

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) were 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_1VCVC words, where C_1 was /b/, /d/ or /k/. Our results show that the values of formant slope and intercept are dependent on: 1) the place of articulation of adjacent consonant, 2) the vowel quality, and 3) the language's vowel system density. Discriminant analysis results show the possibility of language separation on the basis of duration, $F1$ and $F2$ slope and intercept values.

1. INTRODUCTION

Du point de vue phonétique, la réduction vocalique (RV) est définie lorsque les valeurs cibles de référence 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] (Joos [1], Lindblom [2], etc.). Les voyelles non-accentuées (surtout de type périphérique) subissent une RV très marquée par rapport aux voyelles accentuées, en anglais par exemple (cf. Peterson & Barney [3], Lindblom [2], Stevens & House [4], Fourakis [5], Hillenbrand *et al.* [6], Hillenbrand *et al.* [7], etc.). Le débit de parole affecte également les valeurs formantiques. Fourakis [5] et Pickett [8] ont montré qu'en débit rapide, les voyelles ont une durée vocalique beaucoup plus courte qu'en débit normal, et vont être réduites. Lindblom [2] explique que ce sont les différences de durée et non de débit de parole ou d'accentuation qui affectent largement la RV. Pour Gay 1978 [9] et Pols & van Son [10], lorsque les voyelles accentuées et non-accentuées ont la même durée, c'est la non-accentuation qui favoriserait davantage la RV, et non la durée réduite des voyelles non-accentuées. Autrement dit, des voyelles accentuées avec des durées variables ne subiront pas de RV importante.

Par ailleurs, la coarticulation d'une voyelle avec les consonnes et/ou voyelles adjacentes favorise la RV. En effet, différents travaux ont montré que les voyelles produites dans des environnements non-"nul" (autres qu'en contexte [hVd] en anglais, ou en isolation, Stevens & House [4]) subissent une forme de RV comparable à

celle due à la non-accentuation ou aux débits de parole rapides (cf. Lindblom [2], Stevens & House [4], Öhman [11], Fourakis [5], Hillenbrand *et al.* [6], Hillenbrand *et al.* [7], Al-Tamimi [12], Al-Tamimi & Ferragne [13], etc.). Cependant, les effets de la coarticulation sur les voyelles ont été étudiés d'un point de vue statique, c'est-à-dire en ne retenant que les valeurs centrales de formants de ces voyelles.

D'un point de vue dynamique, certains travaux ont étudié cet effet dès le début de la transition vers la voyelle, jusqu'à l'état stable, ou plus loin encore. Les travaux proposés par Hillenbrand *et al.* [6] et par Hillenbrand *et al.* [7] concernent l'étude des changements spectraux inhérents aux voyelles ; i.e. l'évolution formantiques des voyelles. Des mesures acoustiques ont été effectuées à 20% et à 80% (à 70% pour Hillenbrand *et al.* [7]) de la durée de la voyelle. Des tests de classification discriminante ont été appliqués aux données. Les résultats montrent que la durée vocalique et les changements spectraux inhérents aux voyelles apportent plus de précisions à l'analyse discriminante : le taux d'erreur de classification est proche de 11% par rapport au taux d'erreur de 15% lorsque les valeurs formantiques statiques sont utilisées.

Les deux études proposées par Lindblom [2] et Al-Tamimi [12] (entre autres), concernent la caractérisation dynamique des formants vocaliques et plus particulièrement $F2$. Les auteurs ont étudié les pentes des équations du locus (suivant la formule : $F2_{onset} = m * F2_{milieu} + b$, où m et b représente la pente et l'ordonnée à l'origine, respectivement) afin de séparer le lieu d'articulation des consonnes et de caractériser, par la suite, le degré de coarticulation entre les consonnes et les voyelles adjacentes, et le degré de RV. Quand les pentes sont fortes, la coarticulation de la consonne avec la voyelle suivante est maximale ; le degré de coarticulation et la RV sont forts ($m = 1.0$, i.e. résistance minimale de coarticulation lors de l'articulation de la consonne vers la voyelle). À l'inverse, des pentes faibles indiquent l'absence d'effets de coarticulation des voyelles ; le degré de coarticulation et la RV sont faibles ($m = 0$, i.e. résistance maximale de la coarticulation lors de production CV). Leurs résultats montrent que les pentes de l'équation du locus varient en fonction du lieu d'articulation des consonnes adjacentes allant de /g/ (0.95) > /b/ (0.69) > /d/ (0.28).

Les 4 dernières études citées, qui tendent à caractériser la RV d'un point de vue dynamique, s'appuient sur des mesures "statiques" à 2, voire 3 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, la pente obtenue n'est pas exactement la pente moyenne de la transition mais tient également compte des variations de pente.

2. MÉTHODOLOGIE

2.1. Langue, locuteurs et corpus

Trois systèmes vocaliques ont été comparés : l'arabe marocain avec 5 voyelles /i: ə a: u u:/ (Hamdi [14]), l'arabe jordanien avec 8 voyelles /i i: e: a a: o: u u:/ (Bani & Owens [15]) et le français avec 11 voyelles orales /i e ε a ɑ 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 : C_1VC , C_1VCV , ou C_1VCVC . C_1 est l'une des 3 consonnes *phonologiquement* communes aux 3 systèmes /b d k/ ; elle 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). La tâche des locuteurs consistait à produire les voyelles dans des Mots, des Syllabes et en Isolation (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. Le choix de ce type de réalisation est motivé par le fait que cette forme se rapprocherait le plus de la situation de parole produite *normalement*, i.e. avec une articulation naturelle de la voyelle. Le choix a porté sur l'analyse des 3 voyelles communes dans les 3 systèmes /i a u/ (voyelles longues accentuées en AM & AJ), afin de comparer des voyelles 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 extraites

ont été converties en Bark (d'après la formule proposée par Schroeder *et al.* [16]). Les valeurs formantiques de l'onset ont été déterminées en suivant la méthode proposée par Al-Tamimi [11], et correspondent à la valeur distante de 5 ms du début de la transition vers la voyelle. Toutes les valeurs formantiques des voyelles ont été vérifiées manuellement afin de corriger les possibles erreurs d'extraction automatique.

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 du début de la transition 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 (Al-Tamimi [12]),
- H2 : La RV ne sera pas la même en fonction des voyelles : sur $F2$, /u/ sera la voyelle la moins réduite, /i/ la plus réduite, et /a/ se situera entre les deux. Sur $F1$, c'est la voyelle /a/ qui aura les valeurs les plus élevées (d'après les observations de Peterson & Barney [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. En effet, les langues possédant plus de voyelles subissent un degré de coarticulation moindre par rapport à celles possédant moins de voyelles, afin de conserver plus de distinctivité 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) une analyse discriminante avec validation croisée ayant comme paramètres d'entrée : les valeurs de PF et d'OO de $F1$ et $F2$ et 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 basses indiquent une RV importante et vice versa. Ceci peut s'expliquer par le fait que, plus les valeurs de l'état stable sont proches des valeurs de référence, plus la valeur de la PF est élevée.

3.1. Réduction Vocalique et lieu d'Articulation

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/ est ($p < 10^{-6}$). Sur cet axe, la RV évolue de la façon suivante : /b/ > /k/ > /d/ (cf. figure 1).

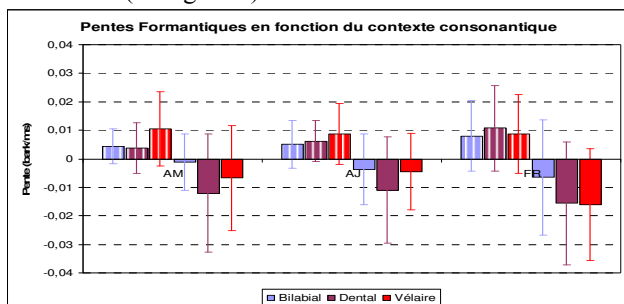


Figure 1 : PF de $F1$ (barres hachurée) et de $F2$ (barres pleines) en AM, AJ et FR en fonction des consonnes.

Les valeurs d'OO sont significativement différentes en fonction du lieu d'articulation consonantique de la façon suivante : /b/ > /d/ > /k/ sur $F1$ ($p < 10^{-6}$) et /d/ > /k/ > /b/ sur $F2$ ($p < 10^{-6}$), (cf. figure 2).

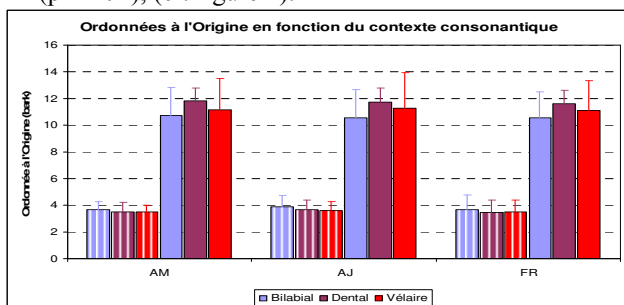


Figure 2 : OO de $F1$ (barres hachurée) et de $F2$ (barres pleines) en AM, AJ et FR en fonction des consonnes.

Les résultats de l'analyse discriminante donnent un taux de classification significatif de 66,10% (53,60%, durée exclue) en AM (χ^2 , $p < 10^{-6}$), de 44,50% (41,10%, durée exclue) en AJ (χ^2 , $p < 10^{-6}$) et de 50,60% (43,20%, durée exclue) en FR (χ^2 , $p < 10^{-6}$).

3.2. Réduction Vocalique et timbre de Voyelles

La RV dépend du timbre vocalique (H 2). Nos résultats montrent que les valeurs de PF, sur l'axe $F1$, sont élevées pour /a/, basses pour /i/ et intermédiaires pour /u/ ($p < 10^{-6}$). Ainsi sur cet axe, la RV évolue comme suivant : /i/ > /u/ > /a/. Sur l'axe $F2$, les valeurs de PF sont élevées pour /u/, basses pour /i/, et intermédiaires pour /a/ ($p < 10^{-6}$). Ainsi sur cet axe, la RV évolue de la façon suivante : /i/ > /a/ > /u/ (cf. figure 3).

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 $F2$ ($p < 10^{-6}$), (cf. figure 4).

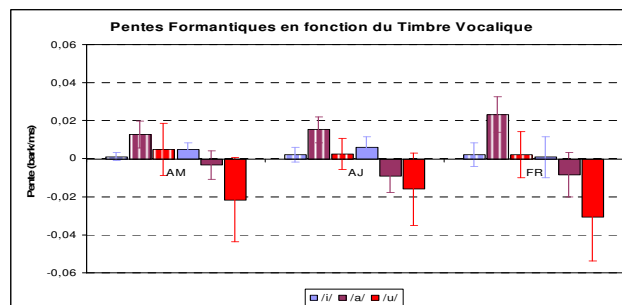


Figure 3 : PF de $F1$ (barres hachurée) et de $F2$ (barres pleines) en AM, AJ et FR en fonction des voyelles.

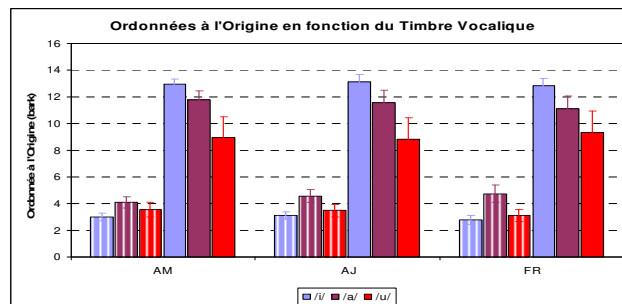


Figure 4 : OO de $F1$ (barres hachurée) et de $F2$ (barres pleines) en AM, AJ et FR en fonction des voyelles.

Les résultats de l'Analyse Discriminante donnent un taux de classification très significatif de 99,80% (98,90%, durée exclue) en AM (χ^2 , $p < 10^{-6}$), 98,40% (98,40%, durée exclue) en AJ (χ^2 , $p < 10^{-6}$) et 98,40% (98,40%, durée exclue) en FR (χ^2 , $p < 10^{-6}$).

3.3. Réduction Vocalique et Différences Translinguistiques

La RV est affectée par la densité des systèmes (H 3). Nos résultats montrent que pour /a/, les valeurs de PF sont élevées en FR, basses en AM et intermédiaires en AJ ($p < 10^{-6}$). Pour /i/, les valeurs de PF sont élevées en AJ, basses en FR et intermédiaires en AM ($p < 10^{-6}$). Pour /u/, les valeurs de PF sur l'axe $F2$ sont élevées en FR, basses en AJ et intermédiaires en AM, mais plus basses en AM sur $F1$ ($p < 10^{-6}$) (cf. figures 3 & 5).

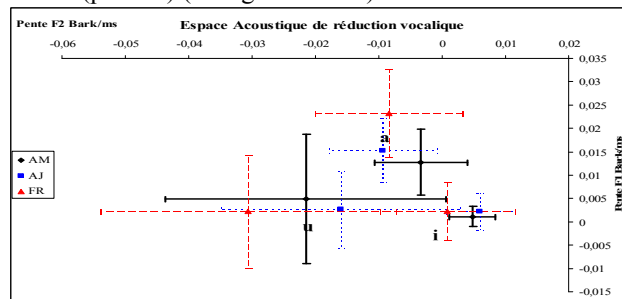


Figure 5 : Espace Acoustique des PF en AM, AJ et FR.

Les valeurs d'OO pour la voyelle /a/ sont élevées en FR, basses en AM et intermédiaires en AJ ($p < 10^{-6}$). Pour la voyelle /i/, les valeurs d'OO sont élevées en AJ, basses FR et intermédiaires en AM ($p < 10^{-6}$). Pour la voyelle /u/,

AM présente les valeurs les plus élevées, FR, les plus basses et AJ des valeurs intermédiaires ($p < 10^{-6}$), (cf. figures 4 & 6).

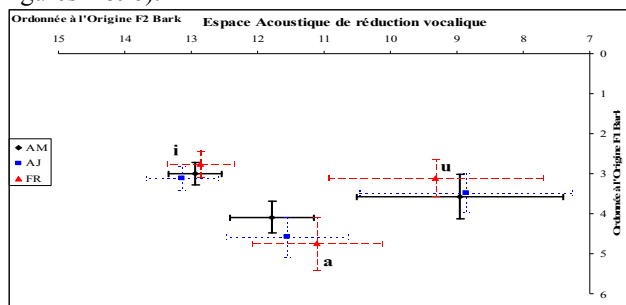


Figure 6 : Espace Acoustique des OO en AM, AJ et FR.

Les résultats de l'Analyse Discriminante ont montré qu'il est possible de discriminer les 3 systèmes. En effet, l'analyse donne un taux de classification significatif de 24,90% (15,40%, durée exclue) pour les consonnes (χ^2 , $p < 10^{-6}$), 64,90% (61,90%, durée exclue) pour les voyelles (χ^2 , $p < 10^{-6}$) et 47,20% (38,70%, durée exclue) pour les langues (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. 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 la comparaison avec de la parole spontanée.

5. REMERCIEMENT

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

BIBLIOGRAPHIE

- [1] Joos, M. Acoustic Phonetics, *Language*, Vol. 24: 1-136, 1948.
- [2] Lindblom, B. On vowel reduction, *Report #29, The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, Sweden*, 1963.
- [3] Peterson, G. & Barney, H. Control Methods Used in a Study of the Vowels, *Journal of Acoustical Society of America*, Vol. 24: 175-184, 1952.
- [4] 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.
- [5] Fourakis, M. Tempo, Stress and Vowel reduction in American English, *Journal of Acoustical Society of America*, Vol. 90 (4): 1816-1827, 1991.
- [6] Hillenbrand, J., Getty, L., Clark, M. & Wheeler, K. Acoustic Characteristics of American English vowels, *Journal of Acoustical Society of America*, Vol. 99: 3099-3111, 1995.
- [7] 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.
- [8] Pickett, J. The flow of speech, In *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*, Ch. 10: 142-150, 1998.
- [9] Gay, T. Effect of speaking rate on vowel formant movements, *Journal of Acoustical Society of America*, Vol. 63 (1): 223-230, 1978.
- [10] Pols, L.; van Son, R. Acoustics and perception of dynamic vowel segments, *Speech Communication*, Vol. 13: 135-147, 1993.
- [11] Öhman, S. Coarticulation in VCV Utterances: Spectrographic Measurements, *Journal of Acoustical Society of America*, Vol. 39: 151-168, 1966.
- [12] 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.
- [13] Al-Tamimi, J. & Ferragne, E., Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In *Proc. of the 9th EUROSPEECH*, pages 2465-2468, 2005.
- [14] Hamdi, R. Étude phonologique et expérimentale de l'emphase en arabe marocain de Casablanca. *Thèse de Doctorat en Sciences du Langage: Lyon2, 1991, p.172*.
- [15] Bani-Yasin, R. & Owens, J. The Phonology of a Northern Jordanian Arabic Dialect, *Zeitschrift der Deutschen Morgenlandischen Gesellschaft*, Vol. 137(2): 297-331, 1987.
- [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.