

# Chapter 5

## Extraction of Emotion From Spectrograms: Approaches Based on CNN and LSTM

**Cecile Simo Tala**

*University of Ngaoundéré, Cameroon*

### **ABSTRACT**

*Speech is the main source of communication between humans and is an efficient way to exchange information around the world. Emotion recognition through speech is an active research field that plays a crucial role in applications. SER is used in several areas of life, more precisely in the security field for the detection of fraudulent conversations. A pre-processing step was done on audios in order to reduce the noise and to eliminate the silence in the set of audios. The authors applied two approaches of the deep learning namely the LSTM and CNN for this domain in order to decide of the approach which saw better with the problem. They transformed treated audios into spectrograms for the model of the CNN. Then they used the technique of the SVD on these images to extract the matrices of characteristics for the entries of the LSTM. The proposed models were trained on these data and then tested to predict emotions. They used two databases, RAVDESS and EMO-DB, for the evaluation of the approaches. The experimental results proved the effectiveness of the model.*

### **INTRODUCTION**

Societies are based on communication that responds to a set of rules allowing everyone to understand and be understood. These communications can be in voice, text, and gesture form. These different forms are intended to translate the thought,

DOI: 10.4018/978-1-6684-8127-1.ch005

### ***Extraction of Emotion From Spectrograms***

to represent it using a set of words chosen from a lexicon, gestures responding to a culture, and articulated sounds to form syllables, words, and sentences. This work is situated in the context of the recognition of emotions during telephone conversations between two people. In these types of interactions, the telephone is the only channel of communication. The conversation to be studied becomes important since our only channel of expression is the voice. It contains a multitude of information about the speaker such as his emotions, his age, his identity, his gender as well as the physiological disorders felt during oral expression. The extraction of this information has given rise to several areas of speech research, in particular the recognition of emotions from the voice. Speech is the most widespread means of exchanging information between human beings all over the world (Kwon, 2021), and attention should be paid to it. However, the most significant factor in human speech is emotion (Nardelli et al., 2015), which can be analyzed for judgments about humans and other expressions. Speech is the most widespread means of exchanging information between human beings all over the world (Kwon, 2021), and attention should be paid to it. However, the most significant factor in human speech is emotion Nardelli et al. (2015), which can be analyzed for judgments about humans and other expressions. Speech is the most widespread means of exchanging information between human beings all over the world (Kwon S., (2021), and attention should be paid to it. However, the most significant factor in human speech is emotion (Nardelli et al., 2015), which can be analyzed for judgments about humans and other expressions.

Emotions or emotional states are fundamental for humans insofar as they permeate humanity consciously and unconsciously in the most varied areas of life. They influence our perceptions, our behaviors, our mental states, and our daily activities such as communication, learning, and decision-making. The importance of emotions in the learning process has been known for a long time (L. Kerkeni et al, 2020), Nowadays the recognition of emotions in a speech signal is one of the most emerging areas of research and plays an important role in applications. in real time where researchers have developed methods to detect emotions from a voice signal (Kwon, 2021; (Mustaqeem and Kwon, 2019; Anvarjon et al., 2020) It paves the way for human-computer interaction (HMI) and plays an important role in many effective services such as call centers and tracking customer emotions to provide better services (Gupta, et al. (2007). In the medical field, speech-based diagnostic systems are developed to assess the extent of depression and distress (Rana et al. (2019), and some emotion recognition systems are designed for healthcare centers to monitor depression. state of the speaker for bipolar patients (Badshah et al., 2019; Wang, et al., 2015) There are so many other applications, such as multimedia search systems, (Roberts et al., 2012) forensic science (Vögel et al., 2018) smart car systems that have as their aim to improve their performance by using an effective emotion recognition system. More and more man is dependent on machines. There

29 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/extraction-of-emotion-from-spectrograms/348350](http://www.igi-global.com/chapter/extraction-of-emotion-from-spectrograms/348350)

## Related Content

---

### Significant Enhancement of Object Recognition Efficiency Using Human Cognition based Decision Clustering

Upendra Kumar and Tapobrata Lahiri (2013). *International Journal of Computer Vision and Image Processing* (pp. 1-15).

[www.irma-international.org/article/significant-enhancement-of-object-recognition-efficiency-using-human-cognition-based-decision-clustering/103955](http://www.irma-international.org/article/significant-enhancement-of-object-recognition-efficiency-using-human-cognition-based-decision-clustering/103955)

### Personalized Content-Based Image Retrieval

Iker Gondra (2009). *Artificial Intelligence for Maximizing Content Based Image Retrieval* (pp. 261-287).

[www.irma-international.org/chapter/personalized-content-based-image-retrieval/4160](http://www.irma-international.org/chapter/personalized-content-based-image-retrieval/4160)

### Classifying Behaviours in Videos with Recurrent Neural Networks

Javier Abellan-Abenza, Alberto Garcia-Garcia, Sergiu Oprea, David Ivorra-Piqueres and Jose Garcia-Rodriguez (2017). *International Journal of Computer Vision and Image Processing* (pp. 1-15).

[www.irma-international.org/article/classifying-behaviours-in-videos-with-recurrent-neural-networks/195006](http://www.irma-international.org/article/classifying-behaviours-in-videos-with-recurrent-neural-networks/195006)

### Construction of 3D Triangles on Dupin Cyclides

Bertrand Belbis, Lionel Garnier and Sebti Fofou (2011). *International Journal of Computer Vision and Image Processing* (pp. 42-57).

[www.irma-international.org/article/construction-triangles-dupin-cyclides/55099](http://www.irma-international.org/article/construction-triangles-dupin-cyclides/55099)

### Face Match for Family Reunification: Real-World Face Image Retrieval

Eugene Borovikov, Szilard Vajda and Michael Gill (2017). *International Journal of Computer Vision and Image Processing* (pp. 19-35).

[www.irma-international.org/article/face-match-for-family-reunification/183658](http://www.irma-international.org/article/face-match-for-family-reunification/183658)