

Registres des tons vietnamiens dans la parole continue

Do Dat TRAN, Eric CASTELLI

International Research Center MICA, IPH - CNRS/UMI-2954 – Grenoble INP - 1 Dai Co Viet, Hanoi, VIETNAM
(Do-Dat.Tran, Eric.Castelli)@mica.edu.vn

ABSTRACT

This paper describes the analysis results about register of Vietnamese tones in continuous speech. The obtained results show that the classification of the registers of Vietnamese tones into two groups (high and low) which is commonly used in isolated words is not appropriate in continuous speech. The register of tone has to be considered in a specific context because of the influence of tonal coarticulation effect.

Keywords: vietnamien, ton, registre, registre relatif.

1. INTRODUCTION

Récemment, la langue vietnamienne a été le sujet de certaines recherches linguistiques, mais la plupart de ces études se focalisent sur l'analyse des caractéristiques en mode de mot isolé [1][2] ou sur certaines des caractéristiques mélodiques se fondant sur des données limitées de la parole [3][4]. En outre, il y a eu peu de recherche sur les variations de ton et sur les registres des tons dans la parole continue, particulièrement concernant les facteurs qui sont liés au contrôle du contour de F0 pour la synthèse de la parole.

Dans les langues tonales, les tons peuvent être divisés en deux groupes, haut ou bas, selon leur registre (quelques fois en trois groupes : haut, milieu, et bas) [1], [5]. Quant à la langue vietnamienne, d'après [1], les ton1, ton3, ton5 sont de type « registre haut » et les ton2, ton4 et ton6 sont de type « registre bas ». Néanmoins, lorsqu'on met en œuvre ce résultat dans la chaîne parlée continue, cela ne fonctionne pas vraiment correctement.

Dans notre étude, un corpus de parole continue a été tout d'abord construit, et puis nous avons analysé les registres des tons dans ce corpus. Ce processus et les résultats obtenus sont présentés dans les sections ci-dessous.

2. CORPUS

Dans notre travail, un corpus de textes de 33 paragraphes et 114 dialogues a été collecté à partir de différentes ressources telles que des histoires, des livres et des documents Web.

Les dialogues sont des conversations courtes contenant de 3 à 6 phrases, sur différents sujets de la vie quotidienne. Les paragraphes se composent d'environ 70 mots. Ces mots ont été extraits à partir de plus grands paragraphes dans des domaines divers, comme par exemple : la culture, le sport, la science, l'économie, la météo,

l'éducation, la santé, etc. (afin d'avoir un grand nombre de mots avec des contextes différents). Les principales caractéristiques des deux types de données sont présentées dans les tableaux 2-7 et 2-8. En combinant les deux types de corpus, un corpus total de textes qui se compose de 630 phrases correspondant à 9289 syllabes (avec 1597 syllabes différentes) a été ainsi construit.

Tableau 2-1 : Deux types de données et leurs caractéristiques

Type de données	Nombre de phrases	Nombre de syllabes	Nombre de syllabes différentes
Dialogue	212	1483	457
Paragraphe	418	7806	1513

Tableau 2-2 : Nombre des tons dans le corpus

Ton 1	Ton 2	Ton 3	Ton 4	Ton 5a	Ton 5b	Ton 6a	Ton 6b
2598	1795	453	845	1402	721	851	624

Une animatrice professionnelle de la télévision de Hanoi a été requise pour lire ce corpus textuel avec un débit normal. Le processus d'enregistrement a été réalisé dans le studio calme du Centre MICA à la fréquence d'échantillonnage de 16 kHz. Un corpus de 38 minutes de signal de la parole a été ainsi construit, correspondant à 31 minutes de corpus de paragraphes et 7 minutes de corpus de dialogues.

Le processus d'annotation comprend deux étapes :

- première étape : tous les signaux de parole sont automatiquement étiquetés au niveau de la syllabe par le système RAP en langue vietnamienne développé au laboratoire GETALP-LIG. L'information exportée d'une syllabe contient le début et la fin du segment de signal de la syllabe ;
- deuxième étape : les fichiers étiquetés sont corrigés manuellement en employant le logiciel Praat¹.

3. REGISTRES DES TONS DANS LA PAROLE CONTINUE

Pour la langue vietnamienne, [1] a utilisé la hauteur du point final du contour de F0 du ton comme critère de classification des registres. A la différence de [1], [5] ont,

¹ Praat <http://www.fon.hum.uva.nl/praat/>

quant à eux, utilisé la hauteur globale (la hauteur moyenne) du ton comme critère de classification des registres en langue cantonaise. La classification de [1] est fondée sur l'impression auditive et cette classification semble bien fonctionner pour les 6 tons isolés. Néanmoins, lorsqu'on met en œuvre ce résultat dans la chaîne parlée continue, en synthèse en particulier, l'intelligibilité et le naturel de la parole produite n'ont pas un niveau acceptable.

En étudiant les contours de F0 des phrases dans le corpus, nous constatons que la forme de la courbe mélodique dépend considérablement des tons utilisés et les tons varient considérablement selon leur environnement tonal immédiat. Nous allons observer les contours de F0 de deux énoncés (figures 3.1 et 3.2).

- Phrase 1 : « Chưa bao giờ khu vực này lại tấp nập đến như vậy » - /cư_x_1 baw_1 zư_2 Xu_1 vuok_6b naj_2 lai_6 tấp_5b nấp_6b den_5b juu_1 vư_j_6/ ; (Cet endroit est plus trépidant que jamais).
- Phrase 2 : « Loại người nhan nhản vẫn lượn lờ ở các chốn đô hội » - /lwai_6 ɲư_xj_2 nan_1 nan_4 vữ_n_3 lư_xn_6 lư_2 ư_4 kak_5b con_5 do_1 hoj_6/ ; (Ce type de gens reste omniprésent dans les (grandes) villes).

Dans les figures, les contours de F0 des syllabes « nập », « như » (dans la phrase 1) et « nhan », « lượn », « chốn » (dans la phrase 2) sont beaucoup influencés par les contours de F0 de leur syllabe précédente. La partie initiale du contour de F0 de ces syllabes est soit haute soit basse, et elle est dépendante du point de terminaison et de la direction du contour de F0 de la syllabe précédente.

En ce qui concerne les registres des tons, nous supposons qu'il y a deux niveaux de registre dans les figures 3.1 et 3.2 : le niveau haut correspond au contour du ton 1, et le niveau bas correspond au contour du ton 2. Le ton dont le registre est haut / bas est présenté par le symbole H / L. Nous pouvons alors remarquer que la différence entre les tons peut se décliner, non seulement en termes de différence de forme, mais encore au niveau du registre du ton. Cependant, il y a des endroits où la classification de [1] n'est pas vraiment correcte. Selon [1] les contours de F0 des syllabes du ton 5a (« đên » /den-5/ et « chốn » /con-5/) sont associés à un registre haut (H), mais sur les figures ils se situent dans la région basse de la gamme de fréquence de F0. Dans un autre cas, le contour de la syllabe du ton 6a « lượn » (/lư_xn-6/) se situe dans la région haute de la deuxième phrase, alors que dans la classification de [1] le ton 6a est associé à un registre bas (L).

Dans la littérature, on divise les tons lexicaux principalement selon leur hauteur et leur forme du contour de F0. Par exemple, en mandarin, quatre tons H (ton1), R (ton 2), L (ton 3) et F (ton 4) ont respectivement des contours de F0 de niveau haut, milieu-montant (mid-rising), bas-immérgant (low-dipping) et haut-descendant (high-falling). Ces tons s'appellent les tons de contours

(contour tones) [6]. Pour quelques langues africaines leurs tons sont connus comme les tons de registre (register tones) car ils ont une spécification de F0 simple, tels que H (high), M (mid) ou L (low). En ce qui concerne la langue vietnamienne, les tons lexicaux devraient être considérés comme les tons de contour. La classification des registres des tons en fonction de la hauteur du point final n'illustre qu'une caractéristique locale des tons, elle ne montre pas la caractéristique globale du contour de F0 des tons.

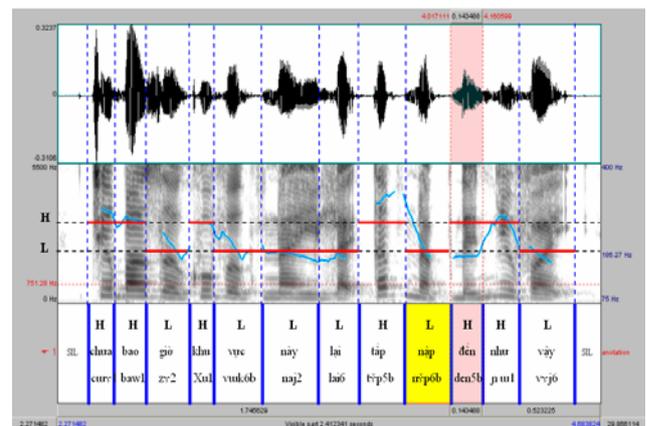


Figure 3.1 : Contour de F0 et transcription de la phrase 1

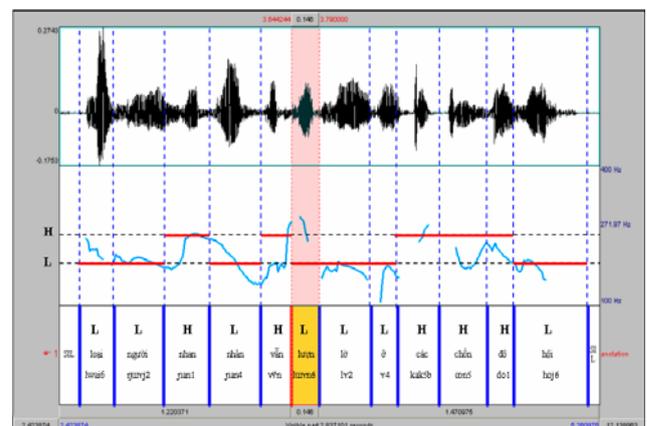


Figure 3.2 : Contour de F0 et transcription de la phrase 2

D'ailleurs, il faut considérer l'influence du phénomène de coarticulation dans la parole continue. Pour le ton 5a, peut-être quelqu'un expliquera que la partie basse du contour de F0 de la syllabe /den-5a/ est causée par l'effet de la coarticulation progressive du ton 6 précédent. Pourtant, pour la syllabe /con-5a/ portant également le ton 5a, le ton précédent (ton 5b qui a un contour montant dans la région de fréquence haute) élève les points initiaux de ce ton 5a mais son contour de F0 reste dans la région basse. Donc, on peut constater que le ton 5a a tendance à retourner à la région de F0 basse avant de s'élever. Quant à la syllabe « lượn » /lư_xn-6/, sa position

élevée est provoquée principalement par l'influence de l'augmentation du ton 3 précédent.

Tout ceci nous porte à croire que la classification de [1] ne s'adapte pas à la parole continue en langue vietnamienne. Ainsi il est nécessaire d'avoir une autre conception pour décrire entièrement les caractéristiques du registre des tons vietnamiens dans la chaîne parlée.

4. REGISTRE RELATIF

En ce qui nous concerne, nous appliquons la conception de [5], c'est-à-dire que la hauteur globale est considérée comme le registre du ton. Afin de réduire l'influence des tons environnants, nous avons analysé le registre relatif de deux tons adjacents, comme cela a été présenté dans les travaux de [5]. Pour une paire de tons voisins (i, j), où i et j dénotent respectivement les identités des tons courants et précédents, le registre relatif est le rapport de registre de cette paire qui est calculé comme suit:

$$R_{(i,j)} = \frac{R_i}{R_j} \quad (4.1)$$

où R_i est le registre (ou la hauteur) du ton i, et R_j est le registre du ton j (ton gauche).

Dans notre étude, le registre du ton est la valeur F0 moyenne. Puisque les points initiaux et finaux ne sont pas stables, ils n'ont pas été employés pour calculer la hauteur des tons. Le rapport de ton global du ton i au ton j ($R_{i,j}$) est obtenu en calculant la valeur moyenne de toutes les occurrences de (i, j) dans notre corpus. En conséquence, avec les six tons et les deux variantes des tons 5 et 6 (ton5b et ton6b), une matrice 8x8 des rapports de registres a été obtenue, et est décrite dans le tableau 4.1.

A partir du tableau 4.1, nous pouvons constater que la différence des registres entre les 6 tons est significative, voir même importante dans certains cas : la différence entre le registre du ton 5b et celui des autres tons est d'au minimum 24%. Il est aisé de constater que le ton 5b présente le registre le plus élevé, puis vient en seconde position le ton 1.

Tableau 4.1 : Matrice des rapports de registres entre deux tons adjacents de toutes les combinaisons de deux tons

Ton i Ton j	1	2	3	4	5a	6a	5b	6b
1	0.99	0.84	0.91	0.82	0.90	0.88	1.24	0.84
2	1.13	0.92	0.96	0.91	0.99	0.93	1.38	0.92
3	1.21	1.05	1.04	1.07	1.06	1.11	1.48	1.09
4	1.13	0.92	0.94	0.90	0.98	0.92	1.39	0.93
5a	1.19	1.08	1.12	1.05	1.07	1.18	1.50	1.18
6a	1.11	0.92	0.90	0.90	0.95	0.92	1.36	0.93
5b	0.84	0.73	0.77	0.68	0.78	0.80	1.10	0.70
6b	1.14	0.91	0.96	0.94	0.98	0.95	1.45	0.94

Tableau 4.2 : Matrice de l'écart-type relatif (en %) de toutes les combinaisons de deux tons

Tone i Tone j	1	2	3	4	5a	6a	5b	6b
1	6	11	9	8	9	10	10	8
2	10	9	13	10	9	11	8	9
3	10	13	11	11	14	13	14	12
4	11	10	10	11	10	10	13	10
5a	10	13	9	14	12	10	14	11
6a	13	12	14	13	11	12	9	14
5b	10	12	9	12	13	13	14	11
6b	11	12	13	11	11	10	9	11

De plus, cette différence dépend de la position du ton dans une combinaison de 2 tons (ton – ton), comme nous allons le montrer avec deux exemples :

- pour la paire comprenant le ton 2 et le ton 4 (figure 4.1), si le ton 4 précède le ton 2 (figure 4.1a) alors le registre du ton 4 est plus haut que celui du ton 2 ($R_{2,4} = 0.92$) ; à contrario, si le ton 4 suit le ton 2 (figure 4.1b) alors son registre est inférieur à celui du ton 2 ($R_{4,2} = 0.91$) ;
- avec une autre paire (ton 2 et ton 5a), quand le ton 2 est avant le ton 5 (figure 4.2a) alors leurs registres sont équivalents ($R_{5a,2} = 0.99$) mais pour un ordre inverse des deux tons (figure 4.2b), le registre du ton 2 est plus haut que celui du 5a ($R_{2,5a} = 1.08$).

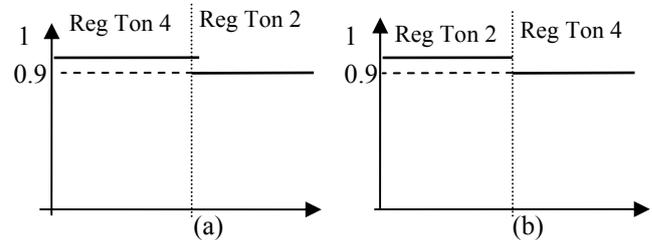


Figure 4.1 : Registre relatif entre le ton 2 et le ton 4

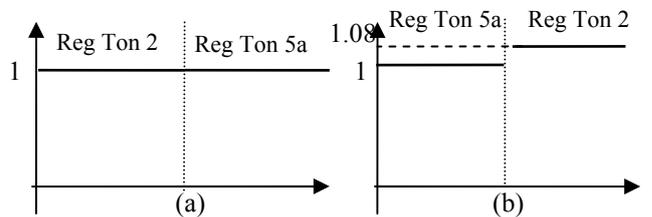


Figure 4.2 : Registre relatif entre le ton 2 et le ton 5a

Afin d'observer plus clairement la différence du registre des tons par rapport à leurs positions relatives dans les paires de tons, nous utilisons l'équation 4.2. Cette équation nous permet de distinguer la différence du registre d'un ton quand il est placé avant ou après un autre ton différent.

$$D = |R_{i,j} - 1/R_{j,i}| \quad (4.2)$$

Tableau 4.3 : Différence du registre des tons par rapport à leur position

<i>Ton j</i>	<i>Ton i</i>	$R_{i,j}$	$R_{j,i}$	<i>D</i>
1	2	0.84	1.13	0.04
1	3	0.91	1.21	0.08
1	4	0.82	1.13	0.07
1	5a	0.90	1.19	0.06
1	6a	0.88	1.11	0.03
1	5b	1.24	0.84	0.05
1	6b	0.84	1.14	0.04
2	3	0.96	1.05	0.00
2	4	0.91	0.92	0.18
2	5a	0.99	1.08	0.07
2	6a	0.93	0.92	0.16
2	5b	1.38	0.73	0.00
2	6b	0.92	0.91	0.18
3	4	1.07	0.94	0.00
3	5	1.06	1.12	0.16
3	6	1.11	0.90	0.01
3	5b	1.48	0.77	0.18
3	6b	1.09	0.96	0.05
4	5	0.98	1.05	0.02
4	6	0.92	0.90	0.20
4	5b	1.39	0.68	0.08
4	6b	0.93	0.94	0.13
5	6a	1.18	0.95	0.13
5	5b	1.50	0.78	0.21
5	6b	1.18	0.98	0.16
6	5a	1.36	0.80	0.10
6	6b	0.93	0.95	0.12
5b	6b	0.70	1.45	0.01

Les résultats dans la colonne D du tableau 4.3 nous montrent qu'il y a certaines paires pour lesquelles la différence des registres de deux tons en fonction de leur position relative (ton2-ton3, ton3-ton4, ton3-ton6a...) est très faible. Cependant, pour certaines paires ton2-ton4, ton2-ton6b, ton5a-ton5b...), la différence est significative (environ 16%, correspondant à 3 demi-ton). Ces résultats semblent nouveaux et sont différents des résultats de ceux présentés dans les travaux de [5] en langue cantonaise où les rapports de registre des paires de tons ne changent pas en fonction de leur position relative.

De plus, l'écart type relatif (de 6% à 14%) dans le Tableau prouve que les rapports de registre sont tout à fait stables. En conséquence, similairement à [5], dans la parole continue en vietnamien, bien que le niveau absolu de F0 des tons puisse changer sensiblement dans des contextes différents, leurs hauteurs relatives par rapport à un autre ton précédent ou suivant restent en grande partie invariables. Ceci semble important pour une bonne perception des tons. Cependant, parce que les tons sont affectés par leurs tons environnants, leur registre sera changé par l'effet de coarticulation.

Grâce à ces résultats, nous avons établi des relations entre ces facteurs en proposant une nouvelle méthode pour produire des contours de F0 (présentée plus en détail dans [7]). Le contour du registre tonal est tout d'abord produit en le calculant à partir des rapports de registres relatifs entre deux tons adjacents. Puis les modèles de tons lui sont superposés. Cette méthode permet à notre système de synthèse de générer des contours de fréquence fondamentale avec un coefficient de corrélation moyen de 0,71 par rapport aux contours originaux. Afin d'évaluer la performance de notre modèle de génération du contour de F0 qui se compose de l'influence des registres relatifs et de l'effet de coarticulation tonale sur la caractéristique naturelle de la phrase synthétique, un test de perception basé sur le test MOS a été réalisé. Les résultats prouvent que pour une langue tonale comme le vietnamien, le registre tonal relatif est un paramètre important pour la génération du contour de F0 [7].

5. CONCLUSIONS

Nous en concluons que la classification du registre des tons en deux groupes (haut et bas) comme habituellement utilisée en mode de mots isolés n'est pas appropriée aux tons en mode dynamique de la parole continue. Le rapport de registre entre deux tons doit être calculé dans des contextes spécifiques à cause de l'effet du phénomène de coarticulation tonale. Ces résultats sont utilisés pour nous permettre d'établir des relations entre les facteurs qui exercent une influence sur les paramètres prosodiques du vietnamien.

BIBLIOGRAPHIE

- [1] Nguyen, Q.C, "Reconnaissance de la parole en langue Vietnamienne", *PhD. thesis INP-Grenoble*, France, June 2002.
- [2] Tran D.D, Castelli E., Serignat JF., Le X.H., Trinh V.L., "Influence of F0 on Vietnamese syllable perception", *Proc. of Interspeech2005, Lisbon*, pp. 1697-1700, 2005.
- [3] Nguyễn T.T.H. et Boulakia G., "Another look at Vietnamese intonation", *ICPhS 1999 San Francisco*, pp. 2399 – 2402, 1999.
- [4] Marc B. "Coarticulation effects in Northern Vietnamese tones", *Proc. ICPhS 2003, Barcelona*, pp. 2673 – 2676, 2003.
- [5] Li Y. , Lee T., and Qian Y., "Analysis and modeling of F0 contours for Cantonese text-to-speech", in *ACM Transactions on Asian Language Information Processing*, Volume 3, No 3, pp 169-180, 2004.
- [6] Xu Y., "Effect of tone and focus on the formation and alignment of F0 contours". *Journal of Phonetics* 27, 55-105.
- [7] Tran D. D., "Synthèse de la parole à partir du texte en langue vietnamienne", *PhD. Thesis INP-Grenoble*, France, December 2007.