# Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect

**Habiba Dahmani [1], Hussein Hussein [2], Burkhard Meyer-Sickendiek [2] and Oliver Jokisch [3]**

(1) Department of Electrical Engineering, University of Mohamed Boudiaf, Msila Algeria
(2) Department of Literary Studies, Free University of Berlin, Berlin, Germany
(3) Institute of Communications Engineering, Leipzig University of Telecommunications (HfTL)

ICALP 2019 – 7th International Conference on Arabic Language Processing

الندوة الدوليه حول المعالجة الآلية للغه العربيه

October 16th-17th 2019, Nancy, France

# Contents

- Objective and introduction
- Emotion recognition on Arabic
- Database (incl. case study Algerian data)
- Experiment
- Results
- Conclusion and future work

# Objective

- Building  a natural Arabic audio-visual database for the computational processing of emotions and affect in speech and language which will be made available to the research community

# Introduction I

- Automatic Recognition of emotions from speech and image: includes understanding of natural language → emotions detection and classification within the speech signal possible.
- The information expressed through speech can be divided into three categories:
  - linguistic information: such as accent, phrase and sentence type
  - paralinguistic information: for example: intention, attitude and speaking style
  - nonlinguistic information: such as age, gender, physical and emotional states of speakers.
- Recognizing human emotions mainly from speech signal is a challenging process for many reasons. In general, studies mention two principal difficulties: audio-video databases and recognition algorithms which will allow for quick and accurate emotion identification.

Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect, H.Dahmani/H.Hussein/B.Meyer/O.Jokisch

# Introduction II

- Developing and creation of reliable database is a requirement for building emotion recognition systems.
- A significant emotion recognition obstacle is the quality of the recorded speech samples (quality, size, and the type of the database).
- The speech corpora are either:
  - acted / simulated
  - induced or
  - natural.
- The most used recognition algorithm and successful classifiers are : k-nearest neighbor (K-NN), Gaussian mixture model (GMM), Hidden Markov models(HMM), support vector machine(SVM), decision tree algorithms, and artificial and deep neural networks (ANN, DNN).

Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect, H.Dahmani/H.Hussein/B.Meyer/O.Jokisch

# Emotion (recognition) in Arabic language

- Arabic language and its dialects are still considered a relatively resource-poor language when compared to other languages such as English
- The Arabic language has three forms:
  - Classical Arabic or literary Arabic language,
  - Modern Standard Arabic (MSA),
  - Colloquial Arabic.
- Concerning Arabic affective computing, there has been a considerable amount of works on the collection of emotional speech.

Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect, H.Dahmani/H.Hussein/B.Meyer/O.Jokisch

# Databases of Arabic emotional speech

| Name | Type | MSA or dialect | Speakers | Linguistic material | Emotions |
|---|---|---|---|---|---|
| **KSUEmotions (2014, 2018)** | simulated | MSA | 20 (10 females and 10 males) | 16 sentences | 5 emotions: neutral, sadness, happy, surprised, and questioning |
| **Egypt (2017)** | acted | MSA | 7 | 500 sentences | 6 emotions: happiness, sadness, fear, anger, inquiry, neutral |
| **REGIM_TES (2015-17)** | acted | Tunisian | | | |
| **Arabic-Natural Audio-Dataset (2018)** | natural | Dialectal: Egyptian, Jordan, Gulf, Lebanese | 6 | | 3 emotions: happiness, anger, and surprise |

- The table shows the scarcity of the Arabic linguistic resources and confirm too that little work has been devoted for the analysis of emotional speech in Arabic but some pioneer work around 2005, 2006, 2011.

➔ Hence, need of such work within Arabic natural language processing (NLP) is motivated.

Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect, H.Dahmani/H.Hussein/B.Meyer/O.Jokisch

# Our database

- The undergoing database consists of a **collected data of spontaneous emotional speech in Arabic language, including MSA and colloquial Algerian**, Tunisian, Lebanese, Jordanian, Syrian, and Egyptian.
- The database consists of **audio-visual recordings** of some Arabic TV talk shows, segmented into broadcasts. The corpus contains **spontaneous and very emotional speech** recorded from **discussions between the guests** of the talk shows.
- So far, We collected about **50 broadcasts of the talk shows** for Arabic (including MSA, and Algerian, Tunisian, Lebanese, Jordanian, Syrian, and Egyptian dialects). The number of programs and dialects can be expanded further to balance the database.

# Selected Arabic Data I

- A brief description of the selected TV show programs is given below:

1. **For Arabic MSA**, we propose the Opposite Direction (Arabic: الاتجاه المعاكس) is a debate TV show handling current events in the Middle East and the Arab world. The topics handled are mostly influenced by political, economic or social topics,
2. **For Arabic Algerian**: "Open Your Heart" ( Arabic:إفتح قلبك) is an Algerian psychosocial reality television talk show. It is the local adaptation of the French series "Y'a que la vérité qui compte." .
3. **For Arabic Tunisian**: I have what I say to you or Andi Mankolek (Arabic : عندي ما نقلك) is a Tunisian version of the French program "Y'a que la vérité qui compte". the same principle of the Algerian show "open your heart.
4. **For Arabic Lebanon**: AalAkid on Future TV, The show was a Lebanese adaptation of the popular French show "Sans aucun doute," presented by Julien Courbet daily on TMC.

# Selected Arabic Data II

5. **For Arabic Egyptian**: A Comedy TV show Program (Candid Camera): you can do it or not (Arabic: قدها ولا مش قدها). Program dumps in the form of a contest. Participants perform tasks in a public place, in front of their friends or family members from three different The program scheduled daily on the screen of Cairo and people TV channel[15].
6. **For Arabic Jordanian**: Clear jammer program( Arabic: تشويش واضح:) Is a satirical comedy program that discusses the latest political, economic and social news at the local and Arab levels as well as the international level in a comical and comic style.
7. **For Arabic Syrian**: lady Nour (in arbic **نور خانمor** Noor Khanam), a young Syrian media anchor on the Syrian television show. This program offers a variety of paragraphs, from mock sketches and sarcastic dialogues.

# Case Study: Algerian Emotional Speech

- We focus in this study on the **Algerian dialect** which is classified by the Algerians themselves as a **mixture of three languages: Arabic, Berber, and French**.
- This dialect is characterized by the multitude including sub-dialects that are clearly variants of Arabic, and other sub-dialects that are non-Arabic which we call the Amazigh dialect.
- Algerian dialect is generally described as an Arabic idiom attached to the **Maghrebian Arabic group** (Algerian, Moroccan, Tunisian and Libyan). However, its morphology, syntax, pronunciation, and vocabulary are quite different from other Arabic dialects.
- The Algerian and the other North African dialects are considered a little **distant from the MSA**

Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect, H.Dahmani/H.Hussein/B.Meyer/O.Jokisch

# Case Study: Algerian Emotional Speech

- A small speech database which **"Red Line", a weekly social program on Al-Shorouk Algerian channel. A talk show where guests are invited to talk.** The show is presented by **three permanent hosts**; anchorwomen who appear on each program, introduces and interacts with the guests. Religious and psychological opinions are present. Social program sometimes deals with sensitive subjects that are difficult to discuss.

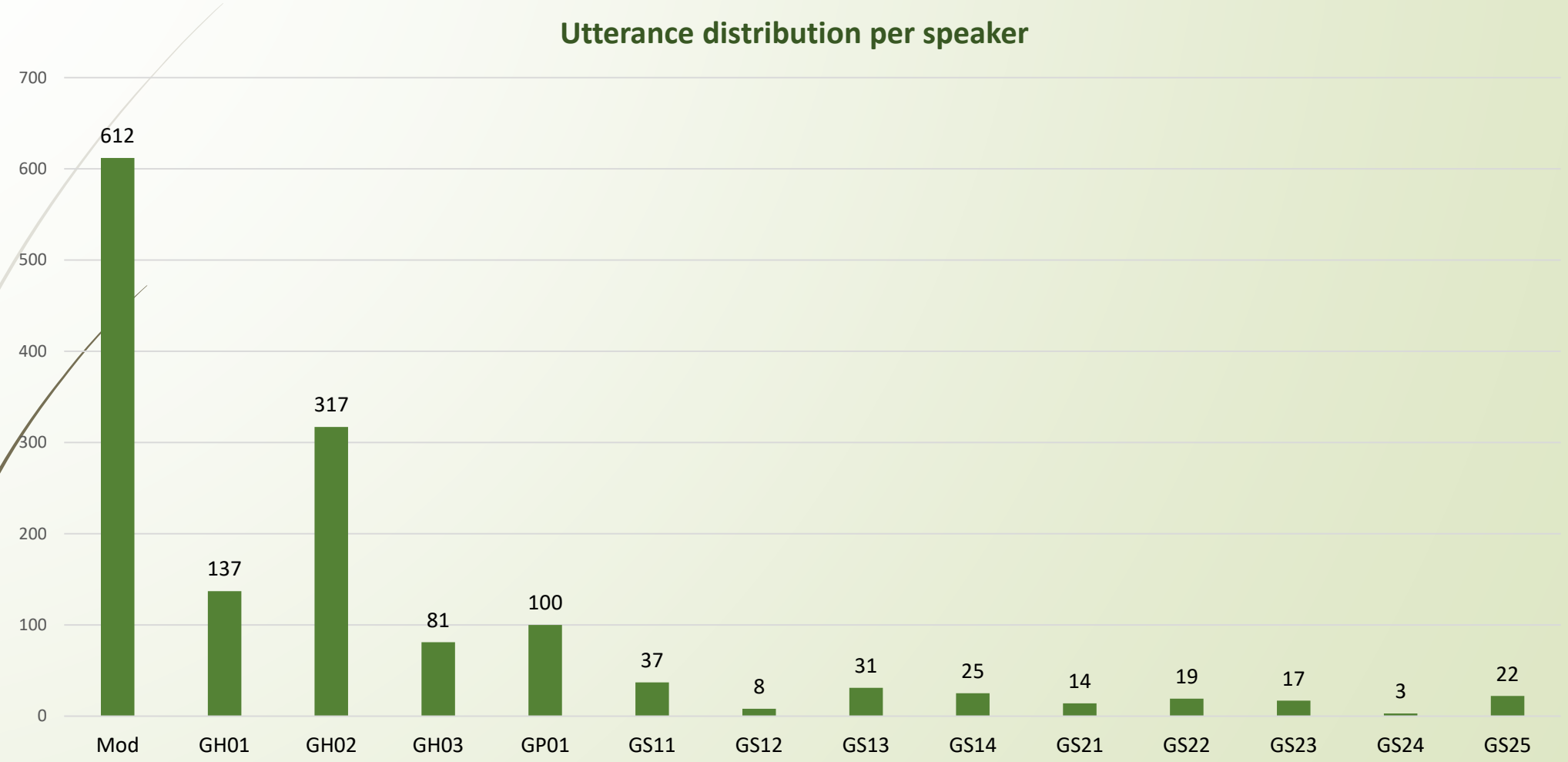
- The **video** files are MPEG-coded image sequences of 352 × 288 pixels with a frame rate of 25 fps. A constant code rate of **1.15 Mbit/s** was used. Recordings were taken with a sampling frequency of **48 kHz and later downsampled to 16 kHz (16 bit)**. These criteria are commonly used for speech databases

- Both MSA and dialects are used for communication.
- The speech corpus consists of records of **two hours** which are segmented in the first step into smaller units or clips containing the whole dialogue between number of talk show guests.

- It consists of **1,443 utterances** from **14 different speakers** among them **5 females** (two little girls and 3 adult females). But it should denoted that there are actually **three dominant speakers**

Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect, H.Dahmani/H.Hussein/B.Meyer/O.Jokisch

# Case Study: Algerian Emotional Speech



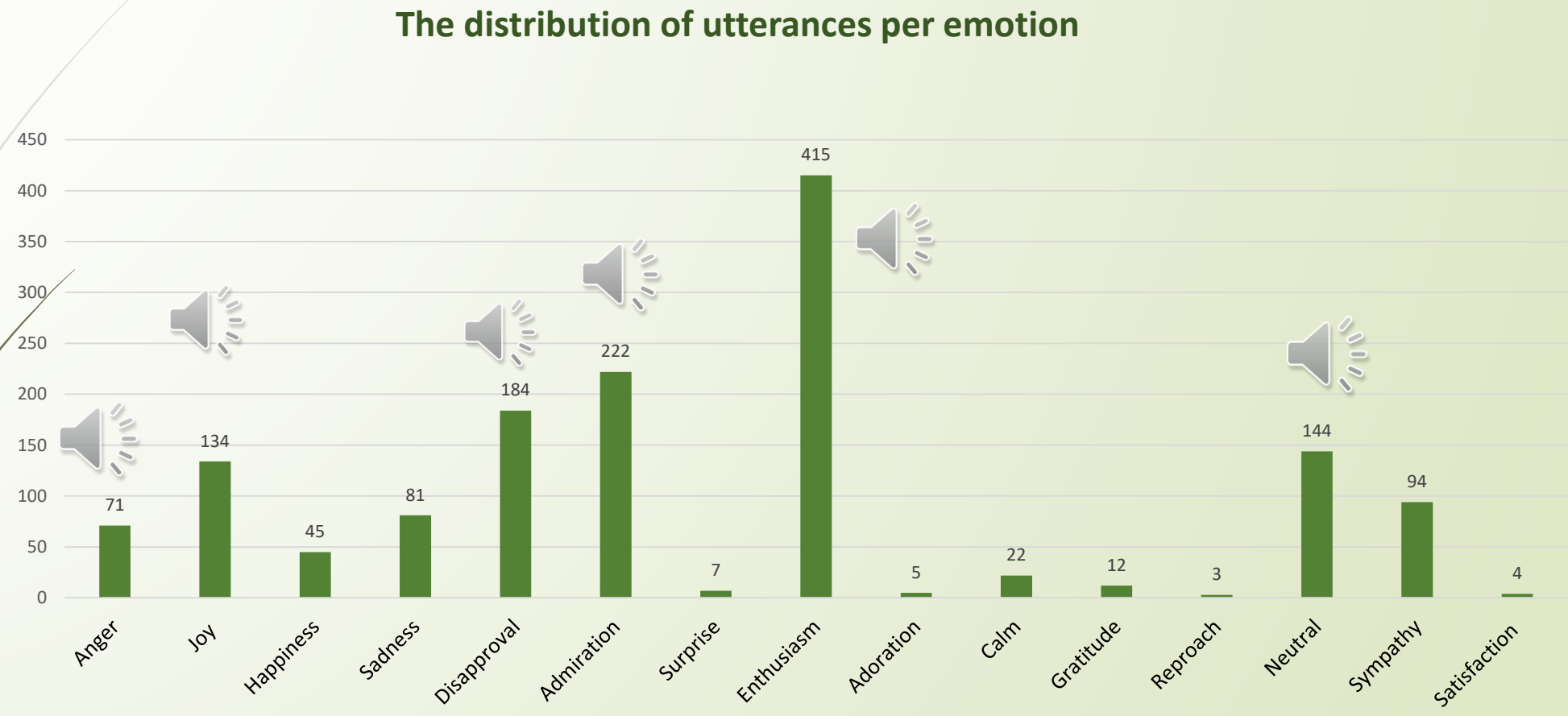

Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect, H.Dahmani/H.Hussein/B.Meyer/O.Jokisch

# Case Study: Algerian Emotional Speech

- **15 emotions were detected in the collected database** and rated for the present data: Anger, Joy, Happiness, Sadness, Disapproval, Admiration, Surprise, Enthusiasm, Adoration, Calm, Gratitude, Reproach, Neutral, Sympathy, and Satisfaction.

- There are about **five emotions that are more present than others** (Enthusiasm, Admiration, Disapproval, Neutral, and Joy) as shown in Fig. 2.

# Case Study: Algerian Emotional Speech




Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect, H.Dahmani/H.Hussein/B.Meyer/O.Jokisch

# Experiment: features

- We perform feature extraction for emotions by using the *openSMILE* feature extraction tool.
- Feature extraction can be split into two categories according to the processing domain:
  - The time-domain features : Zero-Crossing Rate (ZCR) and short-time average energy.
  - Frequency-domain features: pitch or fundamental frequency ($F_0$), spectral features (band energy, spectral roll-off, spectral flux, and spectral centroid),cepstral features (MFCCs), and linear prediction features (LPC).
- The feature set contains features which result from LLDs with the corresponding delta coefficients (ΔLLD) and statistical functionals applied to each of the LLD and ΔLLD. The f statistical functionals are used for every feature: min, max, range, standard deviation and mean.

# Experiment: feature vectors

- **A** (120 features): MFCCs
- **B** (50 features): Energy, pitch, ZCR
- **C** (230 features): Energy, pitch, ZCR, spectral features
- **D** (350 features): Energy, pitch, ZCR, spectral features, MFCCs
- **E** (430 features): Energy, pitch, ZCR, spectral features, MFCCs, LSP
- **F** (384 features): Baseline feature set of the Emotion Challenge in the Interspeech 2009.
- **G** (6373 features): Baseline feature set of the Computational Paralinguistics Challenge (ComParE) in the Interspeech 2013
- **H** (6552 features): The large *openSMILE* emotion feature set with more functionals and more LLD.

# Experiment: Classification

- Following machine learning algorithms with default values using the *WEKA* data mining toolkit are applied for the recognition of emotions:
  1. **IBk**: the Instance-Based (IB) classifier with a number of (k) neighbors is the K-Nearest Neighbours (KNN) classifier using the euclidean distance and 1-nearest neighbour.
  2. **AdaBoostM1**: the boosting algorithm uses the Adaboost M1 method.
  3. **LogitBoost**: The classifier performs additive logistic regression.
  4. **SimpleLogistic**: a classifier for building linear logistic regression models
  5. **RandomTree**: Random trees is a collection of decision trees that considers K randomly chosen attributes at each node [16].
  6. **RandomForest**: The classifier of random forest consists of several uncorrelated decision trees.
  7. **SMO**: The Sequential Minimal Optimisation (SMO) for training a Support Vector Machines (SVM) classifier .
  8. **J48**: The J48 algorithm used to generate a pruned or unpruned decision tree

# Results

- Performance is measured using the **f-measure** which is the **harmonic mean between precision and recall**. Table shows classification results by applying **eight kinds of feature sets and eight classifiers**.
- We **only** used extracted features and classifiers to classify five emotions: **Enthusiasm, Admiration, Disapproval, Neutral, and Joy**.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| SimpleLogistic | 0.42 | 0.33 | 0.39 | 0.46 | 0.46 | 0.42 | 0.42 | 0.42 |
| SMO | 0.42 | 0.20 | 0.39 | 0.47 | 0.48 | 0.40 | 0.42 | 0.44 |
| IBk | 0.42 | 0.33 | 0.36 | 0.42 | 0.42 | 0.41 | 0.34 | 0.46 |
| AdaBoostM1 | 0.20 | 0.23 | 0.20 | 0.20 | 0.25 | 0.20 | 0.23 | 0.20 |
| LogitBoost | 0.38 | 0.35 | 0.39 | 0.44 | 0.44 | 0.40 | 0.42 | 0.40 |
| J48 | 0.33 | 0.30 | 0.34 | 0.34 | 0.37 | 0.32 | 0.34 | 0.40 |
| RandomForest | 0.40 | 0.34 | 0.40 | 0.40 | 0.44 | 0.35 | 0.32 | 0.45 |
| RandomTree | 0.30 | 0.32 | 0.34 | 0.35 | 0.34 | 0.29 | 0.28 | 0.34 |

# Experiment: Results (confusion matrix)

- the confusion matrix in the Table 3 for the five emotions by using the SMO classifier and the feature set E.

| | Admiration | Disapproval | Enthusiasm | Joy | Neutral | Sum |
|---|---|---|---|---|---|---|
| Admiration | **96** | 17 | 83 | 12 | 14 | 222 |
| Disapproval | 14 | **99** | 60 | 2 | 9 | 184 |
| Enthusiasm | 82 | 44 | **219** | 38 | 32 | 415 |
| Joy | 7 | 8 | 50 | **64** | 5 | 134 |
| Neutral | 21 | 19 | 49 | 3 | **52** | 144 |

# Conclusion

- We presented the first steps of our work whose objective is to design and build a new Natural Arabic multimodal spontaneous emotion database for the research community in order to facilitate the research in the field.
- There is insufficient interest in this language regarding the recognition of emotions or regarding all disciplines of artificial intelligence in overall.
- Arabic is one of the oldest languages in the world. It is one of the first widely used languages nowadays. So it is really vital to give a lot of importance to further study this language and its variants which are its dialects.
- The existing MSA and dialectal Arabic speech corpora are very sparse and are of low quality. For some Arabic dialects, speech resources do not exist at all.
- We measured the performance of emotion classification by using a variety of audio features and several classifiers for five emotions in the Algerian dialect.

# Future work

- We propose to use Romanization method for transcription of the speech corpus.
- Continuing to perform different annotations like prosodic, Part-of-Speech tags, and syntactic labels and segmentation in word and chunk levels.
- We will investigate both methods: the dimensional and the category labels for the emotional annotation and use the two classic criteria for assessing the quality of such labels: validity and reliability.
- We will establish different measures of impact and discuss the mutual influence of acoustics and linguistics.
- The classification will be experimented using different levels, word, chunk, and utterances. Classification performance relies on deep learning and pattern recognition techniques

# Thank you

# Case Study: Algerian Emotional Speech

- **Recording Quality:** The video files are MPEG-coded image sequences of 352 × 288 pixels with a frame rate of 25 fps. A constant code rate of 1.15 Mbit/s was used. Recordings were taken with a sampling frequency of 48 kHz and later downsampled to 16 kHz (16 bit). These criteria are commonly used for speech databases