

« Locus equation » pour les consonnes /b/, /d/ et /ɣ/ du vietnamien

Eric Castelli¹ - Anne Hierholtz^{1,2}

¹Centre de recherche international MICA
IP Hanoi - CNRS/UMI-2954 - INP Grenoble
1 Dai Côt Viêt, Hanoi, Vietnam

²Ecole Polytechnique Fédérale de Lausanne
CH-1015, Lausanne, Suisse

Eric.Castelli@mica.edu.vn, Anne.Hierholtz@mica.edu.vn
<http://www.mica.edu.vn>

ABSTRACT

Locus equation measurements are one approach used to characterise vocal tract resonances during stop consonant production, the place of articulation of the consonant and the nature of the vowel-consonant transition. Taking up again previous literature studies, the aim of the present work is to applying locus equation measurements to Vietnamese language, for two stop voiced consonants /b/ and /d/, and for the specific velar voiced fricative consonant /ɣ/. Because Vietnamese is a tonal language, a specific corpus was built, in order to avoid as much as possible tonal co-articulation effects. We also take into account the two specific vowels that present dynamic characteristics. Comparisons with other languages are given at the end of this study.

1. INTRODUCTION

L'appareil de production du signal de parole peut être considéré en première approximation comme un système simple source + filtre (ou excitateur + résonateur). Lors de la production des voyelles, la principale source d'excitation est constituée des vibrations des cordes vocales et celle-ci est pseudopériodique. Les spectres présentent alors des pics bien visibles, appelés formants, qui peuvent être considérés comme la signature de la voyelle. Ces formants représentent les résonances du conduit vocal et dépendent de la géométrie de ce même conduit. En revanche, pendant la production des consonnes sourdes, la source d'excitation est essentiellement bruitée, avec une source de bruit qui n'est pas forcément située au début du conduit vocal, et les spectres directement calculés sur le signal présentent une répartition assez uniforme des fréquences. Dans le cas des consonnes occlusives, le signal est même partiellement nul juste avant le bruit de plosion puisque le conduit vocal est complètement fermé. Il est alors difficile, voire impossible, de déterminer les résonances du conduit vocal. Cependant, que cela soit pendant la production des voyelles ou pendant celle des consonnes, le conduit vocal est toujours présent, avec une géométrie présentant des résonances.

L'idée du locus consiste donc à mesurer les transitions formantiques pendant la production de séquences CVC et de prolonger artificiellement les positions des résonances

dans les consonnes où il est souvent impossible de les déterminer. La position présumée des trois premières résonances (formants F1, F2 et F3) et la prolongation des trois courbes d'évolution au milieu de la consonne s'appellent « le Locus Consonantique » [1, 2]. Chaque consonne de chaque langue peut ainsi être caractérisée. Cependant, le locus ne renseigne pas seulement sur la géométrie du conduit vocal pendant la production de la consonne mais il apporte des informations pertinentes sur la nature de la transition [3].

Le vietnamien est une langue tonale sur laquelle peu d'études ont été menées. Nous pouvons trouver dans la littérature quelques travaux sur la phonétique de la langue mais ses caractéristiques phonologiques restent à étudier car elles permettront de comprendre ses stratégies de production spécifiques. Quelques propriétés spectrales et acoustiques caractérisant les voyelles et les tons sont connues [4, 5] mais aucune étude sur les consonnes n'a encore été menée à notre connaissance. Nous nous proposons d'appliquer l'équation du locus (*locus equation*) à la langue vietnamienne pour caractériser ses consonnes plosives voisées et sa consonne fricative vélaire. Cette caractérisation, outre d'approfondir nos connaissances de la langue, permettra pour une utilisation future de synthétiseurs anthropomorphiques, de mieux piloter les évolutions géométriques du conduit vocal.

2. LANGUE VIETNAMIENNE

Il est généralement admis que le vietnamien présente 9 voyelles acoustiquement différenciables, représentées par les caractères vietnamiens « i/y, ê, e, o, ô, a/ă, ơ/â, u, u » correspondant phonétiquement respectivement à /i/, /e/, /ɛ/, /ɔ/, /o/, /a/, /ɣ/, /tʰ/ & /u/. Les deux caractères « ă » et « â » peuvent être considérés comme une autre transcription des voyelles /ɣ/ et /a/, cependant, de récents travaux ont montré que les voyelles correspondantes notées /ă/ et /ɣ/ présentent des caractéristiques dynamiques [4]. Il y a 21 consonnes en vietnamien standard avec quelques différences dépendantes des dialectes prononcés. Ainsi dans les régions du nord du Vietnam, aucune distinction n'est faite entre les phonèmes /s/ et /ʃ/, /z/ et /z/, et /c/ et /t/. Notons au passage que le vietnamien présente seulement deux lieux d'articulation pour les consonnes plosives, qui sont labiales pour /p, b/

et alvéolaires pour /t, d/, alors qu'en français, il existe les consonnes vélares /k, g/. En effet, les consonnes vélares du vietnamien /ɣ/ et /χ/, respectivement voisée et non voisée, sont considérées non pas comme des plosives, mais comme des fricatives, la consonne /g/ n'existant pas en vietnamien.

3. EXPERIMENTATION

Les consonnes voisées plosives du vietnamien sont /b/ et /d/, transcrites respectivement par les caractères vietnamiens « b » et « đ », suivies par l'une des 11 voyelles vietnamiennes /a/, /ɛ/, /e/, /i/, /ɔ/, /o/, /u/, /ɣ/, /w/, /ã/ et /ɣ̃/. Chaque locuteur vietnamien prononce les syllabes CV (ou CVC) dans la phrase porteuse suivante : « Nôi CV/CVC êm ru » qui veut dire « Prononce CV (ou CVC) doucement ». Plusieurs considérations ont guidé le choix de cette phrase porteuse. Premièrement, le vietnamien étant une langue tonale, pour éviter d'éventuels phénomènes d'interférence avec les tons, nous avons souhaité une phrase prononcée en contexte de tons monotones. La condition idéale pour mesurer les formants dans les voyelles, aurait été de faire suivre la voyelle par une consonne occlusive non voisée /t/ ou /p/, ou bien voisée /d/ ou /b/. Cependant, en vietnamien, les syllabes fermées se terminant par ces consonnes ne se prononcent qu'avec des tons à la dynamique complexe. C'est pourquoi, pour les 9 voyelles normales, nous avons gardé une production de syllabes CV ouvertes, en choisissant le mot suivant dans la phrase tel qu'il commence par une voyelle neutre, pour laquelle nous savons que la géométrie du conduit vocal correspond à un tube de section constante, fermé à la glotte et ouvert aux lèvres. Deuxièmement, nous devons tenir compte du comportement dynamique des voyelles /ã/ et /ɣ̃/ qui n'existent qu'en contexte de syllabes fermées et avec la consonne nasale /n/ comme son final dans le cas des tons monotones. La consonne nasale ne nous permet pas de mesurer correctement les formants de la voyelle, car il est alors difficile de définir le point cible de la voyelle. C'est pourquoi, nous choisissons pour les deux voyelles spéciales « dynamiques », une syllabe CVC avec la consonne finale /t/ et le ton montant (habituellement noté ton 5) pour que cette configuration augmente le nombre de mots au contenu sémantique significatif en vietnamien (chaque exemple de syllabe CV ou CVC correspond à un mot ayant alors une signification). Chaque locuteur prononce donc les deux consonnes /b/ et /d/ associées avec les 11 voyelles, ce qui correspond à 22 syllabes différentes.

Nous étendons alors notre étude à la consonne vélaire fricative voisée /ɣ/ transcrite avec les caractères vietnamiens « g » ou « gh » (« g » avec les voyelles /a/ et /u/, « gh » avec les voyelles /ɛ/ ou /i/, « h » pouvant alors être considéré comme une représentation de la propriété fricative de la consonne). Pour cette consonne, la principale difficulté réside dans la détection d'un relâchement équivalent au « burst » des consonnes

occlusives pour mesurer la valeur du 2^{ème} formant au démarrage (onset) de la voyelle. Chaque sujet prononce des syllabes CV (ou CVC), combinaisons de la consonne fricative vélaire et des 11 voyelles vietnamiennes, ce qui rajoute 11 syllabes aux 22 syllabes précédentes

3.1. Mesures

Chaque séquence de signal de parole est enregistrée 5 fois par 8 locuteurs (4 hommes et 4 femmes) originaires du Nord du Vietnam (vietnamien officiel) avec un ordinateur dans un studio calme et sans bruit (mais qui n'est pas une véritable chambre sourde) à la fréquence d'échantillonnage de 11025 Hz. Une pré-emphase permet de rendre plat le spectre des voyelles. Les fréquences des formants sont mesurées avec le logiciel « Praat », en appliquant un fenêtrage gaussien et en calculant les coefficients LPC avec la méthode de Burg (pour plus de détails, se référer à [7]).

3.2. Résultats

Chaque « locus equation » est générée pour une consonne avec toutes les répétitions de chaque contexte vocalique (55 séquences = 5 répétitions x 11 voyelles). La pente de régression linéaire, l'ordonnée à l'origine (*Y-intercept*), le coefficient de corrélation R^2 et la déviation standard (*standard error of estimate*) SE sont regroupés pour les consonnes plosives initiales /b/ et /d/, pour chaque locuteur dans le tableau 1. Seule la locutrice F1 présente des valeurs sensiblement différentes de la moyenne du groupe d'étude (pente plus faible et Y-intercept plus grand).

Un exemple de tracé pour un seul locuteur féminin (F3) est donné par la figure 1. Les figures 2a et 2b montrent les équations du locus pour les deux consonnes plosives pour tous les locuteurs.

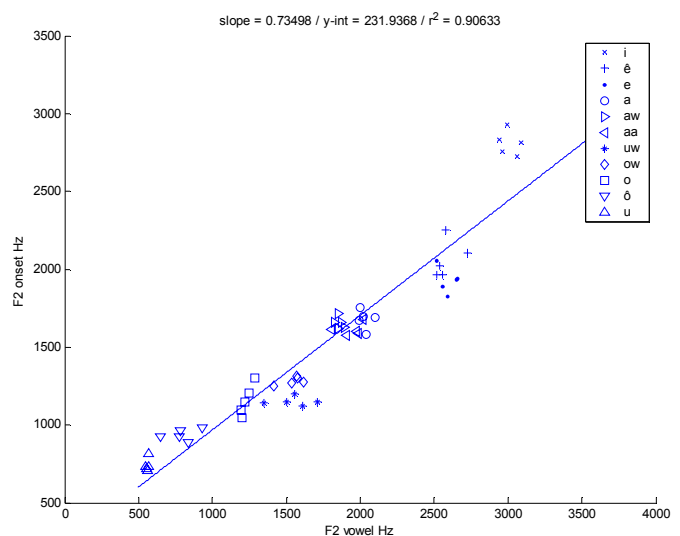


Figure 1 : « Locus equation » de la consonne /b/ (locuteur F3). Les notations « aw, aa, uw et ow » correspondent respectivement aux voyelles ã, â, σ et u.

Table 1 : Pente, Y-intercept, R^2 et SE des plosives /b/ et /d/ pour chaque locuteur/.

Locuteur Fem/Male	Pente		Y-intercept Hz		R^2		SE (Hz)	
	/b/	/d/	/b/	/d/	/b/	/d/	/b/	/d/
F1 (Hien)	0.36	0.21	680	1603	0.79	0.68	43.66	33.52
F2 (Huong)	0.56	0.29	377	1498	0.86	0.72	52.50	42.42
F3 (Yen)	0.73	0.41	232	1464	0.91	0.84	61.56	48.82
F4 (Hien2)	0.59	0.26	530	1564	0.82	0.66	64.15	44.71
M1 (Hai)	0.65	0.39	244	1103	0.85	0.81	55.47	39.10
M2 (Son)	0.53	0.46	331	979	0.81	0.86	49.28	36.45
M3 (Dat)	0.50	0.31	427	1170	0.84	0.76	43.58	34.32
M4 (Ha)	0.57	0.30	299	1223	0.89	0.67	41.03	40.51
Moyenne	0.56	0.33	390	1325	0.85	0.75	51	40
Ecart-type	0.11	0.08	153	235	0.04	0.08	12.01	07.77

Table 2 : Pente, Y-intercept, R^2 et SE de la fricative /ɣ/ pour chaque locuteur.

Locuteur Fem/Male	Pente	Y-intercept Hz	R^2	SE (Hz)
	/ɣ/	/ɣ/	/ɣ/	/ɣ/
F1 (Hien)	0.76	345	0.89	62.25
F2 (Huong)	0.86	331	0.88	73.13
F3 (Yen)	0.85	389	0.95	52.57
F4 (Hien2)	0.66	587	0.76	86.52
M1 (Hai)	0.70	500	0.90	48.13
M2 (Son)	0.75	421	0.91	44.53
M3 (Dat)	0.66	587	0.88	49.84
M4 (Ha)	0.49	689	0.81	48.00
Moyenne	0.71	481	0.87	58.12
Ecart-type	0.12	130	0.06	14.81

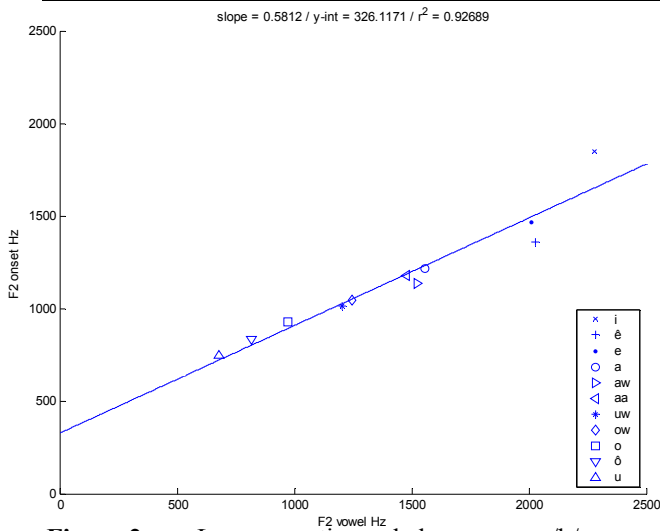


Figure 2a : « Locus equation » de la consonne /b/.

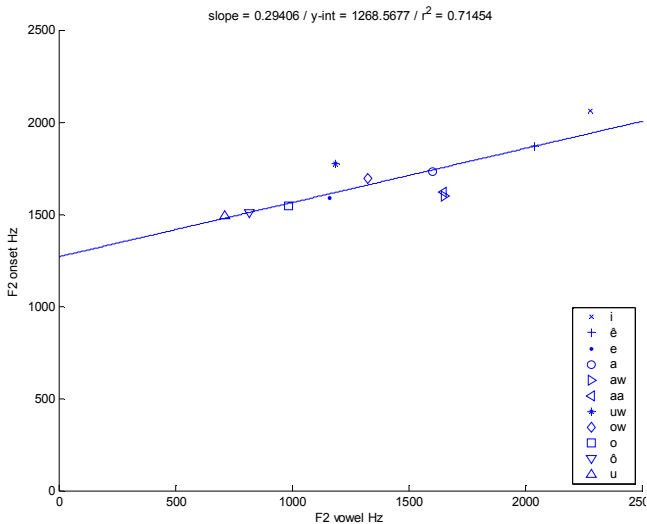


Figure 2b : « Locus equation » de la consonne /d/.

En ce qui concerne la consonne vélaire fricative /ɣ/, la même méthode est appliquée pour caractériser l'équation du locus. Le tableau 2 regroupe les valeurs de pente, « Y-intercept », R^2 et SE. Pour cette consonne aussi, un locuteur (M4) présente des résultats différents de la moyenne des autres locuteurs (pente plus faible et Y-intercept plus grand). Sur la figure 3 est tracé le « locus equation » pour toutes les configurations vocaliques pour tous les sujets.

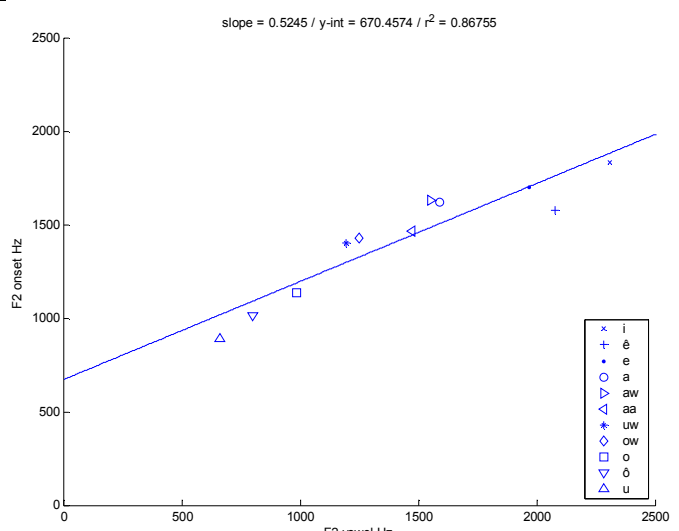


Figure 3 : « Locus equation » de la consonne /ɣ/.

Similairement aux résultats présentés par Sussman et al. [2] pour l'anglais américain pour la consonne plosive /g/, qui présente le même lieu d'articulation que la consonne vietnamienne /ɣ/, deux sous-groupes correspondant aux voyelles antérieures et aux voyelles postérieures (/g/ palatal et /g/ vélaire) peuvent être séparés et caractérisés par deux droites aux pentes différentes.

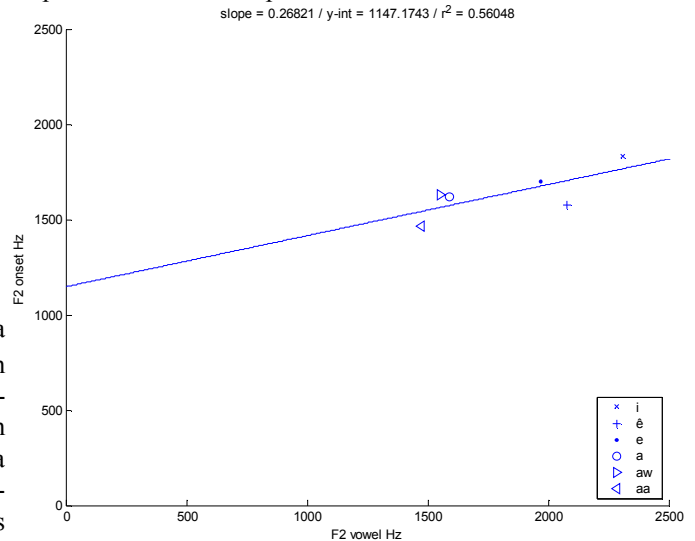


Figure 4a : « Locus equation » de la consonne /ɣ/ pour les voyelles antérieures.

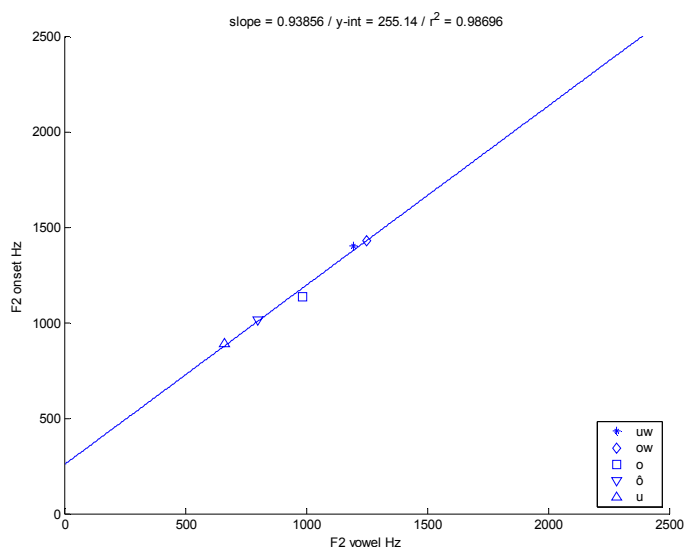


Figure 4b : « Locus equation » de la consonne /y/ pour les voyelles postérieures.

Nous obtenons alors les valeurs pour chaque sous-groupe de $R^2 = 0.98$, $SE = 48\text{Hz}$ pour les voyelles postérieures et de $R^2 = 0.56$, $SE = 180\text{Hz}$ pour les voyelles antérieures. Les figures 4a et 4b tracent les deux locus respectifs.

Comme proposé par Sussman et al. [2], nous utilisons le plan « pente/ordonnée à l'origine » pour représenter les lieux d'articulation des trois consonnes (figure 5). Nous retrouvons bien des zones distinctes qui classifient les deux consonnes /b/ et /d/ et « /y/ voyelles postérieures ». Cependant le lieu d'articulation de la consonne /y/ dans le contexte de voyelles antérieures chevauche ceux des consonnes /b/ et /d/.

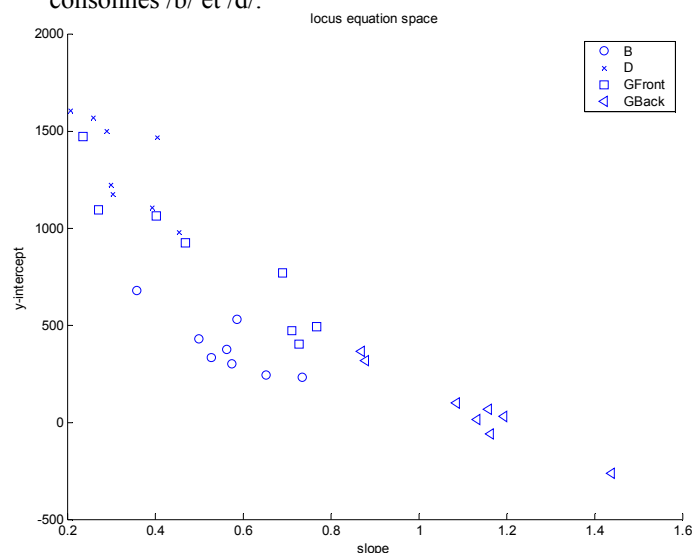


Figure 5 : Lieux d'articulation des consonnes /b/, /d/ et /y/ (dans les deux contextes de voyelles avant et arrière)

4. DISCUSSION

24 équations de locus ont été obtenues, à partir des mesures de 1320 prononciations des consonnes /b/, /d/ et /y/, avec une linéarité excellente et un bon regroupement des points autour de la droite de régression. Sur

l'ensemble de nos données nous obtenons un coefficient de corrélation R^2 de 0.82 ce qui est comparable avec les valeurs trouvées dans la littérature (0.86 pour les données de Sussman et al. [2]). Les lieux d'articulation des 2 consonnes plosives et de la consonne vélaire en contexte de voyelles arrières sont bien distincts. Cependant celui de la consonne vélaire palatale semble plus incertain, du fait peut-être de la difficulté de mesurer le formant F2 dans les transitions avec la consonne fricative.

Nous pouvons comparer nos résultats obtenus pour les consonnes vietnamiennes aux données disponibles pour d'autres langues, comme celles de Sussman et al. [2] sur l'anglais américain (10 locuteurs), de Krull [7] et de Lindblom [1] sur le suédois (respectivement 5 et 1 locuteurs), et de Sussman et al. [8] sur le thaï, l'arabe et l'urdu (respectivement 6, 3 et 5 locuteurs). Le tableau 3 regroupe les valeurs de pente et d'ordonnée à l'origine et permet de comparer les différentes stratégies d'articulation, qui pour Thaï et VN semblent similaires.

Table 3 : Pente, Y-intercept pour différentes langues

Langues	Pente			Y-intercept Hz		
	/b/	/d/	/g//y/	/b/	/d/	/g//y/
Anglais	0.87	0.43	0.66	106	1073	807
Suédois	0.63	0.32	0.95	487	1096	360
Thaï	0.70	0.30	---	228	1425	---
Arabe	0.77	0.25	0.92	206	1307	229
Urdu	0.81	0.50	0.97	172	857	212
Vietnamien	0.56	0.33	0.52	390	1325	670

BIBLIOGRAPHIE

- [1] Lindblom B. "On vowel reduction" *The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm*, Report n°29, 1963
- [2] Sussman H., McCaffrey H. A. & Matthews S.A. "An investigation of locus equations as a source of relational invariance for stop place categorization" *J. Acoust. Soc. Am.* 90, 1309-1325, 1991
- [3] Chenoukh S., Carré R. & Lindblom B. "Locus equations in the light of articulatory modeling" *J. Acoust. Soc. Am.* 102, 2380-2389, 1997
- [4] Castelli E & Carré R. "Production and perception of Vietnamese vowels" *Interspeech-Eurospeech 2005*, Lisbon, Portugal, 2005
- [5] Pham Thi N. Y., Castelli E. & Nguyen Q.C. "Gabarits des tons vietnamiens". *JEP 2002* Nancy, pp 23-26, 2002.
- [6] Boite R. et al. "*Traitement de la parole*" Presses Polytechniques et Universitaires Romandes, 2000
- [7] Krull D. "2nd formant locus patterns & consonant-vowel coarticulation in spontaneous speech" *Phonetic Experimental Research at the Institute of Linguistics*, Univ. of Stockholm, PERILUS VII, 66-70, 1989.
- [8] Sussman H., Hoemeke K. & Ahmed F. "A cross-linguistic investigation of locus equations as a relationally invariant descriptor for place of articulation" *J. Acoust. Soc. Am.* 94, 1256-1268, 1993.