Acted Emotional Corpus of Algerian Dialect Speech : Requirements, Description and Initial Results

Faycal YKHLEF[1], Wahiba BENZABA[1], Ratiba BOUTALEB[1], Abdeladhim DERBAL[1], Djamel BOUCHAFFRA[1] and Farid YKHLEF[2]

[1]ASM division, Centre de Développement des Technologies Avancées, Algiers, Algeria

[2]LATSI laboratory, University of BLIDA 1, Blida, Algeria

E-mail : fykhlef@cdta.dz

***Abstract* -** We present in this paper the first phase of an ongoing project that aims at creating the acted Emotional Speech Corpus of Algerian Dialect. The corpus is named ESCAD and is being recorded by 53 nonprofessional native actors (speakers) and subjectively assessed by 11 referees. The entire corpus, once accomplished, will include wav-files that convey angry, happy, disgust and neutral emotional states. The first phase of the project is restricted to the simulation of angry and neutral emotions along with the identification of unskilled actors. Furthermore, this phase summarizes the main requirements to create acted speech corpus, describe the components of the entire data and highlight the initial assessment results. We have explored the dependent t-test for paired samples to prove the separability between neutral and angry wav-files. In addition, the listening tests allow us to identify the unskilled group of actors among the initial volunteers. These actors will not take part in the simulation of happy and disgust emotions. We can share the first version of the corpus upon request.

***Keywords* -** Acted emotions; Speech corpus; Algerian dialect.

I. INTRODUCTION

Several modalities can be used to recognize emotions. One can mention, text, facial expression, electrocardiogram, speech signal, electroencephalogram, and galvanic skin response. Speech modality has the advantage to be non-invasive, non-intrusive and contact free [1]. In addition to its ability to convey human emotions, the speech waveform carries numerous evidences related to the speaker' gender, age, mood and pathologies [2]. Emotion recognition has been a hot topic in the field of human machine interaction [3, 4]. Indeed, emotions are classified into seven main classes: sadness happiness, disgust, surprise, fear, anger and neutral [5]. Secondary emotions can be extracted from these main classes. The technology of recognizing emotions though speech can be used in numerous applications such as speech analytics, medical psychology, emotional health, assistive robotics and gaming [6]. Pattern recognition and machine learning techniques are employed to recognize speech emotions. An essential condition to design such techniques is the availability of data. Numerous Emotional Speech Corpora (ESC) have been recorded in different languages [7-9]. The emotions comprised in these corpora may vary depending on the type of application and standards. Three main categories of ESC can be distinguished: natural, acted, and elicited [10, 11].

The recording of natural corpora is conducted in real-life circumstances through which speakers are innately disclosing their moods, feelings and emotions. This category of ESC is generally recorded in health facilities while doctors and patients are discussing. The conversations between customers and agents in call centers can also be exploited to create natural ESC [10, 11].

Acted ESC include speech recordings that convey synthetic emotions simulated by skilled actors (speakers). The linguistic content of these recordings must not be biased towards a specific emotion. Elicited ESC include speech recordings that convey synthetic emotions simulated by talented actors without their knowledge. More precisely, actors are pushed to conduct conversations which include several contextual emotions that were defined in advance [10, 11]. During these conversations, actors will be elicited to express a specific emotion unintendedly. The design of speaker and language-independent models is a hot topic in speech emotion

recognition [10]. Consequently, the creation of customized ESC that contain spoken words and sentences in several languages and dialects is needed to assess the performance of language-independent models. Moreover, such corpora must be recorded by a large number of speakers. There has been little research on the recording of Arabic ESC. The Arabic talk shows [7], the Ksuemotions corpus [11], the Modern Standard Arabic data [12], the Tunisian dialect datasets [13], and the audio-visual corpus of AD [3] are examples Arabic language corpora. The recordings introduced in these studies are acted corpora with the exception of the studies presented by S. Klaylat et al. [7] and H. Dahmani et al. [3] which propose a natural emotional corpus.

It is important to note that Arabs are speaking dialects that are different from the Standard Arabic (SA) to communicate in their daily lives. These dialects are the mother tongue language in all Arabic countries. Nevertheless, SA is used in televisions and newspapers. For instance, in Algeria people are employing numerous dialects to speak to each other. Basically, the vocabulary of these dialects can vary significantly according to the geographical location. In fact, people living in Algiers and its boundaries are speaking a dialect which is not the same as the one spoken in the west or the east of Algeria. The dialects spoken in Algeria are composed of words borrowed from Spanish, French, Turkish, Berber (Amazigh) and SA. To the best of our knowledge, no acted emotional speech data of Algerian dialect were created before. Thus, our research group is creating the first ESC that includes speech recordings of the dialect spoken in Algiers and its boundaries. Our ESC is recorded by 53 native actors. The number of actors (speakers) is sufficiently large to assess the performance of speaker-independent models. Our actors are imitating call center agents and are uttering the most common expressions used by these agents to communicate with customers. We have elected voluntary actors (nonprofessional) who have good abilities in mimicking voices and simulating emotions.

The mission of the entire project is to record wav-files that convey four synthetic emotional states, namely, neutral, angry, disgust, and happy. The created data is named “Emotional Speech Corpus of Algerian Dialect” ESCAD. We present in this paper the first phase of the project that consists in simulating angry and neutral emotions and identifying unskilled actors. It should be noted that the creation of speech corpora acted by nonprofessional actors is time consuming and difficult. Consequently, we have divided the process of corpus creation into two phases.

The focus of the first phase is on introducing the guidelines to identify unskilled speakers through the simulation of 50% of the corpus. The establishment of these guidelines is based on the assessment of neutral and angry wav-files (the initial results of the first phase).

The second phase of the project consists in simulating happy and disgusts emotions and assessing the entire data. The actors who are identified as unskilled during the first phase are not going to take part in the simulation of happy and disgust emotions. The solution that we have proposed will help reduce the time taken to complete the recordings and create an appropriate ESC. The first phase of the project is accomplished. However, the second phase is still in progress.

The ultimate purpose of this paper is threefold : (1) establishing the requirements that are necessary to build acted emotional speech recordings, (2) describing the methodology of creating acted data, and (3) assessing the neutral and angry wav-files in order to identify unskilled actors among the initial volunteers. The corpus has been evaluated using both listening assessments and statistical tests. Listening assessments have been conducted by 11 referees. However, the statistical tests use the paired t-test to compare the means of pitch values extracted from neutral and angry wav-files.

The remainder of this paper is structured as follows. The requirements for creating acted corpora are presented in Section II.

The description of ESCAD is provided in section III. The results of assessments are presented in Section IV. Finally, section V concludes the first phase and provides perspectives.

II. REQUIREMENTS FOR CREATING ACTED CORPORA

The requirements that we have established to choose and manage actors, along with selecting sentences are inspired from the work of Burkhardt et al. [5]. Actors are asked to simulate emotions while reading a set of sentences that include the phonemes of AD.

The main requirements are outlined as follows :

A) Actors

- the number of actors who take part in the recording of the corpus must be high. This condition must be satisfied to allow testing speaker-independent models,
- the process of recording must not favorite an actor against another one. Therefore, the same sentences have to be uttered by all actors,
- the same number of wav-files must be recorded for each emotion. Therefore, all sentences have to be used to simulate emotions,
- sentences have to be proclaimed to the actor in advance to avoid articulation effects,
- it is possible for actors to listen to wav-files that convey the prerecorded emotions. This step can help them to easily mimic emotions,
- in certain cases, actors are requested to recall memories associated with the emotion to be simulated,
- to reduce the variability of speech recordings, the microphone must be positioned carefully and consistently so that the distance between the actor (speaker) and the microphone is fixed,

B) Sentences

- sentences must be free from emotional bias. Sentences like "Go out, I don't want to see you again" have to be avoided. The main reason is that it conveys negative feelings and angry emotion.
- vowel-consonant ratio has to be high,
- the recording quality of wav-files should be high,

A summary of the most important ESC is given in Table 1. A special attention is put on acted ESC.

III. CORPUS DESCRIPTION

This section provides a global description of ESCAD. A brief report on AD is given in the first subsection. The second subsection defines the framework (scope) of the corpus. The remaining subsections highlight the main components of the corpus (Fig. 1).

A) Algerian dialect

Depending on the geographical region, several dialects are spoken in Algeria. The vocabulary of these dialects consists of altered SA words and numerous external expressions and words borrowed from French, Spanish, Ottoman Turkish and Amazigh [14, 15].

Algerian people are using dialects to communicate [12]. However, both SA and French language are used in official writing.

We can distinguish two main categories of dialects in Algeria, Hilalian and pre-Hilalian [16]. The dialect spoken in Algiers (the capital of Algeria) and its outskirts, i.e., "Algiers dialect" (AD) belongs to the second category. More details about the hierarchy of dialects can be found in [16]. In this paper, the focus of attention is on the dialect spoken in the capital and its outskirts (the mother tongue language of our actors).

Table 1. Acted Emotional Speech Corpora

Ref.	Language	# of			Fs	ASD
		*speakers**	*emotions*	*sentences*		
[8]	German	10 (5m+5f)	7	10	48	6
[11]	Standard Arabic	23 (10m+ 13f)	5	2	16	4
[12]	Standard Arabic	1f	4	6	44.1	/
[13]	Dialect: Tunisian	12 (6m+6f)	5	12	8	5

ASD: Average sentence duration (s)
Fs: Sampling Frequency (kHz)
***male '**m**', female '**f**'

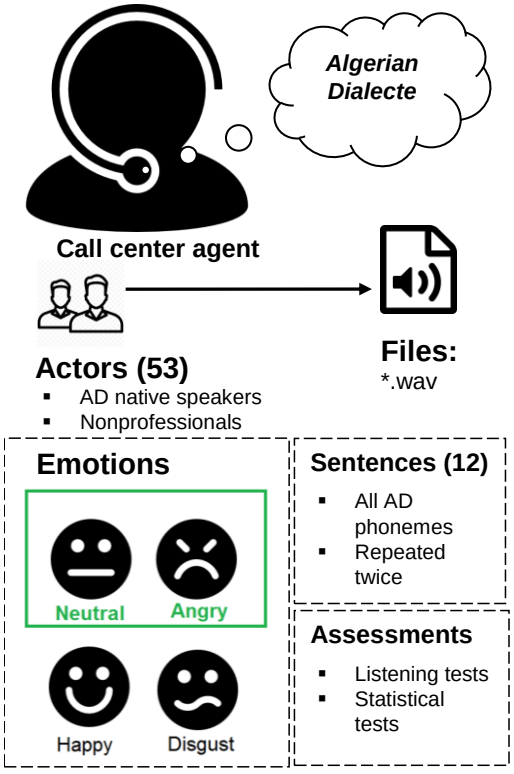

Fig. 1. Overview of ESCAD. Emotions in green color are simulated in the first phase of the project.

The AD encompasses all the SA consonants (a total number of 28 phonemes) besides the three French (FR) phonemes /p/, /g/ and /v/[1] (Fig. 2) [17, 18]. It is worth noting that /θ/ (ث), /ð/ (ذ), and /ðˤ/ (ظ) consonants are uttered as /t/ (ظ), dˤ/ (ض), and /d/ (د), respectively [17]. Depending on the origin of words, either FR vowels or SA ones are employed to compose the desired word.

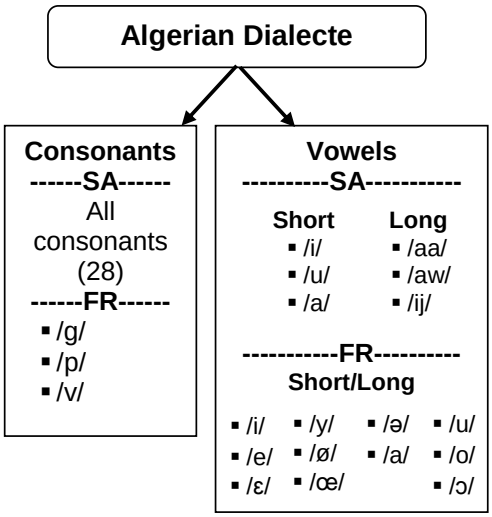

Fig. 2. AD phonemes

To the best of our knowledge, the evolution of formants and fundamental frequencies of AD phonemes has not been studied before. Moreover, an official dictionary of AD has not been established yet. More details about the grammatical and syntaxial guidelines of AD are available in [17].

B) Scope

Being an agent n call centers is an exhausting and stressful profession. Every day, agents are dealing with customers of different kinds. Some of them are nice, others are moody, and a minority may behave offensively. However, agents must provide admirable service by handling customers stress and excessive emotions. Therefore, the monitoring of agents' mood and emotions is vital to preserve the company's reputation. As a rule of thumb, aggressive agents should be fired. However, polite agents must be rewarded.

The design of a machine learning method which can identify this category of agents requires a speech corpus that includes wav-files that convey agents' emotions. Therefore, in this paper, we are creating the first acted ESC of AD that includes expressions spoken by call center agents. Our actors are simulating the aforementioned emotions by mimicking call center agents of electricity distribution companies.

C) Actors (speakers)

One of the most important steps in the creation of acted speech corpora is the choice of actors. It is obvious that the ideal candidates for this task are theatre or movie actors. However, it is not an easy mission to ask professional actors to join our team and convince them to simulate the entire corpus for free. Therefore, our approach consists in seeking ordinary persons (nonprofessional speakers) among the employees of our research center who have good abilities in mimicking voices and acting.

Yet, we have managed to identify 53 talented actors among the employees of our center who accepted to join our team and simulate the desired emotions.

D) Key sentences

We have made an investigation on the main expressions commonly employed by call center agents in Algiers and more precisely those agents working at electricity distribution companies. After that, we have written 12 different expressions (sentences) that involve high vowel-consonant ratio. Furthermore, these expressions encompass all AD phonemes and are free from any emotional bias. The IPA notation have been used to transcribe the entire set of sentences (Table 2).

E) Emotions included in the entire corpus

ESCAD includes four main emotions, namely, neutral, angry, happy and disgust. Happy discloses positive feeling; however, angry and disgust unfold negative sentiments. All these emotions are acted by ordinary actors.

F) Emotions includes in the first phase of the project

The task of mimicking voices or emotions by ordinary individuals is difficult due to two main issues: (i) the actors (speakers) are not available all the time to record the corpus, (ii) moreover, they need many repetitions to mimic the desired emotions.

To overcome these issues, we managed the recording schedule according to our actors' free-time. Furthermore, in the first phase, we have restricted our

[1] IPA phonetic notation [25]

simulations to mimic only two emotions. We have chosen neutral (N) and angry (A) states since they can easily be imitated by nonprofessional speakers. Once we accomplish the first phase of the project, we will explore the results of assessments to establish the main guidelines of actors' performance. Later, the actors who are not skilled will not take part in the recording of happy and disgust emotions.

G) Process of recording

The process of recording has been conducted in the lecture theatre (Fig. 3). The level of noise in this theatre is 35 dB. The number of actors who participated in the simulation of neutral and angry emotions is 53 (Table 3).

Table 2. Key expressions (sentences)

	Sentences	ASD*
1	lafaktyɣ ntaaʕak mahijʃ xaalsˁa لافكتور نتاعك ماهيش خالصة	1,507
2	wijn teskun وين تسكن؟	0,621
3	likip teknik mahijʃ hna ʔasˁbar ħattaa jadduxlu لكيب تيكنيك ماهيش هنا أصبر حتى يدخلو	2,300
4	ʔarwaaħ tʃuwf mʕa lwakala أرواح تشوف مع الوكالة	1,270
5	madam raakı matˁrafikja lkɔ̃tœɣ مادام راكي مطرافيكيا الكونتور	1,734
6	waaʃ naqdar ndirlak xuwja واش نقدر نديرلك خويا؟	1,222
7	xallasˁ lfatˁuwra ʔumbaʕd ʔʃki خلص الفاطورة أومبعد اشكي	1,487
8	lħdijd wal fuwr ʔiliktrijk jaaklu bazzaaf trisiti الحديد والفور الكتريك ياكلو بزاف تريسيتي	2,419
9	ʕandkum yn maas عندكم أون ماس	1,002
10	lɛ̃nstalasið taʕ lgaaz taʕak mahijʃ fi li nuwrm لانسطلاسيون تع القاز تاعك ماهيش في لي نورم	2,246
11	ɣadwa nvirɪfjijwlak lapaan غدوة نفيريفيولك لابان	1,403
12	dˁurk jdʒi lɣispðsaabl ضرك يجي الريسبونسابل	1,198

***ASD:** Average sentence duration (s)

Table 3. Gender and age of actors

Gender	Age (in years)			# of Actors
	min	*Max*	*Mean*	
Male	29	57	40.9	25
Female	23	57	37.9	28

Fig. 3. Process of recording.

At the beginning of the recording, the actor is asked to sit on a chair. Later, the person who is monitoring the recording explains the process of simulation and informs the actor that he is free to use his body language while simulating angry state.

The actor is asked to read a single sentence in one step. The majority of actors have chosen to sit on the chair. However, some of them, have preferred to stand-up and move their hands while simulating angry states. The speech waveform was acquired through an omnidirectional microphone attached to the collar of the speaker's t-shirt. It is worth mentioning that the microphone is designed to operate between 100Hz and 18kHz. Furthermore, its signal to noise ratio and sensitivity are respectively -74 dB and 30dB. This microphone was connected to the sound card of the computer (the brand of the soundcard is SRS Premium). The main requirements of choosing and managing actors, along with selecting sentence were satisfied.Actors were free to repeat the recording of a given sentence if needed. The simulation of neutral states is conducted at first.

After that, our actors were requested to simulate angry mood. For each emotion, the best two repetitions have been recorded at a sampling rate of 44.1kHz. We have exploited the AVS audio editor to record the wav-files [19]. The ".wav" format was used. The number of wav- files we have created is 2543. Unfortunately, 17 files have been corrupted due to manipulation errors.

Therefore, our corpus consists of 2527 wav-files (Table 4).

We have conducted a preprocessing phase that involved two stages. The first stage reduces the recording noise by applying noise removal effect of AVS software. The second stage equalizes and

normalizes the dynamic range of the data. The same software was used to complete the second stage. The wav-files of ESCAD are identified as follows:

- Position from 1 to 2: actor' ID (01 to 57),
- Position 3: actor's gender ('m' or 'f'),
- Position from 4 to 5: sentence ID (01 to 12),
- Position from 6 to 7: sentence recurrence S1 or S2,
- Position 8: emotions ('N' or 'A').

The original wav-files that are recorded without applying preprocessing have also been achieved autonomously in another directory.

Table 4. Overview of ESCAD (Phase 1)

Dialect	Algiers and its outskirts
Corpus type	Acted
Emotions of phase 1	Neutral (N) and Angry(A)
# of actors	53
# of sentences	12
# of recurrences	2
Corpus duration	1h 04min 40s
Sampling rate	44.1kHz
# of wav-files	2527

H) Assessment of the corpus

The goal of this task is to provide answers to these questions:

(i) What is the mathematical proof that neutral and angry wav-files are not alike?

(ii) How to identify actors who are less skilled at mimicking both neutral and angry emotions?

The answer to the first question is provided through statistical tests based on paired t-test [20]. However, the answer to the second question is given by conducting listening tests.

- **Statistical tests**

It has been reported in the literature that prosodic features and speech quality measures can reveal the acoustic distinction between angry and neutral states [24]. To facilitate this task, we have used a single prosodic feature, namely, pitch variations to conduct the test. We have utilized Gonzalez and Brookes method [21] to compute the pitch contour of each wav-file. Next, we have conducted a dependent t-test for paired samples [20] to test the statistical significance of the differences between neutral and angry emotional states. The dependent variable that we have considered is the mean value of the pitch contour (MF). The paired samples consist of MF values computed using the pairs of neutral and angry wav-files (all actors are considered). An initial examination has been led to test the hypothesis of homogeneity of variance and normality of the data by applying Bartlett' tests and Anderson-Darling respectively [22].

The null hypothesis is given as follows:

"the true mean difference between the paired samples is null,"

The alternative hypothesis is the opposite and can be expressed as follows :

"the true mean difference between the paired samples is not null,"

The null hypothesis will be rejected if the difference between sample means is too large or too small. Otherwise, it will be accepted. The significance level we have adopted in our study is 5%.

- **Listening tests**

The assessment of ESCAD has been led by 11 referees (7 females and 4 males) who are native speakers of AD. They do not have listening difficulties or hearing impairments. Listening tests are led separately for each referee (evaluator). We have played the wav-files of ESCAD one after another in a random order through headphones.

The referee had to predict the emotion conveyed by the wav-file being assessed. We have proposed three different tags (classes): (1) angry (a), (2) neutral (n) or (3) other type of emotions (o) (the emotion is neither neutral nor angry) (Fig. 4).

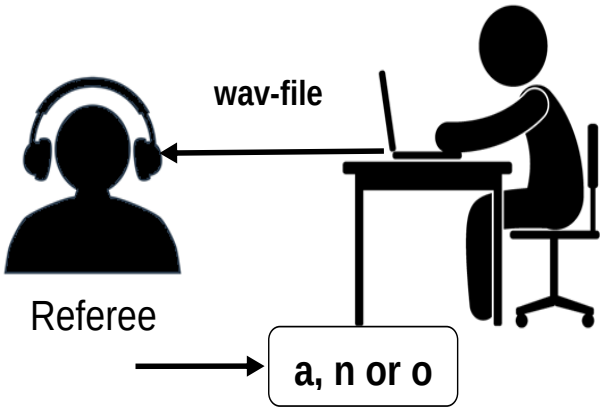

Fig. 4. Listening test

For every single wav-file (WF), we have computed the scores of each tag and assembled them in one vector. The vector is named "S" and is composed of three components, namely, "Sa", "Sn" and "So" respectively. Each component ranges between 0 and 100%. "Sa" denotes the percentage of referees who tagged WF as angry "a" (recognition of

angry), "Sn" denotes the percentage of referees who tagged WF as neutral "Sn" (recognition of neutral), and "Sn" denotes the percentage of referees who decided that WF is conveying another type of emotion "o".

The overall decision (OD) of our referees corresponds to the tag with the highest score (HS). For instance, the third wav-file has been recorded by a female actor to simulate anger.

The scores "Sa", "*Sn" and "So" that we have obtained through the assessment are respectively 85%, 15%, and 0%. This result means that 85% of the referees confirmed that the third wav-file is conveying angry emotion, 15% of them considered this wav-file to be neutral and none of them considered the same file to disclose another type of emotions. The emotion conveyed by the wav-file being tested is correctly simulated (or acted) if the following conditions are satisfied:

(i) The OD of our referees refers to the emotion simulated by the actor,

(ii) the HS must be greater than a precise threshold "Th".

else, the emotion conveyed through this wav-file is wrongly acted.

We have computed the individual assessment of the actor "IAA" by dividing the total number of wav-files recorded by the same actor into the number of correct assessments. Furthermore, this metric is computed for both neutral and angry emotions as follows :

$$IAA_N = \frac{\#\text{of CANV}}{\#\text{of neutral wav-files}} \times 100 \quad (1)$$

$$IAA_A = \frac{\#\text{of CAAV}}{\#\text{of angry wav-files}} \times 100 \quad (2)$$

Where CANV and CAAV are respectively the correct assesments of neutral wav-files and the correct assesments of angry wav-files.

It is also generalized to include both neutral and angry states.

$$IAA_G = \frac{\#\text{of CARAF}}{\#\text{of all wav-files}} \times 100 \quad (3)$$

Where CAAF is the correct assesments of all wav-files.

An actor who has an IAA that is greater or equal to a specific threshold "Th" is considered to be skilled (he is a good actor).

IV. RESULTS AND DISCUSSIONS

A) *Paired t-test*

The paired t-test was carried out using MATLAB software. The obtained results indicate that t-test rejects the null hypothesis at the 5% significance level. This result means that the true mean difference between the paired samples is not null.

In other ways, we can state that the mean value of the dependent variable MF computed using neutral wav-files is different from the one computed using angry wav-files. Therefore, we have provided the statistical proof that neutral and angry wav-files of ESCAD are not alike.

B) *Listening tests (acting performance)*

In order to identify the group of actors who are unskilled in mimicking neutral and/or angry emotions over other actors, we have conducted the listening tests described in section 3.Then, we have computed the IAAN, IAAA and IAAG for each actor by applying a threshold "Th" of 85%. The larger the AAS is, the more brilliant the actor is. Next, the exact percentage of actors who have an IAA greater or equal to "Th" has been estimated. It is worth noting that the threshold we have set in our study is more stringent than the one reported in [13]. The main findings are discussed as follows. We have discovered that all our actors show better performance in simulating neutral sentiments than angry temper (Fig. 5). Furthermore, we have identified 18 actors among the initial volunteers (53 actors) who are less skilled at simulating neutral or angry voice sentiments (Table 5).

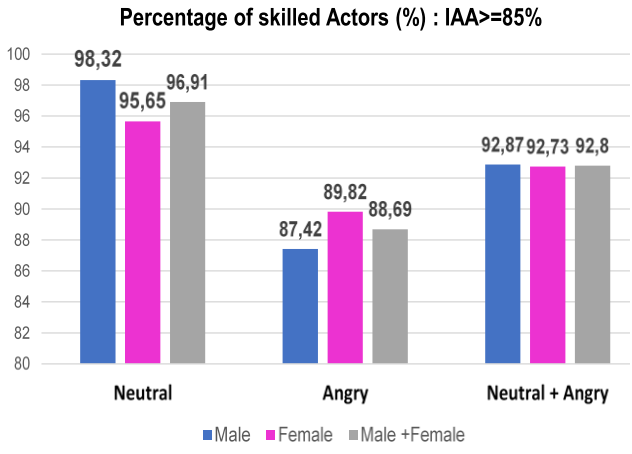

Fig. 5. Global assessment of ESCAD (Skilled Actors: IAA>=85%)

Table 5. Determination of Less skilled Actors (IAA<85%)

Speakers / Emotion	Actors' ID	Gender*
Neutral	{19, 32, 52}	3f
Angry	{6, 12, 1, 14, 25, 23, 13, 16, 20, 35, 29, 22, 2, 11, 34}	7f+8m
Neutral+Angry	{6, 12, 1, 14, 25, 23, 13, 19}	5f+3m

*****male '**m**', female '**f**'

These actors have an IAA which is less than 85%. Most of them are females.

V. CONCLUSION

We have described a project whose mission is to create the acted emotional corpus of Algerian dialect speech. The corpus is named ESCAD and encompasses speech emotions simulated by 53 native actors (speakers). The entire corpus includes four main emotions, namely, neutral, angry, happy and disgust. In order to reduce the time allocated to record and assess the corpus, we have conducted the project in two main phases. The goal persuaded by the first phase is twofold : (i) simulate neural and angry emotions, and (ii) identify skilled actors (speakers) among the initial volunteers. In the second phase, these actors are going to simulate happy and disgust emotions. This phase will allow us to assess the entire data. We have achieved the first phase of the project and have assessed the recorded data based on both listening tests and statistical analysis.

We have explored the paired t-test to prove that neutral and angry wav-files of our corpus are statistically different. Thereafter, 11 referees have conducted the listening tests to identify the groups of unskilled and skilled actors among the initial list of volunteers. We have established stringent criteria for identifying these actors based on their performance in simulating neutral and angry emotions. The group of unskilled actors will not take part in the simulation of happy and disgust emotions. Our next phase will be the simulation of happy and disgust emotions and the assessment of the entire speech corpus.

VI. ACKNOWLEDGMENT

The authors would like to express their gratitude to the actors and the referees who participated in this work. This research was funded by the ATRSS (Agence Thématique de Recherche en Science de la Santé) under grant n°: 64/DFPR/ATRSS/2017.

VII. REFERENCES

[1] M. Hagmueller, E. Rank, and G. Kubin, *Evaluation of the human voice for indications of workload induced stress in the aviation environment* (Eurocontrol experimental Centre, 2006).

[2] F. Ykhlef, W. Benzaba, R. Boutaleb, J.B. Alonso, and Far. Ykhlef, Yet another approach for the measurement of the degree of voice normality : A simple scheme based on feature reduction and single Gaussian distributions, *Proc. 17th IEEE International Symposium on Multimedia*, Miami, FL, USA, 2015.

[3] H. Dahmani, H. Hussein, B. Meyer-Sickendiek, and O. Jokisch, Natural Arabic language resources fore emotion recognition in Algerian dialect, *Proc. International Conference on Arabic Language Processing*, Nancy, France, 2019.

[4] M. Hossain, and G. Muhammad, Emotion recognition using deep learning approach from audio–visual emotional big data, *Information Fusion, 49*, 2019, 69-78.

[5] P. Ekman, Basic Emotions, in T. Dalgleish and M. Power (Eds.), *Handbook of Cognition and Emotion* (John Wiley & Sons, Ltd, 1999) 45-60.

[6] A., Konar, and A. Chakraborty, *Emotion Recognition*: *A Pattern Analysis Approach* (Wiley, 2015).

[7] S. Klaylat, Z. Osman, L. Hamandi, and R. Zantout, Emotion recognition in Arabic speech, *Analog Integrated Circuits and Signal Processing, 96 (2)*, 2018.

[8] F. Burkhardt, A. Paeschke, M. Rolfes, W.F. Sendlmeier, and B. Weiss, A Database of German emotional speech, *Proc.* Interspeech'05, Portugal, 2005.

[9] SR. Livingstone and FA. Russo, The Ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in north American English, *PloS one*, *13(5)*, 2018, e0196391.

[10] Rao. K. Sreenivasa, and G. Shashidhar Koolagudi, *Robust Emotion Recognition using Spectral and Prosodic Features* (Springer Publishing Company, 2013).

[11] A. H. Meftah, Y. A. Alotaibi, and S. Selouani, Evaluation of an Arabic speech corpus of emotions: A perceptual and statistical analysis, *IEEE Access*, *6*, *2018*, 72845-72861.

[12] M. B. Othman, *Analyzing acoustic markers of emotion in Arabic speech*, Master of Philosophy, University of Manchester, Uk, 2017.

[13] M. Meddeb, H. Karray, and A.M. Alimi, Building and analysing emotion corpus of the Arabic speech, *Proc. IEEE International Workshop on Arabic Script Analysis and Recognition*, Nancy, France, 2017.

[14] K. Meftouh, N. Bouchemal, and K. Smaili, A study of a non-resourced language: an Algerian dialect," *Proc.3rd International Worshop on Spoken Languages Technologies for Under-resourced Languages*, Cape-town, South Africa, 2012.

[15] I. Guellil, and F. Azouaou, Arabic Dialect Identification with an unsupervised learning (based on a lexicon) application case: Algerian dialect, *Proc. 19th IEEE International Conference on Computational Science and Engineering*, Paris, France, 2016.

[16] K. Versteegh, *The Arabic language* (Second Edition, Eedinburgh University Press, 2014).

[17] S. Harrat, K. Meftouhy, M. Abbass, K. Hidoucix, and K. Smaili, An Algerian dialect: study and resources, *International Journal of Advanced Computer Science and Applications, 7(3),* 2016, 384-396.

[18] F. Chouireb and M. Guerti, Towards a high quality Arabic speech synthesis system based on neural networks and residual excited vocal tract model, *Signal, Image and Video Processing, 2(1),* 2008, 73-87.

[19] AVS editor : https://www.avs4ou.com, last accessed 01/01/2020.

[20] C. F. J. Wu, and M. Hamada, *Experiments : planning, analysis, and parameter design optimization* (Wiley, 2000).

[21] S. Gonzalez, M. and Brookes, PEFAC - a pitch estimation algorithm robust to high levels of noise, *IEEE Trans. Audio, Speech, Language Processing, 2 (22),* 2014, 518-530.

[22] M.L. Hugh, Multiple comparison analysis testing in ANOVA, *Biochem Med, 21(3),* 2011, 203-209.

[23] A. Castellanos, J.M. Benedi, and F. Casacuberta, An analysis of general acoustic-phonetic features for spanish speech produced with Lombard effect, *Speech Communication, 20,* 1996, 23-36.

[24] F. Vog, *Real-time automatic emotion recognition from speech*, Phd, University of Bielefeld, Germany, 2010.

[25] International Phonetic Association, *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (Cambridge University Press, 1999).