LECTURE AUTOMATIQUE DE TEXTE ET NOMBRES EN ARABE STANDARD : LE SYSTEME ARPHON

Khoudir BENBELLIL, Kamel FERRAT,
Ghania DROUA-HAMDANI, Mourad ABBAS.
Centre de Recherche Scientifique et Technique pour
Le Développement de la Langue Arabe

Résumé

Cet article traite de la synthèse de la parole à partir de textes (TTS) appliquée à la langue arabe. Cette opération consiste en la conversion automatique d'un texte écrit en parole synthétique. La méthode de synthèse adoptée est la concaténation d'unités préstockées (polysons). La lecture automatique concerne les textes Arabes voyellisés contenant aussi bien des mots que des nombres. La parole obtenue est de bonne qualité mais présente cependant quelques «discontinuités» pour lesquelles des procédures de lissage s'avèrent nécessaires.

Mots- clés : Langue Arabe - synthèse de la parole - diphones - polysons.

الملخص

يتناول هذا المقال تركيب كلام منطوق عن طريق تحويل نصّ مكتوب باللغة العربية يحتـــوي على كلمات مشكلة و أعداد .تتمثل طريقة التركيب في ربط متسلسل لوحداتٍ صوتية مخزّنة. يمتـلز الكلام المولّد بالجودة إلاّ أن بعض الحشرجات تتخلله، الأمر الذي يتطلب منا إعـــداد برمجيـــات لإزالتها .

الكلمات المفاتيح : اللغة العربية – تركيب الكلام –وحدات صوتية.

Abstract

This paper deals with speech synthesis from text (TTS) applied to standard Arabic. The adopted method of synthesis is the concatenation of stored units. The automatic reading concerns texts containing vowelised words and also numbers. The resulting speech is of good quality but has some artifacts for which smoothing procedures are necessary.

Keywords: Arabic langage - speech synthesis - diphones - polyphones

1. Introduction

La lecture automatique de texte connaît de nos jours un grand intérêt de la part des chercheurs travaillant dans le domaine de la parole. Les applications potentielles vont du simple jouet pour enfants à l'aide aux aveugles en passant par les messageries vocales. Pour cela, le laboratoire du traitement de la parole du Centre de Recherches Scientifiques et Techniques pour le Développement de la Langue Arabe (CRSTDLA) a entrepris, ces dernières années, la conception d'un système de lecture automatique de textes et nombres pour la langue Arabe Standard. Cette entreprise a été couronnée par l'implémentation d'un lecteur utilisant des unités préenregistrées (polysons). Ce système dénommé ARPHON produit une parole de bonne qualité mais présente quelques « discontinuités » qu'il faut éliminer.

Le développement technologique à son stade primitif a vu les premiers essais de synthèse de la parole, tous basés sur des systèmes mécaniques. Le premier à construire un analogue mécanique du conduit vocal est C. Kratzenstein aux environs de 1780. Il est suivi en 1791 par le Baron Von Kempelen dont les recherches sur la parole, commencées dès 1769, lui ont permis de mettre au point une machine plus élaborée que la précédente pouvant émettre aussi bien des consonnes que des voyelles. En 1922, J. Stewart mis au point le premier système électrique. En 1950, un véritable synthétiseur de la parole portant le nom de Pattern Play-back est né. Il est mis au point aux laboratoires Haskins et fonctionne comme un sonagraphe en sens inverse. En 1953, les synthétiseurs à formants sont développés par W. Lawrence, en Grande-Bretagne (Parametric Artificial Talker ou PAT), par G. Fant, en Suède (Orator Verbis Electrica ou OVE), ainsi qu'au Massachusetts Institute of Technology (MIT) et aux Bell Laboratories [1].

Aujourd'hui, le développement technologique connaît un essor considérable. De nouvelles techniques de traitement du signal ainsi que de nouveaux appareils sophistiqués permettent de mieux maîtriser l'analyse des paramètres acoustico-articulatoires. Ces analyses de plus en plus poussées, permettent la réalisation de synthétiseurs efficaces. Mais on s'aperçoit vite que la synthèse d'une parole presque naturelle exige des algorithmes très complexes et par conséquent, une grande puissance de calcul [2]. Actuellement, la tendance est à l'utilisation de segments de parole naturelle préenregistrés de divers types permettant la génération de la parole par concaténation.

2. Architecture du système ARPHON

La synthèse de la parole comprend essentiellement deux phases :

-une phase de traitement de textes où des règles sont appliquées pour transformer le texte écrit en une suite de polysons ;

-une phase de génération de l'onde acoustique correspondant au texte écrit.

Il existe plusieurs techniques de synthèse de parole. Celle que nous avons adoptée consiste à rechercher dans une base de données sonores (dictionnaire de ARPHON) les segments de signal parole correspondant à la suite des unités élaborées durant la phase du traitement du texte. Les segments sont ensuite juxtaposés les uns à la suite des autres d'où

le nom de synthèse par concaténation. Le système ARPHON comprend donc trois parties essentielles à savoir (fig1) :
- la base de données (le dictionnaire),
- le module du traitement de textes,
- le module de traitement du signal.

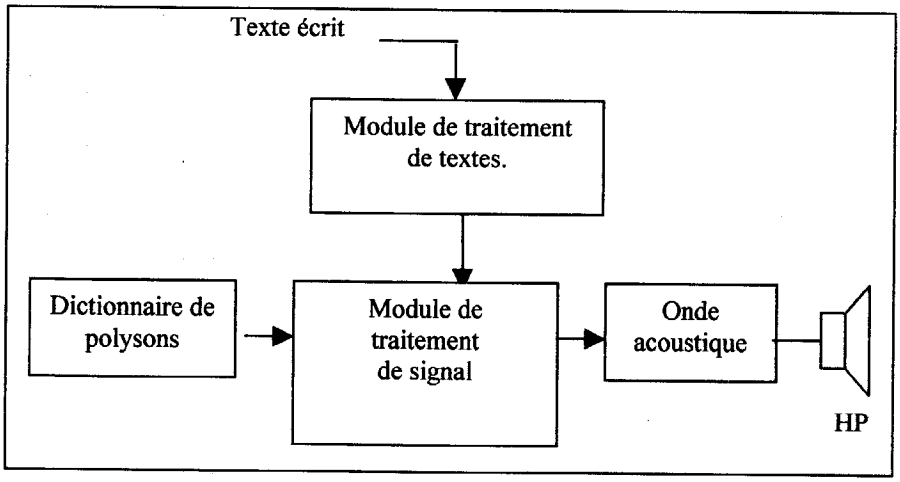


Fig. 1 : Synoptique général du système ARPHON

2.1. Le dictionnaire des polysons

Un diphone est un segment de parole qui commence du milieu d'un phonème et s'étend jusqu'au milieu du phonème adjacent. La classe des polysons est plus large que celle des diphones et comprend aussi des unités ayant un ou plusieurs phonèmes «transitoires» dans leurs parties centrales. La segmentation n'a pas été effectuée sur ces phonèmes car ils ne disposent pas de zone de stabilité spectrale. L'utilisation de telles unités confère à la parole une qualité meilleure car la concaténation s'opère uniquement sur des zones stables. La table2 donne l'inventaire des polysons du dictionnaire. Le programme de lecture automatique recherche les polysons dans le dictionnaire grâce à un index contenant les noms de ces polysons, leurs adresses dans ce dictionnaire ainsi que leurs tailles. La nomenclature de ces polysons se base sur un codage des caractères de la langue arabe standard par un jeu de caractères donné par la table 1. Ce codage est utilisé aussi dans la procédure de transcription de la phase du traitement du texte (fig. 2) permettant ainsi la 'phonétisation' du texte et par suite sa décomposition en polysons.

Harf en arabe	Code adopté
" اللّه " الجلالة لفظ في ل	L
ء	?
ب	b
ت	t
ث	ç
ج	G
ح	H
خ	K
د	d
ذ	D
ر	r
ز	z
س	S
ش	c
ص	$
ض	µ
ط	T
ظ	£
ع	§
غ	E
ف	f
ق	q
ك	k
ل	l
م	m
ن	n
ه	h
و	w
ي	y
فتحة	a
ضمة	u
كسرة	i
ممدودة فتحة	A
ممدودة ضمة	U

Harf en arabe	Code adopté
ممدودة كسرة	I
فتحتين	@
ضمتين	&
كسرتين	=
مد حرف	~
شدة	*
#	Virgule
##	Point final

Tab.1 : Code adopté pour la nomenclature des polysons.

Type de polysons	Mot porteur	Nombre de réalisations
Les unités [aLLA] et [uLLA] pour le mot (اللّه)		2
CC	#ta[CC]ata#	22X22=484
#C	[#C]ata#	22
C#	#kata[C#]	22
V#	#katat[V#]	6
VC	#?at[VC]a#	22X6=132
CV	#a[CV]ta#	22X6=132
VTV	#?at[VTV]ta#	6X6X6=216
VTC	#?at[VTC]a#	6X6X22 = 792
#TV	[#TV]ta#	6X6=36
VT#	#katat[VT#]	6X6 = 36
VTTV	#?atVTTVta#	6X6X6X6 = 1296
CLV	#?a[CTV]ta#	6X6X22 = 792
		Total : 4100 unités

Tab.2 : Inventaire des polysons du système ARPHON.

Dans la table 2, V désigne une des six voyelles [a],[u],[i],[A],[U],[I]. T désigne une des six consonnes à caractère transitoire [l],[r],[w],[y],[§] (ﺀ),[E] (ﺉ). Le reste des vingt

deux consonnes de l'arabe standard sont désignées par C. Pour extraire ces unités, nous avons choisi un ensemble de mots porteurs [3].

Dans notre implémentation, nous disposons d'un dictionnaire de 3760 unités différentes en omettant des 4100 possibles celles qui correspondent à des combinaisons CC inexistantes en Arabe Standard [4].

2.2. Le module de traitement du texte

Dans ce module, comme nous l'avons dit, le texte en entrée est transformé en une chaîne phonétique par un ensemble de procédures (fig. 2) et est ensuite décomposée en polysons. Les textes que le système est en mesure de traiter se composent de nombres, de consonnes, de voyelles et autres symboles (signes de ponctuation, etc.).

2.2.1. Les mots d'exception

Un mot d'exception [5] désigne un mot qui ne se prononce pas comme le suggère son écriture. Ces mots sont recensés sans voyellisation, puis réécrits correctement sous leur forme phonétique dans une table. Nous avons ajouté à ces mots d'autres mots que nous désignons par «invariants» et qui peuvent se trouver dans le texte sans voyellisation comme par exemple ces quelques prépositions (...،من ،على ،إلى ،حتى) (tab. 3).

Une procédure permettant l'extraction de la structure consonantique est appliquée à chaque mot du texte. Par exemple, cette procédure transforme le mot [#Dahab#] en [#Dhb#]. Le mot ainsi transformé est ensuite recherché dans la table des mots d'exception et s'il est trouvé, il est remplacé par son correspondant voyellisé comme indiqué dans la table 2.

2.2.2. Conversion des nombres en textes

Pour effectuer la conversion de nombres en texte, le programme de lecture a besoin d'un petit vocabulaire de nombres composé des mots '$ifr' désignant le nombre 0, 'wAHid' pour le nombre 1, etc. Notre lecteur utilise le vocabulaire suivant (tab. 4).

Le nombre à convertir est d'abord subdivisé en groupes de 9 chiffres et chaque groupe est lui-même subdivisé en groupes de 3 chiffres. On se ramène donc de proche en proche à une combinaison des items du vocabulaire ci-dessus. Par exemple, le nombre 14568 est converti en [#?arba§a§aCar?alfwaKamSami?awaçamAniyawaSittUn#]. Pour la lecture des nombres décimaux, il suffit de lire leurs parties entière et fractionnaire.

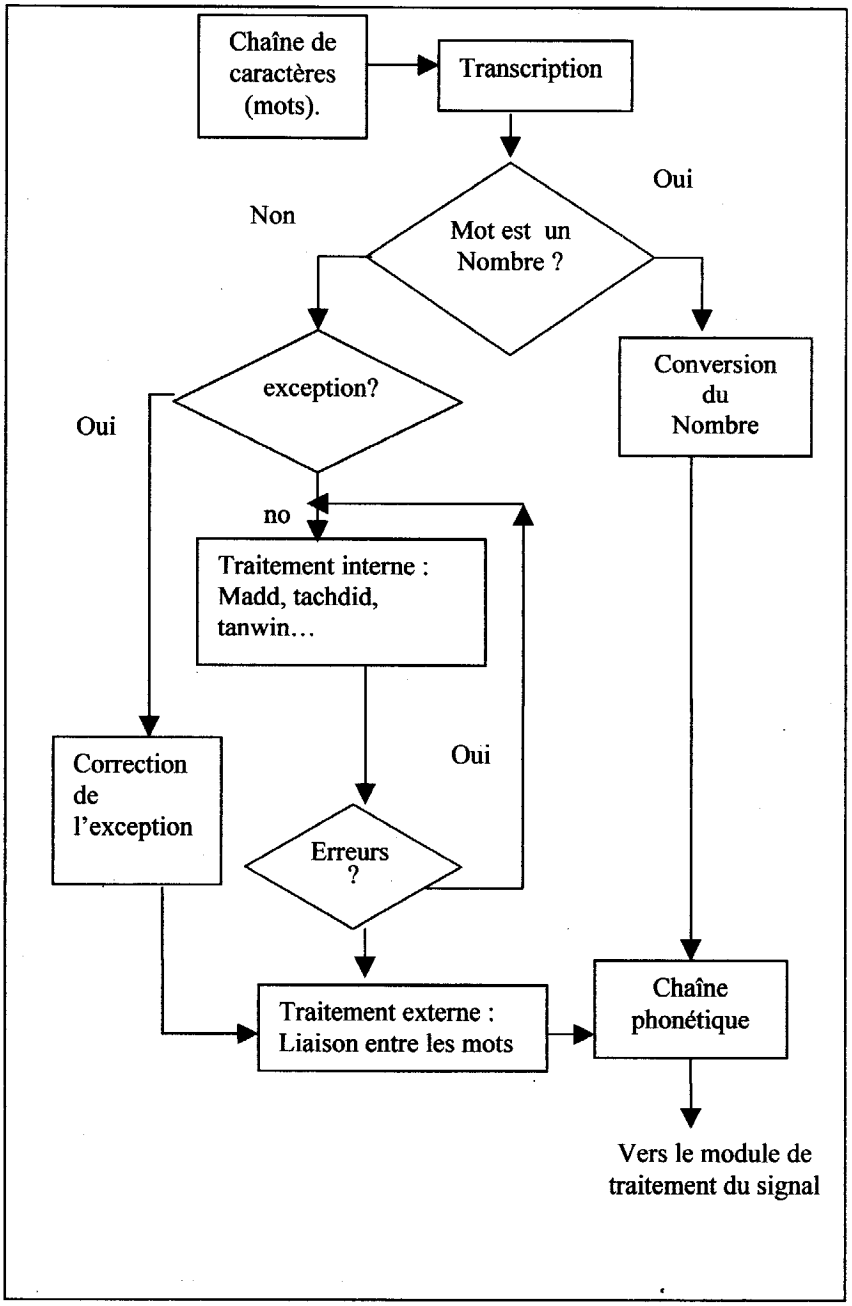


Fig. 2 : Organigramme du traitement de texte.

Mots d'exception	transcription	Corrections
هذا،	hD~	hADA
هؤلاء	h ?l~?	hA?ulA?i
ذلك	Dlk	DAlika
من، على، الى،...	Mn, §l~,~l~,...	nin, §alA, ?ilA,...

Tab. 3 : Exemples de quelques mots d'exception

Nombre	Translation littérale
0	$ifr
1	wAHid
2	?içnAn
3	çalAça
4	?arba§a
5	KamSa
6	Sitta
7	Sab§a
8	çamAniya
9	TiS§a
10	§aCara
11	?iHdA + §aCar
12	?içnA + §aCar
20	§iCrUn
30	çalAçUn
40	§arba§Un
50	KamSun
60	SittUn
70	Sab§Un
80	çamAnUn
90	TiS§Un
100	Mi?a
200	mi ?atAn
1000	?alf
2000	?alfAn
1000000	milyUn
1000000000	milyAr

Tab. 4 : Vocabulaire de la conversion des nombres en texte.

2.2.3. Le traitement des erreurs du texte

Lors de la saisie d'un texte arabe, des erreurs peuvent se produire. A cet effet, nous avons recensé les plus fréquentes. Parmi ces erreurs, nous citons :

-un caractère inconnu, signalé par la procédure de transcription.

-les caractères du madd incompatibles avec les voyelles qui les précèdent. Ainsi le harf madd représenté par l'un des hurufs [~],[w] et [y] doit être précédé respectivement par une fatha, une damma et une kasra.
-le tanwin figurant au milieu d'un mot ou précédé d'un [ال].
-une succession de deux voyelles au moins..
-un mot débutant par une voyelle.
-le harf madd après une consonne et le tachdid après une voyelle.

Lorsqu'une telle erreur est rencontrée, le lecteur la signale en générant un message approprié.

2.2.4. Les traitements internes au mot

Les procédures de traitement au niveau du mot s'inspirent des quatre règles suivantes :

-Règle de l'article défini (ال)

Cette règle intéresse le [ل] du [ال] qui doit être prononcé lorsque la consonne qui le suit est une consonne 'lunaire' (harf qamari). Dans le cas d'une consonne 'solaire' (harf chamsi), cette consonne est géminée et le [ل] n'est pas prononcé.

-Règle du madd

Une voyelle courte ([a], [u], [i]) suivie du caractère [~] est changée respectivement en ([A],[U],[I]) (voir table 2). Le symbole du madd 'ا' doit suivre une voyelle (courte) sauf dans le cas du caractère [لا] qui est équivalent à [ل] suivi du harf madd [ا] et par conséquent, est retranscrit en [la~]. De même, le harf madd à la fin d'un verbe à la troisième personne du pluriel à l'accompli, doit être supprimé. Exemple : [ذهبوا] sera transcrit en [#DahabU#].

-Règle du tachdid

Une consonne suivie du code du tachdid [*] est dupliquée car dans la nomenclature adoptée pour les polysons, une consonne géminée dans une séquence V_1C*V_2 se réalise par la concaténation de deux polysons V_1C et CV_2.

-Règle du tanwin

Les codes suivants [@](فتحتين), [&] (ضمتين) et [=] (كسرتين) dénotent un tanwin. Ils sont changés respectivement en ([a], [u], [i]) suivie de [n].

2.2.5. Les traitements inter-mots

La lecture du texte doit se faire avec un certain naturel. Il ne s'agit pas de lire le texte avec une pause à chaque fin de mot. Des transformations s'imposent sur un mot lorsqu'un mot est suivi d'un déterminant [ال] ou d'une hamza. Dans ces deux cas, il y a lieu de considérer la nature de la hamza, wasl (liaison) ou qat' (coupure). Dans le cas d'un wasl, nous supprimons la chaîne [?a] de la transcription [?al] de l'article défini [ال] du mot qui le suit. Par exemple [#Dahaba ~lwaladu#] sera converti en [#Dahabalwaldu#]. Ceci est réalisé en transformant d'abord la première chaîne en [#Dahaba ?alwaladu#] grâce à la

transcription et nous obtenons la seconde chaîne en exécutant la règle du wasl. Nous illustrons ceci par l'exemple suivant :

تَابَ بَعدَ إثمه ، ثَابَرَ فِي عَمَله فَنَجَحَ.

#tAbaba§da?içmihi#çAbarafI§amalihifanaGaHa##

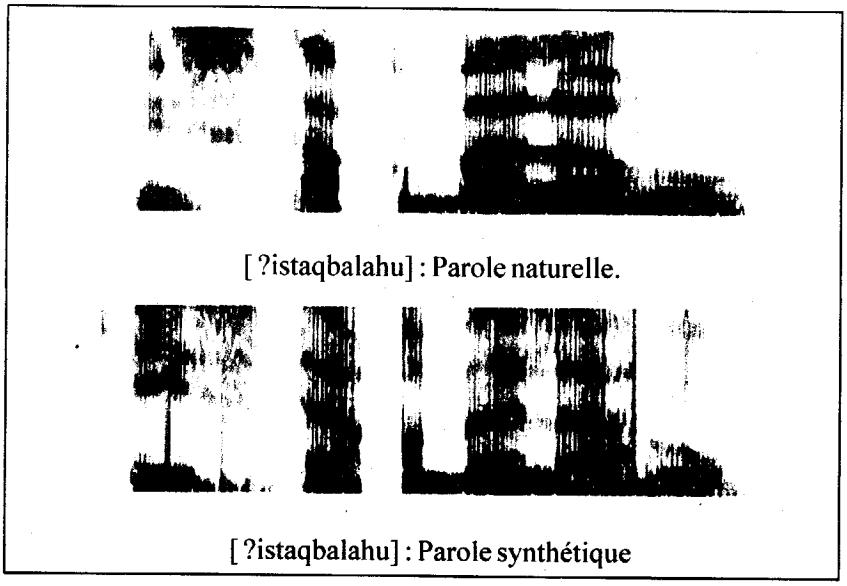
حَاكِمُ هذا البَلَد عَادِلٌ.

#HakimuhADalbaladi§Adilun##

2.3. Génération de l'onde acoustique

Les unités sonores composant le dictionnaire sont des segments de signal parole échantillonné à 10 kHz et codé en PCM (16 bits signés par échantillon).

La méthode de concaténation par polysons a été choisie de manière à réduire au maximum les discontinuités dans les contours formantiques au voisinage des points de concaténation. Il reste toutefois des discontinuités audibles rendant nécessaire l'élaboration de techniques de lissage appropriées. Nous avons deux approches concernant l'application du lissage au lecteur. Dans la première, le lissage s'applique à chaque unité durant la génération (lissage 'inline'). La seconde méthode utilise un dictionnaire de signaux préalablement lissés (lissage 'offline'). Dans ce cas, il faut alors fixer certaines valeurs moyennes pour la fréquence fondamentale et l'amplitude. De même, les durées des phonèmes suivant leurs contextes doivent être calculées. Nous devons alors réajuster tous les segments du dictionnaire suivant ces valeurs et le lissage se fait grâce aux techniques OLA et ses dérivées. L'exemple ci-dessous (fig. 4) montre la parole générée par le système ARPHON comparée à une parole naturelle.



[?istaqbalahu] : Parole naturelle.

[?istaqbalahu] : Parole synthétique

Fig. 4 : Parole générée par ARPHON comparée à une parole naturelle.

3. Conclusion

Le module de traitement de textes améliore notablement la lecture automatique de textes en permettant de dépasser le stade d'une lecture mot à mot. De même, le module de traitement du signal délivre une parole de bonne qualité. Si Dieu le veut, nous entamerons la réalisation d'un module de prosodie qui prendra en charge les caractères de ponctuation, la vitesse d'élocution, l'intonation et l'accentuation ainsi que des techniques de lissage pour atténuer toute discontinuité perceptible au voisinage des points de concaténation.

Bibliographie

[1] Calliope, *La parole et son traitement automatique*, Editeur J.P.Tubach, Masson, Paris, 1989.

[2] S. Lemmetty 'Review of Speech Synthesis'. MSc. Thesis. Laboratory of Acoustics and Audio Signal Processing, Helsinki. University of Technology, Finland, 1999.

[3] M. Guerti, 'Contribution à la synthèse de la parole par diphones en Arabe Standard.' Thèse de Magistère, Université d'Alger, 1983.

[4] A. Hamani, N. Mohallebi, 'Réalisation d'un lecteur automatique en Arabe Standard sous Word 97', Projet de fin d'études 1998, dirigé par A. Cherif-Zahar, K.Ferrat et K. Benbellil.

[5] Y. El-Imam, 'Speech Synthesis in Arabic Language', IEEE Transactions of Speech and Signal Processing, 1989.