

Analyse dynamique de la réduction vocalique en contexte CV à partir des pentes formantiques en arabe dialectal et en français.

Jalaleddin AL-TAMIMI

Laboratoire Dynamique du Langage (UMR 5596)

Institut des Sciences de l'Homme, 14 av. Berthelot – 69007 Lyon, France

Mél: Jalal-Eddin.Al-Tamimi@univ-lyon2.fr - <http://www.ddl.ish-lyon.cnrs.fr/>

ABSTRACT

Linear regression parameters (formant slopes and intercepts) are proposed to measure the degree of vowel reduction in 3 vowel systems: Moroccan Arabic, Jordanian Arabic and French. 10 speakers per language produced a list of vowels in C_1VC , C_1VCV , or C_1VC_2VC words, where C_1 or C_2 was /b/, /d/ or /k/. Our results show that the values of formant slopes and intercepts are dependent on: 1) the place of articulation of adjacent consonants, 2) vowel quality, and 3) the language's vowel system density. Discriminant analysis results show the possibility of language separation on the basis of $F1$, and $F2$ slopes, intercept values, and duration.

1. INTRODUCTION

En phonétique, on parle de réduction vocalique (RV) lorsque les valeurs cibles des voyelles (produites isolément) ne sont pas atteintes, selon le contexte. Les voyelles ont tendance à être "réduites" vers une forme *plus ou moins* centralisée (sur les deux axes $F1$ et $F2$) se rapprochant ainsi d'un schwa [ə] (Lindblom [1]). À débit rapide et en position non-accentuée, les voyelles (surtout de type périphérique) ont tendance à être réduites en durée et par conséquent, subissent une RV très marquée (par rapport aux voyelles produites à débit lent et/ou en position accentuée), comme en anglais par exemple (cf. Peterson & Barney [2], Lindblom [1] en suédois, Stevens & House [3], Fourakis [4], Hillenbrand *et al.* [5], etc.). Pour Lindblom [1], ce sont essentiellement les différences de durée (et de contexte consonantique) et non de débit ou d'accentuation qui affectent la RV, en suédois : plus une voyelle est longue, plus la valeur cible est atteinte, et vice-versa. En revanche, Gay 1978 [6] et Pols & van Son [7] montrent que la non-accentuation et les différences de débit favorisent davantage la RV, lorsqu'ils comparent des voyelles accentuées produites à débit rapide et non-accentuées produites à débit lent, ayant par conséquent la même durée. Par ailleurs, Fourakis [4] a montré que le contexte environnant affecte largement les valeurs centrales de formants (et par conséquent la RV), en comparaison avec le débit de parole ou l'accentuation. Autrement dit, des voyelles accentuées ayant des durées variables et produites à débit normal en contexte CV subiront une forme de RV due majoritairement aux effets des consonnes adjacentes et non de la durée. En effet, la durée de la transition vers la voyelle en contexte CV est un invariant dynamique en production des occlusives voisées et non-voisées (≈ 50 ms), Kent & Moll [8]. Au-delà des 50 ms, c'est la durée de la partie stable de la voyelle qui pourrait influencer la RV, qui est favorisée par la coarticulation d'une voyelle avec les consonnes et/ou voyelles adjacentes. D'après Stevens & House [3], les

voyelles produites dans un environnement "nul" (i.e. en contexte [hVd] ou en isolation) ne sont pas réduites, à l'inverse de celles présentes dans les autres contextes consonantiques (cf. Lindblom [1], Stevens & House [3], Öhman [9] Fourakis [4], Hillenbrand *et al.* [5], Al-Tamimi & Ferragne [10], etc.). Cependant, les effets du contexte consonantique sur les voyelles ont été étudiés d'un point de vue statique, c'est-à-dire en ne retenant que les valeurs centrales des formants vocaliques (Stevens & House [3], Fourakis [4], Al-Tamimi & Ferragne [10]). D'un point de vue dynamique, l'étude de l'effet consonantique sur les voyelles en contexte CV a été établie par Lindblom [1], Hillenbrand *et al.* [5] et Fowler [11], entre autres et cela dès le début de la transition vers la voyelle, jusqu'à l'état stable, ou plus loin encore. Hillenbrand *et al.* [5] ont étudié ces effets dynamiques en caractérisant les changements spectraux inhérents aux voyelles. Des mesures acoustiques ont été effectuées à 20% et à 70% de la durée de la voyelle. Les résultats de leur analyse discriminante montrent que le taux de classification correcte est en moyenne 6,10% supérieur lorsque les valeurs de $f0$ et F_{1-3} à 20% et à 70% sont prises en compte par rapport aux valeurs de $f0$ et F_{1-3} à 50% (94,10% pour le premier test et 88% pour le second). La prise en compte de la durée améliore légèrement le taux d'identification (de 6,10% à 6,30%). Dans l'étude proposée par Lindblom [1], l'auteur a investigué la dynamique des formants vocaliques, et plus particulièrement $F2$, en utilisant les pentes de l'équation du locus (LE) en suédois (suivant la formule : $F2_{onset} = m * F2_{milieu} + b$, où m et b représente la pente et l'ordonnée à l'origine, respectivement). Ses résultats montrent que les pentes de LE varient en fonction du lieu d'articulation des consonnes allant de /g/ (0.95) > /b/ (0.69) > /d/ (0.28), et sont considérés depuis comme un bon indicateur du lieu d'articulation. Modarresi *et al.* [12] ont montré que les valeurs de pentes de LE ne varient pas entre /k/ et /g/ (mais varient en contexte bilabial et dental). Fowler [11] explique que LE sert à caractériser, à la fois, le lieu d'articulation et le degré de coarticulation entre les consonnes et les voyelles : une pente forte ($m = 1$) indique une coarticulation maximale entre consonnes et voyelles (i.e. une résistance minimale de la coarticulation), tandis qu'une pente faible ($m = 0$) indique l'absence de coarticulation entre consonnes et voyelles (i.e. une résistance maximale de la coarticulation). Le lien entre le degré de coarticulation et la RV peut être expliqué suivant la relation linéaire entre $F2_{onset}$ et $F2_{milieu}$: les modifications de valeurs de $F2_{milieu}$ affecteront celles de $F2_{onset}$ et par conséquent celles des pentes de LE.

La caractérisation des effets consonantiques sur les voyelles, la RV, d'un point de vue dynamique par les changements spectraux inhérents aux voyelles ou par les

pentades de LE, s'appuient sur des mesures "statiques" à 2 instants de la durée totale de la voyelle. Je propose, dans ce travail, une méthode de caractérisation de la RV d'un point de vue entièrement dynamique, c'est-à-dire, en représentant la transition par sa droite de régression donnant la pente formantique (PF). La transition part de l'onset et se termine au milieu de la voyelle. En ce sens, le calcul de la PF intègre totalement la durée et prends en compte ses variations.

2. MÉTHODOLOGIE

2.1. Langue, locuteurs et corpus

Trois systèmes vocaliques ont été comparés : l'arabe marocain de Casablanca avec 5 voyelles /i: ə a: u u:/ (Hamdi [13]), l'arabe jordanien d'Irbid avec 8 voyelles /i i: e: a: o: u u:/ (Bani Yassin & Owens [14]) et le français avec 11 voyelles orales /i e ε a α ɔ o u y ø œ/ (AM, AJ et FR, respectivement). 10 locuteurs hommes par système, âgés de 20 à 30 ans et ne présentant aucun trouble du langage ni au niveau articulatoire, ni auditif, ont produit une liste de mots ayant des structures syllabiques de type : C₁VC, C₁VCV, ou C₁VC₂VC. C₁ ou C₂ contient l'une des 3 consonnes *phonologiquement* communes aux 3 systèmes /b d k/ ; qui est suivie de la voyelle étudiée. Les items ont été présentés aléatoirement avec 5 répétitions par locuteur, à l'intérieur d'une phrase porteuse (le protocole expérimental a été adapté aux locuteurs arabophones par l'utilisation du système d'écriture de l'arabe standard, sans vocalisation et avec 2 listes différentes adaptées au lexique des deux dialectes). La tâche des locuteurs consistait à produire les voyelles dans des Mots, des Syllabes et en Isolation à débit moyen et en style non-marqué (ex. [bo:se ~ bo: ~ o:] = "bisou" en AJ). Les enregistrements ont été effectués dans une chambre insonorisée et numérisés directement sur un PC avec un taux d'échantillonnage de 22 kHz, 16 bits, mono. Au total, le corpus rassemble 2196 voyelles en AM, 3258 en AJ et 4800 en FR, segmentées manuellement.

2.2. Analyse et traitement de données

Dans cette étude, seules les données correspondant à la réalisation en Mot ont été utilisées, car cette forme se rapprocherait le plus de la situation de parole produite *normalement*, avec une production naturelle de la voyelle. Seules les 3 voyelles /i a u/ (voyelles longues accentuées en AM & AJ) ont été analysées du fait qu'elles sont communes aux 3 systèmes et non intrinsèquement réduites. Des analyses acoustiques fines des 2 premiers formants de chaque voyelle ont été effectuées avec le logiciel Praat en utilisant l'algorithme d'extraction de formants "Burg" (équivalent à une analyse LPC, auto-corrélation, avec 24 coefficients LPC) avec une fenêtre d'analyse gaussienne de 12,5 ms et un pas de déplacement de 5 ms. Les valeurs formantiques de l'onset ont été déterminées en suivant la méthode proposée par Al-Tamimi [15], et correspondent à la valeur distante de 5 ms du début de la transition vers la voyelle. Les valeurs formantiques, extraites toutes les 5 ms, ont été vérifiées manuellement afin de corriger les possibles erreurs d'extraction automatique. Ces valeurs ont été ensuite converties en Bark (suivant la formule proposée par

Schroeder *et al.* [16]) pour procéder à une normalisation de données entre les locuteurs. Les valeurs de PF et de l'ordonnée à l'origine (OO) - m et b respectivement, dans la formule $y = m \cdot x + b$ - sont obtenues par une analyse de régression linéaire. Le calcul a été effectué de l'onset vocalique jusqu'au milieu de la voyelle pour chaque mesure, où les valeurs de x représentent le temps et les valeurs y, les formants. Ainsi, pour chaque voyelle, une série de 2 valeurs de PF et d'OO représentant F1 et F2 est obtenue. Compte tenu des observations rappelées précédemment, on suppose que :

- H1 : La RV sera affectée par le lieu d'articulation de la consonne adjacente, suivant l'évolution suivante : /k/ > /b/ > /d/, d'après les résultats obtenus de l'équation du locus, Lindblom [1],
- H2 : Les valeurs de PF se seront pas les mêmes en fonction des voyelles : sur F1, /a/ aura les valeurs les plus élevées, /i/ et /u/, les plus basses, sur F2, /u/ aura les valeurs les plus élevées, les plus basses pour /i/, et /a/ se situera entre les deux. (Stevens & House [3]),
- H3 : Les valeurs de PF et d'OO entre les 3 systèmes seront différentes, avec une RV moins importante en FR, et plus importante en AM, puisque les langues possédant plus de voyelles présentent une RV moins importante par rapport à celles possédant moins de voyelles, pour conserver la distinction entre elles, (Manuel [17]).

3. RÉSULTATS ET DISCUSSION

Deux types d'analyses statistiques ont été appliqués aux valeurs des PF et d'OO de F1 et de F2 afin de tester les influences sur la RV : 1) une MANOVA à 3 facteurs (langues, consonnes et voyelles), en utilisant NCSS et 2) 2 Analyses Discriminantes (AD) avec validation croisée ayant comme paramètres d'entrée : les valeurs de PF et d'OO de F1 et F2 en incluant et en excluant la durée, en utilisant SPSS. Les valeurs de PF sont positives ou négatives indiquant la direction de la transition (i.e. une PF négative indique une transition descendante vers la voyelle). En ne prenant en compte que les valeurs absolues de PF, on peut tester nos hypothèses. Des valeurs de PF élevées indiquent une RV moindre et vice versa, car les valeurs cibles se rapprochent de celles de l'état stable vocalique.

3.1. RV vs. Lieu d'Articulation des consonnes

Les résultats montrent que le lieu d'articulation affecte la RV (H 1). Les valeurs de PF sur l'axe F1 sont élevées pour /k/, basses pour /b/ et intermédiaires pour /d/ ($p < 10^{-6}$). Sur cet axe, la RV évolue de la façon suivante : /b/ > /d/ > /k/. Sur l'axe F2, les valeurs de PF sont élevées pour /d/, basses pour /b/ et intermédiaires pour /k/ ($p < 10^{-6}$). Sur cet axe, la RV évolue de la façon suivante : /b/ > /k/ > /d/ (figure 1). Les valeurs d'OO sont significativement différentes en fonction du lieu d'articulation de la façon suivante : /b/ > /d/ > /k/ sur F1 ($p < 10^{-6}$) et /d/ > /k/ > /b/ sur F2 ($p < 10^{-6}$), (figure 2). Les résultats de l'AD donnent un taux de classification des consonnes par langue significatif de 66,1% (53,6%, durée exclue) en AM (χ^2 , $p < 10^{-6}$), de 44,5% (41,10%, durée exclue) en AJ (χ^2 , $p < 10^{-6}$) et de 50,6% (43,20%, durée exclue) en FR (χ^2 , $p < 10^{-6}$).

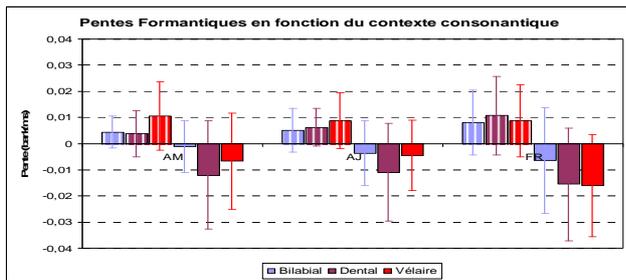


Figure 1 : PF de F1 (barres hachurée) et de F2 (barres pleines) en AM, AJ et FR en fonction des consonnes.

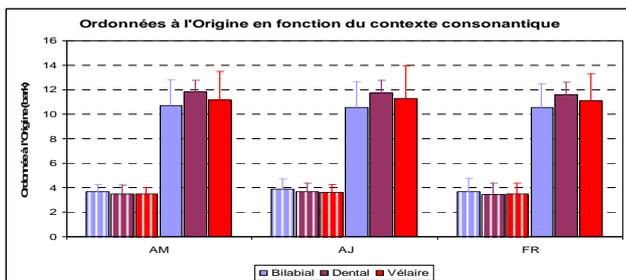


Figure 2 : OO de F1 (barres hachurée) et de F2 (barres pleines) en AM, AJ et FR en fonction des consonnes.

3.2. RV vs. Timbre de Voyelles

Les valeurs de PF, sur l'axe F1, sont élevées pour /a/, basses pour /i/ et intermédiaires pour /u/ ($p < 10^{-6}$). Sur l'axe F2, les valeurs de PF sont élevées pour /u/, basses pour /i/, et intermédiaires pour /a/ ($p < 10^{-6}$). Ceci montre que la voyelle /u/, sur l'axe F2, et la voyelle /a/, sur l'axe F1, subissent le maximum de RV (figure 3).

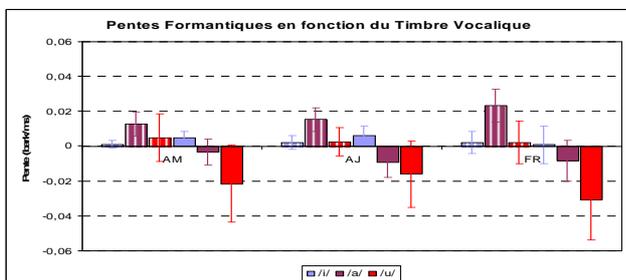


Figure 3 : PF de F1 (barres hachurée) et de F2 (barres pleines) en AM, AJ et FR en fonction des voyelles.

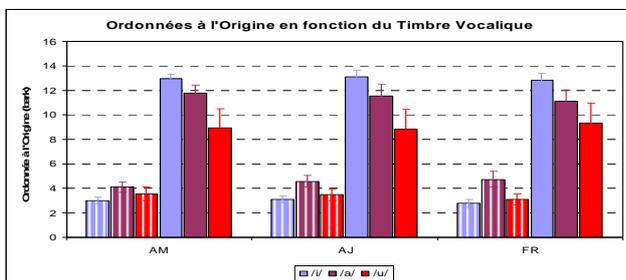


Figure 4 : OO de F1 (barres hachurée) et de F2 (barres pleines) en AM, AJ et FR en fonction des voyelles.

L'étude des effets consonantiques sur les voyelles montrent que, pour les 3 systèmes, la consonne /k/ affecte largement la voyelle /a/ sur l'axe F1 et la consonne /d/ réduit davantage les voyelles /i/ et /u/, sur l'axe F2, ($p < 10^{-6}$), expliquant ainsi une partie de la variabilité.

Les valeurs d'OO sont significativement différentes en fonction des voyelles de la façon suivante : /a/ > /u/ > /i/ pour F1 ($p < 10^{-6}$) et /i/ > /a/ > /u/ pour F2 ($p < 10^{-6}$), (figure 4). Les résultats de l'AD donnent un taux de classification des voyelles par langue très significatif de 99,8% (98,9%, durée exclue) en AM (χ^2 , $p < 10^{-6}$), 98,4% (98,4%, durée exclue) en AJ (χ^2 , $p < 10^{-6}$) et 98,4% (98,4%, durée exclue) en FR (χ^2 , $p < 10^{-6}$).

3.3. RV vs. Différences Translinguistiques

Afin de tester cette hypothèse, des prédictions spécifiques doivent être formulées. En effet à durée identique, sur l'axe F1, une valeur de PF élevée pour /i/ et /u/ et basse pour /a/ indique une centralisation importante. Sur l'axe F2, plus la valeur de PF est basse, plus la voyelle est réduite (i.e. une centralisation pour /i/ et /u/ et antériorisée pour /a/). En observant les figures 3 & 4, toutes ces prédictions sont confirmées. En effet, on observe globalement une RV importante en AM et moins importante en FR, AJ étant intermédiaire ($p < 10^{-6}$). Néanmoins, comme le montre la table 1, il existe des différences importantes au niveau de la durée moyenne de chaque voyelle par langue (calculée de l'onset au milieu temporel). Ceci pourrait affecter le calcul de la PF, car au-delà de 50 ms (durée de transition), c'est la durée de l'état stable qui pourrait affecter la RV.

Table 1 : Durée moyenne et Écart-Type pour chaque voyelle (de l'onset au milieu temporel) par langue.

	/i/		/a/		/u/	
	moyenne	écart-type	moyenne	écart-type	moyenne	écart-type
AM	127,50	32,37	115,07	29,92	78,61	32,38
AJ	110,77	26,72	129,63	21,60	90,41	28,58
FR	76,17	26,20	80,27	12,58	76,24	21,90

Pour pallier ce problème, un second calcul de PF et d'OO a été effectué en normalisant le temps (i.e. en le situant dans un intervalle de 0 à 0,5, où 0 = le début de l'onset et 0,5 = le milieu temporel de la voyelle) afin d'éliminer les différences de durée entre les voyelles longues de l'arabe et les "brèves" du français. Les résultats obtenus montrent que pour /i/, les valeurs de PF sur l'axe F2 sont élevées en AJ et AM et basses en FR ($p < 10^{-6}$). Pour /a/, les valeurs de PF sur les deux axes F1 et F2 sont élevées en AJ, basses en AM et intermédiaires en FR ($p < 10^{-6}$). Pour /u/, les valeurs de PF sur l'axe F2 sont élevées en FR, basses en AM et intermédiaires en AJ ($p < 10^{-6}$) (figures 5). Ceci indique globalement une RV importante en AM, par rapport à AJ et FR, qui présente moins de RV, indiquant l'existence de différentes cibles vocaliques dans les 3 systèmes (cf. également Al-Tamimi & Ferragne [10]), surtout pour la voyelle /a/, où une RV importante est observée en FR par rapport à AJ. Les valeurs d'OO pour la voyelle /i/ sur l'axe F1 sont élevées en AJ, basses en FR et intermédiaires en AM ($p < 10^{-6}$). Sur l'axe F2, elles sont élevées en AJ par rapport à celles en AM et FR (qui se chevauchent). Pour la voyelle /a/ sur l'axe F1, les valeurs d'OO en FR sont élevées, basses en AM et intermédiaires en AJ ($p < 10^{-6}$). En revanche, elles sont élevées en AM, basses en FR et intermédiaires en AJ ($p < 10^{-6}$). Pour la voyelle /u/ sur l'axe F1, AM et AJ présentent les valeurs les plus élevées (qui se chevauchent) et FR, les plus basses ($p < 10^{-6}$). Sur l'axe F2, les valeurs sont élevées en

FR et basses en AJ et AM (les valeurs se chevauchent) ($p < 10^{-6}$), (figures 6).

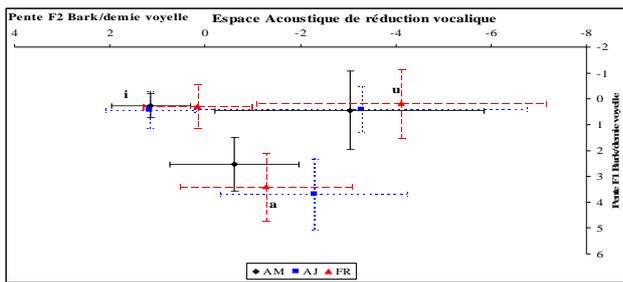


Figure 5 : Espace Acoustique des PF en AM, AJ et FR.

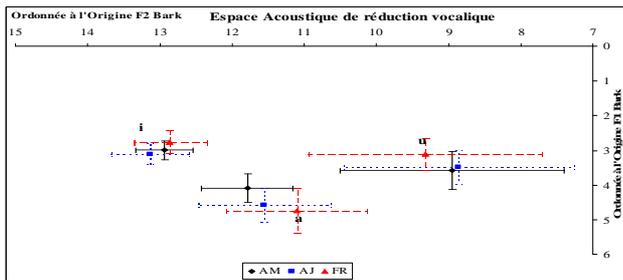


Figure 6 : Espace Acoustique des OO en AM, AJ et FR.

Les résultats de l'AD ont montré qu'il est possible de discriminer les 3 systèmes. En effet, l'analyse donne un taux de classification significatif de 36,7% pour les langues (χ^2 , $p < 10^{-6}$) et 57,40% pour les voyelles (χ^2 , $p < 10^{-6}$). Le taux de discrimination de chaque langue par catégorie vocalique est de : 61,1% pour /i/, 66,1% pour /a/ et 54,3% pour /u/, (χ^2 , $p < 10^{-6}$).

4. CONCLUSION

Dans cette étude, une méthode de mesure de la RV basée sur les valeurs de PF et d'OO obtenues au moyen d'une régression linéaire a été proposée. Les résultats ont montré que les valeurs de PF et d'OO sont influencées par le lieu d'articulation des consonnes adjacentes (H 1), le timbre des voyelles (H 2) et par les différences translinguistiques de type "densité des systèmes" (H 3). L'utilisation de la durée et des valeurs de PF, d'OO, de F1 et de F2 lors d'une analyse discriminante a permis de séparer les 3 systèmes au niveau des consonnes et des voyelles. En excluant la durée du modèle, les taux d'identification restent approximativement les mêmes. La normalisation de la durée à travers les systèmes n'a pas changé les résultats : la RV est plus importante en AM, intermédiaire en AJ et plus basse en FR. Les résultats obtenus montrent qu'il est possible de reconnaître le lieu d'articulation de la consonne adjacente ou le timbre de la voyelle et cela dès le début de la transition. La prochaine étape consistera à analyser les autres voyelles et les autres contextes de réalisation, ainsi qu'à comparer avec de la parole spontanée. L'étude des attentes perceptives des auditeurs de chaque système permettra de mettre en évidence le rôle de ces indices dynamiques en perception.

5. REMERCIEMENT

Je remercie René Carré, François Pellegrino, Emmanuel Ferragne et Christelle Dodane pour leur aide précieuse.

BIBLIOGRAPHIE

- [1] Lindblom, B. On vowel reduction, *Report #29, The Royal Ins. of Tech., Speech Transmission Laboratory, Stockholm, Sweden*, 1963.
- [2] Peterson, G. & Barney, H. Control Methods Used in a Study of the Vowels, *Journal Acoustical Society of America*, Vol. 24: 175-184, 1952.
- [3] Stevens, K. & House, A., Perturbation of vowel articulations by consonantal context: An acoustical study, *Journal of Speech and Hearing Researches*, Vol. 6: 111-128, 1963.
- [4] Fourakis, M. Tempo, Stress and Vowel reduction in American English, *Journal of Acoustical Society of America*, Vol. 90 (4): 1816-1827, 1991.
- [5] Hillenbrand, J., Clark, M. & Nearey, T. Effects of consonant environment on vowel formant patterns, *Journal of Acoustical Society of America*, Vol. 109 (2): 748-763, 2001.
- [6] Gay, T. Effect of speaking rate on vowel formant movements, *Journal of Acoustical Society of America*, Vol. 63 (1): 223-230, 1978.
- [7] Pols, L.; van Son, R. Acoustics and perception of dynamic vowel segments, *Speech Communication*, Vol. 13: 135-147, 1993.
- [8] Kent, R. & Moll, K., Vocal-Tract Characteristics of Stop Cognates, *Journal of Acoustical Society of America*, Vol. 46 (6, part 2): 1549-1555, 1969.
- [9] Öhman, S. Coarticulation in VCV Utterances: Spectrographic Measurements, *Journal of Acoustical Society of America*, Vol. 39: 151-168, 1966.
- [10] Al-Tamimi, J. & Ferragne, E., Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In *Proc. 9th EUROSPEECH*: 2465-2468, 2005.
- [11] Fowler, C., Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation, *Perception & Psychophysics*, vol. 55: 597-610, 1994.
- [12] Modarresi, G., Sussman, H., Lindblom, B., & Burlingame, E. Locus equation encoding of stop place: revisiting the voicing/VOT issue, *Journal of Phonetics*, vol. 33: 101-113, 2005.
- [13] Hamdi, R. Étude phonologique et expérimentale de l'emphase en arabe marocain de Casablanca. *Thèse de Doctorat, Sciences du Langage: Lyon2*: 172, 1991.
- [14] Bani-Yasin, R. & Owens, J. The Phonology of a Northern Jordanian Arabic Dialect, *Zeitschrift der Deutschen Morgenlandischen Gesellschaft*, Vol. 137(2): 297-331, 1987.
- [15] Al-Tamimi, J., 2004, L'équation du locus comme mesure de la coarticulation VC et CV : Étude préliminaire en Arabe Dialectal Jordanien. In *Proc. of 25ème Journée d'Études sur la Parole*, pages 9-12, 2004.
- [16] Schroeder, M., Atal, B., & Hall, J., Optimizing digital speech coders by exploiting masking properties of the human ear, *Journal of the Acoustical Society of America*, Vol. 66: 1647-1652, 1979.
- [17] Manuel, S. Y. (1990). The Role of Contrast in Limiting Vowel-to-Vowel Coarticulation in Different Languages. *Journal of Acoustical Society of America*, Vol. 88(3): 1286-1298, 1990.