamy, 1993). Ce corpus est basé sur de la parole téléphonique dans 11 langues (i.e., anglais, farsi, français, allemand, hindi, japonais, coréen, mandarin, espagnol, tamul et vietnamien) produites par 90 locuteurs différents.

Les expériences utilisant des stimuli en parole naturelle peuvent être groupées en deux catégories, selon le type de matériel linguistique employé par les auteurs. Ainsi, un premier groupe est représenté par les études s'intéressant à la réussite en identification, indépendamment des conditions d'écoute et/ou des particularités linguistiques des idiomes proposées. Le second groupe réunit des expériences utilisant des corpus choisis de telle sorte que l'identification puisse permettre de déterminer des stratégies perceptuelles de nature linguistique, dépendantes des conditions expérimentales ou des spécificités linguistiques des langues considérées dans le corpus (langues d'une même famille, par exemple).

L'une des premières études significatives appartenant à la première catégorie est celle de Lorch *et al.* (1989). Les auteurs ont testé les capacités des auditeurs à identifier des langues inconnues après 20 secondes d'écoute, à faire la différence entre les sons spécifiques à chaque langue, à transcrire des syllabes, voire des mots, et à déterminer la famille des langues testées. Plus tard, Muthusamy *et al.* (1992) ont testé la capacité de sujets monolingues américains à identifier 10 langues inconnues à partir de stimuli dont la durée varie entre 1 et 6 secondes. Bien que la plupart des langues à identifier aient été complètement inconnues des sujets, ces derniers sont parvenus à obtenir des taux de reconnaissance correcte intéressants : pour 1 seconde de parole : 36 % d'identification correcte ; pour 2 secondes de parole : 43 % d'identification correcte ; pour 4 secondes de parole : 51,2 % d'identification correcte et enfin, pour 6 secondes de parole ils obtiennent 54,6 % d'identification correcte. Une étude ultérieure (Muthusamy *et al.*, 1994) a permis de confirmer ces performances même avec des stimuli enregistrés au téléphone. Il s'agissait dans cette étude de tester une population représentée par 10 sujets américains écoutant 10 langues non-familiales ainsi que des sous-groupes de 2 sujets non-anglophones dont la langue maternelle était présente dans le test et qui avait pour tâche d'identifier les 9 autres langues. Les échantillons de parole présentaient une durée de 1 à 6 secondes. Les résultats ont montré que les scores d'identification dépendent de la durée de l'extrait et de la familiarité des sujets avec les langues testées, laquelle est liée au degré d'exposition aux langues en question.

Dans la catégorie des études axées sur le rôle des particularités sonores des stimuli en termes de qualité de la parole ou de spécificités structurelles des langues, citons d'abord les résultats obtenus par Bond *et al.* (1991). Ces auteurs, en testant la capacité des adultes humains à discriminer les langues en condition bruitée où des informations importantes du signal sont absentes, confirment à leur tour les observations faites dans les études antérieures. Ces expériences, réalisées avec des sujets adultes, ont par la suite été étendues à des enfants d'âge scolaire aboutissant à des performances comparables (Stockmal *et al.*, 1994). Une autre expérience menée par Stockmal *et al.* (1996) à partir de stimuli en arabe, japonais, chinois, indonésien, russe et espagnol visait à déterminer les propriétés acoustiques des stimuli permettant aux sujets de différencier ces langues appartenant à des familles linguistiques éloignées. Les sujets ont réussi à grouper les langues sur la base de critères linguistiques tels que les propriétés psycho-acoustiques de la fréquence fondamentale, le type syllabique de la langue ou le débit de parole. Afin de distinguer l'information caractéristique de la langue de l'information caractéristique du locuteur, Bond *et al.* (1998) ont mené une étude utilisant un corpus de 8 paires de langues produites par des locuteurs bilingues. Selon les trois auteurs, les auditeurs naïfs sont capables de différencier l'information spécifique à la langue de celle caractéristique de la voix du locuteur. Afin de préciser leurs résultats, Stockmal *et al.* (1999) ont mené par la suite deux expériences dont l'objectif était de déterminer les indices permettant d'une part, la discrimination du coréen par rapport à d'autres langues attestant une structure rythmique similaire et, d'autre part, la discrimination du coréen par rapport à d'autres langues parlées dans la même zone géographique. Il s'est avéré que les indices utilisés pour la discrimination du coréen aussi bien des langues asiatiques environnantes que des idiomes éloignés montrant des particularités rythmiques analogues, étaient de nature complexe, linguistique et extra-linguistique. Contre toute attente, la première partie de l'expérience – consacrée à l'identification des langues à partir de motifs rythmiques – a montré que les indices exploités pour l'identification des différentes langues sont d'abord de nature géographique avant d'être directement liés aux propriétés phonotactiques des idiomes présentés lors du test ; alors que pour la seconde

partie de l'expérience – consacrée à l'identification par zone géographique – les indices retenus concernent les caractéristiques des voix des locuteurs, les variations de fréquence fondamentale et les particularités de l'inventaire vocalique de chaque langue. Enfin, les travaux de Marks *et al.* (1999) ont confirmé le rôle de la phase d'apprentissage dans l'amélioration des performances en identification linguistique par des sujets humains, hypothèse déjà testée par Ohala *et al.* (1979) comme nous le verrons dans le paragraphe 4. Dans le même ordre d'idées, il semble que les sujets polyglottes bénéficiant, par essence, d'une exposition fréquente à plusieurs langues, tirent un meilleur profit de la phase d'apprentissage et obtiennent des scores d'identification supérieurs à ceux des sujets monolingues.

Enfin, les expériences menées par Vasilescu (2001 et 2002) sur les langues romanes ont montré que quelques secondes de parole sont suffisantes pour que des sujets humains adultes parlant des langues maternelles différentes (français, roumain, japonais et anglais), soient capables d'identifier cinq des langues appartenant à cette famille linguistique (français, italien, espagnol, portugais et roumain), en faisant appel à des stratégies perceptuelles à la fois non linguistiques (le rapport langue maternelle/langues de test, la familiarité antérieure avec les langues) et linguistiques (le rôle potentiel de l'information vocalique). Plus largement, les auditeurs sont capables de fournir des regroupements perceptuels basés sur des critères linguistiques tels que l'appartenance des langues à une famille linguistique ou bien les particularités structurelles communes dues au voisinage géographique (Maddieson & Vasilescu, 2002). Ces observations ont été faites au travers d'un paradigme expérimental proposant à des sujets choisis selon le critère [+/- linguiste] un entraînement sur cinq langues cibles issues de familles linguistiques différentes, suivi d'un test d'identification de ces langues et d'un second test d'évaluation de la similarité des langues cible par rapport à 18 autres langues.

De manière générale, les expériences citées ont mis en évidence des indices discriminants gravitant autour des niveaux linguistiques segmental et suprasegmental. Dans les deux paragraphes suivants nous nous pencherons sur les travaux visant notamment à mettre en évidence les facteurs discriminants relevant des deux niveaux linguistiques.

3. Les expériences d'identification des langues par des traits segmentaux

Nous présentons dans cette partie les très rares travaux qui, à notre connaissance, ont étudié exclusivement le rôle des indices de nature segmentale dans la reconnaissance et/ou la discrimination des langues. La première étude n'est pas à proprement parler une étude en identification des langues, car elle s'intéresse plus spécifiquement à la perception des segments dans une perspective monolingue de manière à mettre en évidence une hiérarchie potentielle des informations segmentales. Cole *et al.* (1996) ont analysé le rôle des segments vocaliques et consonantiques dans la reconnaissance des mots en utilisant plusieurs types de parole : (i) tout d'abord des phrases en parole naturelle, (ii) puis des phrases où toutes les voyelles ont été remplacées par un bruit, et enfin (iii) des phrases où toutes les consonnes ont été converties en un bruit. L'objectif de la démarche est de tester la hiérarchie des indices segmentaux dans la reconnaissance. Les résultats montrent que lorsque les voyelles sont préservées les sujets peuvent reconnaître 56,5 % des mots et 21,5 % des phrases. En revanche, quand seule l'information consonantique est préservée les scores de réussite chutent pour se situer autour de 14,4 % pour les mots et 8 % pour les phrases. Les résultats obtenus permettent ainsi aux auteurs de poser l'hypothèse selon laquelle la reconnaissance des mots dépend plus des voyelles que des consonnes.

Les travaux de Barkat touchant à l'identification des dialectes arabes, (1998, 1999, 2000 et 2001) aboutissent à la conclusion similaire que, parmi les données segmentales, l'information vocalique possède un statut privilégié. L'auteur appuie ses résultats sur deux démarches complémentaires : la première est basée sur les stratégies *déclarées* par les auditeurs eux-mêmes dans une phase d'évaluation postérieure au test et la seconde est fondée sur l'analyse acoustique des indices ainsi identifiés. Les résultats révèlent que les auditeurs, locuteurs arabophones natifs de parlers arabes du Maghreb ou du Moyen Orient, s'avèrent capables de distinguer ces parlers et catégorisent les différents stimuli selon ces deux zones géographiques principales (cf. Section 6.). L'information la plus pertinente qui permet cette division dialectale est liée à l'organisation qualitative et quantitative des systèmes vocaliques, les segments consonantiques, quant à eux, donnant des informations de type socio-linguistique (parlers de citadins vs. de bédouins).

Nous pouvons donc en conclure que les sujets humains sont capables de réaliser avec succès une tâche de reconnaissance et/ou d'identification linguistique même quand certaines informations segmentales sont manquantes dans le signal. Plus précisément, dans une hiérarchie des informations de nature segmentale utilisées, le rôle des segments vocaliques semble déterminant. Une explication possible de ces résultats pourrait se trouver dans les informations temporelles et les variations d'amplitude portées par le vocalisme, ce qui

implique aussi des caractéristiques de niveau prosodique. Toutefois, bien que les résultats déjà obtenus à ce niveau suggèrent la nécessité d'une meilleure exploitation de ce type d'informations, les études portant sur le rôle des indices segmentaux restent peu fréquentes en regard de celles utilisant des indices purement suprasegmentaux.

4. Les expériences d'identification des langues par la prosodie

De nombreuses études exploitant les capacités du système perceptuel humain ont montré que la prosodie est un indice fiable pour la discrimination des langues. Ainsi, Ramus (1999) a construit une série d'expériences, dans le but premier de mettre en évidence certains aspects concernant l'acquisition de la langue par l'enfant, mais aussi dans celui d'apporter de nouvelles informations pour l'identification automatique des langues. Ses travaux montrent le rôle important du niveau suprasegmental et en particulier du rythme dans l'acquisition du langage. Les expériences menées avec des nouveau-nés mettent en évidence le fait qu'ils soient capables de discriminer entre trois langues attestant des caractéristiques rythmiques différentes à partir de stimuli sonores synthétisés ne présentant plus que les seules propriétés rythmiques de ces langues. Etant donnée que l'information suprasegmentale est très significative pour la discrimination des langues, plusieurs approches expérimentales ont été appliquées au secteur de l'IAL et ce afin de mieux circonscrire sa nature. Pour certains auteurs, il a donc été également question de comparer la façon dont les humains se servent des informations prosodiques dans la discrimination des langues avec le traitement de cette composante langagière dans les systèmes automatiques, pendant que d'autres chercheurs tentaient de modéliser les données suprasegmentales afin de les rendre exploitables dans le cadre de l'IAL.

L'un des premiers travaux réalisés dans cette optique est celui de Ohala *et al.* (1979). Les auteurs ont tenté de vérifier comment les sujets humains parviennent à discriminer trois langues (anglais américain, japonais et cantonais), alors qu'ils n'ont accès qu'aux informations concernant les variations de la fréquence fondamentale. Les résultats montrent que la moyenne générale de la réussite se situe à 56,4 % et varie en fonction des paramètres suivants : lorsque les sujets bénéficient d'un entraînement préalable les scores d'identification correcte atteignent 55,1 % d'identification correcte. En revanche, les sujets qui ne bénéficient pas d'une phase d'entraînement obtiennent uniquement 44,1 % de réponses correctes. Enfin, les sujets monolingues réussissent la tâche à 50,3 %, les bilingues à 59,3 % et les trilingues à 53,2 %. Quant au rôle de la durée des stimuli, les stimuli courts (<10 secondes) sont reconnus à 55,1 %, tandis que les stimuli longs (15-25 secondes) sont reconnus à 57,7 %. Les scores obtenus dépendent donc (i) de la durée des stimuli, (ii) de l'apprentissage, (iii) du caractère mono/bilingue des auditeurs et (iv) de la présence de la langue maternelle parmi les langues de test. Auparavant, Atkinson (1968) et Bonte (1975) avaient tenté des expériences perceptives s'appuyant sur des stimuli artificiels où des données prosodiques avaient été préservées, sans que pour autant les résultats soient extrêmement révélateurs. Par la suite, Maidment (1983), se proposait de mettre en valeur plusieurs facteurs, tels que le rôle de l'information prosodique dans l'identification en parole naturelle, le rôle de l'entraînement et les critères d'identification évoqués par les sujets après le test, et, enfin, le rôle des acquis antérieurs, donc du caractère [+/- naïf] de ces derniers. Pour cela l'auteur a employé des stimuli naturels et enregistrés avec un laryngographe en français et en anglais. Les stimuli ont été par la suite présentés à deux groupes de sujets, francophones et « naïfs », qui ont également eu la possibilité de se servir d'une échelle de valeur allant de 1 à 4, pour noter le degré de similarité des différents stimuli entendus. Les 74,7 % de réponses correctes révèlent qu'il est aisé de distinguer entre français et anglais en s'appuyant uniquement sur les caractéristiques prosodiques de ces deux langues.

Ramus *et al.* (1999) ont réalisé une étude qui examine les capacités de discrimination des langues chez les nouveau-nés provenant d'un environnement multilingue. L'anglais et le japonais sont reconnus et discriminés à travers des phrases où les informations préservées concernent tour à tour : (i) le niveau phonotactique ; (ii) le rythme et l'intonation ; (iii) seulement l'intonation ; (iv) seulement le rythme. Il s'est avéré que le rythme est suffisant pour permettre la discrimination des phrases en japonais de celles en anglais. Ainsi, cette étude a confirmé les résultats des travaux antérieurs de Ramus (1999) consistant à simuler la classification des langues par rapport à leur appartenance à un type accentuel au travers d'une analyse discriminante. L'analyse discriminante a été effectuée sur des mesures concernant la segmentation des voyelles et des consonnes présentes dans 8 langues. Ces mesures suggèrent que les types rythmiques reflètent des propriétés phonologiques spécifiques qui, par ailleurs, sont signalées par les propriétés phonéto-acoustiques de la parole. Plusieurs classes de langues correspondant aux types rythmiques sont ainsi obtenues. Par la suite, les expériences perceptuelles menées auprès de nouveau-nés en utilisant le paradigme de la succion non nutritive, ont confirmé la pertinence perceptuelle de ces classes. Ainsi, il semble que les bébés réagissent de façon plus

sensible lorsqu'ils écoutent deux stimuli issus de deux classes rythmiques différentes, que lorsqu'ils écoutent deux stimuli qui appartiennent à la même classe rythmique. Ces résultats sont étayés par des études similaires faisant appel à des sujets adultes et utilisant un paradigme expérimental basé sur la recherche d'intrus.

À la suite d'un test utilisant des stimuli créés avec une méthode similaire à partir d'enregistrements en anglais, en français, en allemand, en japonais et en mandarin, Navrátil (1998) vérifie l'importance de certains types d'informations phonotactiques et prosodiques présentes dans le signal. Il utilise : (i) des échantillons en parole naturelle ; (ii) des échantillons où l'ordre des syllabes a été inversé et (iii) des échantillons où seule l'information concernant la fréquence fondamentale et l'amplitude est présente. Les résultats sont obtenus avec 84 auditeurs adultes appartenant à douze populations différentes. Les résultats révèlent que les informations phonotactique et acoustique sont plus importantes que l'information prosodique et que l'information lexicale est également à prendre en compte. Néanmoins, l'information prosodique semble jouer un rôle plus important pour l'identification de certains types de langues, tels que le mandarin – langue tonale – comparé à celui des autres langues.

D'autres travaux prouvent qu'une information suprasegmentale beaucoup moins saillante peut fournir suffisamment d'indices pour permettre une reconnaissance ou une discrimination linguistiques. Ainsi, Mori *et al.* (1999) ont étudié le rôle de l'information prosodique dans l'identification des langues en utilisant les résiduels d'une analyse LPC (« Linear Predictive Coefficients ») en tant que stimuli auditifs. Ces stimuli comportent donc l'information acoustique liée à la « source » et écartent l'information liée au « filtre » (« *coefficients LPC* ») et donc à l'information segmentale. Les auteurs argumentent ce choix en considérant que les résiduels correspondent à l'information suprasegmentale, en plus de l'information sur le locuteur. Les auditeurs ont réussi à identifier des mots grâce aux résiduels, ce qui a autorisé Mori et ses collaborateurs à affirmer que les indices présents dans les résiduels (i.e., la présence ou l'absence des harmoniques, les variations temporelles d'intensité, entre autres) fournissent suffisamment d'informations pour permettre de différencier entre les catégories majeures de segments, i.e., voyelles et consonnes, ainsi qu'entre les modes d'articulation.

Enfin, d'autres études portent sur la discrimination des *dialectes* grâce à des informations de nature prosodique. Foreman (1999) réalise une expérience en parole naturelle qui porte sur la discrimination dialectale entre l'anglais « standard » (*Mainstream American English*) et l'anglais des afro-américains (*African-American English*). L'auteur parvient à une conclusion élémentaire : l'identification dialectale est plus facile dans le cas où le motif intonatif ne varie pas à l'intérieur de la communauté des locuteurs d'un même dialecte. De plus, il s'est avéré que l'exposition antérieure des sujets aux dialectes analysés joue un rôle considérable dans l'obtention de bons résultats en discrimination. Barkat *et al.* (1999) évaluent le pouvoir discriminant des indices prosodiques et rythmiques dans la différenciation des parlers arabes (et plus précisément les variations de la courbe de la fréquence fondamentale et l'amplitude de certains événements contenus dans le signal). Ces informations sont suffisantes, à elles seules, pour permettre la différenciation des dialectes arabes maghrébins des dialectes arabes moyen orientaux de manière significativement supérieure à la chance, lors de tests réalisés avec des locuteurs naïfs arabophones. Une étude acoustique ultérieure a montré que les variations rythmiques dans les dialectes arabes de l'Ouest *vs.* de l'Est trouvent leur origine dans certaines des propriétés phonologiques des langues en question (i.e., complexité consonantique et proportion de segments vocaliques) – lesquelles influencent effectivement le rythme – permettent aux sujets de discriminer aisément entre ces deux variétés dialectales (Ghazali *et al.*, 2002).

Ces diverses études illustrent le rôle important de l'information suprasegmentale dans l'identification des langues. Ainsi, le niveau suprasegmental semble être acquis très précocement lors de l'apprentissage de la langue maternelle et, plus tard, susceptible de fournir aux sujets adultes des indices discriminants très pertinents permettant l'identification avec succès de langues diverses. En effet, les sujets humains sont capables de discriminer des langues inconnues même si l'information segmentale est absente. Parmi les indices suprasegmentaux, notons le caractère fortement discriminant du rythme et des variations de la fréquence fondamentale.

5. Discussion

La principale conclusion qui s'impose à la fin de ce bilan sur les approches en identification perceptuelle des langues est que, effectivement, l'être humain dispose d'un système perceptuel très robuste face aux tâches d'identification linguistique. En effet, les travaux que nous venons de citer mettent en évidence qu'il a besoin d'un maximum de quelques secondes pour discriminer des langues complètement inconnues, avec des taux de

réussite importants. De plus, il est capable, au cours d'une phase post-test, de fournir des informations sur les critères qui lui ont facilité la discrimination. Cette capacité discriminatoire reste efficace, même quand des informations de nature segmentale et/ou suprasegmentale sont manquantes dans le signal ou même quand la qualité du signal est altérée. L'oreille humaine développe dans ces conditions défavorables des stratégies de compensation en faisant appel à d'autres traits discriminants, et en procédant à une intégration fructueuse des informations préservées dans le signal.

Les études que nous avons passées en revue utilisent généralement des stimuli en parole naturelle d'une durée maximale de 6 secondes. Elles comparent un nombre variable de langues, allant jusqu'à dix. Les populations participant au test sont en règle générale anglophones. Quant à l'étude du rôle des niveaux linguistiques segmental et suprasegmental dans l'identification, nous avons pu noter que la plupart des études s'intéressent préférentiellement au rôle de l'information prosodique, au travers de paradigmes expérimentaux employant des stimuli modifiés. Concernant les stratégies mises en œuvre par les auditeurs afin de traiter l'information prosodique, il semble qu'il existe des principes généraux chez tous les êtres humains, malgré les facteurs différenciateurs tels que la langue maternelle et l'appartenance de cette dernière à un autre type rythmique que les autres langues testées.

Enfin, peu de travaux sont dédiés à l'identification des dialectes ou des langues proches du point de vue typologique. Souvent, les corpus sont constitués de langues appartenant à des familles linguistiques diverses dont les particularités structurelles sont peu contrôlées. Notre approche est différente, en cela qu'elle s'intéresse à la discrimination perceptuelle et automatique de différents dialectes arabes. Dans les paragraphes suivants nous montrons que les auditeurs arabophones sont capables *d'identifier perceptuellement* le dialecte d'un locuteur arabophone donné, même dans les conditions où le degré d'intercompréhension existant entre les deux formes dialectales est très réduit, voire absent. A cette fin, les auditeurs font appel à des *critères linguistiques* de nature segmentale qui, de plus, s'avèrent pertinents pour la *discrimination automatique* de ces variétés linguistiques.

6. Géographie Dialectale du Monde Arabophone

Chaque parler arabe présente des caractéristiques qui lui sont propres. Celles-ci peuvent être d'ordre syntaxique, lexical, segmental et/ou prosodique. Certains parlers ont innové à l'intérieur de chacun de ces domaines et ne présentent aujourd'hui que très peu de caractéristiques communes. En effet, bien que les parlers contigus sur le plan géographique attestent de nombreux traits communs et que l'ensemble des parlers arabes présente un air général de ressemblance sensible aux sujets parlants, l'intercompréhension devient toute relative — voire inexistante — lorsque l'on prend en considération des points situés aux antipodes du domaine. Une vue d'ensemble du domaine linguistique arabe (figure 1) permet de distinguer deux zones principales à l'intérieur du continuum dialectal : à l'Ouest, la zone occidentale regroupant les parlers du Maghreb ; à l'Est, la zone orientale constituée des parlers Moyen-Orientaux (Cohen, 1970 ; Fleish, 1975). Il convient, par ailleurs, de souligner la présence d'une zone intermédiaire aux frontières linguistiques relativement vagues dont les parlers — définis comme transitoires — se caractérisent par des caractéristiques mixtes. Toutefois, les parlers arabes constituent de manière générale des formes linguistiques aisément identifiables par le système perceptuel des locuteurs/auditeurs arabophones.

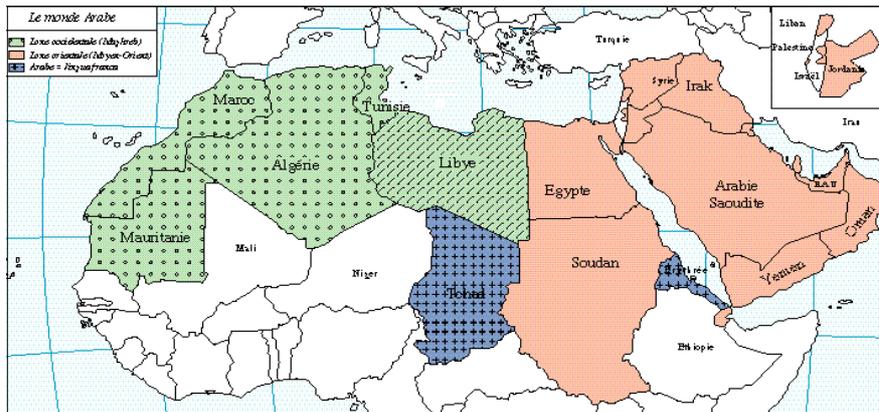


Figure 1. Géographie dialectale du Monde Arabophone en gris clair avec des points, zone OUEST regroupant les parlers du Maghreb ; en gris plein zone EST regroupant les parlers du Moyen-Orient ; en hachuré, zone intermédiaire et en gris foncé avec des croix les zones où l'arabe est une langue véhiculaire, i.e., lingua franca.

6.1. Identification dialectale perceptuelle et détermination expérimentale d'indices discriminants

6.1.1. Corpus et méthode

Les travaux relevant de l'identification dialectale arabe sont peu nombreux. En effet, seules deux études ont abordé le problème de la variabilité interdialectale arabe au niveau phonétique (Abu-Haidar, 1991 ; Rjaibi-Sabhi, 1993). L'idée maîtresse sous-tendant ces travaux est que le système phonologique du dialecte maternel des sujets arabophones transparaît au niveau de leurs productions en arabe standard sous la forme de transferts linguistiques. Ces derniers fournissent une indication pertinente quant à l'origine dialectale du sujet parlant. Notre travail vient compléter les travaux précédemment cités et est basé sur la perception de ces marques linguistiques. Nous partons de l'hypothèse que de la même manière qu'un individu est capable de reconnaître une voix familière à partir d'un échantillon de parole très bref, les individus possédant une connaissance antérieure de langues et/ou dialectes devraient être capables de les identifier dans un délai court, sur la base de certaines propriétés linguistiques discriminantes. D'un point de vue méthodologique, nous avons choisi de travailler directement sur les formes vernaculaires de l'arabe afin de contrôler toute normalisation perceptive vers la variété standard liée à un effet de filtre qui pourrait conduire à la non-perception des traits dialectaux les plus typiques. L'hypothèse que nous avançons ici est que ces caractéristiques linguistiques — si elles s'avèrent perceptibles — pourraient constituer un faisceau d'indices de discrimination pertinents dans le cadre d'une tâche d'identification dialectale. Afin de repérer les indices utilisés par les auditeurs arabophones pour identifier l'origine dialectale de locuteurs s'exprimant en arabe, nous avons mis au point une expérience perceptive d'identification dialectale (Barkat, 1998 et Barkat, 1999). Dans ce cadre, six variétés dialectales représentatives des deux principales zones géographiques du Monde Arabe (i.e., Maghreb vs. Moyen-Orient) ont été sélectionnées et une base de données acoustiques a été enregistrée avec 12 locuteurs natifs (6 maghrébins, d'origine marocaine, algérienne et tunisienne et 6 orientaux, d'origine syrienne, libanaise et jordanienne) arabophones s'exprimant de manière spontanée dans leur dialecte maternel sur la base d'un livre d'images sans texte¹.

Les enregistrements — effectués en chambre insonorisée sur magnétophone DAT — ont été digitalisés à 22 kHz, 16 bits, mono sous Sound Forge©. Quarante-huit échantillons de parole d'une durée variable allant de 7 à 30 secondes² ont été extraits du corpus. Ils ont ensuite été réorganisés en ordre aléatoire et présentés comme matériel de stimulation lors d'une expérience d'identification, chaque stimulus étant précédé d'un numéro de passage³. Les 96 échantillons de parole ont été présentés à 18 sujets arabophones « naïfs »

¹ Mayer, M., (1980), *Frog, where are you ?*, Dial Books for Young Readers, New-York.

² Des recherches traitant de l'identification de l'identité sociale véhiculée par les usages dialectaux ont montré que 10 à 15 secondes sont suffisantes pour reconnaître l'origine sociale et/ou dialectale (Lee, 1971).

³ Dans cette première expérience, nous sommes partis du postulat que les locuteurs arabophones disposent tous d'une connaissance relative des variations de production en fonction de l'origine géographique. De plus, cette connaissance est favorisée par le contact entre ces populations dans la communauté étudiante qui nous a servi de « vivier à sujets ». Etant

(i.e., étudiant(e)s non spécialistes de linguistique), soient : 9 sujets maghrébins et 9 sujets moyen-orientaux originaires des six mêmes pays que ceux représentés dans les stimuli mais n'ayant jamais eu de contact avec les personnes dont les voix ont été utilisées pour le matériel expérimental, et ce, afin d'éviter toute identification dialectale par reconnaissance du locuteur. Les sujets disposaient de quinze secondes pour exécuter les trois tâches suivantes : (i) identifier l'origine dialectale du locuteur entendu en fonction de la zone géographique (i.e., Maghreb *vs.* Moyen-Orient) ; (ii) identifier le pays dont pouvait être originaire le locuteur parmi les 6 choix proposés sur la grille de réponses (table 1) ; (iii) définir — dans la mesure du possible — les indices prosodiques, segmentaux et/ou lexicaux ayant permis l'identification. Les réponses apportées aux trois tâches précédentes devant être reportées sur une grille préformatée distribuée aux sujets en début d'expérience. Outre la détermination expérimentale d'indices discriminants *a priori* pertinents pour l'identification automatique des parlers arabes, nous avons voulu vérifier auprès de nos sujets les hypothèses suivantes : (i) l'identification par zone géographique principale est aisée pour l'ensemble des sujets. Elle constitue de ce fait une réalité linguistique perceptible par des sujets arabophones même non entraînés⁴ ; (ii) les meilleurs scores d'identification sont obtenus pour la distinction du dialecte maternel par rapport aux autres parlers ; (iii) Les résultats les moins probants concernent les parlers proches à l'intérieur d'une même zone, ce qui impliquerait — pour ces parlers spécifiquement — une analyse plus fine des traits distinctifs, couvrant plusieurs niveaux de la langue, voire à plus long terme dans le cadre de l'IAL, l'intégration d'un modèle d'apprentissage particulier (basé, par exemple, sur la détection d'items lexicaux spécifiques).

6.1.2. Identification dialectale par zones géographiques principales

Les résultats de l'identification dialectale par zone révèlent que 97 % des stimuli maghrébins et 99 % des stimuli moyen-orientaux ont été identifiés correctement. Les erreurs de classification, soit respectivement 3 % et 1 %, résultant — pour le cas des stimuli maghrébins — du fait de l'emploi, pourtant impropre, par l'une des locutrices originaire de la zone occidentale (i.e., d'origine algérienne), d'items morpho-syntaxiques appartenant à l'arabe classique. Les parlers moyen-orientaux présentant plus de caractéristiques communes avec l'arabe classique, la perception de certains indices (lexicaux, morphologiques et/ou phonético-phonologique) a conduit certains sujets maghrébins à juger ces productions comme plutôt orientales (Barkat, 2001). Néanmoins, malgré la présence de ces quelques erreurs induites tour à tour par (i) la présence d'emprunts à l'arabe classique et (ii) le caractère bref de certains stimuli (i.e., moins de 2 secondes), les scores d'identification par zone restent très largement supérieurs au hasard. Nous pouvons donc affirmer que la bipartition dialectale du domaine linguistique arabophone en termes de zones géographiques principales (parlers maghrébins *vs.* parlers orientaux) transparait à travers des éléments perceptifs identifiables par l'ensemble des locuteurs arabophones originaires de l'une ou l'autre de ces deux régions. Ainsi, cette bipartition constitue une réalité linguistique bien établie sur le plan perceptuel qu'il serait intéressant d'analyser plus en détail du point de vue linguistique afin de constituer une liste exhaustive des traits dialectaux (lexicaux, phonético-phonologiques, prosodiques et/ou rythmiques, etc.), propres à chacune de ces deux variétés dialectales.

6.1.3. Identification dialectale par pays

Les résultats obtenus par les sujets pour la tâche d'identification par pays confirment, dans la grande majorité des cas, les hypothèses avancées ci-dessus selon lesquelles les meilleurs taux d'identification devraient concerner les dialectes proches des parlers natifs et les parlers appartenant à la même zone géographique que ceux-ci. Il semble donc que les sujets perçoivent — à l'intérieur des stimuli les plus représentatifs d'un parler — certains indices acoustiques propres à chaque variété dialectale qui leur permettent d'effectuer une discrimination fine en termes de « pays ». La table 1 présente les scores d'identification par pays obtenus par l'ensemble des sujets testés sur la totalité des stimuli présentés. Il nous permet d'ores et déjà

donnée cette exposition antérieure non négligeable aux dialectes arabes, nous n'avons pas retenu de période d'apprentissage dans ce protocole. En revanche, nous avons accordé du temps à l'explication de la troisième tâche (i.e., définition des indices permettant la discrimination). Celle-ci a été présentée en termes « communs » (i.e., « sons » typiques pour « indices phonétiques », « mélodie, musique de la langue » pour « indices prosodiques », etc.). Par ailleurs, nous avons établi sur la base d'exemples *ad hoc* un code de transcription exploitable pour l'analyse. Pour ce qui est de la mise en valeur de la prononciation d'un segment spécifique ayant contribué à l'identification dialectale, il a été convenu que les sujets (sachant tous écrire en arabe) devaient transcrire l'item en arabe classique et cercler le phonème (i.e., « la lettre ») caractérisée par une prononciation particulière.

⁴ Les études citées ci-dessus ont parfaitement établi le rôle de la période d'entraînement pour l'amélioration des performances d'identification linguistiques et/ou dialectales. Nous postulons ici que les sujets arabophones possèdent tous une connaissance relative des différentes variétés linguistiques arabes du fait de leur co-habitation sur le territoire français et des émissions radiophoniques diffusées aujourd'hui dans la plupart des dialectes arabes.

d’observer que tous les locuteurs/auditeurs arabophones — à l’exception des sujets syriens (SY) — attestent le taux de reconnaissance le plus élevé pour, au moins, leur dialecte d’origine.

	Sujet MA	Sujet AL	Sujet TU	Sujet SY	Sujet LI	Sujet JO	Tx moyen
Stim MA	94	63	67	33	54	75	64
Stim AL	88	92	48	69	48	65	68
Stim TU	73	85	96	56	83	46	73
Stim SY	31	35	19	90	75	88	56
Stim LI	27	33	38	96	94	100	65
Stim JO	23	46	35	83	85	100	62

Table 1. Matrice de confusion des scores d’identification par pays en fonction de l’origine dialectale des sujets (en %).

6.1.3.1. Identification perceptuelle « intra-zone »

Bien que les sujets syriens aient été plus performants pour la reconnaissance du libanais (i.e., dialecte « *levantin* » linguistiquement et géographiquement proche, mais non-maternel), le score moyen d’identification du parler d’origine (tous sujets, toutes variétés et toutes zones géographiques confondus) avoisine les 94 %, et ne présente pas d’écart significatif d’une origine dialectale à l’autre, comme l’a confirmée l’analyse de variance effectuée sur la base des résultats obtenus pour chacun des sujets lors de l’identification de son dialecte d’origine. A ce niveau de l’analyse, les résultats nous autorisent à nuancer l’approche — souvent tranchée dans les travaux de dialectologie traditionnelle — selon laquelle la définition d’entités linguistiques appelées « *arabe maghrébin* » vs. « *arabe oriental* » ne constitue pas une réalité aussi nettement contrastée. Il semble en effet que les sujets se soient appuyés sur des macro-critères discriminants efficaces pour d’une part, distinguer des parlers maghrébins par rapport à leurs pendants moyen-orientaux et, d’autre part, arriver à une classification plus fine en termes de parler national.

6.1.3.2. Identification perceptuelle « inter-zones »

Les scores associés à la distinction des parlers inter-zones selon l’origine des sujets révèlent une asymétrie importante. Les taux d’identification correcte obtenus par les sujets maghrébins pour la discrimination des parlers moyen-orientaux est de 32 %. Les résultats obtenus pour l’identification des parlers maghrébins plus ou moins proches de leur dialecte d’origine sont de 78 %, la différence observée entre ces deux scores étant significative au plan statistique [$t_{(8)} = 149, p \leq .0001$]. De la même manière, les taux d’identification des sujets orientaux attestent des valeurs nettement différenciées selon qu’il s’agisse de stimuli maghrébin (59 %) ou moyen-orientaux (90 %) [$t_{(8)} = 224, p \leq .0001$]. Par ailleurs, on observe que pour la même tâche (i.e., discrimination par pays des stimuli *n’appartenant pas* à la zone dialectale d’origine), les deux populations font preuve d’un comportement différencié : alors que les sujets moyen-orientaux parviennent à identifier correctement à 59 % les stimuli maghrébins, les sujets maghrébins présentent quant à eux un taux de discrimination des stimuli orientaux de l’ordre de 32 % seulement. Cette dissymétrie dans les proportions de reconnaissance des parlers inter-zones se révèle statistiquement significative [$t_{(8)} = 4, p = .0023$]. Notons enfin que les scores obtenus par les deux populations pour l’identification des 48 stimuli maghrébins — soient 78 % pour les sujets maghrébins et 59 % pour les sujets orientaux — présentent une différence plus significative encore [$t_{(8)} = 184, p \leq .0001$]. On retrouve la même opposition pour la discrimination des stimuli moyen-orientaux, puisque les sujets maghrébins présentent un taux d’identification correcte de 32 % largement inférieur à celui obtenu par les sujets originaires de l’Est (90 %). Cette différence s’avère elle aussi largement significative [$t_{(8)} = 170, p = \leq .0001$].

Ces résultats corroborent deux des trois hypothèses de départ, à savoir : (i) que l’on reconnaît significativement mieux les parlers proches du dialecte maternel, (ii) que la discrimination des parlers appartenant à la même zone géographique pose — en règle générale — moins de problèmes que l’identification des parlers représentatifs de la zone dialectale opposée. De plus, ils montrent, que certains sujets — en l’occurrence les moyen-orientaux — semblent plus performants que les maghrébins pour les tâches d’identification dialectales en termes de pays. Cette différence pouvant être attribuée au fait que la population maghrébine est — en France — supérieure à celle des moyen-orientaux, et que ces derniers étant plus exposés aux différentes variétés dialectales arabes occidentales sont donc plus sensibles aux particularismes dialectaux des parlers maghrébins.

6.1.4. Détermination expérimentale d'indices discriminants pour les parlers arabes

Parmi l'ensemble des critères discriminants dégagés à l'issue de l'expérience perceptuelle et identifiés par les sujets eux-mêmes, la distribution des voyelles dans l'espace acoustique ainsi que la réalisation de l'opposition de durée vocalique (i.e., rapports voyelles longues / voyelles brèves, noté RVL/VB) peuvent être considérés comme pertinentes en vue de la reconnaissance automatique des parlers arabes par zones géographiques principales. En effet, les sujets ont majoritairement constaté qu'il existait bien des différences inter-dialectales typiques de l'une et/ou l'autre des deux zones dialectales, au niveau des réalisations vocaliques, tant au niveau qualitatif qu'au niveau quantitatif. Notre objectif a donc été de caractériser au plan acoustique les différences pressenties au niveau perceptuel.

7. Organisation spectrale et temporelle des systèmes vocaliques : des indices de discrimination dialectale pertinents

7.1. Distribution des segments vocaliques en arabe maghrébin vs. oriental

Le système vocalique de l'arabe standard est généralement décrit comme étant composé de trois phonèmes brefs [i u a], auxquels s'opposent trois voyelles longues de mêmes timbres [i: u: a:]. Toutefois, des constatations concernant l'introduction secondaire de certaines nouvelles voyelles définies comme centrales et/ou d'aperture moyenne ont été faites pour la majorité des dialectes (Ghazali, 1979 ; Barkat, 2000).

7.1.1. Analyse acoustique

Nous avons procédé à l'analyse acoustique de la dispersion vocalique dans les différents parlers présentés ci-dessus. Pour chacune des deux zones dialectales concernées, nous nous intéressons plus particulièrement à la caractérisation qualitative et quantitative des segments vocaliques tels qu'ils sont effectivement réalisés en parole spontanée (i.e., non-lue). Ceci nous permet de caractériser au plan acoustique la distribution des segments vocaliques dans les parlers maghrébins et orientaux et la réalisation de l'opposition de durée, c'est-à-dire le rapport voyelle longue / voyelle brève dans chacun des parlers étudiés. Les analyses rapportées ici ont été effectuées à partir d'enregistrements effectués en chambre sourde sur la base d'un corpus semi-guidé consistant en la traduction spontanée, en arabe dialectal, d'un texte court (i.e., « *La Bise et le Soleil* »). Dans un second temps, nous avons procédé à l'étiquetage manuel par timbre des 1500 segments vocaliques monophongues présents dans les échantillons de parole. Une analyse LPC a ensuite été effectuée au centre de chacune de ces voyelles, de manière à obtenir les valeurs formantiques associées à leur état stable. Ainsi, bien que conscients des nombreux paramètres influençant la réalisation des voyelles (nombre de syllabes dans le mot, nature de ces syllabes, position de l'accent, etc.) nous n'avons pas explicitement tenu compte de ces facteurs de variation co-intrinsèques afin d'établir une représentation globale de la distribution vocalique propre à chacun des parlers et de dégager des informations qu'un modèle de reconnaissance automatique puisse prendre en compte.

7.1.2. Résultats acoustiques

7.1.2.1. Aspect qualitatif

Les analyses acoustiques menées à partir de parole spontanée en arabe maghrébin et oriental nous ont permis de caractériser au niveau phonétique la distribution des voyelles dans l'espace acoustique et d'établir une distinction entre parlers occidentaux privilégiant la génération de voyelles intérieures (i.e., centrales) résultant d'un processus de réduction vocalique et parlers orientaux préférant les positions périphériques. Nous avons vu que les voyelles brèves se distribuent, à l'Ouest vs. à l'Est du domaine, suivant deux schémas de dispersion vocalique distincts : les parlers maghrébins présentent, en effet, une distribution vocalique plus condensée et nettement plus centralisée que celle des parlers orientaux (figure 2), ce qui vient préciser la remarque de Marçais (1977:24) selon laquelle « *les parlers du Maghreb se caractérisent par la ruine considérable de leur matériel vocalique* ».

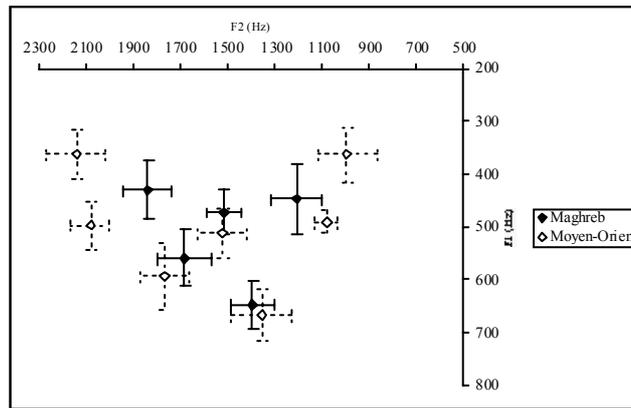


Figure 2. Représentation globale de la dispersion acoustique des voyelles brèves en arabe maghrébin vs. moyen-oriental.

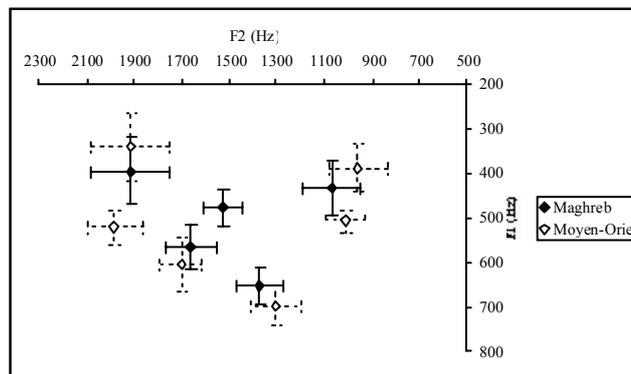


Figure 3. Représentation globale de la dispersion acoustique des voyelles longues en arabe maghrébin vs. moyen-oriental.

Ces deux schémas de dispersion vocale différenciés apparaissent également au niveau du vocalisme long : les voyelles longues du Maghreb apparaissent comme plus centrales en comparaison de leurs correspondantes orientales en arabe oriental (figure 3).

Ainsi, bien que les systèmes phonologiques varient substantiellement d'un dialecte à l'autre à l'intérieur d'une même zone géographique (Barkat, 2000), l'observation comparée de la distribution des segments vocaliques brefs tels qu'ils sont effectivement réalisés en parole spontanée montre que les parlers maghrébins — contrairement aux parlers orientaux — tendent à favoriser la génération de timbres centralisés, ce qui conduit à une certaine réduction des contrastes vocaliques d'origine. Les parlers orientaux en revanche privilégient les positions périphériques et ont tendance à conserver une dispersion vocale maximale tout en en développant par ailleurs de nouvelles voyelles d'aperture moyenne de type [e] et [o] et centrale (i.e., [ə]). Afin de mieux comprendre la dispersion des voyelles brèves dans les parlers du Maghreb, il convient de rappeler qu'un certain nombre d'entre elles ont purement et simplement disparu du système : les voyelles courtes finales, par exemple, déjà instables de par leur position, ne subsistent généralement plus. Certaines autres voyelles brèves positionnées à l'intérieur de syllabes ouvertes peuvent, selon le parler, soit s'abrèger au point de devenir de simples points vocaliques de timbre plus ou moins neutre selon le contexte articulatoire et d'une durée généralement inférieure à 25 ms, soit disparaître complètement du fait d'une règle de syncope touchant — à de rares exceptions près — toutes les voyelles brèves ainsi positionnées. Enfin, il semblerait que les voyelles brèves qui sont conservées — tout en entretenant les oppositions qualitatives de base (i.e., [ɪ ʊ]) — soient articulées à l'intérieur d'un espace plus réduit que celui des parlers orientaux. Outre la caractérisation, sur le plan acoustique, de la distribution des segments vocalique en arabe maghrébin vs. oriental, cette étude nous autorise à considérer la distribution des voyelles brèves dans l'espace acoustique comme un critère potentiellement pertinent pour l'identification automatique des parlers arabes. Par ailleurs, notre travail confirme l'idée déjà avancée par Ghazali (1979) que proposer un système vocalique identique à celui généralement postulé pour l'arabe standard n'est pas compatible avec les faits phonétiques observables en arabe dialectal. Ceci constitue un argument supplémentaire quant au développement de modèles

d'identification basés sur un décodage acoustico-phonétique à partir de parole spontanée tenant compte des variations dialectales et individuelles.

7.1.2.2. Aspect quantitatif

Les dialectes arabes opposent des voyelles longues aux voyelles brèves. Dans les dialectes où cette opposition est clairement attestée, l'opposition de durée est inscrite dans le système phonologique sous-jacent de la langue et permet de distinguer des significations par la seule variation de la quantité vocalique. Les travaux traitant de la durée vocalique en arabe dialectal sont rares et jusqu'alors assez dispersés (Mitleb, 1984 ; Jomaa, 1994 ; Ghazali *et al.*, 1992 ; Angoujard, 1979). Toutefois, bien que ce type de connaissance soit indispensable pour l'élaboration des règles phonétiques significatives liées au phénomène de temporalité, l'étude de l'opposition de la durée vocalique telle qu'elle est réalisée en parole naturelle (et non à l'intérieur de mots isolés comme dans la plupart des études précédemment citées) reste primordiale. En effet, les connaissances ainsi acquises permettraient d'améliorer les performances des modèles de reconnaissance et des systèmes de synthèse de la parole dont l'objectif est, d'une part, de pouvoir fonctionner à partir de parole spontanée dans le cas des systèmes d'identification et d'autre part, d'être le plus proche possible des phénomènes observés en parole naturelle, pour ce qui concerne la synthèse de la parole. Nous avons ainsi retenu les mesures acoustiques des segments vocaliques brèves et longs des six mêmes variétés d'arabe dialectal qu'en 6.1.1. avant d'établir — sur la base de nos données — le rapport moyen VL/VB tel qu'il est effectivement réalisé en parole spontanée, i.e., non-lue. La table 2 présente les valeurs moyennes de durée vocalique pour chacun des parlers étudiés et permet de remarquer — qu'en moyenne — les voyelles brèves et/ou longues présentent d'une zone dialectale à l'autre des différences de durée remarquables. Afin de caractériser au niveau statistique ces écarts, nous avons reporté dans la table 3 les rapports V_L/V_B calculés pour chaque parler, sur la base desquels un T-test a été effectué.

	Durée moyenne des V_B (en ms)	Durée moyenne des V_L (en ms)
Maroc	58	106
Algérie	52	103
Tunisie	51	102
Moyenne Maghreb	54	104
Syrie	61	140
Liban	58	145
Jordanie	54	131
Moyenne Moyen-Orient	58	139

Table 2. Moyenne des durées vocaliques tous timbres confondus par pays et par zone (en ms).

Rapports V_L/V_B moyens par pays (tous timbres et tous locuteurs confondus)			
Maghreb		Moyen-Orient	
Maroc	1.8	Syrie	2.3
Algérie	2.0	Liban	2.6
Tunisie	2.0	Jordanie	2.1
Moyenne des Rapports V_L/V_B par zone			
1.9		2.3	

Table 3. Rapports $R V_L/V_B$ moyens en arabe maghrébin et moyen-oriental (en ms).

L'analyse statistique révèle que les rapports mis en œuvre dans l'une et l'autre des deux zones dialectales sont significativement différents $T(3 ; 2,35) = 2.50$ ($p = .04$), les parlers orientaux attestant des rapports significativement plus élevés que les parlers du Maghreb. Ce résultat signifiant que dans les parlers du Moyen-Orient, l'opposition de durée vocalique s'établit de manière plus contrastée qu'au Maghreb, où l'on constate néanmoins que — même au Maroc — cette opposition subsiste dans nos données.

8. Application à l'identification automatique des parlers arabes

La dernière étape de ce travail a consisté à évaluer dans une tâche d'identification automatique des zones dialectales arabes l'efficacité des paramètres (distribution et quantité vocalique) dont l'expérience perceptuelle et les analyses acoustiques ont révélé la pertinence. Le modèle utilisé est basé sur la modélisation acoustique des systèmes vocaliques, qui est un sous-système de la modélisation différenciée (Pellegrino *et al.* 2000). Il permet d'obtenir par une approche non supervisée des modèles des systèmes vocaliques à partir de données non-étiquetées. Cette pratique présente, entre autres, l'avantage de ne pas être influencée par les connaissances phonologiques, qui dans le cadre de l'arabe, apparaissent souvent comme biaisées compte tenu des phénomènes d'interférence et d'hypercorrection avec la variété haute de la langue (arabe classique et/ou moderne).

8.1. Modélisation acoustique des systèmes vocaliques

En détectant dans le signal les sons possédant une structure formantique, les algorithmes développés sont en mesure de distinguer des segments vocaliques, correspondant aux parties stables des voyelles et des segments non vocaliques (consonnes et parties transitoires des voyelles). A partir des segments vocaliques du corpus d'apprentissage pour chaque zone dialectale, des modèles statistiques de type « mélange de lois gaussiennes » sont estimés dans un espace de paramètre cepstral. Au cours de la phase de test, la zone dialectale identifiée est celle pour laquelle la vraisemblance des voyelles extraites de l'échantillon de test est la plus élevée.

8.2 Expériences

8.2.1. Corpus

Les expériences d'identification automatique sont réalisées à partir d'un sous-corpus des données analysées précédemment et utilisées pour les expériences perceptuelles. Le corpus d'apprentissage est élaboré à partir de la traduction spontanée du texte « La Bise et le Soleil » par 10 locuteurs et locutrices arabophones originaires de différents points du domaine linguistique arabophone. Pour chacun de ces 10 locuteurs, 4 répétitions du texte ont été acquises (correspondant à environ 2 minutes de parole). Les données acoustiques — enregistrées auprès des locuteurs et de locutrices natifs — ont ensuite été digitalisées à 16KHz, 16 bits. La zone occidentale est représentée dans le corpus d'apprentissage à travers des stimuli en arabe algérien et marocain. La zone orientale par des stimuli en arabe égyptien, syrien, libanais et jordanien. Le modèle maghrébin (i.e., MA) est basé sur les réalisations vocaliques de cinq locuteurs différents: un locuteur algérien (originaire de Oran) et quatre locuteurs marocains (originaires de Rabat et Casablanca). La modélisation de l'espace acoustique moyen-oriental a été obtenue à partir des réalisations vocaliques détectées dans des corpus de cinq locuteurs orientaux originaires de Syrie (deux locuteurs originaires de Alep), du Liban (un locuteur originaire de Beyrouth) et de Jordanie (deux locuteurs originaires de Irbid). L'objectif de l'expérience est de confirmer ou d'infirmer la robustesse de l'indice de dispersion vocalique et de l'opposition de durée pour l'identification automatique des parlers arabes par zone géographiques principales (i.e., Maghreb vs. Moyen-Orient). Pour chacune des deux aires dialectales concernées, nous avons élaboré quatre modèles en fixant à chaque fois un nombre de classes gaussiennes différent (5, 10, 15, et enfin 20), de manière à déterminer le nombre de classes optimal pour lequel les meilleurs taux de reconnaissance seront observés, et qui donc donnera un indice de la complexité du système vocalique modélisé, même si le faible nombre de locuteurs employés limitera la portée des expériences.

Pour la phase de test, nous avons utilisé les réalisations dialectales de dix autres locuteurs originaires de différents pays du Maghreb et du Moyen-Orient. Les parlers maghrébins sont représentés par des parlers marocains et algériens (i.e., locuteurs originaires de Touggourt et Jijel pour l'Algérie et Rabat, Tétouan et Casablanca pour le Maroc). La zone orientale est représentée par des réalisations en arabe égyptien (locuteur d'Assouan), syrien (locuteur originaire de Homs), palestinien (Hébron et Haïfa) et enfin, jordanien (locuteurs de Irbid).

Les expériences d'identification automatique ont été réalisées, dans un premier temps, sur la base des quatre répétitions de chacun des dix locuteurs (40 tests). Dans un second temps, nous avons considéré les quatre répétitions de chaque locuteur comme un bloc unique à la fin duquel la décision d'identification

dialectale était prise (10 tests). Les taux obtenus à l'issue de cette seconde expérience correspondent ainsi aux scores atteints pour dix tests (i.e., 1 décision par locuteur). Lors de l'utilisation de 40 tests, deux conditions expérimentales ont été testées de manière à déterminer le pouvoir discriminant de la dispersion vocalique d'une part, et de la durée vocalique d'autre part. Dans un premier temps, le modèle d'apprentissage est élaboré à partir des caractéristiques formantiques seules (i.e., coefficients cepstraux). Dans un second temps, le modèle est appris à partir de l'utilisation conjointe des caractéristiques formantiques et de l'information de durée (i.e., coefficients cepstraux et durée des segments vocaliques). En accord avec les expériences perceptives réalisées, notre hypothèse est que l'utilisation conjointe de ces deux critères discriminants conduit à améliorer les taux de reconnaissance de manière significative.

8.2.2. Résultats

Les résultats révèlent des tendances générales intéressantes. Les modèles composés d'un faible nombre de classes gaussiennes distinguent assez mal les deux zones dialectales représentées dans les différents stimuli : la plupart des échantillons de parole sont ainsi identifiés comme étant de l'arabe maghrébin (cf. le taux d'identification élevé pour la catégorie Maghreb et le faible score correspondant à l'identification de la zone orientale). Cette tendance sous-entend qu'un modèle plus complexe (i.e., comportant un nombre de classes vocaliques supérieur) est nécessaire pour parvenir à caractériser de manière plus fine l'organisation vocalique des parlers orientaux (figure 4). Ainsi, quand la taille du modèle augmente cet effet tend à disparaître et l'on obtient, avec le nombre – ici optimal – de 20 classes vocaliques, un score d'identification par zone de 70 %. L'utilisation du test statistique de Pearson établit que ce score est significativement supérieur à la chance ($P \chi^2 > 3.84$) = .05. Le T-Test effectué sur les taux d'identification moyens obtenus pour chaque zone dialectale ne révèle quant à lui aucune différence significative, ce qui prouve que le modèle est aussi performant pour la discrimination de l'une et l'autre des deux aires dialectales à identifier. Si l'on compare ces résultats avec ceux obtenus à partir d'un apprentissage fondé conjointement sur les caractéristiques spectrales et sur la durée des segments vocaliques, on observe que – dans ce dernier cas et pour le même nombre de composantes gaussiennes – les scores d'identification correcte atteignent 78 %, (figure 5).

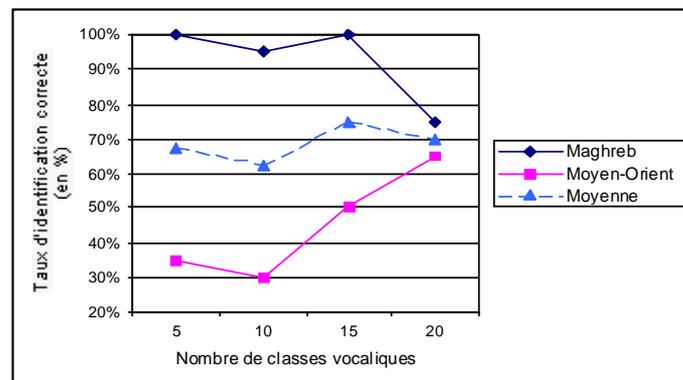


Figure 4. Taux d'identification correcte (en % et pour 40 tests) pour la discrimination par zones en fonction de la taille du modèle (paramètres d'apprentissage : caractéristiques cepstrales).

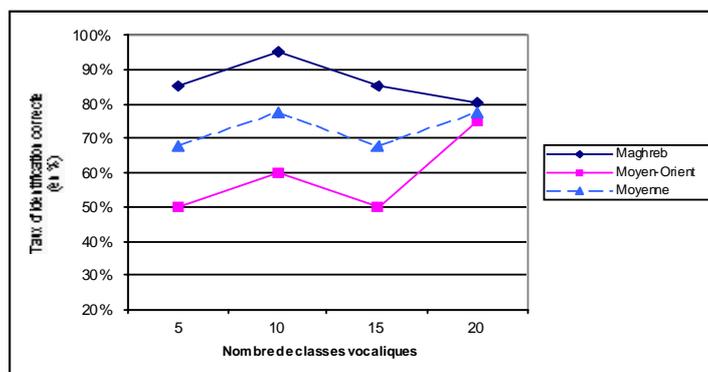


Figure 5. Taux d'identification correcte (en % et pour 40 tests) pour la discrimination par zones en fonction de la taille du modèle (paramètres d'apprentissage : caractéristiques cepstrales et durée vocalique).

Les taux d'identification correcte obtenus avec des modèles de petites tailles sont sensiblement supérieurs lorsque l'on prend en considération l'information de durée vocalique. Alors que le taux correspondant à l'identification de la variété occidentale reste élevé, le score d'identification des variétés orientales passe à 50 % avec 5 classes vocaliques (i.e., classification résultant du hasard), et à 60 % avec 10 composantes. Ces résultats moyens s'expliquent du fait qu'avec un faible nombre de composantes, le modèle n'est pas assez puissant pour permettre une bonne modélisation des systèmes vocaliques orientaux. A l'inverse, avec 20 classes gaussiennes, nous obtenons 78 % d'identification correcte. L'analyse statistique de ce score permet d'écarter l'hypothèse qu'il soit dû au hasard ($P \chi^2 > 3.84$) = .05. Les différences de scores obtenus par le modèle à 20 composantes pour la discrimination du Maghreb vs. du Moyen-Orient se révèlent non-significatives sur le plan statistique, ce qui nous autorise à dire que le modèle est aussi performant pour l'identification des stimuli maghrébins que pour la discrimination des parlers orientaux.

L'utilisation conjointe de plusieurs indices discriminants (i.e., ici la dispersion et la réalisation de l'opposition de durée vocaliques) conduit à une amélioration sensible des scores de reconnaissance passant de 70 % d'identification correcte avec les caractéristiques formantiques des systèmes vocaliques seules à 78 % lorsqu'on prend en compte la durée vocalique. La différence des taux d'identification observée entre l'une et l'autre de ces deux conditions expérimentales présente un écart significatif au plan statistique ($T(39; 1.68) = 1.78$; $p = 0.5$). Ces deux critères peuvent donc être définis comme robustes pour la discrimination automatique des parlers arabes par zone géographique principale. Il convient cependant de rappeler que le nombre de classes vocaliques constitue un facteur essentiel quant aux performances des systèmes d'identification automatique. Compte tenu du faible nombre de données dont nous disposons à l'heure actuelle, nos modèles d'apprentissage étant élaborés à partir des réalisations vocaliques de cinq locuteurs par zone uniquement, nous n'avons pas pu tester les résultats d'identification résultant d'une modélisation plus fine des systèmes vocaliques. Il est, en effet, fort probable qu'un modèle appris avec un nombre supérieur de locuteurs conduirait à de meilleures performances. Nous savons par ailleurs qu'en identification des langues et avec des modèles indépendants du locuteur, la durée des énoncés de test influence grandement les performances des modèles.

Afin de vérifier l'influence de ce facteur, nous avons répété l'expérience d'identification par zone avec 10 tests. Dans cette condition nous avons considéré les quatre répétitions de chaque locuteur comme un bloc unique. La décision d'identification porte donc sur deux minutes de parole continue et non pas sur trente secondes comme dans l'expérience précédente. Les propriétés du modèle utilisé ici correspondent au modèle optimal retenu dans l'expérience précédente (i.e., 20 classes vocaliques).

Avec 10 tests, nous obtenons pour la première condition expérimentale (modélisation des caractéristiques formantiques seules) 70 % d'identification correcte. Dans ce cas, le manque de données ne nous permet pas d'écarter l'hypothèse que ce score puisse être dû au facteur chance (correspondant à 50 %). Toutefois, Lorsque l'on intègre au modèle le paramètre de durée vocalique les scores atteignent 90 % ce qui nous permet de rejeter statistiquement cette éventualité ($P \chi^2 > 3.84$) = .05. Enfin, nous avons vu que sur 40 tests, l'intégration du paramètre de durée conduisait à améliorer de manière significative les scores d'identification (de 70 % à 78 %). Sur 10 tests, bien que les pourcentages obtenus à l'issue de la seconde condition (i.e., 8 MFCC + D) semblent indiquer de meilleures performances lorsque l'on prend en compte le paramètre de durée pour la modélisation des systèmes vocaliques (de 70 % à 90 %), la différence observée s'avérant non-significative au plan statistique.

9. Conclusion

L'objectif du présent travail a consisté à démontrer dans quelle mesure les systèmes d'IAL peuvent tirer avantage des capacités perceptives humaines pour la discrimination linguistique.

Dans une première partie, consacrée aux approches en identification perceptive des langues nous avons passé en revue les démarches expérimentales aspirant à mettre en évidence les performances et les stratégies perceptives humaines lors de différents paradigmes expérimentaux d'identification linguistique. Nous avons ainsi constaté que ces approches permettent de signaler des indices discriminants de nature linguistique, potentiellement adéquats à une modélisation automatique. Notons principalement l'importance des paramètres rythmiques, exploitées avec succès par le système perceptif humain et cela dès le plus jeune âge des sujets. En revanche, les études prenant en compte l'impact des indices de nature segmentale sont encore peu nombreuses et n'exploitent que très peu les potentialités du domaine. Enfin, il n'est pas encore possible de réellement parler d'une exploitation automatique des indices perceptifs. Pour l'heure, une véritable utilisation conjointe des acquis des deux domaines, des études perceptuelles et des approches automatique, n'a pas encore été faite.

Dans une seconde partie, nous avons mis en évidence que des expériences perceptuelles menées auprès de sujets naïfs, peuvent conduire au repérage de critères linguistiques fiables, permettant la discrimination des dialectes arabes. Au travers d'un paradigme expérimental d'identification, conçu afin de déterminer la division perceptuelle du monde linguistique arabe, telle qu'elle est ressentie par des auditeurs/locuteurs arabophones, nous avons constaté le rôle primordial des particularités vocaliques des dialectes arabes dans leur discrimination. Une analyse acoustique ultérieure nous a permis de vérifier la pertinence de cette information et de renforcer l'hypothèse selon laquelle la distribution des voyelles dans l'espace acoustique ainsi que la réalisation de l'opposition de durée vocalique (i.e., voyelles brèves vs. voyelles longues) représentent des critères fiables de discrimination des parlers arabes par zones géographiques principales. Plus précisément, cette fiabilité se vérifie lorsque, modélisés dans un système d'identification automatique, ces critères s'avèrent particulièrement robustes et suggèrent des voies intéressantes à suivre afin d'améliorer les performances automatiques.

Dans une étape ultérieure, nous nous proposons de compléter les expériences portant sur le poids des informations suprasegmentales dans le processus d'identification perceptuelle des dialectes à partir des études déjà réalisées (Barkat *et al.*, 1999 ; Ghazali *et al.*, 2002). Comme dans le cas de l'estimation de la pertinence des systèmes vocaliques, des expériences seront menées avec un système d'identification automatique basé sur la modélisation des unités rythmiques (Farinas *et al.*, 2001 ; Pellegrino *et al.*, en préparation). Enfin, le poids relatif des informations segmentales (systèmes vocaliques) et suprasegmentales (en particulier rythmiques) sera évalué dans des systèmes exploitant la fusion des modèles précédemment cités.

10. REFERENCES BIBLIOGRAPHIQUES

ABU-HAIDAR, F., Variabilité et invariance du système vocalique de l'arabe standard, Thèse de Doctorat nouveau régime en Sciences du Langage de l'université de Franche-Comté, Université de Besançon, 1991.

ANGOUJARD, J.-P., « Allongement et abrègement vocalique en arabe tunisien », *Analyses et théorie*, vol. 1, 1979, p. 1-12.

ATKINSON K., « Language identification from non segmental cues », *Journal of the Acoustical Society of America*, vol. 44, n°378 (A), 1968.

BARKAT, M., « Identification dialectale et détermination expérimentale d'indices discriminants pour les parlers arabes », *Actes des XXIIèmes Journées d'Études sur la Parole JEP'98*, Martigny, 1998, p. 71-74.

BARKAT, M., « Identification of Arabic dialects and experimental determination of discriminative acoustic cues », *Proceedings of International Congress of Phonetic Sciences ICPhS'1999*, San Francisco, Août 1999, p. 901-903.

BARKAT, M., OHALA, J., & F. PELLEGRINO, (1999) « Prosody as a Distinctive Feature for the Recognition of Arabic Dialects » *Proceedings of Eurospeech'99*, Budapest Septembre 1999, pp. 395-398.

BARKAT, M., « Détermination d'indices acoustiques robustes pour l'identification automatique des parlers arabes », Thèse de Doctorat nouveau régime en Sciences du Langage, Université Lumière Lyon 2, 2000.

- BARKAT, M., « Vers l'identification automatique des parlers arabes », *Revue Langue, Linguistique Revue Internationale de Linguistique*, vol. 7, T. Benkirane et M. Ennaji Ed., Maroc, 2001, p. 47-73.
- BARKAT, M., MEUNIER, F., SYNUKYUV, L., MAGRIN-CHANGNOLLEAU, I., « Intelligibility of Time-Reversed Speech in French », *Proceedings of TIPS*, Aix-en-Provence, 8-10 Avril, 2002, sous presse.
- BONTÉ, R., « Can you identify a language from its prosody ? », Unpublished M.A. Thesis, University of California Berkeley, 1975.
- BOND Z., FOKES J., « Identifying foreign languages », *Proceedings of International Congress of Phonetic Sciences ICPhS'1991*, 1991.
- BOND, Z., STOCKMAL, V., MOATES, D., « Same talker, different language », *Applied Psycholinguistics*, Vol. 21, 2000, p. 383-393.
- COHEN, D., « Koiné, langue commune et dialectes arabes », *Janua Linguarum, Série Practica*, Vol 81, Mouton, p.105-25, 1970.
- COLE, R. A., YAN, Y., MAK, B., FANTY, M., BAILEY, T., « The contribution of consonants versus vowels in word recognition of fluent speech », *Proceedings Of International Conference on Acoustics, Speech and Signal Processing*, Atlanta, 1996.
- FARINAS, J., PELLEGRINO, F., « Automatic Rhythm Modeling for Language Identification », *Proceedings of Eurospeech'01*, 2001
- FENNAN, M., Phonologie et morphologie d'un parler arabe marocain de Rabat, Thèse de troisième cycle, Université de Paris V, 1986.
- FOREMAN, C. G. "Dialect identification from prosodic cues" in *Proceedings. of 14th International Congress of Phonetic Sciences*, San Francisco, 1999.
- FLEISH, H., « Arabiyya », *Encyclopédie de l'Islam*, Tome 1, Brill & Maisonneuve, p. 593, 1975.
- GHAZALI S., « Du statut des voyelles en arabe », *Analyses et Théorie*, Vol. 2/3, 1979, p. 199-217.
- GHAZALI, S., BRAHAM, A., « Voyelles longues et voyelles brèves en arabe standard : organisation temporelle », *Actes des 19èmes Journées d'Etudes sur la Parole JEP'92*, Bruxelles, 19-22 mai 1992, p. 1924.
- GHAZALI, S., HAMD, I R., BARKAT, M., « Rhythm variations in Arabic Dialects », *Proceedings of Speech Prosody 2002*, Aix-en-Provence, 11-13 avril, pp. 331-34.
- HOMBERT, J.-M., MADDIESON I., « A linguistic approach to Automatic Language Recognition », *Actes du Congrès International des Linguistes*, 1997.
- JOMAA M., « L'opposition de durée vocalique en arabe : essai de typologie », *Actes des 20èmes Journées d'Etudes sur la Parole JEP'94*, Tregastel, 1-3 juin 1994, p. 395-400.
- LORCH, M., MEARA, P., « How people listen to languages they don't know », *Language Sciences*, Vol. 11, N° 4, p.343-353, 1989.
- OHALA, J.J., GILBERT, B., « On listeners' ability to identify languages by their prosody », *Problèmes de prosodie*, vol. 2, Léon & Rossi (Eds), 1979, p. 123-131.
- MADDIESON, I., VASILESCU, I, Factors in human language identification, *Proceedings of 7th International Conference on Spoken Language Processing*, Denver, Colorado, 2002.
- MAIDMENT, J.A., « Language recognition and prosody : further evidence », *Speech Hearing and Language : work in progress*, UCL N° 1, 1983, p. 133-141.
- MARKS, E. A., Bond Z.S., Stockmal V., « The effect of proficiency in a specific foreign language on the ability to identify a novel foreign language », *Proceedings Of International Congress of Phonetic Sciences ICPhS'1999*, San Francisco, 1999.
- MARÇAIS, P., *Esquisse grammaticale de l'arabe maghrébin*, Maisonneuve, 1975.
- MITLEB, F.M., « Voicing effect on vowel duration is not an absolute universal », *Journal of Phonetics*, vol. 12, 1984, p. 23-27.
- MORI, K., TOBA N., HARADA, T., ARAI, T., KOMATSU, M., AOYAGI, M., MURAHARA, Y., « Human Language Identification with Reduced Spectral Information », *Proceedings of Eurospeech*, Budapest, 1999.

MUTHUSAMY, Y., COLE, R.A., « Automatic segmentation and identification of ten languages using telephone speech" dans Proc. of International Conference on Speech and Language Processing, Banff, Alberta, Canada.

MUTHUSAMY, Y. K., « A Segmental Approach to Automatic Language Identification », PhD dissertation, Computer Science and Engineering, Oregon Graduate Institute of Science and Technology, 1993.

MUTHUSAMY, Y. K., BARNARD E., COLE R. A., « Automatic Language Identification: A Review/Tutorial. », *IEEE Signal Processing Magazine*, Vol. 11, N°4, p.33-41, 1994.

MUTHUSAMY, Y. K., JAIN, N., COLE, R.A., « Perceptual benchmarks for automatic language identification », *IEEE Signal Processing Magazine*, 1994.

NAVRATIL, J., « A perceptual experiment in language identification », *The 8th Czech-German Workshop « Speech Processing »*, Prague, 1998.

PELLEGRINO, F. & ANDRÉ-OBRECHT, R., « Automatic language identification : an alternative approach to phonetic modelling », *European Journal Signal Processing. Special Issue on Emerging Techniques for Communication Terminals*, Eds. Elsevier, Volume 80 issue 7 July 2000

PELLEGRINO, F. , BARKAT-DEFRADAS, M., « Dialectal Identification : The case of Arabic » (en preparation)

POLS, L. C. W., « Flexible human speech recognition », *Proceedings of IEEE Workshop on Automatic speech recognition and understanding ASRU-97*, Santa Barbara, p. 273-283, 1997.

POLS, L. C. W. (1999), « Flexible, robust, and efficient human speech processing versus present-day speech technology », *Proceedings of International Congress of Phonetic Sciences ICPhS'1999*, Vol. 1, San Fransisco, p. 9-16, 1999.

RAMUS, F., Rythme des langues et acquisition du langage, Thèse de Doctorat Nouveau Régime en Sciences Cognitives, EHESS, 1999.

RAMUS, F., MEHLER, J., « Language identification with suprasegmental cues: A study based on speech resynthesis », *Journal of Acoustical Society of America JASA*, Vol. 105, N°1, p.512-521, 1999.

RJIABI-SABHI, S., Approches historique, phonologique et acoustique de la variabilité dialectale arabe : caractérisation de l'origine géographique en arabe standard, Thèse de Doctorat nouveau régime en Sciences du Langage, Université de Besançon.

STOCKMAL, V., MULJANI, D., BOND Z., « Can children identify samples of foreign languages as same or different? », *Language Sciences*, Vol. 16, N°2, p.237-254, 1994.

STOCKMAL, V., MULJANI, D., BOND, Z., « Perceptual Features of Unknown Foreign Languages as Revealed by Multi-dimensional Scaling », *Proceedings of International Conference of Spoken Language Processing ICSLP'1994*, Philadelphia, 1996.

STOCKMAL, V., BOND, Z.S., « Rhythm and region: scaling the perceptual dimensions of Korean », *Proceedings of International Congress of Phonetic Sciences, San Francisco ICPhS'1999*, 1999.

VASILESCU, I., PELLEGRINO, F., HOMBERT, J-M., « Perceptual features for the identification of Romance Languages », *Proceedings of International Conference of Spoken Language Processing ICSLP'2000*, Beijing, China.

VASILESCU, I., Contribution à l'identification automatique des langues romanes, Thèse de Doctorat Nouveau Régime en Sciences du Langage, Université Lumière Lyon 2, 2001.

VASILESCU, I., « Stratégies perceptives en identification des langues », *Actes des XXIVèmes Journées d'Etudes sur la Parole JEP'2002*, Nancy, 2002 (à paraître).