

# Identification des consonnes finales du vietnamien par des locuteurs natifs

Thi-Thuy-Hien Tran, Nathalie Vallée

Département Parole et Cognition, GIPSA-lab

1180, avenue Centrale, BP 25 – 38040 GRENOBLE Cedex 9, France

Tél. : ++33 (0)4 76 82 41 20 - Fax : ++33 (0)4 76 82 43 35

Courriel : [thi-thuy-hien.tran@gipsa-lab.grenoble-inp.fr](mailto:thi-thuy-hien.tran@gipsa-lab.grenoble-inp.fr), [nathalie.vallee@gipsa-lab.grenoble-inp.fr](mailto:nathalie.vallee@gipsa-lab.grenoble-inp.fr)

## ABSTRACT

A great difficulty encountered by Vietnamese subjects, who learn French, is that consonant clusters, which do not exist in Vietnamese, are mispronounced. This problem persists even after several years of practicing, and even when the French clusters correspond to Vietnamese consonant sequences. The general aim of our project is to identify the factors which are the main cause of this problem. In this framework, the paper deals with the perception of syllable-final stops in Vietnamese by 20 native Northern-Vietnamese listeners in order to examine if cues in Vietnamese stop-consonant percepts affect the perception of French consonants. Our findings suggest that specific acoustic characteristics and probably the lexical frequency of final consonants lead the subjects in their responses.

## 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 souvent difficile, en début d'apprentissage, 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]. L'oreille de l'apprenant de langue seconde<sup>1</sup> « se comporte un peu comme une oreille pathologique » en ce que « les habitudes perceptives modelées par le système sonore de la langue maternelle se projettent sur la langue seconde, provoquant ainsi une sorte d'effet de filtrage » [Ost96].

D'autres travaux ont mis en exergue le lien étroit unissant perception et production de la parole des langues secondes [Bor91], [Bor92], [Bes96], [Bes06] et du langage humain en général [Sch02]. Il y aurait, chez les apprenants d'une langue seconde, « systématisation des erreurs de perception et donc de production » [Bor96]. Selon cet auteur, pour qu'un son soit correctement (re)produit, il faut d'abord qu'il soit bien perçu, c'est-à-dire reconnu, identifié par rapport aux autres. Dans ces travaux, la question centrale est de

<sup>1</sup> Le terme *langue seconde* est utilisé ici dans le sens de la « langue dont l'usage est acquis à une étape de la vie ultérieure à l'apprentissage de la *langue première* ou *maternelle* » [Dub01]

comprendre comment la perception des non-natifs influence l'apprentissage d'une langue seconde et notamment ses aspects phonétiques et phonologiques.

Parmi les difficultés de sujets vietnamiens apprenant le français figurent les groupes de consonnes du français qui n'existent pas en vietnamien, langue isolante où les mots sont invariables et indécomposables. Ces groupes sont souvent réalisés déformés par rapport à la cible (ex. « psychique » => [si'fik], « insecte » => [ɛsek'tə]). Cette difficulté résistante perdure même après plusieurs années de pratique et même chez des étudiants de niveau déjà confirmé [Ngu00]. En quoi consistent exactement les erreurs de réalisation et quelles sont leurs implications dans l'acquisition des percepts phonétiques ? Quelles caractéristiques de la langue maternelle gênent la réalisation des clusters consonantiques du français ? L'objectif général de ce travail est de répertorier et comprendre les facteurs responsables de cette difficulté rencontrée par les apprenants vietnamiens, l'hypothèse étant que l'étude de la perception de la parole par des non-natifs peut rendre compte des difficultés que les apprenants tardifs rencontrent avec certains contrastes et segments de la L2 [Bes06].

À part le fait que le vietnamien est une langue tonale, les deux langues possèdent des gabarits lexicaux et des patrons syllabiques très différents. Le vietnamien est une langue monosyllabique sur le plan phonologique mais polysyllabique en partie sur le plan lexical [Mic04]. La différence entre mot simple et mot composé n'existe que par le nombre de syllabes : le mot simple est monosyllabique et les composés comportent de 2 à 4 syllabes. Les composés à deux syllabes sont les plus nombreux et il n'existe pas de composés de plus de quatre syllabes [Tru70]. Par exemple, le mot simple « học » qui signifie *apprendre*, associé à « tập » (*exercer, s'entraîner*), forme le mot composé « học tập » qui porte un sens plus général (*faire des études*).

[Doa99] représente les patrons syllabiques du vietnamien par  $C_1(w)V(C_2)$ , les parenthèses indiquant les constituants optionnels. La diversité des patrons syllabiques est beaucoup plus présente en français, l'attaque et la coda pouvant être complexes, c'est-à-dire occupées par des groupes de consonnes (clusters)  $(C_1)(C_2)(C_3)V(C_4)(C_5)(C_6)(C_7)$  [Rou04]. Ce cas n'est pas rencontré en vietnamien où les séquences de consonnes ne se rencontrent qu'à la frontière de mot ou à la frontière syllabique à l'intérieur d'un mot composé. De fait les séquences de consonnes appartiennent à deux

syllabes différentes. Cependant, 96 % des structures syllabiques du français entrent dans le patron  $(C_1)(C_2)V(C_3)$  qui se rapproche de la structure syllabique du vietnamien  $C_1(w)V(C_2)$ .

Autre dissimilarité entre syllabe du français et syllabe du vietnamien : la réalisation des consonnes qui ferment la syllabe. En effet, les consonnes finales sont réalisées différemment : dans la plupart des cas, la tenue de l'occlusion des plosives en coda (/p/ /t/ /k/) du vietnamien, quelle que soit leur condition de réalisation, n'est pas suivie d'un bruit caractéristique d'explosion rapide et audible [Doa99] et « la phase de relâchement est rarement obligatoire par ces consonnes finales » [Cao85]. Au contraire dans le cas du français, les plosives finales, accompagnées ou non d'une vibration des cordes vocales, sont généralement suivies d'un relâchement audible.

Une étude récente de données acoustiques [Tra09] montre que les consonnes finales du vietnamien comportent des caractéristiques différentes en fonction du type de frontière syllabique qu'elles précèdent : d'une part, les plosives et nasales sont significativement plus longues devant frontière de mot que devant frontière de syllabe à l'intérieur d'un mot composé lexical ; d'autre part, les plosives sont plus souvent non relâchées, sinon avec un burst significativement plus bref et de plus faible amplitude, en finale de syllabe à l'intérieur de mot.

Dans le prolongement de ce travail, l'étude que nous présentons ici a pour objectif d'identifier les corrélats perceptifs des consonnes finales du vietnamien en lien avec un éventuel effet de l'influence du type de frontière syllabique sur la perception des consonnes. En d'autres termes, il s'agit de répondre aux questions suivantes : quels sont les indices qui permettent d'identifier le lieu d'articulation d'une consonne du vietnamien lorsque celle-ci est non relâchée ? Existe-il des différences dans la perception des consonnes en fonction de leur position dans la syllabe ? Dans le mot ?

## 2. MÉTHODOLOGIE

### 2.1. Stimuli

Un ensemble de 55 items sonores naturels de type CVC correspondant soit à un mot monosyllabique, soit à la première syllabe d'un mot composé ont été sélectionnés pour l'étude. Ils comportent l'une des 6 occlusives /p t k m n ŋ/ en contexte de la voyelle la plus ouverte /a/ et de même contexte tonal (ton montant D1). Plus concrètement, dans cet ensemble d'items lexicaux, ces six consonnes se trouvent :

- Soit en position finale  $C_2$  de mots monosyllabiques  $[C_1aC_2]$  (exemples : [tap] « *táp* » *happer*, [kak] « *các* » (marque pluriel), [maŋ] « *máng* » *mangeoire*) ;
- Soit en position finale  $C_2$  de première syllabe  $[C_1aC_2]$  de mots composés lexicaux dissyllabiques  $[C_1aC_2.C_3VC_4]$  (exemples : [mat] dans [mat.za] « *mát dạ* » *content*, [fap] dans [fap.leŋ] « *pháp*

*lệnh* » *ordonnance*, [saŋ] dans [saŋ.taw] « *sáng tạo* » *créer*)

Ces stimuli sont extraits des phrases porteuses d'un corpus lu par un sujet natif masculin, âgé de 27 ans, originaire du Nord du Vietnam.

Les stimuli ont été choisis de manière à ce que leur entourage dans la phrase porteuse soit le plus neutre possible du point de vue articulaire pour faciliter l'extraction de la syllabe : toute consonne sourde en coda est suivie d'une consonne sonore, et inversement, toute consonne sonore est suivie d'une consonne sourde.

Au total, 29 items auditifs CVC terminés par une plosive et 26 items terminés par une nasale ont été retenus pour le test.

Vingt sujets (10 hommes et 10 femmes) originaires du Nord du Vietnam, âgés de 24 à 29 ans, étudiants à Grenoble, ont participé au test. Aucun d'entre eux ne possédait de formation universitaire en phonétique-phonologie. La plupart avaient eu connaissance de notions non approfondies de phonétique lors de leur apprentissage en français langue étrangère au Vietnam.

### 2.2. Déroulement

Nous avons choisi de tester les plosives à part des nasales. Le programme utilisé pour le test est PerceptA de R. Carré. Les sujets avaient pour consigne d'écouter, puis identifier la consonne finale du stimulus présenté parmi 3 solutions proposées (choix forcé) correspondant soit aux 3 plosives /p t k/, soit aux 3 nasales /m n ŋ/ en cliquant la réponse à l'écran avec l'aide de la souris (cf. Fig.1 et Fig.2).

Les sujets n'étant pas familiers des symboles de l'alphabet phonétique international, ceux-ci leur ont été expliqués par l'expérimentateur avant le test. L'interface du programme ne permettant pas d'afficher le caractère spécifique de la nasale vélaire /ŋ/, celle-ci apparaissait sous sa forme orthographique « ng » (cf. Fig.2).

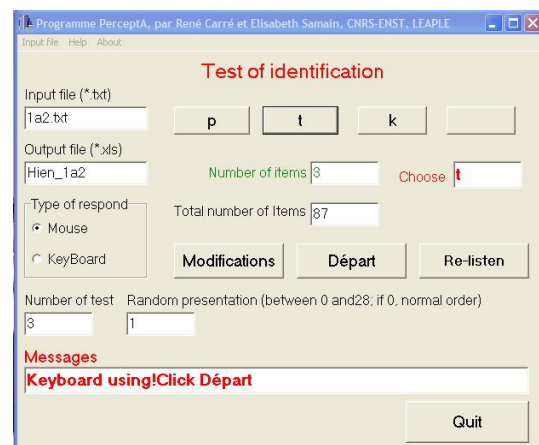
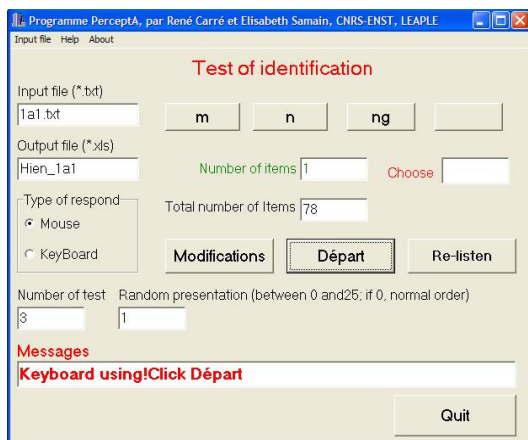


Figure 1 : Interface utilisée pour les plosives.



**Figure 2 :** Interface utilisée pour les nasales.

Le premier test concernait les plosives et était constitué de 3 répétitions des 29 items présentés dans un ordre aléatoire pour chaque sujet. Le second test comportait 3 répétitions des 26 items sélectionnés pour les nasales présentées également de manière aléatoire.

Le test s'est déroulé dans une salle calme du département Parole et Cognition de GIPSA-lab (Grenoble, France) en utilisant un casque binaural Sennheiser HD-25-13.

### 3. RÉSULTATS

Les analyses montrent que pour les plosives en coda, le meilleur score d'identification est obtenu par /t/ et /k/ (cf. Tab.1). La différence de score de bonnes réponses est significative entre /p/ et /t/ ( $p = 0$ ) ; entre /p/ et /k/ ( $p = 0$ ), mais pas entre /t/ et /k/ ( $p = 0,79$ ).

**Table 1 :** Score d'identification des plosives.

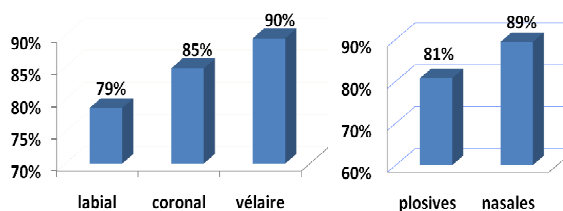
Consonnes	Choix p	Choix t	Choix k
<b>p</b>	<b>74 %</b>	19 %	7 %
<b>t</b>	8 %	<b>84 %</b>	9 %
<b>k</b>	11 %	6 %	<b>83 %</b>

Pour les nasales, la vélaire /ŋ/ obtient le meilleur score d'identification par rapport à /m/ et /n/ (cf. Tab.2). La différence de score de bonnes réponses est significative entre /m/ et /ŋ/ ( $p = 0$ ) ainsi que entre /n/ et /ŋ/ ( $p = 0$ ), mais pas entre /m/ et /n/ ( $p = 0,34$ ).

**Table 2 :** Score d'identification des nasales.

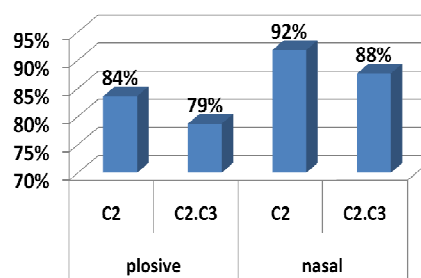
Consonnes	Choix m	Choix n	Choix ŋ
<b>m</b>	<b>85 %</b>	10 %	6 %
<b>n</b>	9 %	<b>86 %</b>	4 %
<b>ŋ</b>	1 %	5 %	<b>95 %</b>

Regroupées par lieu d'articulation, les vélares obtiennent un meilleur score par rapport aux labiales et coronales (cf. Fig.3). Par mode d'articulation, les nasales engendrent moins de confusion de la part des sujets par rapport aux plosives (89 % contre 81 %).



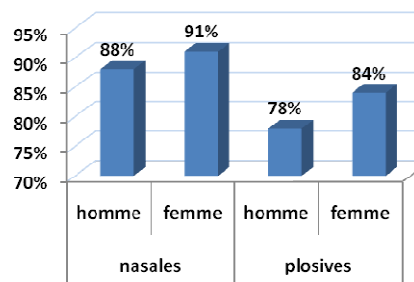
**Figure 3 :** Score d'identification des consonnes finales par lieu et mode d'articulation.

Concernant l'influence du type de frontière syllabique sur le score de bonnes réponses, la figure 4 montre davantage de confusion pour les plosives et les nasales lorsque la coda  $C_2$  précède une frontière intra-mots (devant  $C_3$  de mots composés), et cette différence est significative dans les deux cas.



**Figure 4 :** Score d'identification des consonnes finales selon leur position dans la syllabe.

Les analyses montrent que les meilleurs scores d'identification des plosives ainsi que des nasales sont obtenus par les femmes (cf. Fig.5). La différence de score est différente significativement en fonction du sexe des sujets pour les nasales ( $p = 0,017$ ) et pour les plosives ( $p = 0,003$ ).



**Figure 5 :** Score d'identification des consonnes finales selon le sexe des sujets

### 4. DISCUSSION

Quels sont les indices qui permettent d'identifier le lieu d'articulation d'une consonne lorsque celle-ci n'est pas relâchée ? Pourquoi les nasales ont été mieux perçues en coda par rapport aux plosives ? Est-ce que ces résultats sont dus à l'effet de la fréquence de ces consonnes dans le lexique ? Des indices acoustiques du corpus d'où ont été extraits les stimuli jouent-ils un rôle d'indicateur ?

#### 4.1. Fréquence lexicale

Nous avons travaillé avec un lexique vietnamien de 5 000 entrées que nous avons intégré à la base de données ULSID (UCLA Lexical and Segmental Inventory Database) [Rou04], développée au sein de notre laboratoire. ULSID contient les lexiques syllabés de plusieurs dizaines de langues ainsi qu'un système de requêtes permettant d'interroger les structures lexicales et syllabiques des langues. Dans le lexique vietnamien, nous observons que /t/ et /ŋ/ qui sont les mieux identifiés par les sujets, possèdent les meilleures fréquences d'occurrences dans leur catégorie (plosives et nasales) toutes distributions confondues (cf. Fig.5).

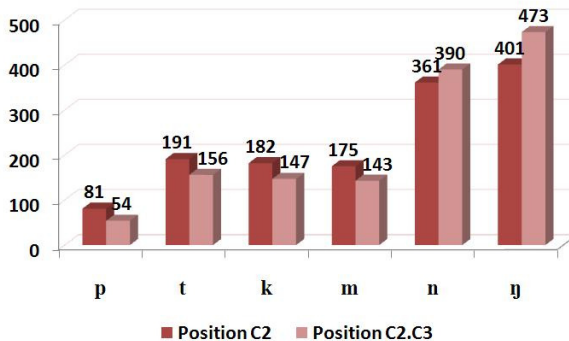


Figure 5 : Fréquence lexicale des consonnes en coda.

Une tendance entre le score de bonnes réponses et la fréquence d'occurrences des consonnes en position de coda dans le lexique a été effectivement observée. Les coefficients calculés ( $R^2$ ) indiquent une corrélation importante entre fréquence d'occurrences et score d'identification des plosives comme des nasales, respectivement 0,76 et 0,89.

La comparaison avec la fréquence de ces consonnes en position initiale montre que les nasales sont particulièrement présentes en position finale de syllabe - C<sub>2</sub> finale de mots monosyllabiques ou finale de première syllabe de mots composés C<sub>2</sub>.C<sub>3</sub> (cf. Fig.6). Ce qui pourrait expliquer un meilleur résultat d'identification des nasales par rapport aux plosives en position de coda.

Les nasales /m/ et /n/ très fréquentes en coda par rapport à la position d'attaque obtiennent des scores d'identification (respectivement 85 % et 86 %) proches de ceux des plosives /t/ et /k/ (respectivement 84 % et 83 %) qui sont cependant plus fréquentes en position d'attaque que de coda. Par conséquent, la fréquence lexicale n'explique pas tout, même si une certaine corrélation entre les scores d'identification et la fréquence dans le lexique a été globalement observée.

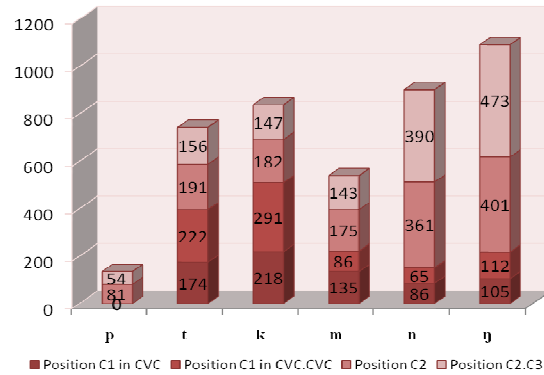


Figure 6 : Fréquence des consonnes en coda en comparaison avec attaque dans le lexique.

Le même type d'analyse dans le lexique, cette fois pour les consonnes en contexte /a/ uniquement et sous le ton montant D1, montre toujours une occurrence plus forte pour les vélaires qui obtiennent les meilleurs scores d'identification. Cependant, la corrélation entre la fréquence d'occurrences et les scores de bonnes réponses n'est plus observée dans le contexte /a/ - D1 ( $R^2 = 0,28$ ).

Quels autres facteurs ont pu guider les sujets dans leur choix de réponse ?

#### 4.2. Indice acoustique – le burst

Malgré la caractéristique non relâchée des plosives finales en vietnamien, nous avons constaté la présence de bursts (dans 34,5 % de cas). Il est intéressant de noter que dans ces cas, les bursts sont de durée brève (en moyenne 4 ms) et d'intensité plus faible (60 dB en moyenne) comparés à ceux produits en position initiale (10 ms et 68 dB en moyenne).

En considérant la présence / absence de burst dans la réalisation de la consonne en coda des stimuli, les scores montrent qu'ils sont significativement meilleurs lorsque le burst est présent et ce quelle que soit la frontière qui suit la consonne (frontière de mot ou de syllabe intramots) (cf. Fig.7). La tendance est identique lorsque sont considérés les lieux d'articulation.

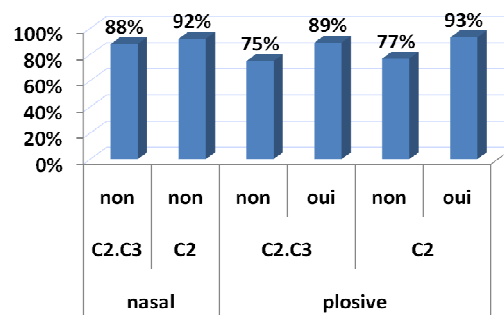


Figure 7 : Score de bonnes réponses en fonction de la présence du burst et du type de frontière.

Ces résultats montrent clairement que le burst facilite l'identification du lieu d'articulation de la consonne.

Pourtant, le nombre de plosives finales relâchées (avec présence de burst) dans les stimuli du test est beaucoup plus restreint par rapport au nombre de plosives non relâchées (cf. Tab.3).

**Table 3 :** Score d'identification en fonction de la fréquence d'occurrences des plosives dans les stimuli du test et de la présence d'un burst.

Consonnes	Burst	Occurrences	Score
<b>p</b>	<b>Non</b>	<b>360</b>	<b>71%</b>
	Oui	120	82%
<b>t</b>	<b>Non</b>	<b>360</b>	<b>76%</b>
	Oui	360	92%
<b>k</b>	<b>Non</b>	<b>420</b>	<b>79%</b>
	Oui	120	97%

La table 3 montre que le score d'identification des plosives non relâchées est au-dessus du seuil du hasard. Ces plosives comportent vraisemblablement d'autres indices acoustiques porteurs d'informations sur leurs lieux d'articulation.

#### 4.3. Indice acoustique – transition formantique

Dans la littérature, plusieurs travaux ont mis en évidence l'importance de la transition formantique pour la perception des consonnes dans des séquences bi-segmentales CV ([Del55], [Lib54], [Ser87]). Pour la transition VC de notre étude, nous avons observé l'évolution des 3 premiers formants  $F_1$ ,  $F_2$ ,  $F_3$ , de la fréquence fondamentale et de l'intensité mesurés à 50 % et 90 % de la durée de la voyelle. À 50 % de sa réalisation, le noyau vocalique est en général dans sa partie la plus stable articulatoirement, qui devrait être la moins influencée par les consonnes de l'entourage. À 90 %, la réalisation vocalique contient la transition entre voyelle et consonne finale. À 100 % de sa réalisation, la voyelle subit encore plus de perturbations pour prendre un point de référence nécessaire au calcul des transitions.

Des analyses statistiques ont été effectuées sous SPSS pour repérer d'éventuels effets d'interaction entre lieux d'articulation et paramètres acoustiques dans cette partie de la transition voyelle - consonne.

Les ANOVA montrent globalement un effet non significatif de la transition de  $F_0$  entre les plosives [ $F(2,28) = 0,63$  ;  $p = 0,54$ ].

Que le lieu soit labial, coronal ou vélaire, les analyses ont également montré un effet non significatif des valeurs de delta  $F_3$  pour les plosives finales [ $F(2,28) = 1,94$  ;  $p = 0,16$ ].

Les mêmes tests effectués pour les valeurs de delta  $F_1$  [ $F(2,28) = 3,8$  ;  $p = 0,04$ ], delta  $F_2$  [ $F(2,28) = 4,55$  ;  $p = 0,02$ ] et delta Intensité [ $F(2,28) = 7,83$  ;  $p = 0,00$ ] montrent des différences significatives de transition en fonction du lieu d'articulation des plosives.

## 5. CONCLUSION

Notre étude montre que des sujets vietnamiens, dans leur langue maternelle, identifient mieux les nasales que les plosives en position de coda. Bien que les plosives soient généralement non relâchées dans cette position, nos résultats montrent que la présence d'un burst même faible augmente le taux d'identification correcte.

Dans le cas des plosives non relâchées, les scores restent supérieurs à 70 %. Des analyses de fréquence lexicale montrent que celle-ci n'explique pas totalement ces scores. Une analyse acoustique des transitions des fréquences formantiques, de la fréquence fondamentale et de l'intensité montre que delta  $F_1$ , delta  $F_2$  et delta Intensité étaient significativement différents en fonction du lieu d'articulation des plosives. Ces résultats sont en conformité avec les études antérieures ([Ser87]) sur le rôle des transitions formantiques dans la perception du lieu d'articulation des consonnes. En l'absence de burst, les consonnes contiennent dans leur transition avec la voyelle précédente des informations sur leur lieu d'articulation qui ont sans doute guidé les sujets dans leur choix de réponse.

Ce travail s'inscrit pleinement dans la vérification de l'hypothèse formulée par [Bes06] : "non native speech perception can also account for the relative difficulties that the late learners have with specific L2 segments and contrasts". Les résultats présentés ici constituent une étape préliminaire à la constitution d'un paradigme expérimental rendant compte de la perception des séquences de consonnes que l'on peut trouver dans les deux langues, à la différence près de la présence d'une frontière syllabique ou frontière de mots. Ces résultats sont actuellement complétés par une série d'expériences portant sur la perception et l'analyse acoustique des productions des consonnes du français par des sujets vietnamiens : analyse perceptive des consonnes simples du français en position initiale, intervocalique et finale, des consonnes en séquences en fonction de leur distribution dans le mot et par rapport à une frontière syllabique. Les résultats seront mis en relation avec des données en production obtenues avec un articulographe électromagnétique EMA® (Carstens) pour des consonnes du français et du vietnamien produites par quatre Vietnamiens natifs.

## REMERCIEMENTS

Cette recherche a bénéficié d'un financement de l'Agence Universitaire de la Francophonie (PC – 411/2460).

Un grand merci à René Carré et Lionel Granjon pour l'aide matérielle qu'ils ont chacun apportée à ce projet et pour les échanges constructifs et discussions très fructueuses. Merci également à tous les Vietnamiens qui ont donné de leur temps pour participer aux tests.

## RÉFÉRENCES

- [Bes95] Best, C. T. (1995), "A direct realist view of cross-language speech perception". In W. Strange (Ed): *Speech Perception and Linguistic Experience*: pp. 171-206. Baltimore, MD: York Press.
- [Bes06] Best, C. T. & al. (2006), "Nonnative and second-language speech perception: Commonalities and complementarities", In M. J. Munro & O.-S. Bohn (eds) *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins.
- [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
- [Bor96] Borrelle, A. (1996), "Systématisation des erreurs de production et donc de perception chez les apprenants étrangers ?", *Revue de Phonétique Appliquée*, N° 118-119, pp. 1-16.
- [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.
- [Del95] Delattre, P. C. & al. (1995), "Acoustic loci and transitional cues for consonants", *Journal of the Acoustical Society of America*, Vol. 27, pp. 769-773.
- [Doa99] Doan, T. T. (1999), *Ngu am tieng Viet* (Tr. La phonétique du vietnamien), Edition de l'Université Nationale, Hanoi.
- [Dub01] Dubois, J. & al. (2001), *Dictionnaire de linguistique*, Larousse, Paris
- [Fle95] Flege, J. A. (1995), "Second language speech learning: Theory, findings and problems". In W. Strange (Ed): *Speech Perception and Linguistic Experience*: pp. 233-272. Baltimore, MD: York Press.
- [Fle97] Flege, J. A. & al. (1997), "Amount of native-language (L1) use affects the pronunciation of an L2". *Journal of Phonetics* 25, pp. 169-186.
- [Lib54] Liberman, A. M. & al. (1954) "The role of consonant-vowel transitions in the perception of the stop and nasal consonants". *Psychological Monographs* 68, pp. 1-13.
- [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.
- [Ost96] Ostiguy, L. & al. (1996), *Introduction à la phonétique comparée. Les sons. Le français et l'anglais nord-américains*. Les Presses de l'Université Laval.
- [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.
- [Sch02] Schwartz, J. L. & al. (2002), "Phonology in a theory of perception – for – action – control", *Phonetics, Phonology and Cognition*, Ed. by. Laks B. et Durand J., Oxford University Press, Oxford, pp. 244-280.
- [Tra09] Tran, T. T. H. & Vallee, N. (2009), "An Acoustic Study of Interword Consonant Sequences in Vietnamese", *Journal of the Southeast Asian Linguistics Society*, pp. 231-249, Pacific Linguistics Publisher.
- [Tru70] Truong, V. C. (1970), *Structure de la langue vietnamienne*, Centre Universitaire des Langues Orientales Vivantes, Paris.