VEABD 1.0: Vowel-Emphatic Algerians Berber Dataset is provided by Dr CHELALI Fatma zohra, Assistant professor at Speech Communication and Signal Processing Laboratory, University of Science and Technology Houari Boumediene of Algiers, ALGERIA.

This data set concerns isolated words containing vowels and Emphatic Berber. In Algeria, the principal Berber-speaking region is Kabylia. In a relatively limited but densely populated surface area, Kabylia alone has two-thirds of Algeria's Berber speakers.

1. *Berber language*

The Berber language is one of the branches of the large Hamito-Semitic linguistic family (or "Afro-Asiatic", according to the American terminology initiated by J. Greenberg), which includes, other than Berber, Semitic, Cushitic, ancient Egyptian, and more distantly, the Tchadic1 group. With all that this notion implies, the Berber language can be considered as the "aboriginal" one of North Africa because currently there is no positive trace of an exterior origin or of the presence of a pre- or non-Berber substratum in this region. As far back as one can go, the Berber language was already installed in its present territory. Particularly, the toponymy has not allowed us to identify, up till now, any kind of pre-Berber linguistic sediment. Despite numerous theories suggested by linguists since the 19th Century in favor of an external origin of the language (Middle East or East Africa), neither prehistoric archeology nor physical anthropology could show the movement of a population coming from elsewhere; it has even been solidly established that man has been present in North Africa, in a continuous manner, for at least a million years [1].

Tamazight (the Berber word for language) covers a vast geographical area: all of North Africa, the Sahara, and a part of the West African Sahel. But the countries principally concerned are, by order of demographical importance: Morocco (35 to 40% of the total population), Algeria (25% of the population), Niger and Mali (Tuaregs) [1].

In Algeria, the principal Berber-speaking region is Kabylia. In a relatively limited but densely populated surface area, Kabylia alone has two-thirds of Algeria's Berber speakers. The other significant Berber-speaking groups are: the Chaouias of the Aures region, having in all likelihood a million people, and the people of the Mzab (in Ghardaia and other Ibadhite cities), having a population of between 150,000 and 200,000. There are in fact other Berber-speaking groups in Algeria, but these are modest linguistic islands of only several thousands to tens of thousands of speakers [1].

Kabyle has three phonemic vowels: open /a/, and close /i/ , /u/ , similarly to Classical Arabic.< e > is used to write the epenthetic schwa vowel [ə] which occurs frequently in Kabyle. Historically it is thought to be the result of a pan-Berber reduction or merger of three other vowels.The phonetic realization of the vowels, especially /a/, is influenced by the character of the surrounding consonants; emphatic consonants invite a more open realization of the vowel, e.g. aẓru = [azˤru] 'stone' vs. amud = [æmud] 'seed'. Often /a, i, u/ are realized as [æ, ɪ, ʊ] .

[1] Berber language, Kabyle. Wikipedia

2. VEABD 1.0: Vowel-emphatic Algerians Berber dataset

VEABD 1.0 contains files in .wav format . The file names are: Lyes, Nabil,Nissa, Sonia,Lila and Tarek. Subjects are undergraduate students from,Electronics and computer science faculty at University of Science and Technology Houari Boumediene of Algiers.

If you use this database for your research refer to it as the VEABD1 dataset, (Fatma zohra Chelali, 2012).

Fatma zohra CHELALI, Amar. DJERADI and Hocine TEFFAHI, « Caractérisation de voyelles et d'emphatiques berbères en vue d'une identification de locuteurs dépendant de texte » The 5th International Conference on Information and Communication Technologies for Amazigh, 5ème conférence internationale sur les TIC pour l'amazighe. 24 et 25 Novembre 2012, Rabat, Maroc. http://www.ircam.ma/fr/index.php?soc=annon&rd=278

We chose six speakers for the Berber language (kabyle) which are: Lyes, Nabil,Nissa, Sonia, Lila and Tarek, all students native to the region of Kabylia. We collected a large number of words containing vowels and Emphatic Berber at different moments pronouncing 10 words containing five (05) emphatic Berber and three words containing vowels.
The sound card we used is called MobilePre USB. MobilePre USB is a preamplifier mobile integrating an audio interface perfect for the records on computer (laptop or mobile).

The database includes speech signals from six (6) different subjects. The speech signals are acquired during 2 s or 3s with different sessions to consider all variations at a sampling rate 16 KHz. After, a manual segmentation with "wavesurfer software" is done to consider only the part of produced voice.

During the recording, each repetition has been analysed to ensure that the entire sequences have been properly stored and avoiding any external interference. The chosen sequences are 05 emphatic Berber as follows:

- **/ص or ş /** in the isolated word **/yessub,issidh/ .**
- / ض or ḍ / in the isolated word **/adhar,assmidh/.**
- **/ ر or ṛ /** in the isolated word **/yerwa,aghrum /.**
- **/ز or ẓ /** in the isolated word **/tizurin,tazalit/.**
- **/ ط or ṭ /** in the isolated word **/attas,amchttouh /.**

❖ Vowels as :

- /**a/** in the isolated word **/tamart/ .**
- /**u/** in the isolated word **/tamurt/.**
- /**i /** in the isolated word **/tanmirt/ .**

The database includes 750 speech signals from six(6) different subjects corresponding to 15 repetitions for each word. The speech signals are acquired during 2 s with different

sessions to consider all variations at a sampling rate 16 KHz. After, a manual segmentation with "software wavesurfer" is done to consider only the part of produced voice.

Parameter	Value
Audio file format	.wav
Sampling rate	16 Khz
Number of bits	16 bits
Corpus	Takbaylit Berber Emphatic
Accent	Algerian Kabyle Language
Number of speakers	06
Number of words	10
Number of repetitions	15

3. Wavesurfer for speech analysis

Wavesurfer software can be used to visualize and analyze sound in several ways. The standard analysis plug-in can display Waveform, Spectrogram, Pitch, Power or Formant panes, as shown in the image below and also for manual segmentation.
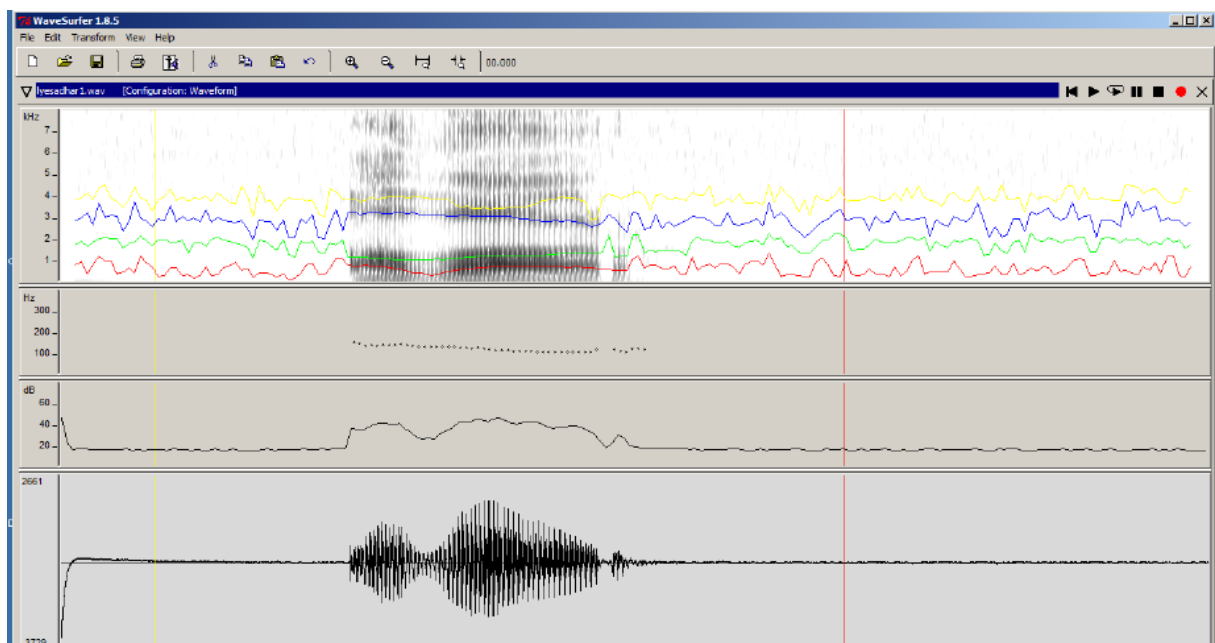
Figure1. Spectrogram, Pitch, Power and Formant representation of sound "Lyes Adhar1.wav" after removing DC component