Gabarits des tons vietnamiens

Pham Thi Ngoc Yen[1], Eric Castelli[1], Nguyen Quoc Cuong[2]

[1]Centre MICA - Institut Polytechnique de Hanoi – 1 Dai Co Viet - Hanoi – Vietnam
Tél. ++84 4 868 30 87 – Fax : ++ 84 4 869 53 19
Mél: ptnyen@vn.refer.org, Eric.Castelli@vn.refer.org
[2]Laboratoire CLIPS-IMAG, équipe GEOD, Univ. Joseph Fourier B.P. 53, 38041 Grenoble cedex 9, France
Tél.: ++ 33 (0)4 76 63 56 51 – Fax : ++33 (0)4 76 63 55 52
Mél: Quoc-Cuong.Nguyen@imag.fr

ABSTRACT

A 135 word corpus uttered by 16 different speakers was build in order to study the shape of the 6 Vietnamese tones. The wavelet method is used to extract the pitch (F0) from a speech signal corpus. General shapes are extracted for each speaker, which will be useful for automatic recognition or for synthesis, and comparisons between men and women show that we can consider no important difference between them. However, we have to separate North speakers form Centre/South speakers.

1 INTRODUCTION

La langue vietnamienne est un mélange d'éléments *mon*, *khmer*, *thaï* et *chinois*. Elle a emprunté un bon pourcentage de mots de base aux langues monotoniques *mon* et *khmer*. Des langues *thaï*, elle a adopté certains éléments de grammaire et leur tonalité. Enfin, bien qu'il soit reconnu que le vietnamien n'a pas ses origines en Chine, le *chinois* a donné au vietnamien l'essentiel de son vocabulaire philosophique, littéraire, technique et gouvernemental, ainsi que son mode d'écriture traditionnel, pendant la période de domination chinoise [Hau53] [Hau54].

Depuis ces 20 dernières années, peu d'études sur la langue vietnamienne ont été menées. Nous pouvons citer les études de l'acoustique vietnamienne de Doan [Doa77], des variations des tons vietnamiens de Han & Kim [Han74], et de la caractérisation des tons de Vu [Vu99]. C'est la raison pour laquelle aucun système de reconnaissance automatique de la parole en langue vietnamienne est disponible alors que plusieurs moteurs de reconnaissance du Mandarin [Yan88] ou du thaï [Tun98] ont été proposés dans la communauté scientifique. Le vietnamien est considéré comme une langue monosyllabique (ou bisyllabique) possédant six tons. Une syllabe peut être répétée avec chacun de ces six tons, ce qui lui donne alors six significations différentes. Néanmoins, cela ne signifie pas que toutes les combinaisons sont possibles : il existe des syllabes qui ne sont prononcées qu'avec quelques tons seulement, comme les syllabes fermées (syllabes qui se terminent par l'une des consonnes /p/ /t/ /k/) qui ne sont prononcées qu'en combinaison des ton5 et ton6. La langue vietnamienne présente 16 voyelles, 21 consonnes et 2 semi-voyelles . Chaque syllabe peut être considérée comme une combinaison de deux parties initiale et finale comme proposé par Doan [Doa77]. La consonne initiale, la partie prétonale et la consonne finale sont optionnelles, c'est-à-dire qu'elles peuvent ne pas exister. Les études de Doan [Doa77] et Han & Kim [Han74] ont montré que les informations concernant le ton sont essentiellement superposées sur la partie finale de chaque syllabe.

2. CORPUS ET MESURE DU CONTOUR DU PITCH

Nous avons réalisé un corpus en vietnamien avec 3 buts : 1) caractérisation des tons ; 2) étude de la reconnaissance des tons; 3) reconnaissance de la parole dans le mode de syllabes isolées pour les applications de commandes vocales de processus simples. Le corpus a été défini pour contenir des syllabes qui comprennent les 16 voyelles du vietnamien, assemblées avec les 6 tons, ainsi que les 21 consonnes initiales. La sélection des mots a été dictée par notre souhait d'utiliser le moteur de reconnaissance pour de la commande vocale sur Internet, avec des commandes du type : "fichier", "ouvrir", "fermer", etc... Nous avons choisi en définitive 135 mots parmi lesquels sont présents 131 mots monosyllabiques et 4 mots bi-syllabiques seulement. Chaque syllabe est prononcée 4 fois dans le mode *mots isolés* par 16 locuteurs différents du Nord (ville de Hanoi), du Centre (villes de Hue et Danang) et du Sud du Vietnam (villes de Ho Chi Minh City et Can Tho) : 5 femmes et 2 hommes du Nord, 2 femmes et 2 hommes du Centre, 3 femmes et 2 hommes du Sud. Nous avons enregistré un ensemble de (4*135*16) syllabes représentant un total de 7860 éléments pour environ 3 heures de parole. Le signal est enregistré dans un studio calme avec une fréquence d'échantillonnage de 16kHz et une résolution de 16 bits. La trame d'analyse est de 64 ms et la trame de décalage de 8 ms.

Pour mesurer la fréquence fondamentale (pitch), nous utilisons une méthode d'analyse par transformée par ondelettes D_yWT (Dyadic Wavelet Transform) proposée

Kadambe et Faye Boudreaux-Bartels [Kad92]. La D_yWT d'un signal x(t) est définie par:

$$D_yWT_x(b,2^j) = \frac{1}{2^j}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-b}{2^j}\right)dt$$

Où $\Psi^*(t)$ est la conjugaison complexe d'une fonction ondelette $\Psi(t)$, $a=2^j$ est le paramètre d'échelle et b est le paramètre de translation. L'algorithme se base sur les hypothèses :

- la période du pitch est estimée par la détermination de l'instant où la glotte est fermée (un événement), puis par la mesure de la durée entre deux événements ;
- les instants de fermeture de la glotte se retrouvent normalement aux points caractérisant des variations brusques ou des singularités du signal vocal ;
- si une fonction ondelette lissée est choisie, alors le maximum local de la fonction D_yWT indique la variation brusque du signal alors que le minimum local indique sa variation lente ;
- le maximum local de la fonction D_yWT sera alors utilisé pour détecter les changements brusques du signal vocal causés par la fermeture des cordes vocales.

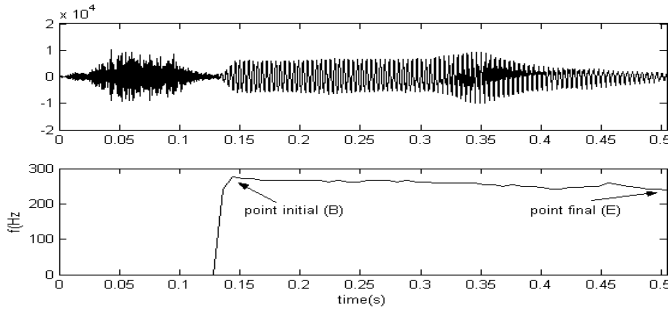

Figure 1: Exemple de ton1 (sujet féminin PNY)

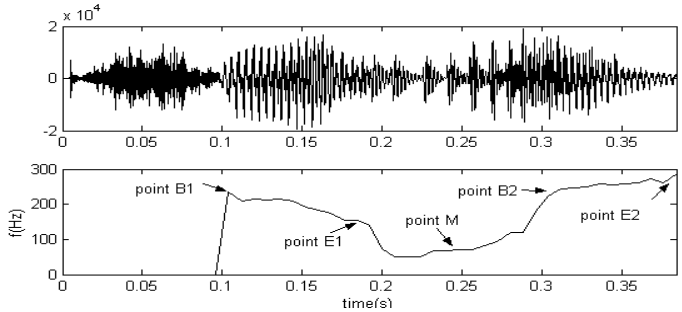

Figure 2: Exemple de ton3 (sujet féminin PNY)

Les figures 1 et 2 montrent deux exemples de contours des tons obtenus. Sur ces contours, afin de caractériser des gabarits aux formes simples, nous mesurons des points caractéristiques : deux points, point initial et point final pour les tons monotones (figure 1), auxquels nous rajoutons des points intermédiaires dans le cas des tons aux allures plus complexes (figure 2). B = begin, E = end et M = middle.

3. GABARITS

L'étude détaillée de nos résultats, nous montre que nous devons tenir compte d'une propriété particulière du vietnamien qui influe sur les variations des tons : il existe deux catégories de syllabes, les syllabes fermées qui se terminent par l'une des trois consonnes /p/, /t/, /k/ et qui ne peuvent se combiner qu'avec les ton5 et ton6, et les autres syllabes, dites syllabes ouvertes, qui quant à elles, peuvent se combiner avec l'ensemble des 6 tons. A partir de cette constatation, nous devons séparer les tons 5 et 6 en deux représentations en fonction des syllabes : ton5a et ton6a pour les syllabes ouvertes et ton5b et ton6b pour les syllabes fermées. Ce qui nous donne au total 8 représentations des six tons.

Les figures 3 et 4 donnent des exemples des 8 représentations des tons pour deux sujets du nord du Vietnam (vietnamien standard). Les courbes en trait plein représentent l'allure moyenne des tons calculée sur toutes les syllabes prononcées par le sujet alors que les courbes en traits pointillés encadrent les variations de ces allures.

Il est courant de regrouper les tons en deux registres : *registre haut* pour les ton1, ton3 et ton5 et *registre bas* pour les ton2, ton4 et ton6 ; la classification se fait sur deux critères : la valeur moyenne du ton et la valeur de son point terminal. Pour le sujet féminin PNY, la valeur moyenne des tons du registre haut est de 260 Hz, alors que la valeur moyenne du registre bas est de seulement de 190 Hz, soit une différence moyenne de 70 Hz.

La durée des tons (c'est-à-dire la durée des parties voisées des syllabes) n'est pas identique : les tons ton1, ton2, ton3, ton4 et ton5a sont longs avec une durée moyenne de 330 ms pour le sujet PNY, alors que les ton5b, ton6a et ton6b sont 3 fois plus courts (durée moyenne de 115 ms pour PNY).

Le contour du ton3 peut être brisé au milieu, c'est-à-dire que les cordes vocales cessent de vibrer dans la partie médiane du ton (F0 = 0Hz) ; cependant ce segment brisé n'est pas obligatoire et est dépendant du locuteur ; en général le ton3 comprend 2 segments : un segment descendant au début et un segment montant à la fin ; le point initial est plus bas que pour le ton1 mais le point final se termine à une fréquence plus élevée ;

3.1 Comparaison entre hommes et femmes

La fréquence moyenne du fondamental de la parole se situe autour de 200 Hz pour les femmes et de 100 Hz pour les hommes, comme dans toute langue. Cependant, il semble que sur l'ensemble de nos sujets féminins, la fréquence fondamentale moyenne évolue dans un espace fréquentiel plus grand que celui de l'ensemble des sujets masculins. Pour les tons, cela a pour conséquence d'augmenter l'espace entre le gabarit minimum et le gabarit maximum pour les femmes, ce qui introduit une plus grande variabilité: 120 Hz pour les femmes et 80 Hz

pour les hommes, les moyennes étant calculées tous sujets et tous tons confondus.
Les pentes moyennes des segments des gabarits sont souvent plus importantes pour les femmes que pour les hommes, ce qui a pour conséquence des différences entre les fréquences des débuts et fins de tons plus importantes.

Cependant, ces différences d'allure entre hommes et femmes restent faibles, d'un point de vue de la reconnaissance, et nous utilisons en entrée de notre système de reconnaissance à base de HMMs, un vecteur caractéristique défini pour les tons, identique pour les hommes et pour les femmes [Ngu01][Ngu02].

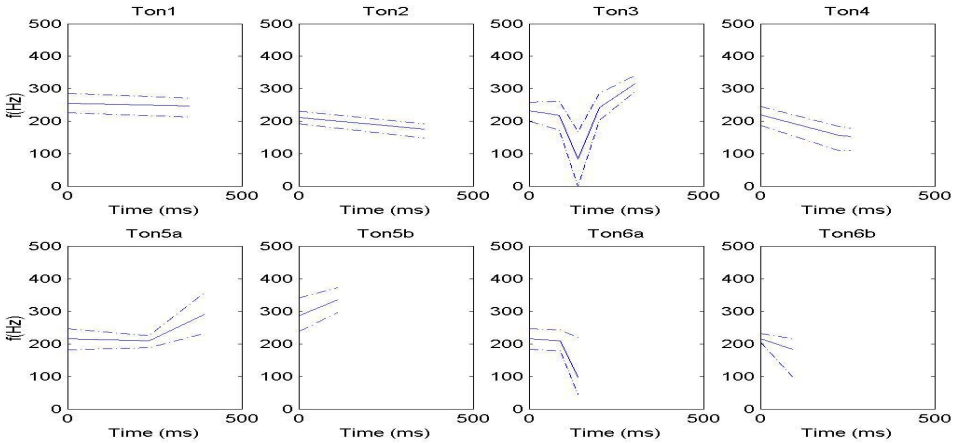

Figure 3 : Les contours des 8 représentations des tons du sujet féminin PNY (pointillés : + grandes variations)

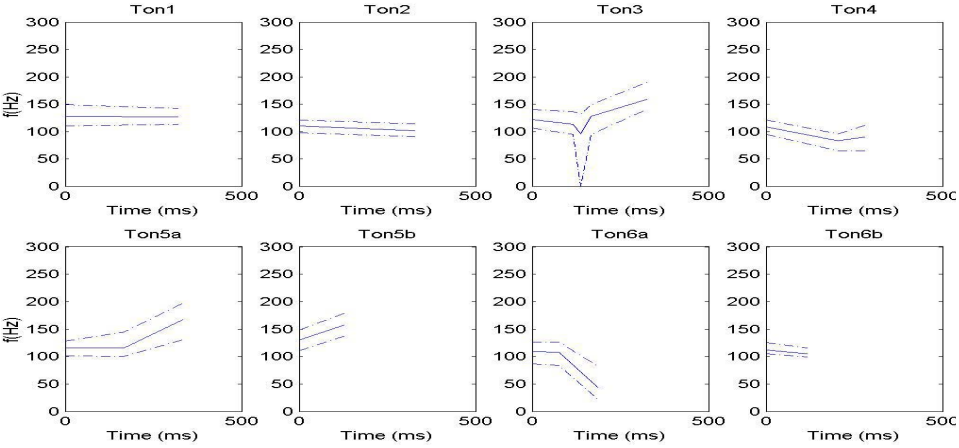

Figure 4: Les contours des 8 représentations des tons du sujet masculin BXH.

3.2 Comparaison entre les dialectes du Nord, du Centre et du Sud du Vietnam.

Il y a peu de différence entre les dialectes du Centre et du Sud du Vietnam. Par contre, des différences sensibles peuvent être remarquées entre ces deux derniers et le dialecte du Nord qui est considéré comme le vietnamien « standard » et officiel. Ces différences se retrouvent tant au niveau des consonnes initiales (la consonne /k/ du nord étant parfois remplacé au sud par son homologue sourde /g/ par exemple), des voyelles (prononcées parfois en diphtongues au nord mais en voyelles uniques au sud, par exemple), que des tons.

Pour les tons, les différences portent sur le ton4. Celui-ci présente une allure spécifique différente de celles des autres tons dans le cas des sujets d'origine nord du Vietnam, mais est prononcé d'une manière quasi identique au ton3 dans le cas des locuteurs du Sud/Centre. On pourrait donc affirmer qu'il y a 6 tons pour le Vietnamien officiel du nord mais seulement 5 tons pour le Vietnamien parlé dans le centre et le sud du pays. Cependant, l'alphabet est identique pour tout le pays, avec 6 notations des tons différentes, nous concluons donc, sans remettre en cause cette définition habituelle à « 6 tons», que pour le sud les tons3 et tons4 sont prononcés de la même manière.

Nous avons constaté aussi que le ton6a des syllabes fermées montre une grande variabilité pour les sujets du sud : certains locuteurs prononcent ce ton avec une allure qui leur est plus ou moins spécifique. Ceci pose un problème en terme de reconnaissance de ce ton, car il

apparaît difficile de le caractériser correctement par un seul vecteur.

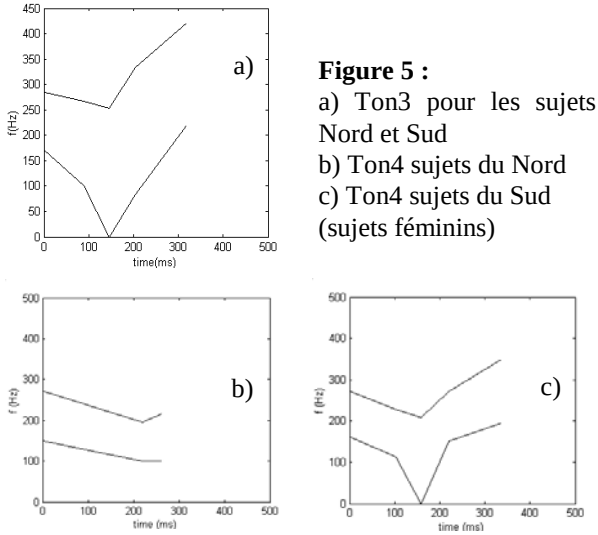

Figure 5 :
a) Ton3 pour les sujets Nord et Sud
b) Ton4 sujets du Nord
c) Ton4 sujets du Sud (sujets féminins)

3.3 Comparaison avec le Mandarin

Comme le vietnamien, le Mandarin, la langue officielle chinoise est une langue monosyllabique à tons. A la différence du vietnamien, le Mandarin n'a que quatre tons lexicaux et un ton neutre (figure 6). Le ton neutre existe seulement dans la syllabe terminale d'un mot multi-syllabique et il n'a pas un contour fixe. En première approximation, les tons du chinois ressemblent à quatre des tons du vietnamien. Cependant, des différences sont notables en terme de complexité et par rapport aux durées.

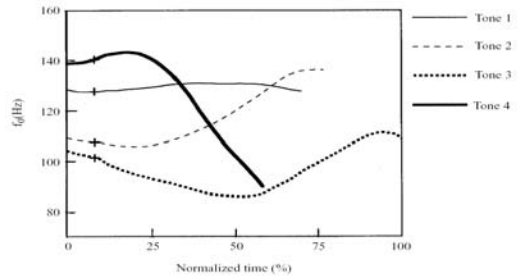

Figure 6 : Contours des quatre tons du Mandarin pour la syllabe [ma] (d'après [XU97])

4. CONCLUSION

Pour les 16 locuteur de notre corpus, nous montrons que les gabarits des 6 tons du vietnamien présentent des allures semblables. Les différences sont faibles entre les homme et les femmes mais un peu plus prononcées entre les locuteurs du Nord et ceux du Centre/Sud, en particulier pour les tons4 et tons6a. A partir ce ces résultats, nous avons défini un vecteur caractéristique spécialement adapté aux tons complexes du vietnamien Nous obtenons des taux de reconnaissance des tons vietnamiens supérieurs à 90 %. En couplant ce système de reconnaissance des tons à un système de reconnaissance acoustique classique basé sur les techniques d'utilisation des vecteurs cepstraux et des modèles de Markov cachés, nous avons réalisé le premier moteur de reconnaissance de la langue vietnamienne officielle du nord pour des applications de commandes vocales en mots isolés avec un taux de reconnaissance d'environ 94 % [Ngu01] [Ngu02]. A l'avenir nous envisageons d'utiliser ces résultats pour réaliser la synthèse du vietnamien grâce à des techniques Mbrola.

Cette étude a été réalisée dans le cadre du programme de coopération universitaire "Centre MICA" entre le laboratoire CLIPS-IMAG (INP Grenoble) et l'Institut Polytechnique de Hanoi.

BIBLIOGRAPHIE

[Hau53] Haudricourt A.G. (1953) "La place du vietnamien dans les langues austroasiatiques", Bulletin de la Société de Linguistique de Paris, 49, 1

[Hau54] Haudricourt A.G. (1954) "De l'origine des tons en vietnamiens", Journal Asiatique, 242,1

[Doa77] Doan T.T. (1977) "Ngu am tiêng viêt (Acoustique vietnamienne)", Nha Xuat Ban Editions.

[Han74] Han M.S. & Kim K.O. (1974) "Phonetic variation of Vietnamese tones in disyllabic utterances tones", Journal of Phonetics, vol. 2, pp 223-232

[Vu99] Vu B.H. (1999) "Ve dac trung co ban cua thanh dieu tiêng viêt o trang thai tinh (Caractérisation des tons vietnamiens dans le mode statique)", Journal of Linguistic Institute of Vietnam, Vol. 6, pp 34-53.

[Yan88] Yang W.J. et al (1988) "Hidden Markov Model for Mandarin Lexical Tone Recognition", IEEE Trans. ASSP, vol36, no 7, pp 988-992

[Tun98] Tungthangthum A. (1998) "Tone Recognition for Thai", Circuits and Systems, IEEE APCCAS, Asia-Pacific Conference, p. 157-160.

[Kad92] Kadambe S. & Boudreaux-Bartels G.F. (1992) "Application of the Wavelet Transform for Pitch Detection of Speech Signals", IEEE Trans. Information Theory, vol. 38, no 2.

[Ngu01] Nguyen Q.C., Pham T.N.Y. & Castelli E. (2001) " Shape Vector Characterization of Vietnamese Tones & Application to Automatic Recognition " ASRU 2001 Madonna di Campiglio, Cdrom.

[Ngu02] Nguyen Q.C. & Castelli E. (2002) "Caractérisation et Reconnaissance automatique des tons du vietnamien" RFIA2002, vol. 2, pp 529-537

[Xu97] XU Y. (1997) "Contextual tonal variations in Mandarin".Journal of Phonetics, vol 25, pp 61-83