UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

MÁSTER UNIVERSITARIO EN
INGENIERÍA DE TELECOMUNICACIÓN

TRABAJO FIN DE MASTER

DEVELOPMENT OF AN IMMERSIVE EMOTION-AWARE SYSTEM

JOSÉ IGNACIO MORA PÉREZ

2023

TRABAJO DE FIN DE MASTER

Título:	DESARROLLO DE UN SISTEMA INMERSIVO PARA LA DETECCIÓN DE EMOCIONES
Título (inglés):	DEVELOPMENT OF AN IMMERSIVE EMOTION-AWARE SYSTEM
Autor:	JOSÉ IGNACIO MORA PÉREZ
Tutor:	CARLOS ÁNGEL IGLESIAS FERNÁNDEZ
Ponente:	PONENTE
Departamento:	Departamento de Ingeniería de Sistemas Telemáticos

MIEMBROS DEL TRIBUNAL CALIFICADOR

Presidente:	—
Vocal:	—
Secretario:	—
Suplente:	—

FECHA DE LECTURA:

CALIFICACIÓN:

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE
INGENIEROS DE TELECOMUNICACIÓN

Departamento de Ingeniería de Sistemas Telemáticos
Grupo de Sistemas Inteligentes

TRABAJO DE FIN DE MASTER

DEVELOPMENT OF AN IMMERSIVE EMOTION-AWARE SYSTEM

SEPTIEMBRE 2023

Resumen

Actualmente, el número de sistemas que utilizan inteligencia artificial y procesamiento de gran cantidad de datos es muy elevado, a diario se desarrollan aplicaciones con nuevas funcionalidades que sirven de apoyo a personas y profesionales en fiderentes ámbitos.

Los sistemas conscientes de las emociones, que son la culminación de la incursión de la inteligencia artificial en el ámbito de las emociones humanas, representan una frontera tecnológica apasionante y transformadora. Estos sistemas están diseñados no solo para comprender las emociones humanas en tiempo real, sino también para responder a ellas, ya se expresen mediante palabras, texto escrito, expresiones faciales o señales fisiológicas.

El potencial impacto de un sistema consciente de las emociones se extiende a diversos ámbitos. En sanidad, ofrece un enfoque revolucionario de la asistencia a la salud mental, con aplicaciones en la detección precoz y la intervención. En la educación, transforma la experiencia de aprendizaje personalizando los contenidos para adaptarlos al compromiso emocional de los alumnos. En marketing, redefine el compromiso adaptando las respuestas a los sentimientos del cliente.

Por eso, en este Trabajo Fin de Master se ha desarrollado un sistema consciente de las emociones que es capaz de analizar emociones en tiempo real y que tambien tiene la capacidad de analizar videos, extrayendo sus característas emocionales desde el video, el audio y el texto. Con el propósito de visualizar los datos recogidos por este sistema, ha sido necesario implementar una aplicación web con varios dashboards en función de si se ha escogido el analisis en tiempo real o el analisis de video. La aplicación permite al usuario visualizar los datos extraidos en forma de gráficos, permiten interpretarlos de forma mucho más visual y haciendo que sean más fáciles de entender.

En resumen, el objetivo de este proyecto ha sido desarrollar un sistema consciente de las emociones a través de una aplicación web que permite el analisis en tiempo real y el analisis de videos, funcionalidades muy utiles para muchos ámbitos.

Palabras clave: Reconocimiento de emociones, Sistema consciente de las emociones, Dashboard, Internet de las cosas, Automatización, Aplicación web, Visualización, Gráficos, Python.

Abstract

Today, the number of systems that use artificial intelligence and process large amounts of data is very high, applications with new functionalities that support people and professionals in different fields are developed every day.

Emotion-aware systems, the culmination of AI's foray into the realm of human emotions, represent an exciting and transformative frontier in technology. These systems are engineered to not only comprehend but also respond to human emotions in real time, whether they are expressed through spoken words, written text, facial expressions, or physiological signals.

The potential impact of an immersive emotion-aware system extends across diverse domains. In healthcare, it offers a revolutionary approach to mental health support, with applications in early detection and intervention. In education, it transforms the learning experience, customising content to match students' emotional engagement. In marketing and customer service, it redefines engagement by tailoring responses to customer sentiments.

Therefore, in this master thesis we have developed an immersive emotion-aware system that is able to analyse emotions in real time and also has the ability to analyse videos, extracting their emotional characteristics from various sources, video, audio and text. To visualise the data collected by this system, it has been necessary to implement a web application with several dashboards depending on whether real-time or video analysis has been chosen. The application allows the user to visualise the extracted data in the form of graphs, allowing them to be interpreted in a much more visual way and making them easier to understand.

To summarise, the objective of this project has been to develop an immersive emotion-aware system through a web application that allows real-time analysis and video analysis, very useful functionalities for many fields.

Keywords: Emotion recognition, Emotion-aware system, Dashboard, Internet of things, Automation, Web application, Visualization, Charts, Python.

Agradecimientos

Quiero aprovechar este espacio para dar mi agradecimiento a las personas más importantes para mi.

A mis padres, porque gracias a vosotros, aquel niño que quería ser Ingeniero de Telecomunicación como papá ha cumplido su sueño. Sin la educación y los valores que me habéis inculcado, sin vuestro apoyo incondicional y sin vuestro sacrificio esto no hubiera sido posible. Siempre habéis sido mis referentes y siempre lo seréis.

A mi familia y amigos, por su apoyo y ánimo durante estos años y por estar siempre orgullosos de mi progreso.

A Carlos Ángel Iglesias por la gran ayuda y los consejos durante el desarrollo del proyecto y por ser un tutor excepcional.

Y en especial a mi abuela Brígida, por su cariño, por sus enseñanzas y por su afán para que estudiara y aprendiera. Si todavía estuviese entre nosotros estaría muy orgullosa de hasta dónde he llegado.

Muchas gracias a todos.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

This chapter introduces the context of the master thesis, including a brief overview of all the different parts discussed in the master thesis. It also breaks down a series of objectives to be carried out during the realization of the master thesis, as well as the motivation for the development of this master thesis. Moreover, it introduces the structure of the document with an overview of each chapter.

1.1 Context

In the modern digital age, human-computer interaction has gone beyond mere practicality. It has developed into a sophisticated understanding of human feelings and attitudes. As a result, emotion-aware systems have become a remarkable technological advancement. These systems are capable of detecting, interpreting, and reacting to human emotions in real time. The immersive emotion-aware system developed is an example of how to combine advanced technologies with the complexity of emotional intelligence. In this master thesis, we explore the importance, features, and potential effects of this web application.

Traditionally, technology has been celebrated for its ability to automate tasks, streamline processes, and improve efficiency. Yet, it often struggled to bridge the gap between machines and human emotions. Enter emotion-aware systems, a new breed of technology that seeks to understand, empathise with, and cater to human emotional states. The developed immersive emotion-aware system offers a digital platform where emotions are acknowledged. Combined with domotics or some Internet of Things (IoT) [8] devices, the system can also be used to improve user experience, decision making and well being.

One of the standout features of the web app is its real-time emotion recognition capabilities. Through a sophisticated blend of computer vision, natural language processing, and data analysis, the system can detect and interpret emotions from the incoming video. The real-time emotion recognition option can also detect whether the user is smiling or not.

Furthermore, the system can analyse a video uploaded to the web application. This functionality can analyse facial expressions in a video, understand emotional cues from audio, and gauge sentiment from text. The Web app stands as a testament to the potential of modern technology to comprehend the most intricate aspect of human communication: emotions.

The emotion-aware web app has a wide range of uses that go beyond a single industry or purpose, its potential is limitless. Healthcare providers can use it to observe their patients' emotional health during telehealth meetings, teachers can assess student participation and feelings in virtual classrooms, and customer service agents can improve user experiences by quickly reacting to emotional signals. Given these points, this system can be applied from marketing to mental health, entertainment to e-Commerce, with no limits.

In this master thesis, we try to offer an immersive emotion-aware system capable of detecting emotions from multiple sources. In addition, to complement and assist the development of this goal, machine learning technologies and some useful libraries have been employed.

1.2 Motivation

The main motivation for the development of this master thesis comes from the growing popularity of artificial intelligence, which provides support for the performance of certain tasks. The emotions that drive our interactions are often overlooked in the digital age, where much of our communication takes place through screens and devices. An emotion-aware system has the potential to revolutionise the way we connect, communicate, and empathise with each other and could have a profound impact on our decisions, relationships, and overall well-being.

Today, digital interfaces are part of many human interactions. We are now at a critical juncture where we can either develop technology that disregards our emotional states, or we can strive to create a new generation of systems that can comprehend, adjust to, and even predict our feelings. The latter option has the potential to revolutionise human-computer interaction, making it more instinctive, empathetic, and meaningful.

The development of an emotion-aware system is a demonstration of our commitment to empathy-driven progress. It is recognition of the fact that emotions are a major factor in our daily lives and that technology can be used to improve them by detecting our emotional states and reacting to them.

Finally, the driving force behind this master thesis is the potential for a positive effect. We are striving to create an emotion-aware system that can improve the mental health of individuals, make digital spaces more inclusive, and enhance the overall well-being of people in various situations. This system could allow people to express themselves more genuinely, to strengthen the bonds between communities, and to create a more emotionally intelligent society.

1.3 Project goals

The objective of this master thesis is, by using the DeepFace library, to create an application that can distinguish facial faces from the rest of the elements of an image or video, analyse the emotions of those faces, and use the audio and text extracted from the video, evaluate machine learning models to distinguish whether a person is sad, happy, or angry among others.

In this master thesis, we are going to make an application to detect emotions in videos. The following steps is followed:

- Study the DeepFace library to be able to extract and use all the functionalities of interest.

- Design and develop a system capable of detecting emotions in videos, audio, and texts.
- Design and development of a feature extraction layer to extract relevant features from raw data.
- Design and development of a visualisation layer to display the data collected by the emotion-aware system.

1.4 Structure of this document

In this section, we provide a brief overview of the chapters included in this document. The remainder of this document is structured as follows.

Chapter 2 covers the main essential concepts used to develop the solution of the master thesis. It introduces and explains the technologies, tools, and resources used in the development of the master thesis.

Chapter 3 provides a comprehensive description of an emotion-aware system that is taken into account when designing each module. Some practical applications of the system and related projects are explained in this chapter.

Chapter 4 outlines the requirement analysis, which is essential for software development. The use cases and the functional and non-functional requirements are explained in this chapter.

Chapter 5 describes the overall architecture of the web application, with the connections between the different layers involved in the development of the master thesis. The project basis and the initial design idea are also discussed in this chapter.

Chapter 6 details the different scenarios that have been run and analyses the results obtained in the system.

Chapter 7 sums up the findings and conclusions found throughout the document and provides an indication of potential future progress that could be made.

CHAPTER 2

Enabling Technologies

This chapter provides a concise overview of the technologies that enabled this master thesis to be realized.

2.1 Programming and development environment Technologies

2.1.1 Python

Python is a well-known high-level programming language that is employed in a multitude of applications, such as web development, scientific computing, data analysis, artificial intelligence, and automation.

Python is known for its simplicity, readability, and flexibility, making it easy to learn and use. It features a vast standard library and a large number of packages that can be easily installed using package managers such as pip. Python supports various programming paradigms, including procedural, object-oriented, and functional programming, and it has a large and active community that contributes to its development and maintenance.

Python's capabilities are far-reaching, extending beyond its core features. It has an extensive standard library full of modules and functions, which provides developers with previously made solutions to common problems, eliminating the need to start from scratch. Package managers such as pip make it easy to add extra features, as there is a vast array of packages created and shared by the ever-growing Python community.

Python is one of the most popular choices for data science due to a combination of factors such as the rich ecosystem of libraries that significantly accelerate the data science workflow, allowing practitioners to efficiently process and explore data, flexibility and versatility that allows data scientists to adopt the most suitable approach for different tasks and facilitates the creation of modular and reusable code, promoting efficient collaboration and experimentation.

2.1.2 Streamlit

Streamlit [9] is an open-source Python library designed to create interactive web applications for data science and machine learning projects. It simplifies the process of turning data scripts into shareable web applications by providing an intuitive interface and powerful features.

Streamlit makes it simple for developers, particularly data scientists and machine learning experts, to quickly and easily convert their data analysis scripts into interactive web applications. By using only a few lines of Python code, you can create interactive dashboards, visualizations, and user interfaces that allow users to interact with data and models without requiring extensive web development knowledge.

One of the main advantages of Streamlit is its ease of use. It provides a simple API that enables users to create intuitive and interactive applications with just a few lines of code. This makes it an ideal tool for small data applications such as the one performed for this

Figure 2.1: Streamlit

work.

Streamlit offers a selection of pre-made components, like graphs and widgets, that can be tailored to your requirements. This makes it simple to add features to your application without having to create intricate code from the beginning.

Due to the advantages of the platform discussed above, all the pre-made components that can be found on it, the positive outcomes, and its user-friendly interface, Streamlit is the optimal choice for the implementation of the application.

2.2 Emotion analysis technologies

2.2.1 DeepFace

DeepFace [1] is a Python library that provides a convenient interface for facial recognition and analysis tasks using deep learning models. It allows developers to perform various operations related to facial attributes, such as facial recognition, emotion detection, age estimation, gender classification, and facial feature extraction. DeepFace aims to simplify the process of working with complex deep learning models for facial analysis by providing pre-trained models and easy-to-use functions.

As shown in the figure above, the library uses pre-trained state-of-the-art models in the background. State-of-the-art models are those models that currently achieve the best possible results for a particular task on a collection of standard datasets used for comparison.

DeepFace library offers a range of deep learning face recognition algorithms that can be used. These include convolutional neural networks, deep belief networks, and restricted Boltzmann machines. Each of these algorithms has its own advantages and disadvantages, and the best one to use depends on the specific application.

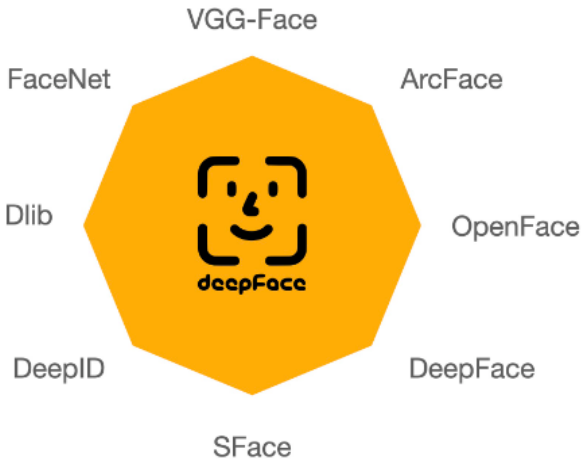

Figure 2.2: DeepFace models [1]

To better understand the different models [10], a description of each one is provided below.

- VGG-Face: The Visual Geometry Group (VGG) is a research group at the University of Oxford that developed a deep convolutional neural network (VGGNet) for image recognition. This model was highly successful in the ImageNet challenge and has since become one of the most widely used models for this purpose. The VGG face recognition model achieves a precision of 97. 78% in the popular Labeled Faces in the Wild (LFW) dataset.

- Google FaceNet: This model was developed by Google researchers. It is based on a convolutional neural network that is trained on a large dataset of faces. The model has the ability to learn the features of a face and can be used to identify a person from a single image or video. FaceNet is a cutting-edge deep learning model for face detection and recognition. It is used for face recognition, authentication, and grouping. It is reported to achieve 99.63% accuracy in the LFW dataset.

- OpenFace: Researchers at Carnegie Mellon University have developed a face recognition model known as OpenFace. This model is heavily based on the FaceNet project, but is lighter and has a more flexible license type. OpenFace achieves a precision of 93. 80% on the LFW dataset.

- Facebook DeepFace: Researchers at Facebook developed a face recognition model called DeepFace. This algorithm was trained on a dataset of four million labelled faces from more than 4,000 people, which was the largest facial dataset available

when it was released. The model is based on a nine-layer deep neural network. The dataset LFW was used to test the accuracy of the Facebook model, which achieved a score of 97 35%.

- DeepID: The DeepID face verification algorithm, developed by researchers at the Chinese University of Hong Kong, is a deep learning-based face recognition model. It was one of the first to use convolutional neural networks and was able to outperform human performance in face recognition tasks. DeepID-based face recognition systems were among the first to surpass human performance in the task. DeepID2 achieved 99.15% on the LFW dataset.

- ArcFace: This model was developed by researchers from Imperial College London and InsightFace together. The ArcFace model has achieved a 99. 4% precision rate when tested on the labeled faces in the wild data set.

- SFace: Recently, the DeepFace library has been augmented with a model that introduces a novel concept with the goal of reducing overfitting issues in imperfect training databases by optimizing the intra- and inter-class distances explained in [11]. This model, known as SFace, has been shown to improve model performance in robust unconfined face recognition, achieving an accuracy of 99. 60% in the LFW dataset.

- Dlib: This last model contains machine learning algorithms and tools for various computer vision and machine learning tasks. It is a versatile library that offers a wide range of functionality, particularly in the areas of facial recognition, object detection and image processing. The Dlib model was not designed by Davis E. King, the main developer of the Dlib image processing library. This algorithm achieved remarkable 99. 38% accuracy on the LFW dataset, making it one of the best-performing models in the area of face recognition.

These models are so impressive that they have demonstrated the ability to analyze images and videos of faces with a level of accuracy that exceeds what human beings can do.

The figure above displays the face detectors included in DeepFace. Each one is discussed in order to determine which is most suitable for the task at hand, elucidating its advantages and disadvantages.

- OpenCV[12]: This detector offers a pre-trained face recognition algorithm that is commonly used in computer vision and image processing applications. OpenCV's face detector is based on the Haar Cascade Classifier and is able to identify faces in images and video streams. One of the disadvantages of this detector is that it needs frontal images to work properly. Otherwise, it will not detect faces correctly.

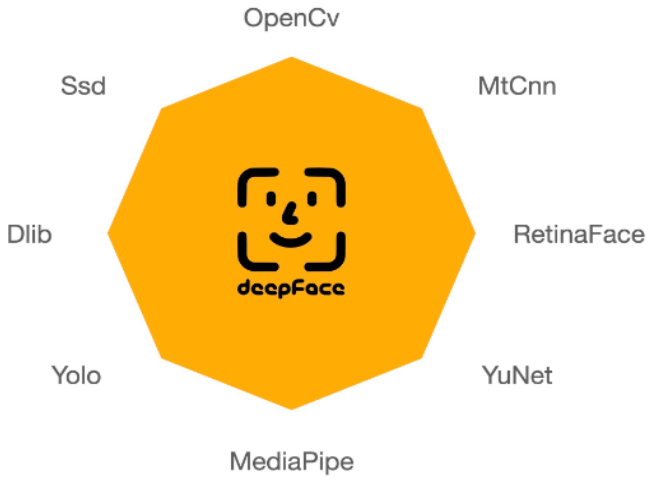

Figure 2.3: DeepFace face detectors [1]

- Dlib[13]: This library, which is one of the most utilized packages for face detection, utilizes a hog algorithm in the background, similar to OpenCV, which is not based on deep learning. Despite this, it has achieved relatively high scores in terms of detection and alignment.

- SSD[14]: SSD, or Single-Shot Detector, is a widely used deep learning-based detector. Its performance is comparable to that of OpenCV, but it does not offer facial landmarks and relies on OpenCV's eye detection module for alignment. Although its detection capabilities are impressive, its alignment score is only average.

- MTCNN[15]: It is a deep learning-based approach that is used to detect faces in images. It is made up of three stages: a Proposal Network, a Refinement Network, and an Output Network. The Proposal Network is used to generate candidate windows, the Refinement Network is used to refine candidate windows, and the Output Network is used to classify windows as face or non-face. MTCNN is a powerful face detector that provides excellent detection results. However, it is slower than OpenCV, SSD, and Dlib.

- RetinaFace[16]: It is recognized as the most advanced deep learning-based face detection model. Its performance under real-world conditions is difficult to achieve. Unfortunately, it requires a lot of computing power. This is why RetinaFace is the slowest face detector compared to the other options.

- Yolo[17]: YOLO and its variants may not be as accurate as two-stage detectors, but they are much faster. YOLO works well with objects of normal size, but struggles to

detect smaller ones. In addition, its accuracy drops significantly when dealing with objects that have large-scale variations, such as faces.

- YuNet[18]: This software was created specifically for real-time applications, providing millisecond-level speed on CPUs and is suitable for mobile and embedded devices. Its design is based on the principles of efficient small models, offering a good balance between accuracy and computational cost.

- MediaPipe[19]: Google developers have created an ultra-fast face detection system that includes six landmarks and the ability to detect multiple faces. This system is based on BlazeFace[20], a lightweight and highly efficient face detector designed for mobile GPU inference. Its remarkable speed makes it suitable for any live viewfinder experience that requires an accurate facial region of interest as input for other task-specific models, such as 3D facial keypoint estimation, facial features or expression classification, and facial region segmentation.

Also, it is important to mention the set of features that DeepFace offers.

- Face verification: The process of verifying faces involves comparing one face to another to determine if they are the same. This is often used to check if a person's physical face matches the one on their ID.

- Face recognition: The objective of this task is to locate a face in an image database. To accomplish this, face verification must be executed multiple times. Deepface has a built-in search capability to perform its task. It will search for the identity of the input image in the database path and will return a list of pandas data frames as output.

- Facial attribute analysis: It is a challenging task due to the complexity of facial features. The task of analyzing facial characteristics involves describing the visual characteristics of faces. This is done to extract attributes such as age, gender, emotion, and race/ethnicity.

- Real-Time face analysis: This feature tests facial recognition and facial attribute analysis with the real-time video feed from your webcam. The problem with this feature is that it makes a five-second countdown, then takes a picture, and, finally, analyzes it. This means that the user will have a result every five seconds but will not have a real-time analysis.

After having examined all the features, the one that fits more in the application and has the optimal potential for the purpose of the application is the facial attribute analysis.

Deepface has a powerful facial attribute analysis component, which can predict age, gender, facial expression (including anger, fear, neutrality, sadness, disgust, happiness, and surprise), gender, age, and race (including Indian, Asian, Latino Hispanic, black, Middle Eastern, and white).

2.2.2 Librosa

Librosa is a Python library specifically designed for audio and music analysis and holds significant value within emotion-aware systems that involve audio data. It provides a rich set of functionalities for extracting, processing, and analyzing audio features.

Librosa has different spectrum features to analyze audio. Mel-Frequency Cepstral Coefficients (MFCCs) are a set of coefficients used to represent the short-term power spectrum of an audio signal. These coefficients are widely used in speech and audio processing for various tasks, including speech recognition, speaker identification, and emotion analysis in emotion-aware systems.

MFCCs are considered powerful feature extraction tools for audio analysis tasks, including emotion recognition. They capture important characteristics of audio signals, such as spectral shape, timbral texture, and rhythmic patterns. In emotion-aware systems, these characteristics help identify emotional cues in speech or acoustic signals.

In conclusion, Librosa's MFCC feature extraction is a beneficial asset for emotion-aware systems. These coefficients offer a concise representation of audio data, capture essential emotional signals, and are resistant to noise and environmental changes. By including Librosa's MFCC extraction in the system's pipeline, it is possible to improve the system's capacity to recognize and respond to emotions expressed through speech and sound signals.

2.3 Data modeling technologies

2.3.1 Numpy

NumPy [21] is a Python library for numerical computing. It introduces the concept of ndarray, an n-dimensional array, which is a highly efficient contiguous block of memory that allows the storage and manipulation of large datasets. This core data structure enables developers to perform various mathematical and logical operations on arrays, making it an essential tool for data processing tasks in an emotion-aware system.

In emotion-aware systems, data is often presented in multiple dimensions, such as video frames, spectrograms, or multichannel audio. NumPy is able to manage these intricate data structures with ease. Provides a simple way to manipulate multidimensional arrays, allowing operations like slicing, filtering, and reshaping. This is essential to extract important

features, such as facial landmarks in videos or spectrogram data in audio analysis, which are necessary for emotion recognition.

In this master thesis, Numpy is used to perform operations on the arrays containing emotional information. Also, this library is used to save and load csv files whose content is the arrays with emotion and gender information.

In addition, NumPy also integrates well with other libraries commonly used in data processing workflows, such as Pandas, for data manipulation.

2.3.2 Pandas

Pandas [22] is another essential Python library that complements NumPy when it comes to data processing and analysis in emotion-aware systems. It is particularly valuable for handling structured data, including tables and time series.

Pandas is well-known for its capacity in data handling and manipulation. It brings in two main data structures, DataFrames and Series, that make it possible for developers to work with structured data without difficulty. In emotion-aware systems, structured data often includes metadata, annotations, or aggregated results, all of which can be efficiently managed and analyzed using Pandas.

Clean and well-structured data are essential in emotion analysis. Pandas provides a variety of functions for data cleaning, transformation, and preprocessing. It can be used to take care of missing values, convert data types, and combine datasets, all of which are essential for guaranteeing the accuracy and reliability of data used for emotion recognition.

Exploratory data analysis is a fundamental step in emotion analysis systems. Pandas simplifies this process by providing intuitive functions for data exploration. It allows developers to calculate summary statistics, create pivot tables, and group data by different attributes. In this master thesis, Pandas is used to generate a DataFrame of emotions and their scores. This allows for the representation of a bar chart with the emotions in the analyzed video.

2.3.3 Google Speech Recognition API

The Google Speech Recognition API [23], also known as the Google Cloud Speech-to-Text API, is a powerful cloud-based service offered by Google Cloud Platform [24]. It enables developers to convert spoken language into text, making it a valuable tool for a wide range of applications, including transcription services. It uses machine learning models and advanced speech recognition technology to accurately translate audio data into text. The primary function of the API is to recognize spoken language in various languages and dialects, with more than 125 languages and variants. It can convert audio from different sources, such as

audio files, streaming data, or real-time audio input from microphones.

Google's Speech Recognition API is renowned for its precision in recognizing speech. It also provides customization options, enabling developers to modify the system to recognize particular terminology or words pertinent to their program.

2.4 Data visualization technologies

2.4.1 Apache echarts

Apache ECharts [25] is an open-source JavaScript visualization tool. It is a free and powerful charting and visualization library that offers easy ways to add intuitive, interactive, and highly customizable charts.

ECharts stands out for its ability to generate interactive and dynamic visualizations. It enables users to build charts that respond to their interactions, such as hovering over data points, clicking to filter data, or zooming in for more detail. This interactivity is especially useful for emotion-aware systems, as it allows users to explore emotional data in a more captivating and informative manner.

Animations can be a great help in understanding data trends and patterns. ECharts offers a wide range of animation effects that can be used for different chart elements. Systems that are sensitive to emotions can take advantage of animations to show changes in emotional states over time or to emphasize particular data points related to emotions.

In this master thesis, ECharts is used to display a line chart that shows the emotions detected over time frame by frame. It is possible to highlight one of the emotions to see how its percentage varies from the beginning to the end of the video. Also, a toolbox is displayed with some features, such as a zoom widget or a magic widget that converts the line chart into a bar chart.

2.4.2 Plotly

Plotly [26] is a Python library that offers a comprehensive suite of tools to create interactive and informative charts and dashboards from data. One of the main difficulties in emotion-aware systems is how to effectively communicate emotional insights to users and stakeholders. Plotly provides developers with the capability to transform raw emotional data into engaging and visually attractive representations. It can be used to create charts and graphs that clearly show emotional patterns, such as facial expressions, sentiment scores, or emotional trends over time.

Plotly was a useful tool for this master thesis to generate pie charts of emotions, gender and detection of smiles. Users can interact with visualizations and charts to gain a

deeper understanding of emotional patterns. This interactivity promotes more engaging and informative experiences for users.

CHAPTER 3

Emotion-aware systems

This chapter provides a comprehensive description of an emotion-aware system that is taken into account when designing each module. Also, some practical applications of the system are explained in detail, so the usefulness of the application is highlighted. To conclude this chapter, a review of some emotion-aware systems is made.

3.1 Introduction

In the rapidly evolving landscape of technology and Artificial Intelligence (AI), a new frontier has emerged that promises to redefine the way humans interact with digital interfaces, emotion-aware systems. These innovative systems represent a profound intersection of artificial intelligence, psychology, and human emotions, with the aim of creating more empathetic, intuitive, and personalized interactions between humans and machines.

As technology has progressed, the aim of making machines more like humans has become a major focus. Machines are great at dealing with data and completing tasks, but they have usually been unable to understand and react to the subtle complexities of human emotions. Now, emotion-aware systems have been developed, which give machines the capacity not only to identify and interpret human emotions but also to adjust their behavior accordingly. This shift in focus marks a revolutionary period in which technology is able to tune into our emotional states, resulting in a stronger bond between humans and machines.

Emotion-aware systems operate on the premise that human emotions play a major role in shaping our perceptions, decisions, and interactions. By incorporating emotional intelligence into technological interfaces, these systems enable a new dimension of interaction that goes beyond simple functionality. They can gauge whether a user is happy, frustrated, engaged, or disinterested and respond accordingly. This synthesis of technology and emotion has the potential to revolutionize countless sectors, from education and healthcare to entertainment and marketing, amplifying the impact of technology on our daily lives.

The functionality of emotion-aware systems is based on advanced technologies such as Natural Language Processing (NLP), affective computing, machine learning, and neural networks. These technologies facilitate the extraction of emotional signs from various modalities, including text, speech, images, videos, and physiological signals. By deciphering linguistic range, vocal tone, facial expressions, and physiological responses, these systems can infer emotional states with remarkable precision.

The applications of emotion-aware systems are wide-ranging and diverse. In healthcare, they can enhance patient engagement and well-being by adapting interventions based on emotional cues. In education, they can create more personalized learning experiences that accommodate individual student emotions and cognitive states. In marketing and customer service, they can tailor interactions to customer sentiments, promoting stronger brand loyalty. These and some other practical applications are explained in more detail in Sect. 3.3.

As the realm of emotion-aware systems is explored, both chances and difficulties come across. The ethical implications of capturing and analyzing individual emotions raise concerns about user confidentiality and data safety. Moreover, the technical intricacies of precisely interpreting emotions across cultural and linguistic backgrounds require meticulous

consideration.

One of the most recognized and influential frameworks for understanding and categorizing human emotions based on universal facial expression is the Ekman Emotion Model [27], developed by psychologist Paul Ekman. This model provides a structured way to classify emotions and their corresponding facial expressions, laying the foundation for cross-cultural and cross-linguistic studies of emotions.

Paul Ekman's research was instrumental in challenging the prevailing belief that emotional expressions were culturally determined. Through extensive cross-cultural studies, he discovered that certain facial expressions are universally associated with specific emotions, regardless of cultural background. Ekman identified six basic emotions that he considered universal across all human cultures: happiness, sadness, anger, fear, surprise, and disgust. Later, Ekman expanded his model to include more nuanced emotions, resulting in a broader range of emotions that humans experience and display.

3.2 Emotion recognition tools

Recent advances in AI have enabled the analysis of emotions to be carried out without human intervention, using emotion recognition devices that do not require specialized equipment, making them more accessible. Most of the emotion recognition tools use common devices such as webcams or microphones. Algorithms for predicting emotions based on facial expressions are mature and considered accurate. There are several techniques for facial emotion recognition that depend mainly on the way of extracting feature data. Two of the technologies that have become most relevant for this analysis are the features based on appearance and the features based on geometry [28]. Both techniques involve extracting certain features from the images, which are then fed into a classification system. The main difference between them lies in the characteristics extracted from the video, images, and the classification algorithm used [29]. Geometric-based techniques identify specific features, such as the corners of the mouth or eyebrows, and extract emotional data from them. On the other hand, appearance-based extraction techniques describe the texture of the face caused by expressions and extract emotional data from changes in gestures [30]. Emotion recognition from speech analysis is an area that is gaining traction nowadays [31]. Additionally, once the audio has been extracted and analyzed, it can be transcribed and the resulting text can also be analyzed [32].

Many studies have used variations of Russel's circumplex model of emotions [2], which provides a way to map out basic emotions in a two-dimensional space based on valence and arousal. This approach makes it possible to identify a desired emotion and measure its intensity simply by looking at two dimensions.

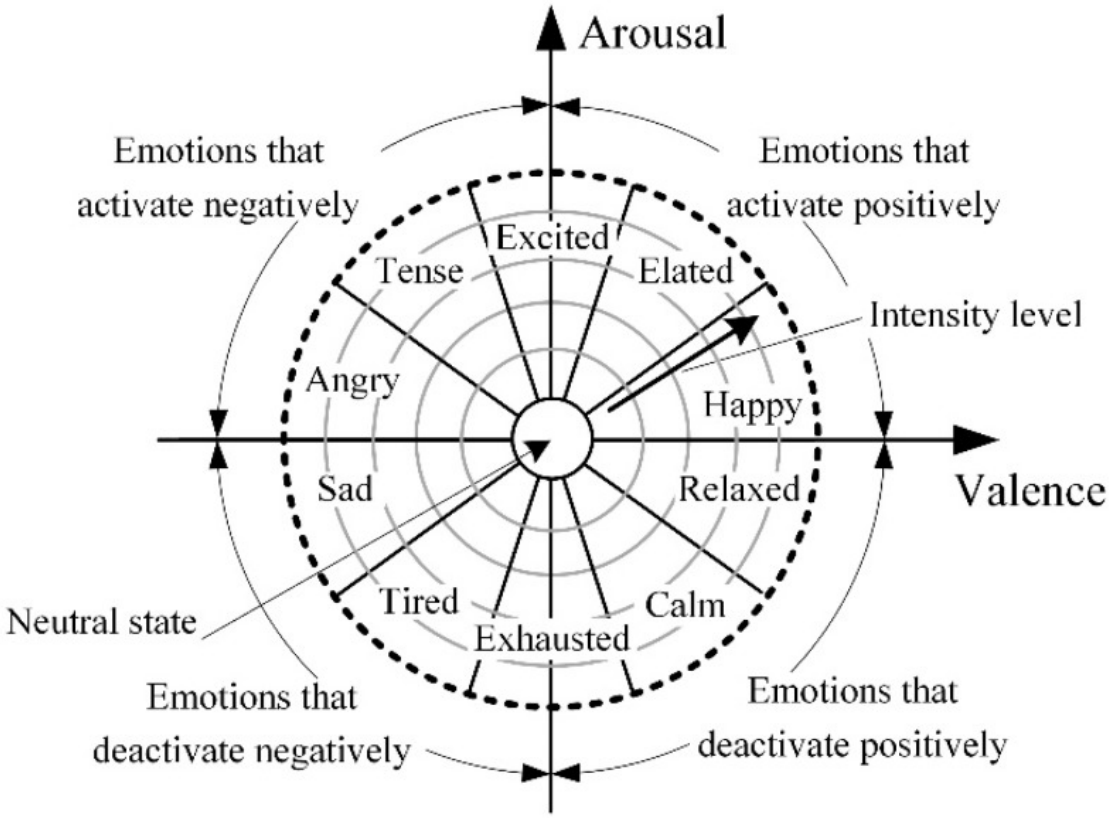

Figure 3.1: Russel's circumplex model of emotions [2]

The model described above makes the classification and evaluation of emotions clear. However, there are still many issues related to emotion assessment, especially the selection of measurement and results evaluation methods. This model provides a useful framework for understanding the relationships between different emotions and how they vary in terms of their emotional qualities. It is important to note that while the model is a valuable tool for conceptualizing emotions, it simplifies the complexity of human emotions and may not fully capture the intricacies of individual experiences.

Russell's circumplex model has been influential in psychology and emotion research, offering insights into emotional patterns and helping researchers and practitioners better understand the diverse range of human emotions. The hypothesis is that emotions of opposite valence, such as happiness and anger, may lead to similar reactions. In contrast, emotions of similar valence, such as anger and sadness, may result in different decisions and reactions.

3.3 Practical applications of emotion-aware systems

3.3.1 Healthcare and well-being

AI has emerged as a transforming force in healthcare and well-being, promising to revolutionize the way professionals diagnose, treat, and manage medical conditions. AI has made an important impact in the field of early disease detection and diagnosis. Its algorithms are capable of analyzing complex medical images with remarkable accuracy, allowing the identification of conditions such as cancer, cardiovascular diseases, and neurological disorders. This precision allows medical professionals to start treatment plans earlier, leading to better patient outcomes and higher survival rates.

In addition, AI empowers the creation of customized treatment plans tailored to individual patients. By analyzing extensive datasets that encompass patient histories, genetics, and medical profiles, AI can recommend treatments that account for unique factors. This advancement not only enhances the effectiveness of interventions, but also minimizes potential side effects, offering a new level of care that was previously unattainable.

Two examples of emotion-aware systems in the medical field are mental health monitoring and counseling and therapy.

- Mental Health Monitoring: Emotion-aware systems can analyze facial expressions and speech to monitor emotional well-being, and identifying signs of stress, anxiety, or depression in patients can prompt timely interventions and support. The use of AI and machine learning algorithms has facilitated the provision of personalized care

that is customized to the needs of a particular patient, with an emphasis on emotional support [33].

- Counseling and Therapy: Virtual therapists or counseling systems can evaluate users' emotions during sessions and provide real-time feedback or interventions based on detected emotional states. An example of a mental health tool designed for counselling is Woebot[34], which is an automated conversational agent designed to provide cognitive behavioral therapy in the format of brief, daily conversations and mood tracking.

3.3.2 Marketing

AI is a key tool in marketing, enabling companies to better understand their customers and develop personalized and automated marketing strategies. The evolution and momentum that AI will bring about in creative, analytical, and technological capabilities will lead to the deployment of marketing that will improve the customer experience and business results.

As well, AI enhances the accuracy and effectiveness of targeted advertising. By analyzing user data and behavior patterns, it is possible to predict which ads are most likely to resonate with specific segments of the audience. This personalization not only increases the likelihood of engagement, but also improves the overall user experience by delivering content that aligns with individual interests.

In the area of content optimization, AI can make the process of creating and delivering relevant content to people more efficient. NLP algorithms are used to analyze language patterns and sentiment, allowing marketers to create messages that emotionally connect with their target audience. Customer engagement and advertising effectiveness are two examples of emotional-aware system use cases.

- Customer Engagement: In the retail sector, emotion-aware systems can be used to assess customers' facial expressions and behaviours to determine their reactions to products or store designs. These data can be used to help retailers improve customer experience. For example, when watching a horror movie or show, the system can detect if the customer is laughing or does not show any fear or surprise emotions. The movie can become more terrifying, or the environment can be altered depending on the AI capabilities available to the user.

- Advertising effectiveness: Marketers can assess emotional responses to advertisements, assisting in the customization of campaigns to evoke the desired emotions and establish a connection with the intended audiences. For example, if a company makes a funny advertisement and wants to prove that it will evoke the right emotion in the audience, they will only need a few volunteers to watch the advertisement while their facial

expressions are analyzed by the application. If the results are as expected, then the ad has a good chance of being successful.

3.3.3 E-Learning

Today, e-learning [35] is an important factor in the educational landscape due to its advantages and potential applications in various academic settings, such as distance learning, self-learning, and traditional face-to-face learning. This has motivated numerous companies and universities to conduct research and development in order to improve different learning methods and improve academic outcomes.

The increasing popularity of e-learning has led to an increase in the number of smart devices in our homes and offices, called IoT [8]. These gadgets are designed to make life easier for people. However, with the wide range of available smart devices, it is essential to use a system that can link them together to maximize their potential. To facilitate the integration of these devices with Web services, task-automation platforms have been developed.

- Student Engagement: In online learning, emotion-aware systems can monitor students' facial expressions and attention levels to gauge their engagement and adapt the content accordingly. An example of an emotion-aware system used for student commitment is [36]. This thesis consists of the implementation of a system capable of detecting the mood of students in the course and during the performance of different activities.

- Feedback and Support: Systems can analyze learners' emotions during assessments or quizzes to provide personalized feedback and support. The thesis cited above can also be an example of this use case because if the system detects lack of attention from the student, then it will send a message to his phone.

3.3.4 Human-Computer Interaction

The synergy between AI and Human-Computer Interaction (HCI) [37] has revolutionized the way people interact with digital systems. AI's capacity to generate personalized user experiences is particularly noteworthy. By analyzing large amounts of user data, AI algorithms can anticipate user preferences and behaviors, customizing interfaces and content to suit individual requirements. This personalization not only increases user satisfaction, but also encourages more meaningful engagement by providing pertinent information and interactions.

The role of AI in HCI extends beyond convenience; AI is revolutionizing the way we interact with technology. It provides personalized experiences, allowing natural communication and promoting inclusion, thus transforming the human-machine relationship. AI is making technology an intuitive and effortless extension of human abilities. As AI advances, its potential to revolutionize HCI is a powerful factor that is influencing the future of technology interaction.

Two of the main characteristics of AI applied to HCI using emotion-aware systems are user experience enhancement and accessibility, both are briefly explained in the following.

- User Experience Enhancement: Emotion-aware systems in HCI can adapt interfaces based on user emotional states, ensuring that content presentation and interactions align with user feelings.

- Assistive Robots for Older Adults: Socially Assistive Robotics (SAR) is a subfield in robotics focused on developing intelligent robots that can offer assistance and support to users [38]. Elderly people living in retirement homes often experience loneliness and seclusion. Socialization and mental stimulation are essential to improve their well-being [39], [40]. SAR has demonstrated to help address this problem by providing companionship to the elderly through conversation and socialization. An example is the pilot study explained in [41].

3.3.5 Automotive

AI is revolutionizing the automotive sector, bringing about a new age of safety, efficiency, and user experience. One of the most important contributions of AI is in the area of autonomous driving. AI-driven systems, such as Advanced Driver-Assistance Systems (ADAS) [42] and self-driving cars [43], are expected to revolutionize transportation by reducing the number of accidents, reducing traffic jams, and providing more convenience.

Integration of AI into the automotive sector creates a safer, smarter, and more efficient future. From self-driving capabilities to predictive maintenance and personalized experiences, AI is propelling the industry toward innovation that aligns with the demands of modern mobility. The development of AI is set to have a major influence on the automotive industry, leading to a new era of transportation.

The use of emotion-aware systems in transportation can be a solution to ensure that the driver is fully focused on driving, making the driver feel comfortable inside the vehicle, and reducing the number of incidents. Driver monitoring and passenger experience are two examples of emotional-aware system use cases.

- Driver Monitoring: An emotion-aware system in automotive driver monitoring lever-

ages AI-driven analysis of facial expressions and behavioral cues to determine the driver's emotional state. The system employs computer vision and machine learning techniques to recognize subtle changes in the driver's facial expressions and body language, capturing signs of distraction, fatigue, stress, or other emotional states. This information is crucial for improving road safety, as emotional states can affect a driver's attention, reaction time, and overall driving behavior. The data obtained helps improve safety by enabling the vehicle to respond to the driver's emotions, ensuring a more attentive and safer driving experience.

- Passenger Experience: Emotion-aware systems can customize car experiences based on passenger emotions, adjusting lighting, entertainment, and climate settings. Using computer vision and machine learning, the system is able to detect slight changes in driver facial expressions and behaviors, recording his/her emotional reactions to different stimuli in the vehicle environment. This data is essential to create a more enjoyable and personalized passenger experience.

3.3.6 Gaming and Entertainment

The history of emotion-aware systems in video games reflects the industry's relentless pursuit of creating immersive experiences that resonate deeply with the emotions of players. In the middle of the decade of 2010, video games began to utilize webcams or specialized gadgets to monitor the facial expressions of gamers. This technology allowed games to recognize smiles, frowns, and other expressions, which then had an effect on in-game events.

Today, AI-driven emotion-aware systems are at the forefront of this development. Machine learning algorithms are used to assess the facial expressions, body language, and physiological signals of players to identify their emotional state. The information obtained is then used to modify the game, character interactions, and storylines in real-time. Adaptive difficulty levels, personalized storylines, and responsive soundscapes all contribute to a more immersive and engaging experience and are explained below.

- Gaming Experience: In video games, emotion-aware systems can analyze the facial expressions, gestures, and physiological signals of the player to gauge their emotional state. This information can be used to adjust the game's difficulty level, pacing, and challenges in real-time. For example, if a player is showing signs of frustration, the game could offer assistance or provide alternative routes to progress.

- Interactive Storytelling: Video games can adapt their storylines and character interactions based on the emotional reactions of the players. Emotion-aware systems

track emotional engagement and tailor dialogue, character behavior, and plot twists to resonate with the feelings of the players, enhancing the sense of immersion.

Examples of games like "Hellblade: Senua's Sacrifice" [44] highlights the ability of AI to portray mental health struggles through the experiences of the protagonist. The game responds to players' emotional cues, offering a deeply empathetic and immersive journey.

3.3.7 Security

The use of emotion-aware systems has revolutionized public safety and security, providing a new level of situational awareness and proactive threat prevention. Using AI, machine learning, and behavioral analysis, these systems are able to detect and interpret emotional signals in real time, leading to improved emergency response, crime prevention, and general public well-being.

- Surveillance and Crowd Management: Emotion-aware systems contribute to improve situational awareness by examining the collective emotional state of a group. By monitoring facial expressions, body language, and other indicators, security personnel can better understand the emotions of those in the crowd. This information helps identify potential tensions, distress, or agitations that could escalate into disruptive situations.
- Security Checks: Airports and public spaces could use emotion-aware systems to assess travelers' emotional states during security checks. Although it is important to avoid bias and respect individual rights, analyzing emotional responses within a larger context can help identify potentially high-risk individuals.

3.4 Related projects

In the ever-changing world of research and innovation, a key element of progressing knowledge is building on the discoveries and successes of those who have gone before. This section looks into related projects, creating a story that links our current master thesis to the previous work that has been done. This section serves as a bridge between the past and the present, providing a comprehensive overview of projects that have contributed to emotion-aware systems. Exploring the methodologies, findings, and implications of prior research in order to build a strong base for this master thesis contribution. In addition, analysis of similarities and differences between ideas provides a clear perspective.

Some of these studies share the use of a dashboard to display emotions experienced during the learning process and have been the inspiration for the development of the emotion-aware system dashboard.

The most relevant research on the subject is presented below.

3.4.1 Emotion-Aware Cyber-Physical Systems

This first project is a Ph.D. thesis [3] developed at the Universidad Politécnica de Madrid that aims to find the optimal management of complex infrastructures focusing on large data centres with emotional awareness. This interdisciplinary work investigates the integration of emotional understanding into systems that bridge the digital and physical worlds. The thesis centres around the notion that recognizing and responding to human emotions can significantly enhance the capabilities and effectiveness of cyber-physical systems.

This Ph.D. thesis presents an insightful exploration of the integration of emotions into technology-driven environments. It underscores the significance of recognizing emotions in cyber-physical systems, discusses the technical underpinnings of emotion recognition, and highlights the diverse applications and implications of this integration.

Moreover, a prototype for supervision was developed. This prototype includes the information the system needs to conclude the predominant emotion of the user by monitoring its main variables. The data model built for the project is designed to make decisions about important issues, such as scheduling or prioritizing. After analyzing the correlation between emotions and productivity, the author claims that the use of a management tool to control users' emotions is clear.

Figure 3.2: Emotion-Aware Cyber-Physical Systems [3]

As shown in Figure 3.2, the prototype presents a dashboard where emotions and users

are represented, and the health of the system is obtained from the information received from the IoT devices distributed in the infrastructure.

In relation to Russell's circumplex model of emotions presented earlier, this Ph.D. analyzed the impact on the productivity of both valence and arousal. On the one hand, the impact of valence on productivity shows a clear trend, as can be seen in Figure 3.3, the higher the valence, the greater the productivity of the user.

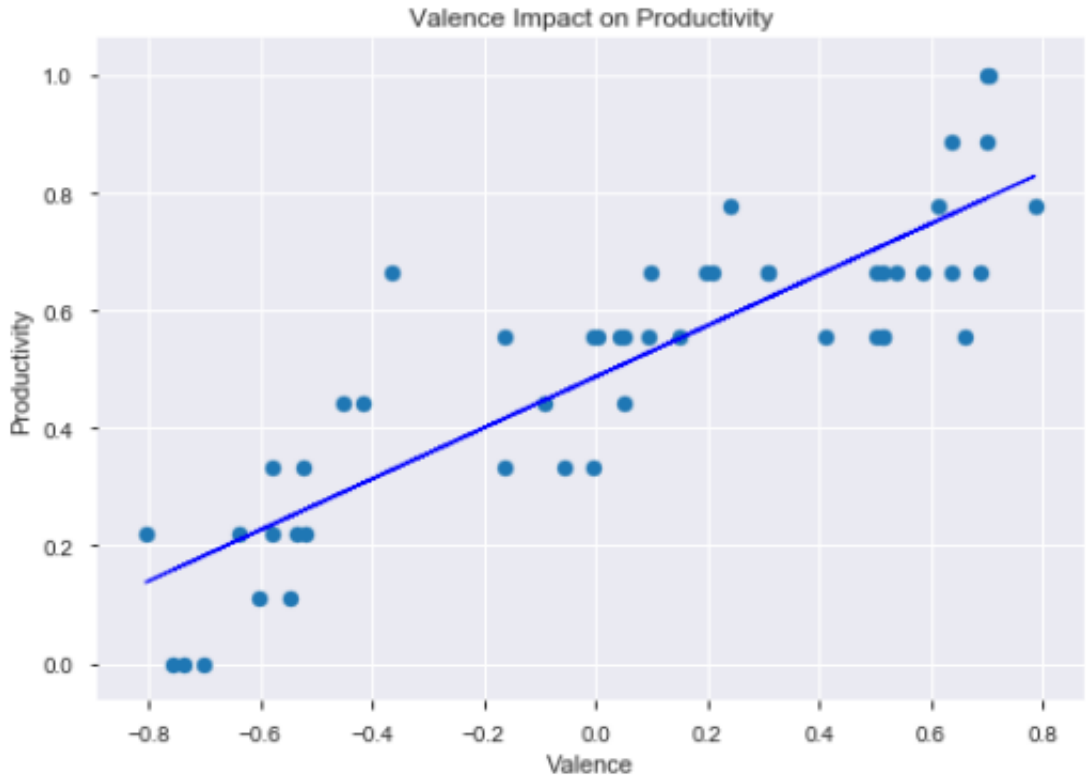

Figure 3.3: Valence Impact on Productivity [2]

However, arousal does not have a clear tendency. Still, when analyzing Fig. 3.4, two different trend lines are represented, showing that the cutoff point occurs when arousal has a value of approximately 0,417. Taking into account human behavior, a high value of arousal or excitement can lead to a high state of nervousness, which can also lead to low productivity, as well as a low value of arousal can lead to a lack of motivation and low productivity due to difficulty in concentration. All this reflection demonstrates that the graph is consistent with situations in real life.

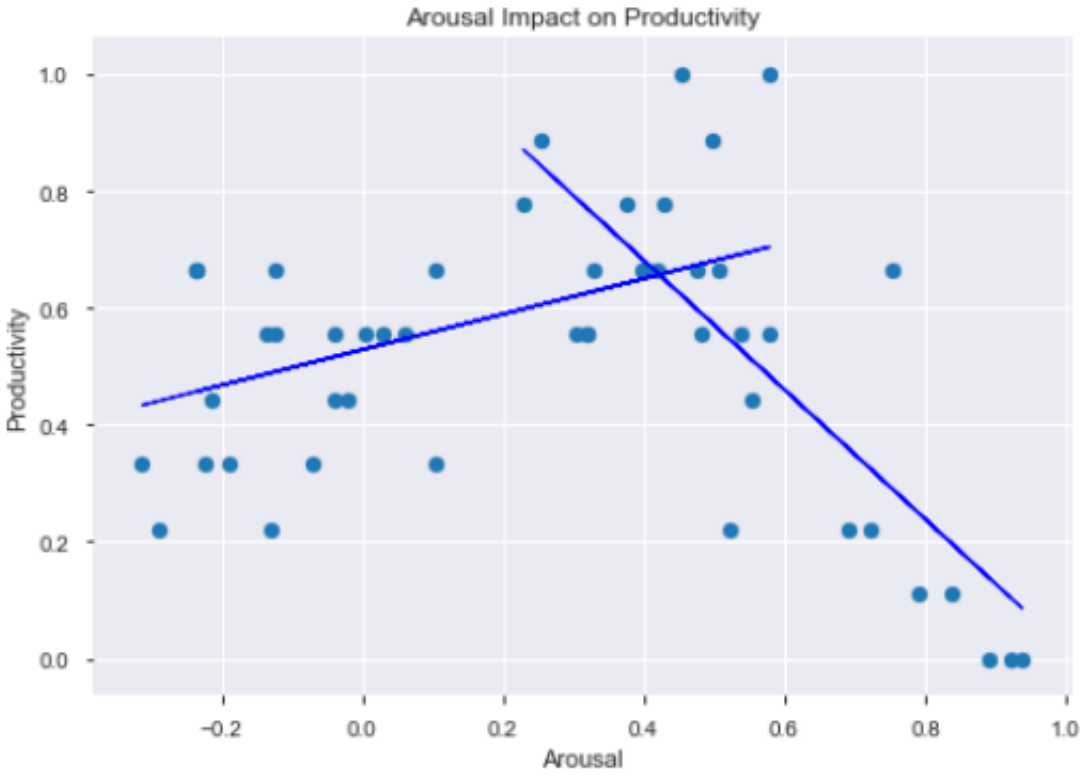

Figure 3.4: Arousal Impact on Productivity [2]

The conclusion of this project is that the emotion-aware system developed is really useful for monitoring the performance of employees or students, decreasing human errors at work or lack of attention at class by being able to analyze the root of the issues and problems of the users while doing their job.

3.4.2 MADE Teacher's Dashboard

The second investigation presents the development of a web application called MADE (Multi-dimensional Emotion Analytics) Teacher's Dashboard [4]. The objective of this project was to gather and display the feelings experienced by students of an online educational platform in order to create a tool that would enable instructors to adjust their lessons to the student's emotional state.

The main motivation for this project is the model of affectivity in education of Kort et al. [45]. This article examines the emotional stages that a student experiences during the course of their learning. By understanding the stage in which a student is at, educators can provide emotional support to help them overcome any challenges they may face. Kort et al. suggested a model of interaction between emotions and learning, a four-part learning spiral model in which emotions shift as the learner progresses through the quadrants and ascends

the spiral. It is widely accepted that students can benefit from the right encouragement and assistance.

One of the bases of the design of this project is the facial microexpressions discovered by Ekman, as already introduced in Sect. 3.1. The application focused simply on the canonical forms of Ekman's categorization of emotions based on facial expression analyzes. Another basis is computer detection of facial expressions of emotions, using a JavaScript library named Clmtrackr [46], which allows the user to detect facial expressions of emotions through a webcam. Clmtrackr is open source and freely available and supports tracking of facial features and detection of emotions in real-time, identifying the six different emotions (disgust, fear, joy, surprise, sadness, and anger).

Like in the first project, this set of emotions emphasizes valence, and other dimensions, such as arousal or dominance, are less explicitly identified.

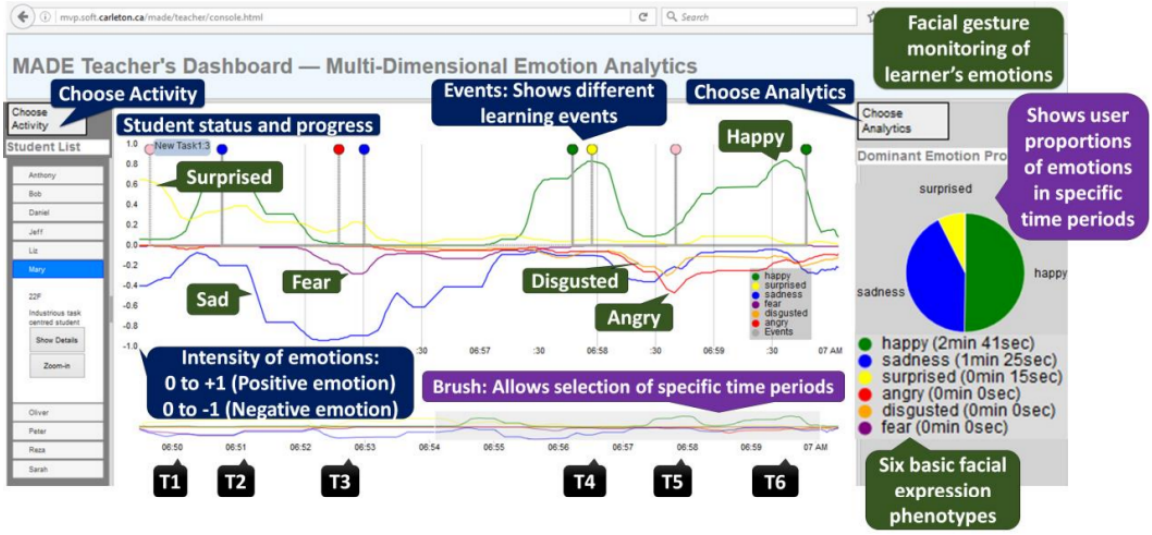

Figure 3.5: The MADE Teacher's Dashboard [4]

The implementation of the dashboard was carried out on an existing mathematical learning platform named MADE Ratio. By utilizing the emotion detection library, it was possible to link the student's emotional state to the learning activity they were engaged in. The visualization, depicted in Fig. 3.5, was a graph that illustrated the student's emotions and how they changed over time, as well as a pie chart that showed the proportion of emotions detected during a certain period.

3.4.3 ZOE Emotion Research Lab

Emotion Research Lab is an organization dedicated to measuring emotions in real-time. They record the emotional response through facial analysis and eye tracking. Its primary

objective is to help brands better understand the behavior of subjects with respect to their products and promotions.

In this particular project, the documentation is really limited due to there are only a few videos showing a tutorial on the application. Zoe [5] is an Emotional Intelligence Platform that automatically detects and recognizes emotions from videos and provides intelligent feedback and insights using facial recognition. This application performs video analysis and surveys using emotional intelligence. Fig. 3.6 is extracted from the platform tutorial video. The figure shows two of the three-page reports generated by the platform. The first page shows the global data of the analyzed videos as doughnut charts showing the percentage of emotional performance, activation, and positive and negative sentiment. It also presents the activated emotional experience showing the percentage of each detected emotion in the video as a doughnut chart and highlighting the main activated emotion and the less dominant emotions.

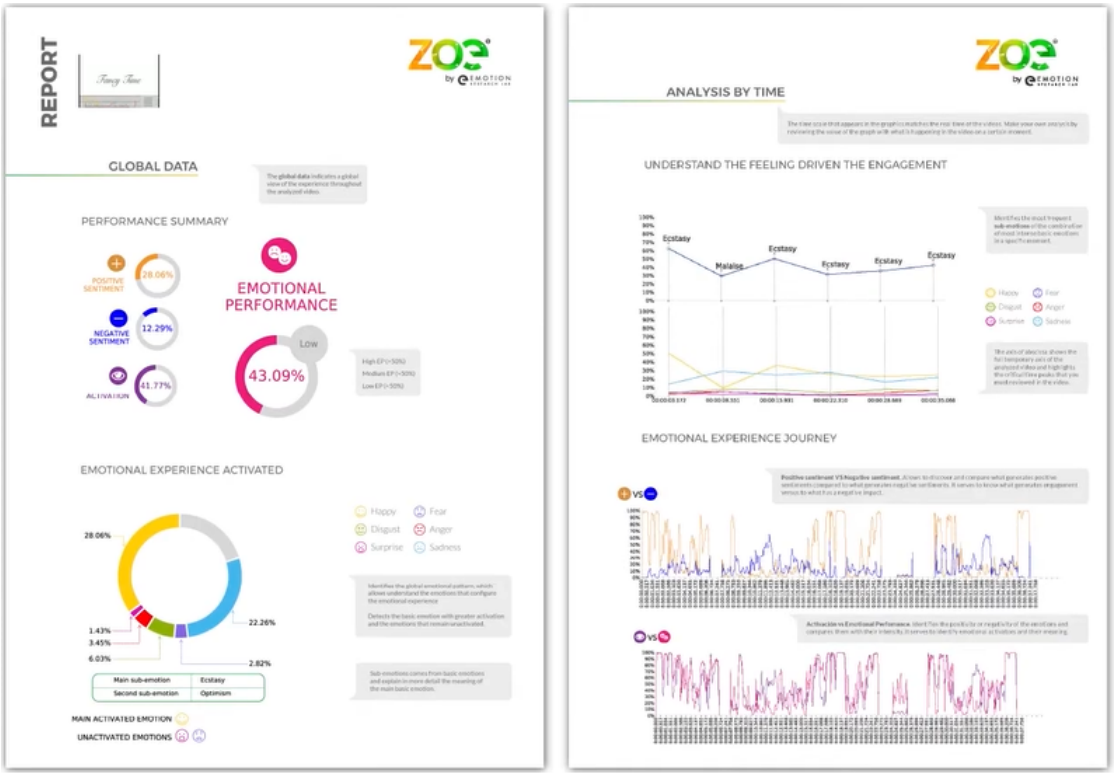

Figure 3.6: Tutorial ZOE Report 1 [5]

The second page of Fig. 3.6 displays the data extracted from the video over time. The first two line charts represent the engagement and the emotions over time, where is easy to understand the correlation between happiness and engagement in this particular example. The third and fourth line charts display the comparison between positive and negative sentiment detected over time and the comparison between emotional performance and activation,

respectively.

Fig. 3.7 shows the last page of the report, where the emotional metrics are explained. This project measures 6 basic emotions, 158 secondary emotions, and 4 emotional metrics.

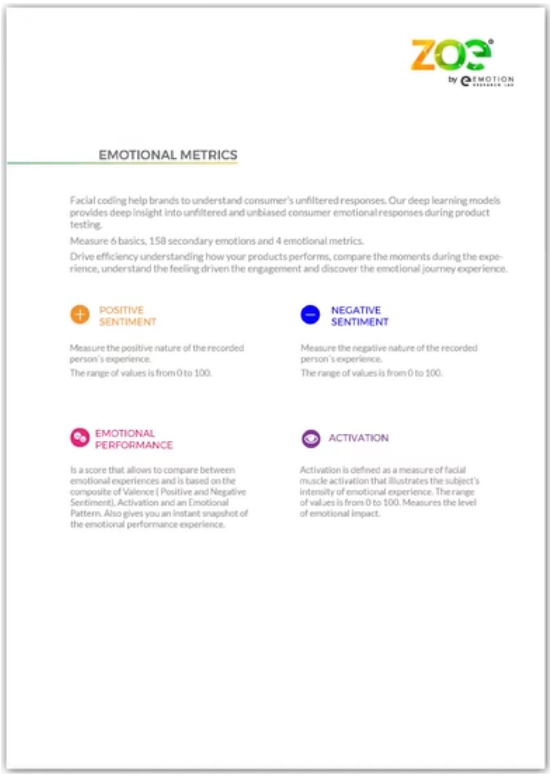

Figure 3.7: Tutorial ZOE Report 2 [5]

The measured emotional metrics and their explanation are as follows:

- Positive sentiment: Measure the positive nature of the recorded person's experience. The range of values is from 0 to 100.
- Negative sentiment: Measure the negative nature of the recorded person's experience. The range of values is from 0 to 100.
- Emotional performance: This is a score that allows one to compare emotional experiences and is based on the composite of valence (positive and negative sentiments). Activation and an Emotional Pattern. Also, gives an instant snapshot of the emotional performance experience.
- Activation: Activation is defined as a measure of facial muscle activation that illustrates the intensity of the emotional experience of the user. The range of values is from 0 to 100. Measures the level of emotional impact.

This project also generates an analyzed video in which all these parameters are displayed on the screen, showing the emotional metrics near the detected face and an emotional bar chart accompanied by an emotional line chart over time on the bottom side of the video. The analyzed video with its graphs is shown in Figure 3.8.

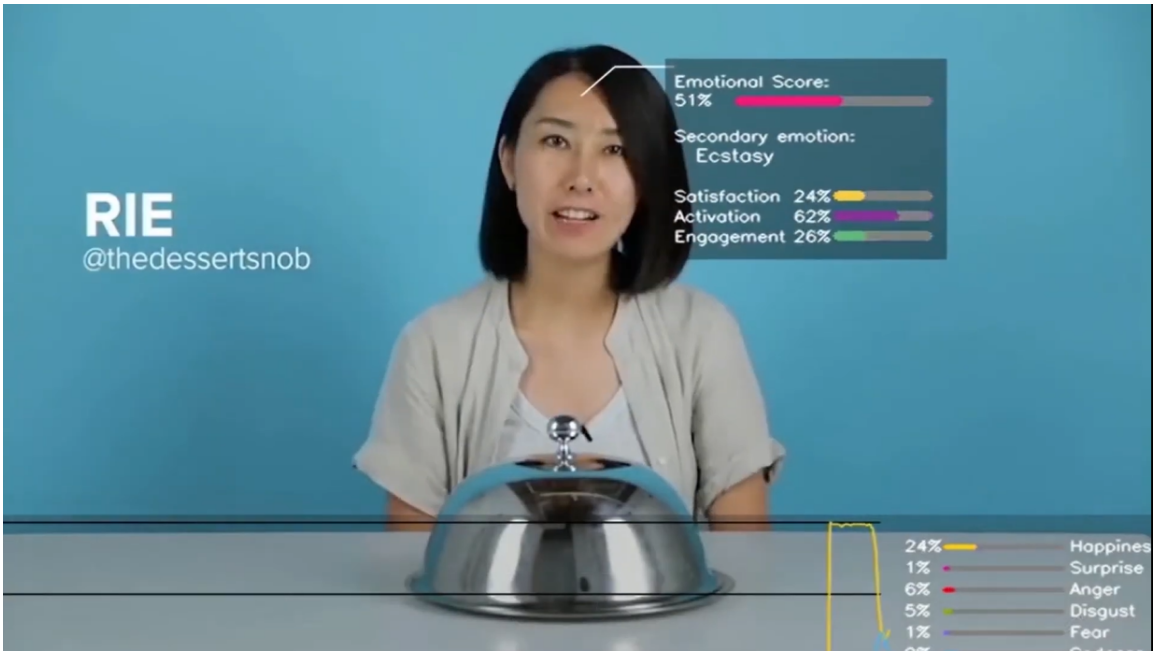

Figure 3.8: Tutorial ZOE analyzed video [5]

Finally, the project also develops a voice transcription, but it does not analyze the emotions from the text; it just transcribes the audio and represents the text. This project is really interesting due to the high capabilities of the analysis and the report generation. Still, the programming language and the source code are private, so it is not possible to know how it is actually developed.

3.4.4 Affdex

Affectiva [6], previously known as Affectiva Affdex, is a pioneering force in the realm of emotion recognition technology. Rooted in the fields of computer vision and artificial intelligence, Affectiva's innovative solutions offer a profound understanding of human emotions through the analysis of facial expressions and behavioural cues.

The goal of the creators of this software is to digitize emotions in order to help people in many aspects of life. There are many possibilities and uses of Affdex because its operation is based on facial coding; in the analysis of images of faces to detect features such as smiles of smugness, frowns, smiles of happiness or raised eyebrows. That is why everything that involves emotions could implement this software to make people's lives easier and more

complete.

Affectiva's facial expression analysis is at the core of its capabilities. Its technology is capable of accurately detecting and interpreting a range of emotions that people show on their faces, from a slight lift of an eyebrow to a fleeting smile. The algorithms are designed to capture subtle movements that express emotions such as joy, sadness, anger, surprise, and more. This analysis goes beyond mere recognition, exploring the intensity, timing, and development of emotional expressions, providing a comprehensive view of emotional involvement.

Researchers and advertisers utilize the AI platform Affectiva emotion to evaluate essential brand media, such as digital ads. As can be seen in Fig. 3.9, the affdex dashboard has different configurable parameters.

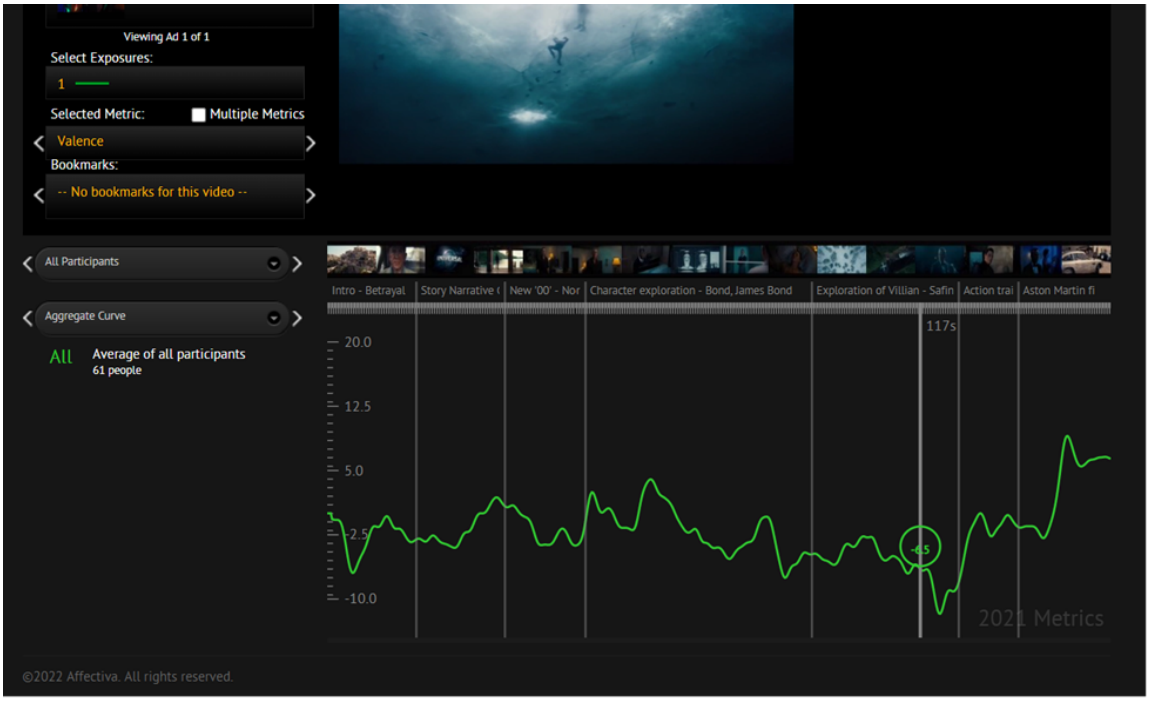

Figure 3.9: Affectiva Emotion Analytics Dashboard - Affdex [6]

The Affectiva dashboard gathers critical insights from user emotion metrics that affect outcomes such as purchase intent and brand lift. There is also a wide array of metrics that advertisers can select to display alone or with multiple metrics at the same time. Additionally, there is an option to compare metrics based on demographic data, such as gender or age. This project is complemented by user responses to surveys about the advertisement that has been chosen.

3.4.5 Emotion-Aware System Design for the Battlefield Environment

In the realm of modern warfare, the convergence of technology and human experience has created a new frontier of research and innovation: emotion-aware system design for the battlefield environment [47]. This cutting-edge field seeks to harness the power of emotion recognition technology to improve decision-making, optimize mission outcomes, and ensure the well-being of military personnel navigating complex and high-stakes scenarios.

The intensity of the battlefield environment, combined with the pressure to make quick decisions, can cause strong emotions. Stress, fear, determination, and focus are all emotions that can have a major influence on the decisions made by soldiers and commanders. When these emotional states are recognized and utilized, the effectiveness and safety of military operations can be greatly improved.

Designing emotion-aware systems lies at the crossroads of psychology, technology, and human-machine interaction. By incorporating sensors, data analysis, and machine learning algorithms, these systems can recognize and interpret emotional signals displayed by soldiers. Facial expressions, physiological indicators, vocal intonation, and even keystrokes can provide information on the emotional states of people in real-time. This abundance of data provides commanders and support teams with the resources they need to make informed decisions that take into account not only tactical strategies, but also the emotional health of their personnel.

3.4.6 Speech Emotion Recognition Project

The last research analyzed is a web application based on the ML model for the recognition of emotions for selected audio files. Speech Emotion Recognition Project [7] is released under the Massachusetts Institute of Technology (MIT) License and is part of the final data mining project for the Institute for Theory and Computation (ITC) Fellow Program 2020.

The field of Speech Emotion Recognition (SER) is a rapidly developing area of research that combines linguistics, signal processing, and artificial intelligence. It focuses on deciphering and interpreting the emotional nuances present in human speech. This evolving discipline is attempting to bridge the gap between human communication and technology, allowing machines not only to understand the words we say, but also to detect the emotional tones that give them meaning.

Typically, the SER task is divided into two main sections: feature selection and classification. The discriminative feature selection and classification method that correctly recognizes the emotional state of the speaker in this domain is a challenging task.

The pipeline of this project is shown in Fig. 3.10. It can be seen that there are two feature selections, MFCCs and Mel-specs. In this case, in the web application, it is possible

to choose which one is used to analyze the uploaded audio.

- MFCCs: These coefficients represent the short-term power spectrum of a sound by transforming the audio signal, so they are considered an important feature for SER. It is the most researched and utilized feature in research papers and open-source projects. For this feature, a Convolutional Neural Networks (CNNs) classifier is used to extract the salient and discriminative features.
- Mel-specs: The Mel scale is significant because it more accurately reflects the way humans perceive sound compared to linear scales. In this case, a trained DenseNet is used as a classifier.

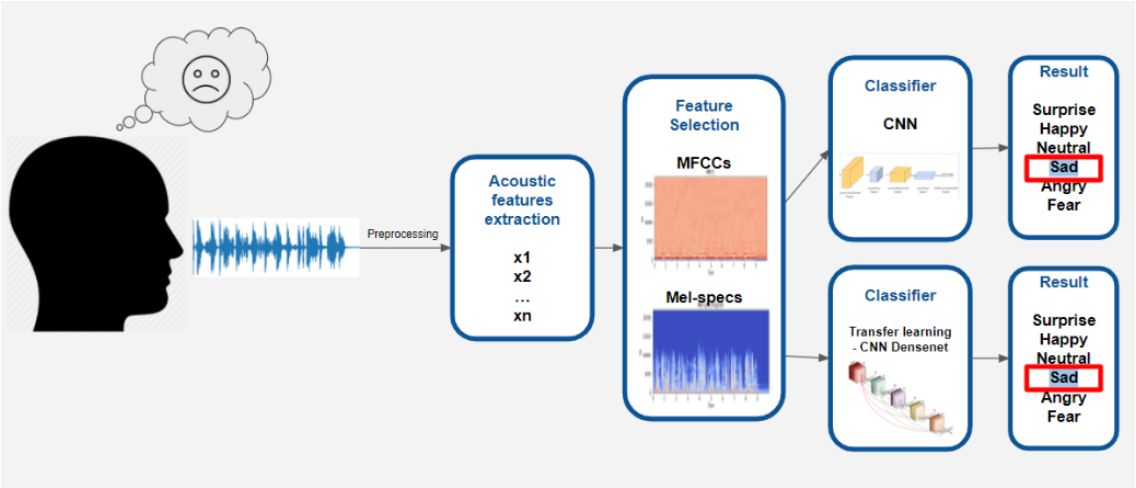

Figure 3.10: Speech Emotion Recognition Pipeline [7]

The web application is developed for Streamlit, like the system developed for this thesis. Fig. 3.11 shows an example of how to use it. First, the user chooses the options available in the sidebar of the web application. There are several options; the first is the application the user wants to utilize; in our analysis, the most relevant is emotion recognition. The second option is the model; as explained before, there are two possibilities, MFCCs and Mel-specs. As our goal is to analyze 7 different emotions, MFCCs needs to be chosen. Finally, there are additional settings that allow the user to choose how many emotions are analyzed and whether or not to analyze the gender of the speaker. If only 3 emotions are selected, the model can only predict positivity, negativity, or neutrality in the voice tone; if 6 emotions are selected, the model predicts one of these emotions: fear, angry, neutral, happy, sad, or surprise; at last, if 7 emotions are selected, the disgust emotion is added to the 6 emotions mentioned above.

Second, the user uploads the audio file to be analyzed; in this example, the audio uploaded shows the angry emotion.

Finally, the web application displays the analysis done, showing the waveform of the signal detected from the audio, an audio player if the user wants to listen to the uploaded audio, two graphs that represent the analysis using both MFCCs and mel-log spectrogram, and the predictions depending on the settings selected on the sidebar.

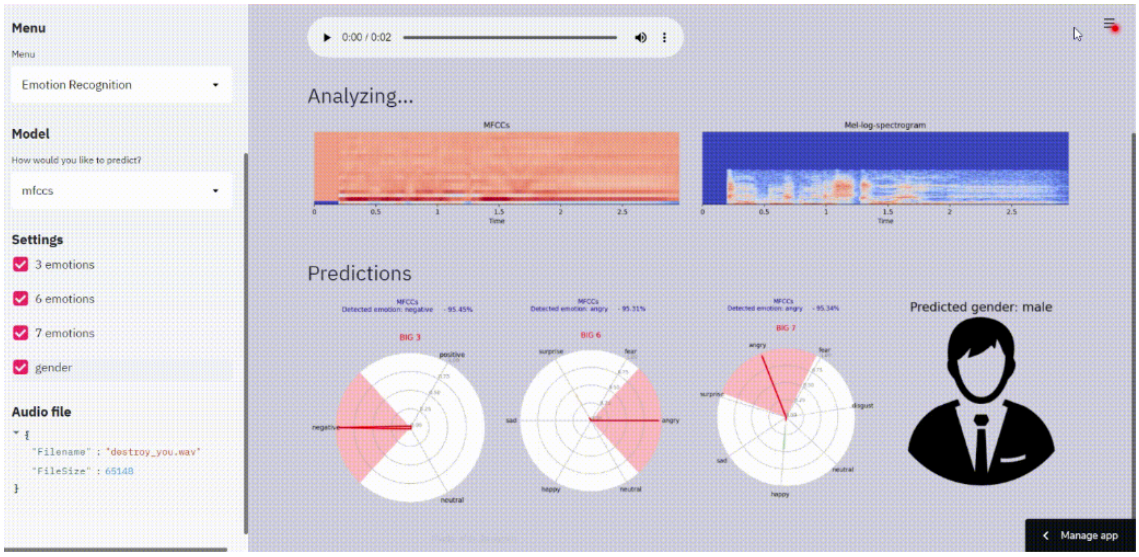

Figure 3.11: Speech Emotion Recognition [7]

In conclusion, this project managed to develop a dashboard in which audios are analyzed with two different models. This application is a really good app for speech emotion recognition, although the predictions displayed are too simple and only show the dominant emotion of the analyzed audio. This project has different pre-training models for emotional analysis, and one of them was used to develop the speech emotion recognition of the application developed for this thesis.

CHAPTER 4

Requirements analysis

This chapter outlines the requirement analysis, which is essential for software development. Use cases are outlined and discussed in order to identify the requirements that the system must satisfy in order to meet expectations. This allows us to focus on the most important features and leave out any less significant functionalities that could be addressed in future projects.

4.1 Use cases

In this section, the use cases are outlined and discussed in order to identify the requirements that the system must satisfy to meet expectations. As can be seen in Fig. 4.1, this diagram illustrates the use cases and the actors involved. The subsequent sections explain the actors involved and how to interact with the system to generate the final list of requirements.

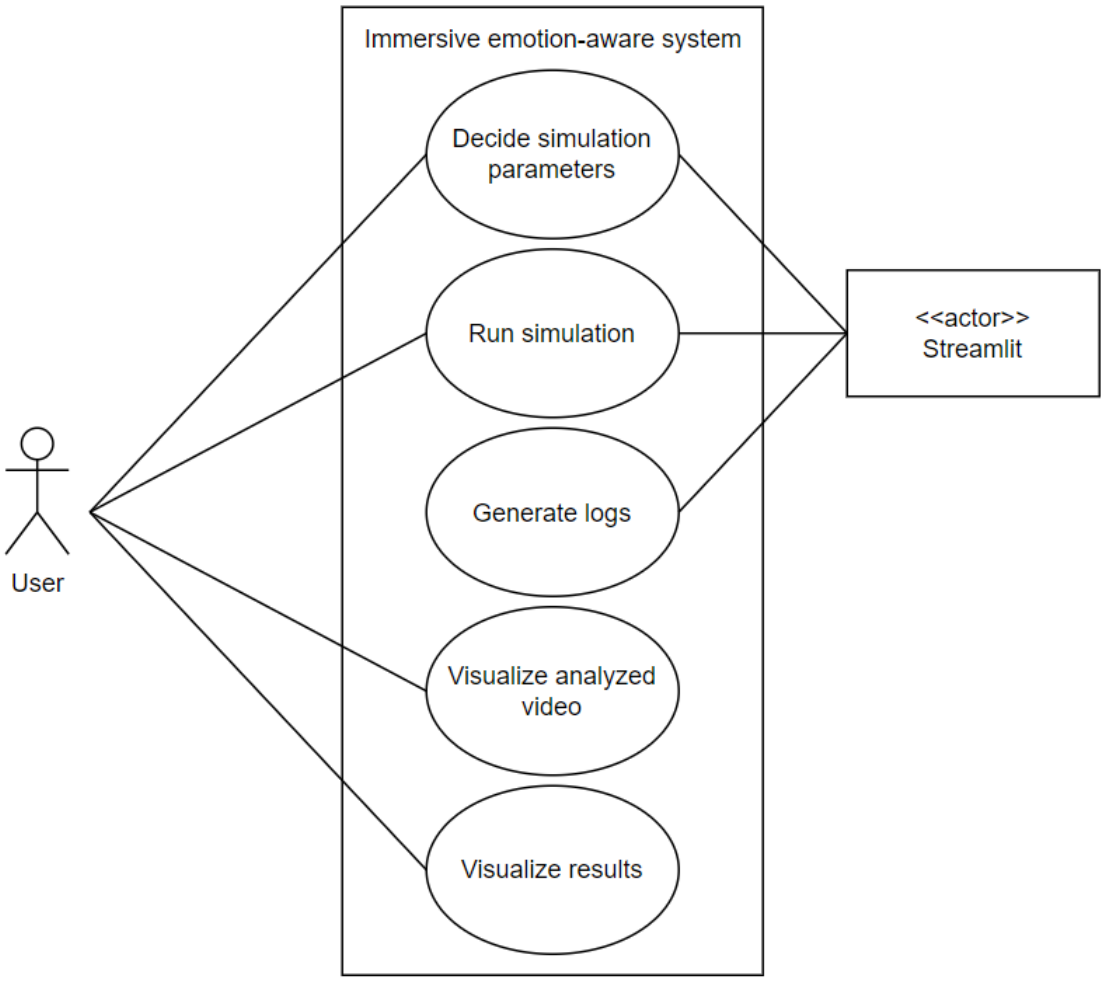

Figure 4.1: Use Cases

The system is used by two actor roles, user and Streamlit, as a secondary actor. The user is the main actor. He uses the system for i) to decide one of the options from the menu, choosing between real-time analysis, uploading a video or visualizing the results of the analysis of a video from the gallery; ii) run the simulation after entering the name under which the experiment is to be saved; iii) to visualize the analyzed video generated by the system adding an emotion bar diagram next to the detected faces and the rectangles where the detected faces are; and iv) to visualize the results, through a visual interface. Also, the

analyzed video and the data extracted from the video in Comma-Separated Values (csv) format are saved. Streamlit helps to support the application and generate the necessary logs for the graphs. The use cases are further detailed in the following subsections.

- Decide menu option, explained in Sect. 4.1.2
- Run simulation, explained in Sect. 4.1.3
- Generate logs, explained in Sect. 4.1.4
- Visualize analyzed video, explained in Sect. 4.1.5
- Visualize charts, explained in Sect. 4.1.6

4.1.1 Actors

The problem involves two actors: Users and the Streamlit application, which is the application framework that supports the emotion-aware system. All their details are included in Table 4.1.

Actor Identifier	Role	Description
ACT-1	User	The main user of the application. He decides between the menu options offered, uploads the video or uses the real-time option and visualizes the results.
ACT-2	Streamlit	The application framework. It is used to host the application and to visualize the generated analysis.

Table 4.1: Use cases

4.1.2 UC1: Decide menu option

The first use case of the project is to decide which of the menu options available in the sidebar the user wants to use. This use case is explained in Table 4.2

4.1.3 UC2: Run simulation

This use case is the second one to perform in the process. The system has multiple options that the user can select. The use case contains, in turn, three additional use cases, one for each of the available options of the system.

Use Case Name	Decide menu option
Use Case ID	UC1
Primary Actor	User
Pre-Condition	The application shows the home screen and the user visualizes the available options.
Post-Condition	-
Flow of Events	1. The user starts the application. 2. The system shows the home page in Streamlit. 3. The user decides which of the menu options in the sidebar he wants to use. 4. The system shows the page of the selected option.

Table 4.2: Decide menu option

- Real-time Emotion Recognition, explained in Sect. 4.1.3.1
- Upload Video for Analysis, explained in Sect. 4.1.3.2
- Analyzed Videos Gallery, explained in Sect. 4.1.3.3

4.1.3.1 Run simulation: Real-time Emotion Recognition

This use case is explained in Table 4.3

4.1.3.2 Run simulation: Upload Video for Analysis

This use case is explained in Table 4.4

4.1.3.3 Run simulation: Analyzed Videos Gallery

This use case is explained in Table 4.5

4.1.4 UC3: Generate logs

After carrying out all the configurations, the application can generate the logs. This use case is explained in Table 4.6

Use Case Name	Run simulation: Real-time Emotion Recognition
Use Case ID	UC2.1
Primary Actor	User
Pre-Condition	The user has selected the "Real-time Emotion Recognition" option.
Post-Condition	-
Flow of Events	1. The user selects the "Real-time Emotion Recognition" option. 2. The system displays the corresponding page. 3. The user gives permission to the application to use his webcam. 4. The system displays the video collected by the webcam with a blue rectangle around the detected faces and a bar chart near them showing the emotions in real time of every detected face. 5. The system also detects smiles and draws a red rectangle around the lips. 6. Finally, the system displays a dynamic table showing the percentage of all the emotions the system can detect.

Table 4.3: Run simulation: Real-time emotional recognition

Use Case Name	Run simulation: Upload Video for Analysis
Use Case ID	UC2.2
Primary Actor	User
Pre-Condition	The user has selected the "Upload Video for Analysis" option.
Post-Condition	-
Flow of Events	1. The user selects the "Upload Video for Analysis" option. 2. The system displays the corresponding page. 3. The user writes the name of the experiment in the "Experiment name" text input widget and uploads the video he wants the system to analyze. 4. The system analyzes the uploaded video, generating the analyzed video and the emotions and gender logs with the necessary data to generate charts; extracts the audio and its emotion logs; and transcribes the extracted audio and its emotion logs.

Table 4.4: Run simulation: Upload Video for Analysis

Use Case Name	Run simulation: Analyzed Videos Gallery
Use Case ID	UC2.3
Primary Actor	User
Pre-Condition	The user has selected the "Analyzed Videos Gallery" option.
Post-Condition	-
Flow of Events	1. The user selects the "Analyzed Videos Gallery" option. 2. The system displays the corresponding page. 3. The user selects the video from which he wants the system to display the analyzed video and the generated graphics. 4. The system displays the analyzed video and the generated. charts.

Table 4.5: Run simulation: Analyzed Videos Gallery

Use Case Name	Generate logs
Use Case ID	UC3
Primary Actor	Streamlit
Pre-Condition	The user has uploaded a video to the system.
Post-Condition	-
Flow of Events	1. The system analyzes the uploaded video. 2. The system generates the logs. 3. The system saves the logs in csv format.

Table 4.6: Generate logs

4.1.5 UC4: Visualize the video analyzed

Once the video has been analyzed after uploading it or the user selects a video from the gallery, the application presents it in the first tab of the analyzed video. This use case is explained in Table 4.7

Use Case Name	Visualize analyzed video
Use Case ID	UC4
Primary Actor	User
Pre-Condition	A video has been analyzed or selected to be displayed.
Post-Condition	-
Flow of Events	1. The user uploads a video to the system for its analysis or selects a video to display its analysis 2. Streamlit displays the analyzed video in the first tab. 3. The user can play the video and visualizes the detected emotions in every frame.

Table 4.7: Visualize analyzed video

4.1.6 UC5: Visualize charts

Once the video has been analyzed after uploading it or the user selects a video from the gallery, the application presents the results. This use case is explained in Table 4.8

Use Case Name	Visualize charts
Use Case ID	UC5
Primary Actor	User
Pre-Condition	A video has been analyzed or selected to be displayed.
Post-Condition	-
Flow of Events	1. The user uploads a video to the system for its analysis or selects a video to display its analysis 2. Streamlit displays all the charts generated from video, audio and text. 3. The user can select the desired tab to see the emotion charts generated from the video; the audio with its emotion pie chart associated, and the text generated from the transcription of the audio and its emotion pie chart generated too.

Table 4.8: Visualize charts

4.2 Requirements

Software engineering distinguishes two different types of requirements: Functional and Non-functional requirements. After recognizing the use cases associated with the issues on which this master thesis is focused, the analysis yields the following requirements.

4.2.1 Functional requirements

Functional requirements are features or functions of the product that developers must implement to enable users to perform their tasks. They detail the actions that the system must perform or must not perform. This section contains the functional requirements for this master thesis.

- **FR1:** *The user* must be able to select one of the options from the menu. The use case described in Sect. 4.1.2 provides this requirement.

- **FR2:** *The user* must be able to run a simulation in real time. The use case described in Sect. 4.1.3.1 provides this requirement.
- **FR3:** *The user* must be able to upload a video to be analyzed by the system. The use case described in Sect. 4.1.3.2 provides this requirement.
- **FR4:** *The user* must be able to select one of the videos from the gallery to see its analysis. The use case described in Sect. 4.1.3.3 provides this requirement.
- **FR5:** *The user* must be able to navigate between tabs to visualize the analyzed video, the audio, the text, and the generated charts. The use cases described in Sect. 4.1.3.2, Sect. 4.1.3.3, Sect. 4.1.5, and Sect. 4.1.6 provide this requirement.
- **FR6:** *The system* should be able to generate the necessary logs from the video. The use case described in Sect. 4.1.4 provides this requirement.

4.2.2 Non-functional requirements

Non-functional requirements are requirements that specify criteria that can be used to judge the operation of a system rather than specific behaviors as functional requirements specify. These do not interfere with the primary functions of the system. The following are the requirements that the project must take into account.

- **NFR1:** *Operation:* The user interface is based on a Web application.
- **NFR2:** *Portability:* The user application is available for all browsers that support Streamlit.
- **NFR3:** *Interface:* The system has to comply with the interfaces (data and protocols) of the external systems with which it interacts: Streamlit.
- **NFR4:** *Resources:* The system limitation in terms of computing resources is determined by Streamlit.
- **NFR5:** *Safety:* No personal data is required to use the application.

CHAPTER 5

Architecture and Methodology

This chapter describes in depth how the system is structured in different modules, how users interact with them and also how the modules interact with other modules by themselves. First, the global architecture of the project is presented and then a detailed explanation of each module is given. The project basis and the initial design idea are also discussed in this chapter.

5.1 Introduction

An introduction of global architecture is needed to understand every aspect and motivation of the actual master thesis. In this way, in this section, the different blocks that participate in the performance of the project are explained, as well as a detailed description of the architecture of the project. Then, a detailed explanation of the layered architecture is included in Sect. 5.2.

Firstly, a general idea of the blocks of the global project is needed. It is mainly composed of the following blocks: two **External Devices**, which are the camera and the microphone; the audio and the video originate from them; the **Video Block** where the video introduced is analyzed using DeepFace [1], detecting all the faces and smiles in it, recognizing the emotions of the user's face and generating the logs; the **Audio Block** where the audio introduced is analyzed, extracting the emotions using NLP and a pre-trained model; the **Text Block** where the audio introduced is transcribed into text and the emotions are extracted using NLP and a pre-trained model as it was done with the audio; and **Streamlit** which is an open source app framework for Machine Learning where the application is hosted, it is responsible of the data visualization providing a user-friendly interface where the user can navigate between the different options provided by the app, analyzing some videos.

This general idea is illustrated in Fig. 5.1.

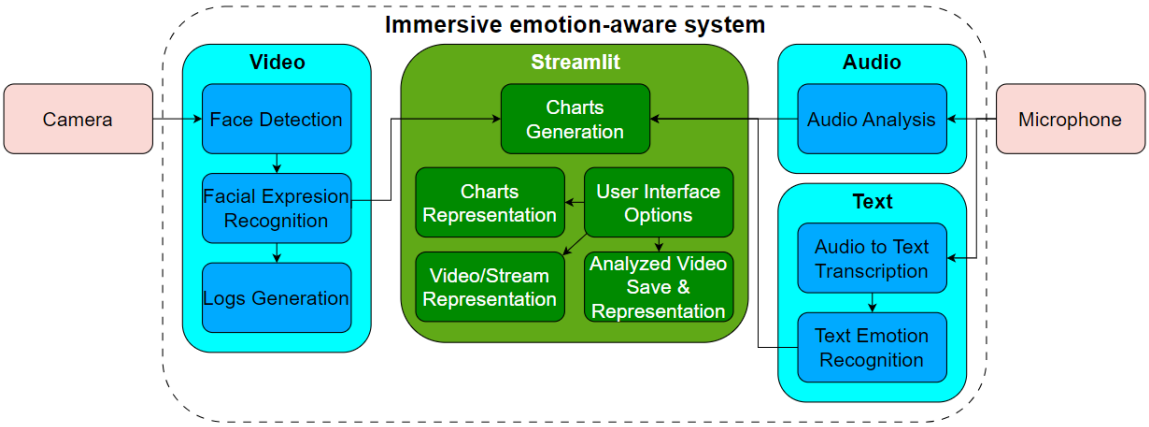

Figure 5.1: Full system architecture

5.2 General Architecture

In this section, the architecture of the project is explained from a global perspective, including an overview of each layer that interacts in the system and describing how they have been implemented. The architecture is designed to facilitate recognition, analysis, and response

to human emotions.

The general architecture of the system is shown in Fig. 5.2.

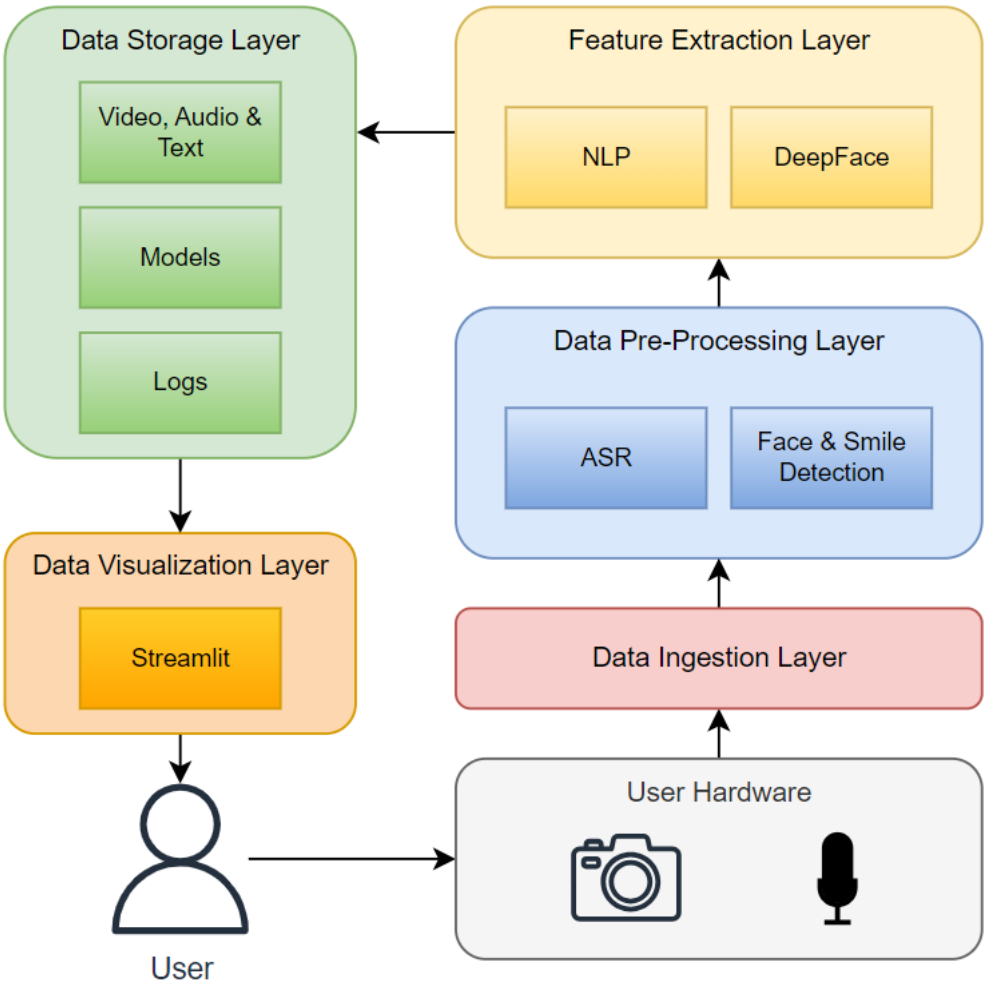

Figure 5.2: General layered architecture diagram

As can be seen in Fig. 5.2, this general architecture is made up of different layers, which are explained in the following sections.

- Data Ingestion Layer, explained in Sect. 5.2.1
- Data Pre-Processing Layer, explained in Sect. 5.2.2
- Feature Extraction Layer, explained in Sect. 5.2.3
- Data Storage Layer, explained in Sect. 5.2.4
- Data Visualization Layer, explained in Sect. 5.2.5

5.2.1 Data Ingestion Layer

The Data Ingestion Layer, often a critical component of a data processing pipeline that serves as the entry point for data into the system, is responsible for collecting and importing data from various sources for further analysis, storage, or processing. This layer is foundational to data-driven systems, ensuring that data are efficiently and securely gathered from diverse origins and made accessible for downstream operations.

In this master thesis, this layer is responsible for collecting the raw video and audio. It connects to the Data Pre-processing Layer (Sect. 5.2.2) that provides the raw video and the extracted audio from the video.

5.2.2 Data Pre-Processing Layer

The Data Pre-Processing Layer, located downstream of the Data Ingestion Layer in a data processing pipeline, plays a critical role in refining and preparing raw data for analysis and modeling. It encompasses a series of operations and techniques aimed at cleaning, structuring, and improving data to ensure its quality and suitability for downstream tasks.

In this layer, one of the most important techniques used is the recognition of faces and smiles in the video. This recognition is carried out using the OpenCV library [12]. Frame-by-frame analysis is performed using this library to recognize faces and smiles in each of the images, painting a blue square on the faces and a red rectangle on the smiles if detected.

Another of the most important techniques is Automatic Speech Recognition (ASR) for the transcription of the audio into text. The Google Speech Recognition API [23] has been used to make this possible. The Google Speech Recognition API is a service provided by Google Cloud [24] that allows developers to convert spoken language into written text. It uses machine learning models and advanced speech recognition technology to accurately translate audio data into text. The primary function of the API is to recognize spoken language in various languages and dialects, with more than 125 languages and variants. It can convert audio from different sources, such as audio files, streaming data, or real-time audio input from microphones.

5.2.3 Feature Extraction Layer

The Feature Extraction Layer is a crucial component in emotion analysis systems. It is responsible for extracting relevant features from the raw data that can be used to characterize and quantify emotional content.

In this master thesis, there are two data sources, video and audio, but from the pre-processing of audio, through its transcription, the text is obtained, so the feature extraction

is also performed on it. In the following sections, the architecture of all the sources analyzed is explained.

5.2.3.1 Video feature extraction

The video architecture for pre-processing and feature extraction is presented in Fig. 5.3.

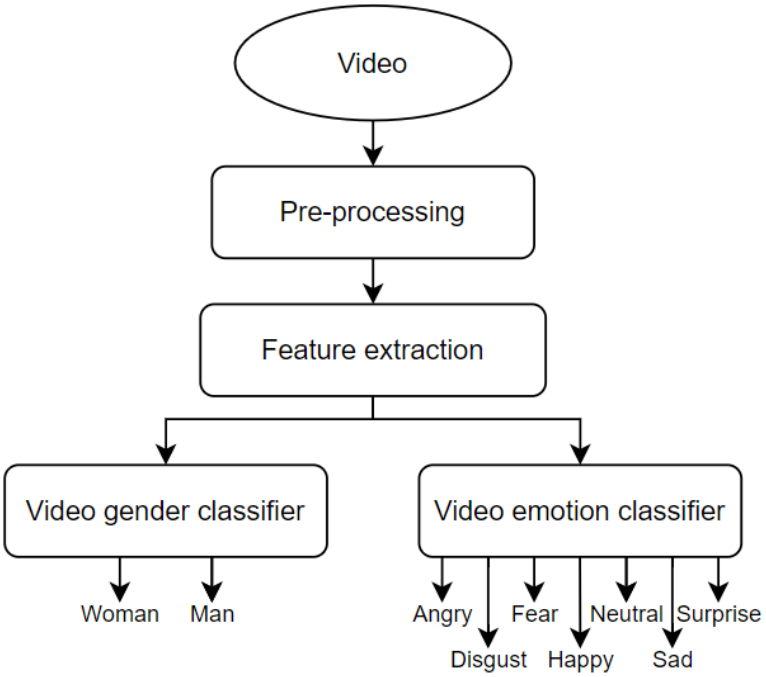

Figure 5.3: Simple video architecture

As can be seen, the extraction of the video features is divided into two blocks, sex and emotion, both of the features classified by DeepFace [1]. The library analyzes the video frame by frame and extracts both gender and emotion from each image. The percentage of each gender and of each emotion is stored in a csv file and later displayed on different pie charts. The genders that DeepFace can detect are “Woman” or “Man”, and emotions are the seven universal facial emotions which are “Angry”, “Disgust”, “Fear”, “Happy”, “Neutral”, “Sad”, and “Surprise”.

Once the emotions are detected on each frame, using OpenCV again, the dominant emotion is written above the top left corner of the detected face rectangle. In addition, a bar chart with the percentage of each emotion is displayed near each detected face. With these two components represented in the video, the generation of the analyzed video is done and is sent later on to the Data Storage Layer.

5.2.3.2 Audio and text feature extraction

Now, the audio and text architecture for pre-processing and feature extraction is presented in Fig. 5.4.

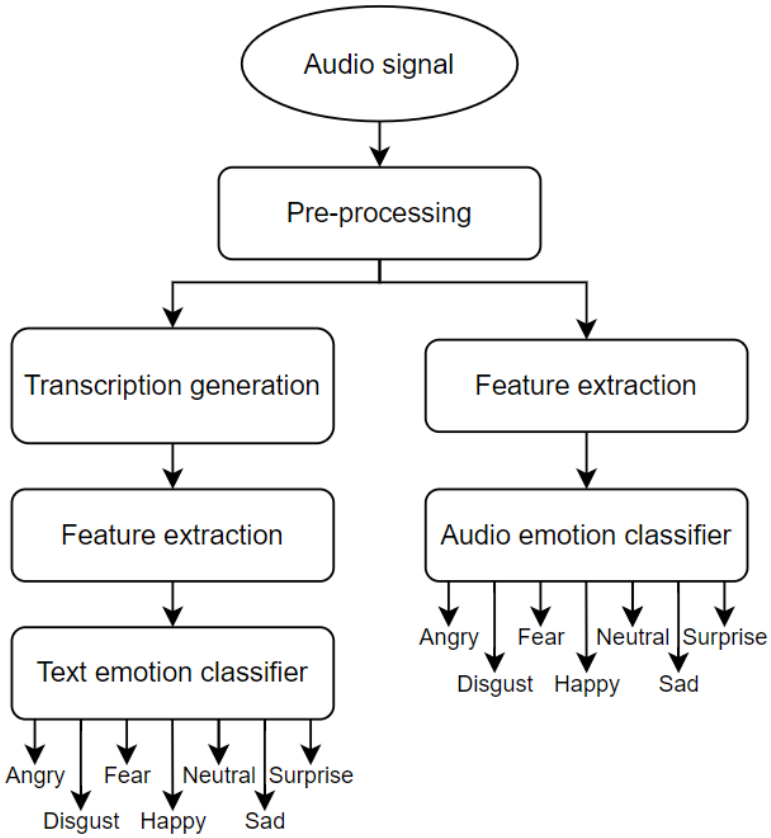

Figure 5.4: Simple audio and text architecture

The main difference between this architecture and the video one is that from the pre-processing of the audio, there is a transcription generation to extract the recognized text from the speech for its further analysis. As can be seen in the figure, the feature extraction of both audio and video extracts the emotions recognized in each of them separately, being able to detect the seven universal facial emotions too.

Both feature extractions are performed by using pre-trained models. The audio emotion classifier model used is the same model used in the speech emotion web app introduced in Sect. 3.4.6, using NLP techniques to analyze the audio and identify the emotions it expresses. The model used is a pre-trained CNN model trained with the following datasets:

- Crowd-sourced Emotional Mutimodal Actors Dataset (Crema-D)
- Ryerson Audio-Visual Database of Emotional Speech and Song (Ravdess)

- Surrey Audio-Visual Expressed Emotion (Savee)
- Toronto emotional speech set (Tess)

The model selected from 3.4.6 is the one that uses MFCCs, which are a representation of the short-term power spectrum of a sound by transforming the audio signal, and are also considered to be an important feature for SER. MFCCs is the most researched and utilized feature in research articles and open-source projects.

In terms of the model used for the extraction of emotions from text, it was trained on a text dataset that was labeled with the corresponding emotions expressed in it. The model is the same as the one used in [48], which is an app for text emotion classifier. The aim of that project is to develop a model that uses NLP techniques to accurately detect emotions in text data.

The machine learning model is trained on the extracted features to predict the emotions expressed in the text data. The model used for this project is a Logistic Regression and MultinomialNB. Logistic regression achieved an accuracy of 62% in the data.

Fig. 5.5 represents a simplified architecture of the immersive emotion-aware system.

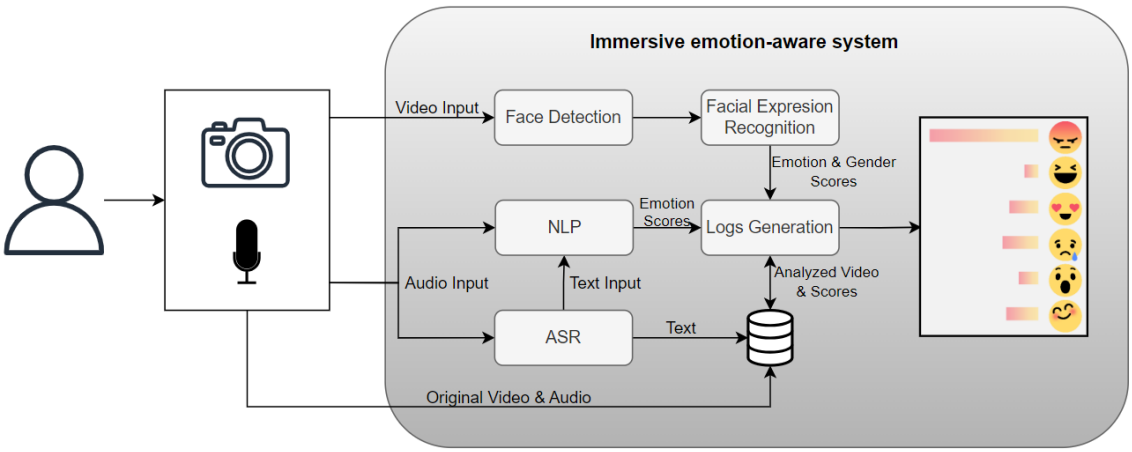

Figure 5.5: Simple architecture

5.2.4 Data Storage Layer

The Data Storage Layer is responsible for efficiently storing and managing the data, ensuring its availability, durability, and accessibility for various downstream tasks. For this master thesis, there are three main directories, which are video, audio, and text.

Inside these three main directories, there are some more directories that separate the raw data from each of the data sources, the emotions recognized, and, only for the video, the analyzed videos edited with the extracted emotions from the original video.

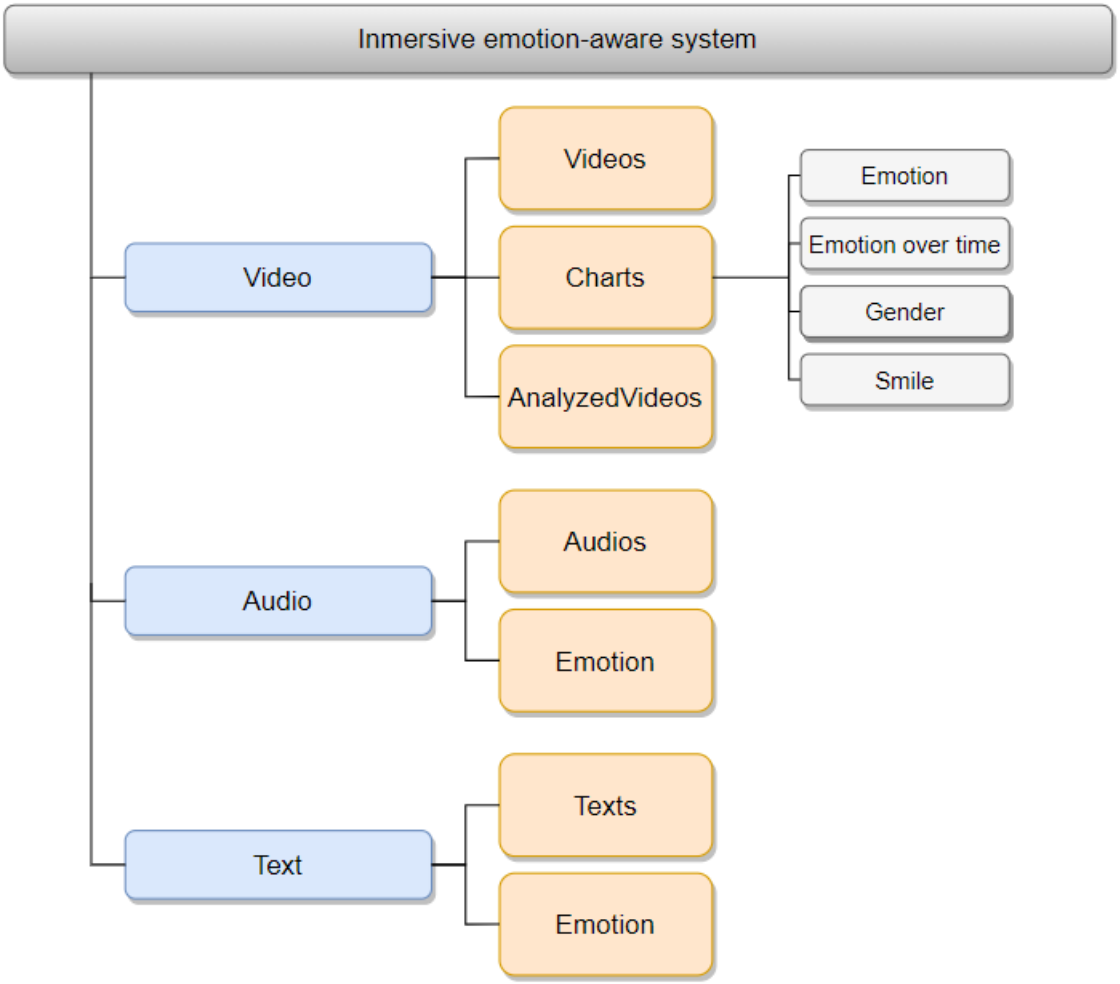

Figure 5.6: Directory structure

Fig. 5.6 shows the directory structure of the project. In this case, the data is stored locally, but if it is necessary to make the application scalable, then using a NoSQL database such as MongoDB [49] could be the best solution.

In the first place, there are three subdirectories in the video folder. The "videos" directory stores the original videos that are uploaded to the application in the "Upload Video for Analysis" option in mp4 format. The original video is visible in the first tab both when it is uploaded to the application and when a video is selected for viewing in the "Analyzed Videos Gallery" option. The "charts" directory stores all the data needed to generate the charts with the information extracted from the video. This directory is also divided into four parts, one for each of the charts displayed, an emotion pie chart, an emotion line chart with the emotion analysis from all the frames, a gender chart and a smile chart. All of this data is saved in csv format due to this format is a data interchange format used when there is a large amount of data, so it is used in the vast majority of databases and commercial applications. Its main use is to move data between programs in tabular form. In this case, the csv file is generated in the "Upload Video for Analysis" option and is used in the "Analyzed Videos Gallery" option. Finally, the "analyzedvideos" directory stores edited videos with rectangles around the detected faces and smiles, the dominant emotion of each frame, and the bar graph showing the percentage of each emotion in each frame.

Secondly, there are two subdirectories in the audio folder. The "audios" directory stores the original audio files extracted from the original video. The audio files are stored in wav format. The "emotion" directory stores the emotions extracted from the audio in csv format as in the videos.

Lastly, the text folder is distributed in the same way as the audio directory. The "texts" folder stores the transcribed text in TXT format, and the "emotions" folder stores the emotions recognized in the text in CSV format.

In addition, there is a models folder where the pre-trained models for audio and text emotion recognition are located.

5.2.5 Data Visualization Layer

The Data Visualization Layer serves as the interface between processed data and human users. It is responsible for transforming complex emotional insights derived from video, audio, or text data into visually comprehensible and actionable representations. Some data visualization techniques used in this emotion-aware system are:

- **Charts:** Pie charts and bar charts can represent emotional trends, results of sentiment analysis, or the prevalence of specific emotions in the data. These visualizations make

it easy to spot emotional changes.

- **Emotion Overlays:** In video data, emotions can be visualized by overlaying emotional labels or color-coded facial expressions onto video frames, making it visually apparent when and where emotions occur.

- **Time Series Plots:** Time series plots can reveal how emotions evolve over time in audio, video, or text data. This helps to identify emotional trends or changes.

As mentioned above, the application consists of three pages apart from the home page. Now, the distribution of the pages and what they display is explained briefly. In Sect. 5.3, the mockups of each page are presented so that the main idea of the project and its possible uses are understood.

- **Home Page:** This is the first page that the user sees when he starts the application. On this page, a welcome message is displayed, the different options that the application has, and an option menu is displayed where the user can choose which of the three options he wants to use. Sect. 5.3.2.1 shows a mock-up of this page.

- **Real-time Page:** This page is very different from the two that follow. For its operation, it is necessary to allow the use of the webcam in the application. Once this option is selected, it shows the video from the webcam edited with the components explained in Sect. 5.2.3.1. The application can detect multiple faces, analyzing all of them and showing the results on the screen.

- **Upload Videos Page:** On this page, as mentioned in previous sections, it is possible to upload a video for the system to analyze all its features. On the right side of the video uploader, there is a single-line text input widget where the user has to type the name under which he wants to save the experiment he is performing for later viewing on the video gallery page.

  The first step is the analysis of the emotions and gender of the video, generating the corresponding files for the creation of the charts and the analyzed video shown in the first tab. Subsequently, the audio is extracted and the emotions are analyzed using MFCCs and transcribed to generate the text in which the analysis is performed using both NLP techniques.

  After finishing the analysis of all sources, in the same tab where the video has been uploaded, both the original video and the analyzed video appear, allowing the user to visualize the frame-by-frame analysis that has been performed. In the next tab, two pie charts are displayed; the first chart shows the detected percentage of each of the

seven emotions that make up the dataset, while the second chart shows a pie chart with the percentage of time in which a smile has been detected in the video compared to the time in which it has not been detected.

The third tab shows a line chart in which all the emotional data collected from the video frame by frame are included. The horizontal axis represents the frame being analyzed, while the vertical axis represents the percentage of emotion detected. Each of the seven emotions that DeepFace can detect is represented by a different colored line, making it easy to distinguish one emotion from the others. Fig. 5.7 is an example of a line graph generated by the application made with the Apache ECharts library.

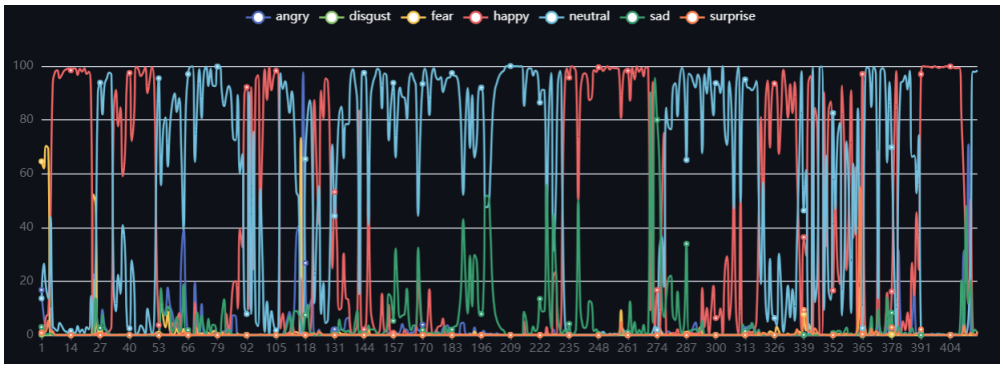

Figure 5.7: Line chart example

The fourth tab shows the gender analysis made by DeepFace. In this tab, a pie chart is included showing in percentage the two genres that the library is able to detect, and next to it, an icon of a man or a woman is shown depending on which of the two genders has been detected.

The last two tabs show the analysis performed on audio and text, respectively. The audio tab displays an audio player in case the user wants to listen to the sound extracted from the video, and a pie chart showing the emotions as it was done in the video tab. Regarding the text tab, the structure is similar to the audio tab showing the text that has been transcribed from the audio and a pie chart showing the percentage of emotions as well.

- **Videos Gallery Page:** This last page is similar to the previous one. The tab structure is the same for both. The main difference is that when selecting this page, instead of a video uploader, a select widget is displayed. This select widget allows the user to choose from the videos he has analyzed and saved in his storage.

5.3 Project basis and design for the development of an Emotion-Aware System

Once the architecture has been explained, the basis and design of the project are presented. By knowing the emotions of users performing different activities, whether for e-learning, teleworking, or for an assistive robot for elderly people, it is possible to make variations in their environment or in the number of responsibilities they are taking on that would allow them to improve their efficiency or quality of life.

The development of this master thesis focuses primarily on the collection of data related to users' emotions, as well as their gender. The user who uploads the video is able to better understand the emotions the people in the video are showing. It is also possible to analyze the emotions that these people show through their speech and the words they are pronouncing, so it is possible to check if the emotions they show through all the sources are the same or if they are faking it.

The motivation and objectives of each part of this development are explained below.

5.3.1 Data collection

Data collection is based primarily on the emotions experienced by the user during the video. However, it is necessary to collect more information to increase the accuracy of the system, such as audio and speech. The tools used to obtain these data are the following:

- Video emotion and gender: video and webcam-based emotion and gender recognition.
- Audio emotion: Mel-Frequency Cepstral Coefficients.
- Text emotion: Natural Language Processing.

5.3.2 Data visualization

Data visualization allows the user to establish cause-and-effect relationships from the collected data. To achieve this goal, a dashboard is created showing visualizations of the evolution of emotions throughout the video.

Once the information to be collected and the basis of the project was clear, it was necessary to design the application. The design of this project began with the creation of mock-ups for each of the pages of the application.

The following sections show each of the mock-ups developed, along with an explanation of the design and the changes introduced for the aesthetic and functional improvement of the application.

5.3.2.1 Home page

The home page is the first page that is displayed when the application is started. Fig. 5.8 shows the design idea for this page.

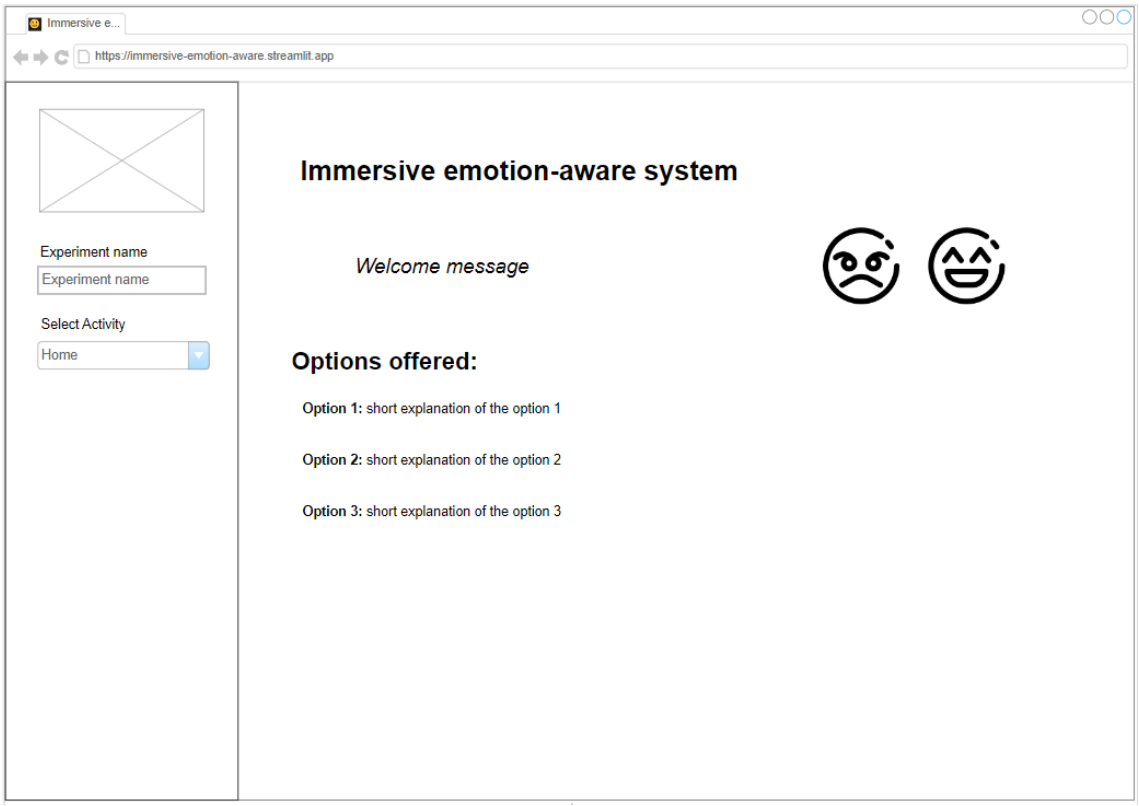

Figure 5.8: Mock-up home page

This page was the easiest to create because its functionality is to present the application and provide a brief explanation of the possible options it offers. Once the user knows the functionalities that the application can offer, he can select from the sidebar menu which of the three options he wants to try. In the following sections, the three options on the sidebar menu are explained in detail.

5.3.2.2 Real-time page

This page was the first one developed. Fig. 5.9 shows the design idea for this page, showing the video from the webcam analyzed in real time together with a series of charts with the emotions detected.

The first idea for this page was to be able to record the video while displaying it on the screen, along with the analysis of emotions. The "Start" and "Stop" buttons on the bottom left of the figure would provide this functionality. The main problem that did not

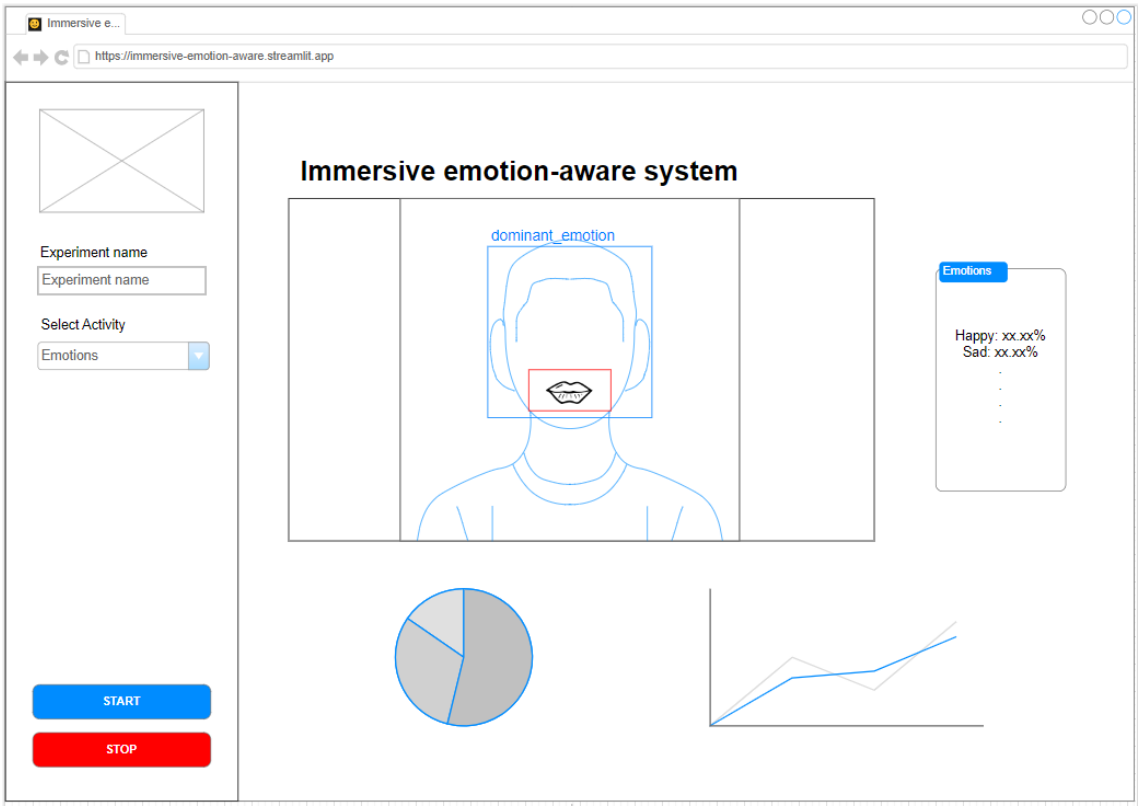

Figure 5.9: Mock-up real-time page

allow this feature to be implemented was the high computational need to analyze the video while saving it, causing problems in the recording and making the video not smooth.

Regarding the rest of the sidebar, the text box with the name of the experiment appears only on the page where the videos are uploaded, and the drop-down widget is replaced by an options menu where the four selectable pages appear.

Also, the idea of displaying charts in real-time was discarded because when analyzing each video frame, the chart would be updated every few milliseconds, making it difficult to see the results displayed on the chart. Despite the difficulty of seeing the charts, a bar chart was added to the analyzed video because in the following pages, it is possible to stop the analyzed video and visualize the chart.

Finally, the interactive table on the right side of the video was displayed, showing the percentage of the emotions detected in real-time.

5.3.2.3 Upload videos page

This page is the one that has required the most work to develop. Fig. 5.10 and Fig. 5.11 show the design idea for this page.

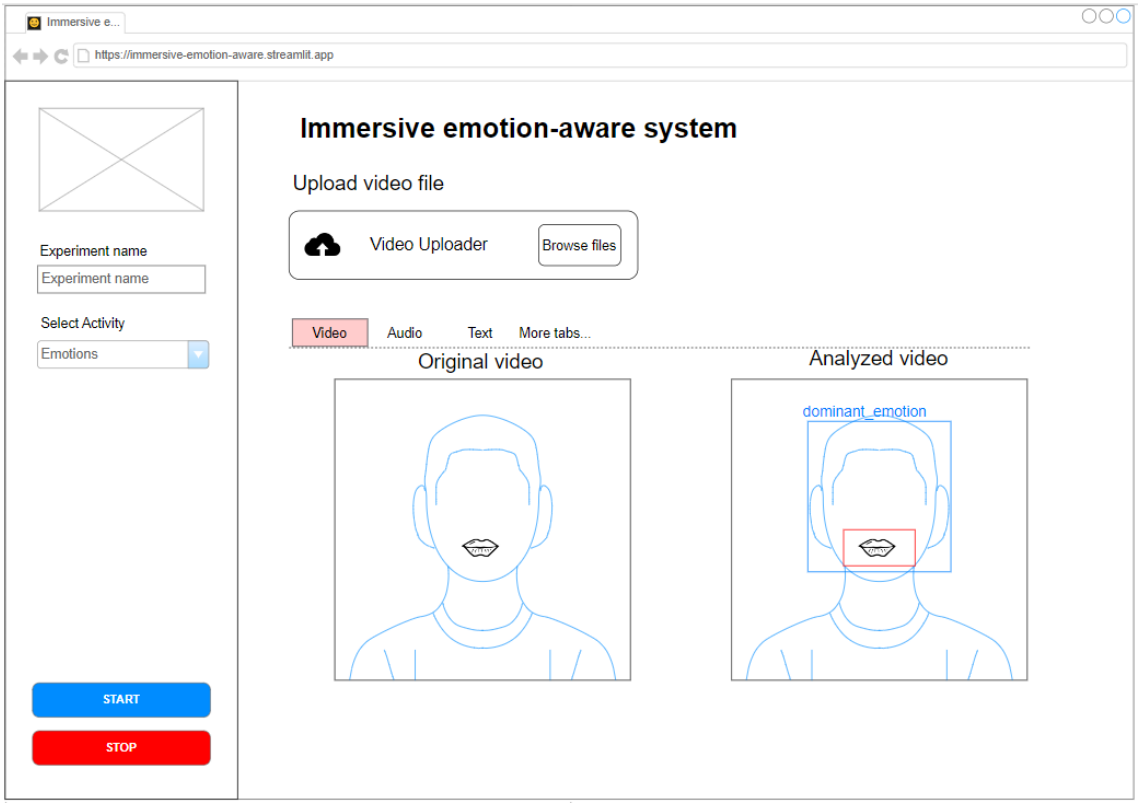

Figure 5.10: Mock-up upload videos page video tab

This page is made up of several tabs that show different charts about the data collected.

The first of the tabs is shown in Fig. 5.10. Once the analysis of the video is done, this tab displays both the original video and the analyzed video. One difference between the mock-up and the application is that the text box with the name of the experiment is displayed next to the video uploader widget, so it is easier for the user to find it because the sidebar can be hidden.

Regarding the tabs that contain the charts, no mock-up was made, but rather the most related charts were grouped so that their values could be displayed at the same time and the user could analyze them easily.

The last two tabs require further analysis. These tabs display the audio, the speech of the user, and their corresponding emotion pie charts.

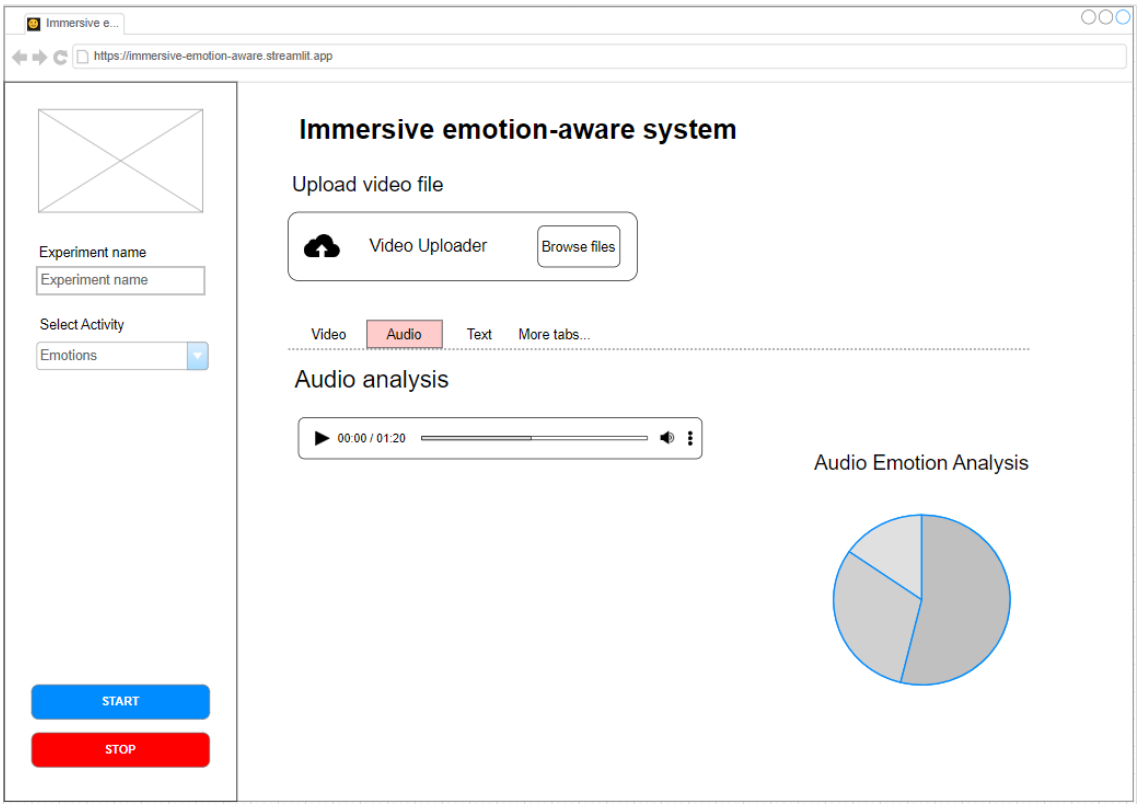

Figure 5.11: Mock-up upload videos page audio tab

As can be seen in Fig. 5.11, an audio player is displayed so that the user can listen to the audio extracted from the uploaded video. Furthermore, once audio emotions are extracted using the MFCCs, a pie chart is displayed on the right side of the page with the percentage of each of the emotions extracted from the audio.

The last tab, which is the text analysis, is similar to the audio one. Instead of displaying an audio player, this tab displays the text transcribed from the audio. The pie chart of emotions is also displayed, but this time NLP techniques are used to extract the emotions

from the text.

5.3.2.4 Videos gallery page

The last page is similar to the previous one. The tab structure is the same for both. The main difference is that when selecting this page, instead of a video uploader, a select widget is displayed. This select widget allows the user to choose from the videos he has analyzed and saved in his storage.

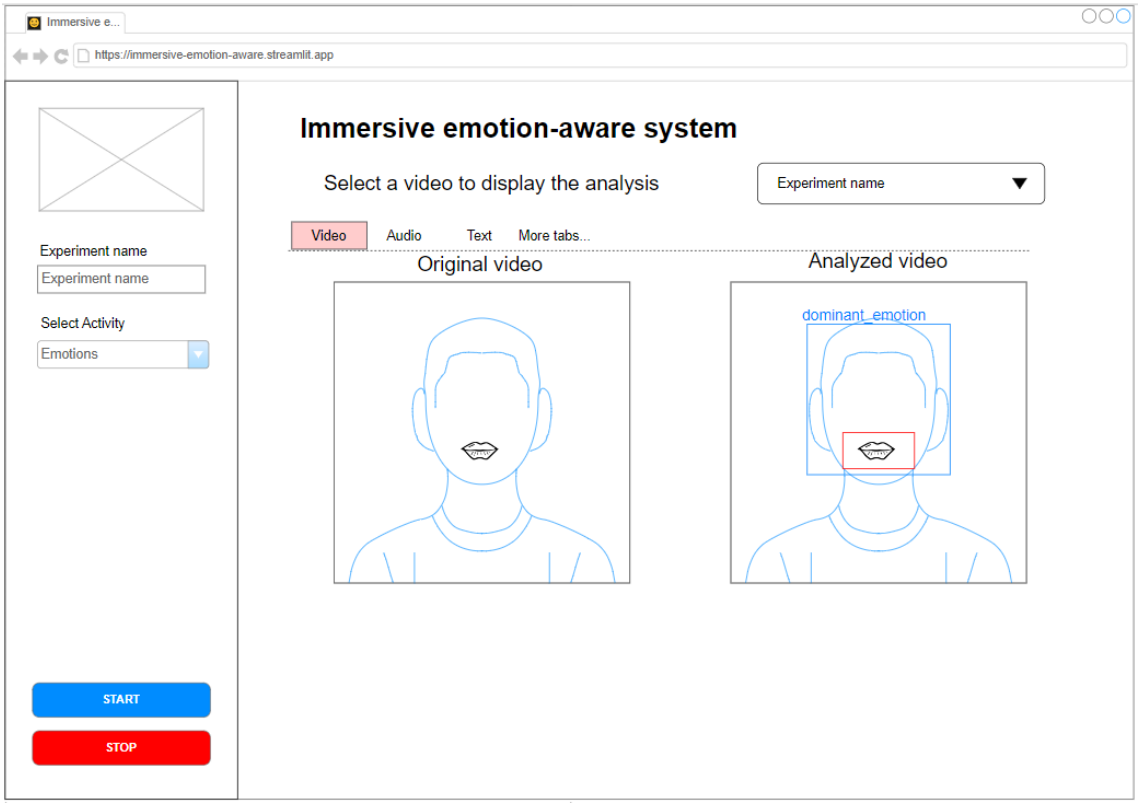

Figure 5.12: Mock-up videos gallery page

As can be seen in Fig. 5.12, the select box widget is displayed on the top right side of the page. The tab structure is the same as the one explained in Sect. 5.3.2.3. This page is very useful in case the user wants to review a video that has already been uploaded to the system without having to analyze it again, thus having to wait until all the frames of the video are analyzed, and the charts are displayed.

CHAPTER 6

Case study

In this chapter we discuss two types of experiment: Real-time emotional analysis and smile detection, which include face and smile detection using OpenCV and facial emotion recognition using DeepFace; and emotional analysis of a video, which analyzes emotions from the video images, from the extracted audio from the video, and from the text transcribed from audio, displaying charts to allow the user to analyze the results.

6.1 Introduction

Knowing that one of the objectives of this project is to design and develop a visualization layer to display the data collected by the emotion-aware system, transforming complex emotional insights derived from video, audio, or text data into visually comprehensible and actionable representations, the best way to demonstrate that this objective has been met is to conduct experiments using the full functionality of the system.

In the following sections, we will study the operation of the emotion-aware system application from the user's point of view. As mentioned above, the user can decide whether to use the real-time emotion recognition option or to upload a video and visualize its analysis. To provide an overall view of the application, both options will be tested in order to obtain concrete results. It is worth noting the possibility that the results obtained by the different sources of information that are analyzed in the case of video analysis may differ from each other if the message that is emitted and the user's facial expression does not show the same emotion.

The purpose of the experiments is to show how the system works from the moment the application is started. The first page that is always shown when running the application is the home page. Therefore, to not repeat this page in each experiment, Fig. 6.1 shows what the home page looks like in the application.

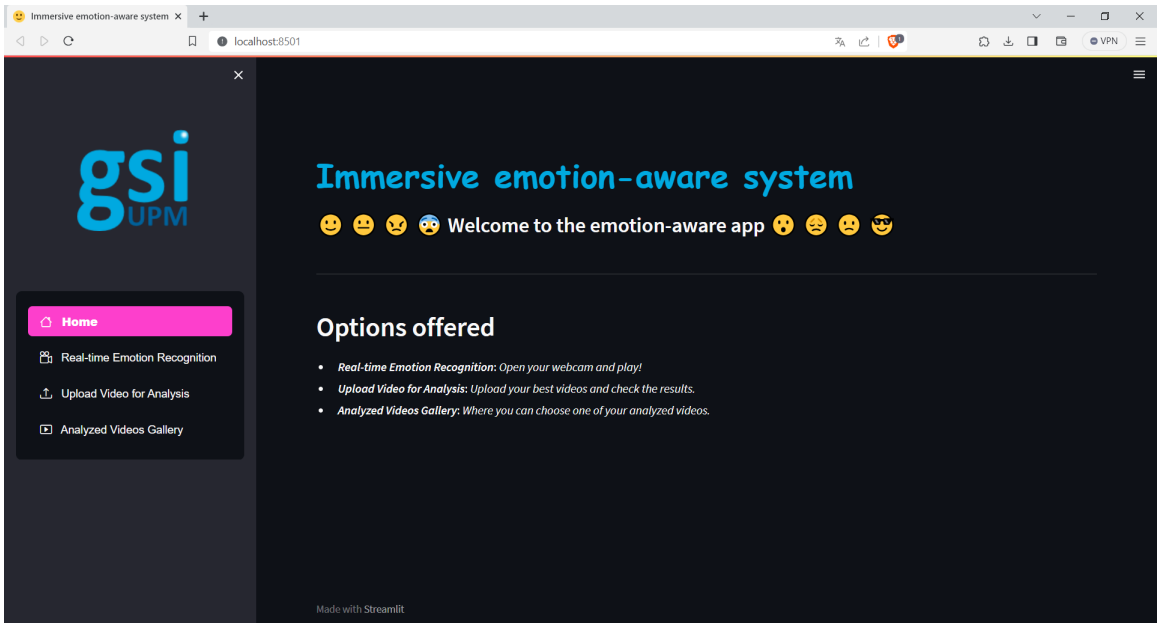

Figure 6.1: Home page

As can be seen in the figure above, the application has a sidebar where the Grupo de Sistemas Inteligentes (GSI) logo and an options menu, which allows the user to select the

desired option, are displayed. This sidebar is the same for all pages and can be collapsed by clicking the 'X' button.

The purpose of this page is to introduce the application and give a brief overview of the available features. Once the user is aware of the capabilities of the application, they can choose from the side menu which of the three options they would like to explore.

6.2 Description of the experiments

This section presents a brief description of the experiments that are performed to demonstrate the operation of the developed application.

- **Experiment 1:** This experiment consists of a test for the real-time emotion recognition option showing a happy emotion and a smile to verify that the system detects the correct emotion on the face and the smile.
- **Experiment 2:** This experiment consists of a test for the real-time emotion recognition option as in the first experiment. In this case, the emotion that the system must detect is neutrality without a smile.
- **Experiment 3:** The last experiment tests the video analysis option. A video will be uploaded to the system, and all the app tabs will be presented in order to show all the analysis done and the extracted video and speech.

6.3 Experiment 1: Real-time emotion recognition detecting happiness and smile

As has been said before, once the user has run the application, the home page is displayed, as can be seen in Fig. 6.1. From the home page, the user can select any of the three options from the option menu, excluding the home page itself.

For this experiment, the selected option is "Real-time Emotion Recognition". When the user clicks on this option, a table with the detected emotions ordered according to their percentage and the video from his webcam is displayed.

The video from the webcam is analyzed before being displayed in the application. If the system does not recognize any faces, the original video is displayed without analysis, as can be seen in Fig. 6.2

However, if the system detects a face, the video from the webcam is analyzed. This analysis consists of detecting all the faces in the video by placing a blue square around them. At the same time, the emotion of each of the detected faces is analyzed. A text with

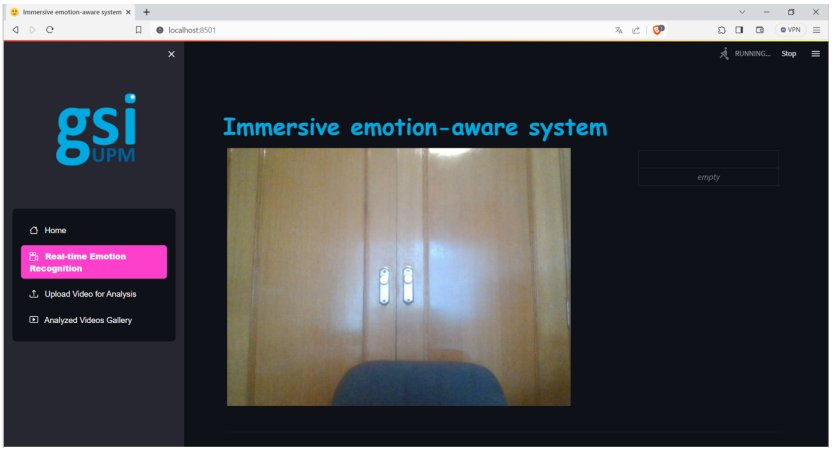

Figure 6.2: Real-time analysis without faces

the dominant emotion in each frame is placed in the upper left corner of the square, as well as a bar chart next to the face showing the percentage of each of the seven emotions that the system is able to detect. Finally, the system is also capable of detecting smiles in a similar way to face detection. In this case, a red rectangle appears around the detected smile on each face. All this analysis can be observed in Fig. 6.3.

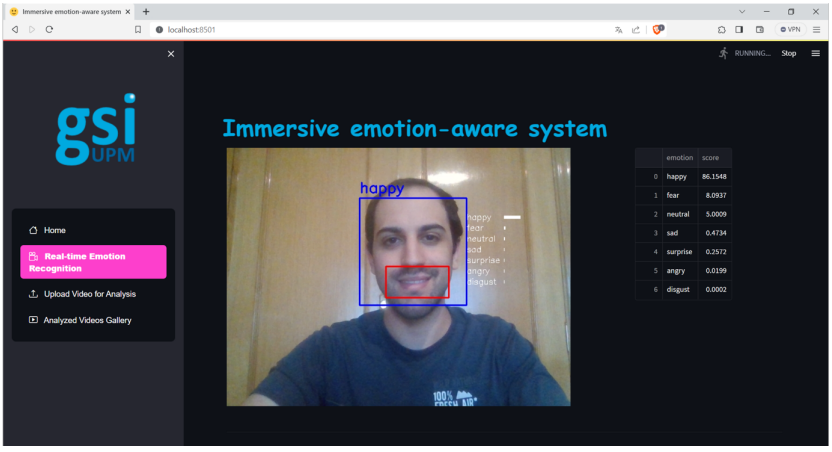

Figure 6.3: Real-time analysis happy emotion

The figure above shows the system response when the user is showing happiness and he is smiling.

6.4 Experiment 2: Real-time emotion recognition detecting neutrality and no smile

The second experiment is an addition to the first one. It shows how the system can identify all seven emotions. In this case, the neutral emotion was the dominant emotion on the user's face. Fig. 6.4 shows the analysis performed by the system.

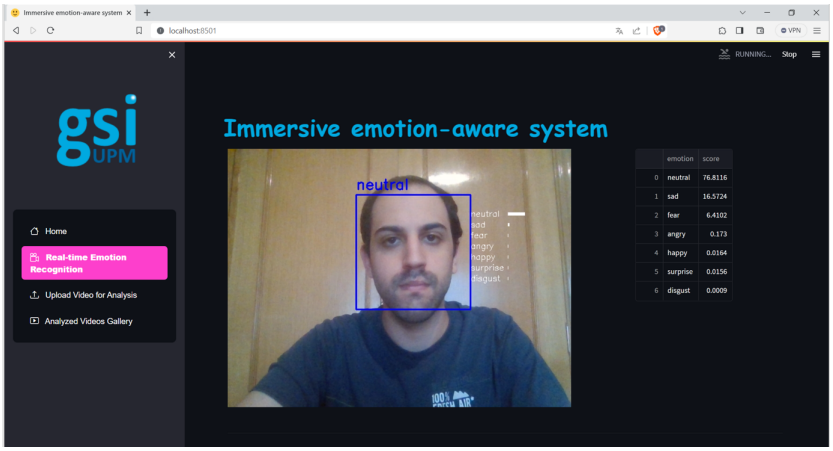

Figure 6.4: Real-time analysis neutral emotion

As can be seen in the figure above, the system detects the user's face. It displays his emotions, but because the user is not smiling, the system does not detect any smile, so the smile rectangle is not displayed as it was in the previous experiment.

6.5 Experiment 3: Video analysis

The third and last experiment shows the full functionality of the emotion-aware system. When the user selects the "Upload Video for Analysis" option on the sidebar, a file uploader that admits mp4, mov and avi formats is displayed. Next to it, a single-line text input widget is also displayed, where the user must write the name under which you want to save the performed experiment.

Once the user uploads a video for its analysis, the back end of the system starts to work on its processing and feature extraction. Meanwhile, in the front end of the system, a spinner widget is displayed so that the user can see that the application is still running and will soon get the video analysis. Fig. 6.5 shows the application when the user has just uploaded a video, and the system is analyzing it to extract its features.

When the system has finished analyzing the video and extracting its characteristics, the entire analysis is presented on the screen divided into tabs for easy navigation through the

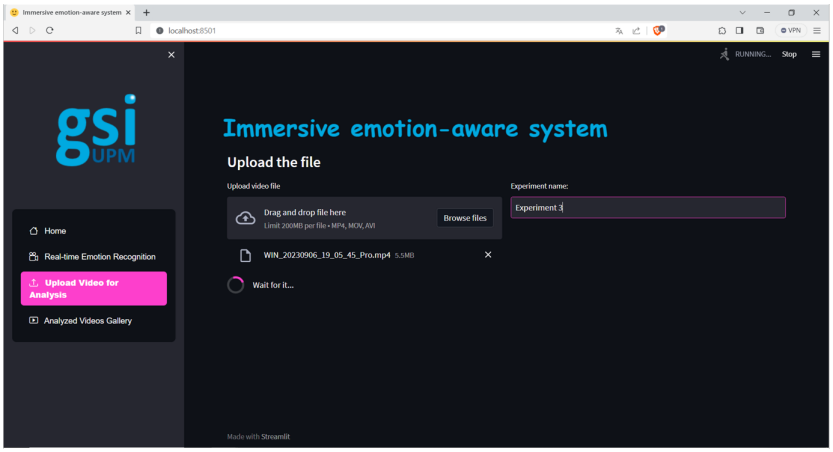

Figure 6.5: Upload page when a video is uploading

application. All data generated by the application are saved using the csv format. This allows the application to re-display the data on the last option offered.

Fig. 6.6 shows the app just after the video analysis is complete. The spinner widget disappears and all the tabs are revealed. As can be seen, the first tab is also displayed showing the original video and the analyzed video together.

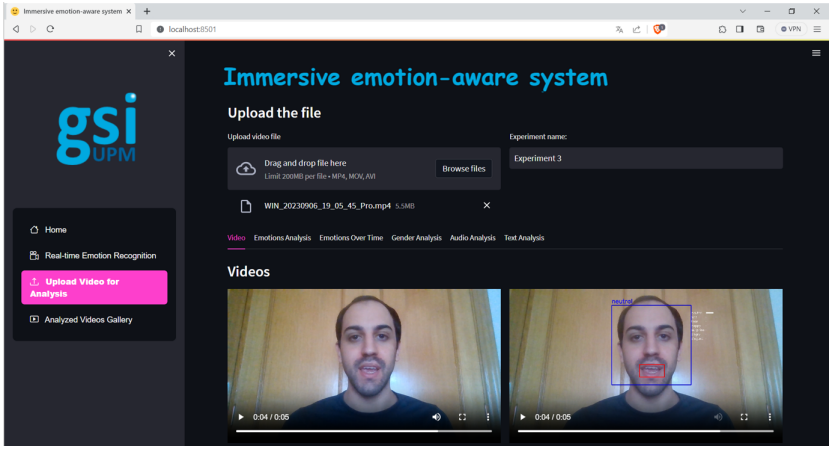

Figure 6.6: Upload page when a video is uploaded

In order to demonstrate that the data are correctly stored and the "Analyzed Videos Gallery" works properly, we are visualizing the results using this option. From now on, the option selected in the app is the one mentioned before, and all the figures are displayed from the stored data.

The "Video" tab is not displayed because it looks exactly the same as Fig. 6.6.

The second tab, which can be seen in Fig. 6.7, shows the emotions and smile analysis detected in the video. The first pie chart represents the emotions extracted from the analysis

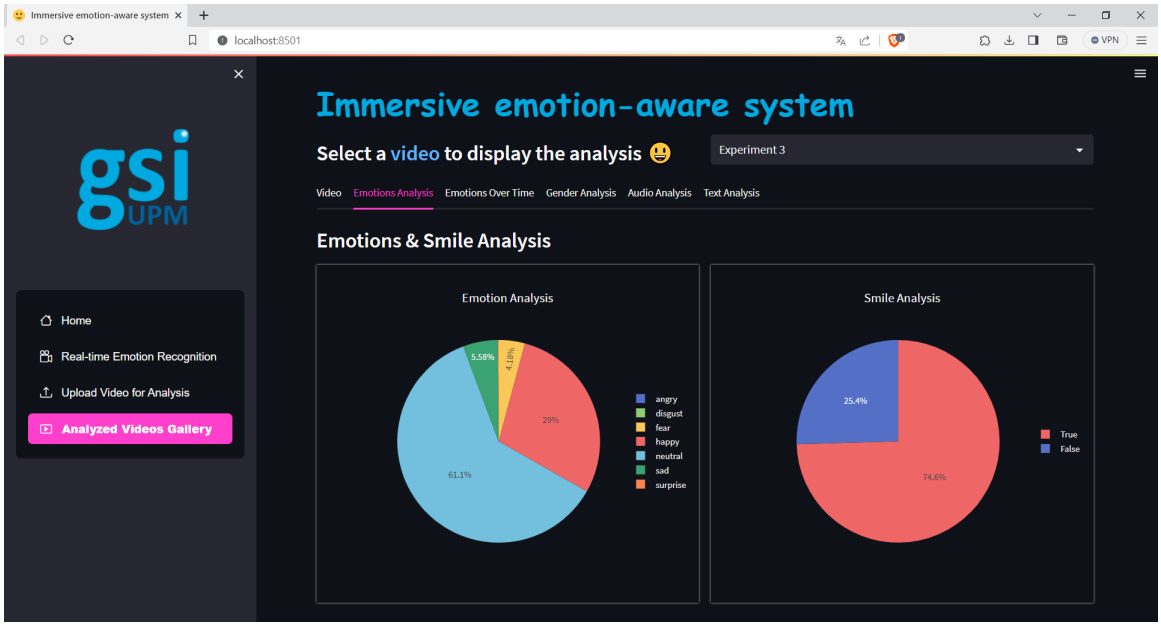

Figure 6.7: Analyzed videos gallery video emotions and smile tab

of the video. In this case, the video analysis showcases that the most common emotion is neutral and the second one is happy with a high percentage. The second pie chart represents the percentage of time in which a smile has been detected in the video compared to the time in which it has not been detected. In this experiment, the time the user is smiling triples the time the user is not smiling.

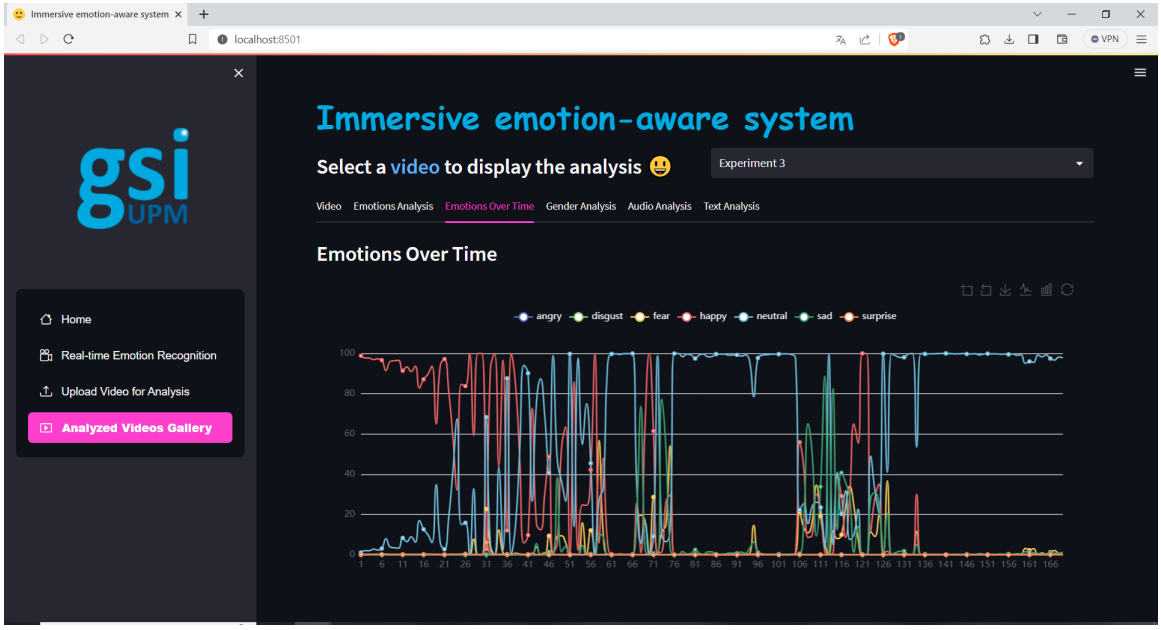

Figure 6.8: Analyzed videos gallery video emotions over time tab

The third tab, which can be seen in Fig. 6.8, shows a line chart in which all the emotional data collected from the video frame by frame are included. The horizontal axis represents the frame being analyzed, while the vertical axis represents the percentage of emotion detected. Each of the seven emotions that DeepFace can detect is represented by a different colored line, making it easy to distinguish one emotion from the others. In this experiment, it is easy to see that the dominant emotion at the beginning of the video is happy, while during the rest of the video the predominant emotion is neutral.

The fourth tab, which can be seen in Fig. 6.9, represents the gender analysis performed by DeepFace. In this tab, a pie chart is included showing in percentage the two genres that the library is able to detect, and next to it, an icon of a man or a woman is shown depending on which of the two genders has been detected. In this experiment, DeepFace determines that the person that appears in the video is a man with 95.5% accuracy and that is the reason why the system displays the man icon.

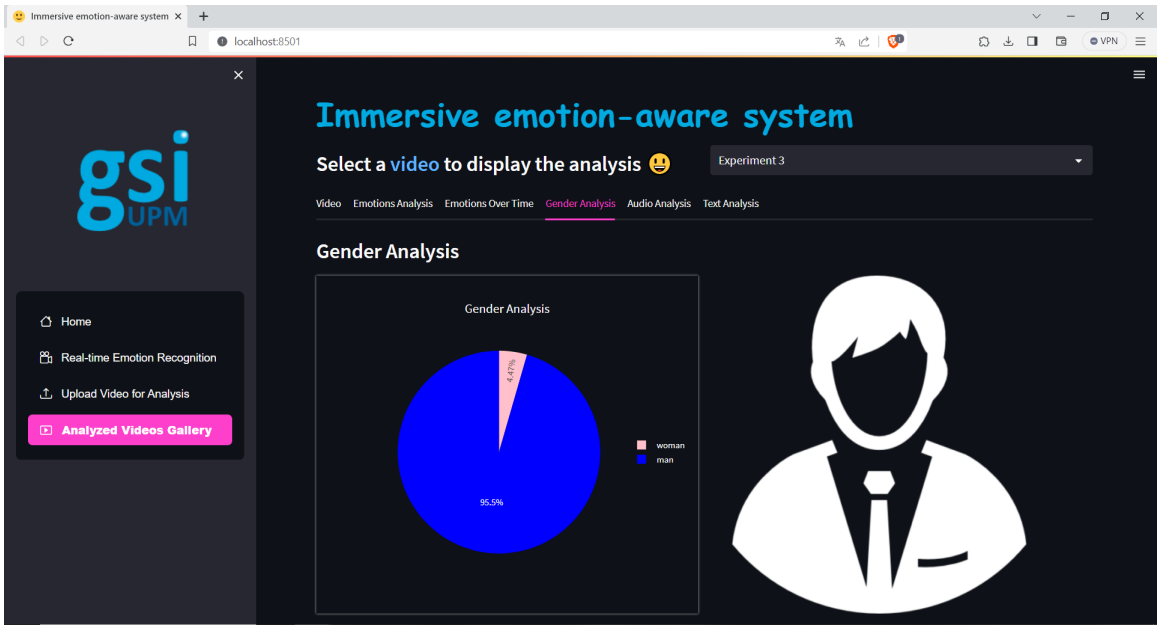

Figure 6.9: Analyzed videos gallery gender analysis tab

The fifth tab displays the audio and its analysis. As can be seen in Fig. 6.10, the audio extracted from the video is displayed in the first column, making it possible for the user to listen to it. Additionally, a pie chart of emotions is displayed next to the audio player. The audio analysis is performed using MFCCs as it was explained in the previous section. In this experiment, the dominant emotion extracted from the voice signal is sad with a really high percentage.

The last tab displays the text and its analysis. As can be seen in Fig. 6.11 the audio transcription generates the text represented in the first column. In this case, the extracted

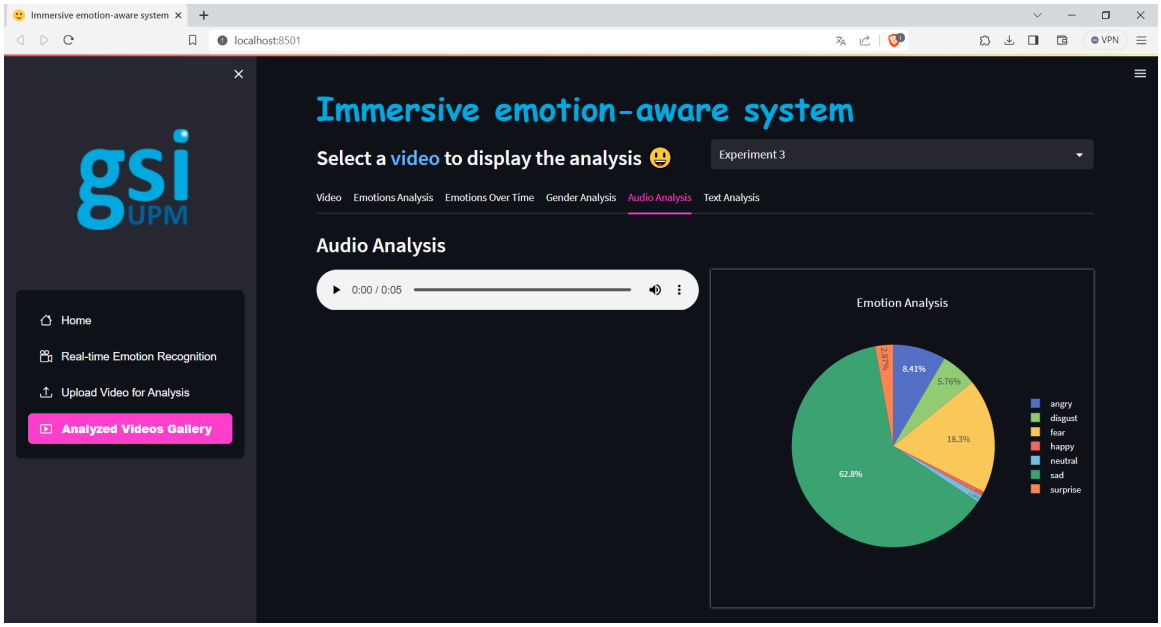

Figure 6.10: Analyzed videos gallery audio analysis tab

text from the audio is "it feels good to test this application". Before using some NLP techniques and a pre-trained model to predict emotions in texts, the result can be seen in the pie chart of the second column. The main emotion detected in this text is happiness with an accuracy of more than 70%.

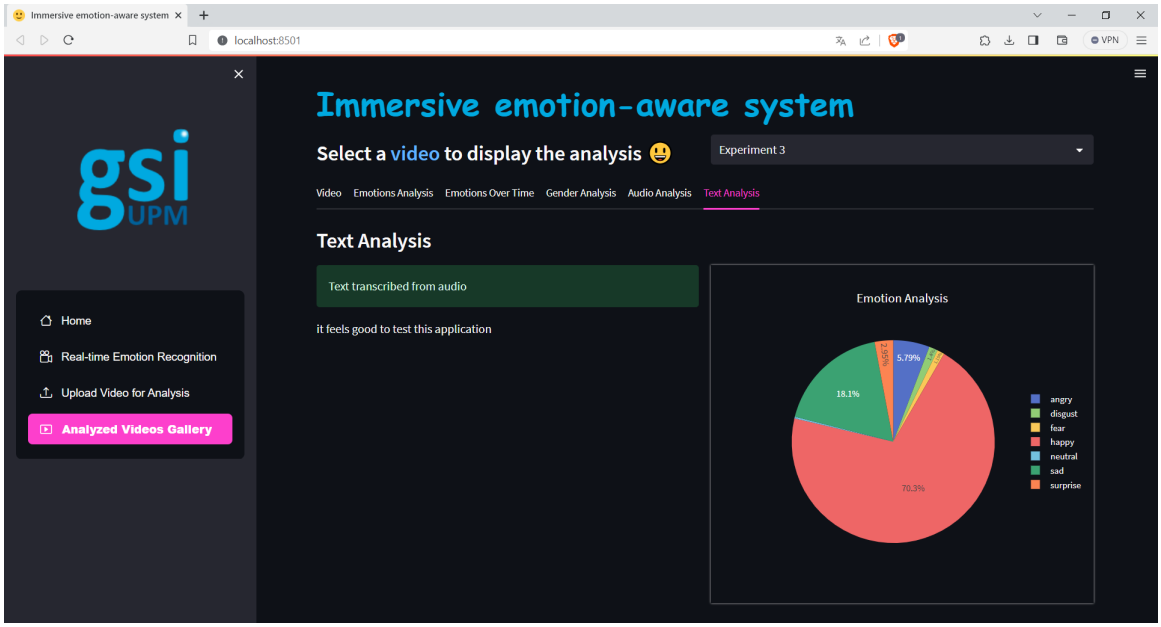

Figure 6.11: Analyzed videos gallery text analysis tab

CHAPTER 7

Conclusions

This chapter describes the conclusions of the master thesis along with the goals achieved and a discussion of future work.

7.1 Achieved goals

The goals achieved for this master thesis are the following.

- We have studied the DeepFace library to be able to extract and use all the functionalities of interest. DeepFace's functionalities cover a broad spectrum of facial analysis tasks, from basic face detection to advanced tasks like emotion analysis and facial recognition. Its versatility and compatibility with popular deep learning frameworks make it a valuable resource for developers, researchers, and organisations looking to harness the power of facial analysis.
- We have designed and developed a system capable of detecting emotions in videos, audio, and texts. Combining the data extracted from all sources provides a much more detailed and accurate analysis, making the usefulness of the system greater than just analysing one source.
- We have designed and developed a feature extraction layer to extract relevant features from raw data that can be used to characterise and quantify emotional content. All extracted features are stored in an orderly fashion so that they are accessible to the system itself or to others.
- We have designed and developed a visualisation layer to display the data collected by the emotion-aware system, transforming complex emotional insights derived from video, audio, or text data into visually comprehensible and actionable representations.
- We have tested the system to verify its correct operation. In the previous chapter, three experiments have been carried out in which all the options offered by the system have been tested, verifying that the system recovers all the information it offers.

7.2 Conclusion

To conclude this master thesis document, we recapitulate the implemented system. We have developed an emotion-aware system that can detect emotions in real-time or through video analysis. As we finish this exploration of the potential and consequences of this system, it is clear that this technology has tremendous potential and influence in many areas.

This system consists of a web interface that allows the user to select whether to analyse the emotions of a video from his own gallery or to analyse his emotions in real time using his webcam. The application uses Streamlit to represent the results and Python to deploy the server and execute the experiments. The architecture of the system is layered, where all layers are crucial for the system to work properly.

The development of this master thesis has been motivated mainly by the growing interest in artificial intelligence. An emotion-aware system has the potential to revolutionise the way we connect, communicate, and empathise with each other and could have a profound impact on our daily relationships.

Immersive emotion-aware systems have the potential to transform several aspects of our lives, from mental health support to education, entertainment, and customer service.As technology progresses and becomes more embedded in our daily routines, these systems offer the potential to create a more understanding and responsive environment that is tailored to our emotional health and preferences.

In addition, the increasing use of smart IoT devices adds to the use of data captured by emotion-aware systems. These smart devices are able to adapt the environment according to the user needs detected. The data generated by emotion-aware systems can provide valuable information. Emotion trends and patterns can be analysed to better understand human behaviour, preferences, and responses.

In conclusion, this master thesis has been designed to make it easy to deploy and use in a variety of applications. All data extracted from the sources are stored in the same format to facilitate a possible connection to another intelligent system. Additionally, the use of Streamlit facilitates a possible deployment in the cloud because Streamlit has a cloud of its own that provides easy management of the application and security.

7.3 Future work

In this section, the possible new features or improvements that could be made to the master thesis are explained.

- Improve the skills of the emotion recognition tools used, being able to detect even more emotions.
- Taking into account the practical application of the system and the environment in which it is deployed, create intelligent automation to improve the user experience.
- Add new features, such as long-term memory of the system, to be able to predict user emotions.
- Increase the options that the system provides, allowing the user to upload an audio file, transcribing it, and extracting both sources' emotions, or allowing the user to upload a daily message expressing his/her emotions.
- Combine the analysis carried out on each of the sources individually to offer the user

a unique result that shows the emotion that the system has detected through the analysis of all the data that have been offered to the system.

APPENDIX A

Impact of this project

This appendix reflects, quantitatively or qualitatively, on the possible impact (positive or negative, direct or indirect, current or future) and the responsibilities related to this thesis. It considers the social (Sect. A.1), economic (Sect. A.2) and environmental (Sect. A.3) impact, as well as ethical implications (Sect. A.4)

A.1 Social impact

The social impacts of an emotion-aware system are wide-ranging, affecting how people interact with technology, organizations, and one another.

Emotion-aware systems can be powerful tools in monitoring and managing mental health. They can detect signs of emotional distress early on, provide strategies to cope with it, and link people to support networks. This could help reduce the stigma associated with mental health and improve overall well-being.

In addition, emotion-aware systems have applications in healthcare and therapy. They can help therapists understand patient emotional responses during sessions, helping with diagnosis and treatment planning. In telehealth, these systems can provide emotional support and monitor patients' well-being remotely.

A.2 Economic impact

The potential economic impact of an emotion-aware system is considerable, with ramifications for a variety of industries both directly and indirectly. Such systems could be a catalyst for innovation, improve productivity, and open up new economic opportunities.

Emotion-aware systems can improve user experiences in a variety of applications, including e-Commerce, entertainment, and customer service. By providing users with more engaging and personalized interactions, these platforms can increase the amount of time and money users spend on them.

Businesses can take advantage of emotion-aware systems to better understand customer preferences. Market research companies can provide emotional analysis of customer reactions to ads and products. Advertisers can customize their campaigns based on emotional information, potentially increasing their advertising efficiency and profits.

A.3 Environmental impact

The project does not provide any significant change from an environmental point of view. The only subject involved is the energy consumption spent on the development and maintenance of the system. Emotion-aware systems often rely on cloud-based infrastructure and data centers to process and store large volumes of data. These data centers consume significant amounts of electricity to cool and run servers, which can have a negative environmental impact. Optimizing data center energy efficiency is critical to reducing the environmental footprint.

On the other hand, emotion-aware systems that enable remote work and virtual meetings can contribute to reducing emissions related to commuting and travel.

A.4 Ethical implications

As we embrace the potential of emotion-aware systems, ethical considerations come to the forefront. Privacy, consent, and data security must be carefully addressed. The responsible development and deployment of these systems require a commitment to ethical standards that protect users' rights and well-being.

Emotion-aware systems require the collection of personal data, including emotional data. It is crucial to obtain informed and explicit consent from individuals before collecting emotional information. Users must fully understand how their data will be used.

Finally, emotion recognition algorithms can exhibit bias, leading to inaccuracies or unfair treatment, especially among diverse demographic groups. Careful testing, monitoring, and mitigation of bias are essential to ensure fairness.

APPENDIX B

Economic budget

This appendix covers the costs that the master thesis requires. These costs include human resources (Sect. B.1), physical resources (Sect. B.2), licenses (Sect. B.3) and total budget (Sect. B.4).

B.1 Human resources

This section estimates the total cost of the human resources required to elaborate the master thesis. The time employed in the designing, developing and testing this system is taken into account.

The estimate of the working time used to carry out this master thesis is calculated on the basis of ECTS credits. The MUIT master thesis consists of 30 ECTS credits, each representing 25 to 30 hours of work. This gives a total of 900 hours of work. Considering that working hours for a person are 8 per business day and that a month has 22 business days, one engineer would spend between 5 and 6 months to elaborate the master thesis.

The average salary of a junior Telecommunications Engineer with the necessary qualifications and experience is 24,000 euros annually. Therefore, the human resources expenses for the master thesis will be 12,000 euros. The expenses associated with the maintenance of the application after the development process is finished are not included in the cost.

B.2 Physical resources

This master thesis provides a web application that could be run on a server or on a personal computer as it is launched on port 8501 by default, but it could be customized.

The personal computer on which the master thesis has been developed costs approximately 450 euros by mid-2023. The technical characteristics are as follows:

- **CPU:** Intel Core i7-1065G7 1.50 GHz
- **Memory:** 8 GB RAM
- **Hard Disk:** 512 GB SSD

If the web application is running in Streamlit Cloud, the resource limit is 1GB and it is free.

B.3 Licenses

The software used in the development of the master thesis is open-source software.

B.4 Total budget

The total budget for the development of the master thesis is approximately 12,450 euros, 12,000 euros from human resources, and 450 euros from physical resources.

Bibliography

[1] Rahul Agarwal. DeepFace: A python library for facial analysis and recognition. https://github.com/serengil/deepface. Accessed on 27/08/2023.

[2] Andrius Dzedzickis, Artūras Kaklauskas, and Vytautas Bucinskas. Human emotion recognition: Review of sensors and methods. *Sensors*, 20(3):592, 2020.

[3] Alberto Corredera Arbide. *Emotion-Aware Cyber-Physical Systems*. PhD thesis, Telecomunicacion, 2019.

[4] Reza GhasemAghaei, Ali Arya, and Robert Biddle. A dashboard for affective e-learning: Data visualization for monitoring online learner emotions. In *EdMedia+ Innovate Learning*, pages 1536–1543. Association for the Advancement of Computing in Education (AACE), 2016.

[5] Emotion Research Lab. Tutorial zoe. https://www.youtube.com/watch?v=1-3eGmN3kBg, 2019. Accessed on 30/08/2023.

[6] Affectiva. Affectiva emotion analytics dashboard. https://www.affectiva.com/product/affectiva-media-analytics-for-ad-testing/, 2023. Accessed on 30/08/2023.

[7] CyberMaryVer. Speech emotion recognition project. https://github.com/CyberMaryVer/speech-emotion-webapp, 2021. Accessed on 03/09/2023.

[8] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE communications surveys & tutorials*, 17(4):2347–2376, 2015.

[9] Streamlit Inc. Streamlit: The fastest way to create data apps. https://www.streamlit.io/. Accessed on 27/08/2023.

[10] Sefik Serengil. Deepface: The most popular open-source facial recognition library, 2021. *https://viso.ai/computer-vision/deepface*, 2021. Accessed on 27/08/2023.

[11] Yaoyao Zhong, Weihong Deng, Jiani Hu, Dongyue Zhao, Xian Li, and Dongchao Wen. Sface: Sigmoid-constrained hypersphere loss for robust face recognition. *IEEE Transactions on Image Processing*, 30:2587–2598, 2021.

[12] Itseez. Open source computer vision library. https://github.com/itseez/opencv, 2015. Accessed: 28/08/2023.

[13] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[14] Yilin Liu, Ruian Liu, Shengxiong Wang, Da Yan, Bo Peng, and Tong Zhang. Video face detection based on improved ssd model and target tracking algorithm. *Journal of Web Engineering*, 21(02):545–568, Jan. 2022.

[15] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, Oct 2016.

[16] Sefik Ilkin Serengil and Alper Ozpinar. Lightface: A hybrid deep face recognition framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 23–27. IEEE, 2020.

[17] Ziping Yu, Hongbo Huang, Weijun Chen, Yongxin Su, Yahui Liu, and Xiuying Wang. Yolo-facev2: A scale and occlusion aware face detector, 2022.

[18] Wei Wu, Hanyang Peng, and Shiqi Yu. Yunet: A tiny millisecond-level face detector. *Machine Intelligence Research*, pages 1–10, 2023.

[19] B. Thaman, T. Cao, and N. Caporusso. Face mask detection using mediapipe facemesh. In *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, pages 378–382, 2022.

[20] Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov, Karthik Raveendran, and Matthias Grundmann. Blazeface: Sub-millisecond neural face detection on mobile gpus, 2019.

[21] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[22] The pandas development team. pandas-dev/pandas: Pandas. `https://doi.org/10.5281/zenodo.3509134`, February 2020. Accessed on 07/09/2023.

[23] Google Speech Recognition API. `https://cloud.google.com/speech-to-text`. Accessed on 27/08/2023.

[24] Google Cloud. Google Cloud. `https://cloud.google.com/`. Accessed on 27/08/2023.

[25] Apache Software Foundation. Apache echarts. `https://echarts.apache.org`, 2017. Accessed on 31/08/2023.

[26] Plotly Technologies Inc. Collaborative data science. Plotly. `https://plot.ly`, 2015. Accessed on 07/09/2023.

[27] Paul Ekman et al. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16, 1999.

[28] Maja Pantic and Marian Stewart Bartlett. *Machine analysis of facial expressions*, volume 558. INTECH Open Access Publisher, 2007.

[29] Nicu Sebe, Ira Cohen, Theo Gevers, and Thomas S Huang. Multimodal approaches for emotion recognition: a survey. In *Internet Imaging VI*, volume 5670, pages 56–67. SPIE, 2005.

[30] Dhwani Mehta, Mohammad Faridul Haque Siddiqui, and Ahmad Y Javaid. Facial emotion recognition: A survey and real-world user experiences in mixed reality. *Sensors*, 18(2):416, 2018.

[31] Monica La Mura and Patrizia Lamberti. Human-machine interaction personalization: a review on gender and emotion recognition through speech analysis. In *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT*, pages 319–323. IEEE, 2020.

[32] Francisca Adoma Acheampong, Chen Wenyu, and Henry Nunoo-Mensah. Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, 2(7):e12189, 2020.

[33] Vidhi Mody and Vrushti Mody. Mental health monitoring system using artificial intelligence: A review. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pages 1–6, 2019.

[34] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR mental health*, 4(2):e7785, 2017.

[35] Joi L Moore, Camille Dickson-Deane, and Krista Galyen. e-learning, online learning, and distance learning environments: Are they the same? *The Internet and higher education*, 14(2):129–135, 2011.

[36] Enrique Sánchez Tolbaños. Design and Development of an Emotion-aware Learning Analytics system based on Machine Learning Techniques and Semantic Task Automation. Master thesis, Universidad Politécnica de Madrid, ETSI Telecomunicación, July 2019.

[37] Gaurav Sinha, Rahul Shahi, and Mani Shankar. Human computer interaction. In *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, pages 1–4. IEEE, 2010.

[38] David Feil-Seifer and Maja J Matarić. Socially assistive robotics. *IEEE Robotics & Automation Magazine*, 18(1):24–31, 2011.

[39] Debanjan Banerjee and Mayank Rai. Social isolation in covid-19: The impact of loneliness, 2020.

[40] K Aung, Mohd Said Nurumal, and WNSW Bukhari. Loneliness among elderly in nursing homes. *International Journal for Studies on Children, Women, Elderly And Disabled*, 2:72–8, 2017.

[41] Hojjat Abdollahi, Mohammad Mahoor, Rohola Zandie, Jarid Sewierski, and Sara Qualls. Artificial emotional intelligence in socially assistive robots for older adults: a pilot study. *IEEE Transactions on Affective Computing*, 2022.

[42] Adnan Shaout, Dominic Colella, and Selim Awad. Advanced driver assistance systems-past, present and future. In *2011 Seventh International Computer Engineering Conference (ICENCO'2011)*, pages 72–82. IEEE, 2011.

[43] Claudine Badue, Rânik Guidolini, Raphael Vivacqua Carneiro, Pedro Azevedo, Vinicius B Cardoso, Avelino Forechi, Luan Jesus, Rodrigo Berriel, Thiago M Paixao, Filipe Mutz, et al. Self-driving cars: A survey. *Expert Systems with Applications*, 165:113816, 2021.

[44] Joseph Fordham, Christopher Ball, et al. Framing mental health within digital games: an exploratory case study of hellblade. *JMIR mental health*, 6(4):e12432, 2019.

[45] Barry Kort, Rob Reilly, and Rosalind W Picard. An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In *Proceedings IEEE international conference on advanced learning technologies*, pages 43–46. IEEE, 2001.

[46] Renan Vinicius Aranha, André Biondi Casaes, and Fátima LS Nunes. Influence of environmental conditions in the performance of open-source software for facial expression recognition. In *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*, pages 1–10, 2020.

[47] Kai Lin, Fuzhen Xia, Chensi Li, Di Wang, and Iztok Humar. Emotion-aware system design for the battlefield environment. *Information Fusion*, 47:102–110, 2019.

[48] SannketNikam. Emotion detection in text using natural language processing. `https://github.com/SannketNikam/Emotion-Detection-in-Text`. Accessed on 07/09/2023.

[49] Shannon Bradshaw, Eoin Brazil, and Kristina Chodorow. *MongoDB: the definitive guide: powerful and scalable data storage*. O'Reilly Media, 2019.

[50] Ignacio Corcuera-Platas. Development of a Deep Learning Based Sentiment Analysis and Evaluation Service. Master thesis, Universidad Politécnica de Madrid, ETSI Telecomunicación, Madrid, January 2018.