STATIC AND DYNAMIC CUES IN VOWEL PRODUCTION: A CROSS DIALECTAL STUDY IN JORDANIAN AND MOROCCAN ARABIC.

Jalaleddin Al-Tamimi

Laboratoire dynamique du Langage – CNRS – Université Lyon 2 (UMR 5596)
14, Avenue Berthelot – 69007 Lyon – France
Jalal-Eddin.Al-Tamimi@univ-Lyon2.fr

ABSTRACT

The aim of this paper is to examine the role of dynamic cues (i.e. formant slopes obtained from a linear regression analysis) in comparison with static one (i.e. vowel targets) in the classification of Jordanian and Moroccan vowels, using Discriminant Analysis. 10 speakers per dialect produced a list of vowels in C_1VC_2, C_1VC_2V, or C_1VC_2VC words, where C_1 and C_2 were either /b/, /d/, /dˤ/ or /k/, and V, each vowel. Results show the possibility of vowel separation between both dialects for a specific consonantal environment. Using dynamic cues improves the correct classification rates of about 5% for Moroccan Arabic and 13% for Jordanian Arabic.

Keywords: Arabic dialects, vowel production, formant slopes, vowel targets, classification.

1. INTRODUCTION

Vowel targets, produced in isolation, are considered as the canonical form of vowels ([7], among others). However, they must be considered as a *"Laboratory Artefact"* [8], because: 1) vowels are mostly produced in coarticulation with consonants according to various syllabic structures, and 2) vowel formants are highly instable due to intra- & inter-individual variability. Some researchers ([13], among others) have described vowels produced in isolation as different from those produced in context, concluding that listeners use different cues to identify vowels in isolation or in context. Thus, they have considered these isolated vowels as "useless" for the identification~discrimination experiments and that dynamic information (formant movements and transitions) are more useful in speech perception.

The aim of this paper is to evaluate the role of static and dynamic cues in the classification of Arabic vowels by Discriminant Analysis. One of the motivations of this work is that the morphological structure of Arabic (a non-concatenative language with a triconsonantal root that exhibits direct consonant~consonant relations [10, 11]) implies that vowels never occur in isolation. We have shown that Arabic speakers have difficulties to produce and perceive vowels in isolation. Preliminary results show that dynamic cues (formant transitions) improve the perception of Arabic vowels ([3]).

We propose to compare the vowel systems of two Arabic dialects: Jordanian and Moroccan Arabic in terms of their static and dynamic representations. The static one is a description of vowel targets at the temporal mid-point; the dynamic one is a representation of vowels by their formant slopes, calculated from onset to temporal mid-point, and obtained from a linear regression analysis. The evaluation of dynamic cues role will be conducted in the basis of vowel classification by Discriminant Analysis. The next step of this research will be to examine the role of these dynamic cues in perception [3].

2. METHOD

2.1. Speech Material

Jordanian Arabic with /i iː eː a aː oː u uː/ ([4]) and Moroccan Arabic with /iː aː ə u uː/ ([6])), (JA & MA, henceforth) were compared. 10 male speakers per dialect (aged 20 to 30) recorded a list of vowels in C_1VC_2, C_1VC_2V, and C_1VC_2VC, where C_1 and C_2 were either /b/, /d/, /dˤ/ or /k/, and V, each vowel. The items were randomly presented 5 times in an adapted carrier sentence (the Modern Standard Arabic script was used without vocalization). The speakers were asked to produce these items with normal rate and non marked style. Recordings were made in a sound-attenuated room, on a PC, with 22050 Hz, 16 bits, mono. We ended up with 986 vowels for MA, and 1432 for JA (JA /i u/ in the /k/ context, and /oː/, in /dˤ/, were not produced by speakers due to technical problems).

2.2. Data Analysis

Data were segmented manually and measurements of the first 3 formant frequencies were carried out with Praat [5], using the "Burg" algorithm with a 12.5ms Gaussian window, and a 5ms step. Formant values extracted every 5 ms were verified manually to prevent automatic error extraction values, and then converted to Barks using the formula proposed by [12], to normalize between speakers.

2.2.1. Static Cues

Formant values at the temporal mid-point were determined to represent vowel targets. Means, standard deviations, and vowel space areas (Convex Hull) were calculated for each vowel by place of articulation.

2.2.2. Dynamic Cues

Formant slope values were obtained from a linear regression analysis, from the onset to the vowel's temporal mid-point (to prevent C_2 effects on vowels). Onset values were determined following the method proposed by [1]: the formant value is measured 5 ms after the vowel transition release. Formant slope (m) and intercept (b) values for each formant were obtained from the formula:

(1) $$F_{ormant} = m * D_{uration} + b$$

Formant slope values obtained are positive or negative; indicating the direction of the transition (i.e. negative value indicates a descending transition to the vowel), and absolute values indicate their steepness: high values indicating a steeper transition.

2.3. Statistical Analysis

To evaluate the importance of static or dynamic cues in the description of Arabic vowels, we conducted 2 types of statistical analysis: 1) a MANOVA with 4 factors (Dialect, Speaker, Consonant, and Vowel), and 2) a cross-validation Discriminant Analysis, where the 3 formant values (without duration) were used for the static cues; and the coefficients of the linear regression (i.e. slope and intercept values) for each formant, plus the slope duration were used for dynamic cues. Discriminant Analysis is used to evaluate to what extent dynamic cues improve vowel separation within and between the two Arabic dialects, and to assess the validity of our results.

3. RESULTS

3.1. Static cues

JA & MA vowels dispersion by consonants are presented in figures 1 & 2. Results show that the consonants' place of articulation affects JA & MA vowels on both axes: on F1, for JA: $F(3)=38.5$; $p<0.001$, and for MA: $F(3)=320.5$; $p<0.001$; on F2, for JA: $F(3)=195.5$; p<0,001, and for MA: $F(3)=342.4$; $p<0.001$. JA & MA vowel dispersion areas obtained from the Convex Hull method indicate the effect of consonants' place of articulation on the vowel dispersion, (see figure 3). MA vowels dispersion is influenced as follows: /k/≥/d/>/b/>/dˤ/, whereas for JA, the effects are as follows: /k/>/dˤ/>/d/>/b/.

Figure 1: JA vowels by place of articulation.

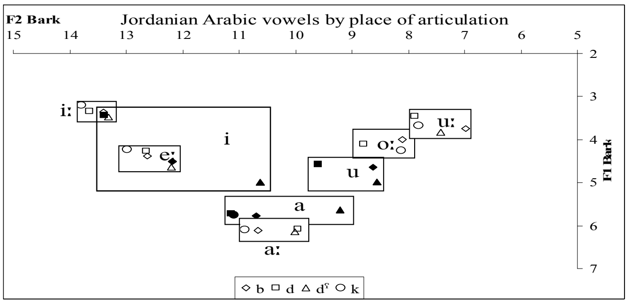


Figure 2: MA vowels by place of articulation.

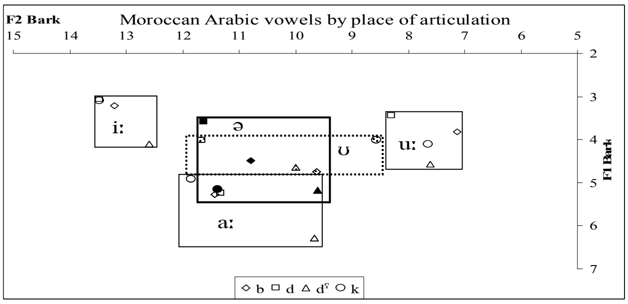


Figure 3: JA & MA vowel areas by place of articulation.

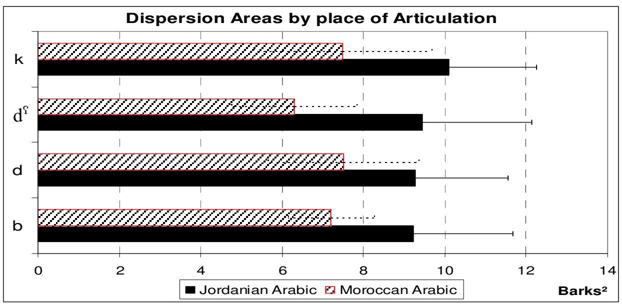


Table 1: MA & JA vowel areas (Barks²) by place of articulation (grey cases indicate missing data).

	MA					JA							
	iː	aː	ə	u	uː	iː	i	eː	aː	a	oː	u	uː
b	0,65	1,39	0,42	4,23	0,94	0,53	1,17	0,77	1,94	1,05	1,28	0,73	1,51
d	0,58	1,10	1,16	0,62	0,59	0,75	0,73	1,19	1,79	0,93	1,50	0,52	0,83
dˤ	0,82	0,76	0,77	2,60	1,62	0,64	0,78	1,06	0,69	2,75		0,65	1,02
k	0,45	1,03	2,61	1,31	1,20	0,49		1,29	1,57	0,67	0,77		2,17

We have also calculated the dispersion areas per vowel (see table 1). Areas dispersion for long vowels indicate that /iː/, in both dialects, presents a lesser area dispersion, while /aː/ & /uː/, an intermediary one. Due to a high degree of inter-speaker variability, MA /ə/ & /ʊ/ have a higher dispersion areas. The results of the Discriminant Analysis indicate the possibility to distinguish JA & MA vowels for each consonantal environment: 44.2% (Box's M=159.6; $p<0.001$) for MA and 32.9% (Box's M=110.8; $p<0.001$) for JA. Although these rates are not so high (i.e. a high confusion degree), they can help us to evaluate the role of dynamic cues. Rates of correct classification of MA & JA vowels per consonant are presented in table 2.

Table 2: Rates of correct classification of MA & JA vowels by consonants, (Box's M; $p<0.001$).

	/b/	/d/	/dˤ/	/k/
MA	82,70%	83,50%	80,40%	75,00%
JA	68,10%	69,70%	83,20%	78,40%

Confusions in the classification are due to the merging of MA's /ə/ & /ʊ/ on the one hand, and the proximity of the JA's /i u/ to /eː oː/, respectively, on the other hand. Results (see figures 1 to 3) show some differences between the two dialects: the MA vowel system is more reduced (i.e. centralized in both axes) than JA's ($p<0.001$), with a correct classification rate of 54.9% (Box's M=104.9; $p<0.001$), between JA & MA vowels. It was also possible to discriminate both dialects as a function of consonants' place of articulation: 56.1% for /b/ (Box's M=28.7; $p<0.001$), 62.5% for /d/ (Box's M=45.6; $p<0.001$), 49.6% for /dˤ/ (Box's M=49.6; $p<0.001$) & 56.3% for /k/ (Box's M=28.5; $p<0.001$).

3.2. Dynamic cues

Here, we characterize MA & JA vowels by their transitions from the onset to the temporal mid-point by a linear regression analysis. Results obtained indicate the dependency of formant slopes & intercepts on the place of articulation of adjacent consonants, on vowels, & on formants (see figures 4 & 5). This representation is based on the intercept values derived at two points: 0ms, and the temporal mid-point for each formant. The line associating both values indicates the steepness of the slope. From the linear regression coefficients, we can derive formant shifts from the onset to the temporal mid-point. For example, JA /iː/ in /dˤ/ environment with a slope duration of 117ms has these formant shifts: -0.41, 2.85 & 0.22 Barks for F_1, F_2 & F_3 respectively (see table 3 & figure 4).

Table 3: Slope, Intercept & Formant shift for F1, F2 & F3 of JA /iː/ in /dˤ/ environment.

	F1	F2	F3
Slope	-0,003	0,024	0,002
Intercept	3,803	11,126	14,898
Formant shift	-0,410	2,852	0,215

In figure 4, we present an example of F_1, F_2 & F_3 slopes for JA /iː/ as a function of the 4 consonants (see <image_file_1.jpg>, and <image_file_2.jpg> for JA, and MAs vowels).

Figure 4: JA formant slopes for /iː/.

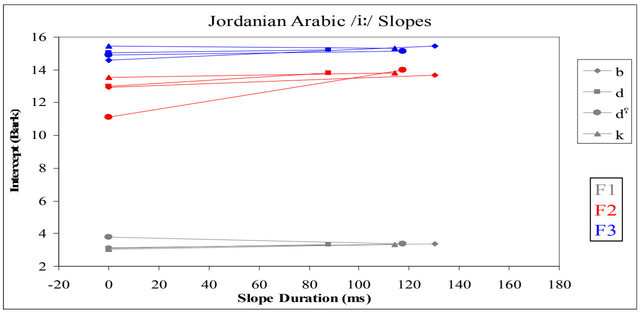
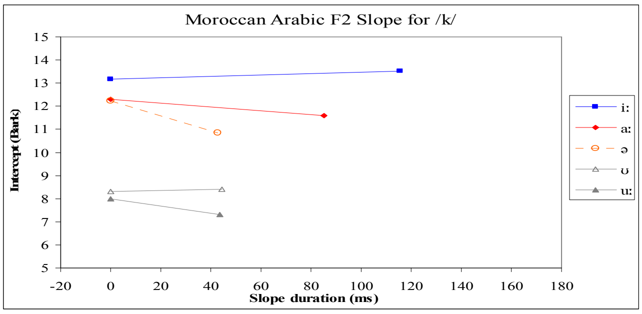


Figure 5: MA F2 formant slopes (in /k/ environment).



In figure 5, we present MA F2 formant slopes in /k/, respectively (for other examples, see <image_file_3.jpg>, <image_file_4.jpg>, & <image_file_5.jpg> for JA F_1, F_2 & F_3 formant slopes per consonant, respectively, and <image_file_6.jpg>, <image_file_7.jpg>, & <image_file_8.jpg> for MAs'). We observe in these examples the direct effects of the place of articulation on the steepness of formant slopes, on the values of the derived intercepts, and on the formant targets. We tested the effects of these differences with Discriminant Analysis. The rates of correct classification indicate the possibility to distinguish JA & MA vowels for each consonantal environment: 52.7% for MA (Box's M=1104.6; p<0.001), and 54.3% for JA (Box's M=1704.2; p<0.001). We present in table 4 the rates of correct classification of MA & JA vowels per consonant.

Table 4: Rates of correct classification of MA & JA vowels by consonants, (Box's M; p<0.001).

	/b/	/d/	/d^ʕ/	/k/
MA	91,20%	88,30%	76,00%	87,20%
JA	87,10%	86,10%	89,00%	92,20%

We observe differences between the linear regression coefficients (slopes and intercepts) for MA & JA: lesser values for both formant slopes and intercepts, indicating a more reduced onset and target values in MA. The rates of correct classification indicate the possibility to discriminate MA & JA vowels with a correct classification rate of 58.5% (Box's M=1558.8; p<0.001). The discrimination between MA & JA as a function of the consonants' place of articulation was possible, with correct classification rates of: 58.5% for /b/ (Box's M=690.7; p<0.001), 63.5% for /d/ (Box's M=611.9; p<0.001), 78.0% for /d^ʕ/ (Box's M=424.2; p<0.001) & 62.5% for /k/ (Box's M=642.9; p<0.001).

4. CONCLUSION & DISCUSSION

We proposed in this paper an evaluation of the role of static and dynamic cues in the classification of MA & JA vowels. Results obtained in both dialects show significant differences in vowels dispersion as a function of the consonants' place of articulation. The use of static cues permitted discrimination of the vowels of MA & JA as a function of consonants' place of articulation (with an average rate of correct classification of 80.4% for MA and 74.85% for JA), and between the two dialects. Dynamic cues improved the visualization of formant shifts, and the correct classification rates of the vowels of each dialect as a function of consonants' place of articulation (with an average rate of correct classification of 85.68% for MA and 88.6% for JA), and between both dialects. We observe an improvement of the correct classification rates of the Discriminant Analysis, when dynamic cues are proposed: an average of 5% for MA and of 13% for JA. The differences observed between JA & MA may be explained in terms of vowel systems density (see [2] & [3]). As these results indicate that dynamic cues improve vowel separation in both MA & JA, we suppose that these dynamic cues may help Arabic speakers to perceptually discriminate, between the vowels of their system. The next step will be to assess the validity of these results in perception. First results indicate that dynamic cues used in a MOA paradigm [9] facilitate Arabic vowels' identification by native Arabic speakers (see [3]).

5. AKNOWLEDGMENT

We are grateful to René Carré, Egidio Marsico, John Ohala and François Pellegrino for their help.

6. REFERENCES

[1] Al-Tamimi, J. 2004. L'équation du locus comme mesure de la coarticulation VC et CV : Étude préliminaire en Arabe Dialectal Jordanien. *Proc. 25ème Journée d'Études sur la Parole*, 9-12.

[2] Al-Tamimi, J. & Ferragne, E. 2005. Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In *Proc. 9th EUROSPEECH*, 2465-2468.

[3] Al-Tamimi, J. (2007, to appear). *Indices dynamiques et perception des voyelles : étude translinguistique en arabe dialectal et en français*. Unpublished PhD Dissertation, University Lyon 2.

[4] Bani-Yasin, R., Owens, J. 1987. The Phonology of a Northern Jordanian Arabic Dialect. *Zeitschrift der Deutschen Morgenlandischen Gesellschaft,* 137(2), 297-331.

[5] Boersma, P., Weenink, D. 2006. Praat. Doing Phonetics by Computer. version 4.4.34.

[6] Hamdi, R. 1991. *Étude phonologique et expérimentale de l'emphase en arabe marocain de Casablanca*. PhD. Dissertation, Université Lyon 2, 172.

[7] Joos, M.A. 1948. Acoustic phonetics. *Language* 24 (Suppl.), 1-136.

[8] Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychology Revue*, 74, 431-461.

[9] Johnson, K., Flemming, E., Wright, R. 1993. The hyperspace effect: Phonetic targets are hyperarticulated, *Language,* 69, 505-528.

[10] McCarthy, J. 1979. *Formal Problems in Semitic Phonology and Morphology*. Ph.D. dissertation, MIT.

[11] Pierrehumbert, J. 1992. Dissimilarity in the Arabic Verbal Roots. *Proc. 23rd Meeting of the Northeastern Linguistic Society, Graduate Student Association, U. Mass. Amherst.* 367-381.

[12] Schroeder, M.R., Atal, B.S., Hall, J.L. 1979. Optimizing digital speech coders by exploiting masking properties of the human ear. *J. Acoust. Soc. Am*. 66, 1647-1652.

[13] Strange, W. 1989. Evolving Theories of Vowel Perception", *J. Acoust. Soc. Am*. 85(5), 2081-2087.