EMO&LY (EMOtion and AnomaLY) : A New Corpus for Anomaly Detection in an Audiovisual Stream with Emotional Context

Cédric Fayet[1,2], **Arnaud Delhay**[1,2], **Damien Lolive**[1,2], **Pierre-François Marteau**[1,3]
[1]IRISA - EXPRESSION Team, Lannion&Vannes, France
[2]Université de Rennes 1, Rennes, France
[3]Université de Bretagne Sud, Vannes, France
{cedric.fayet, arnaud.delhay, damien.lolive, pierre-francois.marteau}@irisa.fr

Abstract

This paper presents a new corpus, called EMOLY (EMOtion and AnomaLY), composed of speech and facial video records of subjects that contains controlled anomalies. As far as we know, to study the problem of anomaly detection in discourse by using machine learning classification techniques, no such corpus exists or is available to the community. In EMOLY, each subject is recorded three times in a recording studio, by filming his/her face and recording his/her voice with a HiFi microphone. Anomalies in discourse are induced or acted. At this time, about 8,65 hours of usable audiovisual recording on which we have tested classical classification techniques (GMM or One Class-SVM plus threshold classifier) are available. Results confirm the usability of the anomaly induction mechanism to produce anomalies in discourse and also the usability of the corpus to improve detection techniques.

Keywords: Audiovisual corpus, Acted emotion, Reading task, Anomaly detection, Unbalanced data

1. Introduction

According to (Chandola et al., 2009), "an anomaly is defined as a pattern that does not conform to an expected normal behavior". Anomaly detection systems have been applied to various domains: intrusion detection (Axelsson, 2000), fraud detection (Abdallah et al., 2016), sensor networks (Park et al., 2010), flight safety monitoring (Li et al., 2011) or video surveillance (Ko, 2008).

In the field of human abnormal behavior detection, the focus is put on tasks like crowd modeling, violence detection (Mehran et al., 2009; Gu et al., 2014), human activity detection (Chaquet et al., 2013). In the corpora used or presented in those works, speech and/or face of the subject are usually not available. On the other hand, speech corpora available in the scope of anomaly detection are focused on disease, stress (Hansen et al., 1997) and detection of depression. For instance, the NKI-CCRT corpus (Clapham et al., 2012) has been built to study speech intelligibility before and after cancer. In (Giraud et al., 2013), a corpus containing multimodal expressions of stress during a public speaking task is presented. While this corpus is not dedicated to anomaly detection, it has the advantage to be multimodal, notably including speech and face of the participant. Such a corpus enables to study jointly facial expressions and speech.

When it comes to emotions, some corpora are also available like the Belfast Naturalistic Database (Douglas-Cowie et al., 2003) containing records of people discussing emotive subjects or the EmoTV1 corpus (Abrilian et al., 2005) containing TV interviews bearing naturalistic non-acted emotionally marked data.

All those corpora have been designed and collected to answer to specific scientific questions. As far as we know, no emotional audiovisual corpus exists containing controlled, acted or natural anomalies to address specifically the anomaly detection question.

Consequently, in this paper, we present a new and complementary multimedia corpus called EMOLY which contains human-centered anomalies. The corpus is composed of 41 subjects who read a tale three times. During the reading, we induce a reaction of the subject which can be seen as an abnormal behavior given the context. The corpus gathers various anomalous reactions in terms of intensity, modality (speech or facial expression) or emotion.

The remainder of the paper is structured as follows. Section 2. details how the corpus is designed while section 4. describes the corpus content. Finally, first experiments using the corpus are presented in section 5.

2. Corpus Design

In its first version, the corpus is constituted of 123 records from 41 participants who read three time the French version of the tale "the handless maiden" ("La jeune fille sans main") from Brothers Grimm.

2.1. Participants

Participants are either Master level students from ENSSAT (engineering school affiliated to University of Rennes 1), Licence level students from Lannion IUT or members of the lab (colleagues, PhD students). 41 participants have been recorded including 11 females and 30 males. Participants have an average age of 22 ± 2 years old. Over the 41 participants, 37 are native french speaker and four are non-native french speaker. 12 participants have had opportunities (theatre, representation, ...) to improve their communication skills during their life. Each participant has taken part to three different recording sessions. Participants are not supposed to know that we study anomalous reactions. The task is presented as the constitution of an emotional corpus.

2.2. Recording Process

To get participants accustomed to the task, the tale is sent before the first recording to each participant and the recordings are done during two separate sessions. A printed version is also available in the recording booth to give the possibility to each participant to read it one more time before

the recording session. For all three recordings, the directions given to the subjects remain the same: we ask each participant to read the tale to the camera as if she/he was telling the tale to a child. The purpose is to get expressive speech and facial expressions from them.
The first recording session is anomaly free and is only used to make the speaker more comfortable with the reading task in front of a camera. Before the recording, each participant completed a form which includes identity, personality assessment and a confidential agreement. During a second session, two recordings are carried out. In the first one, we introduce an anomaly, as explained before, to cause a reaction from the speaker. In the second one, we ask the speaker to act the anomaly of his choice. One week delay is kept between the two sessions and the participants do not know what is going to happen during the second session.
To sum up, for each participant, we obtain three recordings of the reading of a slideshow:

1. Introductory record: This is the first time that a participant discovers the slides and reads them in the context of the recording.

2. Induced record: For a particular slide, we use an anomaly induction to trigger a reaction.

3. Acted record: We ask the participant to act an anomaly for a particular slide. No guideline is given by default for this record. In case the participant asks for directions, we give some basic examples.

At the end of the second session, each participant filled in a broadcast agreement and signed a confidential agreement between them and the lab. This agreement is supposed to avoid any revelation of the true purpose of the corpus to the others participants.

2.3. Identity and Personality Assessments

The corpus is focused on humans who read a tale. Reaction(s) to the anomaly inductor is expected to be related to the speaker. If we want to study his/her reaction, we need to gather information about his personality, his speaking abilities and some standard information. In this purpose, the subject needs to fill in two assessment forms at the beginning of the first session.
The first one, called 'Identity assessment', contains questions relative to the genre, age, profession, social origin, geographical origin (birth country, and actual residence with time spent in each of them). We also asked them to evaluate their own oral abilities by answering to several questions about past training, previous jobs or tasks related to oral presentation and speaking abilities. We finally asked them for an auto-evaluation of their oral abilities on a scale from one to six (one is the lowest).
The second called 'Personality assessment' contains questions that can provide a good estimation of the personality of the subject. We choose to use the Big Five model to describe subject personality. The Big Five model has been proposed to describe personality through five personality traits: openness, conscientiousness, extraversion, agreeableness and neuroticism (John and Srivastava, 1999). To evaluate their personality traits each subject completes the BFI-Fr Questionnaire (Plaisant et al., 2010) adapted to french speakers from which a score (with a range between 1 (lowest) to 5 (highest)) is computed for each Big Five's scale. The figure 1 shows the overall distribution of each score by using box-and-whisker plots. The box extends from the lower to upper quartile values of the data, with a line at the median.

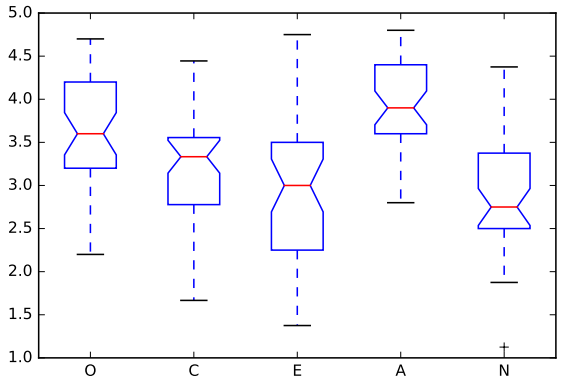

Figure 1: Distribution of the five personality traits (Openness (O), Conscientiousness (C), Extraversion (E), Agreableness (A), and Neuroticism (N)) for the 41 subjects.

2.4. Stimuli

The tale text is manually split into 11 parts, each one is composed of an emotionally consistent part of the story. The different parts are then presented to the subjects as a slide show. Each slide contains the text to read and an illustration in the background (Figure 2). In the normal situation (no anomaly induction), we added a background image related to the textual content to help each participant to act emotionally the content of the slide.
During the second recording, we want to provoke a reaction from the reader (a behavioural anomaly). The reaction needs to be considered as unexpected by someone who watches the record without seeing the slides. To complete this purpose we use an anomaly induction (detailed in section 2.5.).

2.5. Anomaly Induction

Anomaly induction is implemented to produce a variation in the protocol to which the subject is accustomed to. We define three types of anomaly inductions :

- Image : Changing the background with the image of a baby with the head of "Mister Bean", and, introducing some transparency to the text box to increase the visibility of the background.

- Animate : Changing the background with an animated image of the experimenter doing a grimace, and, introducing some transparency to the text box to increase the visibility of the background.

- Text : Rotating the text box multiples time to left and to the right.

Based on these three types of anomalies, we can generate different anomaly inductions, by choosing the slide on

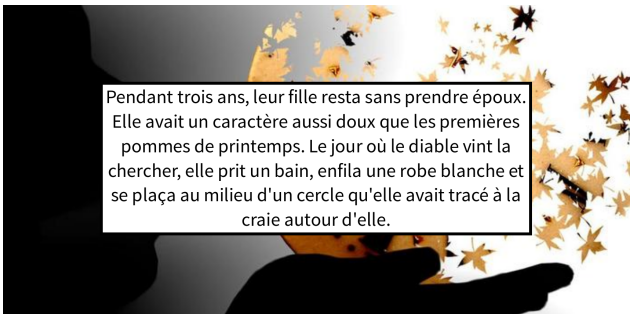

Figure 2: Example of a slide presented to the subjects. This is the sixth slide and it contains a piece of text and an illustration.

which it will be applied and delayed his effect on the slide a certain number of times. An anomaly induction is characterized by his type, the slide in which it will be triggered, and a time delay.

2.6. Hardware

The records have been performed in a recording studio which is composed of two parts: a recording booth and a supervision room. The recording booth is isolated from outside with acoustic insulation materials and a double door. The recording and supervision rooms are separated with soundproofed double glazing so as the supervisor can observe what is happening in the recording room. A green screen was placed in the background of the room from the camera point of view. The video of the face is captured with a webcam (Logitech HD Pro Webcam C920 Webcam). Speech is recorded using a high fidelity headset microphone (DPA 4066-F Omnidirectional Headset Microphone) to ensure high quality in capturing the voice.

3. Annotation

During the second record, with the purpose to induce an anomalous behavior to the participant, we use an anomaly induction described in section 2.5.. And during the third record, we asked to the participant to act an abnormal reaction during the reading of a pre-selected slide. This does not guarantee that the participant will react to the anomaly induction. During the third recording, it is also possible that the speaker either forget to act or act a too weak reaction. Otherwise the reactions can take different forms and intensity. In order to capture the presence and the diversity of anomalies in the recordings, we put an annotation tool in place to get a first human analysis by several annotators.

3.1. Annotation Tool

The annotation tool is a full web service incorporated into the team evaluation web service called PercepEval. The tool has two major screens:

- Timeline annotation. This screen contains three main elements: a video of a subject reading a slide, the slide itself, and an interactive timeline. The interactive timeline gives the possibility to create an annotation by selecting a time range in the video. The number of annotations per video is not constrained. Moreover, two annotation labels can share the same time range.
- Label annotation. To move to the next video, each annotation needs to be filled with a label. This screen contains two lists: one for choosing a label to describe the specific observation that handle the creation of the annotation; and another list, to give an intensity to this observation.

3.2. Annotation Protocol

Before accessing to the first sample, an annotator has access to a page which explains how to use the tool, the context of the record, and what to annotate. The annotator knows that the subject is reading a slide-show of a tale. But he only has access to the slide to annotate. He is asked to annotate reactions that are unexpected during the reading of the slide, by considering voice, facial expression, emotional reaction, and the realization of the task. The slide presented to the annotator does not contain any information about the way that anomaly is produced (acted or induced), and if ever it contains any anomaly.

3.3. Available Labels

The annotator can choose the level in the label hierarchy he wants, to best qualify his observation. The label hierarchy is given below.

- Audio:
  - speech:
    - prosody: rhythm, accent, intonation, intensity;
    - dysfluency: repetition, stuttering;
    - paraverbal: different type of laughing, respiration;
    - different type of silence;
    - timber;
  - others: noise, outdoor event ...
- Emotion: contains sixteen emotions which correspond to the extended list of basic emotions proposed in (Ekman, 2005).
- Video:
  - face: we use the FACS system (Ekman and Friesen, 1978) as label, but, we create a hierarchical representation by regrouping all the Action Units (AU) corresponding to a specific face part into a node (lips, eye, eyebrows, cheek, eyelid, nose, chin, neck, jaw, tongue, neck, glabella);
  - head movement;
  - eye movement;
  - physiology; blushing;
  - others: agitated subject, outdoor event ...
- Task: the reader's words are not link with the text contained in the slide, missing a sentence, sentence is not read at the right time/order

Each annotation is associated with an intensity using 4 values: slight, marked or pronounced, severe, maximum.

		Slide								
		10		9		8		7		
	Time (s)	0	3	0	3	0	3	0	3	
Method	Image	2	2	2	2	0	0	0	0	8
	Text	4	5	5	4	1	0	1	0	19
	Animate	3	1	2	4	2	0	1	1	14
	Total	9	8	9	10	3	0	2	1	41
		17		19		3		3		

Table 1: Number of anomaly inductions used by the experimenter during the building of the corpus. Each anomaly induction is identified by its type, slide and time.

4. Dataset Content

4.1. Anomaly Types

In our corpus, we have induced and acted anomalies. The first type, *induced* anomaly, corresponds to the reaction of the subject during the second recording that may happen when we triggered an event during the reading of the slide. Table 1 contains the number of different anomaly induction methods used during the creation of the corpus.
The second type, *acted* anomaly, corresponds to an abnormal behavior acted by the subject during the third recording. Before the beginning of the third recording, the experimenter asks to the subject to act an abnormal behavior during the reading of a specific slide (the slide is chosen by the experimenter). In total, the slides 7, 8, 9 and 10 have been used, respectively, 3, 3, 18 and 18 times. One subject has acted an abnormal behavior for two different slides, and has been counted twice.

4.2. Annotators

The eleven annotators are members of the EXPRESSION research team at IRISA. The team focuses on studying human language data conveyed by different media: gesture, speech and text. Three annotators have completed each more than twenty annotations, one have completed ten annotations and the others have completed less than six annotations. Each sample has been seen at least by one annotator.

4.3. Annotation of Anomalies

Table 2 presents the number of annotations for induced (by considering each type of anomaly inductor) and for acted samples. Annotations have been grouped into 6 categories: paraverbal, verbal (by splitting the audio node into this two categories), face (by using the node face available in the video node), emotion and task (by using their respective node), and a "nothing" category when we found no annotation for a sample.
First, we notice that annotators have detected nothing unusual for 14 samples, while an anomaly induction was triggered during the reading of the slide. In details, those 14 samples correspond to 5 samples with inductors of the "Image" type, 2 with the "Text" type, and 7 in the case of acted anomaly.
"Image" inductor has been used for eight participants (Table 1). For five recordings out of eight (Table 2), the annotators have noticed no abnormal reaction. This lack of subject's reaction has been noticed by the experimenter during the recording. Consequently, this inductor type has been used less than the others.

	Induced			**Acted**
	Image	**Animate**	**Text**	
Verbal	1 (12.5%) 1.0 1.0 ±0.0	4 (28%) 7 1.75 ±0.83	17 (89%) 1 1 ±0	14 (33%) 24 1.71 ±0.95
Paraverbal	1 (12.5%) 3 1.33 ±0.47	7 (50%) 12 1.72 ±1.16	16 (84%) 58 1.78 ±1.13	28 (69%) 42 1.5 ±0.78
Face	1(12.5%) 1 1 ±0	8 (57%) 15 1.875 ±1.05	16 (84%) 28 1.78 ±1.14	32 (76%) 53 1.65 ±0.87
Emotion	0 0 0	9 (64%) 13 1.44 ±0.68	14 (73%) 25 1.78 ±0.94	12 (28.57%) 17 1.41 ±0.64
Task	1 (12.5%) 3 3 ±0	1 (7%) 1 1 ±0	4 (21%) 4 1 ±0	17 (40.47%) 27 1.58 ±1.23
Nothing	5 (62.5%) 5 1	0 0 0	2 (10%) 2 1	7 (16%) 7 1

Table 2: Number of annotated cues identified by the annotator for the abnormal slide (acted and induced). The anomalous slides are grouped into columns by the type of their anomaly inductor. The rows correspond to label categories. In a cell, the first line corresponds to the number of samples which have at least one annotation and the percentage that it represents compared to all the samples that share the same type of anomaly inductor. The second line corresponds to the total number of annotations that used this category. And the third line corresponds to the mean (and std. dev.) of the numbers of annotations of this category per sample annotated in this same category (*i.e.* line2/line1).

For "Text" and "Animate" induced anomalies, the annotators have tagged abnormal phenomenon by choosing "Verbal&Paraverbal", "Face", and "Emotion" labels for more than half of the sample set. In this case, observable phenomena show that the inductive event had measurably impacted on this three channels. In the case of acted anomaly, we can see that the importance of the "Emotion" category decreases, and the importance of the "Task" category increases. That can be explained by the variety of acted anomalies and by the fact that subjects choose to act an anomaly by doing a variation in their voice or facial expressions, more than doing a variation in the emotion channel. Our interpretation is that completely acting an emotion is more challenging than just doing a variation in the voice or in the facial expression.
By analyzing the annotation, we find that, in general, annotators have done a precise description of their observations. They have chosen leaf labels, instead of choosing preferably general labels at the top of the hierarchy. And they have precisely selected the time range in which the phenomena occur.

4.4. Anomaly Example

Examples of speech signals are given on Figure 3. Speech signals correspond to the same text: "*Implorant le pardon de sa fille, il se mit à aiguiser sa hache*" (Imploring his daughter's forgiveness, he began to sharpen his axe.). The

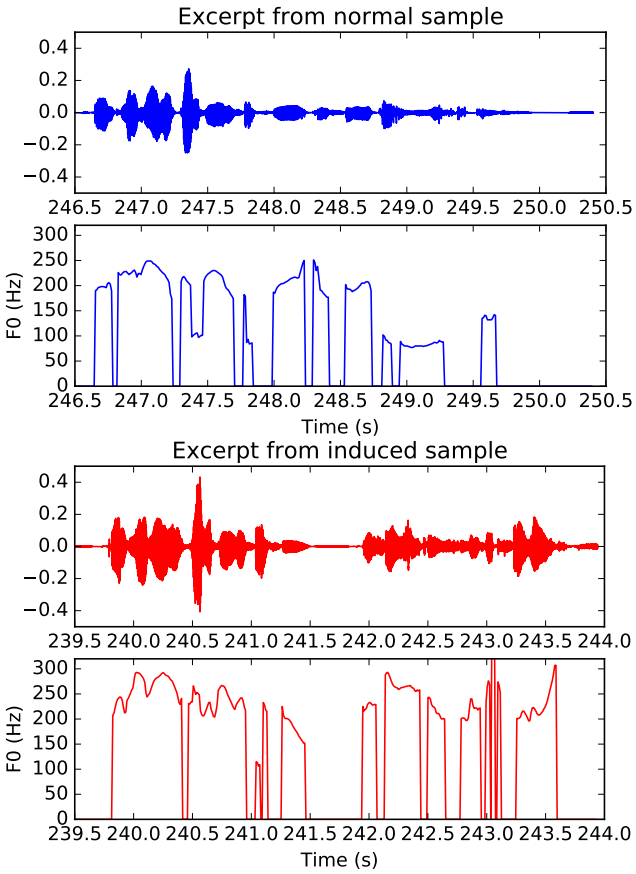

Figure 3: Speech analysis for the reading of the same sentence during the normal session (top, blue curve) and the induced session (bottom, red curve).

Figure 4: Face reaction at the beginning of the slide containing the anomaly induction for the normal session (left) and the induced session (right).

bottom part of Figure 3 contains the speech signal and the F0 contour corresponding to the induced anomaly sample. At the top of Figure 3, the speech signal and the F0 contour are from the same slide but extracted from the normal recording. The induction process on this example makes the speaker laugh while speaking. This explains the big acoustic differences between the two samples.

Figure 4 shows the face of a subject who reacts to an anomaly inductor (right) and the face of the subject for the same sequence during the normal session (left). It clearly shows a smiling expression when the subject discovers the anomaly induction.

4.5. Available Materials

The corpus is composed of three records of 41 subjects. The recording is done through a web-interface using web RTC (chrome implementation) to capture and to align the audio-video signals. The codecs used are VP9 (with a variational fps) for the video encoding and Opus codec (with an audio sampling at 48kHz) for the audio encoding.
After the recording, we check manually the quality of the records. We have used *ffmpeg* to correct mis-alignments and fixed the sampling rate of the video file at 25 fps (with a resolution up to 1080p), and the audio sampling at 44kHz. Finally, the corpus contains 1353 samples, including 83 samples *i.e.* 6% of the total population, where the experimenter either tries to induce an anomalous reaction, or asks to act an abnormal behavior. By considering only the samples that received at least one annotation, we get 68 samples *i.e.* 4.4% of the total population. The samples that have been proposed to the annotation process are either those where an anomaly inductor has been triggered or those for which the subject was supposed to act an anomalous behavior.
Audio, video features and meta-data are available from the team website[1].

5. First Experiment

A preliminary experiment has been conducted using the anomaly detection framework described in (Fayet et al., 2017). This framework is a 2-step pipeline: first an unsupervised classifier which assigns an anomaly score to each sample, and then a threshold classification is used to label each sample as normal or anomalous.
As mentioned in section 4.5., the reading of a slide corresponds to a sample. Each reading either with the induced or acted reaction is considered to belong to the anomalous class and others samples are assigned to the normal class.
Samples are represented by using only audio-based features. Those features are 24 prosodic features, based on low level features like pitch, first two formants, energy and duration of voiced/unvoiced segments with a sliding analysis window size of 40 ms and a step of 10 ms. They are extracted by using the PRAAT software (Boersma and Weenink, 2016). From these low-level features, we derive the final feature vectors that summarize their time evolution by computing the mean, maximum, minimum and entropy for each of them.
We compare here a Gaussian Mixture Model (GMM) and a OneClass-SVM (OCSVM). Hyper-parameters of the models are tuned by using the BIC score for GMM (Steele and Raftery, 2010) and an unsupervised score classifier for OCSVM (Caliński and Harabasz, 1974). For the GMM, the number of components is setup to 10 and each component has its own diagonal covariance matrix. For the OCSVM, we chose the RBF kernel and found, on the training data, ν equals to 0.3 and γ equals to 0.0001.
Results in Table 3 seem to indicate that a GMM and OCSVM approches are able to separate the normal samples from anomalous ones; with an advantage to the GMM

[1] https://www-expression.irisa.fr/results-and-resources/corpus/emoly/

	GMM	OC-SVM
Acted	0.771 $\pm$0.052	0.730 $\pm$0.069
Induced	0.677 $\pm$0.064	0.624 $\pm$0.097
Acted+Induced	0.717 $\pm$0.047	0.683 $\pm$0.056

Table 3: Mean area under ROC curve (ROC-AUC score) and standard deviations.

approach. By comparing the results for acted anomalies to induced ones, we can notice that acted anomaly are better detected through our anomaly chain than induced ones. The reason could be that reactions caused by induced anomalies are more subtle and nuanced, sometimes under control of the speaker, and then more difficult to detect than acted behaviors.

Moreover, by considering the percentage of "Face" labels in Table 2, it seems that the facial expressions, hence the video signals, contain some useful information about the anomalous reactions. So, by using features extracted from the video signal, we expect an improvement in our results.

6. Conclusion and Future Work

In this paper we have presented a new corpus usable to study anomalous behaviors during expressive interactions using both speech and facial expressions. This first version of the corpus contains various kind of anomalies and is composed of records from 41 subjects (11 females and 30 males), totalling about 8.65 hours of records. More information on the corpus is available on the team website[2].

One major objective of this work was to collect and test the reactions of subjects to anomaly induction and also check the ability to automatically detect those anomalies. First experiments seem to show that unsupervised classifiers are usable to separate anomalous samples from normal ones. We confirmed that the protocol is able to induce anomalous reactions from the subject and also that a subject with none or small guidelines is able to act one.

As a next step, it will be interesting to get more annotators and to expand the annotation to all the sample of the corpus. It could enable a comparison between the performance of the annotators and the automatic detection.

7. Acknowledgements

This research has been financially supported by the French Ministry of Defense - Direction Générale pour l'Armement and the région Bretagne (ARED) under the MAVOFA project.

8. Bibliographical References

Abdallah, A., Maarof, M. A., and Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, 68:90–113.

Abrilian, S., Devillers, L., Buisine, S., and Martin, J.-C. (2005). EmoTV1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In *Proceedings of the 11th International Conference on Human-Computer Interaction*.

Axelsson, S. (2000). Intrusion detection systems: A survey and taxonomy. Technical report, Chalmers University of Technology, Göteborg, Sweden.

Boersma, P. and Weenink, D. (2016). PRAAT: doing phonetics by computer, December.

Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15.

Chaquet, J. M., Carmona, E. J., and Fernández-Caballero, A. (2013). A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*, 117:633–659.

Clapham, R. P., van der Molen, L., van Son, R., van den Brekel, M. W., Hilgers, F. J., et al. (2012). NKI-CCRT Corpus-Speech Intelligibility Before and After Advanced Head and Neck Cancer Treated with Concomitant Chemoradiotherapy. In *Proceedings of the 8th international conference on Language Resources and Evaluation (LREC)*, volume 4, pages 3350–3355.

Douglas-Cowie, E., Campbell, N., Cowie, R., and Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech communication*, 40(1):33–60.

Ekman, P. and Friesen, W. V. (1978). *Manual for the facial action coding system*. Consulting Psychologists Press.

Ekman, P., (2005). *Basic Emotions*, pages 45–60. John Wiley & Sons, Ltd.

Fayet, C., Delhay, A., Lolive, D., and Marteau, P.-F. (2017). Big Five vs. Prosodic Features as Cues to Detect Abnormality in SSPNET-Personality Corpus. In *Proc. Interspeech 2017*, pages 3281–3285. Springer.

Giraud, T., Soury, M., Hua, J., Delaborde, A., Tahon, M., Jauregui, D. A. G., Eyharabide, V., Filaire, E., Le Scanff, C., Devillers, L., et al. (2013). Multimodal expressions of stress during a public speaking task: Collection, annotation and global analyses. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 417–422. IEEE.

Gu, X., Cui, J., and Zhu, Q. (2014). Abnormal crowd behavior detection by using the particle entropy. *Optik-International Journal for Light and Electron Optics*, 125(14):3428–3433.

Hansen, J. H., Bou-Ghazale, S. E., Sarikaya, R., and Pellom, B. (1997). Getting started with SUSAS: a speech under simulated and actual stress database. In *Eurospeech*, volume 97, pages 1743–46.

John, O. P. and Srivastava, S., (1999). *The Big Five trait taxonomy: History, measurement, and theoretical perspectives*, volume 2, pages 102–138. Guilford.

Ko, T. (2008). A survey on behavior analysis in video surveillance for homeland security applications. In *Applied Imagery Pattern Recognition Workshop, 2008. AIPR'08. 37th IEEE*, pages 1–8. IEEE.

Li, L., Gariel, M., Hansman, R. J., and Palacios, R. (2011). Anomaly detection in onboard-recorded flight data using cluster analysis. In *IEEE/AIAA 30th Digital Avionics Systems Conference*, pages 4A4–1–4A4–11. IEEE.

[2]https://www-expression.irisa.fr/results-and-resources/corpus/emoly/

Mehran, R., Oyama, A., and Shah, M. (2009). Abnormal crowd behavior detection using social force model. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 935–942. IEEE.

Park, K., Lin, Y., Metsis, V., Le, Z., and Makedon, F. (2010). Abnormal human behavioral pattern detection in assisted living environments. In *Proc. of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments*, page 9.

Plaisant, O., Courtois, R., Réveillère, C., Mendelsohn, G., and John, O. (2010). Validation par analyse factorielle du big five inventory français (bfi-fr). analyse convergente avec le neo-pi-r. In *Annales Médico-psychologiques, revue psychiatrique*, volume 168, pages 97–106. Elsevier.

Steele, R. J. and Raftery, A. E. (2010). Performance of Bayesian model selection criteria for Gaussian mixture models. *Frontiers of Statistical Decision Making and Bayesian Analysis*, 2:113–130.