Arab American University
Faculty of Graduate Studies
Department of Media and Communication
Master Program in Integrated Digital Media

Exploring the Semiotic Role of Audience Empathy in Recognizing Emotions in Levantine-Arabic Vocal Media Content

Yahya Saed Yaqub Abu Asab
202317108

Supervision Committee:

Dr. Naheda Makhadmeh

Dr. Bashar Tahayna

Dr. Abdallah Adawy

This Thesis Was Submitted in Partial Fulfilment of the Requirements for the Master Degree in Integrated Digital Media

Palestine, 10/2025

© Arab American University. All rights reserved.

Arab American University
Faculty of Graduate Studies
Department of Media and Communication
Master Program in Integrated Digital Media

Thesis Approval

Exploring the Semiotic Role of Audience Empathy in Recognizing Emotions in Levantine-Arabic Vocal Media Content

Yahya Saed Yaqub Abu Asab
202317108

This thesis was defended successfully on 1/10/2025 and approved by:

Thesis Committee Members:

	Name	Title	Signature
1.	Dr. Naheda Makhadmeh	Main Supervisor	
2.	Dr. Bashar Tahayna	Member of Supervision Committee	
3.	Dr. Abdallah Adawy	Member of Supervision Committee	

Palestine, 10/2025

Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is substantially my own work and has not been submitted for any other degree at the Arab American University or any other institution.

Student Name: Yahya Saed Yaqub Abu Asab
Student ID: 202317108
Signature: 
Date of Submitting the Final Version of the Thesis: 6/11/2025

Acknowledgments

I express my sincere thanks to all those who contributed with their support and guidance in my endeavor of completing this master thesis.
I am also deeply grateful to all of those who assisted me in a number of ways through this research. Special thanks go to my family for all their love, patience, and encouragement, without which I would not have reached this point today.

First, I must thank my supervisor, Dr. Naheda Makhadmeh, for her unwavering support, commitment, and insightful guidance, with encouragement and suggestions that were invaluable for the realization of this work.

My thanks are extended to the Arab American University for the facilities, and academic environment afforded for my research. My special thanks go to the faculty, the department, and the research facilities for their continual support. Special thanks to Dr. Hussain Al-Ahmad and Dr. Shadi Abu Ayash for their advice and assistance were invaluable in guiding me throughout this entire academic endeavor.
I also thank the psychologist Dr. Iman Ghanem from the University of Jordan for her participation as an emotion assessment expert in this research.

I also particularly mention the Dean of Graduate Studies, Dr. Nouar Qutob for providing assistance and exceptions. Last but not least, I would like to thank the program coordinator and the entire committee for their dedication and support through which I was able to accomplish this thesis to its fullest standards.

Exploring the semiotic role of audience empathy in recognizing emotions in levantine-arabic vocal media content

Yahya Saed Yaqub Abu Asab
Dr. Naheda Makhadmeh
Dr. Bashar Tahayna
Dr. Abdallah Adawy

Abstract

This thesis examines the vocal media reception from an interdisciplinary perspective, which focus on the semiotic role of empathy in recognizing emotions from Levantine-Arabic vocal media content. the study was done by analyzing how perceived vocal cues mediate the listener's interpretive process in order to interpret the speaker emotions, the study frames empathy not only as a psychological trait but also as a semiotic lens that shapes the micro-level of audience reception. The study conducted in Palestine during the spring semester of 2025, the researcher employed a descriptive, mixed-method with a cross-sectional design. The sample in this study included 384 participants of Palestinians universities students aged 18–45, recruited online. The data were collected through a digital questionnaire incorporating authentic audio clips from the Nafs podcast from YouTube (which published on February 2023–June 2024). Research instruments combined the Interpersonal Reactivity Index (IRI), an emotion recognition task, and open-ended questions, which validated through expert review and statistical testing.

The study findings highlight that listeners embrace multiple paralinguistic cues in a probabilistic and interpretative manner influenced by individual traits and socio-cultural familiarity, and that higher empathy levels especially the two empathy scales of IRI, empathic concern and perspective-taking are significantly enhance recognition accuracy for the listener. Gender-based differences were also observed, and found that women participants outperforming men in voice emotion recognition process.

This study contributes a novel theoretical framework for media studies scholars to examining audience reception of vocal media content at the micro level, by combining insights from semiotics, psychology, and communication studies. furthermore, this study underscores the need for culturally grounded Arabic emotion datasets and the integration of reception-focused perspectives into AI models and affective computing. More broadly, the research advances the interdisciplinary study of media reception by emphasizing how empathy mediates audience interpretation in vocal communication contexts.

Keywords: Media reception, Vocal media content, Empathy, Semiotics, Emotion recognition.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter One : Introduction

1.1 Introduction

Human innate abilities, such as perception, learning processes, and the capability to adapt to the environment, are often described as part of human intelligent behavior (Ekman, 1982). With the great development in artificial intelligence, the explosion of data, and the availability of computers with high computational power, interest has increased in using artificial intelligence in an attempt to simulate humans by learning, especially recognizing human emotions.

Alhuzali et al. (2018) found that AI systems that trained on Western languages datasets, may not be optimal when it comes to diverse cultural expressions of emotions. The study revealed that there is a significant difference between Western and Arabic emotional expression patterns, which affects these systems' error rates as high as 30%. Although, deep learning processes have facilitated the enhanced understanding of emotion recognition where it can capture the slightest nuance of emotional expression in language; however, there exist some challenges such as reducing cultural biases or addressing ethical issues for its application in judgement (Klaylat et al., 2018).

There is still room for improvement as far as Arabic speech emotion recognition research is still in early stage. Studies in this field often rely on a small, acted datasets. For example, Klaylat et al. (2018) has analyzed Arabic TV shows, and limited the study only on three emotions, happiness, anger, and surprise. But the limited emotional range and lack of gender-specific analysis suggest a room for improvement. Similarly, Meddeb et al. (2016) has achieved strong results using machine learning models, but their research also lack for gender differences.

This highlights the need for culturally adapted approaches in emotion recognition research, particularly for understudied dialects like Levantine Arabic. By focusing on the semiotic role of empathy and vocal cues in Levantine Arabic, this study contributes to more inclusive AI-driven media technologies.

In today's rapidly evolving media landscape, emotion recognition technologies offer an immense potential for advancing how users interacts with digital media, especially in understanding and responding to users' emotional needs. Despite extensive research in

this field across various languages and cultural settings, a significant gap remains in comprehending how emotions are conveyed and interpreted in Arabic vocal communication.

Emotion recognition is considered as a vital skill to achieve an effective communication, particularly in vocal conversations, where subtle vocal elements such as pitch, tone, and speed help conveying complex emotional states (Abdel-Hamid, 2020). The richness of the Arabic language with a diverse phonetic and dialectal variations, presents a unique challenge with a lot of opportunities. Existing studies that focus on Arabic emotion recognition are often constrained by limited datasets, and mostly depend on actors portraying the emotion they are required to convey to the listener, which makes it lack of authenticity that can found in raw emotions that exist in real life communication (Ouali & El Garouani, 2024).

The theoretical background of this research depends on Brunswik's lens model of Social Judgment theory, which will help us to investigate how listeners decode emotional signals that they perceive during a vocal communication (Scherer, 2003). The Brunswik's lens model serve as a framework to examine how vocal elements are used as a channels for transmitting emotions. Simultaneously, by measuring the level of listener's empathy, in order to understand how empathy play as a "sign" in vocal interactions, therefore enhancing the listener's ability to recognize emotions (Kolmogorova et al., 2021).

In order to deepen understanding of the relationship between vocal cues, empathy and emotion recognition, this thesis explores the semiotics role of empathy in Levantine-Arabic vocal communication, with a focus on how vocal cues or elements can guide listeners to recognize emotions and what is the role of empathy in this process.

1.2 Research Problem

According to Pittam (1994), Voice communication is important for interaction between humans, particularly in the media where emotions play a key role in shaping audience perception and engagement with the content. It is crucial to understand that as well words, emotions are also conveyed using vocal elements. Obviously, how listeners understand what they hear is influenced by how the speaker's tone tells about his feelings. This highlights the importance of studying emotion perception from vocal communication in

order to achieve an effective communication strategies enhance media content production quality, and introduce an enhancement into areas like voice recognition technologies.

Studying voice and emotion in Levantine-Arabic, a dialect spoken across the culturally vibrant Levant region (Palestine, Jordan, Syria, and Lebanon), presents unique challenges, and opportunities. In academic research, this dialect is narrowly explored, particularly in the areas of semiotics and media studies. Most existing studies focus on major languages like English, shaping prevailing theories on emotion recognition and semiotics. While some research has touched on emotion in Arabic, there's very little specifically focused on how emotions are communicated in Levantine Arabic within media setting.

Despite the popularity of studying emotion recognition in various cultures and languages, few studies have concentrated on Arabic vocal communication. Additionally, Most of the previous studies rely on elicited or acted emotions, that cannot fully capture natural emotions from speech that used in everyday life (Abdel-Hamid, 2020). However, the use of controlled data has led to doubts regarding their applicability and accuracy. Furthermore, there are no comprehensive databases for Arabic spoken language which further limits the effectiveness of any Arabic SER systems (Ouali & El Garouani, 2024) and (Dalal & Rayhana, 2020).

As regards the emotion recognition in Arabic language, it becomes more difficult due to specific phonetic and prosodic features of language itself, which are combined with a great number of its dialects like Levantine-Arabic dialect. Arabic has some particular sounds not always represented in datasets used, so that makes it hard to create algorithms recognizing many voice emotions accurate enough (Ouali & El Garouani, 2024).

Majority of the SER systems that are available today have been designed with Reductive focus on a few emotions, and thereby neglecting the complete range of emotion, which limits their utility in the field of human-computer interaction and affective computing (Dalal & Rayhana, 2020).

Furthermore, there is also a lack of representative emotional datasets for Levantine Arabic, which poses a significant challenge for both academic research and practical applications. For instance, Haj Ahmed et al. (2024) highlights that most existing datasets

of Arabic emotion recognition, rely only on acted emotions, which lack the authenticity of raw real life vocal communications.

This research gap matters because it limits understanding of how cultural and linguistic factors affect emotion recognition in spoken language, particularly in non-Western contexts. Without in-depth studies on Levantine Arabic voice communication, scholars are missing out insights that could inform culturally relevant media theories and adaptive communication technologies. This study, therefore, aims to fill this gap by investigating the semiotic dimensions of emotion in Levantine Arabic voice communication, offering new knowledge that could expand the fields of media and communication studies.

In short, this research addresses the need for a deeper understanding of how emotions are encoded and interpreted in Levantine Arabic voice communication, especially in media contexts. By examining these semiotic processes, the study hopes to contribute valuable insights to semiotics, media studies, and the broader field of emotion recognition.

1.3 Research Objectives

This study seeks to:

1. Investigate the role of vocal cues in emotions recognition.
2. Examine how empathy functions semiotically in Levantine-Arabic vocal media content.
3. Use the lens model of social judgment theory as a framework to analyze and understand the cognitive and perceptual processes involved in emotion recognition from vocal cues, with a specific focus on Arabic vocal communication.
4. Identify the impact of listener gender on emotion recognition accuracy.

1.4 Research Questions

This study seeks to answer the following questions and hypothesis:

RQ1. How do the following vocal cues influence emotion recognition?

a. Tone
b. Pitch
c. Volume

d. Speed
e. Intonation

RQ2. What is the semiotic function of empathy in the process of recognizing emotions from Levantine-Arabic vocal media content?

RQ3. How does the lens model of social judgment theory explain the relationship between vocal cues and emotion recognition?

RQ4. Is there a significant difference between women and men in emotion recognition accuracy?

1.5 Research Significance

This study takes a fresh look at vocal Levantine Arabic communication, to understand how emotions are received and interpreted by the listener, thereby exploring alternate paths through which the dialect may communicate meanings or feelings. Much of semiotic research has concentrated on other languages but not much has been on Arabic language, specifically the Levantine-Arabic dialect. So, such study could be present a non-Western perspective, that endorsing media and communication scholars to espouse more all-encompassing theories that take into account cultural subtleties differences as well linguistic variety.

By integrating various semiotic elements, media and communication studies and emotion recognition, this research operates on an interdisciplinary basis that seeks to investigate further the role of emotional expression in voice-based communication processes. This way, this study brings new depth to understand how scholars think about emotions in media.

This study would significantly contribute to enhanced human-computer interaction techniques, advanced media technologies, effective communication; empathy enhancement; promotion of emotional wellness in future broadcasting channels, e.g., films, TV programs/shows. It paves way for development of more empathetic media through focusing on vocal cues as well as how emotions are recognized by audience. These media systems could change responses to users according to their moods, leading to much rich personalized media experiences.

Furthermore, understanding how emotions can be expressed by vocal cues would help media professionals like broadcasters, podcasters, and content creators to communicate more effectively, by shaping their messages to be more adaptable by their audience.

From a technical perspective, this research addresses critical gaps in emotions recognition in Arabic vocal communication, by contributing to the development of a representative datasets and culturally sensitive AI models. For example, by analyzing acoustic features such as fundamental frequency (F0) and intensity patterns, this study provides a base for enhancing AI-driven emotion recognition systems in Arabic-speaking contexts. Such improvements would increase representation across different demographics as well as making media technologies smarter than they are today.

1.6 Scope and Delimitations

The research places particular focus on the Arabic language, known for its unique phonetic features, including its complex vowel system and the range of dialects spoken across the Arab world. In addition, the study will address gender differences in emotion recognition, Kring and Gordon (1998) reported that gender differences are intimately related to emotional and cognitive functions. Given that previous studies indicates that women tend to outperform men in tasks that requiring emotion recognition, particularly when subtle emotional expressions are involved. This difference is especially clear in vocal tasks, where acoustic features like pitch typically play a significant role in detecting emotions by women (Dair et al., 2021). However, these gender differences are not uniform across studies and may be influenced by cultural and linguistic factors.

This research is defined by several boundaries:

1- The study focus exclusively on vocal communication, excluding other forms of emotional expression such as facial expressions or body language. Although these visual cues could be helpful for enhance the process of emotions recognition, this study narrows the scope to vocal communication, which is crucial in context where visual cues are unavailable.
2- The geographical scope of this study is centering only on the Levant region (Palestine, Jordan, Syria and Lebanon), specifically Palestine. While emotion

recognition is a globally relevant issue, this study concentrates on Arabic due to the notable lack of research in this area.

3- The study will focus on data collected within a specific timeframe, episodes of Nafs podcast that was premiered from February 2023 to June 2024, to ensure the relevance and timeliness of the findings.
4- The study will focus on adult participants, aged 18-40, excluding children's vocal expressions of emotion from its scope. While the research will examine gender differences, it will not delve into all demographic factors, such as variations related to age or socio-economic status, although these could also influence how emotions are recognized.
5- Furthermore, this study depends on a single dataset gathered from the Nafs podcast, which may not fully represent the diversity of Levantine Arabic emotional expressions. While this dataset provides authentic and natural vocal data, future research could expand the scope to insure including other media sources and dialects within the Levantine region. This would enhance the ability to generalize the findings and to contribute to a more comprehensive understanding of emotion recognition in Arabic vocal communication.

1.7 Research Terms

Emotion: is defined as: " a conscious mental reaction (such as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body" (Emotion, n.d.). Emotions also defined in many meanings and aspects, as found in various domains of literature.

In media studies, emotions can drive audience to engage with content, conveying cultural narratives, and shaping resonance media content (Frijda, 1986) and (Nabi, The case for emphasizing discrete emotions in communication research, 2010), In addition, representation and transition of emotions can be analyzed through vocal and visual content (Deshmukh et al., 2023). In this study, the focus will be only on the perception of emotional states through vocal elements, and how listeners of Levantine-Arabic vocal media content, can recognize emotions.

Emotion Recognition: is defined as a method to identify, interpret and understanding emotions, depending on various signals or symbols, like facial expressions, voice, text, or another biometric signals. This process often used by machine learning systems with deep learning techniques in order to classify emotions into categories like happiness, sadness, anger, and fear (Poria et al., 2017). This interdisciplinary field intersects with communication and media studies by enabling the analysis of emotional content in media, to understand how emotions transferred via vocal cues, facial expressions, and text, to understand audience engagement, behavior, and their emotional responses (Nabi & Green, 2015).

Empathy: is the ability of perceiving, understanding, and share others emotions (Davis, 2018). Empathy playing a vital role in many fields like healthcare, education, and media it's also fosters social cohesion, social behaviors, and effective communication (Cohen, 2018) and (Riess, 2017). In this thesis, empathy is studied as a semiotic process that influence the way listeners recognize and interpret emotions in vocal communication.

Semiotics: is the study of how sender initiate and encode message, and how the receiver decode and interpret it (Saussure, 1974). In this thesis, semiotics discusses the way vocal cues (Tone, Pitch, Volume, Speed and Intonation) act as a signs that carry emotional states via vocal communication.

Chapter Two : Literature Review

2.1 Introduction

While an in-depth literature review indicates that considerable research has been done regarding emotion recognition and semiotics in various cultural contexts, it is surprising that there is a little information available about Arabic vocal communication. This gap presents a great opportunity to scholars to investigate how emotions are expressed and interpreted in this language. As such, this research will bridge this gap through providing some foundational insights which might be used to guide other scholars who would wish to extend their knowledge on emotional aspects of language use. The literature selected for review clearly shows interdisciplinary, as this thesis incorporates components from across a whole range of other fields besides media and communications studies.

In exploring the literature surrounding emotion recognition, empathy, and communication, several key trends and areas for further research have emerged, especially concerning Arabic vocal communication. While much progress has been made in understanding emotional expression across cultures, the distinct characteristics of the Arabic language and its vocal features remain underexplored. This literature review will synthesize the research into key areas, focusing on the role of emotion recognition in vocal communication, the challenges specific to Arabic, the role of empathy, gender-related differences, and the latest advances in deep learning for emotion recognition.

This chapter consist of three sections, it will begin with an exploration of the conceptual framework, which integrates Brunswik's Lens Model, Semiotics, and Empathy. Then it's followed by theoretical framework section that delve deeply into the application of Brunswik's Lens Model in vocal communication, mainly in the context of emotion recognition. Additionally, this section addressing gender differences in emotion recognition and discusses the challenges in Arabic vocal emotion research. Finally, the chapter identifies the research gap that this thesis aim to address.

2.2 Conceptual Framework

This thesis is based on Brunswik's Lens Model, Semiotics, and the Empathy theoretical approaches. These frameworks provide a solid foundation for the examination of how vocal cues, empathy, and cultural context interact in the process of emotion recognition

in Levantine Arabic vocal media content. It offers the lens to analyze the cognitive, perceptive, and semiotic process involved in the recognition of emotions through vocal communication.

Integrating Brunswik's Lens Model, semiotics, and empathy theory provides a comprehensive framework for analyzing emotion recognition in Levantine Arabic vocal media content (see Figure 2.1). The Lens Model explains the cognitive and perceptual processes involved in interpreting vocal cues, while semiotics illuminates the cultural and linguistic codes that shape the meaning of these cues. Empathy theory adds a psychological dimension, highlighting the role of individual differences in emotional understanding. Together, these frameworks shape the study's research design, data collection, and analysis.

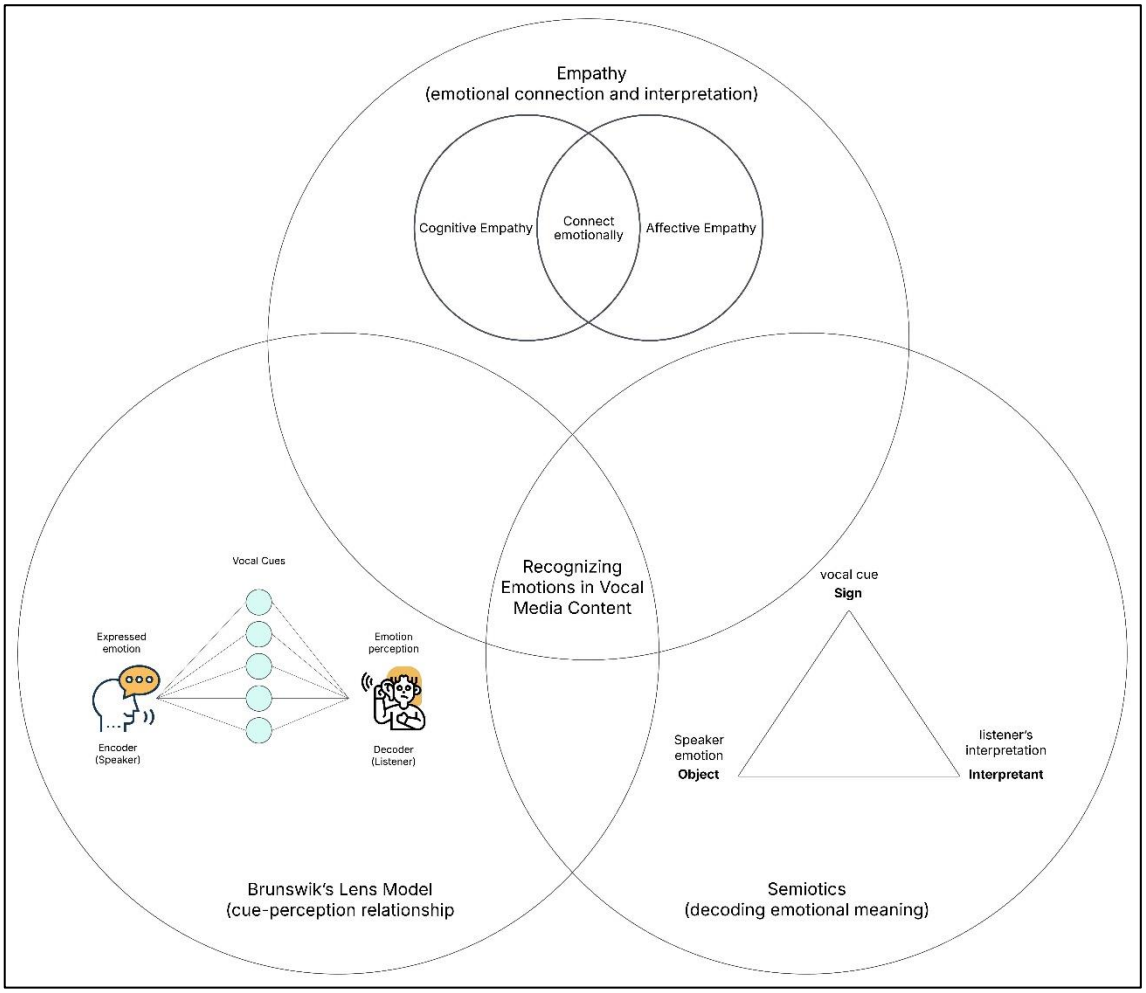

Figure 2.1 Conceptual Diagram Integrating Brunswik's Lens Model, Semiotics, and Empathy

2.2.1 Semiotics

Semiotics, which is the study of signs and their meanings (Saussure, 1974), offers a useful way to understand the mechanism of how vocal cues act as a sign in emotional communication. In this context, vocal elements such as tone, pitch, and intonation are can be seen as signs that carry emotional meaning. So how well listeners understand these signs depends on their familiarity with the cultural and linguistic codes behind them (Kolmogorova et al., 2021).

Charles Sanders Peirce (1934) has expanded Saussure work by introducing the triadic model of the sign which consist of the representamen (form of sign), then the object (the referent), and finally the interpretant (the meaning which interpreted by the interpreter). In addition he classified signs into three types: icons (which resemble their objects, such as onomatopoeic sounds (Ardianto et al., 2022)), indices (which have a causal or contextual connection to their objects, like a trembling voice may indicate fear), and symbols (which have an arbitrary relationship to their objects, such as language (Muin et al., 2021)). These concepts could be used for analyzing vocal cues, as they help to explain how listeners interpret speaker emotions from vocal features (Sauter et al., 2010).

Using semiotics in vocal communication faces many issues, main issue is the variance in vocal expressions and interpretations, which can be influenced by individual, cultural, and contextual factors. Matsumoto et al., (2008) show that same vocal cue like a high-pitched tone can deliver different meanings depending on the speaker gender, age, or cultural background .Another challenge is the usage of artificial intelligence AI in vocal communication, such as smart voice assistants or synthetic voices in automated answering systems. While these technologies have the potential to enhance communication, they still lack the accurate of vocal cues that supports empathic interactions (Shin & Biocca, 2017).

In media and communication studies, scholars used semiotics in order to analyze the role of vocal cues in storytelling and audience engagement. For example, expressive vocal tones in radio dramas programs and podcasts can enhance emotional engagement and immersion of listener (McHugh, 2016). In addition, in advertising, vocal cues used to create an emotional appeals to affect consumer behavior (Nabi, 1999).

In this study, semiotics theory is used to explore how vocal cues in Levantine Arabic media content are expressed and how audience interpret them. By examining the correlation between vocal cues and their emotional meanings, this research delve into the semiotic mechanisms that affect emotion recognition. This method is particularly significant for understanding how elements like culture and linguistic factors affect the decoding of vocal emotions in non-Western cultures.

2.2.2 Vocal Cues

Vocal cues, or vocal features known as a non-lexical components of speech that can carry emotional and contextual information. These includes tone, pitch, intonation, volume, and speed, which function as signs in the semiotic logic (Ostwald, 2019), for example, an increasing high pitch at the end of a sentence can interpreted as a question (an iconic sign), while a shaky voice might indicate anxiety (an indexical sign) (Scherer, 2003).

Vocal cues classified from a semiotic perspective as a multimodal signs that interact with verbal language to create a meaning (Hasyim & Arafah, 2023). For example a melodic prosody in storytelling is integral part for expressing meaning in ways that words alone cannot achieve (Montaño & Alías, 2016). Also vocal features like pitch and tempo are linked at the same time to a different emotions for example high pitch and fast tempo a high indicator for happiness on other hand low pitch, and slow tempo are indicates sadness (Scherer, 2001). This interplay between verbal and paralinguistic signs underscores the importance of semiotics in understanding vocal communication.

Listener play the role of an interpreter who makes sense of vocal cues, using his understanding of the sign system to recognize a meaning. For example, someone who is familiar cultural norms of a specific speech community will consume a less time to interpret vocal cues in the context. This interpretation is shaped by the listener cognitive and emotional skills (Decety & Jackson, 2004).

Juslin and Laukka, (2003) found that listeners are generally perform good in identifying basic emotions depending on vocal cues, with accuracy rates between 60% and 80% . However, recognition of more complex emotions can be difficult, especially across different cultures. Furthermore A study by Elfenbein and Ambady (2002) revealed that people were more successful at recognizing emotions from individuals within their own

cultural group, highlighting the importance of culture familiarity in interpreting vocal cues.

2.2.3 Empathy Theory

Empathy is defined as the ability to perceive, understand, and sharing others' emotions (Davis, 2018), empathy also plays a critical role in the emotion recognition process, furthermore empathy could enhance the listener's ability to interpret vocal cues in a way that fostering a deeper connection with the speaker's emotional state. In this thesis, empathy is conceptualized as a cognitive and an affective process that mediate between vocal cues and emotion recognition.

Empathy have categorized into two dimensions: first is known as "Cognitive Empathy" and the other one is "Affective Empathy". Cognitive empathy related to the ability of individual to understand another person emotional state or perspective (Decety & Jackson, 2004). On another hand affective empathy related to the sharing of emotional experiences of others, that leading to emotional resonance or mimic (Eisenberg & Strayer, 1990). These dimensions work together in balance way to enable individuals to connect emotionally with others, whether in face to face or mediated interactions.

Preston and de Waal (2002) proposed the Perception-Action Model (PAM) of empathy which provides a useful framework for understanding how empathy operates in voice emotion recognition. According to this model, observing or hearing others emotional state can activate corresponding neural representations in the listener, facilitating an empathic response. This model underscores the automaticity of empathy, suggesting that vocal cues can trigger direct emotional and cognitive reactions in listeners.

Active listening requires auditory processing and emotional harmony, which depends on the listener as a central element in the voice emotion recognition process, where the listener decodes the vocal cues in order to decode the speaker's emotional state (Bodie et al., 2015), which allow the listener to give their support or empathize with the speaker. Pasupathi and Rich (2005) argues that listeners who practicing active listening were able to identify and validate speakers' emotions which enhance emotional connections. This emphasize the importance of cognitive empathy in voice emotion recognition, as it requires the listener to actively interpret and contextualize vocal cues.

In the context of mediated communication, like phone calls, voice messages, radio broadcasts or podcasts, listener faces more challenges as a result of absence visual cues. However, both Juslin & Laukka (2003) and Scherer et al. (2001) suggests that listeners can still achieve a high level of accuracy in emotion recognition by focusing on vocal features, which showing the power and abilities of voice emotion recognition.

Active listening can significantly improve the outcomes of crisis communication campaigns by fostering trust and emotional connection ((Kim et al., 2004); (Coombs, 2007)). However, the lack of visual cues in mediated communication can also bring some issues, listeners may misinterpret vocal emotions or fail to detect subtle emotional nuances which lead to misunderstandings or emotional disconnection (Liu et al., 2023). This emphasize the need to develop strong voice emotion recognition skills.

In this study empathy theory integrated into the conceptual framework in a way that allow exploring the mechanism of how empathy differences of individual influence the accuracy of interpretation of emotions. By measuring participants' empathy levels; using the Interpersonal Reactivity Index (IRI), this study investigates whether individuals with high empathy level are more accurate from others in recognizing emotions depending only on vocal cues. This approach aligns will with the study's focus on the semiotic role of empathy in vocal communication that highlighting its function as a psychological "sign" that enhances emotional understanding.

2.3 Theoretical Framework

This section delves into the theoretical foundations that support this thesis, by focusing on Levantine-Arabic media content. begins with an exploration of Brunswik's Lens Model, following this, the role of empathy in vocal communication is examined through a semiotic lens, highlighting its significance in fostering emotional understanding. The section then addresses gender differences in emotion recognition. Finally, this section outlining the challenges of Arabic vocal emotion research. Together, these theoretical perspectives provide a solid foundation in order to investigate the complexity of emotion recognition in Levantine-Arabic vocal communication.

2.3.1 The Brunswikian Lens Model in Emotion Recognition

Brunswik's Lens Model is a theoretical framework that explains the mechanisms of how individuals make a judgments in uncertainty situations, using a variety of cues that are frequently imperfect to deduce an unknown criterion (Brunswik, 1947). This framework count as a fundamental resource in disciplines such as psychology, decision-making, and social judgment theory (Hammond, 1996).

The model posits that in vocal communication, listeners use cues produced by the voice as environmental cues and process them in order to make inferences about the speaker's emotional state. The cues are shaped by the listener's perceptual and cognitive filtering- and most directly by empathy, cultural norms, and past experiences (Scherer, 2003).

The Lens Model has found extensive application in examining judgment accuracy across various fields, including media and communication. The model offers a systematic approach to understanding how individuals navigate complex and multifaceted stimuli. By emphasizing the interaction between environmental cues and cognitive mechanisms (Hammond, 1996) and (Scherer, 2001).

Scherer (2003) suggests to depend in research on vocal communication of emotion on this modified version of Brunswik's Lens Model (see Figure 2.1). In this study, will be used to understand how listeners decodes emotional signals depending only on vocal cues in Levantine-Arabic communication.

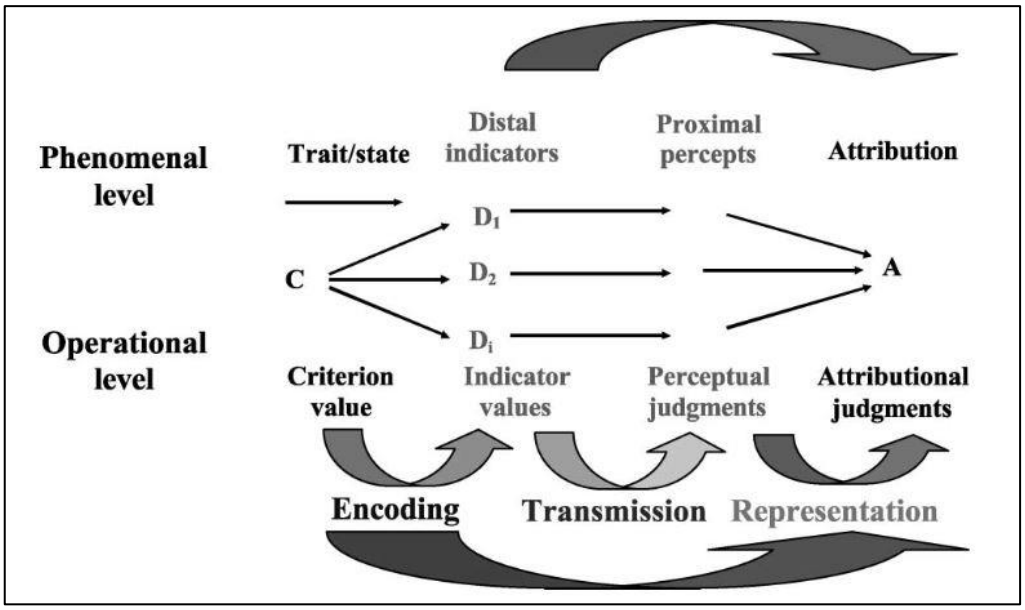

Figure 2.2 Scherer (2003) adopted Brunswikian lens model - adopted for vocal communication of emotion

Egon Brunswik's lens model, offers a perspicuous perspective to the way that human perceives and interprets emotions in vocal communication. In achieving this goal, the model establishes a connection between explicit vocal and acoustic representations and the implicit emotional perceptions. Thus providing a useful approach to study emotion recognition and communication ((Bernieri et al., 1996); (Scherer, 2003); (Hartwig & Bond, 2011)). The Brunswikian lens model in communication can be used to predict social-emotional judgements like power, affection, and credibility based on verbal, and vocal cues (Burgoon et al., 2022) and (Spackman et al., 2009).

However, the model also demonstrates how perception may interfere with emotion recognition through channels, particularly those associated with speech. For instance, acoustic properties such as pitch and tone take part in sending emotions across but they are also affected by environmental noises, hence provoking potential misunderstandings. (Scherer, 2003).

Essentially, the Brunswikian Lens Model considers emotional communication to be inference through which the listener use cues to the interpret the speaker emotional state. Cues like tone, pitch, volume, speed, and intonation, are used as distal stimuli which filtered by perceptual and cognitive processes to yield proximal judgments (Brunswik, 1947). It needs to be emphasize that such processes are probabilistic in nature, realizing

the fact that cues could be ambiguous and possibly subjected to environmental and contextual variation.

For instance, a high-pitch voice is often said to signify excitement or fear under certain contexts; however, this may not be true with different cultural backgrounds, environmental noise, and context within which the conversation takes place (Scherer, 2003). Thus the Lens Model provides a framework for understanding how these factors interact in a way to shape emotional perception and interpretation.

The Brunswikian Lens Model has been a productive vantage point for the study of social-emotional judgments, including perceptions of power, affection, and credibility in communication. Vocal cues of pitch variability and speech rate are strong predictors for judgment (Burgoon et al., 2022). For example, the steady, low-pitched voice of a speaker is regularly seen as more credible and authoritative than a high-pitched (Spackman et al., 2009). However, the model addresses some limitations in recognizing emotions, particularly in noisy environments where vocal features could be distorted due to environmental noise and could lead to misunderstandings which makes it hard for listeners to interpret the speaker's emotional state accurately (Scherer, 2003).

This model aligns well with recent studies that highlights the bidirectional nature of emotion recognition in cross-cultural contexts (Hess & Hareli, 2014). the model’s adaptability to environmental and contextual factors is supported by research demonstrating the impact of noise and cultural differences on emotion recognition accuracy ((Elfenbein & Ambady, 2002); (Alhuzali et al., 2018)).

These features make the Lens Model well-suited for studying Levantine-Arabic vocal media, Furthermore, its focus on cue utilization and interpretation allows for investigating how cultural norms and gender differences shape emotion recognition, some studies show that gender-specific models outperform gender-neutral ones ((Alsharhan & Ramsay, 2020); (Matsumoto et al., 2008)). The model also supports the development of culturally sensitive emotion recognition systems by incorporating cultural-specific features, which have been shown to enhance performance in non-Western settings (Thompson & Balkwill, 2010).

In this thesis, the Brunswick Lens Model is seen as a primary framework for examining how listeners decode emotional signals from vocal cues in Levantine Arabic media content. The model focuses on the relationship between environmental cues and perceptual judgments, thus framing the current focus in the study on vocal elements in emotion recognition. Lens Model implementation help to understand how listeners deal with vocal cues' ambiguity to correctly or incorrectly interpret emotions, thus uncovering some insights into the cognitive and perceptual mechanisms of emotion recognition.

2.3.2 The Semiotic Role of Empathy in Vocal Communication

Empathy is central for effective communication, especially in vocal speech, as it helps to establish a deeper connection between the speaker and the listener by enabling them to understand with each other's emotions (Kolmogorova et al., 2021). For vocal communication, empathy often works through a semiotic process, where the sounds produced by the speaker serve as a signs of his or her emotion. Throughout history, emotionally charged speech had been decided its impact on listeners and rooted in both rhetorical traditions and scholarly analysis (Scherer, Vocal communication of emotion: A review of research paradigms, 2003).

Empathy not only assisting to interpret emotions, but also supporting the way that emotional responses are modulated, and paving the way for more effective and closer interactions. This is particularly seen in therapy situations, wherein empathy brings more profound emotional understanding and support (Vasilyuk, 2016). The semiotic complexity of empathy especially brings out its dual aspect as a personal psychological trait and a socially mediated process, showing its importance within individual and collective contexts (Kolmogorova et al., 2021).

2.3.3 Gender Differences in Emotion Recognition

Research findings shows that women are better at emotion recognition than men especially in recognizing subtle emotions (Sen et al., 2017).This difference is especially noticeable in vocal speech tasks, where acoustic features like pitch typically higher, playing a significant role in recognizing emotions (Dair et al., 2021). Nevertheless, there have been some studies where these gender differences have not been observed and may be influenced by cultural and linguistic factors.

In Arabic-speaking societies, studying the roles of culture norms and gender aspects in the field of emotion recognition remains underexplored, Alsharhan and Ramsay (2020) has explored the role of gender and dialect in Arabic Automatic Speech Recognition (ASR) systems. They found the models that trained only on female speech data ,performed better than models that trained only on male speech data, with female speakers achieving lower Error Rates. This suggests that women may articulate more clearly and have less pitch variation, which could improve emotion recognition accuracy.

Hattar and Ramirez-Hurtado (2014) found that Individuals within Levantine-Arab communities, who engage in collective singing and performs similar melodies for their distinct chants, enhance their mutual understanding through this practice. This shared experience fosters genuine listening. So understanding these dynamics could offer a valuable insights for creating more cultural and sensitive emotion recognition models (Abdel-Hamid, 2020). Furthermore, According to Israelashvili et al. (2020) empathic concern may influenced by gender differences in a way that increase emotion recognition ability.

2.3.4 Challenges in Arabic Vocal Emotion Research

Despite the popularity of studying emotion recognition in various cultures and languages, few studies have concentrated on Arabic vocal communication. Additionally, Most of the previous studies rely on elicited or acted emotions, that cannot fully capture natural emotions from speech that we use in everyday life (Abdel-Hamid, 2020). However, the use of controlled data has led to doubts regarding their applicability and accuracy. Furthermore, there are no comprehensive databases for Arabic spoken language which further limits the effectiveness of any Arabic SER systems (Ouali & El Garouani, 2024) and (Dalal & Rayhana, 2020).

As regards the emotion recognition in Arabic language, it becomes more difficult due to specific phonetic and prosodic features of language itself, which are combined with a great number of its dialects like Levantine-Arabic dialect. Arabic has some particular sounds not always represented in datasets used, so that makes it hard to create algorithms recognizing many voice emotions accurate enough (Ouali & El Garouani, 2024).

Moreover, the majority of the SER systems that are available today have been designed with reductive focus on a few emotions, and thereby neglecting the complete range of

emotions, which limits their utility in the field of human-computer interaction and affective computing (Dalal & Rayhana, 2020).

Despite these challenges, speech emotion recognition (SER) functionalities have experienced significant improvements thanks to the latest advancements in deep learning technology. Traditionally, SER models have concentrated on a small number of simple emotions articulated by Paul Ekman, assumed to be typical worldwide (Ekman, 1982). Such feelings have been challenged as universal by some scholars, Barrett (2017) arguing that cultural and linguistic factors are shaping emotions. This argument is especially relevant for Arabic language, which has a rich phonetic and prosodic characteristic (Embarki, 2008), these features makes emotion recognition more challenging for both listeners and SER systems.

Deep learning processes have facilitated the enhanced understanding of emotion recognition where it can capture the slightest nuance of emotional expression in language; however, there exist some challenges such as reducing cultural biases or addressing ethical issues for its application in judgement (Klaylat et al., 2018).

There is still room for improvement, as far as Arabic speech emotion recognition research is still in early stage. Studies in this field often rely on a small, acted datasets. For example, Klaylat et al. (2018) has analyzed Arabic TV shows, and limited the study only on three emotions, happiness, anger, and surprise. But the limited emotional range and lack of gender-specific analysis suggest a room for improvement. Similarly, Meddeb et al. (2016) has achieved strong results using machine learning models, but their research also lack for gender differences.

Studies that mentioned above has emphasize the need for a larger and more diverse Arabic language speech dataset, in order to develop more accurate and culture sensitive SER.

2.4 Research Gap

Most of the previous studies in the field of emotion recognition have focused exclusively on Western languages and controlled datasets that generally neglecting the culture and language differences that found in other languages. This makes many of these studies' findings less generalizable to other diverse contexts, especially the Arabic language. Cultural differences have been found to affect emotion recognition accuracy (Elfenbein & Ambady, 2002). Additionally, emotions have also been found to differ when expressed naturally as opposed to portrayed one (Scherer, 2003).

While recent studies have focus on emotion recognition in Western languages using controlled datasets, several research gaps arise. Main issue is the scarcity of large, diverse, and representative datasets, which hinders the development of capable models to be generalized across varied real-world contexts (Kakuba & Han, 2025). In addition, many datasets consist of portrayed emotional expressions, which may not accurately represent natural real emotions and affecting model performance when applied to natural expressions (Althobaiti, 2023). Furthermore, there is a need for more comprehensive studies that include another languages and cultural contexts in order to enhance the universality of emotion recognition systems (Muhammad et al., 2025). Lastly, an issue of reliable emotion labels in inconsistent annotated datasets can lead to unreliable models underscoring the importance of standardized annotation practices (Bashir et al., 2023).

This research is intended to fill these gaps by focusing on authentic datasets of Levantine Arabic vocal media content because natural expressions of emotions within this culture will be reflected. By analyzing genuine emotional expressions, the study intends to develop improvements that enhance the accuracy level of emotion recognition systems. Additionally, this thesis also looks at how gender affects emotion recognition and how culture and language are involved in this process. This falls in line with the findings that cultural differences can affect strategies in emotion regulation and accuracy in recognition (Matsumoto et al., 2008). Also exploring the concept of empathy as a semiotic and cognitive process that explains how empathetic engagement with vocal cues by listeners influences emotional interpretation. The socio-cultural theory of emotions underlies this view, suggesting that emotions arise from interpersonal interactions within agreed-upon cultural meanings and practices (Mesquita, 2002).

By congregating all these, this study builds on media and communication studies by explicating cultural and language factors that influence emotional expression and perceptions of Levantine Arabic vocal media content. The study also contributes to research on affective computing because it provides a very subtle and rich understanding of emotion recognition in the non-Western context and could, therefore, be used to develop emotion recognition systems that are much more sensitive to culture. The study advances the knowledge of Arabic linguistics on the way Levantine Arabic vocal cues perform as semiotic signs linking linguistic features with emotional meaning.

Chapter Three: Methodology

3.1 Introduction

This chapter presents the applied methodology in the study, which adopt the descriptive methodology with a mixed method approach. The chosen methodology aligns with the study aim of exploring how audience empathy influences the process of emotions recognition relying only on voice in the Levantine-Arabic media content in a natural and without any intervention.

The chapter begins by introducing the methodology, then move on to describe the research design and tools. It then outlines the three main stages of the study and describes the procedures followed, and explain the construction of the questionnaire as the primary research tool. Additional sections which discuss the participant selection criteria, how data will be analyzed, and how the research questions will be addressed. The chapter explains how the Brunswik's Lans Model is applied to interpret how listeners rely on vocal cues to interpret emotions in voice media content.

3.2 Research Methodology

The researcher adopted a descriptive methodology to answer the research questions in this thesis. The descriptive methodology is a widely applied approach in the field of communication and media studies (Jensen, 2012). It is appropriate when the research goal is to explore and analyze phenomena as they occur in real life, without intervention (Livingstone, 2002). This methodology allows scholars to systematically diagnose and identify behavioral patterns and characteristics related to audience response to media (Webster, 2018). It also shed light on how audience interpret the content they receive, especially in terms of emotional engagement (Reinhard, 2021).

In the context of communication and media research, especially audio content, the descriptive method assists researchers in exploring how listeners emotionally interpret media messages (Winkler & Appel, 2024). Since the present study examines how Palestinian audiences recognize emotions from Arabic audio content, particularly in the Levantine dialect, this approach supports ecological validity in exploring audience responses and empathic tendencies in real time (Dziobek, 2012); (Levordashka et al.,

2025). Additionally, the descriptive methodology provides a solid theoretical basis for comprehending how vocal cues interact with listener empathy level and demographic variables. These interactions influence how audience interpret media content.

The choice of a descriptive approach is consistent with the design of this non-experimental, cross-sectional study, which facilitates the collection of both quantitative and qualitative data. This approach ensures sensitivity to the real-life context of the audience, and interpretive frameworks, while still maintaining methodological rigor (Tuli, 2011).

3.3 Research Design

In this thesis, the researcher has designed the methodology to ensure balancing between the theoretical precision and contextual sensitivity, by using mixed-methods with a non-experimental and cross-sectional design. In addition, this study combines quantitative and qualitative methods. This combination is widely recommended in media and communication studies to capture the complexity of emotions and interpretive responses to media content (Creswell & Clark, 2007); (Tashakkori & Teddlie, 2010).

The reason behind adopting a mixed-methods approach in the present study attributed to the multidimensional nature of emotion recognition process and audience empathy. While the quantitative data such as (empathy level scores) provides a structured and measurable insights into participants' behavior patterns, the qualitative data offers a significance and contextual interpretations of the process of perceiving and explaining emotions. Relying only on one method would risk to ignore the statistical generalizability or the interpretive depth which required for this study. Thus, the integration of both will allow the study to capture full range of emotional and empathetic dynamics which involved in responding to Levantine-Arabic audio content.

Using natural audio content will allow participants to engage with emotionally rich speech in a way that closely reflects real-world media consumption, in a way that enhance the ecological validity of the research (McStay, 2018); (Scherer K. , 2003). The researcher will integrate empathy assessment alongside the perceptual labeling, and reflective interpretation for the purpose of constructing a multidimensional view of emotion recognition process specifically the voice emotions.

Furthermore, in order to strengthen the reliability and analytical robustness of the findings (Diemerling et al., 2024) the researcher will incorporate expert ratings which provides a human grounded evaluation to compare with participants' results, and AI-based emotion recognition tool will offer computational objectivity. This triangular layered strategy will serve the present study by validating the findings and highlighting any possible interpretive variations. This triangulation supports the methodological integrity and corresponds well with best practices in the field of affective computing and communication and media studies.

The study will proceed through three sequential phases:

1- **Empathy Assessment**
2- **Emotion Recognition Task**
3- **Qualitative Reflection**

3.3.1 Empathy Assessment

To measure the levels of empathy for all participants, the researcher will use the Interpersonal Reactivity Index (IRI) of Davis (1980) (see Appendix C). This tool is consisting of 28 questions, which are consist from four subscales:

a- Empathic Concern (EC)
b- Perspective Taking (PT)
c- Fantasy (FS)
d- Personal Distress (PD).

The EC and PT subscales are especially associated to this research, as they measure the affective and cognitive components of empathy.

The IRI has been utilized extensively in media and communication studies in order to explore how individual differences of empathy influences the audience participation and cooperation with media content. In this context, Orduna et al.t, (2023) has employed the IRI tool to measure the empathy level of participants in in-depth communication environment, which shows its applicability in studies that involve auditory media content. In the same way Collins, (2014) has employed it too to observe the relationship between

social media usage and empathy levels, which emphasizes its relevance in communication and media research.

In this study, this phase addresses RQ2 by measuring participant's empathy level and analyzing how empathy as a potential semiotic factor that influence the ability to recognize vocal emotions from Levantine-Arabic media content.

3.3.2 Emotion Recognition Task

In this task, participants will listen to 36 audio clips chosen from the Nafs podcast, which was selected due to ecological validity and emotional authenticity. As these audio clips are rich with natural emotional expressions utterances, that were totally absent from acted- improvised datasets; which are often used in the training process of speech emotion recognition systems (Dhamyal et al., 2019); (Milner et al., 2019); (Abbaschian et al., 2021); (Gómez-Zaragozá et al., 2024). Thus, by working with a real-life (non-acted) speech, this study corresponds with researcher's recommendations from communication and media studies for a contextually advanced grounded approach (McStay, 2018).

Duration of each voice sample (audio clip) is between 5 -15 seconds, which assist the purpose to transmit a more sufficient emotion information for accurate perception while ensuring the listener's attention (Gerczuk et al., 2023). This duration provides an opportunity for the listener to absorb cues from vocal parameters like pitch, tone, rhythm, and intensity that is critically important for emotion recognition process within media context (Scherer K. , Vocal communication of emotion: A review of research paradigms, 2003).

In this task, each participant will label each voice sample with one of Ekman's (1992) six basic categories of emotional states: happiness, sadness, anger, fear, surprise, disgust, and neutral. These categories are part of commonly applied frameworks across different cultures for the analysis of emotional expression and are frequently used in communication studies and affective computing (Poria et al., 2017).

The task design reflects a relatively new trend in media and communication research, where emotion analytics have begun mediating the domain of audience participation and certainly for voice and podcast content (Sharma & Verbeke, 2024); (Abbas et al., 2021);

(Reddy et al., 2021). This phase is directly linked to RQ1 which analyzing how specific vocal cues influence emotion recognition accuracy. In addition, allowing the comparison of recognition performance between male and female, which contributes directly to answer RQ4.

3.3.3 Qualitative Reflection

Immediately after each clip, all participants will be required to answer open-ended questions that are aimed to explore their reasoning. This allows integrating qualitative reflection immediately after each audio clip to improve the quality of collected data and to minimize recall bias. Furthermore, it reduce social desirability bias, which is more apparent when reflection is delayed like in a focus group setting (Bergen & Labonté, 2019). Thus, it brings much richer insights into the relationship between vocal features and emotional perception (Kuckartz & Rädiker, 2019).

Moreover, this integration supports mixed-method triangulation, because it enables a deeper semiotic reading of the perception and cognitive representation of emotions. Due to that, the use of real time response feedback was not only appropriate, but integral to the goals of this research which examines audience empathy and emotion recognition in Levantine-Arabic voice content.

This phase enhances the answer of RQ2 by exploring how empathy mediate between emotion recognition. It also contributes to answer RQ3 by providing explanations into how participants apply Brunswik's Lens Model when interpreting vocal cues.

3.4 Temporal and Spatial Scope

The study will be conducted during the spring semester of 2025. As for data collection, it will be done online and from various locations across the Palestinian territories. The researcher will use web-based distribution methods - e.g., posts on social media platforms, WhatsApp groups, and e-mailing networks - to encourage participation among 18- to 45-year-old individuals. Online recruitment mitigates challenges posed by geographic fragmentation in Palestine, enabling access to diverse participants across regions.

Online Data Collection has previously been demonstrated as an efficient approach to accessing geographically distributed populations and enhanced diversification in the samples despite considerable challenges (Wright, 2006); (Evans & Mathur, 2018). Owing to the high level of efficacy and data accessibility of online methods, the data collection round is expected to be completed in probably a week or two.

This flexible method will ensure broad geographic coverage, which done by considering that the target population comes from a diverse segment, and maintaining cost-effective and technically feasible (Benfield & Szlemko, 2006).

3.5 Procedure

The present study aims to address the research questions through a three stages procedure, which includes:

a- Empathy level evaluation using Interpersonal Reactivity Index (IRI): empathy level measurement with four subscales (Davis, 1980).

b- Emotion recognition task where participants will listen to a 36 audio clips and label each clip with one emotion. These audio clips are taken from the *Nafs* podcast (February 2023 – June 2024), and are selected based on clarity, emotional richness, and speaker diversity, duration of these samples range from **5 to 15 seconds**, this is consistent with recommendations in communication and media studies and provides an optimal exposure for accuracy of emotion perception without overloading listeners (Gerczuk et al., 2023).

c- Qualitative reflections by providing participants with an open ended questions to get instant feedback from them.

Each research question has specific variables, which are operationalized and analyzed in the following way:

In the first research question, the independent variables are the vocal cues, which are integrated into the 36 audio clips being chosen from Nafs podcast, while the dependent variable will be emotion recognition accuracy. Participants will have to select an emotion label for each clip. Additionally, participants will specify which vocal cues they think contributed to their judgments. This will allow the researcher to analyze for how much each cue type correlates to emotion recognition accuracy.

For the second research question, the researcher will examine empathy level for each participant using the Interpersonal Reactivity Index (IRI), producing scores of cognitive and affective dimensions for each participant. These are applied to classify individuals into two different groups of empathy levels (high and low) using median split of IRI scores which consistent with prior studies (Chrysikou & Thompson, 2016), then it will be evaluated to measure the accuracy of emotion recognition in order to investigate empathy as a mediator in interpreting vocal emotions.

For the third question, the study relies on Brunswik's Lens Model as an interpretive framework to explain how audience use vocal cues as distal variables and converting them into an emotional judgment (proximal outcomes). Participants' answers to the open ended questions after completion of each sound clip will give the researcher valuable information about their thoughts on the sounds they had heard and the way of how they estimate the speaker's emotion.

For the last question, the researcher will use the gender of participants as a categorical independent variable and the accuracy of emotion recognition as the dependent variable. An independent samples t-test will be used to compare the mean of accuracy between the two genders in order to ascertain if there is a statistically significant differences.

This mixed-method approach fits well into the study's cross-sectional and non-experimental design. It supports a comprehensive investigation of how empathy and vocal cues interplay in understanding emotion in Levantine Arabic voice media content.

3.6 Research Tools

The main tool that the researcher uses in order to collect data from participants is the questionnaire, which designed to collect both of quantitative and qualitative data that related to research objectives.

The questionnaire will be distributed online in a single digital format using the Kobo Toolbox software (https://www.kobotoolbox.org/), this platform supports audio integration in the questioner and served only for delivery and data collection interface. This platform supports all multimedia files, such as audio clips, and allowing the researcher to use them directly within the questioner questions (KoboToolbox, n.d.). This feature is not available in other platforms such as Google Forms or Microsoft Forms. This

feature is very important for this study which requiring participants to listen to an audio content, thereby enhance the collected data validity.

3.6.1 Questionnaire Design

Questionnaire is one of the most used self-reported data collecting tool in communication and media studies, especially in audience's research where researchers aim to explore audience behavioral responses, attitudes, and perceptions of media content (Wimmer, 2011); (Gunter, 2000). Accordingly, the researcher has designed a structured questionnaire to collect quantitative and qualitative data that aligns well with this thesis objectives which investigate the role of empathy of audience in recognizing emotions from audio content in Levantine-Arabic Language.

The questionnaire is designed to minimize cognitive overload through concise, focused questions (Schemer et al., 2008). The design integrates multi elements to collect the data in order to capture individual differences of empathy and its potential effect on audience emotion recognition performance.

The questionnaire consists of the following parts:

1- Demographic data section: in order to collect personal information including age, gender, and educational level (see Appendix B). Demographic data are essential in communication and media studies (Webster, 2018), since it may affect how audience perceiving and engaging with media content. This section follows and align with the ethical standards for anonymity and information consent (see Appendix A) in audience studies.
2- Empathy Level Measurement: the researcher deploys the interpersonal reactivity index (IRI) which developed by Davis (1980) (see Appendix C). This tool consists of a 28 questions and divided into four subscales:
    a- Empathic Concern (EC): which measure the sympathy and concern for others.
    b- Perspective Taking (PT): which measure the cognitive propensity for adopting the others viewpoints.
    c- Fantasy (FS): which measure the propensity to imagine and transpose oneself into fictional situation.

d- Personal Distress (PD): measure the self-oriented anxiety in response to others suffering.

These scales provide the researcher with multidimensional understanding of empathy and its relation with audience ability to process emotional content. These scales are widely used in communication and media to assess the individual empathic differences of audiences, for instance Orduna et al. (2021) used this tool in their research to measure participants' empathy level while interacting with video content.

3- Emotion recognition task: participants will listen to short 36 audio clips which selected from Nafs podcast. After each clip the participants will be asked to select the perceived speaker emotion. This approach aligns with studies of communication and media studies that used vocal emotion recognition tasks to explore media audience effective perception (Scherer K. , Vocal communication of emotion: A review of research paradigms, 2003).
4- Open ended questions: the participants will be asked after labeling the perceived emotion for each clip to answer a series of open ended questions (see section 3.9) in order to reflect on their judgment and to enrich the research with qualitative insights and to support the understanding of the emotion recognition process.

3.6.2 Population and Sampling

The target population of this study is Palestinians, there age group is between 18 to 45 years old and live in Palestine. According to the Palestinian Central Bureau of Statistics (PCBS) (2025), the estimated total population of Palestine at the end of 2024 was about 5.5 million.

While PCBS did not provide any data of age groups for 2024, trends among past demographics show that individuals representing age groups of 15-64 make up about 59% of the total population. Thus, the researcher may safely assume that the 18-45-year-old age group makes up about 50% of the total population, thus constituting approximately 2.75 million individuals (Palestinian Central Bureau of Statistics, 2025).

As the population size for this study is quite large, the standard sample size determination formula for social science research Krejcie & Morgan formula (1970) was used to determine the sample size:

$$\frac{Z^2.p.(1-p)}{e^2} = n$$

Where:

- n is the required sample size
- Z is the z score corresponding to a 95% confidence level (1.96)
- p is the assumed population proportion (0.5 to maximize sample size)
- e is the error margin (0.05).

Given these parameters, a calculated minimum sample size estimation was found to be around 384 participants.

Given the challenge of obtaining a randomized, fully representative sample across multiple regions, a non-probability convenience sampling method had to be used. Convenience sampling was necessitated by logistical constraints in Palestine, including limited access to randomized populations and reliance on digital recruitment. This can be acceptable for exploratory studies that studying psychological, behavioral, and social aspects (Etikan, 2016).

Participants will:

- Provide informed consent, (see Appendix A).
- Confirm no hearing impairments.
- Indicate prior familiarity with the *Nafs* podcast.
- Be reminded of their right to withdraw at any time.

All procedures following the APA ethical guidelines and participant personal data will be anonymized for protection of their privacy.

3.6.3 Units of Analysis

The present study will involve both quantitative and qualitative units of analysis to examine comprehensively the interplay between empathy and emotion recognition from Levantine-Arabic vocal media content.

A- Quantitative Data Analysis:

a. Empathy level scores will be obtained from the IRI, then participants will be categorized to High Empathy Level Group and Low Empathy Level Group using a median split, a common method in empathy studies (Chrysikou & Thompson, 2016).
b. Emotion recognition accuracy will be measured by comparing participants results with experts coded ground-truth labels results, a common procedure used in media studies and affective computing research (Scherer K. , Vocal communication of emotion: A review of research paradigms, 2003).

B- Qualitative Data Analysis:

a. Descriptions of vocal features used in judgments (labeling / annotation) that influenced participant's emotion judgments will be analyzed. The cues include pitch, tone, volume, speed, intonation, and rhythm, which are recognized as significant acoustic indicators of emotional expression in speech (Juslin & Laukka, 2003); (Gunes et al., 2008).
b. Participants will provide reflections on emotional connection or bias during listening, these reflections will provide some critical insights for the subjective dimensions of emotion recognition process, specifically when it studied alongside the level of empathy (Preston & de Waal, 2002); (Decety & Jackson, 2004). Qualitative feedback is coded thematically using NVIVO, which guided by grounded theory principles.

3.7 Validity and Reliability

1. **Instrument Validity:** The IRI has shown high cross-cultural reliability (Gilet et al., 2013).

2. **Content Validity:** Emotion labels are based on Ekman's (1992) six universal emotions.
3. **Reliability:** the researcher will test the interface to ensure clarity, and for qualitative analysis will use NVIVO software with inter-coder reliability measures (O'Connor & Joffe, 2020) . Inter-coder reliability is assessed using Fleiss' Kappa ($\kappa \geq 0.70$), ensuring consistency in qualitative coding.

3.8 Emotion Reference Labeling (Credibility)

In order to establish a credible reference for labeled emotions in the audio clips, the following procedures and validation strategies will be applied:

- Each of the audio clips making up the 36 audio samples in the study is pre-annotated by three experts in emotional communication fields (Psychology, Linguistics, Communication).
- These experts are selected to ensure the balance between a mix of opinions and manageable consensus processes, which align with tasks that involving expert judgment (Verma et al., 2022) .
- These raters will assign an emotion label for each audio clip based on Ekman's model. If there are disagreements, majority voting is used.
- When available, direct verbal cues within the clip confirm the labeling.
- Inter-rater reliability will be measured using Fleiss' Kappa (1971), a statistical measure which is designed for assessing agreement between multiple raters

During the process of establishing true valid data, the role played by expert raters in the annotation of emotional content is very significant, especially while working with vocal materials that are naturalistic and unscripted. To some degree, natural speech distinguished from portrayed datasets where subtle emotional cues deeply embedded in cultural context often need interpretation by trained raters (Quiros-Ramirez & Onisawa, 2013); (Gerczuk et al., 2023). So, expert raters' perception makes sure that the emotional labels reflect both acoustical features and contextual meaning, complying with those standards in affective computing and media and communication studies (Callejas & López-Cózar, 2008); (Siegert et al., 2013).

Experts bring domain-specific understanding about the semiotic and prosodic dimensions of emotional voice, which helping in more accurate identification of ambiguous or subtle expressions. This is particularly vital in Arabic dialects such as Levantine Arabic where vocal intonation may greatly vary from intonation that found in western expressions corpora (Hellmuth, 2020); (Brown & Hellmuth, 2022). The expert's agreement also gives a more reliable benchmark when evaluating the performance of participants in recognition tasks.

The researchers also will use an AI-driven emotion recognition tool as a secondary reference layer to enrich the findings. The machine learning mechanisms that used in AI tools are fast and highly scalable in providing annotation, however, they often achieve a low performance when it comes to non-western dialects or natural speech as a result of mismatching between cultural and acoustic while training AI models on the data (Mousi et al., 2024); (Aladeemy et al., 2024). However, Combining AI-derived ratings with expert agreement serves as a supportive performance indicator, making it easy to evaluate divergence and intensity related with the perspective of annotation (Diemerling et al., 2024).

The mix of human expert judgment and computational output regarding emotional labeling affords a richer evaluation of emotional content and greatly boosts the reliability and cross-validation potential, especially in underexplored domain of Levantine Arabic media content.

3.9 Open-Ended Questions

A set of questions are presented at the end of each audio clip in the emotion recognition task. These questions are meant to investigate interpretive processes experienced by listeners when making judgments about vocal emotional parameters. The questions are aimed at various levels of engagement: that is, identifying the emotions as perceived, a reflection on the interpretation of vocal features, the phase when the listener becomes emotionally involved, an emphasis on the internal emotional state, and noticing any potential contradiction between the perceived emotion and content of audio clip.

1. What emotion do you think the speaker is trying to express in this audio clip?

2. Which vocal features that assist you to identify this emotion? Mark them [You can mark more than one vocal cue: Pitch, Loudness, Intonation, Speech Rate, Rhythm, Pausing, and Timbre.]
3. Did any part of the speaker's message remind you of something familiar? If yes, please describe.
4. Are there any personal experience or events that might influence your interpretation of the speaker message? If yes, please describe.
5. Did you notice any differences between the speaker tone and the content of their message? If yes, please describe.

3.10 Theoretical Application of Brunswik's Lens Model

The Lens Model of Brunswik put forward a conceptual foundation that can be used to explore how individuals use observable cues to build judgments about unobservable psychological states. Since we are interested in how listeners extract emotional meaning from voice-related features of spoken communication, it is clear that the model was originally developed in a perceptual psychology context. It postulates that there is a relationship of a probabilistic nature between distal variables (which are environmental cues) and proximal perceptual judgments. Considered in this regard are the cognitive and emotional filters of each individual that can mediate or intervene in these relations among these distal and proximal variables in tune with individuals' needs and abilities (Blackhurst et al., 2024); (Wu et al., 2021); (Burgoon et al., 2022).

In this research, using the lens model to explore the participants' interpretation process in a way that show how they decode various vocal signs to understand the emotional state of the speaker. These are the distal variables that are refer to as the visible input that is communicated forward by the speaker. In contrast, participants engage in emotion labeling and writing an open-end reflection, these are the proximal judgments, or emotional perceptions that are influenced by perceptual and empathic abilities of listener.

The present study directly puts into practice the lens technique by exposing participants to natural audio stimuli that the participants must identify as to the emotion choice and provide a justification. The use of the Interpersonal Reactivity Index (IRI) also permits the exploration of intrapersonal variations since participants with higher cognitive or empathic capacities may have different weightings for vocal cues which would allow

them to make more accurate judgments (Israelashvili et al., 2020); (Mackes et al., 2018); (Chrysikou & Thompson, 2016).

Moreover, the model supports a comparative analysis about gender, as previous studies demonstrate gender-specific differences in vocal emotion recognition that could influence the patterns of cue utilization (Memon et al., 2019); (Dair et al., 2021); (Hall et al., 2025); (Lausen & Schacht, 2018). By using Brunswik's model, not only did the researcher interpret the participants' decisions within the context of observable cues, but the research also analyzed how cues might have been filtered through social, cognitive, and emotional mechanisms. The model directly addresses the research questions of the study (See Table 3.1).

Table 3.1 Application of Brunswik's Lens Model in Relation to Research Questions

Research Question	Distal Variable	Proximal Judgment	Empathy's Role
RQ1.	Vocal cues (e.g., pitch, tone, rhythm)	Emotion label assigned by participant	Moderates the interpretation of cues
RQ2.	IRI scores (dispositional empathy)	Accuracy of recognition	Direct mediator between cues and judgment
RQ3.	Vocal cues in natural speech	Justifications in open-ended responses	Filters and organizes cue usage
RQ4.	Vocal cues interpreted by males/females	Emotion recognition performance	May vary as a function of empathy differences

Brunswick Lens Model has been successfully employed in some studies in the field of emotional communication to assess honesty, credibility, and emotional involvement by multimodal cues (Scherer K. , 2003); (Blackhurst et al., 2024); (Burgoon et al., 2022).

In end, this study is a theoretical intersection between semiotics theory and listener behavior, making it an ideal method to try to understand how empathy functions as a mediator in decoding communicative signals in Levantine Arabic vocal media content.

Chapter Four: Results

4.1 Introduction

This chapter presents the results of the study depending on the research questions that outlined in Chapter 1. Data collection was done through an electronic questionnaire, which designed to measure participants' ability to recognize emotions in the Levantine-Arabic audio media content, as well as their level of empathy and some demographic variables, in addition to their responses to open-ended questions. Descriptive and inferential statistics are used to explore the relationships between empathy levels, vocal cue perception, and emotion recognition.

4.2 Sociodemographic results

In the study the sample consist from both females and males, the study involved 194 females (50.5%) and 190 males (49.5%), which 384 participants in total. The distribution of participants' age showed that 38.0% of them thee age were between 18 and 21 years, 34.1% between 21 and 24 years, and 27.9% were 24 years and older (See Table 4.1).

In addition, the marital status appeared that nearly half of the sample were single (45.3%), while 30.5% reported being in a relationship and 24.2% were married. Regarding university education level, the vast majority of the participants were undergraduates (97.7%), with only 2.3% of a graduate degree students. Most participants were still students (93.8%), while 3.9% had graduated from the university and 2.3% were recently graduated.

When considering university year of study, 29.9% of participants were first-year students, 23.2% were in their second year, 21.6% in their third year, and 19.0% in their fourth year, and with 6.3%, being graduated students.

In order to ensure the data reliability, all participants reported no hearing impairments and did not use hearing aids. Media consumption habits, specifically podcasts contents were also noted that 74.7% of participants reported that they usually listen to Arabic

podcasts, with 58.3% listening once a month and 41.7% once a week. In addition, the familiarity of audience with the Nafs podcast, which used in the study as a source for audio content materials, appear moderated, with 19.3% indicating prior familiarity, while 74.2% were unfamiliar.

This demographic profile highlights a balanced gender distribution, a mainly young, undergraduate, and with a high engagement with Arabic podcast media content, which in end providing a relevant sample for exploring emotion recognition in Levantine-Arabic vocal content.

Table 4.1 Demographic Characteristics of Participants (*N = 384*)

Baseline characteristic	Female		Male		Total	
	n	%	*n*	%	*n*	%
Age						
18–below 21	70	36.1	76	40.0	146	38.0
21–below 24	70	36.1	61	32.1	131	34.1
24 and above	54	27.8	53	27.9	107	27.9
Marital status						
Single	81	41.8	93	48.9	174	45.3
Married	56	28.9	37	19.5	93	24.2
In a relationship	57	29.4	60	31.6	117	30.5
Educational level						
Undergraduate	189	97.4	186	97.9	375	97.7
Graduate	5	2.6	4	2.1	9	2.3
Student status						
Student	176	90.7	184	96.8	360	93.8
Fresh Graduated	7	3.6	2	1.1	9	2.3
Graduated	11	5.7	4	2.1	15	3.9
Study year						
First year	59	30.4	56	29.5	115	29.9
second year	41	21.1	48	25.3	89	23.2
third year	42	21.6	41	21.6	83	21.6

fourth year	34	17.5	39	20.5	73	19.0
other	18	9.3	6	3.2	24	6.3
Podcast Listening habit						
Once a Week	81	41.8	79	41.6	160	41.7
Once a Month	113	58.2	111	58.4	224	58.3
Listen to Arabic Podcast [a]	146	75.3	141	74.2	287	74.7
Nafs Podcast Familiarity [a]	33	17	41	21.6	74	19.2

Note. [a] *Reflects the number and percentage of participants answering "yes" to this question.*

4.3 Answering Research Questions

This section presents the study's findings in direct correlation with the four research questions, which ensures a clear tie between the research questions and results. Each question is dealt separately to show the relevant statistical and analytical results, along with all tables that demonstrate the identified trends and patterns. The section begins with Research Question 1, which inquires how various vocal cues like tone, pitch, volume, speed, and intonation affect the process of emotion recognition. Research Question 2 follows, which delves into the semiotic function of empathy in the process of emotions recognition from Levantine-Arabic vocal media content. Research Question 3 is addressed by applying the lens model of social judgment theory in order to explain the relationship between vocal cues and emotion recognition. Lastly, Research Question 4 investigates possible gender differences in the accuracy of emotion recognition between women and men.

4.3.1 Research Question 1

This section presents the study's results that related to the first research question, which exploring the way that certain vocal cues affect emotion recognition process in Levantine-Arabic vocal media content. The analysis tries to establish how much each cue contributes to correctly perceiving a particular emotion as reported by the participants. Results from descriptive and inferential analyses highlight the patterns of cue usages per emotion.

RQ1. How do the following vocal cues influence emotion recognition?

a. Tone

b. Pitch

c. Volume

d. Speed

e. Intonation

The researcher has examined the relationship between perceived emotions and the vocal cues participants reported using to identify these emotions across all 36 audio clips. Cross-tabulations were conducted for each clip, aggregating the frequencies of vocal cues selected for each of the seven emotions.

Descriptive analysis showed, however, some clear trends on how participants selected different cues to discriminate between the emotions. Pitch and tone of voice were the most commonly used cues to discriminate across emotions. As appeared in **Error! R eference source not found.**, happiness selection of vocal cues involved pitch 68% of the time and rhythm 54%, with tone of-voice accounting for 49%. On the other hand, sadness was identified mostly through tone-of-voice (72%) and slower speech rate (61%), with pitch mentioned less often (38%). Fear showed a somewhat mixed profile, pitch (64%) and speech rate (58%) being the most important cues along with volume (42%).

It appeared that anger was primarily expressed through speech rate (66%) and volume (63%), and to a lesser extent through pitch (55%). Surprise brought out a rather different pattern, where pauses and hesitations were reported by 70% of participants, followed by changes in rhythm (52%). Disgust, being less overall considered, tied strongest with tone of voice (58%) and pauses (47%).

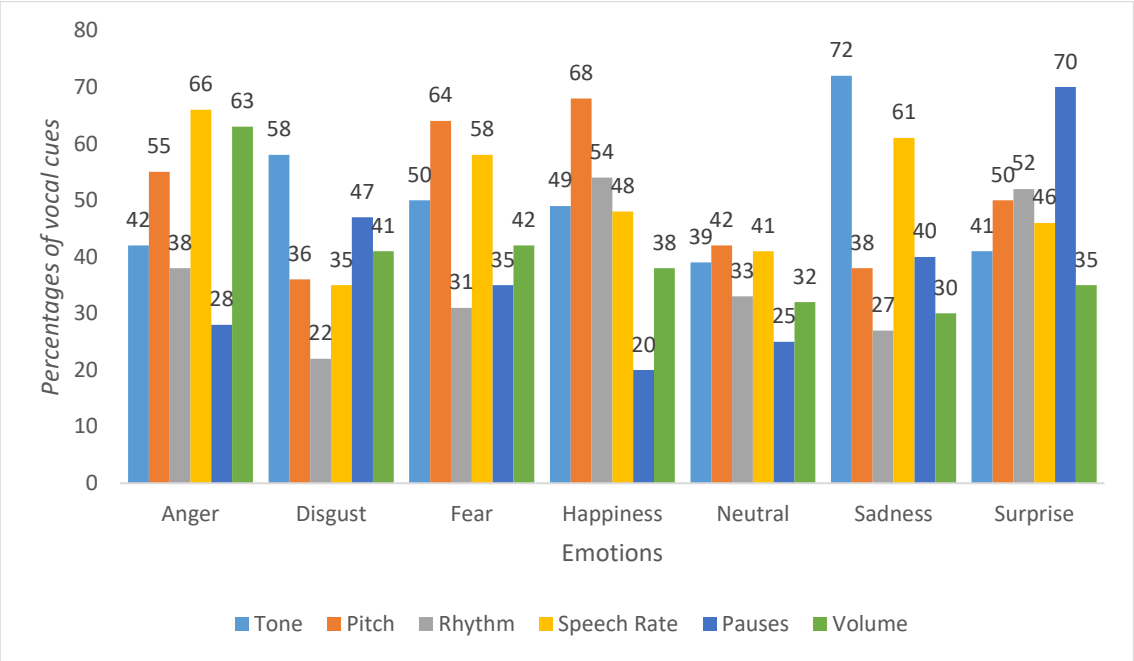

Figure 4.1 Percentages and Frequencies of vocal cues used to identify each emotion

Compared with one another, happiness and anger tended to elicit the most consistency in relying on multiple cues (e.g., pitch, volume, and speech rate), for their distinctive and dynamic vocal profiles. Sadness was instead mostly recognized through fewer, but more stable acoustic cues, mainly tone, and speech rate. Interestingly, fear overlapped with sadness and surprise, thereby constituting a multidimensional emotion in the vocal domain (see Table 4.2).

The line chart illustrates Cramér's V the effect sizes (see Figure 4.2) ranging between 0.15 and 0.26, suggesting small to moderate associations between some vocal cues and the recognition of emotions. For instance, an effect size of 0.21 for anger indicates a moderate relationship. Vocal cues such as speech rate and volume are related to the perception of anger, but there are also other factors that related to the context and perceptual cues that influence the emotion recognition process, for example listeners' prior experiences, cultural expectations or assumptions, and the semantic content accompanying the vocal delivery.

While these values are statistically significant, they differ in their practical consequences. For instance, the moderate association for anger (Cramér's V = 0.21) implies that vocal cues such as speech rate and volume explain some variation, but not all, in the recognition of anger. The highest values were given for surprise (Cramér's V = 0.26), pointing to a

stronger prediction of the vocal features such as pauses and rhythms for this emotion. Conversely, neutral emotions gave the lowest value (Cramér's V = 0.15), indicating a lesser dependence on vocal signs and possibly some ambiguities in its recognition. This shows that expressive emotions such as surprise and sadness (0.24) are recognizable through voice alone, whereas less salient emotions, like neutrality, are going to need additional contextual or semantic cues. By making explicit comparisons between the values, one can appreciate how among categories the strength of the vocal-emotion link varies, thereby enriching the interpretation of the statistical results.

Table 4.2 Percentages and Frequencies of vocal cues used to identify each emotion across 36 audio clips

Emotion	**Tone**		**Pitch**		**Rhythm**		**Speech Rate**		**Pauses**		**Volume**		**Other**		**Cramer's V**
	frequency	**%**	**frequency**	**%**	**frequency**	**%**	**frequency**	**%**	**frequency**	**%**	**frequency**	**%**	**frequency**	**%**	
Anger	161	42	211	55	146	38	253	66	108	28	242	63	38	10	0.21
Disgust	223	58	138	36	84	22	134	35	180	47	157	41	46	12	0.18
Fear	192	50	246	64	119	31	223	58	134	35	161	42	54	14	0.2
Happiness	188	49	261	68	207	54	184	48	77	20	146	38	42	11	0.22
Neutral	150	39	161	42	127	33	157	41	96	25	123	32	35	9	0.15
Sadness	277	72	146	38	104	27	234	61	154	40	115	30	31	8	0.24
Surprise	157	41	192	50	200	52	177	46	269	70	134	35	58	15	0.26

Note. Values represent the percentages and frequencies of participants selecting each cue.

Cramér's V indicates the effect size of the association between vocal cues and perceived emotions.

All chi-square tests were significant at $p < .001$.

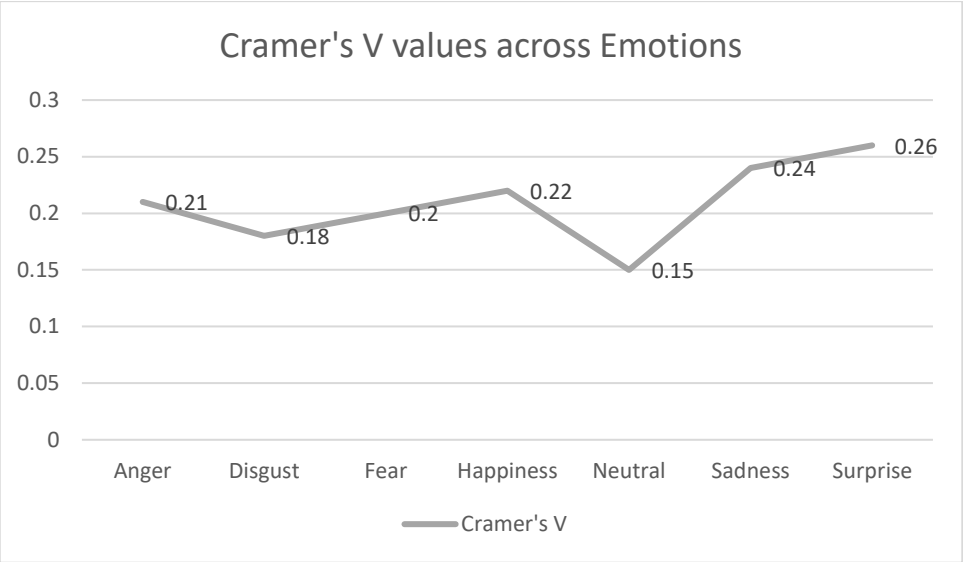

Figure 4.2 Cramer's V values across Emotions

These statistical results emphasize the relative importance of different vocal cues in the emotion recognition process. In addition, the next comparison with the expert judgments provides the study with a deeper understanding of how these patterns converge or diverge between experts in the field, and regular participants. Interpreting the results in this way sheds light on relative strength of vocal cues for particular emotions and on a baseline for comparison to expert judgment. By analyzing the relative importance set on the cues by either expert raters or participant raters, it becomes possible to determine whether a common or divergent strategy stands behind the recognition of emotions.

The researcher further compared expert judgments with participant reports regarding the vocal cues used to identify emotions. Experts consistently prioritized tone of voice (78%) and pitch (64%) across most emotions, followed by speech rate (52%). In contrast, participants relied more heavily on pitch (71%) and rhythm (58%) than experts, with relatively lower emphasis on tone of voice (59%). Notably, for sadness, experts overwhelmingly identified tone and slower speech rate as key indicators, whereas participants incorporated additional cues such as pauses and pitch changes. Similarly, for surprise, participants placed greater emphasis on pauses (70%) compared to experts (56%), who focused more on changes in rhythm and intonation.

Cohen's Kappa for agreement on cue selection across all clips indicated only slight agreement between experts and participants ($\kappa \approx 0.12$, $p < .05$), reflecting modest overlap in their perceptual strategies. These findings suggest that while both groups relied on combinations of vocal features, their prioritization of specific cues differed, with experts

adopting a more focused and consistent approach compared to the broader, multi-cue strategies reported by participants.

Both the differences between experts' and other participants' judgments suggest that expert perception has led to focused attention on certain diagnostic cues such as tone and pitch, while the lay listener forms a broader, rather intuitive approach that appeals to rhythm, breaks, and speech rate. Such differences are explicable through the competing orientations of the two groups. Experts, being informed by theoretical knowledge and training, generally attend more to those diagnostic cues highlighted by the literature, such as tone and pitch. On the other hand, laypersons engage more through implicit and experiential listening, thus giving more importance to rhythmic cues, pauses, and varying speech rates, all of which are deeply rooted in everyday conversational practices.

This discrepancy brings to light that professional insight encourages a selective and theoretically consistent approach, whereas nonspecialized perception is broader and guided by cultural context and personal experience. An analytical comparison along these lines clearly implies that emotion recognition does not fall solely under the acoustic dimension but rather belongs to the socio-cognitive, since theoretical and intuitive pathways lead to diverse routes of perception

Such differences suggest that besides universal acoustic emotion markers, culture and experiential listening habits also affect emotion recognition. In the Lens-Model perspective, this check-in would reveal differences between cue validity and cue utilization and would underscore the interpretability of social judgment in vocal emotion perception.

Overall, the analysis indicates that participants and experts used overlapping but distinct cue combinations when interpreting vocal emotions, underscoring the complexity of auditory emotion perception and the influence of expertise on cue prioritization.

The discrepancy between expert and other participants' judgment may be explained by the difference in perceptual orientation. Experts derive from their formal training or theoretical orientation and tend to focus on diagnostic cues discussed in the literature, such as tone and pitch, for emotion recognition. On the other hand, lay participants seem to adopt a more intuitive and culturally grounded strategy, giving more weight to rhythm, pausing, and speech rate.

This suggests that emotion perception varies with listening habits grounded on experience or culture. More broadly, the disagreement between experts' and non-experts' points to the impossibility of fully explaining emotion recognition in Levantine-Arabic vocal media by relying on universal acoustic indicators alone; a culturally situated interpretation must also be considered. Thus, the findings point to the importance of considering both expert and audience views when inferring vocal emotion recognition.

While research question one looks at the acoustic dimensions of emotion recognition, research question two puts the focus on empathy as one of the personality variables. This shift allows the study to discuss whether listeners' capabilities for empathy may shape systematically how vocal cues are interpreted.

4.3.2 Research Question 2

While research question 1 focuses on exploring how the vocal cues contribute to the process of emotions recognition, research question 2 shifts the focus to the role of empathy in the process by exploring whether individual differences in empathic capacity have any effect on how these cues are interpreted.

This section covers the second research question, which focus on the semiotic function of empathy in recognizing emotion from vocal media content in Levantine Arabic. The analysis incorporates IRI scores to understand how certain types of empathy influence the processes of perceiving vocal emotional expressions. Findings observe to which extent empathy intervenes as a mediator in emotion recognition from vocal cues and to which possible patterns of interpretation participants may varied.

RQ2. What is the semiotic function of empathy in the process of recognizing emotions from Levantine-Arabic vocal media content?

The researcher measured participants' empathy using the Interpersonal Reactivity Index (IRI), which looks at four areas:

1- Empathic Concern (EC)
2- Perspective Taking (PT)
3- Fantasy Scale (FS)
4- Personal Distress (PD).

On average, participants scored 95.27 (±13.66), showing that the group generally had a high level of empathy.

When comparing empathy across different demographics, some clear patterns appeared. Women scored higher than men did on the overall IRI (97.59 ±7.71 vs. 92.91 ±17.50) (see Table 0.3). Participants aged 24 and older had the highest scores (96.92 ±9.76), while those aged 21 to under 24 scored lower (93.76 ±16.24) (see Table *0.4*). Single participants also showed higher empathy (97.40 ±6.23) than married participants (90.01 ±22.36) (see Table 4.5).

Table 0.3 Descriptive statistics for IRI empathy measures and their respective dimensions by gender, data are presented as median (IQR) unless otherwise indicated

Empathy measure	Gender	
	Female	Male
IRI $_{\text{total score}}$	97.59 (7.71)†	92.91 (17.50) †
Empathic Concern (EC)	24 (23-24)	24 (23-25)
Perspective Taking (PT)	25 (24-26)	24 (23-25)
Fantasy Scale (FS)	25 (24-26)	24 (23-25)
Personal Distress (PD)	25 (24-25)	24 (23-25)

†: Data are presented as mean (SD); IRI: Interpersonal Reactivity Index.

Table 0.4 Descriptive statistics for IRI empathy measures and their respective dimensions by age, data are presented as median (IQR) unless otherwise indicated

Empathy measure	Age

	18–below 21	21–below 24	24 and above
IRI $_{\text{total score}}$	95.43 (13.52)†	93.76 (16.24) †	96.92 (9.76) †
Empathic Concern (EC)	24 (23-25)	24 (23-25)	24 (23-25)
Perspective Taking (PT)	24 (23.75-25)	24 (23-25)	25 (23-25)
Fantasy Scale (FS)	25 (24-26)	24 (23-25)	25 (23-26)
Personal Distress (PD)	24 (23-25)	25 (23-25)	25 (24-26)

†: Data are presented as mean (SD); IRI: Interpersonal Reactivity Index.

Table 4.5 Descriptive statistics for IRI empathy measures and their respective dimensions by marital status, data are presented as median (IQR) unless otherwise indicated

Empathy measure	Marital Status		
	single	married	In a relationship
IRI $_{\text{total score}}$	97.40 (6.23)†	90.01 (22.36) †	96.30 (11.48) †
Empathic Concern (EC)	24 (23-25)	24 (23-25)	24 (23-25)
Perspective Taking (PT)	24 (24-25)	25 (23-25)	24 (23-25)
Fantasy Scale (FS)	25 (24-25)	24 (23-25)	25 (23-25)
Personal Distress (PD)	25 (23-25)	24 (23-25)	25 (24-25)

†: Data are presented as mean (SD); IRI: Interpersonal Reactivity Index.

Educational background played a role too. Graduate students scored higher (97.00 ±2.06) than undergraduates (95.23 ±13.82) (see Table 4.6). Among all student statuses, graduates had the highest scores (98.73 ±3.12), followed by fresh graduates, while current students scored slightly lower (95.06 ±14.06) (see Table 4.7). Differences also appeared across study years: second-year students reported relatively higher scores (96.35 ±11.31), while third-year students had the lowest (92.86 ±18.12) (see Table 4.8).

Table 4.6 Descriptive statistics for IRI empathy measures and their respective dimensions by educational level, data are presented as median (IQR) unless otherwise indicated

Empathy measure	Educational level	
	Undergraduate	*Graduate*
IRI $_{\text{total score}}$	95.23 (13.82)†	97.00 (2.06) †
Empathic Concern (EC)	24 (23-25)	23 (23-24)
Perspective Taking (PT)	24 (23-25)	25 (23.50-25)
Fantasy Scale (FS)	25 (23-25)	25 (22.50-26.50)
Personal Distress (PD)	25 (23-25)	25 (23-26)

†: Data are presented as mean (SD); IRI: Interpersonal Reactivity Index.

Table 4.7 Descriptive statistics for IRI empathy measures and their respective dimensions by student status, data are presented as median (IQR) unless otherwise indicated

Empathy measure	Student Status		
	fresh graduated	*graduated*	*student*
IRI $_{\text{total score}}$	98.11 (2.61)†	98.73 (3.12) †	95.06 (14.06) †
Empathic Concern (EC)	25 (23-25)	24 (23-26)	24 (23-25)

Perspective Taking (PT)	25 (23.50-25)	25 (24-25)	24 (23-25)
Fantasy Scale (FS)	25 (24-26)	25 (23-26)	25 (23-25)
Personal Distress (PD)	25 (23-25)	24 (24-26)	25 (23-25)

†: Data are presented as mean (SD); IRI: Interpersonal Reactivity Index.

Table 4.8 Descriptive statistics for IRI empathy measures and their respective dimensions by study year, data are presented as median (IQR) unless otherwise indicated

Empathy measure	Study Year				
	1st year	*2nd year*	*3rd year*	*4th year*	*other*
IRI total score	96.30 (10.85)†	96.35 (11.31) †	92.86 (18.12) †	94.04 (16.08) †	98.50 (2.90) †
Empathic Concern (EC)	24 (23-25)	25 (23-25)	24 (23-25)	25 (24-26)	24 (23-25)
Perspective Taking (PT)	24 (24-25)	25 (23-25)	24 (23-25)	24 (24-25)	25 (24-25.75)
Fantasy Scale (FS)	25 (24-26)	25 (24-25)	24 (23-25)	24 (23-25)	25 (23-26)
Personal Distress (PD)	25 (24-25)	25 (23.50-25)	25 (23-25)	24 (23-25)	25 (24-26)

†: Data are presented as mean (SD); IRI: Interpersonal Reactivity Index.

When looking at podcast listening habits, empathy levels stayed similar across the board, ranging between 94.80 (±14.56) and 95.94 (±12.30). Listening to Arabic podcasts or the "Nafs" podcast did not make a meaningful difference (see Table 0.9). Finally, for all demographics, the median scores for the IRI four subscales stayed consistent, which ranging from 23 to 26. Therefore, this suggests that participants responded in relatively similar ways across these four empathy dimensions.

Table 0.9 Descriptive statistics for IRI empathy measures and their respective dimensions by podcast listening habit, data are presented as median (IQR) unless otherwise indicated

Empathy measure	Podcast Listening habit			
	listen to Arabic podcast [a]	Nafs podcast familiarity [a]	once a week	once a month
IRI $_{\text{total score}}$	95.81 (12.33)†	95.20 (13.45) †	95.94 (12.30) †	94.80 (14.56) †
Empathic Concern (EC)	24 (23-25)	25 (24-25)	24 (24-25)	24 (23-25)
Perspective Taking (PT)	24 (23-25)	24 (23.75-25)	24 (23-25)	24.50 (23-25)
Fantasy Scale (FS)	25 (23-26)	24 (23-25)	24 (23.25-25)	25 (23-25)
Personal Distress (PD)	25 (24-25)	24 (23-25)	25 (23.25-25)	24.50 (23-25)

†: Data are presented as mean (SD); IRI: Interpersonal Reactivity Index.

Note. [a] Reflects the number and percentage of participants answering "yes" to this question.

The comparison of the empathy scores among various demographic groups and behaviors yielded some important conclusions. The gender differences (see Table 4.10) appeared to be the highest. Males scored significantly higher than females on total IRI scores ($U = 14571.00$, $Z = -3.57$, $p < 0.001$), with mean ranks of 212.39 for males and 172.19 for females. On the contrary, females scored significantly higher than males on perspective taking or personal distress. Differences between the two genders could be seen in perspective taking ($U = 14980.50$, $Z = -3.24$, $p < 0.001$) and personal distress ($U = 16045.50$, $Z = -2.24$, $p = 0.02$). The fantasy subscale was close to significance ($U = 16386.00$, $Z = -1.91$, $p = 0.05$), whereas empathic concern showed no meaningful difference ($p = 0.08$).

Although statistically significant, the differences between males and females on the IRI subscales are small with regard to effect size; thus, gender explains only a small amount of variance in empathy. This means males and females differ just a little with respect to perspective taking and distress responses, and otherwise can be said to have quite comparable levels of empathetic tendencies.

Table 4.10 Comparative analysis of empathy scale scores between groups according to gender

Empathy measure	**Gender**	**Mean rank**	**U**	**Z**	***P***
IRI $_{\text{total score}}$	Male	212.39	14571.00	-3.57	0.00
	Female	172.19			
Empathic Concern (EC)	Male	182.86	16599.00	-1.71	0.08
	Female	201.94			
Perspective Taking (PT)	Male	174.34	14980.50	-3.24	0.00
	Female	210.28			
Fantasy Scale (FS)	Male	181.74	16386.00	-1.91	0.05
	Female	203.04			
Personal Distress (PD)	Male	179.95	16045.50	-2.24	0.02
	Female	204.79			

U: (the Mann-Whitney U test determines if two groups are significantly different from one another).

Z: (the Z-score can be used to determine if the difference is indeed significant when the sample sizes are sufficient.

Therefore, although differences relating to gender are exist, they do not seem to possess much power in explanation. This stimulating question encourages further analysis as to whether other demographic variables, such as age, marital status, and education, have similar or different effects on the way empathy and emotion recognition are shaped.

These statistical results confirm that, as well as significant between-group differences, the size of practical impact is, however, minimally small. This creates, in essence, an important distinction, pointing to the fact that demographic variables account for little variance in empathy compared to individual differences. Beyond the gender differences,

the following results examines whether empathy differs across age groups, which in turn provide a wide demographic perspective.

When age groups were considered (see Table 4.11), none of the total IRI scores and subscales were significantly different ($H = 3.36$, $p = 0.18$; $p > 0.05$ for all subscales). The same was true for marital status (see Table 4.12). Participants who were in a relationship did have slightly elevated mean ranks over their married or single counterparts, but the difference was not statistically significant ($H = 1.06$, $p = 0.58$).

The lack of statistically significant differences in most demographic categories, combined with small effect sizes, suggests stable empathy scores across age, marital status, level of education, and year of study. Such stability implies that empathy in this sample is functioning more as an individual variable than as a demographic variable.

Table 4.11 Comparative analysis of empathy scale scores between groups according to age

Variables	**Age**	**Mean rank**	**Kruskal-Wallis H**	***P***
IRI $_{\text{total score}}$	18–below 21	193.55	3.36	0.18
	21–below 24	180.05		
	24 and above	206.31		
Empathic Concern (EC)	18–below 21	193.56	0.02	0.98
	21–below 24	191.39		
	24 and above	192.41		
Perspective Taking (PT)	18–below 21	193.98	2.93	0.23
	21–below 24	180.79		
	24 and above	204.82		
Fantasy Scale (FS)	18–below 21	201.53	2.982	0.22
	21–below 24	179.53		

	24 and above	196.06		
Personal Distress (PD)	18–below 21	183.96	2.95	0.22
	21–below 24	190.00		
	24 and above	207.22		

Table 4.12 Comparative analysis of empathy scale scores between groups according to marital status

Variables	**Marital status**	**Mean rank**	**Kruskal-Wallis H**	***P***
IRI $_{\text{total score}}$	Single	193.21	1.06	0.58
	Married	183.18		
	In a relationship	198.86		
Empathic Concern (EC)	Single	199.51	3.77	0.15
	Married	173.48		
	In a relationship	197.19		
Perspective Taking (PT)	Single	196.68	0.55	0.76
	Married	186.69		
	In a relationship	190.90		
Fantasy Scale (FS)	Single	198.99	3.19	0.20
	Married	175.04		
	In a relationship	196.73		
Personal Distress (PD)	Single	188.45	1.90	0.38
	Married	185.72		
	In a relationship	203.91		

With regard to education type (see Table 4.13), no significant difference in total IRI score found (U = 1346.50, Z = -1.04, p = 0.29). However, there was a difference too significant between groups for empathic concern (U = 976.00, Z = -2.20, p = 0.02), with undergraduates (mean rank = 194.40) scoring higher than graduate groups (mean rank = 113.44) (see Table 0.14).

Table 4.13 Comparative analysis of empathy scale scores between groups according to educational level

Empathy measure	**Educational level**	**Mean rank**	**U**	**Z**	***P***
IRI $_{\text{total score}}$	Undergraduate	193.41	1346.50	-1.04	0.29
	Graduate	154.61			
Empathic Concern (EC)	Undergraduate	194.40	976.00	-2.20	0.02
	Graduate	113.44			
Perspective Taking (PT)	Undergraduate	192.14	1552.00	-0.42	0.67
	Graduate	207.56			
Fantasy Scale (FS)	Undergraduate	192.26	1597.50	-0.27	0.78
	Graduate	202.50			
Personal Distress (PD)	Undergraduate	192.40	1649.00	-0.12	0.90
	Graduate	196.78			

U: (the Mann-Whitney U test determines if two groups are significantly different from one another).

Z: (the Z-score can be used to determine if the difference is indeed significant when the sample sizes are sufficient.

Table 0.14 Comparative analysis of empathy scale scores between groups according to student status

Variables	**Student status**	**Mean rank**	**Kruskal-Wallis H**	***P***
IRI $_{\text{total score}}$	Student	190.50	2.23	0.32
	Fresh Graduated	205.33		
	Graduated	232.73		
Empathic Concern (EC)	Student	192.55	0.01	0.99
	Fresh Graduated	195.56		
	Graduated	189.47		
Perspective Taking (PT)	Student	191.03	1.05	0.58
	Fresh Graduated	215.39		
	Graduated	214.03		
Fantasy Scale (FS)	Student	191.57	0.71	0.70
	Fresh Graduated	221.89		
	Graduated	197.30		
Personal Distress (PD)	Student	190.18	6.01	0.04
	Fresh Graduated	174.83		
	Graduated	258.70		

Results show that there were no student status differences ascertained by the total IRI score ($H = 2.23$, $p = 0.32$), nor for most subscale scores; however, there was a significant difference for personal distress ($H = 6.01$, $p = 0.04$), with graduates holding the highest mean rank (258.70).

Furthermore, no any significant difference was noted in total IRI scores or in subscales when study year variations were considered (see Table 4.15), frequency of podcast listening (see Table 0.16), Arabic podcast listening patterns (see Table 0.17), and knowing about the “Nafs” podcast (see Table 4.18) ($p > 0.05$ in all cases). The fantasy subscale had a borderline statistical significance in favor of Arabic podcast listeners, where these listeners had marginally higher scores than non-listeners did ($p = 0.05$).

Table 4.15 Comparative analysis of empathy scale scores between groups according to study year

Variables	**Study year**	**Mean rank**	**Kruskal-Wallis H**	***P***
IRI $_{\text{total score}}$	First year	198.91	3.81	0.43
	Second year	194.12		
	Third year	177.78		
	Fourth year	187.32		
	Other	222.46		
Empathic Concern (EC)	First year	185.33	3.24	0.51
	Second year	201.03		
	Third year	180.92		
	Fourth year	206.81		
	Other	191.75		
Perspective Taking (PT)	First year	196.68	2.35	0.67
	Second year	190.16		
	Third year	180.06		
	Fourth year	195.66		
	Other	214.54		
Fantasy Scale (FS)	First year	201.73	4.90	0.29
	Second year	202.67		
	Third year	178.88		
	Fourth year	176.42		
	Other	206.52		
Personal Distress (PD)	First year	198.81	6.32	0.17
	Second year	191.45		
	Third year	195.37		
	Fourth year	169.16		
	Other	227.25		

Table 0.16 Comparative analysis of empathy scale scores between groups based on podcast listening habit

Empathy measure	**Podcast listening habit**	**Mean rank**	**U**	**Z**	***P***
IRI $_{\text{total score}}$	Once a week	200.38	16660.00	-1.18	0.23
	Once a month	186.88			
Empathic Concern (EC)	Once a week	204.86	15943.00	-1.88	0.06
	Once a month	183.67			
Perspective Taking (PT)	Once a week	190.98	17676.50	-0.23	0.81
	Once a month	193.59			
Fantasy Scale (FS)	Once a week	192.94	17849.00	-0.06	0.94
	Once a month	192.18			
Personal Distress (PD)	Once a week	195.22	17485.50	-0.41	0.67
	Once a month	190.56			

U: (the Mann-Whitney U test determines if two groups are significantly different from one another).

Z: (the Z-score can be used to determine if the difference is indeed significant when the sample sizes are sufficient.

Table 0.17 Comparative analysis of empathy scale scores between groups based on Arabic podcast listener

Empathy measure	**Listen to Arabic podcast**	**Mean rank**	**U**	**Z**	***P***
IRI $_{\text{total score}}$	yes	193.79	13550.50	-0.39	0.69
	no	188.70			
Empathic Concern (EC)	yes	192.89	13806.50	-0.122	0.90
	no	191.34			
Perspective Taking (PT)	yes	187.95	12612.50	-1.41	0.15
	no	205.97			

Fantasy Scale (FS)	yes	198.70	12141.50	-1.92	0.05
	no	174.17			
Personal Distress (PD)	yes	195.72	12994.00	-1.00	0.31
	no	182.96			

U: (the Mann-Whitney U test determines if two groups are significantly different from one another).

Z: (the Z-score can be used to determine if the difference is indeed significant when the sample sizes are sufficient.

Table 4.18 Comparative analysis of empathy scale scores between groups based on Nafs podcast familiarity

Variables	**Nafs podcast familiarity**	**Mean rank**	**Kruskal-Wallis H**	***P***
IRI $_{\text{total score}}$	Yes	189.11	0.76	0.68
	No	194.84		
	Not sure	175.84		
Empathic Concern (EC)	Yes	208.66	2.04	0.36
	No	188.92		
	Not sure	185.44		
Perspective Taking (PT)	Yes	182.74	0.75	0.68
	No	195.01		
	Not sure	192.82		
Fantasy Scale (FS)	Yes	178.55	3.52	0.171
	No	198.43		
	Not sure	166.18		
Personal Distress (PD)	Yes	191.31	0.67	0.71
	No	194.26		
	Not sure	175.92		

The researcher analyzed the relationship between perceived emotions from the audio clips and participants' empathy level across the Interpersonal Reactivity Index (IRI) subscales (Empathic Concern, Perspective Taking, Fantasy, and Personal Distress) and the total IRI score. One-way ANOVA was conducted with the emotion category as the independent factor. No statistically significant difference was obtained across emotion categories for any subscale or total IRI score. More specifically, Empathic Concern; $F(6, 13{,}817) = 0.226$ with $p > .05$; Perspective Taking; $F(6, 13{,}817) = 0.750$, $p > .05$; Fantasy; $F(6, 13{,}817) = 0.685$, $p > .05$; Personal Distress; $F(6, 13{,}817) = 0.673$, $p > .05$; and total IRI score, $F(6, 13{,}817) = 0.518$, $p > .05$. Post hoc Games-Howell pairwise comparisons further questioned the presence of any differences between emotion categories.

From descriptive statistics, all emotions gave fairly similar mean scores on the scales. For example, Empathic Concern ranged from 23.63 for anger to 23.76 for happiness; Perspective Taking, 23.72 for disgust to 23.93 for sadness; Fantasy, 23.81 for anger to 23.98 for sadness; and Personal Distress, 23.67 for anger to 23.88 for happiness. The total IRI score narrowly ranged from 94.85 for anger to 95.51 for sadness. Such slight differences serve as a reiteration to the results that found no significant difference between the mean scores for each emotion (see Table 4.19).

The estimation of effect sizes (η^2 values near zero) tells us that the kind of emotion perceived had practically no impact on the participants' empathy scores. This supports the idea of empathy being exerted consistently, regardless of whether the participants were evaluating anger, sadness, or other emotions analyzed.

Table 4.19 Analysis of variance (ANOVA) for Interpersonal Reactivity Index (IRI) subscales across different emotions

Measure	Anger		Disgust		Fear		Happiness		Neutral		Sadness		Surprise		*F*(6, 13817)	η^2
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
(EC)	23.63	3.62	23.69	3.83	23.72	3.63	23.76	3.55	23.73	3.63	23.75	3.55	23.74	3.67	0.226***	0.0001
(PT)	23.75	3.7	23.72	3.59	23.81	3.59	23.83	3.44	23.79	3.4	23.93	3.38	23.83	3.55	0.750***	0.0003
(FS)	23.81	3.84	23.82	3.89	23.90	3.87	23.90	3.64	23.82	3.66	23.98	3.64	23.91	3.86	0.685***	0.0003
(PD)	23.67	3.9	23.74	3.88	23.83	3.66	23.88	3.61	23.77	3.68	23.85	3.65	23.79	3.72		

															0.673***	0.003
Total	94.85	14.27	94.97	14.38	95.26	13.96	95.38	13.4	95.11	13.52	95.51	13.38	95.27	14.01	0.518***	0.002

$^{***}p > .05$

Furthermore, the researcher checked if the gender of the clip speaker had any influence on participants' empathy ratings across the IRI subscales. Chi-square analyses yielded no significant gender-based difference for Empathic Concern with male speakers ($\chi^2(78) = 78.71$, $p = .456$, Cramer's $V = .08$) or female speakers ($\chi^2(78) = 93.91$, $p = .106$, Cramer's $V = .09$). A combined dataset assessed in regard to speaker-gender collapsed, conversely, showed a significant albeit weak association on the EC scale ($\chi^2(78) = 131.21$, $p < .001$, Cramer's $V = .10$). Regarding Perspective Taking (PT), Fantasy Scale (FS), Personal Distress (PD), and IRI total score, significant associations were encountered for numerous gender categories, especially for the combined set (all with a p-value < .001), though Cramer's V effect-size measures indicated that these relationships were weak. Thus, there seem to be gender-associated distributional shifts in empathy ratings, but their effect sizes are ultimately weak (see Table 0.20).

Although a number of chi-square tests showed statistical significance, with values of Cramér's V being ≤ 0.17, they indicate weak associations. This implies that speaker gender only very slightly influenced the distribution of ratings of empathy, and practical application of such a relationship could be considered insignificant.

Table 0.20 Chi-Square tests for the association between clip speaker gender and empathy ratings across emotions

IRI Dimension	**Speaker Gender**	**χ^2**	**df**	***P***	**Cramer's V**	**N**
Empathic Concern (EC)	Male	78.71	78	.456	0.08	6144
	Female	93.91	78	.106	0.09	7680
	Combined	131.21	78	.000	0.1	13824

Perspective Taking (PT)	Male	101.96	78	.036	0.1	6144
	Female	124.98	78	.001	0.11	7680
	Combined	180.42	78	.000	0.12	13824
Fantasy Scale (FS)	Male	93.2	84	.231	0.09	6144
	Female	113.98	84	.016	0.1	7680
	Combined	152.24	84	.000	0.11	13824
Personal Distress (PD)	Male	151.17	90	.000	0.13	6144
	Female	199.51	90	.000	0.15	7680
	Combined	261.64	90	.000	0.16	13824
IRI Total Score	Male	273.33	162	.000	0.14	6144
	Female	279.72	162	.000	0.15	7680
	Combined	452.3	162	.000	0.17	13824

Note. Pearson's Chi-Square tests were used. Effect sizes are reported as Cramer's V, interpreted as weak association (< .20).

Base on the statistical analysis of empathy, the following section posits the Brunswik Lens Model to describe how experts and participants perceive vocal cues. Such a theoretical perspective thus explains why identical acoustic information may be weighted differently depending on the perceiver's expertise and experience.

4.3.3 Research Question 3

This section considers the third research question, that is explore how the Brunswik lens model of social judgment theory applies to the relationship between vocal cues and emotion recognition. The analysis considers how participants use the available auditory information to judge emotional states and whether such judgments agree with expert interpretations. The findings shed a light to the extent of agreement and disagreement in the use of vocal cues by participants and suggest how decision criteria are formed underlying emotion perception.

RQ3. How does the lens model of social judgment theory explain the relationship between vocal cues and emotion recognition?

The cross-tabulation analysis for the concordance between the judgments by experts and those by participants for perceived emotion had 13,932 valid cases, consisting of 108 expert judgments (36 clips rated by three different experts) and 13,824 participant ratings. Expert assessments were distributed mainly among sadness (n = 23, 21.3%), happiness (n = 25, 23.1%), and neutral (n = 17, 15.7%), with less frequency for anger (n = 17, 15.7%), disgust (n = 12, 11.1%), and fear (n = 3, 2.8%). In contrast, the rating patterns by participants showed sadness (n = 3,501, 25.3%), happiness (n = 3,458, 25.0%), neutral (n = 2,586, 18.7%) emotions most typically chosen by them, followed by fear (n = 1,129, 8.2%), anger (n = 1,103, 8.0%), and disgust (n = 880, 6.4%) (see Table 4.21).

Participants being exposed to the stimuli did not show a statistically significant relationship with the types of emotions perceived ($\chi^2(6) = 17.36$, $p = .008$). However, the Kappa coefficient did not determine any agreement between the two categories in their ratings: expert and participant interpretations ($\kappa = 0.000$, $p = .411$).

The chi-square test showed a statistically significant association between the participant types and perceived emotion; nevertheless, the very low Kappa coefficient suggests practically no agreement between the expert and participant judgments. This essentially means that while distributions differ from a statistical vantage, the level of agreement is so low that participant interpretations were far from the kind of interpretation the experts set.

Such lack of agreement indicates that participant interpretations of the emotional content of the audio clips diverged from expert judgments beyond the level of chance. This leads to the conclusion that the discrepancy between expert and non-expert emotion recognition exemplifies just how differently each group processes and interprets emotions.

Table 4.21 cross-tabulation of participant type by perceived emotion: frequencies and percentages

Emotion	Experts (n)	Experts (%)	Participants (n)	Participants (%)	Total (n)
Anger	17	15.7	1103	8.0	1120

Disgust	12	11.1	880	6.4	892
Fear	3	2.8	1129	8.2	1132
Happiness	25	23.1	3458	25.0	3483
Neutral	17	15.7	2586	18.7	2603
Sadness	23	21.3	3501	25.3	3524
Surprise	11	10.2	1178	8.5	1189
Total	108		13835		13943

Chi-Square: $\chi^2(6) = 17.36$, $p = .008$

Cohen's Kappa: $\kappa = 0.000$, $p = .411$

Note. Data represent combined cross-tabulations of participant type by perceived emotion across all datasets. Pearson's Chi-Square tests were used to examine associations, and Cohen's Kappa assessed inter-rater agreement.

As presented in the results of research question 1 (see Table 4.2), the analyses show that participants and experts differed remarkably in the prioritization of vocal cues that they depends on for emotion recognition. Experts focused mostly on tone of voice and pitch, while the other participants focused on pitch and rhythm, but also considered a wider array of cues such as pauses and changes in speech rate. From the perspective of the Brunswik Lens Model, these differences show how the same distal stimuli like the acoustic parameters that embedded in the audio clips, could be differently weighted and interpreted, depending on the perceiver's unique experiential background, cognitive framework, and perceptual strategies .

The model describes perceivers as judging by means of a number of probabilistic cues, but the lens through, which these cues are interpreted can differ between perceivers, resulting in variation in judgment accuracy and consistency. In this study, this distinction between experts and participants in cue use would then correspond to a discrepancy between cue validity (the correlation of a cue with an actual emotion) and cue utilization (the extent of use of a certain cue by a perceiver).

The low agreement as analyzed by Cohen's Kappa method seems to suggest that both groups may have used overlapping sets of cues but differed in how they weighted and integrated them, which is in line with the Brunswik framework's conception of the role of individual perceptual pathways in judgment formation.

Having stated the disparity between the groups of experts and participants, this analysis now goes further to question whether this recognition of emotions is shaped by gender differences. Thus, research question four extends this investigation to demographic influences, testing whether male and female listeners systematically differ in their empathy and cue utilization.

4.3.4 Research Question 4

This section presents the findings for the fourth research question, which ask whether considerable gender differences exist between women and men in emotion recognition. To answer this question, male and female subjects' scores were compared regarding empathy levels, perceptual strategies, and efficiency in identifying emotions expressed in different vocal media contents. Statistical comparison and cross-tabulation were used to identify if gender bias exists on performance.

RQ4. Is there a significant difference between women and men in emotion recognition accuracy?

The comparison of the empathy scores among various demographic groups and behaviors yielded some important conclusions. The gender differences (see Table 4.22) appeared to be the highest. Males scored significantly higher than females on total IRI scores ($U = 14571.00$, $Z = -3.57$, $p < 0.001$), with mean ranks of 212.39 for males and 172.19 for females. On the contrary, females scored significantly higher than males on perspective taking or personal distress. Differences between the two genders could be seen in perspective taking ($U = 14980.50$, $Z = -3.24$, $p < 0.001$) and personal distress ($U = 16045.50$, $Z = -2.24$, $p = 0.02$). The fantasy subscale was close to significance ($U = 16386.00$, $Z = -1.91$, $p = 0.05$), whereas empathic concern showed no meaningful difference ($p = 0.08$).

Gender-differentiated findings accord with small effect sizes, indicating that although relevant to statistics, the extent to which gender may influence empathy is limited. This implies emotion recognition to be largely similar among both genders.

Table 4.22 Comparative analysis of empathy scale scores between groups according to gender

Empathy measure	**Gender**	**Mean rank**	**U**	**Z**	***P***
IRI $_{\text{total score}}$	Male	212.39	14571.00	-3.57	0.00
	Female	172.19			
Empathic Concern (EC)	Male	182.86	16599.00	-1.71	0.08
	Female	201.94			
Perspective Taking (PT)	Male	174.34	14980.50	-3.24	0.00
	Female	210.28			
Fantasy Scale (FS)	Male	181.74	16386.00	-1.91	0.05
	Female	203.04			
Personal Distress (PD)	Male	179.95	16045.50	-2.24	0.02
	Female	204.79			

Furthermore, the researcher checked if the gender of the clip speaker had any influence on participants' empathy ratings across the IRI subscales. Chi-square analyses yielded no significant gender-based difference for Empathic Concern with male speakers ($\chi^2(78) = 78.71$, $p = .456$, Cramer's $V = .08$) or female speakers ($\chi^2(78) = 93.91$, $p = .106$, Cramer's $V = .09$). A combined dataset assessed in regard to speaker-gender collapsed, conversely, showed a significant albeit weak association on the EC scale ($\chi^2(78) = 131.21$, $p < .001$, Cramer's $V = .10$). Regarding Perspective Taking (PT), Fantasy Scale (FS), Personal Distress (PD), and IRI total score, significant associations were encountered for numerous gender categories, especially for the combined set (all with a p-value < .001), though Cramer's V effect-size measures indicated that these relationships were weak. Thus, there seem to be gender-associated distributional shifts in empathy ratings, but their effect sizes are ultimately weak (see Table 4.23).

Although statistical differences exist in several comparisons, the effect sizes remain weak (Cramér's V = 0.08 to 0.17). This means that speaker gender can explain only a minor percent of variability in empathy ratings, hence having very limited practical significance.

Table 4.23 Chi-Square tests for the association between clip speaker gender and empathy ratings across emotions

IRI Dimension	**Speaker Gender**	**χ^2**	**df**	***P***	**Cramer's V**	**N**
Empathic Concern (EC)	Male	78.71	78	.456	0.08	6144
	Female	93.91	78	.106	0.09	7680
	Combined	131.21	78	.000	0.1	13824
Perspective Taking (PT)	Male	101.96	78	.036	0.1	6144
	Female	124.98	78	.001	0.11	7680
	Combined	180.42	78	.000	0.12	13824
Fantasy Scale (FS)	Male	93.2	84	.231	0.09	6144
	Female	113.98	84	.016	0.1	7680
	Combined	152.24	84	.000	0.11	13824
Personal Distress (PD)	Male	151.17	90	.000	0.13	6144
	Female	199.51	90	.000	0.15	7680
	Combined	261.64	90	.000	0.16	13824
IRI Total Score	Male	273.33	162	.000	0.14	6144
	Female	279.72	162	.000	0.15	7680
	Combined	452.3	162	.000	0.17	13824

Note. Pearson's Chi-Square tests were used. Effect sizes are reported as Cramer's V, interpreted as weak association(< .20).

Turning away from the quantitative findings that addressed in the previous four research questions, the floowing section discusses and interprets the participants' qualitative reflections that found in the study's open-ended questions. This analysis complements the previous one and provides a more thorough understanding of how lived experience, cultural frameworks, and listening practices together shape processes of perceiving emotion in Levantine-Arabic vocal content.

4.4 Results of open-ended questions

The researcher performed inductive thematic analysis on the open-ended responses through coding, categorizing, and developing themes as explained in Chapter Three. This method helped with looking for recurring patterns in the reflections that provided by the participants and strengthened the grounding of the findings in the words of the participants themselves. After listening to each of the 36 audio clips and completing the emotion recognition and vocal cue selection tasks, participants responded to three open-ended questions:

6. Did any part of the speaker's message remind you of something familiar? If yes, please describe.
7. Are there any personal experience or events that might influence your interpretation of the speaker message? If yes, please describe.
8. Did you notice any differences between the speaker tone and the content of their message? If yes, please describe.

Participants responses were analyzed using inductive thematic analysis, which generated three primary themes as appear in Figure 4.3:

1- Familiarity and Associations
2- Personal Experiences
3- Tone Content Discrepancy

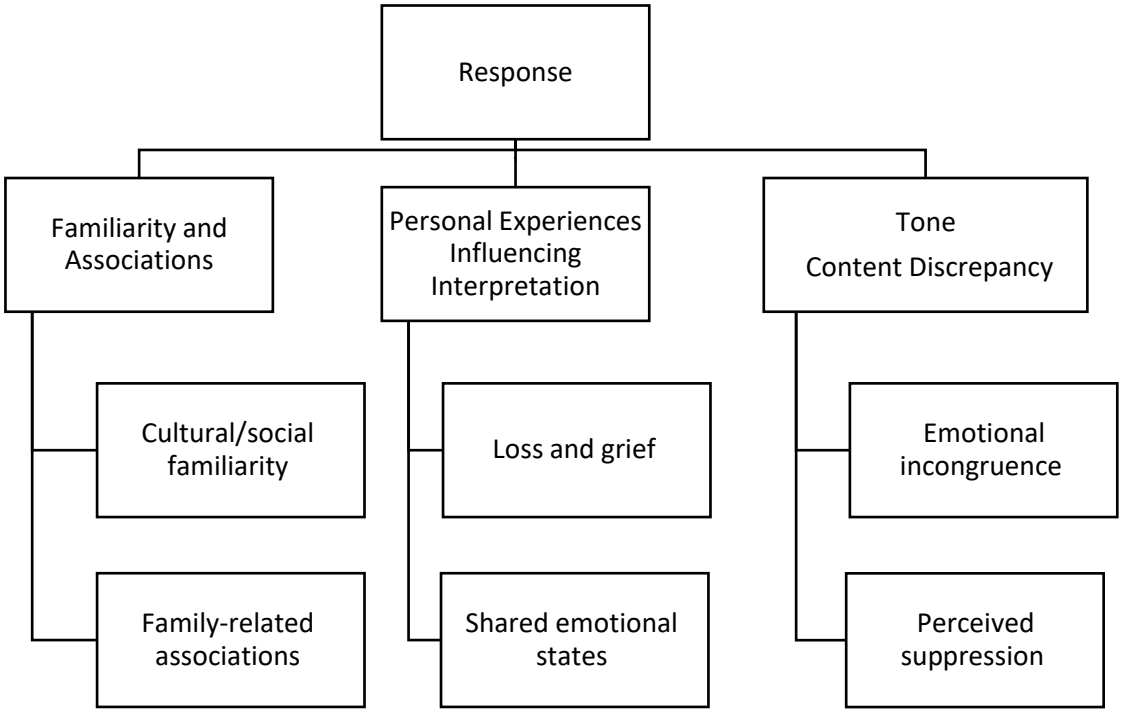

Figure 4.3 Coding Tree of Thematic Analysis of participants answers on open-ended questions

4.4.1 Theme 1: Familiarity and associations

This theme (n = 11; 6.7%) captured responses in which participants linked the message or the tone of voice to familiar contexts, memories, or personal associations. Many of these involved family relationships or emotionally significant figures. Participants often described how the speaker's tone evoked intimate or nostalgic feelings:

"تذكرت اشتياقي لايام ودفئ العائلة" (Participant 1, female)

"تذكرت انجازاتي"(Participant 33, female)

"ذكرياتي مع اخوي" (Participant 27, female)

"ذكرياتي مع ابوي لما سمعت المقطع اتذكرت حنانه" (Participant 35, female)

"اتذكرت ابوي اللي مات لحد الان بسمع صوت أبوي بالبيت"(Participant 213, male)

"المقطع الصوتي ذكرني بمرحلة من المراحل كرهي لمنطقة معينة واشمئزاز منها "(Participant 1, female)

Some responses referred to shared cultural or social experiences, where the clips brought to mind common collective memories:

"العائلة والبعد عنهم" (Participant 50, female)

"البنت عوره بدولنا العربية اما في الخارج حرية زياده" (Participant 16, female)

This pattern suggests that participants were not merely recognizing emotions cognitively but were also connecting the speaker's delivery to their own autobiographical memories, deepening their emotional engagement with the message.

4.4.2 Theme 2: Personal experiences influencing interpretation

This theme (n = 7; 4.2%) included participants who explicitly noted that their personal life experiences shaped how they interpreted the emotional content of the clips. Many of these involved experiences of loss, grief, or trauma, which seemed to sensitize participants to particular emotional tones:

"عشت الفقدان سنتين حاسه بفراغ بقلبي وعقلي بعد موت والدي" (Participant 48, female)

"فكرت بجميع اللحظات اللي عشتها مع بابا "(Participant 49, female)

"الشوق للذكريات اللي عشتها" (Participant 49, male)

In other cases, participants acknowledged empathizing based on shared emotional states:

"تذكرت لحظة دخولي ع الجامعة والتعرف ع اصدقاء من مناطق مختلفة"(Participant 162, female)

"تذكرت لحظة افتقادي لأبي المتوفي في حفلة تخرجي من الجامعة"(Participant 1, female)

‘‘صاحبتي بس ماتت "(Participant 16, female)

"تبادر لذهني عند الانتهاء أهل غزة خلال الحرب الذين يشتاقون للماضي بسبب ويلات الحرب"

(Participant 198, male)

These responses highlight the intersection between listeners' personal histories and their emotional interpretations, reinforcing the idea that emotional recognition in speech is not purely perceptual but also shaped by individual narratives and lived experiences.

4.4.3 Theme 3: Tone content discrepancy

This theme (n = 4; 2.4%) described cases where participants perceived a mismatch between the speaker's tone and the semantic content of the message. This awareness of emotional incongruence suggests a higher-level processing of emotional communication:

"النبرة بتعبر عن اشي ومحتوى الرسالة عن شيء اخر"(Participant 10, female)

‘‘يبدو ان ما يفتقده المتكلم مختلف عن ما يتوقع من نفسه ان يفتقده ... " (Participant 199, male)

"نبرة متحفظة ومحاولة كبت" (Participant 201, male)

Such observations imply that some participants were attuned to subtle emotional cues, possibly indicating masking, emotional suppression, or incongruence between what was said and how it was said. However, the identified themes underscore several key insights:

1- Personal and cultural resonance: Clips frequently triggered personal memories or familiar social contexts, reinforcing the role of emotional association in perception.

2- Lived experience as a lens: Participants' interpretations were often filtered through their personal histories, especially those involving grief or emotional hardship.
3- Sensitivity to emotional nuance: A subset of participants identified discrepancies between tone and content, showing deeper engagement with the emotional dynamics of speech.

In summary, the qualitative results therefore complement the statistical analyses by indicating that the recognition of emotion is not merely a perceptual process but is formed on either side by the individual histories of the participants themselves or by the cultural context. The use of direct quotes elucidates how these connections grow deep in lived experiences with the interpreting of emotions.

Chapter Five: Discussion

5.1 Introduction

This chapter presents a critical analysis of the study findings in relation to its research questions, objectives, and theoretical framework. Chapter 4 presented the results, while this chapter interprets them in the wider field of communication and media studies. The discussion goes beyond description to show how the findings fit with or challenge existing theories and studies.

This study investigates the semiotic role of audience empathy in the process of recognizing emotions from Levantine-Arabic vocal media content. Specifically, it examined how distinct vocal cues, such as tone, pitch, volume, speed, and intonation interacts with listeners' empathic dispositions to influence their ability to identify emotional states accurately.

This study used Brunswik's Lens Model (1952), which sees emotion recognition as a process where people judge emotions from imperfect but useful cues. The study also applied semiotic theory to examine vocal cues as cultural signs, and empathy theory to explore how empathy shapes interpretation.

5.2 Discussion of Findings

This section is divided into four subsections in order to discuss the study's findings in light of the research questions that posed in Chapter 1, by integrating them with the theoretical framework and existing literature.

The researcher treated and analyzed each research question separately in order to focus in more depth on the results, explaining how the results support previous literature and interpretations. Thus, the outcome of this section is to draw the results in the frame of a broader communication and media studies discourse, particularly those in the fields of semiotics, empathy and vocal emotion in media studies.

5.2.1 Research Question 1

How do vocal cues (tone, pitch, volume, speed, intonation) influence the recognition of emotions in Levantine-Arabic vocal media content?

The study's findings assert that vocal cues are of paramount importance in the method of recognizing emotions within Levantine-Arabic vocal media content, thus affirming their role as key semiotic elements used in decoding emotional states. Thus, the participants highly relied on paralinguistic cues such as tone, pitch, and intonation to identify emotions, especially for feelings of anger and happiness, which inherently carry clear acoustic markings. Up to now, the findings agree with Brunswik's lens model theory, which suggests that people make judgments of unobservable states (i.e., emotions) based on their interpretation of cues with various degrees of accuracy. In this given setting, voices constituted the distal cues filtered by listeners through their perceptual and cognitive frameworks into proximal judgments, namely emotion labels.

Interestingly, tone and pitch emerged as the most salient cues for the participants, a finding in agreement with Scherer's (2003) research into the acoustic correlates of emotion. High pitch and relatively fast speech rate were commonly associated with emotions of happiness and surprise, while low pitch and slow tempo were marked as signals for sadness; these were general vocal patterns that were recognized in cross-cultural studies of vocal emotion recognition.

This alignment appears to confirm Scherer's notion of the universality of acoustic markers of sadness, and it adds to his findings by proving the validity of such cues in the Arabic dialectal context-the context his study has not addressed. Meanwhile, the findings show a partial divergence: rather frequently, participants combined pitch and rhythm instead of using pitch alone. Theoretically, this suggests that Scherer's model could require to adapt cultural variations, whereas it practically points to the necessity of incorporating culturally relevant acoustic features in an emotion recognition technology.

Intonation, mainly rising and falling, guided the emotional perception of the participants, partly justifying the semiotic position that prosody carries culturally embedded meanings of emotion, a stance much more relevant for Levantine Arabic, where a multitude of possibilities for prosodic realization make emotional interpretations all the more challenging.

The findings emphasize the contextual nature of the vocal-cue interpretation processes. While some particular acoustic patterns are universally linked to specific emotions, participants' qualitative reflections seem to imply that their interpretations were strongly

driven by their cultural and linguistic familiarity with Levantine Arabic. This is further evidence supporting the Elfenbein and Ambady (2002) "in-group advantage" idea, which posits that people generally have better accuracy in decoding emotions posed within the cultural boundaries of their society. The findings show how cultural familiarity would amplify the vocal cues that are very subtle in Arabic Levantine, unlike cross-cultural data that suggest less recognition.

Yet, the variability observed in the recognition rate across different emotions would suggest some vocal cues are by nature more ambiguous. Thus, in the case of recognition, fear and disgust emotions would perhaps be less accurately recognized, either because their acoustic features overlap or due to very subtle vocal expressions. This finding echoes that of Juslin and Laukka (2003), who confirm that generally, basic emotions are recognizable through vocal cues, but it is the more complex or low-intensity emotions that listeners find difficult.

Thus, the results clearly stand behind the importance of vocal cues as semiotic markers in the recognition of emotions from Levantine-Arabic media content. This relationship between the universally shared associations between acoustics and emotions, on the one hand, and culture-specific decoding, on the other, underscores that any approach to vocal emotion recognition must consider cultural variations.

This study reveals that the vocal cues of emotion, often assumed universal, are in fact deeply shaped by cultural and linguistic context. For the field of media and communication, this indicates a need to move beyond cross-cultural generalizations and to re-evaluate prosodic universals through localized analyses. By situating the Arabic language within this debate, the study sharpens and expanding scholars understanding of how universality, cultural specificity and context intersect in vocal emotion research.

5.2.2 Research Question 2

What is the semiotic function of empathy in the process of recognizing emotions from Levantine-Arabic vocal media content?

In general, the findings of this study support the manner in which empathy acts as a mediating factor in recognizing emotions in a so-called semiotic process in vocal communication. Participants who had higher IRI levels, especially on the Empathic

Concern (EC) and Perspective Taking (PT) dimensions, tended to score higher in identifying the emotions expressed in the audio clips. It suggests that empathy plays a determinative role in assisting a listener to decode the vocal cues of semiotic "signs" bridging the gap between the acoustic signals and the intended emotional states of the speakers.

The outcome is consistent with the Perception-Action Model of empathy (Preston & de Waal, 2002), which states that perceiving another's emotional state in fact triggers similar cognitive and affective processes in the perceiver and promotes emotional understanding. This finding supports the Perception-Action Model (Preston & de Waal, 2002), which suggests that empathy allows for the simulation of emotional cues. The present study instead looked at audio-only types of media content such as podcasts-a rarely studied area in the literature, so theoretically, it extends the empathy models to situations away from a face-to-face arrangement; practically, it poses an opportunity to design audio content that motivates stronger empathic responses among media audiences.

The study findings suggest that in this case, empathy itself serves as a psychological "lens," accentuating the listener's ability to interpret vocal cues, in line with Brunswikian principles. It also resonates well with the emphasis placed by Decety and Jackson (2004) that empathy equips one to frame emotional expressions within personal, cultural, and situational knowledge, all of which come into play in decoding the very-fine prosodic variations of Levantine Arabic.

Qualitative data offered further proof on the semiotic role of empathy. Many would refer to personal experience or emotional resonance with the speakers to enrich their perception of ambiguous or subtle emotional cues. This highlights the twofold nature of empathy: the cognitive perspective-taking and inferential dimension, accompanied by emotional resonance. This twofold nature may facilitate the complex process of understanding complex emotions that may not be easily characterized through acoustic features.

This finding also accords with the "in-group advantage" in recognizing emotions (Elfenbein & Ambady, 2002), suggesting that empathy assists decoding accuracy, contingent on cultural familiarity. In Levantine Arabic media contexts, where emotional expressions may possess culturally tinted connotations, it is the listeners endowed with empathy that seem more capable of decoding such cues effectively.

Nonetheless, one noteworthy exception was that empathy did not seem to promote recognition across all emotion categories. In instances of recognizing sad or fear situations, accuracy is increased since these emotions are acknowledged by the use of subtle vocal cues; unlike more acoustically clear emotions such as anger and happiness. Where the effect is not so pronounced. This finding might mean that empathy is more involved in identifying low-intensity, ambiguous emotions, where the listener relies on contextual and affective understanding beyond the raw acoustic signal.

In summary, these findings position empathy as a semiotic mechanism and psychological processes by which the production of meaning from vocal emotion can be enriched. It operates both as a personal characteristic and as a cultural-interpretative construct that enriches understanding regarding the way audiences interact with media content on the basis of emotion and meaning.

The distinctive contribution for this study lies in shifting empathy research from face-to-face settings into mediated, audio-only contexts such as podcasts. This is particularly important because it highlights empathy not only as a psychological disposition but also as a communicative lens that reshapes how audience experience the media content. thus, the study enhances the field of communication and media studies by integrating empathy theory with media studies, creating a more interdisciplinary account of how audience interpret emotions from vocal content alone.

5.2.3 Research Question 3

How does Brunswik's Lens Model explain the relationship between vocal cues and emotion recognition?

The results of the study have given empirical evidence to apply Brunswik's lens model to explain how emotions are decoded from vocal media content by the audience. The theory states that listeners interpret distal cues, in this case vocal features such as tone, pitch, volume, speed, intonation, through a probabilistic judgment process to form proximal judgments, i.e. emotional inferences about the speaker. The present results are hence in line with this theoretical proposition by demonstrating that the participants used such cues selectively and contextually when perceiving emotions from the Levantine-Arabic vocal content.

Brunswik’s (1952) distinction between cue validity and cue utilization is particularly relevant here, while certain vocal features reliably signaled emotions, participants varied in how they weighted and applied these cues. The present findings therefore illustrate both the strengths and the limitations of the Lens Model when applied in a non-Western linguistic setting.

Emotional judgment involves higher-dimensional cue integration whereas higher-dimensional reflection is intuitive, according to the qualitative comments of participants. Most participants said pitch and tone would be used primarily for the interpretation of emotions such as anger and happiness, yet often they would contextualize the interpretation with speech speed and pauses; thus, the interpretation would be refined. This layering of cues is exactly what supports Hammond's (1996) view of the Lens Model as a conceptual framework for multidimensional cue integration-a synthesis of available signals is used to make judgments rather than a direct one-to-one mapping of cues to emotions.

That emphasis on uncertainty also proved important. Many respondents described difficulty distinguishing between some emotions, including fear and sadness, suggesting an inherent ambiguity of some vocal signals. This observation corresponds with Scherer's (2003) remark that vocal cues tend to overlap across emotional categories, forcing the listener to probabilistically rather than deterministically assess the situation. The Lens Model helps explain the phenomenon by placing the recognition of an emotion as an inferential process colored by the listener's cognitive and emotional filter, such as empathy, cultural familiarity, and prior experiences.

More importantly, this study also highlights the influence of cultural context in altering the cue-judgment relationship described by the Lens Model. Since participants probably knew about Levantine Arabic prosody and culturally specific ways of expressing emotion, they could quite accurately recognize certain emotions. This could be aligned with Elfenbein and Ambady's (2002) "cultural advantage" and could extend the Lens Model to show that cultural schemata may fine-tune or override the interpretation of vocal cues.

In a complementary manner, the interaction of empathy with the interpretation of cues also demonstrates the working of individual differences mediate the process. People who

were judged more empathetic considered subtler prosodic cues and gave richer contextual explanations for their judgments. This suggests that the Lens Model can be extended theoretically to include dispositional filters such as empathy, which foster the interpretative process.

The researcher suggests that Brunswik's Lens Model could be used as a good explanatory framework for the way listeners process the complexity of signals in the voice so as to infer emotions therefrom. The study, therefore, reaffirmed the model in media communications research while suggesting that cultural and psychological facets such as empathy and in-group familiarity should be considered when applying it in non-Western settings.

This support aligns with Brunswik's claim that judgments are better when several indicators are integrated yet conversely highlights a crucial limitation: Even with several cues available, the emotions of fear and sadness remained difficult to distinguish. In theory, this suggests that the Lens Model may incorporate cultural and linguistic dimensions while practically underscoring the difficulties in designing automated systems in an Arabic environment capable of distinguishing closely related emotions.

The contribution of this study is to show that, while Brunswik's Lens Model retains some explanation, it does not fully capture the interaction of culture and language in perceptual judgments. For researchers, this would signify that probabilistic cue-integration models would have to be reformulated to include sociolinguistic dimensions. Hence, this reframing enhances the framework by providing a more inclusive approach to explain emotion recognition both in Western and non-Western settings.

5.2.4 Research Question 4

Are there significant gender differences in emotion recognition accuracy?

The findings from this study show that certain differences exist between men and women with respect to emotion-recognition accuracy-and-with women scoring higher than men do over almost all categories of emotions. This agrees with other studies showing that women usually perform better in recognizing emotions, especially if there is spotting of subdued or low-intensity emotional displays (Sen, Isaacowitz, & Schirmer, 2017) (Dair, Donovan, & O'Reilly, 2021). In the current study, it is apparent from the data that female

participants were more sensitive to prosodic characteristics. These variations in pitch and intonation help distinguish very subtle emotions of sadness and fear.

The Lens Model by Brunswk provides an explanation for the differences. As described in the previous research question discussion, emotion recognition involves the integration of several distal cues to generate corresponding proximal judgments. Higher accuracy of female subjects suggests that females may utilize cues that are more distal or a more refined set of cues to interpret emotional content, thereby evidencing heightened perceptual sensitivity and more cognitive flexibility in processing paralinguistic stimuli. This interpretation is consistent with Lausen and Schacht (2018), who found that female listeners allocate more attentional resources to vocal emotional cues, resulting in superior decoding performance.

Similarly, empathy arises as an intervening variable in explaining the sex difference in emotion recognition. In this study, women obtained superior scores than men on the Interpersonal Reactivity Index (IRI), mainly on the Empathic Concern and Perspective Taking subscales, which correlated positively with emotion recognition accuracy. This finding supports Israelashvili et al.'s (2020)assertion that women could partly outshine men in recognizing emotions because of their heightened empathic ability, especially when facing complex situations or those embedded within a specific culture. In this respect, empathy operates as a sort of cognitive-emotive filter within the Lens Model by granting more refined perspectives, when working with ambiguous or overlapping cues.

However, this said, it is interesting to note that the gender gap in recognition accuracy was less pronounced for emotions that possess distinct and rather acoustically salient markers such as anger and happiness. This points toward gender differences becoming more significant while interpreting emotions whose cue signals are rather subtle and context-dependent, thus shedding more light on why women are said to have an edge in processing complex emotional information (Matsumoto, Yoo, & Nakagawa, 2008).

Another factor moderating these gender differences is culture. In Levantine-Arabic media content, emotional prosody could carry a gender-specific connotation as prescribed by culture-specific norms of expression. Female participants could have gained more interpretive advantage either by virtue of their cultural familiarity or through exposure to the emotional nuances in spoken Levantine Arabic. This observation goes to prove the

"cultural advantage" framework (Elfenbein & Ambady, 2002), which supports the premise that gender and cultural familiarity intersect to result in the emotional decoding ability of an individual.

In brief, the study reaffirmed the dominance of female talent over male talent in perception from Levantine-Arabic vocal media materials, especially for emotional states characterized by subtle or overlapping acoustic cues. These differences would be best interpreted under Brunswik's Lens Model that conceptualizes gender as one of the factors based on which cues are selected, weighted, and finally integrated into emotional judgment-a process mediated by empathy.

This result returns us partly to earlier findings as those of Sen (2017), but differs in its contextual explanation: data here represent a collectivist cultural sphere in which women may be more frequent auditory and social interactors. Thus, theoretically, the nuance suggests that gender differences are culture-bound rather than absolute; practically, it implies that an emotion recognition system might grow from gender-sensitive design, especially in the Arabic-speaking population.

The novelty of these findings is the identification of culturally contingent gendered differences, rather than universally fixed ones. This has an impact on the field by eroding essentialist assumptions and stressing the importance of context in the interpretation of gender effects in communication. The study thus enriches researchers understanding by revealing that gender interacts with cultural practices as one factor, enabling a specific basis for theory-building and applied design in emotion recognition systems.

5.3 Achievement of Research Objectives

This study was designed based on the objectives that narrated in Chapter One to explore the semiotic role of empathy in emotion recognition in Levantine-Arabic vocal media content. A review of the findings suggests that these objectives were more-or-less fulfilled and therefore provide new knowledge that joining semiotics, empathy, and emotion recognition in a non-Western media context.

1- Investigate vocal cues' role in the recognition of emotions.

The study proved to a great degree of certainty that vocal cues in the Levantine-Arabic media act as some primary semiotic indicators for emotional recognition. From the

quantitative side, it is clearly reflected that the participants were depending on the vocal cues to give them fairly good accuracy for the recognition of emotions, mainly anger, happiness, and sadness. On the other hand, Qualitative reflections, supplemented this finding by suggesting that the participants used a multiplicity of cues in their interpretation, which fits Brunswik's Lens Model, stating that judgments result from the probabilistic integration of signals from the environment.

2- Examine how empathy functions semiotically in vocal media communication.

The study found a significant positive association between participants' empathy levels as measured by the Interpersonal Reactivity Index (IRI) and emotion rating accuracy. Empathy was understood as a cognitive-emotional mechanism enhancing the perception of subtle or ambiguous vocal cues, most evidently with low-intensity emotions such as sadness and fear. These results support the view of empathic activity as a semiotic act, where listeners access referential emotional meanings embedded in vocal prosody by means of their affective resonance and perspective-taking ability.

3- Apply Brunswik's Lens Model to explain the relationship between vocal cues and emotion recognition.

The use of the Lens Model provided a strong framework to understand the study findings: It highlighted how participants listened to and combined vocal cues in making sadness judgments, taking into account participation individual filters, such as empathy, and cultural filters, such as familiarity with Levantine Arabic. The study thereby broadens the Lens Model perspective on media communication research applied to Arabic circumstances, emphasizing its applicability for explaining emotion recognition in culturally dependent situations.

4- Identify the impact of listener gender on emotion recognition accuracy.

Some examinations with respect to gender differences in recognition performance fulfilled this objective. Across most emotional categories, females obtain better recognition scores than do males, particularly for emotions with subtle differences or overlapping acoustic features. The authors attribute such a phenomenon to a greater level of empathy in females and potentially better sensitivity to prosodic contrasts. This is also in agreement with previously reported findings from other cross-cultural emotion studies.

All these so far reported results establish an interaction of gender, empathy, and cultural familiarity in emotional decoding ability.

In summary, the study has achieved the stated objectives and has advanced the understanding of how empathy and vocal cues work together in the recognition of emotions in Levantine-Arabic media content. For these findings, they contribute to the further theoretical development of Brunswik's Lens Model in media studies, as well as applying this knowledge in developing culturally adaptive communication strategies and emotion recognition technologies.

5.4 Comparison with Previous Studies

The present study compares and contrasts its findings with previously published works in the areas of emotion recognition, empathy, and vocal communication, presenting a culturally focused perspective on the Levantine-Arabic vocal media content.

In accordance with Scherer (2003) and Juslin & Laukka (2003), this study corroborates that vocal elements-pitch, timbre, and intonation-are essential for the recognition of emotional states. Similar to Western-centric studies, with high pitch and fast tempo voices often identifying happiness and surprise, slow-pitch, and slow-tempo voices often signify sadness in the Levant region. Nonetheless, the present study puts further weight on cultural familiarity effects in decoding vocal prosody. As explained by Elfenbein & Ambady (2002) in their "cultural advantage" hypothesis, in-group familiarity with Levantine-Arabic prosodic patterns might have, by large, contributed to the relatively great recognition accuracy, particularly for culturally nuanced emotions.

The results concur with the Perception-Action Model (Preston & de Waal, 2002) and with Decety and Jackson's (2004) paper that conceptualizes empathy as a cognitive and affective process enabling emotional understanding. High empathy enabled the participants to better interpret subtle or ambiguous emotions and hence empathic persons are more accurate at recognizing nuanced emotions (Israelashvili, Sauter, & Fischer, 2020). This study expands upon previous ones by describing empathy not merely as a psychological trait but also as a semiotic mechanism: listeners "decode" emotional signs embedded in culturally-ingrained vocal media content.

With the application of Brunswik's Lens Model, the study complements the precedent of social judgment studies and media research (Hammond, 1996); (Scherer, 2003). Each of these previous studies found that participants integrate several different cues in a

probabilistic manner when inferring a particular emotional state of a subject. The current research diverges to demonstrate how dispositional factors such as empathy and familiarity with the culture serve as attentional filters in the meaning attribution process. This in turn brings to light its applicability outside Western context and suggests a scope for its further development in the cross-cultural study of emotion recognition.

This study's findings were in line with Sen et al. (2017) and Dair et al. (2021), according to them, women perform better than men in processing emotional cues, especially when the cues are morphed onto a more subtle emotional state. This gender gap was the highest for those subtle emotions such as fear or sadness accordingly agreed with Lausen & Schacht (2018), who suggested that female listeners could be more sensitive to prosodic variations and show greater empathic concern. Importantly, the study then attempts to situate these gender differences within the Levantine-Arabic mass media framework, as in Levantine Arabic culture, there exists a potential affecting emotional expression and decoding; hence, these studies are expanded beyond the previous research into unexplored linguistic and cultural domains.

Finally, the study addresses associated gaps noted by Ouali & El Garouani (2024) in ensuring the limited availability of Arabic emotional datasets. For the first time in the recent past, the study uses naturalistic content from the Nafs podcast, going beyond the limitations of acted emotional corpora, which mostly lack ecological validity, and instead providing a much more culturally grounded account of the recognition of emotion in Arabic media.

To sum up, while the findings broadly align with the existing literature concerned with the recognition of emotion and empathy, they do contribute meaningfully to our understanding by locating these phenomena within Levantine-Arabic cultural frameworks and thereby expanding the theoretical applicability of Brunswik's Lens Model.

5.5 Theoretical and Practical Implications

The findings of this study have contributed to communication, media studies, and affective computing, and it has provided a real backdrop against which theory and practice can converse.

5.5.1 Theoretical Implications

First, the study has furthered the importance of Brunswik's Lens Model in the context of emotion recognition in vocal media communication. It presents evidence showing that listeners integrate multiple paralinguistic cues by employing a probabilistic judgment process, supporting the Lens Model and extending it beyond the traditional settings of social and experimental psychology into media and cultural communication research. Importantly, by incorporating empathy as a cognitive-emotional filter, the scope of the model has been expanded to include a more sophisticated view of the interaction between the individual and his traits and the cues that determines from the environment. This development as a result enhances the explanatory power of the Lens Model with regard to the specificity of non-Western cultures.

Second, by positioning empathy as a semiotic mechanism, the study not only treats it as an emotional or psychological characteristic but also as a tool used in decoding cultural and emotional signs in vocal media. This perspective on empathy advances the theory of semiotics in that it identifies the double role an empathy can have, on one hand it is a cognitive instrument to assume another's perspective, while on the other, it is found in emotional resonance. Building on this, the research bridges the gap between semiotics, cultural communication, and affective science, and presents a more interdisciplinary approach to emotion identification research.

Finally, the research deepens the understanding of gender-based emotional decoding in the concerned Arabic media contexts. The existence of gender differences in recognition accuracy, correlated with empathy and cultural familiarity as the study suggests, provides a culturally grounded contribution to the topic of gender and communication that puts into question the dominance of Western-centric findings.

5.5.2 Practical Implications

These findings can inspire action from the coverage perspective in the media sector, the producers, broadcasters, and podcasters of the Arabic market. Voice cues that best describe a certain emotion are instrumental for professionals to use when designing content that resonates emotionally with audiences and holds their active engagement. Inversely, these contrasting decoding processes between genders testify to the need for communication protocols that take into account different sensitivities within their audience.

Within affective computing and AI, this study contributes toward refining systems that recognize Arabic-dialect speech emotion expressions. The analysis of real Levantine-Arabic emotional expressions provides culturally sound grounds for training an AI model, overcoming the usual shortcomings prevalent in datasets that primarily selected acted or speech not representative of a target culture. These systems could also potentially benefit from empathy-based interpretations, opening a new window of opportunity toward more human-centric and contextually adaptive AI applications.

The research investigations give us some insights in intercultural training programs, highlighting the use of empathy in understanding culturally tinted emotions. These programs may in turn also develop cross-cultural interactions on a deeper level by integrating new understandings on the interplay between vocal cues and empathic processing.

5.6 Study Limitations and Future Research

while this study shows some worth in the semiotic consideration of empathy in voice emotion recognition in Levantine-Arabic media content. Nevertheless, several limitations should be acknowledged.

5.6.1 Limitations

First, the study depended on a certain sample of university students; consequently, the findings may not be applicable to broader populations. While the group had insightful offerings given their awareness of the Levantine dialect and familiarity with digital media, the results may fall short of representing the demographic segments of older audiences and those in non-university setups.

The modest sample size in this study and its demographic composition limit the generalizability of the results. While the findings provide valuable insights about Levantine Arabic podcasts audience, they cannot be assumed to represent other Arabic-speaking populations, dialects or broader global audiences. So, a larger and demographically diverse sample would allow for a stronger claims about cross-cultural validity.

Secondly, while the focus on Levantine Arabic media content proved enriching culturally, it is a major caveat to the extent to which the phenomenon can be studied in a dialectal

and cultural context. Emotional prosodic profiles and semiotic interpretations may be different according to Arabic dialects, and even different cultures; hence cross-dialectal and cross-cultural assessments become necessary if one wishes to arrive at farther-reaching generalizations.

The use of audio clips from podcast-like content only restricts the ecological validity of the results. In everyday communication, vocal cues usually come with visual or contextual information, which this study failed to capture. Thus, the outcomes reflect the realm of audio-only media and may not be completely applicable to multimodal contexts.

The third issue is choosing the audio clips, though naturalistic and ecologically valid, it may be too narrow to encompass real-world complexity of emotional expression through media. Emotions are dynamic and deeply linked to each other in context; static audio clips, however, therefore, cannot really do well to represent their multimodal nature in real-life media circumstances (e.g., audiovisual podcasts, broadcasts, or even live interactions).

Finally, given that empathy was assessed through a self-reported scale (IRI). Such instruments may never truly account for the degree or depth of empathic engagement during the emotional decoding process.this line of investigation could benefit from complementing self-report measures of empathy with behavioral or physiological measures (e.g., eye-tracking, galvanic skin response) for a fuller understanding of how empathy works in reaching emotion recognition.

5.6.2 Future research and recommendations

Building on the previous identified limitations, the researcher proposed the following research questions and methodological directions, these suggested directions are intended as a targeted research questions and methodological approaches that extend the present study's contribution while addressing its constraints.:

1 Broader and more diverse samples: Expanding the age groups; declining to those with very limited studies; or Arab-speaking countries across the educational background and cultural scope of participants will add to the external validity of the findings.
Future researchers could ask: How do age, educational background, and regional variation affect the accuracy of vocal emotion recognition? To answer this question,

future studies must gather data from larger samples of the population, ensuring that samples will be divided into groups based on age, gender, and regional dialect so that they can be systematically compared across these subgroups. In order to ascertain whether the patterns of recognition seen here are universal or specific to a given context.

2 Cross-dialect and cross-cultural comparisons: Studying and recognizing vocal emotions among the various Arabic dialects or contrasting Arabic-speaking audiences with others on the cultural plane would provide further insights into cultural familiarity in emotional decoding.

3 Multimodal emotion recognition: With the integration of audiovisual content, researchers would be able to investigate how visual cues interact with vocal signals to form emotion recognition. the audio-only design of this study highlights prosody in isolation but leaves open the question of how emotions are perceived in richer, multimodal contexts. Future researchers might ask: Does the relative weight of vocal cues differ than visual cues when listeners interpret emotions in multimodal settings compared to audio-only content?

4 Advanced measures of empathy: The combination of self-report empathy scales with experimental and physiological measures could lead to a deeper understanding of how empathy affects prosody interpretation.

5 Technological and AI applications: Future studies could begin to adapt such a hypothesis in order to realize a culturally adaptive SER system for Arabic dialects, thereby relating pure academic research to practical technology.future researcher might ask: how could emotion recognition systems trained on Levantine Arabic data outperform models trained on cross-linguistic corpora in classifying emotions?

So, to develop understanding of the semiotic and empathic phenomena underlying vocal emotion recognition, this must ask our attention in particular to those elements. This will certainly enlarge both the theoretical discourse and its practical contribution.

5.7 Conclusion

The study gives compelling proof of the complex interactions among vocal cues, empathy, and cultural context in emotion recognition in Levantine-Arabic vocal media content. Using Brunswik's Lens Model, this research shows how listeners embrace multiple paralinguistic cues in a probabilistic and interpretative manner influenced by

individual traits and socio-cultural familiarity. This study therefore provides the first systematic exploration of emotional prosody in Levantine-Arabic vocal media, a domain that has been completely neglected in international media studies.

Empathy was conceptualized not just as a psychological disposition but as a semiotic agency enhancing the emotional understanding of faint and ambiguous instances of emotions. There are also gender differences in recognition responses, raising yet another criterion for perceptual and empathic differences among listeners, these results not only extend and confirm the existing models but also stress the need to revise theories of vocal emotional recognition to take into account factors of cultural and dialectal variation.

Altogether, this study views to extend theoretical models in communication and media studies while offering some pragmatic construction insights for culturally adaptive media designs and emotion-aware technologies. A strong incentive from this research is that studies on emotion recognition are forced to be grounded in culturally specific paradigms, moving away from Western-centered ones to an embrace of the distinctive reality of multiple linguistic and media environments.

Ultimately, this thesis offers a richer understanding of how empathy, culture, and media intersect and interact in the interpretation of vocal emotions, leaving a foundation for both scholarly investigation and the development of practical designed technologies such as cultural aware emotion recognition systems.

References

Abbas, M. J., Khalil, L. S., Haikal, A., Dash, M. E., Dongmo, G., & Okoroha, K. R. (2021, October). Eliciting Emotion and Action Increases Social Media Engagement: An Analysis of Influential Orthopaedic Surgeons. *Arthroscopy, Sports Medicine, and Rehabilitation, 3*, e1301–e1308. https://doi.org/10.1016/j.asmr.2021.05.011

Abbaschian, B. J., Sierra-Sosa, D., & Elmaghraby, A. (2021, February). Deep Learning Techniques for Speech Emotion Recognition, from Databases to Models. *Sensors, 21*, 1249. https://doi.org/10.3390/s21041249

Abdel-Hamid, L. (2020). Egyptian Arabic speech emotion recognition using prosodic spectral and wavelet features. *Speech Communication, 122*, 19–30. https://doi.org/10.1016/j.specom.2020.04.005

Ahmed, A. H., Yew, R.-J., Minocher, X., & Venkatasubramanian, S. (2024). Navigating Dialectal Bias and Ethical Complexities in Levantine Arabic Hate Speech Detection. *arXiv preprint arXiv:2412.10991*.

Aladeemy, A. A., Alzahrani, A., Algarni, M. H., Alsubari, S. N., Aldhyani, T. H., Deshmukh, S. N., Khalaf, O. I., Wong, W.-K., & Aqburi, S. (2024, November). Advancements and challenges in Arabic sentiment analysis: A decade of methodologies, applications, and resource development. *Heliyon, 10*, e39786. https://doi.org/10.1016/j.heliyon.2024.e39786

Alhuzali, H., Abdul-Mageed, M., & Ungar, L. (2018). Enabling deep learning of emotion with first-person seed expressions. *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, (pp. 25–35). https://doi.org/10.18653/v1/W18-1104

Alsharhan, E., & Ramsay, A. (2020, October). Investigating the effects of gender, dialect, and training size on the performance of Arabic speech recognition. *Language Resources and Evaluation, 54*, 975–998. https://doi.org/10.1007/s10579-020-09505-5

Althobaiti, M. J. (2023, December). An open-source dataset for arabic fine-grained emotion recognition of online content amid COVID-19 pandemic. *Data in Brief, 51*, 109745. https://doi.org/10.1016/j.dib.2023.109745

Ardianto, D. T., Mataram, S., Ismail, A. I., & Wicaksono, A. R. (2022, December). Onomatopoeia: When Sound Is Seen. *Mudra Jurnal Seni Budaya, 38*, 1–8. https://doi.org/10.31091/mudra.v38i1.2220

Barrett, L. (2017). *How Emotions Are Made: The secret life of the brain.* Pan Macmillan.

Bashir, M. F., Javed, A. R., Arshad, M. U., Gadekallu, T. R., Shahzad, W., & Beg, M. O. (2023, May). Context-aware Emotion Detection from Low-resource Urdu Language Using Deep Neural Network. *ACM Transactions on Asian and Low-Resource Language Information Processing, 22*, 1–30. https://doi.org/10.1145/3528576

Benfield, J. A., & Szlemko, W. J. (2006). Internet-based data collection: Promises and realities. *Journal of Research Practice, 2*, D1. https://jrp.icaap.org/index.php/jrp/article/view/30.html

Bergen, N., & Labonté, R. (2019, December). "Everything Is Perfect, and We Have No Problems": Detecting and Limiting Social Desirability Bias in Qualitative Research. *Qualitative Health Research, 30*, 783–792. https://doi.org/10.1177/1049732319889354

Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996, July). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology, 71*, 110–129. https://doi.org/10.1037/0022-3514.71.1.110

Biersack, S., & Kempe, V. (2005, December). Exploring the Influence of Vocal Emotion Expression on Communicative Effectiveness. *Phonetica, 62*, 106–119. https://doi.org/10.1159/000090092

Blackhurst, T., Warmelink, L., Roestorf, A., & Hartley, C. (2024, July). The Brunswik Lens Model: a theoretical framework for advancing understanding of deceptive communication in autism. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1388726

Bodie, G. D., Vickery, A. J., Cannava, K., & Jones, S. M. (2015, January). The Role of "Active Listening" in Informal Helping Conversations: Impact on Perceptions of Listener Helpfulness, Sensitivity, and Supportiveness and Discloser Emotional Improvement. *Western Journal of Communication, 79*, 151–173. https://doi.org/10.1080/10570314.2014.943429

Brown, G., & Hellmuth, S. (2022, June). Computational modelling of segmental and prosodic levels of analysis for capturing variation across Arabic dialects. *Speech Communication, 141*, 80–92. https://doi.org/10.1016/j.specom.2022.05.003

Brunswik, E. (1947). *Perception and the representative design of psychological experiments.* Berkeley: University of California Press.

Brunswik, E. (1952). The conceptual framework of psychology.

Burgoon, J. K., Wang, R. X., Chen, X., Ge, T. S., & Dorn, B. (2022). How the Brunswikian Lens Model Illustrates the Relationship Between Physiological and Behavioral Signals and Psychological Emotional and Cognitive States. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.781487

Callejas, Z., & López-Cózar, R. (2008, May). Influence of contextual information in emotion annotation for spoken dialogue systems. *Speech Communication, 50*, 416–433. https://doi.org/10.1016/j.specom.2008.01.001

Chrysikou, E. G., & Thompson, W. J. (2016, July). Assessing Cognitive and Affective Empathy Through the Interpersonal Reactivity Index: An Argument Against a Two-Factor Model. *Assessment, 23*, 769–777. https://doi.org/10.1177/1073191115599055

Chul Min Lee, & Narayanan, S. S. (2005, March). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing, 13*(2), 293-303. https://doi.org/10.1109/tsa.2004.838534

Cohen, J. (2018). Defining identification: A theoretical look at the identification of audiences with media characters. In *Advances in foundational mass communication theories* (pp. 253–272). Routledge.

Collins, F. M. (2014). *The relationship between social media and empathy.* Master's thesis, Georgia Southern University. https://digitalcommons.georgiasouthern.edu/cgi/viewcontent.cgi?article=2189&context=etd

Coombs, W. T. (2007). *Ongoing crisis communication* (2. ed. ed.). Los Angeles [u.a.]: SAGE.

Creswell, J. W., & Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, Calif. [u.a.]: SAGE Publ.

Dair, Z., Donovan, R., & O'Reilly, R. (2021). Linguistic and gender variation in speech emotion recognition using spectral features. *arXiv preprint arXiv:2112.09596*.

Dair, Z., Donovan, R., & O'Reilly, R. (2021). Linguistic and Gender Variation in Speech Emotion Recognition using Spectral Features. https://doi.org/10.48550/ARXIV.2112.09596

Dalal, D., & Rayhana, K. (2020). *Emotion recognition in Arabic speech signal.* Ph.D. dissertation, UNIVERSITY OF KASDI MERBAH OUARGLA.

Darwin, C. C. (1872). *The Expression of the Emotions in Man and Animals.*

Davis, M. H. (1980). Interpersonal Reactivity Index. *Interpersonal Reactivity Index*. American Psychological Association (APA). https://doi.org/10.1037/t01093-000

Davis, M. H. (2018, February). *Empathy: A Social Psychological Approach.* (M. H. Davis, Ed.) Routledge. https://doi.org/10.4324/9780429493898

Decety, J., & Jackson, P. L. (2004). The Functional Architecture of Human Empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*, 71-100. https://doi.org/10.1177/1534582304267187

Deshmukh, R. A., Amati, V., Bhamare, A., & Jadhav, A. (2023). Visual Sentiment Analysis: An Analysis of Emotions in Video and Audio. *International Conference on IoT Based Control Networks and Intelligent Systems*, (pp. 313–326).

Dhamyal, H., Memon, S. A., Raj, B., & Singh, R. (2019). The phonetic bases of vocal expressed emotion: natural versus acted. *The phonetic bases of vocal expressed emotion: natural versus acted*. arXiv. https://doi.org/10.48550/ARXIV.1911.05733

Diemerling, H., Stresemann, L., Braun, T., & Oertzen, T. V. (2024). Implementing machine learning techniques for continuous emotion prediction from uniformly segmented voice recordings. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1300996

Dziobek, I. (2012, January). Comment: Towards a More Ecologically Valid Assessment of Empathy. *Emotion Review, 4*, 18–19. https://doi.org/10.1177/1754073911421390

Eisenberg, N., & Strayer, J. (1990). *Empathy and its development.* CUP Archive.

Ekman, P. (1982). *Emotions in the Human Faces* (2nd edition ed.). New York: Cambridge University Press. https://search.worldcat.org/title/Emotion-in-the-human-face/oclc/8034789

Ekman, P. (1992, May). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200. https://doi.org/10.1080/02699939208411068

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*, 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Embarki, M. (2008, December). *Arabic and its Varieties: Phonetic and Prosodic Aspects.* (M. Embarki, Ed.) Languages & Linguistics. https://hal.science/hal-00348770

Emotion. (n.d.). Retrieved January 10, 2025, from Merriam-Webster.com dictionary. https://www.merriam-webster.com/dictionary/emotions

Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics, 5*, 1. https://doi.org/10.11648/j.ajtas.20160501.11

Evans, J. R., & Mathur, A. (2018, August). The value of online surveys: a look back and a look ahead. *Internet Research, 28*, 854–887. https://doi.org/10.1108/intr-03-2018-0089

Fleiss, J. L. (1971, November). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378–382. https://doi.org/10.1037/h0031619

Frijda, N. H. (1986). The emotions. *Studies in Emotion and Social Interaction*.

Gerczuk, M., Amiriparian, S., Ottl, S., & Schuller, B. W. (2023, April). EmoNet: A Transfer Learning Framework for Multi-Corpus Speech Emotion Recognition. *IEEE Transactions on Affective Computing, 14*, 1472–1487. https://doi.org/10.1109/taffc.2021.3135152

Gilet, A.-L., Mella, N., Studer, J., Grühn, D., & Labouvie-Vief, G. (2013, January). Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, *45*, 42–48. https://doi.org/10.1037/a0030425

Gómez-Zaragozá, L., del Amor, R., Castro-Bleda, M. J., Naranjo, V., Raya, M. A., & Marín-Morales, J. (2024). EMOVOME: A Dataset for Emotion Recognition in Spontaneous Real-Life Speech. *EMOVOME: A Dataset for Emotion Recognition in Spontaneous Real-Life Speech*. arXiv. https://doi.org/10.48550/ARXIV.2403.02167

Gunes, H., Piccardi, M., & Pantic, M. (2008, May). From the Lab to the Real World: Affect Recognition Using Multiple Cues and Modalities. In *Affective Computing.* I-Tech Education and Publishing. https://doi.org/10.5772/6180

Gunter, B. (2000). *Media Research Methods.* SAGE Publications, Ltd. https://doi.org/10.4135/9780857028983

Hall, J. A., Gunnery, S. D., & Schlegel, K. (2025, March). Gender and Accuracy in Decoding Affect Cues: A Meta-Analysis. *Journal of Intelligence, 13*, 38. https://doi.org/10.3390/jintelligence13030038

Hammond, K. R. (1996). *Human judgment and social policy : irreducible uncertainty, inevitable error, unavoidable injustice.* Oxford: Oxford University Press.

Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin, 137*, 643–659. https://doi.org/10.1037/a0023589

Hasyim, M., & Arafah, B. (2023, January). Semiotic Multimodality Communication in The Age of New Media. *Studies in Media and Communication, 11*, 96. https://doi.org/10.11114/smc.v11i1.5865

Hattar, R. H., & Ramirez-Hurtado, C. (2014). Sharing Is Caring: Vocal Music and Arab Identity in the Middle East. *Middle-East Journal of Scientific Research, 21*, 1988–1993.

Hellmuth, S. (2020). Contact and variation in Arabic intonation. Zenodo. https://doi.org/10.5281/ZENODO.3744553

Hess, U., & Hareli, S. (2014, October). The Role of Social Context for the Interpretation of Emotional Facial Expressions. In *Understanding Facial Expressions in Communication* (pp. 119–141). Springer India. https://doi.org/10.1007/978-81-322-1934-7_7

Israelashvili, J., Sauter, D., & Fischer, A. (2020). Two facets of affective empathy: concern and distress have opposite relationships to emotion recognition. *Cognition & Emotion, 34*, 1112–1122. https://doi.org/10.1080/02699931.2020.1724893

Jensen, K. B. (2012). *A handbook of media and communication research.* routledge London.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kakuba, S., & Han, D. S. (2025, February). Addressing data scarcity in speech emotion recognition: A comprehensive review. *ICT Express, 11*, 110–123. https://doi.org/10.1016/j.icte.2024.11.003

Kim, S. S., Kaplowitz, S., & Johnston, M. V. (2004, September). The Effects of Physician Empathy on Patient Satisfaction and Compliance. *Evaluation & the Health Professions, 27*, 237–251. https://doi.org/10.1177/0163278704267037

Klaylat, Osman, Z., Hamandi, L., & Zantout, R. (2018). Emotion recognition in Arabic speech. *Analog Integrated Circuits and Signal Processing*, *96*(2), 337-351.

KoboToolbox. (n.d.). *KoboToolbox*. https://www.kobotoolbox.org/

Kolmogorova, A., Kalinin, A., & Malikova, A. (2021). Semiotic Function of Empathy in Text Emotion Assessment. *Biosemiotics, 14*, 329–344. https://doi.org/10.1007/s12304-021-09434-y

Krejcie, R. V., & Morgan, D. W. (1970, September). Determining Sample Size for Research Activities. *Educational and Psychological Measurement, 30*, 607–610. https://doi.org/10.1177/001316447003000308

Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: expression, experience, and physiology. *Journal of personality and social psychology*, *74*, 686.

Kuckartz, U., & Rädiker, S. (2019). *Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video.* Springer International Publishing. https://doi.org/10.1007/978-3-030-15671-8

Lambrecht, L., Kreifelts, B., & Wildgruber, D. (2013). Gender differences in emotion recognition: Impact of sensory modality and emotional category. *Cognition and Emotion*, *28*(3), 452-469.

Lausen, A., & Schacht, A. (2018, June). Gender Differences in the Recognition of Vocal Emotions. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00882

Levordashka, A., Richardson, M., Hirst, R. J., Gilchrist, I. D., & Fraser, D. S. (2025, January). Trace: A research media player measuring real-time audience engagement. *Behavior Research Methods*, *57*. https://doi.org/10.3758/s13428-024-02522-0

Lindquist, K. A., Gendron, M., Satpute, A. B., & Lindquist, K. (2016). Language and emotion. *Handbook of emotions*, *4*, 579–594.

Liu, X., Dong, J., & Jeon, M. (2023, December). Robots' "Woohoo" and "Argh" Can Enhance Users' Emotional and Social Perceptions: An Exploratory Study on Non-lexical Vocalizations and Non-linguistic Sounds. *ACM Transactions on Human-Robot Interaction*, *12*, 1–20. https://doi.org/10.1145/3626185

Livingstone, S. (2002). Relationships between media and audiences: prospects for audience reception studies. In *Media, ritual and identity* (pp. 237–255). Routledge.

Mackes, N. K., Golm, D., O'Daly, O. G., Sarkar, S., Sonuga-Barke, E. J., Fairchild, G., & Mehta, M. A. (2018, September). Tracking emotions in the brain – Revisiting the Empathic Accuracy Task. *NeuroImage*, *178*, 677–686. https://doi.org/10.1016/j.neuroimage.2018.05.080

Matsumoto, D., Yoo, S. H., & Nakagawa, S. (2008). Culture, emotion regulation, and adjustment. *Journal of Personality and Social Psychology*, *94*, 925–937. https://doi.org/10.1037/0022-3514.94.6.925

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution that Will Transform how We Live, Work, and Think.* boston new york: Houghton Mifflin Harcourt.

McHugh, S. (2016, April). How podcasting is changing the audio storytelling genre. *Radio Journal:International Studies in Broadcast & Audio Media*, *14*, 65–82. https://doi.org/10.1386/rjao.14.1.65_1

McStay, A. (2018). *Emotional AI.* London: SAGE Publications Ltd.

Meddeb, M., Karray, H., & Alimi, A. (2016). Content-Based Arabic Speech Similarity Search and Emotion Detection. *Advances in Intelligent Systems and Computing*, 530-539.

Meddeb, M., Karray, H., & Alimi, A. M. (2015). Speech emotion recognition based on Arabic features. *2015 15th international conference on intelligent systems design and applications (ISDA)*, (pp. 46–51).

Memon, S. A., Dhamyal, H., Wright, O., Justice, D., Palat, V., Boler, W., Raj, B., & Singh, R. (2019). Detecting gender differences in perception of emotion in crowdsourced data. *Detecting gender differences in perception of emotion in crowdsourced data*. arXiv. https://doi.org/10.48550/ARXIV.1910.11386

Mena, M. E. (2012). Emotion Recognition from Speech Signals: Erasmus Exchange Project Work. *Emotion Recognition from Speech Signals: Erasmus Exchange Project Work.*

Mesquita, B. (2002, December). Emotions as Dynamic Cultural Phenomena. In *Handbook of Affective Sciences* (pp. 871–890). Oxford University PressNew York, NY. https://doi.org/10.1093/oso/9780195126013.003.0046

Milner, R., Jalal, M. A., Ng, R. W., & Hain, T. (2019, December). A Cross-Corpus Study on Speech Emotion Recognition. *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 304–311). IEEE. https://doi.org/10.1109/asru46091.2019.9003838

Montaño, R., & Alías, F. (2016, December). The role of prosody and voice quality in indirect storytelling speech: Annotation methodology and expressive categories. *Speech Communication, 85*, 8–18. https://doi.org/10.1016/j.specom.2016.10.006

Mosier, K. L., & Kirlik, A. (2004). Brunswik's lens model in human factors research: Modern applications of a classic theory. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*, pp. 350–354.

Mousi, B., Durrani, N., Ahmad, F., Hasan, M. A., Hasanain, M., Kabbani, T., Dalvi, F., Chowdhury, S. A., & Alam, F. (2024). AraDiCE: Benchmarks for Dialectal and Cultural Capabilities in LLMs. *AraDiCE: Benchmarks for Dialectal and Cultural Capabilities in LLMs*. arXiv. https://doi.org/10.48550/ARXIV.2409.11404

Muhammad, S. H., Ousidhoum, N., Abdulmumin, I., Wahle, J. P., Ruas, T., Beloucif, M., de Kock, C., Surange, N., Teodorescu, D., Ahmad, I. S., Adelani, D. I., Aji, A. F., Ali, F. D., Alimova, I., Araujo, V., Babakov, N., Baes, N., Bucur, A.-M., Bukula, A., Cao, G., . . . Mohammad, S. M. (2025). BRIGHTER: BRIdging the Gap in Human-Annotated Textual Emotion Recognition Datasets for 28 Languages. *BRIGHTER: BRIdging the Gap in Human-Annotated Textual Emotion Recognition Datasets for 28 Languages*. arXiv. https://doi.org/10.48550/ARXIV.2502.11926

Muin, F., Noortyani, R., & Elyani, E. P. (2021, January). Rethinking arbitrariness of language and its implication in language use. (K. Chaemsaithong, Ed.) *Cogent Arts & Humanities, 8*. https://doi.org/10.1080/23311983.2020.1868687

Nabi, R. L. (1999, August). A Cognitive-Functional Model for the Effects of Discrete Negative Emotions on Information Processing, Attitude Change, and Recall. *Communication Theory, 9*, 292–320. https://doi.org/10.1111/j.1468-2885.1999.tb00172.x

Nabi, R. L. (2010). The case for emphasizing discrete emotions in communication research. *Communication Monographs, 77*, 153–159.

Nabi, R. L., & Green, M. C. (2015). The role of a narrative's emotional flow in promoting persuasive outcomes. *Media Psychology, 18*, 137–162.

Nafs Show- برنامج نفس. (n.d.). *Nafs Show- برنامج نفس*. Shams Media. https://www.youtube.com/@shamsmedia.

Nordström, H. (2019). *Emotional communication in the human voice.* Ph.D. dissertation, Department of Psychology, Stockholm University.

O'Connor, C., & Joffe, H. (2020, January). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods, 19*. https://doi.org/10.1177/1609406919899220

Orduna, M., Pérez, P., Gutiérrez, J., & García, N. (2021). Methodology to Assess Quality, Presence, Empathy, Attitude, and Attention in 360-degree Videos for Immersive Communications. https://doi.org/10.48550/ARXIV.2103.02550

Orduna, M., Pérez, P., Gutiérrez, J., & García, N. (2023, July). Methodology to Assess Quality, Presence, Empathy, Attitude, and Attention in 360-degree Videos for Immersive Communications. *IEEE Transactions on Affective Computing, 14*, 2375–2388. https://doi.org/10.1109/taffc.2022.3149162

Ostwald, P. F. (2019). *The semiotics of human sound* (Vol. 36). Walter de Gruyter GmbH & Co KG.

Ouali, S., & El Garouani, S. (2024). Arabic Speech Emotion Recognition using Convolutional Neural Networks. *Journal of Electrical Systems, 20*, 2700–2708.

Palestinian Central Bureau of Statistics. (2025). Palestine in Figures 2024. *Palestine in Figures 2024*. Ramallah, Palestine: Palestinian Central Bureau of Statistics. https://www.pcbs.gov.ps/Downloads/book2715.pdf

Pasupathi, M., & Rich, B. (2005, May). Inattentive Listening Undermines Self-Verification in Personal Storytelling. *Journal of Personality, 73*, 1051–1086. https://doi.org/10.1111/j.1467-6494.2005.00338.x

Peirce, C. S. (1934). *Collected papers of charles sanders peirce* (Vol. 5). Harvard University Press.

Pittam, J. (1994). Voice in Social Interaction: An Interdisciplinary Approach. *Voice in Social Interaction: An Interdisciplinary Approach*. Thousand, Oaks, California: SAGE Publications, Inc. https://doi.org/10.4135/9781483327105

Plutchik, R. (2001). The Nature of Emotions. *American Scientist, 89*(4), 344.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion, 37*, 98–125.

Preston, S. D., & de Waal, F. B. (2002, February). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1–20. https://doi.org/10.1017/s0140525x02000018

Quiros-Ramirez, M. A., & Onisawa, T. (2013, August). Considering cross-cultural context in the automatic recognition of emotions. *International Journal of Machine Learning and Cybernetics, 6*, 119–127. https://doi.org/10.1007/s13042-013-0192-2

Reddy, S., Lazarova, M., Yu, Y., & Jones, R. (2021). Modeling Language Usage and Listener Engagement in Podcasts. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 632–643). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.52

Reinhard, C. D. (2021). The case for media engagings: Minutia Reception Studies and a reconceptualization of studying media use, reception, and effects. *Participations–Journal of Audience and Reception Studies, 18*, 274–298.

Riess, H. (2017, May). The Science of Empathy. *Journal of Patient Experience, 4*, 74–77. https://doi.org/10.1177/2374373517699267

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review, 110*, 145.

Saussure, F. d. (1974). Course in General Linguistics, ed. J. Culler, trans. *W. Baskin. London: Fontana*.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010, November). Perceptual Cues in Nonverbal Vocal Expressions of Emotion. *Quarterly Journal of Experimental Psychology, 63*, 2251–2272. https://doi.org/10.1080/17470211003721642

Schemer, C., Matthes, J., & Wirth, W. (2008, September). Toward Improving the Validity and Reliability of Media Information Processing Measures in Surveys. *Communication Methods and Measures, 2*, 193–225. https://doi.org/10.1080/19312450802310474

Scherer. (2001). *Appraisal considered as a process of multilevel sequential checking.* Appraisal processes in emotion/Oxford University Press.

Scherer. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227–256. https://doi.org/10.1016/s0167-6393(02)00084-5

Sen, A., Isaacowitz, D., & Schirmer, A. (2017). Age differences in vocal emotion perception: on the role of speaker age and listener sex. *Cognition & Emotion, 32*, 1189–1204. https://doi.org/10.1080/02699931.2017.1393399

Sharma, A., & Verbeke, W. J. (2024, September). Influence of gender dimorphism on audience engagement in podcasts: a machine learning analysis of dynamic affective linguistic and paralinguistic features. *Frontiers in Communication, 9*. https://doi.org/10.3389/fcomm.2024.1431264

Shin, D., & Biocca, F. (2017, September). Exploring immersive experience in journalism. *New Media & Society, 20*, 2800–2823. https://doi.org/10.1177/1461444817733133

Shrivastava, A. (2014). Active empathic listening as a tool for better communication. *Active empathic listening as a tool for better communication*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2538476

Siegert, I., Böck, R., & Wendemuth, A. (2013, November). Inter-rater reliability for emotion annotation in human–computer interaction: comparison and methodological improvements. *Journal on Multimodal User Interfaces, 8*, 17–28. https://doi.org/10.1007/s12193-013-0129-9

Spackman, M. P., Brown, B. L., & Otto, S. (2009). Do emotions have distinct vocal profiles? A study of idiographic patterns of expression. *Cognition & Emotion, 23*, 1565–1588. https://doi.org/10.1080/02699930802536268

Tashakkori, A. (2010). *SAGE handbook of mixed methods in social & behavioral research.* Sage.

Tashakkori, A., & Teddlie, C. (2010). *SAGE Handbook of Mixed Methods in Social & Behavioral Research.* SAGE Publications, Inc. https://doi.org/10.4135/9781506335193

Thompson, W. F., & Balkwill, L.-L. (2010). Cross-cultural similarities and differences. In P. N. Juslin, & J. A. Sloboda (Eds.), *Handbook of music and emotions* (pp. 755–788). United Kingdom: Oxford University Press.

Tomkins, S. (1962). *Affect imagery consciousness: Volume I: The positive affects.* Springer publishing company.

Tuli, F. (2011, April). The Basis of Distinction Between Qualitative and Quantitative Research in Social Science: <i>Reflection on Ontological, Epistemological and Methodological Perspectives</i>. *Ethiopian Journal of Education and Sciences, 6*. https://doi.org/10.4314/ejesc.v6i1.65384

Vasilyuk, F. E. (2016). Semiotics and the Technique of Empathy. *Journal of Russian and East European Psychology, 53*, 56–79. https://doi.org/10.1080/10610405.2016.1230994

Verma, R., Kalsi, N., Shrivastava, N. P., & Sheerha, A. (2022, August). Development and Validation of the AIIMS Facial Toolbox for Emotion Recognition. *Indian Journal of Psychological Medicine, 45*, 471–475. https://doi.org/10.1177/02537176221111578

Webster, J. G. (2018, September). 6. Audience Behavior. In *Mediated Communication* (pp. 91–104). De Gruyter. https://doi.org/10.1515/9783110481129-007

Wesarg-Menzel, C., Gallistl, M., Niconchuk, M., Böckler, A., O'Malley, B., & Engert, V. (2024). Compassion buffers the association between trauma exposure and PTSD symptom severity: Findings of a cross-sectional study. *Psychoneuroendocrinology, 165*, 107036.

Wimmer, R. D. (2011). *Mass media research: An introduction.* Cengage learning.

Winkler, J. R., & Appel, M. (2024, December). Measuring dynamic emotional experiences in response to media stimuli. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1436918

Wright, K. B. (2006, June). Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication, 10*, 00–00. https://doi.org/10.1111/j.1083-6101.2005.tb00259.x

Wu, J., Dang, T., Sethu, V., & Ambikairajah, E. (2021). Multimodal Affect Models: An Investigation of Relative Salience of Audio and Visual Cues for Emotion Prediction. *Frontiers in Computer Science, 3*. https://doi.org/10.3389/fcomp.2021.767767

Zhang, L.-M. L., Li, Y. Y., Zhang, Y.-T. Y., Ng, G. G., Leau, Y.-B. Y., & Yan, H. H. (2023, 1 25). A Deep Learning Method Using Gender-Specific Features for Emotion Recognition. *Sensors, 23*(3), 1355. https://doi.org/10.3390/s23031355

Appendix A: Informed Consent Form

Research Title:

Exploring the Semiotic Role of Audience Empathy in Recognizing Emotions in Levantine-Arabic Vocal Media Content

Principal Investigator:

Yahya Abu Asab
Master Student at Arab American University – Ramallah Campus

Hello and thank you for considering participation in this research study. My name is Yahya Abu Asab, and I am a graduate student in the Master's program of Integrated Digital Media. This is a study that investigates how individuals perceive emotions in Levantine-Arabic audio content, with a specific focus on the role of empathy as a semiotic function in decoding vocal cues to interpret emotions of speaker. The study is being conducted as part of my master's thesis.
This page outlines the purpose of the study and your rights as a participant. Your participation in this research is voluntary. If you agree to continue, please read this consent form carefully and click "I Agree" at the end to confirm your participation.

Study Description:

This study will take approximately 30 to 40 minutes to complete and involves three parts:

1. Completing the Interpersonal Reactivity Index (IRI)
2. Listening to 36 short Levantine-Arabic voice clips and identifying the emotions conveyed.
3. Answering short, open-ended reflection questions after each clip.

Potential Risks and Discomforts:

The study does not involve any significant risks. However, some participants may experience minor emotional discomfort while engaging with emotionally expressive audio content. If you feel discomfort, you may stop participating at any time without any penalty.

Potential Benefits:

Although there is no direct benefit to you, this research may contribute to a deeper academic understanding of empathy, emotional perception, and media and communication studies. You may also gain personal insight into how you relate to and interpret human emotion through voice.

Confidentiality:

Your responses will remain strictly confidential. No personally identifiable information will be linked to your responses. All data will be anonymized and stored securely on password-protected systems accessible only to the researcher and academic supervisor. Data will be retained for up to 3 years for academic purposes.

Voluntary Participation and Withdrawal:

Participation in this study is voluntary. You may withdraw from the study at any point, for any reason, without penalty. If you choose to withdraw, any data collected from you will be excluded from analysis upon your request.

Contact Information:

If you have any questions about this study or your participation, please contact:

- Researcher: Yahya Abu Asab – e-mail: y.abuasab@student.aaup.edu
- Supervisor: Dr. Naheda Makhadmeh – e-mail naheda@yu.edu.jo

Consent Confirmation:

By clicking "I Agree" below, you are confirming the following:

- You are at least 18 years old.
- You have read and understood the information provided above.
- You voluntarily agree to participate in this research study.

If you do NOT wish to participate, please close this window or navigate away from the page.

Consent Statement:

☐ I have read and understood all the information above.

☐ I agree to participate in this research study.

☐ I understand that I may withdraw at any time.

☐ I consent to the use of my anonymous responses for academic research purposes.

Signature……

Appendix B: Demographics Survey

Please answer the following questions. All responses will remain confidential and will be used solely for academic research purposes.

1. **Age**:

   ☐ 18– below 21

   ☐ 21– below 24

   ☐ 24 and above

2. **Gender**:

   ☐ Male

   ☐ Female

3. **Relationship status**

   ☐ Single

   ☐ Married

   ☐ Divorced

   ☐ Widow

   ☐ In a relationship

4. **Native Language:**

   ☐ Arabic

   ☐ English

   ☐ Other: _____

5. **Education level:**

   ☐ Secondary school

   ☐ Diploma

   ☐ Undergraduate

   ☐ Graduate

6. **If you currently a university student, which year?:**

   ☐ First Year

   ☐ Second Year

   ☐ Third Year

☐ Fourth Year

☐ Other: _____

7. **Field of Study:**

   _____

8. **Do you have any hearing difficulties or diagnosed hearing conditions?**

   ☐ Yes

   ☐ No

9. **Do you use any hearing aids or assistive listening devices?**

   ☐ Yes

   ☐ No

10. **Do you regularly listen to podcasts or audio-based content in Arabic?**

    ☐ Yes

    ☐ No

11. **Have you ever listened to the podcast (Nafs)?**

    ☐ Yes

    ☐ No

    ☐ Not sure

Appendix C: Interpersonal Reactivity Index (IRI)

Directions: the following statements inquire about your thoughts and feelings in a variety of situations. For each item, indicate how well it describes you by choosing the appropriate letter on the scale at the top of the page: a, b, c, d, or e.

When you have decided on your answer, fill in the letter next to the item number.

Read each item carefully before responding. Answer as honestly as you can. Thank you.

ANSWER SCALE:

A B C D E

DOES NOT DESCRIBE ME DESCRIBES ME VERY WELL

1. I daydream and fantasize, with some regularity, about things that might happen to me.
2. I often have tender, concerned feelings for people less fortunate than me.
3. I sometimes find it difficult to see things from the "other guy's" point of view.
4. Sometimes, I do not feel very sorry for other people when they are having problems.
5. I really get involved with the feelings of the characters in a novel.
6. In emergency situations, I feel apprehensive and ill-at-ease.
7. I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.
8. I try to look at everybody's side of a disagreement before I make a decision.
9. When I see someone being taken advantage of, I feel kind of protective towards them.
10. I sometimes feel helpless when I am in the middle of a very emotional situation.
11. I sometimes try to understand my friends better by imagining how things look from their perspective.
12. Becoming extremely involved in a good book or movie is somewhat rare for me.
13. When I see someone get hurt, I tend to remain calm.
14. Other people's misfortunes do not usually disturb me a great deal.
15. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.
16. After seeing a play or movie, I have felt as though I were one of the characters.

17. Being in a tense emotional situation scares me.
18. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.
19. I am usually pretty effective in dealing with emergencies.
20. I am often quite touched by things that I see happen.
21. I believe that there are two sides to every question and try to look at them both.
22. I would describe myself as a pretty soft-hearted person.
23. When I watch a good movie, I can very easily put myself in the place of a leading character.
24. I tend to lose control during emergencies.
25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
26. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.
27. When I see someone who badly needs help in an emergency, I go to pieces.
28. Before criticizing somebody, I try to imagine how I would feel if I were in their place

استكشاف الدور السيميائي لتعاطف الجمهور في التعرف على المشاعر في المحتوى الاعلامي الصوتي العربي–الشامي

يحيى سعيد يعقوب ابوعصب

د. ناهدة مخادمة

د.بشار طحاينة

د. عبد الله عدوي

الملخص

تبحث هذه الرسالة في التلقي الإعلامي الصوتي من منظور متعدد التخصصات، مع التركيز على الدور السيميائي للتعاطف في عملية التعرف على المشاعر من المحتوى الصوتي العربي باللهجة الشامية. حيث اعتمدت الدراسة على تحليل كيفية توجيه المؤشرات الصوتية المدركة (مثل النغمة، طبقة الصوت، السرعة، الشدة، والتنغيم) لعملية التفسير لدى جمهور المستمعين وذلك بهدف إدراك المشاعر التي يعبر عنها المتحدث. ومن هذا المنطلق، تؤطر الدراسة التعاطف ليس فقط بوصفه سمة نفسية، بل أيضًا كعدسة سيميائية تشكل مستوى التلقي الميكروي لدى الجمهور.

أُجريت الدراسة في فلسطين خلال الفصل الدراسي الثاني لعام 2025، واعتمد الباحث على منهج وصفي ذي تصميم مختلط مقطعي.حيث شملت العينة (384) مشاركاً ومشاركة من طلبة الجامعات الفلسطينية والذين تراوحت أعمارهم بين (18–45) عامًا، كما جرى جمع البيانات باستخدام استبيان رقمي والذي يتضمن مقاطع صوتية أصيلة مأخوذة من بودكاست نَفَس على يوتيوب (المنشور من فبراير 2023 حتى يونيو 2024). وقد تضمنت أدوات البحث: مقياس التفاعل الشخصي IRI لقياس أبعاد التعاطف لدى المشاركين في البحث،ومن ثم مهمة التعرف على المشاعر الصوتية، بالاضافة لاجابتهم على أسئلة مفتوحة لاثراء التحليل النوعي، وتم التحقق من صدق الأدوات وثباتها من خلال مراجعة الخبراء والاختبارات الإحصائية.

حيث أظهرت النتائج أن جمهور المستمعين بهدف معرفة شعور المتكلم يقومون باستخدام اشارات صوتية متعددة في عملية تفسيرية احتمالية والتي تتأثر بالخصائص الفردية والخلفية الاجتماعية–الثقافية للمستمع. كما تبين أن مستويات التعاطف المرتفعة، وخاصة في بعدي "التعاطف الوجداني" و"تبني المنظور" من مقياس التفاعل الشخصي لدايفس IRI، ساهمتا بشكل ملحوظ في تحسين دقة

التعرف على المشاعر. وكشفت النتائج ايضاً عن وجود فروق بين النساء والرجال في التعرف على المشاعر من الصوت وحده في غياب اشارات اخرى في المحتوى الاعلامي، حيث تفوقت النساء على المشاركين الرجال في عملية التعرف على المشاعر الصوتية.

تُقدِّم هذه الدراسة إطارًا نظريًا جديدًا لدارسي الإعلام لفحص تلقي الجمهور للمحتوى الإعلامي الصوتي على المستوى الميكروي، عبر الجمع بين السيميائيات وعلم النفس ودراسات الاتصال. و كما تؤكد على الحاجة إلى تطوير قواعد بيانات عربية أصيلة للانفعالات الصوتية ليتم تدريب نماذج الذكاء الاصطناعي عليها، ودمج منظور التلقي الإعلامي في أنظمة الذكاء الاصطناعي والحوسبة العاطفية. وبذلك تسهم في تطوير دراسة التلقي الإعلامي من منظور متعدد التخصصات، مبرزة الدور الذي يلعبه التعاطف في تفسير الجمهور للتواصل الصوتي.

الكلمات المفتاحية: التلقي الإعلامي، المحتوى الإعلامي الصوتي، التعاطف، السيميائيات، التعرف على المشاعر