Chapter One

General Introduction[1]

1.1 Introduction[2]

Emotions are subjective personal feelings (Ekman, 1992a), which can be expressed either verbally (e.g., words, sentences) or non-verbally (e.g., facial expressions, gestures, prosody). Understanding others' emotions is crucial for effective daily communication and social interactions (Jensen, 2014; Jensen & Pedersen, 2016). Since the publication of Charles Darwin's seminal work *The Expression of the Emotions in Man and Animals* (1872; reprint in 1998), the topic has gained widespread attention in fields like linguistics, biology, psychology, neuroscience, etc.

A controversial issue in human emotion recognition is whether it is universal or culture- and/or language-specific. As a pioneer in affective science, Charles Darwin argued that the production and perception of emotions are biologically determined and universal, inherited through the human genome (1872; reprint in 1998). In contradistinction to this, Harre's (1986) social constructivist theory asserts that emotions are shaped exclusively by culture and language. More recently, Elfenbein and Ambady (2002b) proposed their dialect theory, such that while emotion recognition is universal in principle, it becomes less accurate across cultures due to "nonverbal accents", or cultural differences in expression. To date, there is an increasing consensus that cross-cultural emotion recognition results from an interplay between universal, cultural, and linguistic factors (Elfenbein, 2013; Elfenbein, Mandal et al., 2002; Mesquita & Frijda, 1992).

[1] Chapters 2 to 5 have been written as independent articles, including introduction, methodology, and conclusion sections. Therefore, overlaps between these parts are unavoidable.
[2] With gratitude to my co-authors, parts of this chapter were based on Liang et al. (2025).

How can emotion be defined? Although the concept of emotion is complex and multifaceted, it can be defined in a way that offers nuances for understanding. Emotion is a subjective personal feeling arising from events or affairs, causing physical and mental changes (Izard, 2010; Widen & Russell, 2010). According to Scherer (2009), emotion is a dynamic process resulting from individuals' evaluation of important affairs depending on their needs, goals, and values. Based on Scherer's Component Process Model (CPM) of emotion (Scherer, 2001), emotion includes five interrelated components—appraisal processes, autonomic physiology, action tendencies, motor expression, and subjective feeling. The terms *emotion* and *affect* are sometimes used interchangeably to describe feelings, but they are not identical. Emotions are multifaceted feelings that combine different aspects such as psychology, cognition, and behavior. On the other hand, affect is a broader term that includes, but is not limited to, emotions and moods (Shuman & Scherer, 2014).

How does the brain process emotions? Emotions are processed through interactive neural networks in different brain regions (e.g., hippocampus, hypothalamus, and thalamus), known as the Papez circuit, highlighting that emotion processing is not an isolated activity (Papez, 1937). Further to this, Panksepp (1998) added two additional regions—the amygdala and the prefrontal cortex, which are involved in the processing of emotions. Also, other neural networks, like the limbic system and prefrontal regions, are involved in the dynamic processing of emotions (Celeghin et al., 2017). Currently, research on affective neuroscience has employed sophisticated neuroimaging technologies, such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and positron emission tomography (PET) (Cimino, 2002; Dehghani et al., 2023; Gu et al., 2019; Lim et al., 2024; Palomero-Gallagher & Amunts, 2022). For example, using fMRI, Gu et al. (2019) discovered that basic emotions are active in different yet overlapping regions.

How important are emotions? Emotions are intricately related to perception, memory, and decision-making (Palomero-Gallagher & Amunts, 2022; Turnbull & Salas, 2021). Turnbull and Salas (2021) demonstrate that emotions affect cognitive processing, such that positive emotions enhance working memory and facilitate problem-solving skills, whereas negative emotions interfere with working efficiency. Meanwhile, emotions can be regulated by connectivity-based neurofeedback (Dehghani et al., 2023). Findings from affective neuroscience have practical implications for mental health, psychology, and human-computer interaction (HCI) (Hudlicka, 2008;

Jungilligens et al., 2022; Okon-Singer et al., 2015; Renna et al., 2017; Rolls, 1990).

This dissertation is about the way emotions are expressed in speech and whether these emotions are being recognized. The languages involved are Dutch and Korean, which are typologically different in terms of culture and language. The overall question is whether these languages express emotions in a universal way or not, and whether listeners recognize the intended expression of emotions. We will come back to this question after having discussed the main parts that play a role in answering this overall research question.

When emotions in speech can be recognized by listeners, what cues do they use? Acoustic cues, such as pitch, amplitude, spectral distribution, duration, and laryngeal setting, are pivotal (Banse & Scherer, 1996). By acoustically extracting the above parameters, this dissertation aims to identify both universal and language-specific cues in emotional speech recognition, providing a comprehensive understanding of the production and perception of vocal emotions across Dutch and Korean.

1.1.1 Theories of emotion

Emotions can be examined from either a discrete approach (Ekman, 1992b; Izard, 1977), a dimensional approach (Russell, 1980), or an integration of both (Laukka, 2004).

1.1.1.1 Discrete (basic) emotion theory

The discrete emotion theory proposes that different emotions are characterized by specific physiological and behavioral features (Ekman, 1992a, b; Izard, 1992). Thus, this theory identifies a small set of emotions, referred to as basic emotions (Ekman, 1992b). Basic emotions are biologically conditioned and exhibit specific acoustic patterns that can be recognized above chance levels across cultures (Laukka et al., 2013; Laukka & Elfenbein, 2021).

1.1.1.2 Dimensional emotion theory

In contrast, the dimensional theory classifies emotions based on fundamental dimensions such as arousal, valence, intensity, and potency, providing a complex framework for understanding emotions (Russell, 1980; Russell & Barrett, 1999). According to the circumplex model, emotions are primarily defined by the first two dimensions mentioned—arousal and valence (Russell, 1980). Arousal (high-arousal vs. low-arousal) is the physiological change experienced by the speaker (Russell & Barrett, 1999), while valence determines the emotion's positive (pleasant) or negative (unpleasant) nature. For instance, anger is typically a high-arousal and negative emotion, whereas tenderness is regarded as a low-arousal and positive emotion. To understand how arousal and valence interact, it is essential to analyze these two dimensions together. However, two dimensions are insufficient to fully describe emotional nuances (Larsen & Diener, 1992). To describe the subtle differences between emotions from the same family, it is necessary to include other dimensions, such as intensity (Brehm, 1999) and potency (Russell & Mehrabian, 1977). Intensity distinguishes variations between emotions from the same type of emotion (e.g., hot anger vs. cold anger), underscoring the strength and magnitude of an emotion (Bänziger & Scherer, 2005; Brehm, 1999; Larsen & Diener, 1987; Sonnemans & Frijda, 1994; Wright et al., 1983). Potency refers to the cognitive appraisal of a person's power or influence over a situation (Lazarus & Smith, 1988). Integrating these dimensions offers a more comprehensive understanding of the complex nature of emotional states.

We assigned the distinction between basic and non-basic emotions to the dimensional approaches, and called the dimension "basicness". We did so, as basicness classifies emotions in two general subsets, like arousal and valence.

1.1.1.3 Comparing the discrete and dimensional approaches

The discrete and dimensional approaches are not mutually exclusive. Instead, they describe emotions from different perspectives, enhancing our understanding by emphasizing different facets of emotional experience. On the one hand, the discrete emotion approach categorizes emotions as distinct and individual states such as anger, fear, and happiness. It posits that basic emotions are biologically inherited and have unique acoustic patterns, which serve specific adaptive functions and can be reliably differentiated from one another. On the other hand, the dimensional approach regards emotions as varying dimensions, such as arousal and valence (Russell, 1980). Additional dimensions have been proposed, such as intensity (Brehm, 1999) and potency (Russell & Mehrabian, 1977), to handle more detailed nuances of emotions. The integration of both approaches provides a comprehensive framework for a more holistic understanding of emotions.

1.1.2 Empirical studies on cross-cultural emotion recognition

Over the past few decades, numerous studies have investigated emotion recognition across different cultures (Elfenbein & Ambady, 2002b; Juslin &

Laukka, 2003; Pell, Monetta et al., 2009; Scherer et al., 2001). These studies have mostly adopted experimental designs using either a "one-to-many" approach—presenting stimuli recorded by a single group of speakers to several groups of listeners (Beier & Zautra, 1972; Scherer et al., 2001; Van Bezooijen et al., 1983); or a "many-to-one" approach, presenting stimuli recorded by several groups of speakers to a single group of listeners (Chronaki et al., 2018; Kramer, 1964; Pell, Monetta et al., 2009; Thompson & Balkwill, 2006). Additionally, some studies have used a fully balanced design, presenting stimuli from two or more groups of speakers to the same number of listener groups (Albas et al., 1976; Jiang et al., 2015; Paulmann & Uskul, 2014; Sauter et al., 2010).

Taken together, earlier studies aimed to determine to what extent vocal emotion recognition is universal or culture-/language-specific, and have concluded that emotions are decoded above chance cross-culturally, in line with the universality hypothesis (Elfenbein, 2013; Elfenbein & Ambady, 2002b; Scherer et al., 2001). Furthermore, previous studies reveal an in-group advantage, such that individuals recognize emotions produced in their native language more accurately than those in an unknown language, indicating the existence of language-specific prosodic cues in vocal emotion expressions (Pell, Paulmann et al., 2009). However, prior research in this area has mostly employed unbalanced experimental designs and predominantly focused on basic emotions (Ekman, 1992b). Consequently, empirical research on cross-cultural emotion recognition by listeners and speakers from typologically different languages and cultures remains scarce.

Moreover, most studies on cross-cultural emotion recognition have relied on acted rather than spontaneous speech due to the challenges in controlling the verbal content in spontaneous speech (Jiang et al., 2015; Paulmann & Uskul, 2014; Pell, Monetta et al., 2009; Thompson & Balkwill, 2006; Van Bezooijen et al., 1983), with a few exceptions employing spontaneous speech (Chung, 1999). Additionally, most studies have used pseudo-utterances to eliminate semantic cues that might affect emotion recognition.

Another relevant field is research on emotion classification in speech on the basis of specific acoustic parameters, and classifying vocal emotions via a constellation of acoustic parameters. In affective computing, frequently used machine learning models include Support Vector Machine (SVM), LDA (Linear Discriminant Analysis), Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), and Convolutional Neural Network (CNN) (Ezhilarasi & Minu, 2012; Lee & Narayanan, 2005; Luengo et al., 2005; Pallewela et al., 2024; see Ververidis & Kotropoulos, 2006 for a review).

Classification rates vary depending on a number of factors, such as the selection of acoustic features and the size of the corpora. To address the limitations of each model, Ezhilarasi and Minu (2012) proposed hybrid systems that integrate both SVM models and deep learning neural networks (DNNs) to improve classification accuracy. Furthermore, Laukka et al. (2011) suggested that classification rates can be improved by combining speech with facial expressions and physiological signals. However, due to the limited number of corpora, current findings of automatic speech recognition may not be generalizable to different languages, cultures, and vocal emotions.

This dissertation employs a balanced two-by-two design, with speakers and listeners from different languages and cultures—Dutch and Korean, aiming to better examine the in-group advantage in a cross-cultural setting. Balanced design is especially important in evaluating the in-group advantage. However, due to the difficulty of recording acted speech, there are limited corpora of vocal emotion expressions. For example, the VENEC corpus is a large database of vocal emotion expressions (Laukka et al., 2010). It includes 19 different emotions and a total of 6,500 stimuli, recorded by 100 professional voice actors from five countries.

1.1.3 Research on affective neuroscience

Emotions are intricately related to perception, memory, and decision-making (Palomero-Gallagher & Amunts, 2022; Turnbull & Salas, 2021). Turnbull and Salas (2021) demonstrate that emotions affect cognitive processing, such that positive emotions enhance working memory and facilitate problem-solving skills, whereas negative emotions interfere with working efficiency. Meanwhile, emotions can be influenced by connectivity-based neurofeedback (Dehghani et al., 2023). Findings from affective neuroscience have practical implications for mental health, psychology, and human-computer interaction (HCI) (Hudlicka, 2008; Jungilligens et al., 2022; Okon-Singer et al., 2015; Renna et al., 2017; Rolls, 1990).

1.2 Research methodology

I used the stimuli from the Demo (Dutch emotion)/Koremo (Korean emotion) corpus (Broersma et al., 2025).[3] This corpus was specifically developed for cross-linguistic comparison and is more balanced than the existing corpora in several respects. First, the two sub-corpora contain a comparatively large

[3] The scenarios and corpus are publicly available via Radboud University at https://doi.org/10.34973/5kg3-9852

number of emotions (eight emotions) which were balanced in arousal (high-arousal vs. low-arousal) and valence (positive vs. negative), and with an equal number of basic and non-basic emotions (see Table 1.1). Second, the eight emotions were expressed by a large number of actors from two typologically different languages (eight Dutch and Korean actors), with the same number of females and males in each language group, accounting for gender-related differences in prosodic expression of emotions (Klatt & Klatt, 1990). Each actor produced the same emotions twice, resulting in a total of 256 portrayals (8 emotions × 8 actors × 2 tokens × 2 languages). Third, the pseudo-sentence /nuto hɔm sɛpikɑŋ/ is phonologically compatible in both Dutch and Korean.[4] Using a pseudo-sentence can eliminate verbal semantic processing. Further, since vowels are considered to carry more affective meanings than consonants (Majid, 2012), listeners are more affected by vowel duration than consonant duration. In this study, the pseudo-sentence /nuto hɔm sɛpikɑŋ/ was created with a roughly equal number of vowels (/u/, /o/, /ɔ/, /ɛ/, /i/, /ɑ/) and consonants (/n/, /t/, /h/, /m/, /s/, /p/, /k/, /ŋ/). Fourth, the same elicitation technique (the Stanislavski technique) was used by both the Korean and Dutch stage directors. The elicitation methods and recording procedures were the same in both languages. For more details of the elicitation and recording procedure, refer to Chapter 2.

Table 1.1. The eight emotions included in this project in a valence-by-arousal grid (reproduced from Goudbeek & Broersma, 2010b, p. 2212); basic emotions are indicated by *.

		Valence	
		Positive	Negative
Arousal	High	Joy*	Anger*
		Pride	Fear*
	Low	Tenderness	Sadness*
		Relief	Irritation

[4] According to Goudbeek and Broersma (2010b), the pseudo-sentence /nuto hɔm sɛpikaŋ/ is phonologically legal in both Dutch and Korean. However, the low-mid vowel [ɔ] does not exist in Korean (Shin, 2015) and was pronounced as high-mid vowel [o].

1.3 The current study

Emotions play a pivotal role in human communication, especially in a cross-cultural setting (Jensen, 2014; Trampe et al., 2015). Although emotions can be recognized above chance across cultures, individuals recognize emotions more accurately when they are expressed by members from the same or similar cultural/linguistic group than by members from a typologically different group. This is referred to as the in-group advantage. However, although research in this field has produced fruitful insights, there remain significant gaps that prevent us from fully understanding how vocal emotions are affected by cultural and linguistic factors. First, most previous studies have either used a "one-to-many" approach, presenting stimuli recorded by a single group of speakers to several groups of listeners (Beier & Zautra, 1972; Scherer et al., 2001; Van Bezooijen et al., 1983); or a "many-to-one" approach, presenting stimuli recorded by several groups of speakers to a single group of listeners (Chronaki et al., 2018; Kramer, 1964; Pell, Monetta et al., 2009; Thompson & Balkwill, 2006). Consequently, it results in an unbalanced design that renders it difficult to test the in-group advantage. Second, most prior research has focused on basic emotions (Cordaro et al., 2016; Laukka et al., 2016). Therefore, current knowledge on emotions cannot be generalized to non-basic emotions. Moreover, the limited number of emotions leads to unbalanced designs in terms of arousal and valence. Third, most studies investigate emotions from a discrete approach (Ekman & Friesen, 1969; Ekman et al., 1987; Laukka et al., 2013), although a limited number of studies explore emotions from a dimensional approach (Barrett, 1998; Laukka et al., 2005; Mozziconacci, 2002; Russell, 1980). Thus, subtle differences between emotions remain unclear, particularly when emotions share similar features. To bridge the gap, I employ the Demo/Koremo corpus (Broersma et al., 2025) using the "two-by-two" design with listeners and speakers from typologically different cultures and languages—Dutch and Korean, and includes a relatively large number of emotions balanced for arousal and valence.

Specifically, Dutch is a stress-accent language with binary trochees, which has a rather restricted pitch range (Gussenhoven, 1993). In Dutch, word stress is employed to differentiate between identical segment strings, for example, *KAnon* /ˈkanɔn/ "list of saints" versus *kaNON* /kaˈnɔn/ "large gun". Dutch utterances employ two prosodic units above the word level: Intonational Phrase (IP) and Phonological Phrase (PP) (Gussenhoven, 2005). In contrast, Korean does not have minimal stress pairs. Although there are controversies regarding the rhythm classification of Korean, most studies tend to classify it as a syllable-timed language (Arvaniti, 2012). Korean utterances are divided into two prosodic units above the word level, namely Intonational Phrase (IP) and—different than Dutch—the Accentual Phrase (AP) (Jun, 2005).

The overarching goal of this study is to investigate cross-cultural vocal emotion recognition from both the discrete and the dimensional approaches, focusing on the influence of culture-/language-specific factors, acoustic cues, and emotional dimensions (including also emotional intensity), and basicness on recognition accuracy. To address this broad issue, I conducted four studies targeting several sub-questions. Collectively, the results of these four studies contribute to a better understanding of the perception of vocal emotions in a cross-cultural setting.

1.4 Two approaches

The chapters in this dissertation adopt two approaches—a discrete and a dimensional one. The discrete approach targets the categorical perception of separate emotions, which may hinge on subtle differences between discrete emotions (Chapters 3 and 4). The dimensional approach, as implemented in the present dissertation, makes a three-way overall dimensional characterization of emotions based on arousal (high-arousal vs. low-arousal), valence (positive vs. negative), and basicness (basic vs. non-basic) (Chapters 2 and 5). The discrete approach presents a more precise and perhaps more subtle understanding of the distinctions between emotions, while the dimensional approach aims to distinguish overall underlying properties. For instance, both anger and irritation are negative emotions, whereas they differ in terms of arousal. Anger is a high-arousal emotion with intense energy, while irritation is a low-arousal emotion with mild energy (Spielberger et al., 1995). But do the dimensions cover all discrete emotions adequately? Integrating these two approaches might provide a comprehensive framework to study the complexity of emotions. Therefore, in Chapter 5, we combine these two approaches by analyzing confusion matrices based on the eight emotions, illustrating emotions that are easily misclassified.

1.5 Overview of the dissertation

The rest of this dissertation consists of five chapters, each addressing various aspects of vocal emotion recognition. The final chapter (Chapter 6) summarizes the research chapters 2 to 5 with a discussion and an integration of the results, highlights the dissertation's significant contributions, and, of course, addresses questions about the limitations of our empirical studies and provides suggestions for future research.

Chapter 2 "Investigating cross-cultural vocal emotion recognition with an affectively and linguistically balanced design" investigates recognition of vocal emotions by listeners from two different cultures and languages, i.e., Dutch and Korean, and examines the so-called in-group advantage in vocal emotion recognition. The in-group advantage hypothesis predicts that listeners recognize vocal emotions produced in their native language more accurately than when expressed in an unknown language. Regardless of the applicability of the in-group advantage, we predict that listeners recognize vocal emotions produced in their native language and in the unknown language above chance, which would show that the expression and perception of vocal emotion has at least a universal component. Finally, the chapter examines the influence of arousal, valence, and basicness on vocal emotion recognition, within and across cultures. As explained above in § 1.2, these are three dimensions that are part of the dimensional approach to emotion production and perception. In our study, we dichotomize the eight target emotions into subsets of four, i.e., high- vs. low-arousal, positive vs. negative valence, and basic vs. non-basic. We will examine whether some subsets (quadruplets) are easier to recognize cross-culturally than others. For instance, basic emotions may be easier to recognize, both within and across cultural divides, than non-basic emotions. We also predict that confusions in the recognition of vocal emotions will be more frequent within than across dimensional quadruplets. In this chapter, the focus is limited to only the accuracy of the emotion identification, and the similarity structure of the emotions is not investigated. This similarity structure will be examined in a later chapter, where the comparison between machine and human identification of emotions and confusion matrices will be presented.

Chapter 3 "Interpreting the intensity of vocal emotions across cultures" analyzes the intensity ratings by Dutch and Korean listeners, as collected in Study 1. The starting point for this chapter is that Intensity should be added (and studied in more detail) as a separate dimension of (vocal) emotions to capture emotional states that cannot be adequately described by the traditional dimensions of Arousal, Valence, and Potency. Emotions are always expressed with different levels of intensity (Mesquita & Frijda, 1992). Intensity refers to the strength of emotions perceived by receivers, and people tend to respond to emotions with higher intensity than those with lower intensity. Moreover, individuals usually give higher intensity ratings to emotions expressed by members from the same or similar culture/linguistic group than by members from a typologically different group, which is known as the in-group bias

(Kommattam et al., 2019). This chapter examines whether there exists an in-group bias in intensity ratings across accurate and inaccurate trials. Finally, as in the preceding chapter, we examine the effect of arousal, valence, and basicness on intensity ratings. I will do this first for all responses, and then repeat the analysis for the subset of correctly identified emotions (which should have higher intensity ratings).

Chapter 4 "Classifying emotions from acoustic parameters" examines the role of a large number of acoustic cues in recognition accuracy, focusing on the influence of emotion, speaker language, and gender on recognition accuracy. Since we are interested in the vocal (rather than verbal) expression of emotion, this chapter will target the effects of prosodic parameters only. These are properties of human speech that cannot be tied down to specific individuals' speech sounds (phonemes) but are characteristic of larger speech units, such as syllables, phrases, clauses, sentences, and even paragraphs. Specifically, we will examine the role of vocal pitch, acoustic intensity (loudness), articulatory setting (vocal timbre, as conveyed by formants and spectral tilt), and harmonicity (noisiness of the voice). I examine whether recognition accuracy can be reliably predicted by a constellation of acoustic parameters, and compare the recognition accuracy between machine classifiers and human listeners.

This chapter is not only about the effects of acoustic parameters on recognition accuracy but also, and more crucially, on the (cross-cultural) confusion of emotions. If an emotion is not correctly identified, then what is it mistaken for? This opens a window on cross-cultural misunderstanding in the signaling of emotions. So the real research question is not about the accuracy of the emotion perception per se, but on the effects of acoustic parameters on the identification of emotions (whether correct or confused).

Chapter 5 "Recognizing vocal emotions in unfamiliar languages" focuses on cross-cultural vocal emotion recognition by American English and French listeners who had no knowledge of Dutch or Korean. This chapter investigates the relative contributions of the Universality hypothesis, Cultural Proximity, Linguistic Proximity, and emotional dimensions to emotion recognition. According to the Universality hypothesis, some emotions are universally recognized by people across cultures and languages. People from a similar cultural background can recognize emotions more accurately than those from a different one (Elfenbein & Ambady, 2003a). In the vocal domain, listeners find it much easier to identify emotions expressed in a language typologically similar to their native language than to a different one. Furthermore, emotional dimensions, such as arousal, valence, and basicness, affect emotion recognition. However, it remains unknown how these factors affect recognition accuracy of emotions. Therefore, I aim to study to what extent universal, cultural, linguistic, and emotional dimensions affect the perception of vocal emotions. To achieve this goal, I selected American English listeners and French listeners, since English is a stress-timed language, which is prosodically/rhythmically close to Dutch; French is a syllable-timed language, which is prosodically/rhythmically similar to Korean. By comparing the recognition accuracy between these two groups, we can find out how the above factors affect the perception of vocal emotions. First, I tested whether both listener groups recognized vocal emotions above chance. Second, I tested recognition accuracy in Dutch recordings by both groups of listeners. Third, I tested whether French listeners outperform American English listeners in Korean recordings, since French and Korean are syllable-time languages, which share similar prosodic/rhythmic patterns. Finally, I examined the role of arousal, valence, and basicness in vocal emotion recognition.

To answer the above questions, I conducted three perception experiments and one acoustic analysis of stimulus materials. In the first experiment, Dutch and Korean listeners were asked to listen to each stimulus and identify the emotion it conveyed by choosing one of the eight emotions listed on screen. In the second experiment (using the same moment of data collection), the listeners were asked to estimate the intensity of the emotion as experienced/expressed by the speaker. The third experiment tested the perception of vocal emotions by American English and French listeners. Similar to the first study, these two groups of listeners were asked to listen to the same corpus and select the emotion they thought the stimulus expressed and estimate the intensity of the emotion as expressed by the speaker. Fourth, I acoustically analyzed each of the 256 portrayals based on 17 acoustic parameters, and further examined the correlations between acoustic parameters and recognition accuracy. Moreover, I compared recognition accuracy predicted by machine and human listeners.

Chapters 2 to 5 address the following four main research questions:
Chapter 2: Do Dutch and Korean listeners recognize vocal emotions above chance in Dutch and Korean, and is there an in-group advantage in vocal emotion recognition?
Chapter 3: Is there an in-group bias in intensity ratings of Dutch and Korean vocal emotions by Dutch and Korean listeners?
Chapter 4: How do acoustic parameters of vocal emotions vary across emotions, speaker language, and gender in Dutch and Korean?
Chapter 5: Is cross-cultural/language vocal emotion recognition in unfamiliar languages affected by Universality, Cultural Proximity, Prosodic Proximity, and emotional dimensions?