Emotion Recognition in Moroccan Arabic:
Leveraging Deep Learning for Multilingual Sentiment Analysis

Mhamed-Amine Soumiaa
http://orcid.org/0000-0003-2152-2185
National School of Applied Sciences, Hassan First University, Berrechid, Morocco

ABSTRACT

This article explores speech emotion recognition (SER) for Moroccan Arabic (Darija) using deep learning. A dataset of 2,000 labeled audio samples across five emotions (happy, neutral, sad, angry, fearful) was collected from 80 speakers. Darija's linguistic diversity and regional variation pose challenges for emotion detection. After audio preprocessing and feature extraction (MFCCs, HNR, ZCR, F0, etc.), a ResNet-152 model was trained, achieving 90.12% accuracy. Angry and fearful emotions showed the highest recognition rates. This work highlights the potential of deep learning in under-resourced dialects and paves the way for emotion-aware Arabic applications in health, education, and virtual assistants.

INTRODUCTION TO SPEECH EMOTION RECOGNITION (SER)

Speech Emotion Recognition (SER) is a field of research that focuses on identifying and classifying emotions from spoken language. This area of study has gained significant importance in recent years due to its potential applications in various industries, ranging from healthcare to human-computer interaction (HCI). Emotion detection in speech helps bridge the gap between machines and humans, enabling more natural and empathetic interactions. By analyzing speech features such as pitch, tone, tempo, and rhythm, SER systems can infer emotional states like happiness, anger, sadness, and fear, providing a deeper understanding of human emotional expression.

In the context of artificial intelligence (AI), SER is seen as a crucial step towards developing more intelligent systems that can engage in human-like conversations. These systems are designed to be sensitive to emotional cues, enhancing user experiences by providing responses tailored not only to the content of the conversation but also to its emotional context. For example, a virtual assistant that can detect frustration in a user's voice may offer more supportive and helpful suggestions, rather than simply providing generic responses.

DOI: 10.4018/408156

This chapter explores the application of SER techniques in Moroccan Arabic dialect, known as Darija. Although the technology has been widely explored in other languages, including English and Mandarin, there has been relatively little focus on Arabic dialects, particularly in the context of emotional speech analysis. This work addresses the gap by presenting a comprehensive study that utilizes deep learning techniques to classify emotions in speech recorded in Moroccan Arabic. By developing a robust dataset and leveraging modern AI techniques, this research makes significant contributions to the emerging field of sentiment analysis within the Arabic-speaking world.

IMPORTANCE OF STUDYING EMOTIONS IN SPEECH

The recognition of emotions in speech is a critical component in understanding human communication. Emotions are deeply embedded in the way we communicate, influencing the tone, pitch, and rhythm of our speech. This makes emotion recognition essential not only for improving human-computer interactions but also for a variety of applications, including health diagnostics, customer service, and even social robotics.

In healthcare, for example, emotion recognition can play a significant role in the diagnosis and treatment of mental health disorders. Speech patterns are often affected by emotional states such as depression, anxiety, and stress, and recognizing these patterns can aid in identifying patients who need immediate attention or intervention. Similarly, in educational settings, AI systems capable of detecting emotions can provide real-time feedback to students, identifying when they are frustrated or confused, and offering assistance accordingly.

Furthermore, in customer service, where automated systems interact with clients through voice-based interfaces, emotion recognition allows for more responsive and empathetic interactions. For instance, an automated system that detects frustration in a customer's voice could prompt the system to escalate the conversation to a human agent, ensuring that customer concerns are addressed more efficiently.

Recognizing emotions in speech also helps in improving the naturalness and relevance of interactions in artificial intelligence. Traditional AI systems often respond to the literal content of a conversation, but by considering emotional cues, these systems can respond in a way that feels more personal and understanding. Emotion-sensitive systems can interpret not only what is being said, but also how it is being said, enriching the communication experience for users.

However, achieving accurate emotion recognition is a complex task, particularly when dealing with the wide variety of emotional expressions present across different languages, dialects, and cultures. This complexity is compounded by the diverse range of emotions that can be expressed in speech. Understanding the subtle variations in tone, pitch, and tempo that signify different emotions requires a deep understanding of both the linguistic features of speech and the emotional context in which it is spoken.

In the case of Moroccan Arabic, this challenge is particularly pronounced. The dialect, known as Darija, is rich in variations, influenced by Arabic, French, Spanish, and Amazigh, among other languages. The nuances of emotional expression in Darija are unique and often differ from those found in other Arabic dialects. As such, a dedicated effort to build a dataset of emotional speech in this dialect is essential for advancing sentiment analysis in the region and ensuring that AI systems can accurately interpret the emotional content of Moroccan Arabic speech.