

Synthèse Vocale Arabe Sous MBROLA

Zouhir ZEMIRLI — Salima KHABET

Laboratoire MCS – Institut National d'Informatique
BP 68M Oued-smar
16309 ALGER
ALGERIE
z_zemirli@ini.dz, khabet_s@hotmail.com

RÉSUMÉ. L'objet de cet article est de décrire la modélisation et la prise en compte de connaissances phonétiques, phonologiques, morpho-lexicales et syntaxiques nécessaires au développement d'un système complet de synthèse vocale à partir de textes arabes voyellés (*ARAVOICE*) basé sur le moteur de synthèse du projet Internet MBROLA. La génération des signaux acoustiques à partir de la chaîne prosodico-phonétique produite par *ARAVOICE* est confiée au synthétiseur *MBROLA* et la *base de diphones AR1*.

ABSTRACT. The object of this article is to describe the modelling and the taking into account of knowledge phonetical, phonological, morpho-lexical and syntactic necessary to the development of a complete Arabic Text To Speech System starting from diacritized Arab texts. *ARAVOICE* is based on the engine of synthesis of project Internet MBROLA. The generation of the acoustic signals starting from the prosodic phonetic chain produced by *ARAVOICE* is entrusted to synthesizer *MBROLA* and the *AR1* *diphone databases*.

MOTS-CLÉS : Synthèse vocale arabe, analyseur morpho-syntaxique, groupe syntaxique, groupe de souffle, modèle de durée, contours prosodiques.

KEYWORDS: Arabic TTS, morphosyntactic analyzer, syntactic grouping, group of breath, model of duration, prosodic contours.

1. Introduction

Le projet Internet MBROLA vise à favoriser les collaborations internationales quant à la réalisation de voix de synthèse dans nombre de langues et de dialectes [1]. Ces voix sont gratuites, libres d'utilisation pour des applications non commerciales et non militaires. 28 langues existent et pour chaque langue, une ou plusieurs voix sont disponibles (59 voix en tout). La base de diphones AR1 est utilisée pour le développement du système *ARAVOICE*. La figure 1 représente la chaîne de traitement du système de synthèse vocale à partir de textes arabes voyellés. La notation des exemples décrits est celle des phonèmes utilisés par la base de diphones AR1 (<http://lcts.fpms.ac.be/synthesis/>)

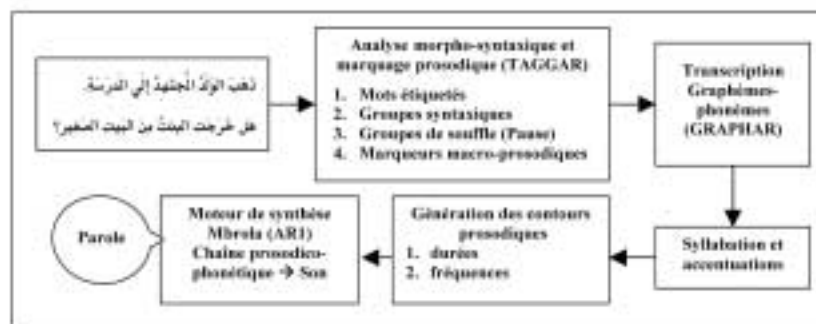


Figure 1. Chaîne de traitement du système *ARAVOICE*

La génération automatique de la chaîne prosodico-phonétique a nécessité la conception et le développement des composantes suivantes : Un système de prétraitement de textes arabes voyellés, un

système de conversion graphèmes-phonèmes, un étiqueteur morphosyntaxique (TAGGAR) destiné au système de synthèse vocale à partir de textes arabes voyellés, la modélisation de la durée des phonèmes, la modélisation de l'accentuation des mots et la génération des contours prosodiques. Nous présentons notre étiqueteur morphosyntaxique et le système de groupage syntaxico-prosodique. L'étiqueteur fournit les informations nécessaires d'une part, à la bonne phonétisation des mots d'un texte et d'autre part, au calcul du contour prosodique de la phrase à synthétiser pour la génération des pauses et des contours intonatifs.

Le module GRAPHAR réalise la phonétisation des textes à synthétiser. GRAPHAR effectue un contrôle préalable de l'écriture des mots en terme de cohérence de voyellation puis, les traitements nécessaires aux entités non lexicales à l'aide d'analyseurs syntaxiques et de lexiques généraux et dédiés. La conversion graphèmes-phonèmes est assurée par des règles morpho orthographiques et phonologiques.

Afin de produire un signal de synthèse aussi naturel et intelligible que possible, nous nous sommes intéressés à plusieurs facteurs prosodiques à travers l'observation de la durée des phonèmes en fonction des structures syllabiques, des pauses et du phénomène de déclinaison. Nous exposerons la méthodologie d'extraction des données à partir des corpus acquis pour le modèle de durée, l'accentuation des mots et la génération de l'intonation.

Nous ferons une évaluation du système ARAVOICE couplé au système de synthèse MBROLA et nous discuterons les résultats obtenus et l'apport des outils développés pour la génération de la chaîne prosodico-phonétique produisant une synthèse intelligible et comportant des contours mélodiques la rendant très acceptable.

2. Analyse morphosyntaxique et groupage syntaxique TAGGAR

TAGGAR est un analyseur morphosyntaxique spécialement développé pour la synthèse vocale arabe à partir de textes arabes voyellés. Il a pour objectif de produire une analyse syntaxique en vue de son utilisation en temps réel dans un système de synthèse vocale. Il faut que l'analyseur soit adapté aux contraintes inhérentes au fonctionnement d'un système de synthèse vocale, en terme de performances mais également en terme de résultats (rapide, robuste et déterministe). A partir d'étiquettes morphosyntaxiques, il réalise un groupage syntaxique et un calcul des contours prosodiques des phrases afin d'améliorer la qualité de la synthèse vocale produite. Cet outil est important dans un système de synthèse de la parole à partir du texte, car l'insertion des pauses et la génération des marqueurs prosodiques, ne peut être faite que si l'on dispose d'un minimum d'information grammaticale sur chaque mot de la phrase. Les textes soumis à l'entrée de notre système sont correctement voyellés. L'étiqueteur a pour rôle d'identifier la catégorie d'un mot donné (à partir d'un texte voyellé) et de lui affecter une étiquette morphosyntaxique (verbe inaccompli, nom accusatif, etc.). Tout système d'analyse morpho-lexicale se compose de deux parties essentielles qui jouent un rôle très important, notamment une base de données lexicales qui contient et décrit les ressources lexicales et un analyseur.

2.1. Ressources lexicales

Les besoins en ressources lexicales nécessaires au développement d'un système de synthèse vocale de bonne qualité ne sont plus à démontrer. Des projets comme Multex [2] ou BDLEX [3] ont grandement contribué à l'amélioration du naturel des synthèses vocales au niveau européen. Les ressources électroniques en langue arabe ne sont pas encore disponibles en vue d'un traitement automatique bien que de nombreux travaux de laboratoire aient été réalisés [4]. Nous avons défini pour notre travail des ressources, certes partielles, mais que nous jugeons suffisantes pour atteindre nos objectifs. Les ressources lexicales sont composées de quatre lexiques : le lexique relatif aux schèmes générateurs, le lexique des éléments affixés, le lexique des mots outils et le lexique des mots spécifiques.

Chaque lexique a pour rôle de fournir toutes informations grammaticales nécessaires à l'analyseur. Le lexique des schèmes générateurs est relatif uniquement aux verbes. Le lexique des mots outils recouvre l'ensemble des particules (les prépositions, les conjonctions de coordination, les particules d'interrogation,

etc.). Le lexique des mots spécifiques contient l'ensemble des mots non dérivables tels que les noms empruntés, les noms propres, etc. Enfin, le dernier lexique contient l'ensemble des éléments affixés.

Le lexique des schèmes verbaux est subdivisé en trois sous lexiques correspondant aux trois temps du verbe. Trois modules de conjugaison sont utilisés pour générer les formes conjuguées des schèmes des trois sous lexiques. Il comporte 72 schèmes pour l'accompli (53 schèmes pour la voix active et 19 schèmes pour la voix passive), 55 schèmes pour l'inaccompli (36 schèmes pour la voix active et 19 schèmes pour la voix passive) et 44 schèmes pour l'impératif. Les 171 schèmes recensés se divisent en deux catégories : Les schèmes sains qui sont des schèmes pour les verbes sains trilitères ou quadrilitères dont la dérivation produit des formes régulières sans traitement particulier et, les schèmes de surface qui sont des schèmes créés de manière artificielle pour pouvoir générer les formes dérivant des verbes faibles.

2.2. Choix du jeu d'étiquettes

Débili, dans ses travaux [5] sur l'étiquetage grammatical et la voyellation automatique, définit 606 étiquettes qui se répartissent en : 50 proclitiques, 9 enclitiques et 547 formes dites simples (297 noms, 218 verbes et 32 particules). Nos textes sont entièrement voyellés et notre premier objectif est de regrouper les mots en groupes syntaxiques. Les groupes syntaxiques auront comme élément central un **verbe** ou un **nom** qui seront reliés par des **particules** de coordination, de relation, du génitif, des pronoms, etc. Ceci permet de dégager trois grandes classes d'étiquettes : les particules, les verbes et les noms. Un verbe sera caractérisé par son temps (accompli/inaccompli et impératif) et son mode (actif ou passif) qui renferment des informations utiles à la génération de la prosodie. Ces informations formeront des sous-classes d'étiquettes associées aux verbes. Les flexions casuelles des noms : sujet, accusatif et génitif ainsi que leur type déterminés/indéterminés jouent un rôle déterminant dans la classification des étiquettes de cette classe. Partant de ces réflexions, nous avons défini pour nos besoins 35 étiquettes grammaticales. Elles se répartissent en trois grandes catégories :

- 4 étiquettes pour 28 catégories de particules (interrogation, réponse, condition, etc.).
- 16 étiquettes pour les verbes (6 pour l'accompli, 6 pour l'inaccompli et 4 pour l'impératif)
- et 15 étiquettes pour les noms (5 pour le cas sujet, 5 pour le cas accusatif et 5 pour le cas génitif).

2.3. Hiérarchie d'analyse morphosyntaxique

L'ordre de traitement des mots d'un texte est important car il permet de minimiser les erreurs d'étiquetage. Le traitement se fait dans l'ordre suivant : Analyse des mots outils et des mots spécifiques, analyse des formes verbales et enfin, analyse des formes nominales.

2.4. Ambiguïté d'interprétation

L'étiquetage de textes arabes même entièrement voyellés peut aboutir à des cas d'ambiguïtés d'étiquetage. C'est le cas des homographes. Par exemple, **لن** est soit un Harfnasb (particule de nasb) ou un verbe à l'accompli (il se plaint). Des règles de désambiguïsation sont utilisées pour assigner une étiquette en fonction des contextes. **لن** sera étiqueté comme un HarfNasb dans la phrase : **تلعنون لن السماء لا** **لن** المرين من الهم الشديد. **لن** شمر ذهابا ولا فطنة.

2.5. Applications de TAGGAR

Nous allons dans ce qui suit décrire de manière succincte quelques applications directes de l'analyseur TAGGAR telles que le groupage syntaxique, l'insertion des groupes de souffle et la détermination de l'accentuation locale des mots ou la génération du contour intonatif. L'objectif de cet article n'étant pas de détailler les applications de TAGGAR mais de montrer son apport dans un système de synthèse vocale.



2.5.1. Etiquetage et groupage syntaxique

L'étiquetage de la phrase : شرب طفلًا لطيفًا الحليب اللذيذ وأكل رجلاً البرقاقة المتعشة. fournit le résultat suivant : اللذيذ (NomDetAccusatif-27) الحليب (NomIndSujet-30) لطيفًا (NomIndSujet-30) طفلًا (NomIndSujet-30) شرب (VerbeAcc-49) البرقاقة (NomDetAccusatif-27) وأكل (VerbeAccPref-50) رجلاً (NomIndSujet-30) المتعشة (NomDetAccusatif-27). (Ponct-65). A partir des quatre groupes syntaxiques défini (£ : groupe verbal, □ : groupe nominatif sujet, § : groupe nominatif accusatif et S : groupe nominatif génitif), nous réalisons le *groupage syntaxique* (£شرب) (طفلًا لطيفًا) (§الحليب اللذيذ) (£وأكل) (رجلاً) (§البرقاقة المتعشة).

2.5.2. Insertion des groupes de souffle

Un groupe de souffle peut être défini comme étant un ensemble de syllabes ou de mots pouvant être prononcé sans pause. Un groupe de souffle est composé au minimum d'un groupe syntaxique. Cela implique qu'une pause n'est jamais insérée à l'intérieur d'un groupe syntaxique tel que nous l'avons défini. Le modèle de découpage en groupe de souffle est fondé sur les signes de ponctuation, la nature des groupes syntaxiques et le nombre de mots depuis la dernière pause. Après la phase de regroupement syntaxique, nous pouvons réaliser l'insertion des *pauses* entre les *groupes de souffle*. (شرب) (طفلًا لطيفًا) (شرب) (الحليب اللذيذ) # (وأكل) (رجلاً) (البرقاقة المتعشة). Le symbole # indique l'insertion d'une pause obligatoire.

2.5.3. Accentuation et intonation

Les groupes syntaxiques définis permettent de définir des remontées intonatives après les groupes syntaxiques. Ils servent pour la réinitialisation de la pente intonative qui caractérise la fréquence fondamentale des phonèmes générés. Les étiquettes produites sont utilisées pour définir aussi bien la syllabe accentuée des mots mais aussi la hauteur de cette accentuation.

2.6. Evaluation

A partir d'un corpus de 5563 mots, l'étiqueteur morphosyntaxique a engendré un taux d'erreur de 2% sur les étiquettes qui a entraîné 1% d'erreurs sur les frontières de groupes syntaxiques. Environ 2% des pauses de groupe de souffles ne sont pas correctement placées. Nous illustrons dans ce qui suit quelques causes d'erreurs d'étiquetage. Le mot بطاقات dans la phrase بطاقات كثيرة. peut être interprété comme le mot بطاقات (des cartes) ou bien بطاقات + ب (avec des ressources). Cela n'a aucune incidence sur le regroupement, on obtient le même groupe nominal génitif. (بطاقات كثيرة).

Le processus d'étiquetage morphosyntaxique repose sur la priorité de l'étiquetage des verbes sur les noms. Certains noms ont une structure qui s'apparente à celle des verbes. Le mot لفتّر est étiqueté comme un verbe alors qu'il correspond à un nom sujet indéterminé.

Cette erreur n'a aucune incidence sur le regroupement syntaxique ni sur l'insertion de la pause après un groupe de souffle. Par exemple, la phrase : وإن اجتمع شخصان أو أكثر من أهل الحضارة. est regroupée correctement en : (وإن) (اجتمع) (شخصان) (أو أكثر) # (من أهل الحضارة).

La langue arabe comporte un certain nombre de mots qui sont inflexibles, leur flexion casuelle ne subit aucun changement par l'influence d'un régissant. Par exemple, le nom propre de femme حدّام /HaDami/ est inflexible dans les trois cas. Dans جاءت حدّام (/HaDami/ est au cas sujet), مررت بحدّام (/HaDami/ est au cas génitif), et dans رأيت من بين الأشجار حدّام (/HaDami/ est au cas accusatif). Les règles contextuelles permettent de bien étiqueter حدّام dans les trois cas.

Les textes soumis à notre système de synthèse vocale ARAVOICE sont supposés entièrement et correctement voyellés. Néanmoins, des omissions de la voyelle finale dans des mots ont engendré quelques erreurs d'étiquetage des mots. D'autres mots ont été rectifiés facilement à l'aide des constituants du mot et de règles contextuelles. Par exemple, le mot ياتنكر bien qu'étant amputé de sa voyelle finale est correctement étiqueté comme NomDetGenitPref.



3. Transcription graphèmes phonèmes GRAPHAR

La synthèse correcte d'un texte doit apporter une bonne transcription phonétique et une bonne prosodie adaptée à la structure syntaxique et au contenu sémantique du document. Un prétraitement est nécessaire avant d'effectuer la conversion graphème-phonème. Certains chiffres correspondent à des nombres (1999), d'autres à des notations de dates (1998/01/01) (98/06/30) ou d'heure (12:30). Des unités monétaires diverses sont utilisées (€, \$, "د.ج." /d.Z./ "dinar algérien"). Des abréviations d'unités de mesures usuelles sont utilisées "كغ" /kG/ ("kilogramme"), "طن" /t./ ("tonne"); Des sigles particuliers à un domaine donné sont utilisés. وقت /w.t./ [6]. Le module de transcription graphème-phonème d'ARAVOICE utilise une approche à base de règles de transcription, un lexique général et un lexique particulier propre à l'utilisateur. La transcription orthographique-phonétique est réalisée par une grammaire basée sur des règles contextuelles de transcription. Une règle de cette grammaire indiquera le lien existant entre le graphème et le phonème associé. Afin de symboliser le mieux possible la situation des éléments en question dans une chaîne orthographique et leur rôle dans la transcription, une règle est de la forme : CG + CT + CD → /SP/. Contexte Gauche + Caractères à Transcrire + Contexte Droit = /Séquence Phonétique/. La séquence phonétique peut indiquer une action complémentaire à effectuer. Le nombre de règles étant assez important, nous avons classé les graphèmes pouvant figurer dans CG et CD en 9 classes. 0 : Tout caractère, 1 : Consonne, 2 : Consonne lunaire, 3 : Consonne solaire, 4 : Voyelle, 5 : Voyelle courte, 6 : Voyelle longue, 7 : Consonne emphatique, 8 : Consonne non emphatique. Les règles de transcription ont pour rôle de rendre compte des phénomènes réguliers de la langue arabe. L'étude morphologique, phonétique et phonologique de la langue arabe standard, nous a permis de constater que la forme orthographique de la langue arabe est très proche de la forme phonétique à quelques exceptions près. La prononciation dépend du contexte immédiat du graphème à transcrire. Chacune des règles de prononciation opère pour convertir un graphème (ou groupe de graphèmes) en un phonème (ou groupe de phonèmes), selon le contexte de réalisation. L'ensemble des règles du module de conversion graphème-phonème est réparti en deux catégories : les règles morpho orthographiques et les règles phonologiques. Dans ce qui suit, nous allons citer quelques règles de transcription graphème-phonème et des exemples où elles sont déclenchées.

3.1. Voyelles courtes, voyelles longues et madd

Le traitement de l'emphase consiste à choisir les voyelles emphatiques, brèves ou longues qui doivent suivre les consonnes emphatiques. Les cinq lettres dites de prolongement 'ا' /a/, 'و' /w/, 'ي' /j/, 'هـ' /h/ (alif maqoura en fin de mot) et 'هـ' (alif mamdouda) sont traitées.

7a0	→	/a/	حرب	/d.a.raba/
Suwa#	→	/uu/	يونس	/daxaluu/
0A0	→	/?aa/	أمن	/?aamana/

3.2. Traitement de la Hamza

Le traitement de la hamza est assez particulier. Il existe deux types de hamza :

1. la hamza stable, qui se prononce toujours et se manifeste comme un [ʔ] n'importe où dans un mot ; Le support graphique de la hamza varie selon sa position et la voyelle qui la précède. A l'initiale, elle a toujours le support alif ('أ', 'إ' ou 'هـ') et elle est toujours suivie d'une voyelle : إبراهيم /?ibraahim/ ("prénom Ibrahim"), أسبوع /?usbuuH/ ("semaine") ou أهن /?ayna/ ("ou?").
2. la hamza instable, que l'on ne trouve qu'en début de mot, et qui se réalise comme un coup de glotte que si le mot en question est en début de phrase ou se trouve après une pause. Ailleurs dans l'énoncé, la hamza et sa voyelle ne se prononcent pas.

La hamza instable se rencontre principalement au début de l'article : ال, au début de certaines formes verbales : ارتفعت et au début de certains noms : اسم, عين. Dans les exemples ci-dessous, l'action 9 indique que la hamza et sa voyelle sont éliminées et le déplacement de la consonne qui suit la voyelle de la hamza à la fin du mot précédent. Cette action a pour rôle d'équilibrer les syllabes des deux mots et d'éviter que deux consonnes n'apparaissent en début de mot.



5-45 → 9 كازينجا /kaana-ḡistiḡbaala/ → /kaanas-tiqbaalu/

Le traitement de la phrase 1419 هـ الموافق ل 23 رمضان 1999 م الموافق 10 حنفي produit la transcription suivante : ?al?axad-HaSara-Zanfii-?alfun-wa-tisHa-mi?ata-wa-tisHa-wa-tisHuuna-milaadil-muwaafiq-li-TalaaTa-wa-HiSruuna-ramaQaan-alfun-wa-?arbaHa-mi?ata-wa-tisHa-HaSara-hiZrii.

4. Génération des durées phonémiques

La génération des durées des phonèmes est basée sur un modèle multiplicatif selon la formule : $Dp=DI*CS*(CRA)$. La durée d'un phonème Dp est calculée en multipliant la durée intrinsèque (DI) par un coefficient syllabique (CS) et un coefficient d'allongement ou de réduction phonémique (CRA). DI est la moyenne du phonème du corpus étudié. CS est le coefficient de réduction syllabique qui est fonction du nombre de syllabe du mot. CRA est le coefficient de réduction ou d'allongement phonémique qui dépend du contexte du phonème (gémination ou situation pré pausale) [7].

5. Syllabation et accentuation

La syllabation des mots obéit à des règles simples, déterministes et universellement reconnues [8] et [9]. La syllabe en langue arabe obéit à deux règles : Le noyau syllabique est une voyelle et deux consonnes ne peuvent se suivre sauf en fin de mot et devant une pause. Ceci, permet d'obtenir deux syllabes ouvertes : CV kataba et CVV maata, et trois syllabes fermées : CVC maktab, CVVC salaam, CVCC hiat. On trouve également les appellations de syllabes lourdes (CVV et CVC) et de syllabes sur-lourdes (CVVC et CVCC). La syllabe de type CV est considérée comme courte, les autres longues.

Les règles de placement de l'accent local ont fait l'objet de nombreux travaux qui divergent sur son apparition en finale dans un mot [8], [10] et [11]. Notre étude sur l'accentuation nous permet de proposer les règles suivantes pour le placement de l'accent local. Si le mot ne comporte que des syllabes courtes : Attribuer l'accent à la première syllabe s'il ne s'agit pas d'un préfixe de type 'و' /wa/ ou 'ف' /fa/. Attribuer l'accent à la seconde syllabe si le mot est préfixé. Si la dernière syllabe d'un mot est sur-longue alors elle portera l'accent local. Sinon la syllabe accentuée (longue ou sur-longue) est l'antépénultième du mot. La dernière syllabe est exclue du processus d'accentuation.

Soit la phrase : نحن مسرورون ان نقدم لكم نظامنا لتطبيق الآلي للتصويع العربية. sa syllabation et son accentuation est : naX&nu*~mas*ruu&ruun*~?an*~nu*qa&di*ma*~la&kum~ni*Maa&ma*nan*~naa&Ri*qaI*~?aa&lii*~lin*nu*Buu&Bil*~Ha&ra*bi*ja~. '&' indique que l'accent est porté par cette syllabe, '*' indique une syllabe non accentuée et '-' est le séparateur de mots.

6. Génération des contours intonatifs

La phrase ذهب لوتذ إلى المدرسة. (ذهب لوتذ إلى المدرسة). Sa syllabation, son accentuation et son marquage prosodique est : EDa&ha*bal*wa&la*du*§?ia&lal* madd&ra*sa*ti*##.

Les remontées intonatives calculées à partir du module TAGGAR sont résumées dans le tableau 1.

Nature du groupe syntaxique	R1 = remontée intonative
GV (E)	20
GNS (□)	15
GNA (§)	10
GNG (§)	15

Tableau 1. Valeur de la remontée intonative après un groupe syntaxique



et bien perçues. Les pauses insérées automatiquement entre les groupes de souffle ont eu un impact significatif sur l'amélioration du confort d'écoute et de réduction de la charge mentale des auditeurs pour les longues phrases. Le modèle prosodique proposé se situe à deux niveaux : syntaxique et local. L'étude des groupes syntaxiques nous a permis également de définir un modèle prosodico-syntaxique basé sur les variations de fréquences d'un groupe à un autre. C'est ainsi que des facteurs de remontées intonatives ont été calculés pour chaque groupe syntaxique. Au niveau local, nous avons élaboré un modèle accentuel très fin basé sur les règles d'accentuation en fonction de la nature et du nombre de syllabes qui composent le mot. Le phénomène de déclinaison commun à toutes les langues a été élaboré à partir des données syntaxico-prosodique et du nombre de syllabes qui composent les groupes syntaxiques. Le phénomène de déclinaison de la fréquence fondamentale tout au long d'une phrase est perceptible. Cette déclinaison toutefois n'est pas totalement linéaire mais décroît du premier groupe syntaxique vers le dernier groupe syntaxique avec des réinitialisation entre les groupes. La pente de la déclinaison semble également corrélée à la longueur de la phrase.

Le système ARAVOICE a été jugé d'une qualité, d'un naturel et d'une assez bonne acceptabilité par les personnes voyantes. L'auditeur handicapé visuel a, quant à lui, très bien accepté le système. L'accentuation locale des mots est très bien perçue et améliore de façon significative le contour mélodique des mots. L'intonation bien qu'acceptable manque quelque peu de naturel, cela s'explique par le fait que notre modèle prosodique ne prend pas encore en compte les phénomènes micro prosodiques.

9. Bibliographie

- [1] Dutoit T., Pagel V., Pierret N., Bataille F., Van Der Vrecken O. (1996), "The MBROLA project : towards a set of high quality speech synthesizers free of use for non commercial purpose", *ieslp'96*.
- [2] Véronis, J.; Hirst, D.; Espresser, R.; Idé, N. (1994), "NL and speech in the MULTTEXT project", AAAI'94 Workshop on Integration of Natural Language and Speech, Seattle, Washington, pp. 72-78.
- [3] De Calmès M., Pérennou G. (1998), BDLEX : a Lexicon for Spoken and Written French. First International Conference on Language Resources & Evaluation, Grenade, 1129-1136
- [4] Dichy J., Braham A., Ghazali S., Hassoum M. (2002), "La base de connaissances linguistiques DIINAR.1", International Symposium on The Processing of Arabic, Tunis, 45-56.
- [5] Débill F., Achour H., Souici E. (2002), "La langue arabe et l'ordinateur : de l'étiquetage grammatical à la voyellation automatique", *Correspondances de FIRMC*, N° 71, pp. 10-28.
- [6] Zemirli Z., "SYNTHAR +: Arabic Text to Speech Synthesis under Multivoix ", *TSI*, VOL17, N°6 / 98, pp. 741-761.
- [7] Zemirli Z., Obrecht R.A., Henni A., Sellami M. (2001), " Prediction of the sound duration in an Arabic Text To Speech System ", *in proceedings of SPECOM'01*, Moscow, 29-31 October 2001, pp 241-245.
- [8] El-Ani S. "Arabic Phonology. An Acoustic and Physiological Investigation", The Hague, Netherlands: Mouton and Co.
- [9] Rajouani A. et al. (1997), "A perceptual study of Arabic intonation ", *Proceedings ESCA Workshop on Intonation : Theory, Models and Applications*, pp. 279-283, Athènes, Greece.
- [10] Koufoughli D. "Contribution à l'étude de l'accent en arabe littéraire", *Annales de l'Université d'Abidjan, série H*, Vol. IX, pp 124-125, 1976.
- [11] Zaki A., Rajouani A., Najl M. "Synthesizing Intonation of Standard Arabic Language Using Neural Networks", *Proceedings of Eurospeech'01 Conference*, Volume 1, pp. 541-544, 2001.
- [12] Gibbon D., Moore R., Winski R. (1997), "*Handbook of Standards and Ressources for Spoken Language Systems*", Editions Moutons de Gruyter, Berlin & New York.