Arabic Vowels Acoustic Characterization

Mohamed Farchi, Tahiry Karim, Ahmed Mouhsen

Arabic Vowels Acoustic Characterization

Mohamed FARCHI[1], Karim TAHIRY[2], Ahmed MOUHSEN [3]
simo.farchi@gmail.com
IMMII Laboratory, Faculty of Sciences & Technics, University
Hassan First, Settat, Morocco

ABSTRACT: A sentence is constructed using basic word units. Each word is composed of syllables and each syllable being composed of phonemes, which in turn, can be classified as vowels or consonants. Vowels may occupy the center or nucleus of a syllable, as opposed to consonants, which occur in marginal positions in a syllable (onsets and codas).Vowels appear early in speech development and are central to understanding the acoustic properties of speech. The aim of this work is to enrich the automatic speech recognition system, by the classification of Arabic vowels based on normalized energy in the formant frequency bands.

Keywords: Automatic speech recognition, Arabic vowels, formants, normalized energy bands.

1 INTRODUCTION

Speech consists of acoustic pressure waves created by the voluntary movements of anatomical structures in the human speech production system. These waveforms are broadly classified into voiced and unvoiced speech [1]. Vowels are voiced sounds, which are produced with periodic vocal fold vibration and without constriction in the vocal tract, they are characterized by resonance in throat, the resonance frequencies of the vocal tract are the formants. The major challenge in describing the articulation of vocalic sounds is to define the position of the tongue, and the majority of the following instrumental techniques, when applied to the task of vowel description, aim to provide different types of information on the location and movement of the tongue during vowel description [2]. Historically, vowels are classified according to tongue height (high, mid, low) and advancement (front, central, back) and lip configuration (round, neutral, spread).

Many researchers have studied the three rules for relating changes in tongue height, tongue advancement, and lip rounding to change in formant frequencies. Hixon et al. (2008) [3] reported that the first formant frequency (F1) varies inversely with tongue height. Therefore, the lower the tongue at the major point of constriction for the vowel, the higher is F1. This rule is stronger for the front compared to the back vowels. The second rule states that the second formant (F2) increases and F1 decreases with increasing tongue advancement. This rule is stronger for high compared to low vowels. The third rule states that all formant frequencies decrease with increased lip rounding, with the major effect on F2. This rule is also stronger for high compared to low vowels.

Normally, the specification of vowel acoustic structure is by measuring the centre frequency of the first three or four formant resonances (Fant, 1960) [4].The first two formants often being sufficient (Kent & Read, 1992) [5], particularly in languages such as English, which do not contain vowel distinctions that depend solely on contrasts in lip position (Ladefoged, 2001) [6]. F1 values relate to tongue position in the vertical domain, and F2 values to the front-back dimension. Iskarous (2010) [7] reported that, at least for steady-state vowel productions, formant structure can provide significant information on location and degree of lingual constriction, as well as presence or absence of lip-rounding.

The characterization of vowels can be performed in terms of time, frequency and energy. Vowel duration values, meanwhile, are usually obtained by measuring from the onset of the second formant (F2) to its offset, although Blomgren and Robb (1998) [8] draw attention to the difficulties in determining vowel duration. 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 [9]. 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 [10].

In this work, we carry out an acoustic study of Arabic shortvowels. The studied parameter is the energy contained in the formant 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 vowels (/a/, /i/ and /u/). Five Moroccan speakers (three male and two female) were invited to pronounce syllables CV (C: consonant and V: vowel) with short vowels. We chose to work with isolated syllables in place of words to reduce the influence of other phonemes on the vowel studied. For the consonant C associated with the vowel V studied, we chose /A/: / ء /

because its production induces minimal stress on the vocal tract.

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.

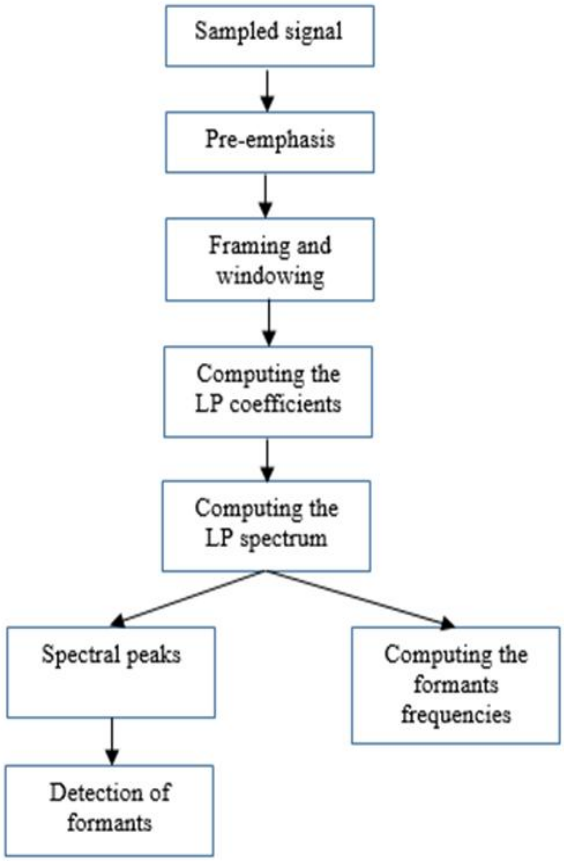

fig 1 : Chart of the detection procedure of formants with LPC [6].

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

By analyzing the tendency of the first three formants (F1, F2 and F3) for Arabic vowels (/a/, /i/, et /u/), three frequency bands were observed. Likewise, the pitch frequency (F0) band must be considered since the vowels are voiced sounds. Dividing the signal into these frequency bands helps to capture its acoustic changes. The four frequency bands considered in this work are summarized in Table 1:

Table 1: Formants frequency bands

	F0 Band	F1 Band	F2 Band	F3 Band
/a/	0–400 Hz	500–800 Hz	1000–1500 Hz	2200–2800 Hz
/i/	0–400 Hz	100–400 Hz	2000–3000 Hz	2800–3400 Hz
/u/	0–400 Hz	300–600 Hz	600–1100 Hz	2400–3000 Hz

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. 512point fast Fourier transform (FFT) is then computed. Themagnitude spectrum for each frame is smoothed by a 20-pointmoving average taken along the time index n. From thesmoothed spectrum X(n,k), peaks in frequency formants are selected as :

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

Where the formant index b represent formants. 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:

$$E_{bn}(n) = \frac{E_b(n)}{E_T(n)} \qquad (2)$$

Where Ebn (n) is the normalized formant energy b in the frame n, ET (n) is the overall energy in the frame n and Eb (n) is the formant energy b in the frame n.

3 RESULTS

The objective of this section is to examine the normalized energy in the formant frequency bands of Arabic vowels (/a/, /i/ and /u/).The results obtained are summarized in (Figs 2, 3 and 4). All vowels have a significant energy in the pitch frequency band since they are voiced sounds. Furthermore, low energy was observed in third formant band for the three vowels, which can be explained by the first two formants are the most important for the acoustic structure of vowel production.

We can see from Fig. 2 and 3 that, the maximum of energy is concentrated in first and second formant bands for Arabic vowels /a/ and /u/. For vowel /a/, the energy in second formant band exceeds the energy in first formant band. Figure 4 show that the most energy is concentrated in first formant band for vowel /i/, this behaviour is due to the total overlap between the first formant band and the pitch band.

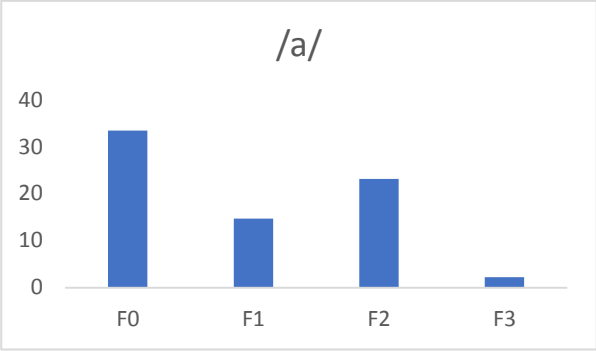

fig.2. Energy formant bands of vowel /a/.

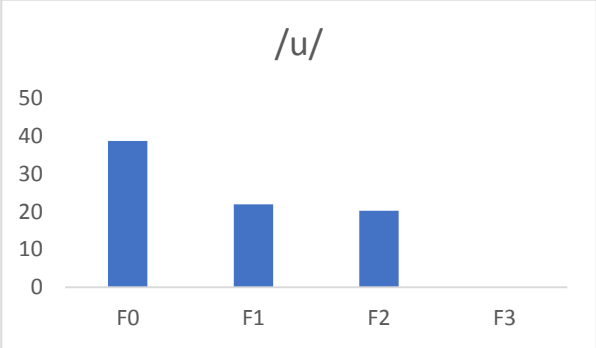

fig.3. Energy formant bands of vowel /u/.

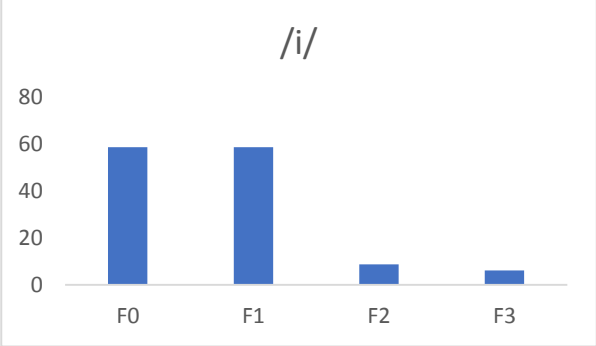

fig.4. Energy formant bands of vowel /i/.

4 CONCLUSION

This study characterizes the Arabic vowels (/a/, /i/ and /u/) according to acoustic cues such as the energy distribution in formant bands. The obtained results show that the energy in first formant F1 band can classify the three vowels (/a/, /i/ and /u/)/. The energy in the second formant F2 band can classify /i/ from /a/ and /u/. In sum, it has been possible throughout this work to characterize Arabic vowels. The findings of this study would constitute a support for speech recognition.

Bibliographie

[1] Prica, B., Ilic, S., "Recognition of Vowels in Continuous Speech by Using Formants" SER.: ELEC. ENERG. vol. 23, no. 3, December 2010, 379-393.

[2] Clark, J., Yallop, C., & Fletcher, J, "An introduction to phonetics and phonology (3rd edition)" London: Blackwell, 2007.

[3] Hixon, T. J., Weismer, G., & Hoit, J. D, "Preclinical speech science: Anatomy, physiology, acoustics, perception" San Diego, CA: Plural Publishing Inc, 2008.

[4] Fant, C. G. M, "The acoustic theory of speech production" The Hague: Mouton, 1960.

[5] Kent, R. D, "The biology of phonological development" In C. Ferguson,L. Menn, & C. Stoel-Gammon (Eds.) Phonological development: Models,research, implications, 1992, (pp. 65–90). Maryland: York Press.

[6] Ladefoged, P, "Vowels and consonants: An introduction to the sounds of the world's languages" London: Blackwell, 2001.

[7] Iskarous, K, "Vowel constrictions are recoverable from formants" Journalof Phonetics, 2010, 38(3), 375–387.

[8] Blomgren, M., & Robb, M, "How steady are vowel steady-states" Clinical Linguistics and Phonetics, 1998, 12, 405–415.

[9] 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.

[10] James R. Sawusch, "Effects of Duration and Formant Movement on Vowel Perception," 1996 Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP-96), October 3-6, Philadelphia.