Vowel Epenthesis, Acoustics and Phonology Patterns in Moroccan Arabic

Azra N. Ali [1], *Mohamed Lahrouchi* [2], *Michael Ingleby* [1]

[1] School of Computing and Engineering, University of Huddersfield, Huddersfield, England.
[2] CNRS, Paris-8, Paris, France.

`a.n.ali@hud.ac.uk, Mohamed.lahouchi@univ-paris8.fr, m.ingleby@hud.ac.uk`

Abstract

In Moroccan Arabic it is widely accepted that short vowels are mostly elided, resulting in consonant clusters and consonant geminates. In this paper we present evidence from our exploratory timing study that challenges this widely accepted principle. We work with minimal pairs of singleton consonants vs. geminates (e.g. /bka/ vs. /bəkka/) that reveals a presence of a vowel insertion between the clusters in word initial position in singleton cases. The length of the vowel insertion (epenthetic vowel) and silent pause of the stop consonant is greater than of a noise material. The epenthetic vowel is present in isolated words and in sentence context too. In this paper we also provide phonetic correlates in the minimal pairs - between epenthetic vowel and lexical vowel, between singleton and geminate consonants, and contrast these with other Arabic dialect phonetic timing studies.

Index Terms: Moroccan Arabic, acoustic analysis, geminates

1. Introduction

In this paper we investigate epenthetic vowel insertion between onset consonant clusters and some of the temporal characteristics of consonants in minimal pairs (singleton and geminates) and the surround vowels in Moroccan Arabic (MA). Little attention has been devoted to such research. MA, a dialect of Arabic is quite different from Modern Standard Arabic (MSA). Often Arabic speakers from the Middle-East do not recognize what is being said by a Moroccan speaker. This is because Moroccan Arabic is influenced by French and Berber dialects such as Tashlhiyt (rich in consonant clusters, even vowel-less words). In Moroccan Arabic the short vowels /a/ and /i/ of an 'underlying' *fusha* are mostly elided but /u/ can be retained depending on whether or not the following consonant is labial. This vowel elision accounts for a number of consonant cluster sequences, word initial, word medial and word final. MSA does not allow word initial consonant clusters of two or more segments, but such clustering is widely attested in MA, and must be licit for any realistic set of generative rules. In fact, even word-initial geminates are attested in MA, as well as a copious supply of word-internal clusters inherited from MSA and loan-words. It is not clear whether or not these clusters and geminates with diverse origins behave in the same way. In particular, is not clear whether any of these attested clusters behave like those of MSA.

There are few phonetic timing studies in Arabic dialects that have investigated minimal pairs, singleton vs. geminates (vowels and consonants) in equivalent surroundings: Iraqi Arabic [1], Jordanian Arabic [2], Lebanese and Palestinian Arabic [3, 4] and Saudi Arabic [5, 6]. These studies have shown that duration plays an important role in distinguishing singleton and geminates consonants and their impact on the duration of the surrounding vowels.

The most interesting early findings were made by Baothman et al. [5,6], who confirmed empirically in a large corpus of spoken Saudi Arabic that it has no branching onsets. She also found evidence in recordings and spectrograms that a phonologically active *sukuun* is present between clusters. For example, in the final cluster of Arabic بَحْر (baḥr = 'sea') a schwa vowel /ə/ can be heard in some speech [5, 6, 7] and even detected as vowel material in spectrograms. Baothman's evidence for the phonological reality of orthographic *sukuun* in clusters is; firstly, triggering of devoicing processes at word-internal cluster sites, and secondly a detectable lengthening of clusters and geminates by a unit of short-vowel duration. The spectrogram in Figure 1 below, shows an epenthetic vowel within the consonant cluster /ʁð/ in the form of a segment with phonation at central vowel frequencies, between consonant clusters and in geminate consonants.

Baothman has developed a modern representation of Arabic speech patterns using element phonology and a coda-less constituent structure, and has been able to model the well-known coarticulation processes of Arabic using this framework. She derived a stress-prediction algorithm that is much simpler than those based on the syllabification in the western classical tradition,
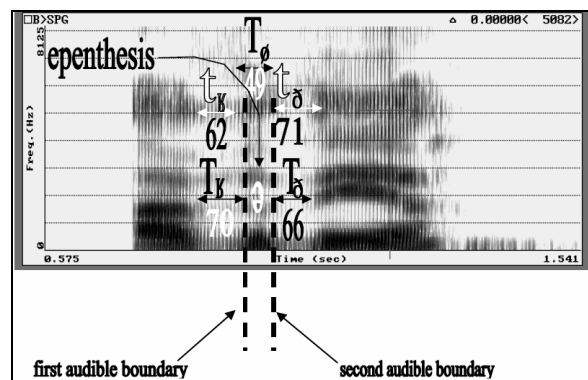

Figure 1: An epenthetic vowel material [6:.90].

A more recent study by Gouskova and Hall [4] has shown an epenthetic vowel between consonant clusters in Lebanese and Palestinian Arabic. They used 30 minimal/near minimal pairs and have shown visible epenthetic vowel occurring within word final consonant clusters; for e.g./bikr/→ /bikər/, /rikb/→ /rikəb/ etc..

The duration of the epenthetic vowel varied across words and speakers. But, in summary, they have shown that epenthetic vowels are phonetically further back than lexical vowels, but no significant difference were found between the duration of the epenthetic vowel and lexical vowel. This conflicts with our data and findings, as discussed further in §3.1 below.

Little attention has been devoted to Moroccan Arabic, although Boudlal [8] and Gafos [9] have conducted timing studies. Boudlal focused, however, on syllable length rather than the detection of vowel insertion and singleton/geminate duration contrasts. Gafos used electromagnetic articulography to examine inter-consonantal temporal overlap in MA clusters (CC), in word initial, word-medial and in word final positions. His findings showed there were less stop-stop consonant overlap in word initial positions and the overlap patterns are speaker specific.

Thus, our work is motivated by the studies of [4, 5, 9]. Many of the MA words are derived from MSA and therefore, there is no reason why in MA there should be a vowel before the geminate consonant but not in the singleton cases. For example, in MSA minimal pair /bakā/ and /bakkā/ is represented in MA as /bka/ and /bəkka/, no vowel before the singleton consonant /k/. It has become widely accepted with little proof that short vowel are elided in the singleton case and thus resulting in a branching onset (consonant cluster). In our exploratory study, we show that a schwa vowel is detectable between consonant clusters in word initial position even in singleton cases

2. Method

2.1. Word set

We reviewed 30 frequent minimal pairs (30 singletons and 30 geminates) consonants in MSA. From these we established which existed in MA, a sample shown in Table 1. In total there were 19 singleton cases which also existed in MA, 11 in geminate cases and 10 in both (singleton and geminate) which we list separately in Table 2. Audio recordings were done of only the words listed in Table 2. Table 1 and 2 omit some words with their glosses due to page restrictions.

Table 1. *Comparison of minimal pairs between MSA and MA.*

Singleton		Geminates	
MSA form	MA form	MSA form	MA form
/samina/	/smən/	/sammana/	*none*
/ʕatˤasa/	/ʕtˤəs/	/ʕatˤtˤasa/	*none*
/bakā/	/bka/	/bakkā/	/bəkka/
/fahima/	/fhəm/	/fahhama/	/fəhhəm/
/kataba/	/ktəb/	/kattaba/	*none*
/xalā/	*none*	/xallā/	/xəlla/
/banā/	/bna/	/bannā/	*none*
Etc……..			

A total of 144 tokens were formed: 18 words *x* 4 repetitions (twice in isolation and twice in a phrase context) *x* 2 participants. The carrier sentence for the target word was /ʕawəd _____ mərra xʷra/ (repeat _____ once again). The recordings were done in a sound booth at the CNRS centre using Sony ECM-MS907 microphone. Adobe Audition 1.0 was used and the recordings were digitized at a sampling rate of 16000 Hz. For each utterance a file containing a waveform display and a spectrogram for each token was made.

Table 2. *Data set - minimal pairs (singleton and geminates).*

	Singleton		Geminates	
	MSA	MA	MSA	MA
h	/fahima/	/fhəm/	/fahhama/	/fəhhəm
k	/bakā/	/bka/	/bakkā/	/bəkka/
k	/sakat/	/skət/	/sakkat/	/səkkət/
k	/sakan/	/skən/	/sakkan/	/səkkən/
l	/ʕalam/	/ʕləm/	/ʕallam/	/ʕəlləm/
m	/samaʕ/	/sməʕ/	/sammaʕ/	/səməʕ/
ʃ	/naʃita/	/nʃətˤ/	/naʃʃata/	/nəʃʃətˤ/
sˤ	/wasˤala/	/wsˤəl/	/wasˤsˤala/	/wəsˤsˤəl/
r	/barad/	/brəd/	/barrad/	/bərrəd/

2.2. Speakers

In this exploratory study we used two male speakers of Moroccan Arabic.

2.3. Measurements

Praat software [10] was used to produce a waveform display and spectrogram for each token. Firstly, we viewed the spectrogram of singleton cases, establishing whether there was any vowel material in the onset consonant cluster for singletons, for both speakers. Secondly, we extracted segmental durational measurements (in ms) for:

- epenthetic and lexical vowels
- vowels following C (singleton) and CC (geminates)
- consonants - C (singleton) and CC (geminates)

The acoustic segments were identified from both waveform and spectrograms (wideband) by positioning cursor at time points in the waveform as well as the segment onset and offset on spectrograms. The total utterance duration was measured from the onset of the first segment to the closure of last segment. The criteria used for identifying epenthetic vowel were via vocalic complex patterns in the waveform and then checked patterns in the accompanying spectrogram, in particular the vowel formants.

3. Results

We first analyzed the phonetic duration of isolated words and then later analyzed the target word embedded in phrase context. Results detailed in §3.1 to §3.2 are for isolated words.

3.1. Vowel insertion

In most cases we found vowel epenthesis similarly to that reported in Saudi Arabic by Baothman [6] and in Lebanese and Palestinian Arabic by Gouskova & Hall [4]. Spectrograms are provided in our forthcoming full paper. We found a vowel in the onset consonant clusters in word initial positions in most of the singleton pairs, for e.g., /bka/ → /bəka/, /nʃətˤ/ → /nəʃətˤ/ etc. In three of the words in the singleton cases, however, all s+C clusters e.g. /skən/, it was difficult to find any traces of epenthetic vowel material between the consonant clusters for both speakers. In the corresponding s+CC geminates, the vowel duration preceding the geminate was considerably reduced in comparison to geminates

following consonants other than s. In Table 3 below, we contrast the durational measurement (ms) of epenthetic vowel (**Və**) and the preceding lexical vowel in geminates (V1). In Table 3, means and standard deviation (in parentheses) are for words excluding the s+C cases since no corresponding vowel insertion was found in singleton cases.

Table 3. *Average vowel duration (ms) – singleton and geminates.*

Singleton	CVəCVC (ms)	Geminate	CV1CCV(C) (ms)
/fhəm/	32.75	/fəhhəm	49.75
/bka/	30.0	/bəkka/	55.75
/skət/	-	/səkkət/	39.0
/skən/	-	/səkkən/	40.0
/sməʕ/	-	/səmməʕ/	34.25
/nʃətˤ/	73.5	/nəʃʃətˤ/	73.25
/ʕləm/	86.75	/ʕalləm/	65.0
/wsˤəl/	92.5	/wəsˤsˤəl/	78.25
/brəd/	49.5	/bərrəd/	57.5
Mean (SD)	**60.83 (27.82)**	**Mean (SD)**	**65.85* (14.15)***

Və – epenthetic vowel   * excludes s+C cases

On average the epenthetic vowel varied across words, ranging from 30ms to 92.5ms. Statistically, exploratory ANOVA showed no significant difference between the duration of the epenthetic vowel in singleton and the lexical vowel duration preceding geminates (F = .626, *df* = 1, *p* = .433). Comparison of the epenthetic vowel was made with surrounding vowels: Table 4 shows the mean duration time (ms) and the standard deviation in parentheses.

Table 4. *Mean duration (ms) and standard deviation (in parentheses) for epenthetic vowel and surrounding vowels.*

Singleton – CCV1(C)		Geminate – CV1CCV2C	
Və	V1	V1	V2
60.83 (27.82)	92.22 (48.61)	65.85* (14.15)*	89.79 (47.71)

Ve – vowel epenthesis   * excludes s+C cases

The duration of the epenthetic vowel on average was 92% of the duration of the short lexical vowel preceding the geminate consonant (geminate-V1). We refer V1 in geminate as a short vowel as studies have shown that the vowel preceding the geminate consonant is shortened to some degree. For example, below in (1) and (2) in Lebanese Arabic study by Khattab [3], it is clear that on average the vowel preceding the geminate was 66.15% (86ms/130ms x 100%) of the duration of the full lexical vowel following the geminate (CC).

vowel pre-singleton C= 88ms ; pre-geminate CC= 86ms   (1)
vowel post-singleton C=127ms; post-geminate CC= 130ms   (2)

In our study, the duration of the epenthetic vowel on average was 92% of the duration of the short lexical vowel (geminate-V1) and 65.96% of the full lexical vowel (singleton V1), established from the data in Table 4. But results were also noted [4] with Lebanese Arabic, the epenthetic vowel on average was 90% of the full lexical vowel and no significant difference was found between the epenthetic and lexical vowel duration.

We also noted that the epenthetic and lexical vowel duration varies across speakers. Epenthetic vowel ranged from ~40ms to ~60ms and lexical vowel ranged from ~58ms to ~65ms. Such variation of epenthetic vowel and lexical vowel across speakers was also noted in the Lebanese Arabic study; epenthetic vowel ranging from ~40ms to 120ms and lexical vowel ranging from 70ms to 125ms.

3.2. Consonants – singleton vs. geminates

The duration for single consonant vs. geminate consonants for both speakers were pooled as shown in Table 5. On average the duration time for geminates was twice the duration of singletons, with the exception of /rr/ which was almost 4 times greater than the singleton /r/. Khattab [3] in Lebanese Arabic noted that the singleton to geminate ratio for consonant taps and trills (/r/) was 1:7.50.

Table 5. *Average consonant duration (ms) – singleton and geminates.*

Singleton	CCVC (ms)	Geminate	CVCCVC (ms)	Ratio C to CC
/fhəm/	72.5	/fəhhəm	143.75	1:1.98
/bka/	77.75	/bəkka/	177.75	1:2.28
/skət/	97	/səkkət/	204.75	1:2.11
/skən/	98.5	/səkkən/	211.25	1:2.14
/sməʕ/	89.5	/səmməʕ/	194.5	1:2.17
/nʃətˤ/	120	/nəʃʃətˤ/	208.5	1:1.74
/ʕləm/	74.25	/ʕalləm/	177.5	1:2.39
/wsˤəl/	121	/wəsˤsˤəl/	183.5	1:1.52
/brəd/	28	/bərrəd/	110	1:3.93
Mean (SD)	**90.89 (28.56)**	**Mean (SD)**	**170.31 (48.36)**	**1:1.87**

Statistically, exploratory ANOVA showed significant difference between the duration of singleton and geminates (F = 79.43, *df* = 1, *p* < .0001). The above singleton/geminate ratio findings are similar to Saudi Arabic [5, 6] Iraqi Arabic [1] and Lebanese Arabic [3].

3.3. Sentence context

The findings from target word embedded in the sentence context shows; firstly, that the epenthetic vowel in sentence context is longer than in isolated word cases. Secondly, the vowel preceding the consonant geminate in s+C cases is shorter than a full lexical vowel. Thirdly, the duration of the geminate consonant is slightly reduced but still remains twice the duration of the singleton consonant. Full data analyses and comparison with isolated words and relevant statistical data are provided in our forthcoming full paper.

4. Discussion

Previous studies with Saudi Arabic [5, 6] and Lebanese Arabic [4] found an epenthetic vowel between consonant clusters. In this exploratory MA study, we have also found vowel epenthesis between consonant clusters in word initial position (in singleton cases of the minimal pairs). No significant differences were found in the duration between epenthetic and lexical vowel, this confirms the findings of [4]. However, the durations of epenthetic schwa were low in some cases, as was also noted by Davidson [11 (working with English speakers articulating Czech consonant clusters) and Ridouane [12] (working with Tashlhiyt consonant clusters).

Dell and Elmedlaoui [13] argue that vowel elements (in Tashlhiyt) other than the realization of /a/, /i/ and /u/ within consonant clusters do not play any role in syllable structure. Several researchers have opposed their idea, and state that epenthetic schwas are part of syllable nucleus. In particular, Coleman [14] states that epenthetic vowels are phonetic realizations of syllable nuclei and "where no epenthetic vowel is evident, it can be regarded as hidden by the following consonant, according to a gestural overlap model". This probably explains why no epenthetic vowel was evident in MA for s+C clusters (in singletons). Although, phonotactically, s+C is permissible in English and French and both of our participants speak French too, from a syllabic point of view, Kaye [15] and other researchers would argue that s+C clusters cannot be treated as branching onsets, because word initial /s/+stop clusters violate the Sonority Sequencing Principle. But element phonology can treat s+C clusters as branching onsets.

Baothman's [5,6] epenthetic vowels were modeled in her element phonology as phonologically active *sukuuns*. In our exploratory study we did not find signs of phonological activity of *sukuun* in geminate consonants, nor vowel epenthesis in s+C clusters. But, there were epenthetic vowels in other word initial consonant clusters, even in the cluster /br/ → /bər/, which is licit in English, French and even in MA. Such finding was also noted in Spanish [16], there was a vowel epenthesis in the cluster /gr/ in the word 'grupo' ('group'). The cluster /gr/ is licit in Spanish and yet in the spectrogram [16:51] the epenthetic vowel was present and independent element which did not constitute part of the following consonant. Also, Coleman [14:735] notes that in Tashlhiyt, /r/ is pronounced "with brief vocoids and sometimes after the tap, and between the individual closure of trills". This supports our current findings in MA.

Our results with consonant geminates support other Arabic dialect studies [1,2,5], that on average the geminate consonant is about twice the duration of the singleton.

Currently, we are extending our exploratory study with a much larger corpus of words. More singleton/geminate consonant contrasts and consonant geminates word initially (e.g. /ddrb/), word-medial (e.g. /bddl/) and word finally (e.g. /sdd/). Two and three consonant clusters word initially (e.g. /ktab/), word-medial (e.g. /saktin/) and word finally (e.g. /sakt/), plus, more s+C clusters and more participants' recordings. Thus epenthetic vowels and *sukuuns* may become more evident which may support for a CV phonology for Moroccan Arabic. Also, with a larger corpus of recording we will be able to establish phonetic correlates between intended schwa and transitional schwa. In our extended work/paper we will also show how various phonological theories can account for epenthetic and transitional schwa.

5. Conclusions

The findings from our exploratory phonetic timing study on MA challenges some of the widely accepted principles, for e.g. it challenges the standard hypothesis that short vowels in Modern Standard Arabic are elided in Moroccan Arabic. The research in this paper has raised a number of interesting questions both from phonetic and phonological perspective.

6. Acknowledgements

The findings in her work have motivated the current and future work with Moroccan Arabic and other Arabic dialects.

7. References

[1] Hassan, Z. M. "Gemination in Swedish and Arabic with a Particular Reference to the Preceding Vowel: An Instrumental and Comparative approach". Fonetik TMH-SPSR Proc., 44:81-85, 2002.

[2] Ham, W. "Phonetic and Phonological Aspects of Geminate Timing", Routledge, London, 2001.

[3] Khattab, G. "A phonetic Study of Gemination in Lebanese Arabic", ICPhS XVI Proc., 153-158, 2007.

[4] Gouskova, M. and Hall, N. "Levantine Arabic Epenthesis: Phonetics, Phonology and Learning", Variation, Gradience and Frequency in Phonology Workshop, Poster, 2007.

[5] Baothman, F. and Ingleby, M., "Representing Coarticulation Processes in Arabic Speech", in S. Boudelaa. (ed.). Perspectives on Arabic Linguistics XVI, John Benjamin Publishing, Amsterdam, 95-102, 2002.

[6] Ingleby, M. and Baothman, F., " Empty Nuclei in Arabic Speech Patterns and the Diacritic Sukuun", in S. Boudelaa. (ed.). Perspectives on Arabic Linguistics XVI, John Benjamin Publishing, Amsterdam, 83-94, 2002.

[7] Harris, J., "English sound structure", Blackwell, Oxford, 1994.

[8] Boudlal, A. "Constraint Interaction in the Phonology and Morphology of Casablanca Moroccan Arabic", Unpublished doctoral thesis, Université Mohammed V.

[9] Gafos, A., Hoole, P., Roon, D and Zeroual, C. "Variation in Overlap and Phonological Grammar in Moroccan Arabic Clusters", In Laboratory Phonology X, Mouton de Gruyter, Berlin/New York.

[10] Boersma, P. Weenink, D. 2006. Praat: Doing Phonetics by Computer (Version 5.0.08) [Computer program]. Retrieved January 1, 2008, from http://www.praat.org/.

[11] Davidson, L. "Phonotactics and Articulatory Coordination Interact in Phonology: Evidence from Nonnative Production", Cognitive Science, Vol. 30, 2006, 834-862.

[12] Ridouane, R. "Voiceless, Vowel-less Syllables in Tashlhiyt Berber: Phonetic and phonological Evidence", [PDF file] retrieved April 2, 2008 http://lpp.univ-paris3.fr/equipe/rachid_ridouane/.

[13] Dell, F. and Elmedlaoui, M. "Syllables in Tashlhiyt Berber and in Moroccan Arabic", Kluwer, 2002.

[14] Coleman, J. "The nature of Vocoids Associated with Syllabic Consonants in Tashlhiyt Berber", ICPhS XIV Proc, 735-738, 1999.

[15] Kaye, J., "Do you Believe in Magic? The Story of s +C Sequences". SOAS Working Papers in Linguistics, 2: 293-313, 1992.

[16] Ramirez, C.J. "Acoustic and Perceptual Characterization of the Epenthetic Vowel between the Clusters Formed by Consonant + Liquid in Spanish", Laboratory Approaches to Spanish Phonetics and Phonology Proc, 48-61, 2006.