



# Étude acoustique des segments dans les séquences consonantiques du vietnamien

Thi Thuy Hien Tran, Nathalie Vallée

## ► To cite this version:

Thi Thuy Hien Tran, Nathalie Vallée. Étude acoustique des segments dans les séquences consonantiques du vietnamien. VIIème Rencontre des Jeunes Chercheurs en Parole, Paris, Jul 2007, Paris, France. pp.140-143, 2007. <hal-00335155>

**HAL Id: hal-00335155**

**<https://hal.archives-ouvertes.fr/hal-00335155>**

Submitted on 28 Oct 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Étude acoustique des segments dans les séquences consonantiques du vietnamien

Thi Thuy Hien TRAN, Nathalie VALLÉE

Département Parole et Cognition de GIPSA-lab  
1180, avenue Centrale – BP 25 – 38040 Grenoble, Cedex 9  
Tél. : ++33 (0)4 76 82 41 20 - Fax : ++33 (0)4 76 82 43 35  
Courriel : thi-thuy-hien.tran@gipsa-lab.inpg.fr

## ABSTRACT

In this paper, we study the acoustic realization of Vietnamese final stops (/p/ /t/ /k/) and nasals (/m/ /n/), often unreleased, in the consonant sequences, starting from the corpus of 140 monosyllabic and dissyllabic items. This is the first stage of discovering the responsible factors explaining for the difficulty of producing the French consonant clusters faced by Vietnamese learner.

## 1. INTRODUCTION

En situation d'apprentissage, un sujet confronté au système phonético-phonologique d'une autre langue éprouve souvent des difficultés avec les unités sonores de cette langue qui n'existent pas dans sa langue maternelle. Il est très difficile, au début, d'entendre correctement les sons de cette langue nouvelle. Or qui n'entend pas, a du mal à produire ces unités sonores mal perçues [Bor92].

Parmi les difficultés de sujets vietnamiens apprenant le français figurent les groupes de consonnes du français qui n'existent pas en vietnamien, langue dont la structure syllabique ne comporte pas de séquences de consonnes. Ceux-ci sont réalisés déformés par rapport à la cible et cette difficulté résistante perdure même après plusieurs années de pratique [Ngu00]. Ils sont souvent remplacés par une des unités qui existent dans le système de la langue maternelle en omettant une consonne du groupe (ex. « psychologie » => [sikoloji]). Ils sont séparés par un son vocalique ajouté entre les consonnes du groupe pour faciliter leur articulation (ex. « tableau » => [ta-bə-lo]). On les rencontre aussi prononcés avec la structure syllabique du vietnamien (ex. « cyclo » adapté « xich lô » [sik-lo] (par soucis de complicité, les tons ne sont pas indiqués) (/k/ non relâché en coda [k<sup>h</sup>], /l/ attaque de syllabe affectée de la tonalité correspondante de la langue de départ)).

En quoi consistent exactement les différences entre ces deux réalisations et quelles sont leurs implications dans l'acquisition des percepts phonétiques ? L'objectif général de ce travail est de comprendre les facteurs responsables de la difficulté à réaliser les clusters consonantiques du français par des apprenants vietnamiens.

Parmi les séquences consonantiques possibles dans les deux langues figurent les séquences /plosive + consonne/, /nasale + consonne/ (plosive et nasale occupant une position de coda). Cependant une des particularités du

vietnamien est que les consonnes plosives /p t k m n ŋ/, quelles que soient les conditions de leur réalisation, ne sont pas relâchées en coda de syllabe, la tenue de l'occlusion n'étant pas suivie d'un bruit caractéristique d'explosion rapide et audible [Doa99]. Cette particularité dans la production des occlusives finales du vietnamien n'est pas contrastive, elle ne change en aucun cas le sens du mot.

Dans le cas du français, les consonnes occlusives finales de mot (plosives et nasales /p t k m n ŋ ɲ b d g/), accompagnées ou non d'une vibration des cordes vocales (voisement), sont réalisées par une obstruction totale (occlusion) en un point du conduit vocal, maintenue pendant un certain laps de temps, puis généralement suivie d'un relâchement audible (le bruit d'explosion).

Les deux langues connaissent une autre dissemblance : elles possèdent des patrons syllabiques et des gabarits lexicaux très différents. Dans le cas de la syllabe, Michaud [Mic04] représente les patrons syllabiques du vietnamien par (C<sub>1</sub>)(w)V(C<sub>2</sub>), les parenthèses indiquant les constituants optionnels. La diversité des patrons syllabiques est beaucoup plus présente en français : (C<sub>1</sub>)(C<sub>2</sub>)(C<sub>3</sub>)V(C<sub>4</sub>)(C<sub>5</sub>)(C<sub>6</sub>)(C<sub>7</sub>) selon Rousset [Rou04]. Cependant, 96 % des structures syllabiques du français entrent dans le patron (C<sub>1</sub>)(C<sub>2</sub>)V(C<sub>3</sub>) qui correspond au gabarit lexical le plus rencontré en vietnamien. Mais langue isolante, le lexique du vietnamien ne connaît les séquences de consonnes que dans les mots composés. Ces suites de consonnes en vietnamien ne se réalisent pas comme des clusters du français, car elles ne sont pas produites avec enchaînement du point de vue articulatoire ([Tru70], [Doa99], [Ngu00]).

Dans cette problématique, nous menons une étude permettant de mieux comprendre les différences de réalisation des plosives et nasales du vietnamien en fonction de leur distribution en considérant deux contextes possibles pour la coda :

- en finale de mot : C<sub>1</sub>VC<sub>2</sub>
- en finale constituant 1 de mot composé : C<sub>1</sub>VC<sub>2</sub>-C<sub>3</sub>VC<sub>4</sub>

C'est une première étape dans la compréhension des relations entre perception et production des consonnes finales pour des Vietnamiens confrontés à l'apprentissage d'une langue étrangère comme le français.

## 2. MÉTHODOLOGIE

### 2.1. Corpus et locuteurs

L'étude porte sur 35 items monosyllabiques et dissyllabiques comportant l'une des 5 occlusives finales /p t k m n/ en contexte de la voyelle la plus ouverte /a/, de structures CV, CVC et CVC.CVC et de même contexte tonal. Concrètement, dans cet ensemble d'items lexicaux, les cinq consonnes se trouvent :

- En position initiale  $C_1$  de mots monosyllabiques de structure ouverte  $[C_1a]$  ou fermée  $[C_1aC_2]$  (ex : *ta, ka, ma, na, tak, kat, mat, mak, nat ...*)
- En position finale  $C_2$  des mots monosyllabiques  $[C_1aC_2]$  (ex : *tak, kat, mat, mak, nat ...*)
- En position finale  $C_2$  de la première syllabe  $[C_1aC_2]$  de mots composés lexicaux dissyllabiques  $[C_1aC_2.C_3VC_4]$  (Ex : *ap.C<sub>3</sub> ; at.C<sub>3</sub> ; ak.C<sub>3</sub> ; am.C<sub>3</sub> ; an.C<sub>3</sub>*)

Ces items ont ensuite été insérés dans une phrase porteuse construite de manière à ce que l'environnement des phonèmes à étudier soit le plus neutre possible du point de vue articulatoire pour les phonèmes qui entourent les 35 mots sélectionnés : « *Bạn sẽ gặp từ ... xuất hiện trong bài khoá* » (Tu vas rencontrer le mot ... apparaître dans le texte) [ban sẽ gặp từ ... swɔ̃t hien cɔ̃j baj ɣwa]

Le corpus est constitué de 4 répétitions des 35 mots insérés dans la phrase porteuse et mises en ordre aléatoire. Les 140 phrases du corpus ont été lues à voix haute par un sujet féminin, âgé de 26 ans, originaire du Nord du Vietnam.

L'enregistrement s'est déroulé dans la chambre sourde du Département Parole et Cognition de GIPSA-lab (Grenoble, France) avec l'enregistreur numérique Marantz PMD 670, un micro AKG C1000S à directivité cardioïde. Le corpus a été numérisé à 44.1 kHz sous format .wav.

### 2.2. Traitement de données

Les données ont été traitées avec le logiciel PRAAT pour mesurer les paramètres acoustiques des consonnes concernées dans le corpus de 140 mots. Les paramètres mesurés sont :

- durée totale de la consonne,
- durée de la voyelle /a/ suivie ou précédée des consonnes  $C_1, C_2, C_2.C_3$ ,
- durée du VOT,
- durée de l'occlusion,
- durée et amplitude du bruit de l'explosion,
- transitions de  $F_0, F_1, F_2$  et intensité des  $C_2$  et  $C_2.C_3$  pendant les 3 dernières périodes avant occlusion.

Le choix des transitions VC pour les codas a été fait en raison de la réalisation fréquente de plosives non relâchées dans cette position. Les caractéristiques des consonnes en coda sont identifiées « grâce aux incurvations de formants vocaliques qui se trouvent dans le segment attribué à la voyelle précédente » [Cao85].

### 2.3. Alignement et mesures

La durée totale des plosives en  $C_1$  a été mesurée en prenant l'intervalle de temps entre la dernière strie périodique de la voyelle finale /tu/ du mot [ttu] alignée avec la fin du tracé formantique visible sur le spectrogramme jusqu'à la première strie de périodicité régulière de voyelle [a] qui suit  $C_1$ .

En position finale  $C_2$ , les curseurs ont été positionnés aux endroits qui suivent directement la descente de dernière crête périodique de la voyelle [a] jusqu'au début du segment fricatif de la consonne /s/ du mot [swɔ̃t] qui suit.

La durée de l'occlusion a été définie dans l'intervalle temporel entre le début de l'occlusion et le début du relâchement repéré sur la courbe d'intensité. En position finale, le début de l'occlusion a été déterminée à la dernière crête périodique de la voyelle /a/ précédente.

Les mesures de VOT et de durée du bruit d'explosion des plosives ont été prises sur des spectrogrammes à larges bandes. Le VOT des  $C_1$  a été mesuré en prenant l'intervalle de temps entre le début du bruit de détente et la première strie de périodicité régulière de la voyelle suivie. Pour les consonnes en position finale  $C_2$  devant  $C_3$ , le VOT a été mesuré entre le début du burst lorsque celui-ci est présent et le début de la vibration des cordes vocales des consonnes  $C_3$  lorsque celles-ci sont voisées /m n z/.

Les mesures d'intensité du bruit d'explosion sont données en dB relatifs. Les moyennes sont basées sur les 4 répétitions. Les mesures d'intensité ne peuvent pas être prises dans les cas où le burst est absent (plosives non relâchées).

Les pentes de  $F_0, F_1, F_2$  et intensité pendant les 3 dernières périodes n'ont été calculées que pour des consonnes en coda ( $C_2$  et  $C_2.C_3$ ) car elles sont non relâchées [p<sup>h</sup>], ce qui n'est pas le cas des initiales. Les pentes de  $F_0, F_1, F_2$  et intensité pendant les 3 dernières périodes ont été obtenues en calculant la différence de valeur de  $F_0, F_1, F_2$  et intensité au temps  $T_2$  (fin de l'intervalle) et leur valeur au temps  $T_1$  (début de l'intervalle) divisée par la durée de l'intervalle ( $T_2 - T_1$ ).

### 2.4. Analyse statistique

Les données des mesures acoustiques ont été ensuite analysées en utilisant le logiciel SPSS (*Statistical Package for the Social Sciences*). Des tests d'analyses de variance (ANOVA) ont été effectués.

- L'effet de la position dans la syllabe ( $C_1, C_2, C_2.C_3$ ) sur les paramètres acoustiques des consonnes plosives (durée totale du segment, durée de l'occlusion, durée du burst, intensité du burst).
- La recherche d'effet d'interaction entre la position de la consonne  $C_2$  ou  $C_2.C_3$  et le lieu d'articulation de la consonne (labial, coronal, vélaire) et les paramètres acoustiques des consonnes en position finale : évolution de l'intensité, de la fréquence fondamentale et des deux

derniers formants de la voyelle mesurés juste avant la closure. Ces paramètres ont été sélectionnés en raison du caractère non relâché de 19,42 % des plosives réalisées dans cette position. Ce test concerne uniquement les consonnes en coda  $C_2$  et  $C_2.C_3$ .

### 3. RÉSULTATS

Les analyses montrent que les consonnes sont plus longues en position initiale qu'en position finale d'un item lexical que celui-ci soit monosyllabique ou dissyllabique. Les labiales atteignent une durée plus longue que les coronales et vélaire en toutes positions. Les plosives sont plus brèves que les nasales en attaque et coda de mots. Les plosives ont une phase de fermeture plus longue en initiale ( $C_1$ ) qu'en finale ( $C_2, C_2.C_3$ ).

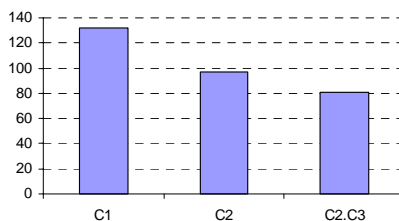


Figure 1: Durée moyenne des plosives (ms)

Concernant la durée des plosives en position finale ( $C_2$  de mots simples et  $C_2.C_3$  de mots composés), nos analyses statistiques montrent que la différence au niveau de leur durée n'est pas significative ( $p = 0,308$ ), ni leur durée de l'occlusion ( $p = 0,926$ ) et écarte donc toute différence de durée entre une consonne plosive en finale de mot et une consonne en finale de syllabe 1 d'un mot composé lexical. Il est à signaler que pour les nasales en position de coda, leur durée est significativement dissemblable ( $p = 0,001$ ).

La valeur du VOT varie en fonction de la position des plosives dans la syllabe. La figure 2 montre que le VOT de /p t k/ en moyenne est plus élevé en initiale qu'en coda. Cet allongement du VOT accompagne l'allongement de la consonne en position initiale. Le VOT moyen des plosives en  $C_1$  est d'environ 16,3 ms et la gamme des valeurs va de 10,9 ms (cas de [t]) à 21,8 ms (cas de [k]) (cf. Fig. 2).

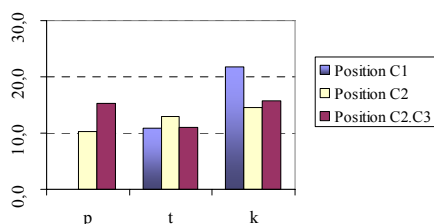


Figure 2: Moyenne des VOT en fonction des consonnes

Le VOT positif des occlusives sourdes vietnamiens dépend des lieux d'articulation de l'occlusive, si la voyelle qui le suit est toujours la même. En effet, le VOT moyen de consonne vélaire [k] a des valeurs plus élevées par rapport aux autres lieux d'articulation bilabiales et coronales, quelle que soit la position dans la syllabe, le mot. Cependant, la différence de VOT entre le [p] et le [k] est estompée en finale de syllabe 1 de mot polysyllabique (respectivement 15,3 ms et 15,8 ms).

Les consonnes perdent leur burst davantage en position  $C_2.C_3$  (27,3%) qu'en  $C_2$  (12,5%) et jamais en  $C_1$ . En ce qui concerne les plosives en coda, traditionnellement décrites comme non relâchées : soit elles sont réalisées sans burst et aucune trace de détente n'apparaît sur le spectrogramme (cf. Fig. 3) ; soit elles comportent un burst de durée plus brève (le burst le plus long de /ka/ est de 5,5 ms en  $C_2.C_3$ ) et peu intense (56,7 dB en  $C_2.C_3$  par rapport à 65,9 dB en  $C_1$ ) et pourraient être considérées comme une réalisation affaiblie de consonne plosive pleine ; soit le burst est visible mais laryngalisé (*creaky voice*) lors qu'il précède une glottale (le cas de /k/ dans la figure 4) :

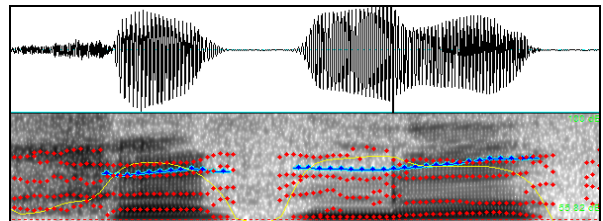


Figure 3: Signal et spectrogramme avec la durée du mot [χat nuɾk] où le burst est absent.

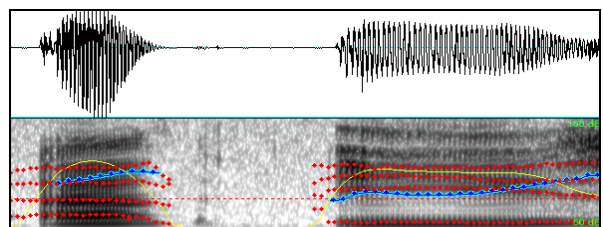


Figure 4: Représentation acoustique du mot [ʔak-ʔi] (laryngalisation)

Les consonnes en position initiale  $C_1$  présentent plus d'énergie dans le bruit de détente qu'en position finale du mot  $C_2$  ou de syllabe 1 de composé lexical  $C_2.C_3$ . La consonne coronale [t] possède le burst le plus intense en  $C_1$ , [t] est articulée avec plus d'énergie (71,5 dB) par rapport aux autres consonnes et aux autres positions. Mais l'intensité du burst des consonnes de même position finale de mot ou de syllabe 1 ( $C_2$  et  $C_2.C_3$ ) ne présente pas de différences significatives ( $p = 0,558$ ).

Le calcul des ANOVA à doubles facteurs à mesures répétées montre globalement un effet non significatif de la transition de  $F_0$  final entre la position  $C_2$  et  $C_2.C_3$  [ $F(1,74) = 0,131$  ;  $p = 0,719$ ]. Les différences de  $\Delta F_0$  sont significatives entre les catégories lieux (labial, coronal et vélaire) [ $F(2,73) = 9,532$  ;  $p = 0,00$ ] que ce soit en  $C_2$  ou  $C_2.C_3$ . Mais il n'y a aucune interaction entre la position et le lieu de la consonne [ $F(2,70) = 1,441$  ;  $p = 0,244$ ].

Les résultats d'analyse nous ont montré que  $F_1$  final des plosives ne connaît pas de différence significative entre position  $C_2$  et  $C_2.C_3$  [ $F(1,74) = 2,203$  ;  $p = 0,142$ ]. La transition de  $F_1$  est différent selon le lieu des plosives [ $F(2,73) = 4,024$  ;  $p = 0,022$ ] mais pas pour les nasales [ $F(1,46) = 0,573$  ;  $p = 0,453$ ]. Que le lieu soit labial, coronal ou vélaire, les analyses ont montré un effet non significatif des valeurs de  $F_1$  final pour les plosives et les nasales entre

C<sub>2</sub> et C<sub>2</sub>.C<sub>3</sub>, aucune interaction entre deux facteurs position et lieu de consonnes.

Les mêmes tests ont été effectués pour les valeurs de delta F<sub>2</sub> final selon le lieu des plosives. Il n'y a pas de différence significative de F<sub>2</sub> selon la position quand la consonne est labiale (p = 0,342) et coronale (p = 0,623), par contre la différence est significative selon la position quand la consonne est vélaire (p = 0,044). La valeur de F<sub>2</sub> final des nasales ne connaît pas de différence significative entre position C<sub>2</sub> et C<sub>2</sub>.C<sub>3</sub> (p = 0,391) ni de différence selon le lieu de consonne (p = 0,531). Que le lieu soit labial ou coronal, les analyses ont montré un effet non significatif des valeurs de F<sub>2</sub> final entre C<sub>2</sub> et C<sub>2</sub>.C<sub>3</sub>, aucune interaction entre deux facteurs position et lieu de consonnes (p = 0,99).

#### 4. DISCUSSIONS

Les analyses avec la variable dépendante durée du bruit d'explosion montrent une différence non significative entre cette durée des C<sub>1</sub> et celle des C<sub>2</sub>.C<sub>3</sub> (p = 0,694). Par contre, la durée du burst des C<sub>2</sub> comparée à celle du groupe des {C<sub>1</sub>/C<sub>2</sub>.C<sub>3</sub>} est significativement différente (p = 0,008). Il s'agira de prolonger cette étude par un test perceptif pour valider si ces corrélats sont des corrélats perceptifs du lieu et de la distribution des consonnes (plosives et nasales).

Pour les nasales, on observe des durées différentes en fonction de leur distribution mais pas d'autres différences significatives sur les autres paramètres acoustiques. Il faudra donc prolonger par l'analyse spectrale des nasales. Les formants sont probablement un indice acoustique fort du lieu de la nasale quelle que soit sa position (initiale ou finale, ou encore finale devant consonne de syllabe 2 de mot composé).

Si des différences acoustiques entre les positions de coda (C<sub>2</sub>) finale de mot et (C<sub>2</sub>.C<sub>3</sub>) finale de syllabe 1 de composé lexical ont été mises en évidence (durée totale et durée de l'occlusion), il s'agira de prolonger cette étude en approfondissant le rôle d'une unité accentuelle en vietnamien. Dans cette langue, le rôle de l'accentuation est resté assez confus en l'absence de beaucoup de recherches particulières sur ce sujet. Une étude expérimentale très détaillée, réalisée par Do (1986) (citée par [Ngu00]) a mis en évidence la présence d'un accent qui tombe sur la dernière syllabe d'un mot composé. Il s'agira de regarder et d'analyser les syllabes accentuées des composés dissyllabiques pour observer le comportement des plosives non relâchées dans le contexte accentuel dans le prolongement de cette étude.

En effet, dans la mesure où de l'énergie est apportée par l'accent, si celui-ci frappe effectivement la syllabe 2 des composés lexicaux, celle-ci n'est pas manifeste au niveau de l'énergie du burst. Sur ce point, il s'agira aussi de mesurer les réalisations en C<sub>3</sub> par rapport à la position C<sub>1</sub> et C<sub>4</sub> par rapport à C<sub>2</sub>.

Faute de temps, ce travail est une étude monolocuteur. Il sera nécessaire de prolonger cette étude pour écarter les caractéristiques individuelles.

S'appuyant sur les résultats obtenus dans cette première étape, il s'agira d'établir un paradigme expérimental rendant compte de la perception des séquences de consonnes. Ce paradigme devra prendre également en considération les analyses tirées d'observations sur les deux langues comme la dissymétrie importante (et presque inexistante en français) entre consonne initiale et consonne finale en vietnamien (seule une faible proportion de l'inventaire consonantique peut apparaître en position finale), aussi les séquences de consonnes possibles, en considérant la frontière syllabique, pouvant être différentes dans les deux langues.

Plus généralement, les résultats expérimentaux pourraient constituer la base d'une réflexion méthodologique en phonétique corrective.

#### RÉFÉRENCES

- [Bor92] Borrelle, A. (1992), "Les rapports entre perception et (re)production dans l'acquisition des langues secondes et/ou étrangères", *Cahiers du Centre Interdiscipline des Science du langage*, Université de Toulouse, pp.29-41
- [Cao85] Cao, X.H. (1985), *Phonologie et linéarité. Réflexions critiques sur les postulats de la phonologie contemporaine*. Société d'études linguistiques et anthropologiques de France. N° 18, SELAF, Paris.
- [Doa99] Doan, T.T. (1999), *Ngu am tieng Viet* (Tr. La phonétique du vietnamien), Edition de l'Université Nationale, Hanoi.
- [Mic04] Michaud, A. (2004), "Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese", *Phonetica* 61, pp.119-146.
- [Ngu00] Nguyen, T.B.M. (2000), *Regards sur l'enseignement de la phonétique dans la formation des étudiants en FLE à l'Université Pédagogique de Ho Chi Minh ville*, Thèse de Doctorat en Sciences du langage, Université de Rouen, France.
- [Rou04] Rousset, I. (2004), *Structures syllabiques et lexicales des langues du monde. Données, typologiques, tendances universelles et contraintes substantielles*, Thèse de Doctorat en Sciences du langage, Université Stendhal, Grenoble.
- [Ser87] Serniclaes, W. (1987), *Etude Expérimentale de la perception du trait de voisement des occlusives du français*, Thèse de Doctorat en Sciences Psychologiques et Pédagogiques, Institut de Phonétique, Université Libre de Bruxelles.
- [Tru70] Truong, V.C. (1970), *Structure de la langue vietnamienne*, Centre Universitaire des Langues Orientales Vivantes, Paris.