❏   1163

Energy distribution in formant bands for arabic vowels

Mohamed Farchi[1], Karim Tahiry[2], Soufyane Mounir[3], Badia Mounir[4], Ahmed Mouhsen[5]
[1,2,3,5]IMMII Laboratory, Faculty of Sciences & Technics, University Hassan First, Settat, Morocco
[4]LAPSSII Laboratory, Graduate School of Technology, University Cadi Ayyad, Safi, Morocco

Article Info	ABSTRACT
	The acoustic cues play a major role in speech segmentation phase; the extraction of these indexes facilitates the characterization of the speech signal. In this work, we aim to study Arabic vowels (/a/, /a:/, /i/, /i:/, /u/ and /u:/), especially the long ones. We are interested in characterizing this type of vowels in terms of time, frequency and energy. The cues extracted and analyzed in this work are: segment length, voicing degree and formants values.

Corresponding Author:

Karim Tahiry,
IMII Laboratory, Faculty of Sciences & Technics,
University Hassan First,
FST of Settat, Km 3, B.P: 577 Road of Casablanca, Settat, Morocco.
Email: karim.tahiry@gmail.com

1.   INTRODUCTION

The production of a vowel is characterized by a maximum opening of the vocal tract without constriction or noise production or silence. A periodic vibration of the vocal cords, which characterize voiced sounds, always accompanies this production [1]. The standard Arabic language has six vowels, three short (/a/, /i/, /u/) and three long (/a:/, /i:/, /u:/) [2]. These vowels differ from those of other languages (English, Spanish...) in terms of number and vocalic quantity.

The characterization of vowels can be performed in terms of time, frequency and energy. Kimiko Tsukada (2009) studied the time characterization of vowels. He presented a comparative study between long and short vowels in Standard Arabic, Japanese and Thai. He reported that the duration of long vowels represent the double of the short vowels duration. He also noticed that the ratio between the duration of short and long vowels differ significantly for the three languages [3]. Alghamidi (1998) conducted a comparative study between long vowels and short ones in terms of frequency for some Arabic dialects (Egyptian, Sudanese, Saudi). He determined that, for studied dialects, F1 and F2 formants of long vowels are different from those of short vowels. The vocalic triangle formed by long vowels includes the short ones [2], [4]. Alotaibi and Hussein (2009) analyzed the vowel formants of standard Arabic. They confirmed that the long vowels formants F1 and F2 are peripheral to those of short ones. They also showed that the values of the formants F1 and F2 help to classify the vowels: the vowel / a / and /a:/ are characterized by a high value of F1 (F1> 500Hz) and the vowel / i / and / i: / have a high value of F2 (F2>1500 Hz) [5]. Sawusch (1996) investigated the effects of duration on vowel perception in normal American-English speakers. He summarized that vowel duration was not a strong perceptual cue to vowel identity but was used by listeners when other sources of information were distorted [6]. Mohammad Abuoudeh and Olivier Crouzet (2014) examined Vowel length impact on locus equation parameters for Jordanian Arabic. They observed

that the vowel length systematically influences the locus equation data, and the variations of vowel length are associated with modifications of spectral configuration [7].

In this work, we carry out an acoustic study of Arabic long vowels compared to those shorter. The studied parameters are summarized in: vowel production time, its formants and the variation of energy contained in the F1 and F2 bands. This paper is organized as follows: we begin by describing the methods and tools used and the experiments carried out. Then we present and discuss the results and we close by a conclusion.

2. METHOD

2.1. Corpus

We constructed a corpus of Arabic language. It consists of short and long vowels. Five Moroccan speakers (three male and two female) were invited to pronounce syllables CV (C: consonant and V: vowel) with short and long vowels. We chose to work with isolated syllables in place of words to reduce the influence of other phonemes on the vowel studied. We can then expand freely the length of the vowel to examine his behavior. For the consonant C associated with the vowel V studied, we chose /A/: / ء / because its production induces minimal stress on the vocal tract. Table 1 shows the syllables of the corpus.

Table 1. Arab Corpus of Long and Short Vowels

Vowel a	Vowel i	Vowel o
/أ/ : /A/	/ئ/ : /I/	/ؤ/ : /U/
/ئا/ : /AA/	/ئي/ : /II/	/ؤو/ : /UU/
/ئاا/ : /AAA/	/ئيي/ : /III/	/ؤوو/ : /UUU/
/ئااا/ : /AAAA/	/ئييي/ : /IIII/	/ؤووو/ : /UUUU/
/ئاااا/ : /AAAAA/	/ئيييي/ : /IIIII/	/ؤوووو/ : /UUUUU/

2.2. Formants extraction method

To construct our corpus, we used the vocal sounds process tool "Praat" to achieve our records in a noise-isolated room, with a sampling frequency of 22050 Hz. We used "Praat" to isolate and determine the duration of each vowel. We used linear predicting coding method "LPC" to extract the first four formants. Figure 1 shows the pre-treatments of the speech signal in order to extract the formants.
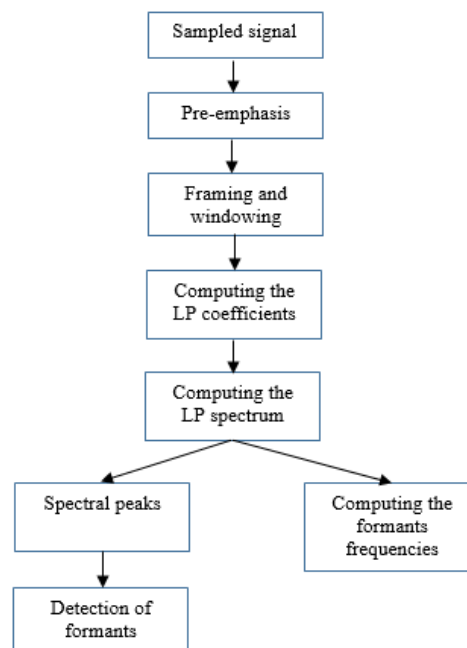

Figure 1. Chart of the detection procedure of formants with LPC [8]

For our experiments, the speech data was sampled to the frequency of 22050 Hz. All coefficients have been computed from pre-emphasised speech signal using 512 points Hamming windowed speech frames. Then the linear prediction coefficients are calculated. The LPC model supplies a smoothed spectral, the peaks of the spectral envelope correspond to the formants.

2.3. Energy formants

The speech sampled at 22050Hz is divided into time segments of 11.6 ms with an overlap of 9.6 ms. Each segment is Hanning windowed and followed by zero-padding. 512 point fast Fourier transform (FFT) is then computed. The magnitude spectrum for each frame is smoothed by a 20-point moving average taken along the time index n. From the smoothed spectrum $X(n,k)$, peaks in two frequency formants (250-850 Hz and 750-2300 Hz) are selected as:

$$E_b(n) = \sum_k 10 \, log_{10}(\, |X(n,k)|^2) \qquad (1)$$

Where the formant index b represent first and second formant (F_1 and F_2). The frequency index k ranges from the DFT indices representing the lower and upper boundaries for each formant. Then, for each frame, the normalized energy band was calculated by:

$$Ebn \, (n) = \frac{E_b(n)}{E_T(n)} \qquad (2)$$

Where $E_{bn}(n)$ is the normalized formant energy b in the frame n, $E_T(n)$ is the overall energy in the frame n and $E_b(n)$ is the formant energy b in the frame n.

3.   RESULTS

Figure 2 represent spectrograms of short and long vowels. It can be seen that both short and long vowels are voiced sounds even if the duration production of long vowels increases.
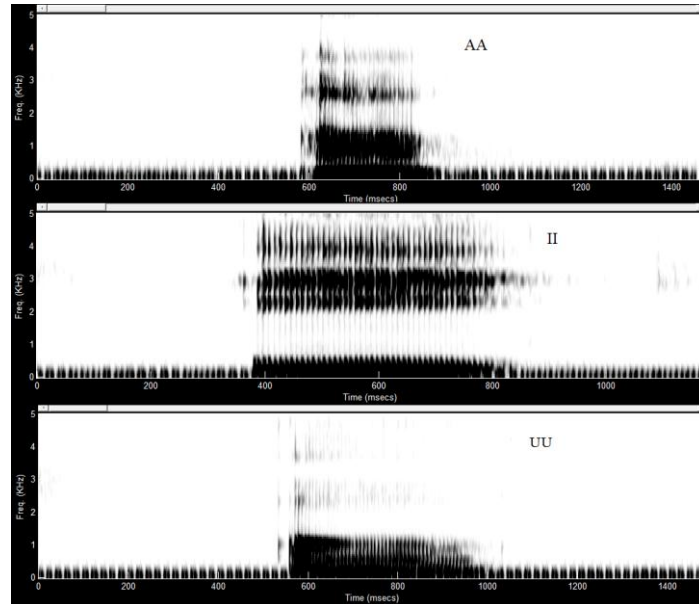

Figure 2. Long vowels /AA/ /II/ /UU/ spectrograms

To minimize the effect of the variability intra-speaker, we calculated the average of each formant for each speaker. We, then, set the average for the five speakers. We performed the same way to calculate the energy variation contained in F1 and F2 bands.

Figure 3 show that the duration of long vowels represent the double of the short ones. This result is consistent with those of Alghamdi [2], Tsukada [3] Alotaibi [5].

Figure 4 summaries the variation of energy contained in F1 band with the increase of the production duration of the three vowels. It can be seen that for /a/ and /i/, the energy in F1 band increases when the duration of vowel production increases. This behavior is the opposite in the /u/ vowel. It is also shown that the energy contained in this band is lower for /a/ and higher for /i/. This behavior can be explained by the fact that the place of articulation of /i/ is in the back of the vocal tract (near to the vocal folds). The energy in F1 band is then more important. For /u/ vowel, we observed that the F1 band energy decreases with the increase of the /u/ production duration. This behavior is due to the limitation of space between the back of the tongue and palate if producing /u/ takes longer.

Figure 5 shows the variation of energy in F2 band. We noticed that the /i/ production duration has no effect on the energy variation: no difference between short and long vowel is noted. For /a/ and /u/ vowels, the F2 band energy decreases rapidly when producing long vowels /a:/ or /u:/ and remains constant even when the duration of long vowels /a:/ and /u:/ increases.

It can also be seen that the energy contained in this band is higher for /a/ due to its place of articulation in the back of the tongue: the energy in F2 band depend on the area between the teeth and the place of articulation. The expansion of this region during production of /a/ leads to more important energy in F2 band.
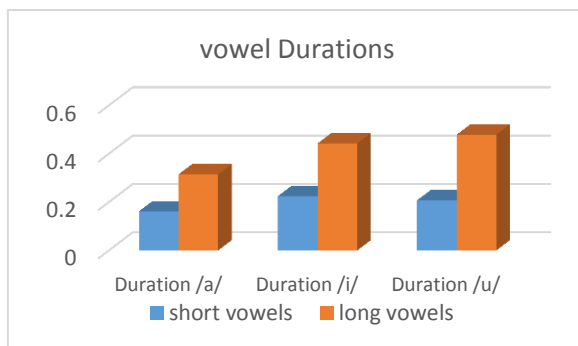

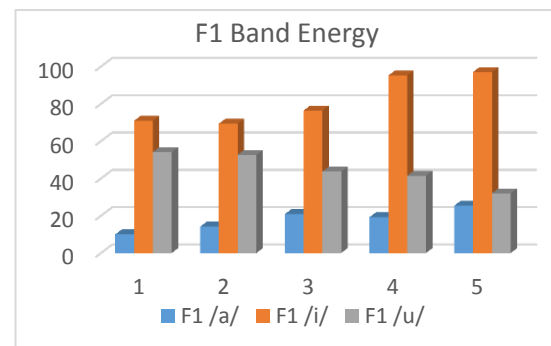Figure 3. Short and long vowels durations

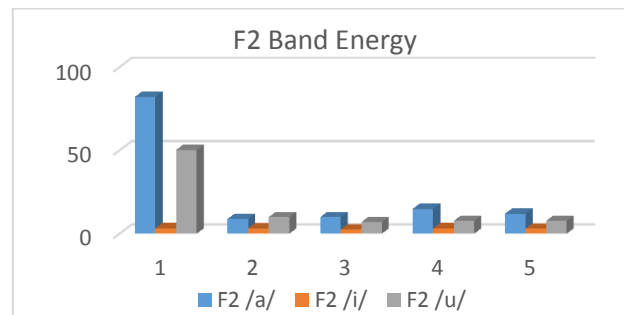Figure 4. F1 band energy variation vs production duration of vowels /a/, /i/ and /u/

Figure 5. F2 band energy variation vs production duration of vowels /a/, /i/ and /u/

4. CONCLUSION

This study compares the long and short Arabic vowels in terms of production duration and energy distribution in F1 and F2 bands. The obtained results show that the long vowels are voiced sound even when their duration production increases in time. The comparison between short and long vowels in term of production duration reveals that long vowels are twice long than short vowels. For each vowel (/u/, /a/ or /i/), the energies contained in F1 and F2 bands vary when producing long vowels. When the production duration of long vowels increases, the F2 band energy remains constant for all vowels while the F1 band energy increases or decreases depending on the vowel produced.

REFERENCES

[1]   Andrew W.Howitt, "Vowel Landmark Detection," *International Conference on Spoken Language Processing*, 4:628-631, Octobre 2000.
[2]   Mansour. M. Alghamdi, "A Spectrographic Analysis of Arabic Vowels: A Cross-Dialectal Study," *Journal of King Saud University*, vol. 10, pp. 3-24, 1998.
[3]   Kimiko Tsukada, "An Acoustic Comparison of Vowel Length Contrasts in Standard Arabic, Japanese and Thai," *2009 International Conference on Asian Languages Processing*, DOI 10.1109/IALP.2009.25, IEEE.
[4]   Djoudi, M., "Contribution à l'étude et à la Reconnaissance Automatique de la Parole en Arabe Standard," these de doctorat, Université Nancy I, Nancy, France, 1991.
[5]   Alotaibi Y., Hussain A., "Speech Recognition System and Formant Based Analysis of Spoken Arabic Vowels," In: *Proc. First International Conference*, December, FGIT, Jeju Island, Korea, pp. 10–12, 2009.
[6]   James R. Sawusch, "Effects of Duration and Formant Movement on Vowel Perception," *1996 Proceedings of the Fourth International Conference onSpoken Language Processing (ICSLP-96)*, October 3-6, Philadel-phia.
[7]   Mohammad Abuoudeh, Olivier Crouzet, "Vowel Length Impact on Locus Equation Parameters: An Investigation on Jordanian Arabic," *COLIPS. 15th Annual Conference of the International Speech Communication Association (ISCA; Interspeech 2014)*.
[8]   Gargouri, D., Frikha, M., Kamoun, M. A., Ben Hamida, A., "A Comparative Study of an All Pole Speech Analysis for Formant Extraction," *Third International Conference on Systems, Signals & Devices*, Sousse, Tunisia, 2005.

BIOGRAPHIES OF AUTHORS

Mohamed Farchi was born in 1993. He received the engineering degree in Electrical Engineering and Power Systems from Mohammadia School of Engineers, Rabat, Morocco, in 2015. He is currently a Ph.D student in Engineering, mechanical, Industrial Management and Innovation (IMMII) Laboratory research Laboratory, Faculty of Sciences & Technics, Hassan First University with a thesis on automatic speech recognition.

Karim Tahiry was born in Settat, Morocco, in 1988. He received the Ph.D. degree from the Faculty of Sciences and Technics, Hassan first University, Morocco in 2018 in Electronics and Telecommunications. In 2011, he received the Master degree in Automatic, Signal processing, Industrial Computing from University Hassan first, Settat, Morocco. Member of Engineering, mechanical, Industrial Management and Innovation (IMMII) Laboratory. Her research interests include speech recognition, signal processing.

Soufyane MOUNIR was born in El borouj, Morocco, in 1984. Assistant Professor at National School of Applied Sciences, University Hassan first since 2014. Member of Engineering, mechanical, Industrial Management and Innovation (IMMII) laboratory. His research interests include speech recognition, signal processing and security of VoIP networks.

Badia Mounir was born in Casablanca, Morocco, in 1968. Engineer degree (1992) in "Automatic and Industrial computing", The Mohammadia School of engineering, Rabat, Morocco. Assistant Professor at Graduate School of Technology, University Cadi Ayyad since 1992. Habilitaded to supervise research (HDR) since 2007 and professor of higher education (PES) since 2017. Member of Laboratory of Process, Signals, Industrial Systems, informatic (LAPSSII) Laboratory. Her research interests include speech recognition, signal processing, energy optimization and modeling.

Ahmed MOUHSEN was born in May 1960. He received the Ph.D. degree in electronics from the University of Bordeaux I, France. He is currently a Professor of Electronics in FSTS university Hassan 1st Settat Morocco. He is involved in the design of hybrid, active, and passive microwave electronic circuits, digital systems and IoT.