LANGUAGES OF EMOTION

Freie Universität Berlin

Does Gender Affect Recognition of Vocal Emotions?

Evidence from Persian speakers living in a collectivist society

Fachbereich Philosophie und Geisteswissenschaften
der Freien Universität Berlin

Dissertation
zur Erlangung des akademischen Grades
Doktorin der Philosophie (Dr. phil.)
Doctor of Philosophy (Ph.D.)

vorgelegt von
Niloofar Keshtiari

M.Sc. in Clinical Linguistics
M.A. in General Linguistics
M.A. in Teaching English as a Foreign Language

Berlin, 2015

Erstgutachterin: Prof. Dr. Gisela Klann-Delius, Freie Universität Berlin
Zweitgutachter: Prof. Dr. Dr. Friedemann Pulvermüller, Freie Universität Berlin

Tag der Disputation: 28.09.2015

Dedicated to
my mother & the memory of my father,
with gratitude and love.

Acknowledgements

The completion of this dissertation would never have happened without the support and the contribution of many people. I would like to take this opportunity to express my sincere gratitude to some of them.

First and foremost, I would like to express my deepest respect and gratitude to my supervisor Prof. Dr. Gisela Klann-Delius, for being a constant source of support, encouragement, and constructive advice. Her exceptional dedication, generosity and caring have constantly extended far beyond the duty of an academic supervisor. I am also very grateful to Prof. Dr. Dr. Friedemann Pulvermüller for his support and for accepting to be the second reviewer of this thesis. Many thanks go to all the members of the dissertation committee (Prof. Dr. Dr. Kondrad Ehlich, Prof. Dr. Stefan Müller, and Dr. Kora Kim) for taking the time to read and comment on the manuscript.

I would also like to thank my co-author Michael Kuhlmann, to whom I could always turn for advice concerning statistical analysis and experiment setups. My experiments would also be unrealistic to accomplish without his technical support. A very special thank you to Dr. Silke Paulmann, from University of Essex in England, for her invaluable guidance in various sections of my project. I am indeed grateful to Professor Farzad Sharifian from Monash University in Melbourne, Australia for his expert advice on the cultural framework of my project.

My very special thanks goes to all the great people I had the privilege to meet during my time in the Cluster of Languages of Emotion at Freie Universitaet Berlin who all contributed to the successful completion of this dissertation project, in particular Farifteh Tavakoli Borazjani, Dr. Luna Beck, Dr. Francesca Citron, Dr. Dar Meshi, Dr. Evgeniya Kirilina, Dr. James Murphy, Dr. John Kluempers, Martin Schultze, Dr. Georg Hosoya, Susan Ullrich, Liila Taruffi, Shuang Guo, Dr. Moritz Lehne, Dr. Elif Alkan Härtwig, Arash Aryani, Laura Sebastian, Filomena Prisco, Pascal Michaelis, Christa Bös, Carlota Klein, Dr. Thomas Stodulka, Dr. Burkhard Meyer-Sickendiek, and Dr. Chun-Ting Hsu. I am very grateful to the brilliant secretaries at the cluster of Languages of Emotion: Regula von Schintling, Richard Praetorius, and Nicole Steindorf-Rentzsch, without their help my paper work would have been utter chaos. I would of course like to thank all the numerous participants who took part in my experiments and diligently accomplished the experiments. I am also hugely indebted to everybody who was helping me to establish the Persian emotional speech database, collect data, recruit participants and run the studies in different countries: Dr. Reza Nilipour, Dr. Saeid Sheikh Rohani, Dr. Yahya Modarresi Tehrani, Jörg Dreyer, Hossein Namvar Arefi, Dr. Hassan Ashayeri, Dr. Arsalan Kahnemuyipour, Dr. Afsaneh Fazly, Dr. Masood Ghayoomi, Dr. Fatemeh Izadi, Trudi Falamaki Zehnder, Karl Brian Northeast, Prof. Klaus Scherer. I am very grateful to the Iranian actors — Mithra Zahedi and Vahid Etemad —without their collaboration it was not possible to establish the Persian emotional speech database. Also, special thanks to Andreas Christian Götz who kindly accepted to translate the English summary of this dissertation into German.

Lastly and most importantly, I am forever indebted to my mother, and my brother who encouraged and helped me at every stage of my personal and academic life, and longed to see this achievement come true. Similar thanks go to the memory of the first teacher of my life, my father, who developed a passion for learning in me.

Structure of the Dissertation

This is a cumulative dissertation presented in the form of a collection of manuscripts that are at the time of thesis submission either published or submitted for publication. Chapter 1 of this dissertation brings about an overview of related theories and concepts. Chapter 2, Chapter 3, and Chapter 4 are dedicated to the three manuscripts. The bibliographic details for each manuscript and their order of appearance in the thesis are set out below.

- Chapter Two:
  Keshtiari, N., Kuhlmann, M., Eslami, M., & Klann-Delius, G. (2014). Recognizing emotional speech in Persian: A validated database of Persian emotional speech (Persian ESD). *Behavior research methods*, 1-20. Doi 10.3758/s13428-014-0467-x

- Chapter Three:
  Keshtiari, N., Kuhlmann, M., & Klann-Delius, G. (2014). Gender effects on the recognition of emotional prosody: Evidence from Persian Language (Farsi). *Submitted*

- Chapter Four:
  Keshtiari, N., Kuhlmann, M., & Klann-Delius, G. (2014). Gender effects on the recognition of emotional speech: Evidence from a collectivist culture. *Submitted*

And finally Chapter 5 concludes this dissertation with a general discussion as well as implications and suggestions for future research.

Table of Contents

Summary

Emotion perception and expression are an integral part of human interaction. We human beings communicate emotions in a rich and sophisticated way that sets us apart from all other species on this planet. In fact, communicating and understanding emotions is vital for human social interaction. Human beings' major channels of communication, the routes through which the content of communication flows, are vocal and visual modality. Emotions portrayed via vocal channel (i.e., lexical content and/or prosody) carry a rich source of information about a speaker's emotions and are influenced by different factors. While previous research has yielded a wealth of information regarding the recognition of vocal expression of emotions, a number of essential factors influencing these processes have been neglected. *Language* and *culture*[1] are instances of these overlooked factors.

Language is a crucial factor in transmission of emotions. It was revealed that vocal expressions of emotion manifested in emotional prosody[2], are inherently dynamic and are integrated with linguistic properties of the language under study. Besides, researchers believe that linguistic-specific features of a language (e.g., rhythm) affect vocal expressions of emotion (i.e., emotional prosody). Rhythm is an important aspect of prosody which pertains to the way languages are organized in time. The postulated classification of stress-timed rhythm and syllable-timed rhythm is the foundation of phonetic research on rhythm. While this distinction is not totally uncontroversial, it is regularly used to make a prosody-wise division between languages, which are said to be (primarily) stress-timed, or (primarily) syllable-timed. Research has revealed a stark[3] contrast between *stress-timed languages* (where the amount of time to say a sentence depends on the number of stressed syllables) and *syllable-timed language* (where the amount of time to say a sentence depends on the total number of the syllables. In other words, stress-timed languages (e.g., German, English) tend to have an equal amount of time elapsed between stressed syllables and do not place equal stress on every syllable[4]. While studying recognition of emotional prosody, this distinction of rhythmic division of time is important as during vocal communications, in

[1] In the context of this dissertation, the term "culture" refers to shared elements (e.g., rituals and habits, belief and value system, social institutions, and behavior patterns) within any interactive collectivity of people as well as the basis of the individual preferences in daily life.

[2] Emotional prosody encompasses non-verbal aspects of human language (i.e., modulations in the acoustic parameters of speech, such as intensity, rate and pitch) and provides a rich source of information about a speaker's emotions and social intentions.

[3] One such contrast is durational variability in consonant sequences.

[4] For more explanations see section 1.3 of the dissertation.

conjunction with linguistic decoding, listeners attend to changes not only in pitch, loudness and voice quality, but also to changes in rhythm, to form an impression about the speaker's emotional state. Yet, despite the large amount of research on the recognition of vocal emotions in stress-timed languages (e.g., English, German), no such study has been performed on a syllable-timed language (e.g., Persian, Turkish and French).

Besides language, *culture* is another influential factor in emotional processes. In fact, culture and emotion are intermingled with each other. On the one hand, emotions and emotional experience are largely influenced by culture. Namely, how emotions are perceived, expressed, and regulated differ as a function of cultural norms of each society. On the other hand, emotions reinforce the cultural themes of a given society. As such, in any given cultural context, behaviors and emotional experiences that fit with and enhance the goals of that specific culture are reinforced. Cultural variation in the perceived importance of social interactions is manifested in *norms* for decoding and displaying emotions. These norms are called *emotional display rules*. Emotional display rules are cultural norms observed as an important part of any culture; they refer to culturally prescribed rules learnt early in life via socialization. These culturally shared norms and unwritten codes dictate *when*, *how* and *to whom* individuals should express their emotional experiences. They may magnify, demagnify, neutralize, qualify, or mask an emotional response. Further research has documented that these norms regulate the display of emotion in cases in which display of specific emotions might be potentially disruptive to social interaction, as well as the decoding of emotions when recognizing these emotions would disrupt social interaction. Accordingly, cultural variation in the accuracy of emotion recognition is attributed to the norms that cultures impose on their members as for recognition and expression of specific emotions. This idea suggests that for a full understanding of emotional processes we should consider the cultural context where emotions are perceived, expressed, and experienced.

To study emotions within cultural context, it is essential to make use of stable and meaningful dimensions of cultural variability. To date, *Cultural dimensions theory* developed by Hofstede in 1980 is a widely-accepted model for describing cross-cultural communication. In this context *dimension* refers to an aspect of a culture that can be measured relative to other cultures. Cultural dimensions theory describes five different dimensions including a) Individualism-Collectivism, b) Power Distance, c) Uncertainty Avoidance, d) Masculinity-Femininity, and e) Long Term Orientation-Short Term Normative Orientation. Among the dimensions of cultural variability, *individualism-collectivism* which is conceived as two poles of one dimension is a meaningful predictor of cultural variations in emotional norms (i.e., display rules). Individualists tend to emphasize and enhance their personal goals, interests and values over the society

they belong to. Self-enhancement, the need for individual autonomy, detachment from others and functioning based on personal choices are main features of individualistic cultures. Considering these features, individualist cultures regard emotions as essential personal experiences whose expression is any individual's right. In these cultural contexts, individuals tend to consider emotions as inner states vented spontaneously. Individualist cultures reinforce outward displays of emotion. In contrast to individualism, collectivism is featured by interdependent relationships which focus on social cohesion and a group's harmony. In a collectivist culture the concept of self is bound to surrounding social context in which the self cannot be split from others and the goal is to keep oneself tied to others. Collectivists identify themselves as members of a group they belong to, therefore they emphasize and enhance group goals, interests and values over those of each individual member of the group. In these cultures emotions are regarded as interactive experiences which mirror the social context rather than the individual's inner self. In collectivistic cultures expression of emotion is controlled, due to the point that it is grounded in evaluation of the relationship between the self and the others.

Despite the influential role of culture in emotion processes, the existing studies on vocal emotions have targeted members of individualist cultures, but no study has investigated the recognition of vocal emotions in a collectivist culture where the language of the stimuli is the same as the mother tongue of the participants.

Gender differences in recognition and expression of emotion is the subject which has gained increasing attention over the past decades. In spite of this expanded attention, conflicting findings have left the exact nature of these presumed differences an open question. A number of psychologists and anthropologists (e.g., Michael Eid, Batja Mesquita, and David Matsumoto) argue that for a comprehensive understanding of gender differences in recognition and expression of emotion and to achieve more accurate results, these studies should be replicated with regard to the influential factors (i.e., language and culture). To this end, this dissertation aims at examining the role of gender in the recognition of vocal emotions in a syllable-timed language (i.e., Persian) spoken in a collectivist society (i.e., Iran). In doing so, three studies are conducted which are briefly described below.

The first study pertains to generating an emotional speech database. To date, numerous databases of vocal expressions of the basic emotions have been established in several languages including English, German, Chinese, Japanese, Russian, as well as many other languages. However, this has not been achieved in Persian. Research has documented that emotional prosody can be affected by the linguistic features of a language, and that it is essential for researchers to generate valid emotional stimuli that is suitable for the linguistic background of the participants of a study. Considering this issue, recording

samples of all the basic emotions are required for a comprehensive study on emotional speech for the language under study. Therefore, in this study we aimed to fill this gap for Persian language by generating a robust set of validated stimuli. Such stimuli can minimize the influence of individual bias, and can avoid subjectivity in stimulus selection in future studies of Persian emotional speech.

In studying emotional speech, various researchers have often prepared their own, study-specific lists of sentences without validating the emotional lexical content. Although, to conduct a study on emotional speech, a set of validated sentences is required to separate the impact of lexical content from prosody on the processing of emotional speech. Therefore, three experiments (with 1,126 participants) were conducted to generate a set of validated sentences (lexical material) in colloquial Persian. The output of these experiments is a set of 90 validated Persian sentences that were reliably associated either with one particular emotion (i.e., anger, 17 sentences; disgust, 15; fear, 15; sadness, 14; and happiness, 15) or with no emotion at all (neutral, 14 sentences). The sentences were articulated by two native Persian speakers with acting experience (one male, one female) in three conditions: (1) congruent (emotional lexical content articulated in a congruent emotional voice), (2) incongruent (neutral sentences articulated in an emotional voice), and (3) baseline (all emotional and neutral sentences articulated in neutral voice). The speech materials comprise about 470 sentences. The validity of the database was evaluated by a group of 34 native speakers in a perception study. Previous works suggest that vocal emotions are recognized almost four times above chance level. Therefore, in order to select the best possible exemplars, a minimum of five times chance performance in the seven-choice emotion recognition task (i.e., 71.42 %) was set as the cutoff level. Hence, 468 utterances were regarded as valid portrayals of the target emotions. Acoustic analysis of the valid emotional utterances revealed differences in pitch, intensity, and duration, the attributes that may help listeners to correctly classify the intended emotion. The final output of this study is a database called *Persian Emotional Speech Database (Persian ESD).* This database is generated to be used as a reliable material source (for both text and speech) in future cross-cultural or cross-linguistic studies of emotional speech, and it is available to researchers for academic research purposes. Researchers from several scientific fields may find our database useful, but it is primarily aimed at psychologists, neuroscientists, linguists, and computer scientists. The stimuli of the following studies are also selected from the vocal portrayals of the *Persian ESD*.

In the second study we examined the role of gender in the recognition of emotional prosody in an under investigated linguistic context (i.e., a syllable-timed language). As mentioned earlier language is a crucial factor in transmission of emotions. During vocal communications, in conjunction with linguistic decoding, listeners attend to changes in not only in pitch, loudness and voice quality, but also to changes

in rhythm, to form an impression about the speaker's emotional state. Nevertheless, in examining the role of gender in the recognition of vocal emotions extant studies neglected the rhythmic division of time, and almost exclusively focused on stress-timed languages (i.e., German and English). This reveals a gap in research in gender effects on the recognition of emotional prosody in languages with a different rhythmic division of time. Therefore, we aimed to fill this gap by examining the role of gender in the recognition of emotional prosody in native speakers of Persian, a syllable-timed language. In a behavioral experiment, 66 native speakers of Persian were asked to recognize the emotional prosody of a set of validated vocal portrayals. These vocal portrayals were a set of sentences with emotionally neutral lexical content, intoned in one of the five intended emotions (anger, disgust, fear, happiness, and sadness) by a male and a female actor. Findings of the study indicated that recognition of emotional prosody significantly differed as a function of gender of the participants, that is, female participants outperformed their male counterparts in recognizing all the intended emotional categories. Our results with Persian speakers are in line with the results of German and English speakers. Since Persian is a syllable-timed language, our findings suggest that the nature of the rhythmic division of time in a language (i.e., stress- vs. syllable-timed) does not influence the decoding of emotional prosody.

The third study dealt with gender effects on the recognition of emotional speech in an under investigated cultural context. The existing literature on gender and emotional speech has only targeted speakers of a specific cultural background (i.e., individualist). Given that emotional processes are largely influenced by culture, and that socio-cultural variables mediate and moderate gender differences in emotional functioning, results of the existing studies might not apply to members of collectivist cultural groups. Therefore, in the third study, we questioned the effect of gender on the recognition of emotional speech on the members of an under investigated cultural context (Persian speakers living in a collectivist society: i.e., Iran). In this experiment 70 native speakers of Persian were asked to recognize the emotional prosody of a set of emotional lexical content articulated in a congruent emotional voice (encompassing the five emotional categories: anger, disgust, fear, happiness, and sadness) intoned by a male and a female actor. Findings of the study revealed a significant effect of gender in favor of females in the recognition of emotional speech. In line with previous studies with regard to the great sensitivity of the members of collectivist cultures in perception of the emotional cues, findings of the current study revealed very high overall accuracy rates 95% (i.e., almost close to six times chance performance) by members of a collectivist culture. In a previous emotion recognition study performed on members of an individualist culture, an overall accuracy rate of 70% (i.e., almost close to five times chance performance)

was reported. A content based comparison between the findings of these two studies indicated that members of a collectivist culture show higher sensitivity to vocal emotional cues.

Together the present studies underscore the role of gender in the recognition of vocal emotions in a syllable-timed language (i.e., Persian) spoken in a collectivist society. In general, the output of the first study is the first comprehensive database of Persian emotional speech (Persian ESD) encompassing a meaningful set of validated lexical and vocal stimuli, conveying five emotional meanings (i.e., anger, disgust, fear, happiness, and sadness). Since the database covers the three conditions of (a) congruent, (b) incongruent, and (c) baseline, it provides the unique possibility to separately identify the effect of prosody and lexical content on the identification of emotions in speech. The database could also be used in neuroimaging and clinical studies to assess a person's ability to identify emotions in spoken language. Additionally, this database can open up new opportunities for future investigations in speech synthesis research, as well as in gender studies. Findings of study two and three revealed a significant female advantage in the recognition of emotional prosody (i.e., neutral lexical content articulated in a congruent emotional voice) and emotional speech (i.e., emotional lexical content articulated in a congruent emotional voice) respectively. One reason for female advantage might be that females are more socially oriented than males and are more likely to define themselves in relational terms. Besides this, biological (i.e., difference in specific hormones such as estrogen and oxytocin) as well as social factors (i.e., socio-culturally constructed roles that a given society considers appropriate for men and women) can be the probable reasons behind female advantage in the recognition of vocal emotions. In addition to the mentioned issues, in the Gender Inequality Index (GII) of the United Nations Development Program it is well documented that gender inequality in Iran is very high[1]. Yet, researchers investigating gender discrimination in Iran, report that by employing effective coping strategies, a large number of Iranian women have managed to partly deal with cultural barriers caused by gender inequality. Among these coping strategies, developing *interpersonal skills* (i.e., verbal communication, non-verbal communication, listening skills, negotiation, problem solving, decision making, and assertiveness) is reported to be one of the most effective ways Iranian women have employed to manage the gender inequality. In fact recent investigations have revealed that Iranian women show higher interpersonal skills compared to Iranian men. Having high levels of interpersonal skills may be another reason for the female advantage in vocal emotion recognition among Iranians. Besides, findings of the third study indicated that members of a collectivist culture show higher sensitivity to vocal emotional cues. Research has documented that due to the very restrictive emotional display rules within collectivist cultures, the contextual cues provided by the

[1] For detailed explanations please see section 1.8.2 of this dissertation.

interaction partners in these cultures are very subtle. As a result, to avoid breakdown in interpersonal communication and enhance interpersonal harmony in such cultural contexts great sensitivity is needed for in the perception and interpretation of the implicitly conveyed emotional cues. Accordingly, restrictive emotional display rules may be the cause of the high sensitivity to emotional vocal cues in Iranian participants.

All in all, the findings of this dissertation support the assumption that accuracy in the recognition of vocal emotions (i.e., emotional prosody and emotional speech) differs significantly as a function of decoder (participant) gender. In particular, there is an overall recognition advantage for female participants over male participants in vocal emotion recognition task. Furthermore, our findings suggest that the nature of the rhythmic division of time in a language (i.e., stress- vs. syllable-timed) does not affect the decoding of emotional prosody. Additionally, our results provide empirical evidence for the influential role of cultural context in the recognition of emotional speech.

Zusammenfassung

Die Wahrnehmung und das Ausdrücken von Emotionen sind feste Bestandteile menschlicher Interaktion. Wir Menschen kommunizieren Emotionen auf eine hochentwickelte Weise, was uns von allen anderen Arten auf diesem Planeten unterscheidet. Es ist eine Tatsache, dass das Kommunizieren und Verstehen von Emotionen von entscheidender Bedeutung für unsere soziale Interaktion ist. Die menschlichen Hauptkanäle der Kommunikation, durch die Inhalte übermittelt werden, sind sprachliche und visuelle Modalitäten. Via Sprachkanal übermittelte Emotionen (d.h. lexikaler Inhalt und/oder Satzmelodie) sind eine reichhaltige Quelle von Informationen über das emotionale Befinden des Sprechers, und unterliegen vielseitigen Einflüssen. Während vorausgegangene Forschung eine Vielfalt an Informationen bezüglich der Wahrnehmung sprachlich ausgedrückter Emotionen erbracht hat, wurde eine Vielzahl von essentiellen Faktoren vernachlässigt, die diese Prozesse beeinflussen. Sprache und Kultur sind Beispiele für diese übersehenen Faktoren. Im Kontext dieser Dissertation bezieht sich der Begriff "Kultur" auf gemeinschaftlich geteilte Elemente (z.B. Rituale und Bräuche, Glaube und Wertesystem, Sozialeinrichtungen und Verhaltensmuster).

Sprache ist ein Kernbestandteil der Übermittlung von Emotionen. Es wurde erkannt, dass sich der stimmliche Ausdruck einer Emotion, der sich in einer emotionalen Satzmelodie (Prosodie) manifestiert, eine ihm innewohnende Dynamik aufweist und linguistische Eigenheiten der untersuchten Sprache beinhaltet. Die emotionale Satzmelodie umfasst nonverbale Aspekte der menschlichen Sprache (z.B. Modulationen der akustischen Sprachparameter wie Lautstärke, Tempo und Tonlage) und stellt eine reiche Informationsquelle über die Emotionen und sozialen Absichten des Sprechers dar. Daneben glauben Wissenschaftler, dass die linguistisch-spezifischen Merkmale einer Sprache (z.B. Rhythmus) den stimmlichen Ausdruck der Emotionen (d.h. die emotionale Satzmelodie) beeinflussen. Der Rhythmus ist ein wichtiger Aspekt der Prosodie, der die Weise ausdrückt, in der Sprachen zeitlich organisiert sind. Die postulierte Klassifikation in stress-timed (Zeitaufwand für die Aussprache eines Satzes hängt von der Anzahl der betonten Silben ab) und syllable-timed (Zeitaufwand für die Aussprache eines Satzes hängt von der Gesamtanzahl der Silben ab) Rhythmus ist das Fundament der phonetischen Rhythmusforschung. Während diese Unterscheidung nicht unumstritten ist, wird sie regelmäßig benutzt, um eine prosodie-weise Unterteilung zwischen Sprachen vorzunehmen, die als primär stress- oder syllable-timed gelten. Die Forschung hat einen gänzlichen Gegensatz zwischen stress- und syllable-timed Sprachen nachgewiesen. Trotz des bisher großen Aufwandes an Forschung auf dem Gebiet des Verständnisses gesprochener

Emotionen in stress-timed Sprachen (z.B. Englisch, Deutsch), wurde noch keine solche Studie an einer syllable-timed Sprache (z.B. Persisch,Türkisch, Französisch) durchgeführt.

Neben Sprache ist auch die Kultur ein einflussreicher Faktor auf emotionale Prozesse. Tatsächlich sind Kultur und Sprache miteinander vermischt. Einerseits sind Emotionen und emotionale Erfahrungen wesentlich durch die Kultur beeinflusst, wie Emotionen wahrgenommen, ausgedrückt und geregelt werden, unterscheidet sich abhängig von den kulturellen Normen einer jeden Gesellschaft. Andererseits bestärken Emotionen die kulturellen Belange einer gegebenen Gesellschaft. In einem vorgegebenen kulturellen Kontext werden dazu passende Verhaltensweisen und emotionale Erfahrungen verstärkt, die die Ziele dieser Kultur betonen. Die kulturellen Unterschiede in der wahrgenommenen Wichtigkeit des sozialen Zusammenspiels manifestieren sich in Normen, und zeigen die Entschlüsselung und Darstellung von Emotionen. Diese Normen heißen *emotional display rules*. *Emotional display rules* sind kulturelle Normen, die als wesentlicher Bestandteil einer jeden Kultur wahrgenommen werden, sie beziehen sich auf kulturell vorgegebene Regeln, die bereits frühzeitig im Rahmen der Sozialisierung gelernt werden. Diese kulturell geteilten Normen und ungeschriebenen Gesetze diktieren wenn, wie und zu wem der Einzelne seine emotionalen Erfahrungen ausdrücken sollte. Diese mögen eine emotionale Antwort verstärken, abschwächen, neutralisieren, qualifizieren oder maskieren. Weitergehende Forschung hat dokumentiert, dass diese Normen die Zurschaustellung von Emotionen in Fällen, in denen das Zeigen spezifischer Emotionen das soziale Zusammenspiel potentiell stören könnte, regulieren, sowie auch die Entschlüsselung von Emotionen, die bei ihrer Erkennung die sozialen Wechselspiele stören würden. Kulturelle Schwankungen in der Genauigkeit der Emotionswahrnehmung werden den Normen zugeschrieben, welche Kulturen ihren Mitgliedern zur Wahrnehmung und zum Ausdruck bestimmter Emotionen auferlegen. Dieser Gedanke suggeriert, dass wir für ein vollkommenes Verständnis emotionaler Prozesse den kulturellen Kontext mit einbeziehen sollten, in dem Emotionen wahrgenommen, ausgedrückt und erfahren werden.

Für das Studium von Emotionen im kulturellen Kontext ist es essentiell, haltbare und aussagekräftige Bezugsmaßstäbe der kulturellen Vielfalt zu verwenden. Zur Zeit ist die *Cultural dimensions theory*, 1980 von Hofstede entwickelt, ein allgemein akzeptiertes Modell zur Beschreibung interkultureller Kommunikation. In diesem Kontext bezieht sich *dimension* auf einen Aspekt einer Kultur, der relativ zu anderen Kulturen messbar ist. Die *Cultural dimensions theory* beschreibt fünf verschiedene Dimensionen: a) Individualism-Collectivism, b) Power Distance, c) Uncertainty Avoidance, d) Masculinity-Femininity, e) Long Term Orientation - Short Term Normative Orientation. Unter den Dimensionen kultureller Vielfalt ist *individual-collectivism*, der als zwei Seiten einer Dimension begriffen wird, ein aussagekräftiger Prädiktor

für die kulturelle Vielfalt in emotionalen Normen, d.h. Ausdrucksregeln. Individualisten neigen dazu, persönliche Ziele, Interessen und Werte hervorzuheben und über die Gesellschaft, denen sie angehören, zu erhöhen. Selbsterhöhung, das Bedürfnis nach individueller Selbstbestimmung, die Abtrennung von Anderen und eine Arbeitsweise, basierend auf persönlichen Entscheidungen, sind Hauptmerkmale individualistischer Kulturen. Unter Berücksichtigung dieser Merkmale betrachten individualistische Kulturen Emotionen als essentielle persönliche Erfahrungen, deren Ausdruck eines jeden individuelles Recht ist. In diesem kulturellen Kontext neigt der Einzelne dazu, Emotionen als inneren Status zusehen, der sich spontan äußert. Individualistische Kulturen bestärken das Zeigen von Emotionen nach außen. Im Gegensatz zum Individualismus zeichnet sich der Kollektivismus durch voneinander abhängigen Beziehungen aus, welche sich auf das soziale Gefüge und Gruppenharmonie ausrichten. In einer kollektivistischen Kultur ist das Konzept des Ichs in den umgebenden sozialen Kontext eingebunden, in dem das Ich nicht von anderen getrennt werden kann und das Ziel es ist, selbst mit den anderen verbunden zu bleiben. Kollektivisten verstehen sich selbst als zugehöriges Gruppenmitglied, dafür betonen und heben sie Gruppenziele, Interessen und Werte über die eines jeden einzelnen Gruppenmitglieds. In diesen Kulturen werden Emotionen als interaktive Erfahrung gesehen, welche den sozialen Kontext eher widerspiegeln, als das innere individuelle Ich. In kollektivistischen Kulturen ist die Expression von Emotionen kontrolliert, da sie sich auf die Bewertung der Beziehung zwischen dem Ich und den anderen stützt.

Trotz der einflussreichen Rolle der Kultur auf emotionale Prozesse, zielen die vorhandenen Studien zu stimmlichen Emotionen auf Mitglieder individualistischer Kulturen ab, aber keine Studie hat die Wahrnehmung stimmlicher Emotionen in einer kollektivistischen Kultur untersucht.

Geschlechtsunterschiede in Wahrnehmung und Ausdruck von Emotionen sind das Thema, welches zunehmend Beachtung in den letzten Jahrzehnten findet. Trotz der steigenden Beachtung haben widersprüchliche Ergebnisse die genaue Natur der dieser vermuteten Unterschiede als Frage offen gelassen. Eine Vielzahl von Psychologen und Anthropologen (z.B. Michael Eid, Batja Mesquita und David Matsumoto) argumentieren, dass für ein umfassendes Verständnis der Geschlechtsunterschiede in Wahrnehmung und Ausdruck von Emotionen und für das Erreichen genauerer Ergebnisse, diese Studien mit Rücksicht auf die beeinflussenden Faktoren (d.h. Sprache und Kultur) wiederholt werden sollten. Diese Dissertation zielt darauf, die Geschlechterrolle in der Wahrnehmung stimmlicher Emotionen in einer syllable-timed Sprache (d.h. Persisch), welche in einer kollektivistischen Gesellschaft (d.h. im Iran) gesprochen wird, zu untersuchen. Hierfür wurden drei Studien geleitet, die im Folgenden kurz beschrieben werden.

Zur ersten Studie gehört die Generierung einer Datenbank zur emotionalen Sprache. Heute sind zahlreiche Datenbanken von stimmlichen Äußerungen der Grundemotionen in mehreren Sprachen inklusive Englisch, Deutsch, Chinesisch, Japanisch, Russisch sowie anderen etabliert. Wie auch immer, für Persisch wurde dies noch nicht erreicht. Forschungsergebnisse dokumentieren, dass die emotionale Prosodie durch linguistische Merkmale einer Sprache beeinflusst werden kann, und dass es essentiell für die Forscher ist, sichere emotionale Stimuli zu generieren, die für den linguistischen Hintergrund der Studienteilnehmer geeignet sind. Es ist zu bedenken, dass für die untersuchte Sprache Sprachaufzeichnungen von allen Grundemotionen für eine belastbare Studie zur emotionalen Sprache erforderlich sind. Hierfür wird in dieser Studie das Ziel gesteckt, diese Lücke für die persische Sprache durch die Generierung eines robusten Satzes validierter Stimuli zu füllen. Derartige Stimuli können den Einfluss individueller Vorlieben minimieren, und können Subjektivität in der Auswahl der Stimuli in zukünftigen Studien zur emotionalen persischen Sprache verhindern.

Die emotionale Sprache studierend, haben verschiedene Forscher oft ihre eigenen, studienspezifischen Satzlisten erarbeitet, ohne jedoch den emotional-lexikalen Inhalt zu validieren. Eine Studie über emotionale Sprache zu leiten, erfordert einen Satz validierter Aussprüche, um die Auswirkung des lexikalen Inhalts von der Prosodie auf die Sprachverarbeitung zu separieren. Hierzu wurden drei Versuchsreihen geleitet (1126 Teilnehmer), um einen validierten Satz an Aussprüchen in persischer Umgangssprache zu generieren (lexical material). Der Output dieser Versuchsreihen ist ein Satz von 90 validierten persischen Aussprüchen, die zuverlässig, entweder mit einer bestimmten Emotion (z.B. Zorn 17 Aussprüche, Ekel 15, Angst 15,Traurigkeit 14, Heiterkeit 15), oder ohne jegliche Emotion (neutral 14) assoziiert sind. Die Aussprüche wurden von zwei persischen Muttersprachlern (ein Mann, eine Frau) mit Schauspielerfahrung in drei Konditionen artikuliert: (1) kongruent (emotionaler lexikaler Inhalt kongruent mit emotionaler Stimme gesprochen), (2) inkongruent (neutrale Sätze mit emotionaler Stimme gesprochen) und (3) Grundlinie (alle neutralen und emotionalen Sätze mit neutraler Stimme gesprochen). Das gesprochene Material umfasst etwa 470 Aussprüche. Die Validität der Datenbank wurde in einer Wahrnehmungsstudie von einer Gruppe von 34 Muttersprachlern bewertet. Voraus gegangene Arbeiten legen nahe, dass das Erkennen stimmlicher Emotionen fast vierfach über der Zufallsrate liegt. Um die bestgeeigneten Exemplare zu selektieren, wurde hierfür die Mindestanforderung der Treffsicherheit im "seven-choice emotion recognition task" (d.h. 71,42%) auf fünf gesetzt. Daher wurden 468 Äußerungen als gültige Porträts der Zielemotionen berücksichtigt. Akustische Analysen der ausgewählten Äußerungen enthüllten Unterschiede in Tonlage, Laustärke und Dauer, jene Attribute, die dem Hörer helfen können, die beabsichtigte Emotion korrekt einzuordnen. Das finale Ergebnis dieser Studie ist eine Datenbank

namens *Persian Emotional Speech Database (Persian ESD).* Diese Datenbank wurde generiert, um als zuverlässige Materialquelle (beides, Text und Sprache) in zukünftigen interkulturellen oder interlinguistischen Studien zu emotionaler Sprache benutzt zu werden, sie ist Forschern für akademische Forschungszwecke zugänglich. Forscher von mehreren wissenschaftlichen Richtungen mögen unsere Datenbank nützlich finden, sie ist aber primär an Psychologen, Neurowissenschaftler, Linguisten und Computerwissenschaftler gerichtet. Die Stimuli der folgenden Studien sind ebenfalls von stimmlichen Porträts der Persian ESD ausgewählt.

In der zweiten Studie untersuchten wir die Geschlechterrolle in der Wahrnehmung emotionaler Prosodie in einem untererforschten linguistischen Kontext (d.h. syllable-timed). Wie eingangs erwähnt, ist Sprache ein Kernbestandteil der Emotionsübertragung. Bei stimmlicher Kommunikation in Verbindung mit linguistischer Entschlüsselung achten die Hörer nicht nur Änderungen der Tonlage, Lautstärke und Stimmqualität, sondern auch auf Änderungen im Rhythmus, um einen Eindruck von der emotionalen Situation des Sprechers zu erlangen. Nichtsdestotrotz, die Geschlechterrolle in der Wahrnehmung stimmlicher Emotionen untersuchend, vernachlässigten die übrigen Studien die *rhythmic division of time*, und fokussierten sich fast ausschließlich auf stress-timed Sprachen (d.h. Deutsch und Englisch). Dies zeigt eine Lücke in der Erforschung geschlechtsabhängiger Effekte in der Wahrnehmung emotionaler Prosodie bei Sprachen mit verschiedener *rhythmic division of time*. Deshalb zielen wir darauf ab, diese Lücke durch die Erforschung der Geschlechterrolle bei der Wahrnehmung emotionaler Prosodie bei Muttersprachlern in Persisch, einer syllable-timed Sprache, zu füllen. In einer Verhaltensstudie wurden 66 persische Muttersprachler gebeten, die emotionale Prosodie eines Satzes validierter Stimmporträts zu erkennen. Diese Stimmporträts waren ein Satz von Aussprüchen mit emotional neutralem lexikalem Inhalt, intoniert in einer von fünf beabsichtigten Emotionen (Zorn, Ekel, Angst, Heiterkeit und Trauer), gesprochen von je einem männlichen und weiblichen Schauspieler. Die Erkenntnisse der Studie deuten darauf hin, dass die Wahrnehmung emotionaler Prosodie signifikante geschlechtsspezifische Unterschiede zeigt, weibliche Teilnehmer übertrafen ihre männlichen Gegenstücke in allen intendierten Kategorien. Unsere Ergebnisse mit persischen Sprechern liegen in einer Reihe mit den Ergebnissen deutscher und englischer Sprecher. Da Persisch eine syllable-timed Sprache ist, deuten unsere Erkenntnisse darauf hin, dass die Natur der *rhythmic division of time* einer Sprache (d.h. stress- vs. syllable-timed) die Entschlüsselung emotionaler Prosodie nicht beeinflusst.

Die dritte Studie befasste sich mit geschlechtsabhängigen Effekten auf emotionale Sprache in einem untererforschten kulturellen Kontext. Die vorhandene Literatur zu geschlechtsspezifischer und emotionaler Sprache ist nur auf Sprecher eines spezifischen kulturellen Hintergrunds gerichtet (d.h.

individualistisch). Da emotionale Prozesse stark durch die Kultur beeinflusst werden und soziokulturelle Variablen Geschlechtsunterschiede im emotionalen Funktionieren vermitteln und mäßigen, mögen die Ergebnisse existierender Studien nicht direkt auf Mitglieder kollektivistischer Kulturen übertragbar sein. Deshalb hinterfragten wir in der dritten Studie den Effekt des Geschlechts von Mitgliedern einer untererforschten Kultur auf die Wahrnehmung emotionaler Sprache (persische Sprecher, die in einer kollektivistischen Gesellschaft leben, d.h. Iran). In diesem Experiment wurden 70 persische Muttersprachler gebeten, die emotionale Prosodie eines Satzes von Aussprüchen mit emotionalem lexikalen Inhalt, welcher mit kongruenter emotionaler Stimme (die fünf emotionalen Kategorien umfassend: Zorn, Ekel, Angst, Heiterkeit und Trauer) von je einem männlichen und weiblichen Schauspieler intoniert wurde, zu erkennen. Die Erkenntnisse aus dieser Studie enthüllten einen signifikanten geschlechtsabhängigen Effekt zugunsten der Frauen. In einer Reihe mit vorangegangenen Studien unter Beachtung der großen Empfindlichkeit von Mitgliedern einer kollektivistischen Kultur gegenüber der Wahrnehmung emotionaler Zeichen, enthüllten die Erkenntnisse der laufenden Studie eine sehr hohe Gesamtgenauigkeitsrate von 95% (d.h. fast die sechsfache Treffsicherheit) von Mitgliedern einer kollektivistischen Kultur. In einer früheren Emotionswahrnehmungsstudie, durchgeführt an Mitgliedern einer individualistischen Kultur, wurde eine Gesamtgenauigkeitsrate von 70% (d.h. fast die fünffache Treffsicherheit) beschrieben. Ein inhaltsbasierender Vergleich der Ergebnisse dieser beiden Studien indiziert, dass Mitglieder einer kollektivistischen Kultur eine höhere Empfindlichkeit gegenüber stimmlichen emotionalen Zeichen zeigen.

Die vorliegenden Studien unterstreichen zusammen die Rolle des Geschlechts in der Wahrnehmung stimmlicher Emotionen in einer syllable-timed Sprache (d.h. Persisch), welche in einer kollektivistischen Gesellschaft gesprochen wird. Generell, das Ergebnis der ersten Studie ist die erste ausführliche Datenbank emotionaler persischer Sprache (Persian ESD), die einen Satz aussagekräftiger, validierter lexikaler und stimmlicher Stimuli umfasst, die fünf emotionale Bedeutungen übermitteln (d.h. Zorn, Ekel, Angst, Heiterkeit und Trauer). Da die Datenbank die drei Konditionen (a) kongruent, (b) inkongruent und (c) Grundlinie abdeckt, bietet sie die einmalige Möglichkeit, die Wirkung von Prosodie und lexikalem Inhalt auf die Identifikation von Emotionen in gesprochener Sprache getrennt zu identifizieren. Die Datenbank kann ebenso im Neuroimaging und klinischen Studien zur Abschätzung der Fähigkeiten einer Person, Emotionen in gesprochener Spreche zu identifizieren, benutzt werden. Zusätzlich kann diese Datenbank neue Möglichkeiten für zukünftige Untersuchungen zur Forschung an Sprachsynthese eröffnen, genau so auch für Geschlechterstudien. Die Erkenntnisse aus den Studien zwei und drei enthüllen einen signifikanten weiblichen Vorteil in der Erkennung emotionaler Prosodie (d.h. neutraler lexikaler Inhalt,

artikuliert mit kongruenter emotionaler Stimme) bzw. emotionaler Sprache (d.h. emotionaler lexikaler Inhalt, artikuliert mit kongruenter emotionaler Stimme). Ein Grund für den weiblichen Vorteil könnte sein, dass Frauen stärker sozial orientiert als Männer sind und sich selbst eher über relationale Belange definieren. Neben diesem Punkt können auch biologische (d.h. hormonelle Unterschiede wie Östrogen und Oxytocin) oder soziale Faktoren (d.h. soziokulturell geschaffene Männer- und Frauenrollen, die von der Gesellschaft anerkannt werden) mögliche Gründe für den weiblichen Vorteil in der Wahrnehmung stimmlicher Emotionen sein. Zusätzlich zu den erwähnten Faktoren ist in dem *Gender Inequality Index (GII) of the United Nations Development Program* gut dokumentiert, dass die Geschlechtsunterschiede im Iran sehr groß sind. Ausführliche Erläuterungen finden Sie in Abschnitt 1.8.2 dieser Dissertation. Nun, Forscher, die die geschlechtliche Diskriminierung im Iran untersuchen, berichten, dass bei Anwendung effektiver Nachahmungsstrategien, eine große Zahl iranischer Frauen es geschafft hat, mit den durch Ungleichheit verursachten kulturellen Barrieren umzugehen. Über diese Nachahmungsstratien hinaus wird von der Entwicklung zwischenmenschlicher Fähigkeiten (d.h. verbale und nonverbale Kommunikation, Hörfähigkeiten, Negoziierung, Problemlösung, Entscheidungsfindung und Selbstbehauptung) berichtet, dass es einer der effektivsten Wege ist, mittels derer es die iranischen Frauen angestellt haben, mit der Ungleichheit umzugehen. Neuere Untersuchungen brachten hervor, dass iranische Frauen höhere zwischenmenschliche Fähigkeiten zeigen, verglichen mit den Männern. Das Vorhandensein höherer zwischenmenschlicher Fähigkeiten mag ein weiterer Grund für den weiblichen Vorteil im Erkennen stimmlicher Emotionen sein. Daneben deuten die Erkenntnisse aus der dritten Studie darauf, dass Mitglieder einer kollektivistischen Kultur eine höhere Sensitivität gegenüber stimmlichen Hinweisen auf Emotionen zeigen. Die Forschung hat gezeigt, dass aufgrund der sehr restriktiven emotionalen Ausdrucksregeln in kollektivistischen Kulturen die kontextuellen Hinweise, die von den interagierenden Partnern bereitgestellt werden, in diesen Kulturen sehr subtil sind. Als ein Ergebnis, den Zusammenbruch zwischenmenschlicher Kommunikation zu verhindern und die zwischenmenschliche Harmonie in solchen kulturellen Kontexten zu bestärken, ist großes Einfühlungsvermögen nötig für die Wahrnehmung und Interpretation der eingeschlossenen übertragenen emotionalen Signale. Dementsprechend mögen die restriktiven emotionalen Ausdrucksregeln der Grund für die hohe Empfindlichkeit der iranischen Teilnehmer gegenüber emotionalen stimmlichen Signalen sein.

In ihrer Gesamtheit stützen die Erkenntnisse aus dieser Dissertation die Annahme, dass die Exaktheit der Wahrnehmung stimmlicher Emotionen (d.h. emotionale Prosodie und emotionales Sprechen) sich signifikant als Funktion des Geschlechts des Decoders (Teilnehmer) unterscheidet. Im Besonderen gibt es einen generellen Wahrnehmungsvorteil für weibliche Teilnehmer gegenüber den männlichen bei

stimmlich emotionalen Wahrnehmungsaufgaben. Desweiteren legen unsere Ergebnisse nahe, dass die Natur der *rhythmic division of time* in einer Sprache (d.h. stress- vs. syllable-timed) die Entschlüsselung emotionaler Prosodie nicht beeinflusst. Zusätzlich unterstützen unsere Ergebnisse die empirische Evidenz für die einflussreiche Rolle des kulturellen Kontexts auf die Wahrnehmung emotionaler Sprache.

Übersetzung: Andreas Götz

Chapter 1: INTRODUCTION*

* As this is a cumulative dissertation based on individual studies and articles, more introductory information is given at the beginning of the individual chapters describing the studies. To avoid boring the reader with too much repetition, this general introduction is therefore kept relatively short.

Introduction

Interest in the study of expression of emotions in voice and face was sparked in the 19th century by the emergence of the evolutionary biology following the contributions by Darwin (1872). The empirical research on the effect of emotions on the voice, however, emerged at the beginning of the 20th century while psychiatrists were trying to diagnose emotional disorders via electroacoustic analysis (e.g., Skinner, 1935). The invention of Telephone and radio also resulted in growing scientific concern with the communication of speaker states by means of vocal cues in speech (Herzog, 1933). Although, systematic investigations on expression of emotions commenced in 1960s when psychiatrists renovated their interest in diagnosing emotional states through vocal expressions (Hargreaves, Starkweather, & Blacker, 1965), linguists and phoneticians discovered the effect of pragmatic information in speech (Mahl & Schulze, 1964), non-verbal communication scientists explored the scope of various bodily channels to carry emotional signals (Feldman & Rime, 1991; Scherer, 1982; Harper, Wiens, & Matarazzo, 1978), emotion psychologists classified the expression of emotion in various modalities (Tomkins, 1962; Izard, 1971; Ekman, 1972, 1992), and engineers specializing in processing of acoustic signals started to make use of ever more refined and practical technology to investigate the effects of emotion on the voice (Lieberman & Michaels, 1962; Williams & Stevens, 1969). In recent years, however, emotion is at the center of interdisciplinary interest as the meeting point of social-humanistic and biological sciences (Davidson, Scherer, & Goldsmith, 2003). Neuroscientists, psychologists, linguists, anthropologists, and philosophers have started to devote increasing attention to interdisciplinary study of emotions and their effects on human behavior and society. Interdisciplinary study of emotions faces the need to have a clear understanding of the theories on emotion (Scherer, 2005). What follows pertains to this issue.

1.1 Theories on emotion

Despite the diversity of emotion theories developed over the past century—they can be classified based on the fundamental question that is central to the theory, i.e., the specific question that each theory seeks to answer. Following this point, one can classify emotion theories into three groups: a) theories with an evolutionary prospect, b) theories with a cognitive basis, and c) theories with a neuro-

psychophysiological perspective[1]. Prior to proceeding with the related literature on theories of emotion, it is necessary to have a preliminary definition of emotions. Although emotions are difficult to define and measure (Plutchik, 1980) and researchers disagree as to whether emotions are best conceptualized as categories (Ekman, 1992), dimensions (Russell, 1980), prototypes (Shaver, Schwartz, Kirson, & O'Connor, 1987), or component processes (Scherer, 2001), a number of researchers would perhaps agree that "emotions are relatively brief and intense reactions to goal-relevant changes in the environment that consist of many subcomponents: cognitive appraisal, subjective feeling, physiological arousal, expression, action tendency, and regulation" (Scherer, 2000, p.138). As a result, for instance, an event may be appraised as harmful, evoking feelings of fear and physiological reactions in the body; people might express this fear verbally and nonverbally and might act in specific ways (e.g., running away) rather than others. In what follows, I will briefly discus each of the three aforementioned perspectives.

1.1.1 The evolutionary perspective

In the evolutionary perspective, the main question considers the phylogenetic development of emotions. Namely, how emotions evolved and what is the biological function of each emotion? Darwin in his book "The expression of emotion in man and animals" (1872) laid the groundwork for the notion that emotions are reaction patterns formed through evolution. As such humans should, more or less, perceive and produce a same set of emotions[2]. Within this approach, emotions are regarded as phenomena with important survival functions. In other words emotions evolved due to the point that humans, as species, faced adaptive problems. Besides, as humans share an evolutionary past with other species, we should await to observe resemblances in the emotions of tightly-related species. Several researchers (e.g., Paul Ekman, Carroll Izard, Alan Fridlund, Sylvan Tompkins, and Robert Plutchik) have taken an evolutionary approach to emotions.

[1] Besides the three mentioned theories, there exists a forth theory of emotions, called *the social constructivist theory of emotion*. In this dissertation we do not describe this theory in details as it is not directly related to our investigations. However, in brief, this theory argues that although emotions are built upon one or more biological systems of behavior, the functional significance of emotion is to be found primarily within the sociocultural system (Gergen, 1985). Based on this theory emotions are among the roles society creates and individuals enact. Emotions are social constructions, and they can be fully understood only upon a social level analysis (Gergen, 1985).

[2] Paul Ekman and Carroll Izard were among the researchers who conducted the first major empirical test of Darwin's hypotheses by examining whether people from disparate cultures can reliably identify the emotions conveyed via certain expressions (see Ekman, 1992). Their discovery, that a small number of emotions are cross-culturally recognized, was a major discovery in research on psychological universals (Shariff & Tracy, 2011).

According to the evolutionary perspective, the key to understanding emotions is to study their functions (Izard, 1993). Therefore, to understand emotions we should consider how they reflect the environment in which they were developed and to which they were adapted. In this regard Oatley and Jenkins (1996) believe that, about 200,000 years ago humans' environment of evolutionary adaptedness was that of semi nomadic hunter–gatherer small groups of people living with each other in large families. They suggest that, most emotions are apparently adapted to living this life style, which involved cooperating in activities such as hunting and raising children. Several of the activities are associated with basic survival problems such as searching for food, avoiding predators, competing for resources, and taking care of children. These problems, therefore, needed particular types of adaptive reactions. A number of researchers (Plutchik, 1994; Scott, 1980) suggest that these adaptive reactions are the prototypes of emotions as seen in humans.

This view of emotions is tightly related to the so-called concept of *fundamental*, *basic*, or *primary* emotions, that is, the notion that there exist a small number of innate, discrete, and universal emotion categories from which all other emotions may be derived (e.g., Ekman, 2007, 2008; Izard, 1992; Plutchik, 1980). Therefore, each basic emotion can be described in terms of an appraisal of goal-relevant events that have recurred during evolution (Power & Dalgleish, 1997). Instances of such appraisals are provided by Oatley (1992, p.55): happiness (sub goals being achieved), anger (active plan frustrated), sadness (failure of major plan or loss of active goal), fear (self-preservation goal threatened or goal conflict), and disgust (gustatory goal violated). Basic emotions can be regarded as fast and economical algorithms (Gigerenzer & Goldstein, 1996) dealing with central life issues under conditions of limited time, knowledge, or computational capacities. However, as human's relation to the environment changed through the time, emotional expressions lost their primary function and acquired a secondary function in emotional communication among individuals (Shariff & Tracy, 2011; Darwin, 1872). For a better understanding of this issue let's consider the emotion *Fear.* Detection of potentially threatening stimuli elicits a number of responses including heavier breathing, the redistribution of blood in preparation for rapid movement, and a marshaling of attentive resources to promote hypervigilance. These are the primary functions of fear which enhance animals' ability to escape a hunter or other possible threats (Darwin, 1872). The facial muscle movements that together constitute a fear expression originally emerged as part of this adaptive emotional response. The widened eyes of a frightened individual increase the scope of his/her visual field and the speed of the eye movements, allowing expressers to better identify (potentially threatening) objects in their periphery (Susskind *et al.,* 2008). Components of the fear expression therefore may be as much a part of the adaptive emotional response as the frightened

affect and quickened heart rate. Although, as human's relation to the environment changed through the time, fear gained a communicative function (a secondary function): signaling alerts of possible threat, and appeasing potential aggressors (Susskind *et al.,* 2008)[1].

Evolutionary psychologists consider human emotions to be best adapted to the life our ancestors led in nomadic foraging bands (Hess & Thibault, 2009). According to the evolutionary theory of emotions, different emotions evolved at different times. Primal emotions, such as fear, presumably evolved among our pre-mammal ancestors. Filial emotions, such as a human mother's love for her offspring, seem to have evolved among early primates. Social emotions[2], such as jealousy, shame, guilt, and pride, evolved among social mammals (Carmen *et al*., 2013; Hess & Thibault, 2009).

Ekman and Cordaro (2011) believe that specific features are found in most (nearly all) basic emotions. These features include: "1) Distinctive universal signals, 2) Distinctive physiology, 3) Automatic appraisal, 4) Distinctive universals in antecedent events, 5) Presence in other primates, 6) Capable of quick onset, 7) Can be of brief duration, 8) Unbidden occurrence, 9) Distinctive thoughts, memories, and images, 10) Distinctive subjective experience, 11) Refractory period filters information available to what supports the emotion, 12) Target of emotion unconstrained, 13) The emotion can be enacted in either a constructive or destructive fashion" (Ekman & Cordaro, 2011, p. 366).

Pursuing Darwin, Plutchik (1962, 1980) founded his model on an emotion wheel including eight basic emotions, with four pairs of opposites, i.e., joy and sadness, fear and anger, disgust and acceptance, surprise and anticipation. Based on Plutchik's theory, it is implausible to undergo opposite emotions in the same moment, i.e. one cannot, for instance, feel fearful and angry at the same time. Although the number of basic emotions differs from researcher to researcher (Scherer, 2000), happiness, anger, fear, sadness, disgust, and surprise are the emotions usually included in the list of basic emotions (c.f. Cornelius, 1996).

[1] For a detailed list of the primary and secondary function of emotions see Shariff and Tracy (2011).

[2] Social emotions (e.g., embarrassment, guilt, empathy, shame, and jealousy) can be defined as those emotions which require the representation of the mental states (e.g., their opinion or feelings) of other people (Burnett *et al*., 2009; Hareli & Parkinson, 2008). For instance the three emotions of embarrassment, shame, and jealousy necessarily depend on other people's thoughts feelings or actions, as experienced, recalled, anticipated or imagined (Hareli & Parkinson, 2008; Planalp, 1999; Parkinson, 1996). As a result, researchers (e.g., Burnett *et al*., 2009) believe that, the development of social emotions is tightly linked with the development of social cognition, the ability to imagine other people's mental states, which generally develops in adolescence (i.e., the period between puberty onset and the attainment of a stable adult role). Studies have revealed that that children as young as 2 to 3 years of age can express emotions resembling guilt (Hoffman, 2001). However, while five-year-old children are able to imagine situations in which basic emotions would be felt, the ability to describe situations in which all social emotions might be experienced does not appear until seven years of age (Pons, Harris, & de Rosnay, 2004; Harris *et al.,* 1987).

More recently, however, Ekman and Cordaro (2011) mention that there is evidence for universality in the following seven emotions: anger, fear, surprise, sadness, disgust, contempt, and happiness[1].

It is to be acknowledged that the concept of basic emotions has been the subject of controversy (cf. Ortony, & Turner, 1990; Turner, & Ortony, 1992; Panksepp, 1992; Panksepp & Panksepp, 2000). So far research has documented both positive (Levenson, 1992) and negative (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000) views regarding the concept of basic emotions. Arguably, the strongest evidence of basic emotions is derived from studies of facial expression of emotions (Ekman, 1973, 1992). Regardless of the existing controversy and the fact that this approach leaves little room for cognitive aspects of emotions, the concept of basic emotions has been appealing to various researchers (e.g. Plutchik, 1980; Ekman, 1972, 2003, 2008, LeDoux, 2000; Ekman & Cordaro, 2011).

1.1.2 The cognitive perspective

The cognitive perspective is one of the dominant approaches to emotion theory today. Very often the beginning of the cognitive approach is traced back to the work of Magda Arnold (1960) (c.f. Cornelius, 2000). Although one could say that basically, this line of research was shaped by the debate between defenders of the sensory-feedback approach to emotion theory (e.g. James, 1884) and Cannon and Bard's theory of emotion (Cannon, 1927; Bard, 1928). Whereas the James-Lange theory considered sensory processing to elicit bodily changes which are assumed to be experienced as emotions (e.g., increase in heart rate), the Cannon-Bard theory contradicted this idea, claiming that bodily changes were not specific enough to account for all distinct emotions (e.g., both anger and fear might elicit an increased heart rate, but are nevertheless two different emotions). Alternatively, Cannon and Bard suggested that physiological changes and subjective experience arise simultaneously and that the same pattern of

[1] Ekman and Cordaro (2011) define each of these seven emotions as follows: "**Anger**: the response to interference with our pursuit of a goal we care about. Anger can also be triggered by someone attempting to harm us (physically or psychologically) or someone we care about. In addition to removing the obstacle or stopping the harm, anger often involves the wish to hurt the target. **Fear**: the response to the threat of harm, physical or psychological. Fear activates impulses to freeze or flee. Often fear triggers anger. **Surprise**: the response to a sudden unexpected event. It is the briefest emotion. **Sadness**: the response to the loss of an object or person to which you are very attached. The prototypical experience is the death of a loved child, parent, or spouse. In sadness there is resignation, but in can turn into anguish in which there is agitation and protest over the loss and then return to sadness again. **Disgust**: repulsion by the sight, smell, or taste of something; disgust may also be provoked by people whose actions are revolting or by ideas that are offensive. **Contempt**: feeling morally superior to another person. **Happiness**: feelings that are enjoyed, that are sought by the person. There are a number of quite different enjoyable emotions, each triggered by a different event, involving a different signal and likely behavior. The evidence is not as strong for all of these as it is for the emotions listed above" (Ekman & Cordaro, 2011, p.365).

response (i.e., a stimulus is relayed from the thalamus to internal organs and the cortex) underlies all emotions (Breedlove, Watson, & Rosenzweig, 2010). Based on their theory, the thalamus plays a dominant role by sending sensory information to the cortex for interpretation and at the same time sending activation signals via the hypothalamus to the rest of the body. However, other types of cognitive theories on emotion became more prominent as neither approach served to be sufficient as a comprehensive theory of emotion (Breedlove *et al.,* 2010).

The main assumption of the cognitive approach and its related tradition of research is that emotions and thoughts are inseparable. More explicitly, cognitive theories on emotion build their work on a concept called *appraisal*, the process by which objects, events, or persons are judged as being either positive or negative (Oatley & Johnson-Laird, 2014) triggering an emotional response within us. Appraisal theory asserts the existence of a causal relationship between appraisals and emotions, meaning that activation of a specific appraisal pattern brings about the corresponding emotion (Scherer, Schorr, & Johnstone, 2001; Demir, Desmet, & Hekkert, 2009). As Cornelius (2000, p.21) claims, "These patterns provide the link between particular characteristics of the person or organism, his or her learning history, temperament, personality, physiological state and particular characteristics of the situation in which the person or organism finds him or herself. The notion of appraisal, (...), goes hand in hand with the idea that emotions are action tendencies". Altogether, the cognitive theories on emotion suggest that the cognitive state of an individual is involved in eliciting an emotion.

The cognitive approach to emotion is still appealing to researchers and is still growing (Oatley, Parrott, Smith, & Watts, 2011; Keltner, Oatley, & Jenkins, 2013). Among the modern supporters of this approach are, for instance, Lazarus (1991) and Scherer (2000, 2001). Scherer (2000) has developed the component process model (CPM). In the framework of the component process model, Scherer describes emotions "as an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organisms" (Scherer, 2001, p.93). In fact, Scherer suggests to use the term *emotion* just for those moments during which various organismic subsystems are synchronized or coupled to produce an adaptive reaction to an event which is regarded as essential to the individual's well-being (Scherer, 2001). He further mentions that as in other areas of organismic functioning, emotion interacts with phylogenetically older response mechanisms like reflexes and fixed action patterns. Scherer (2001, 2005) suggests that, emotion is a theoretical construct that consists of five components[1] corresponding to

[1] These components correspond do the aspects or features of emotion which have been postulated by most major emotion theorists (Scherer, 2001). For more information in this regard see Kleinginna and Kleinginna (1981).

five distinctive functions i.e., the cognitive component (appraisal), the neurophysiological component (bodily symptoms), the motivational component (action tendencies), the motor expression component (facial and vocal expression), and the subjective feeling component (emotional experience). For a list of these functions, the systems which subserve them and the respective emotion components see table 1.1. As stated by (Scherer, 2005) any change in the state of these components originates from appraisals. According to CPM, the appraisal component is significant due to the point that it serves the check of external or internal stimuli based on their relevance, their implications, the coping potential, and the normative significance (Scherer, 2009).

Contrary to the evolutionary approach, the component process model of emotion does not introduce a limited number of emotions, rather it believes in the indeterminate number of emotions which are constrained by general conditions of life and of social organization (Scherer, 2009).

Table 1.1 Relationships between the functions and components of emotion and the organismic subsystems that subserve, proposed by Scherer (2005, p. 697)

Emotion function	Emotion component	Organismic subsystem (and major substrata)
Evaluation of objects and events	Cognitive component (appraisal)	Information processing (CNS)
System regulation	Peripheral efference component	Support (CNS, NES, ANS)
Preparation and direction of action	Motivational component	Executive (CNS)
Communication of reaction and behavioral intention	Motor expression component	Action (SNS)
Monitoring on internal state and organism environment interaction	Subjective feeling component	Monitor

Note: CNS = central nervous system; NES = neuro-endocrine system; ANS = autonomic nervous system; SNS = somatic nervous system.
The organismic subsystems are theoretically postulated function units of networks.

Scherer's component process model is specifically important for research on acoustic features of speech and emotion (Scherer *et al.,* 2001). As with the help of this model, he could hypothesize various detailed physiological predictions that can be associated with certain emotions (Scherer, Ladd, & Silverman, 1984). For example, the emotion of anger is usually conveyed with high pitch and a fast speech rate across various cultures (see table 1.2 for detailed example of Scherer's hypotheses)[1]. Recently, Scherer, Clark-

[1] As in this table only the three acoustic parameters are considered, it does not have much discriminatory power.

Polner, Mortillaro (2011) argue that, besides the mentioned elements (i.e., pitch, intensity and speech rate) other elements of speech which might be cross-culturally associated with certain emotions in vocal emotion recognition tasks is an open question for further investigations.

Table 1.2 Predicted emotion effects for selected acoustic parameters according to the component process theory by Scherer (1986)

Acoustic Parameter	Emotions				
	Anger	Disgust	Fear	Happiness	Sadness
F0	>	<>	>	>	<
Intensity	>	>	>	>	<
Speech Rate	>	?	>	>	<

Note: F0 = fundamental frequency; < = decrease; > = increase; ? = no prediction made
All predictions were made in comparison to "normal" speech.
No general predictions for the emotion of pleasant surprise have been made.

Three cognitive theories of emotion which continue to develop productively are *the action-readiness theory* (Frijda, 1986, 2007; Frijda & Parrott, 2011), *the core-affect theory* (Russell, 2003), and *the communicative theory* (Oatley, Johnson-Laird, 1987, 2011). Research has documented (Oatley & Johnson-Laird, 2014) that these theories concur in postulating an initial automatic appraisal which does not require conscious processing, followed by a secondary appraisal which includes conscious reflection on the meaning of the emotion, and a third phase of appraisal which is social (i.e., when emotions are verbally confided to others). However, referring to the situations where emotions are generated by non-cognitive mechanisms such as electrical stimulation of the brain or use of psychoactive substances (e.g., LSD), a number of researchers (Roseman 2013; Moors 2013) have questioned the role of cognition in generation of all emotions. Obviously, to reach more comprehensive results, more studies are to be carried out in this realm.

1.1.3 The Neuro- and Psychophysiological Perspective

The main concern of psychophysiological and neurophysiological theories on emotion is to find out which physical central nervous and peripheral-psychological processes and structures associate with an emotion. Psychophysiological and neurophysiological theories on emotion are interested to describe which central nervous system brain processes and brain structures are involved in the origin of a specific

emotion (Papez, 1937 ; MacLean, 1952; Panksepp, 1998), and tackle the question of whether different emotions correlate with specific peripheral-physiological patterns. In brief, these theories of emotion seek to clarify the relationships between emotions and the brain structure (Panksepp, 2007b). To this aim, a large number of studies have been conducted on both animals and humans (e.g. Panksepp & Watt, 2011; Panksepp, 2008, 2005a, 1990; LeDoux, 1996; Adolphs, Tranel, Damasio, & Damasio, 1994, 1995). For example, early theories focused on subcortical structures which might be involved in emotional functions. A pioneer in this realm was Bard (1928). While Bard assumed that the hypothalamus plays a crucial role in emotional processing, Papez proposed a much more complex anatomical model of emotions (Papez, 1937). According to Papez's view, various brain structures might *mediate* various emotion components. Within Papez's model, he postulated that the mechanism which embosses the function of central emotions is comprised of the hypothalamus, the anterior thalamic nuclei, the cingulate gyrus, the hippocampus, and interconnections between these structures. Besides, he suggested that functions of emotional evaluation and expression are mediated mainly by the hypothalamic component of the circuit, whereas the elaboration of emotional experience is mediated mainly by the cingulate gyrus, which is the cortical component of the circuit.

More recently, Panksepp (2005b, 1988) has categorized seven basic *emotional operating neural circuits* or *primary-process affective systems* in the brain. These *circuits* which are homologous neuro-evolutionary foundations for emotional experience include: 1) SEEKING (desire), 2) FEAR (freeze/flight), 3) RAGE (frustration), 4) LUST/PLEASURE (passion), 5) CARE (empathy, nurturing), 6) PANIC

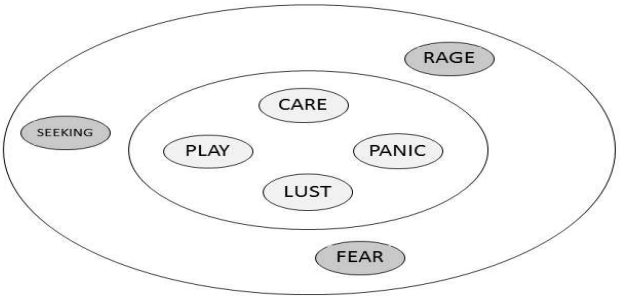

Fig.1.1 The Seven basic Emotional Operating Neural Systems

(distress/sadness), and 7) PLAY (joy) system[1] (cf. Panksepp & Watt, 2011). The basic emotional operating neural circuits are depicted in Figure 1.1. It is to be noted that in this figure, the inner four circuits represent what Panksepp describes as the social emotions, while the outer three circuits reveal the more basic emotional and motivational processes[2].

As a Darwinian neuro-evolutionist, Panksepp holds that these seven emotional operating neural circuits shared by all mammals, are embedded in ancient regions of the brain; they are evolutionary memories built in the nervous system at a fundamental level (Panksepp, 2008). In fact he believes emotions are essential to our survival. Categorizing seven basic emotional operating neural circuits in the brain does not imply that Panksepp discounts other emotions – such as jealousy. Rather, he believes that *raw emotional feelings* are produced via these seven brain emotional operating neural circuits, and that they are *the building blocks* of other emotions (Panksepp, & Watt, 2011).

Although, results of a number of existing studies conducted on the basis of the neurophysiological theories (e.g., Panksepp, & Biven, 2011; Panksepp, 2007a, 2007b; Panksepp, 2005b; Ciompi, & Panksepp, 2005; Knight, Smith, Cheng, Stein, & Helmstetter, 2004; LeDoux, 2000; Damasio, 1999; Gray, 1990) have documented that mainly the subcortical brain regions (i.e. amygdala, the hypothalamus, the basal frontal brain, and the brainstem) are responsible for emotions, scientists are not yet able to map out explicitly each specific emotion to a particular brain region. Definitely, to reach more comprehensive results, further investigations are to be performed in this realm.

Taken together, each of the three discussed approaches to emotion has their own arguments as well as counter arguments. Continuing these discussions and searching for further details on the theoretical plausibility or empirical sufficiency of these approaches is not in the focus of this dissertation. In fact, having this general understanding of the mentioned theories on emotion, instead of searching for further details on the theoretical plausibility or empirical sufficiency of these three approaches to emotion, it is essential to select the appropriate theory which fits the research goals. In this dissertation the aims of upcoming studies are: **a)** to generate a comprehensive emotional speech database for Persian, **b)** to investigate the accuracy rates of Persian speakers in vocal emotion recognition tasks, and **c)** to draw a content-based, culture- and language-wise comparison with previous similar studies performed on a different language and culture. Therefore, it is essential to select a set of emotions which are reported to

[1] Panksepp notation is to use capitals – thus CARE for instance, when he is specifically referring to the CARE Emotional Operating Neuro Circuit (System), as opposed to when using this term in a more conventional way.

[2] Figure 1.1 is adapted from Ross (2010, p.140, Figure 4.1).

be biologically innate and cross-culturally recognizable. Following a number of researchers (LeDoux, 2000; LeDoux & Phelps, 2008; Panksepp, 2008) we assume that human beings possess a set of biologically innate emotions produced via a core emotion system[1]. Taking into account that expressions of *basic emotions* (i.e., anger, disgust, fear, happiness, and sadness) can be recognized pan-culturally from the face, a number of researchers (Pell, Monetta, Paulmann, & Kotz, 2009; Thompson, & Balkwill 2006; Van Bezooijen, Otto, & Heenan, 1983) have empirically shown that, regardless of an individual's culture or linguistic ability this set of basic emotions can be recognized from a speaker's voice. Similarly, Matsumoto (2006) argues that with regard to emotion communication, all individuals start with a same base of universal, pan cultural expressions. Then we learn how to manage and modify these emotional expressions based on cultural display rules. Consequently, in this dissertation we take up the concept of basic emotions, and with a focus on a set of five emotional categories which are well recognized across various languages and cultures i.e., anger, disgust, fear, happiness, and sadness (Izard, 2007; LeDoux, & Phelps, 2008; Panksepp, 2007a, 2008; Ekman, & Cordaro, 2011), we will explicitly examine the role of gender in the recognition of vocal emotions. In this dissertation the term *vocal emotions* encompass two categories i.e., 1) emotional speech: emotional lexical content articulated in congruent emotional prosody and, 2) emotional prosody: neutral lexical content conveyed in emotional prosody).

1.2 Vocal expression of emotions

The most advanced level of vocal expression includes voluntary control over the exact acoustic patterns of vocal expression. It includes the capability to learn vocal patterns by imitation, as well as the production of new patterns by invention. These abilities are vital in the uniquely human creation of language (Juslin & Laukka, 2003). Arguably, human voice is an extremely complicated means of communication. Using our voice, not only we transfer linguistic meaning, but also we communicate our emotions (Karpf, 2006). Communicating emotions via voice, we can use linguistic cues, prosodic cues, or nonverbal vocalizations such as sighs, sobs, and laughter. By linguistic cues we mean lexical content of the spoken utterance such as words, phrases, and sentences which carry emotional meaning. But what are prosodic cues? Psychoacoustic parameters of speech such as Pitch, or fundamental frequency (F0), intensity or loudness, together with rhythm and speech rate or duration, form suprasegmental

[1] In this dissertation, we adopted this viewpoint as a point of departure. However, it is to be mentioned that despite a sizeable amount of research on the concept of *basic emotions*, this concept is still under debate (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000; Panksepp, 1992; Ekman, 1992).

parameters of speech or *prosody*. The term prosody was first introduced by Monrad-Kohen (1947), who defined it as "that faculty of speech which conveys different shades of meaning through of variations in stress and pitch-irrespective of the word and grammatical construction". Crystal (1995) states that pitch (or F0, measured in Hz) is the most important suprasegmental feature of speech. We use pitch variation to express non-lexical information. For instance, in a question in English (as well as in many other languages such as Persian and German), pitch is often raised near the end of a sentence. However, a declarative sentence is usually marked by a falling pitch near the end. Intensity or loudness (measured in dB) is the second important aspect of prosody. Whereas a raised voice is often associated with anger, a lowered voice may reveal sadness or fear felt by the speaker. Duration or speech rate (measured in s or ms) is the third of the three basic parameters of prosody (Van Lancker & Sidtis, 1992). Duration is attributed to the time structure, or tempo of speech units such as syllables, words, or pauses.

According to Shih and Kochanski (2002), prosody serves four different functions: 1) to express lexical meanings (e.g. in tone languages like Chinese), 2) to pass on non-lexical information via intonation (as in questions and declarative sentences), 3) to express discourse functions, e.g. new information is mostly accented (i.e., prominent in terms of differential loudness, or of pitch, or length, or of a combination of these) whereas old information is often deaccented; and 4) to convey emotions, e.g. happy surprise is conveyed in high speed and high pitch. Taken together, we can classify prosody in two different categories: a) linguistic prosody, and b) emotional prosody. *Linguistic prosody* includes non-verbal aspects of human language and provides us with a rich source of linguistic information, whereas *emotional prosody* encompasses non-verbal aspects of human language and provides a rich source of information about a speaker's emotions and social intentions (Banse & Scherer, 1996; Wilson & Wharton, 2006).

Aside from emotional prosody, as mentioned earlier, emotions are also conveyed verbally through the lexical content of the spoken utterance. These two channels of information (prosody and lexical cues) are inextricably linked and may reinforce or contradict each other (e.g., by conveying sarcasm or irony; Pell & Kotz, 2011). Therefore, to interpret the intended meaning or the emotions and the attitudes of a speaker, listeners should effectively monitor both prosodic and lexical information (Tanenhaus & Brown-Schmidt, 2007).

1.3 Language and rhythmic division of time

In linguistics, *rhythm* is defined as the repetition of patterned sequences of elements, which often vary in prominence (Fraisse, 1982). Rhythm is an important aspect of prosody (Wells, 2006) pertaining to the way languages are organized in time (Patel & Daniele, 2003)[1]. The postulated classification of stress-timed rhythm and syllable-timed rhythm (Abercrombie, 1967) is the foundation of phonetic research on rhythm (Abercrombie, 1967; Lehiste, 1977; Bertinetto, 1989). While this distinction is not totally uncontroversial (see Jassem & Witten, 1984; Vihman, Nakai & De Paolis, 2006, for details), it is regularly used to make a prosody-wise division between languages, which are said to be (primarily) stress-timed, or (primarily) syllable-timed (Roach 1982)[2].

According to Windfuhr (1979) there is a stark contrast[3] between stress-timed languages (where the amount of time to say a sentence depends on the number of stressed syllables) and syllable-timed languages (where the amount of time to say a sentence depends on the total number of the syllables) (see Patel & Daniele, 2003, for a detailed description). In other words, stress-timed languages (e.g., German, English) tend to have an equal amount of time elapsed between stressed syllables and do not place equal stress on every syllable. An example that makes this idea clear is as follows: a) *I eat pizza*, and b) *I might have been eating pizza*. It might be surprising to non-native English speakers that these two sentences take exactly the same duration of time to say. This is because in English the important, or meaningful words in a sentence, are stressed. In both sentences given above, the only words that are stressed are "I", "eat", and "pizza" (Doan, 2013, P.17). Whereas English, German and Russian are widely acknowledged by linguists as examples of stress-timed languages (Windfuhr, 1979), Persian, Turkish and French are regarded as instances of syllable-timed languages.

Language is a crucial factor in transmission of emotions (Scherer, Banse, & Wallbott, 2001). Likewise, Pell *et al*. (2009) claim that, vocal expressions of emotion manifested in emotional prosody, are inherently dynamic and integrated with linguistic properties of the language under study. Respectively, Pell (2001) mentions that, linguistic-specific features of a language (e.g., rhythm) affect vocal expressions of emotion (i.e., emotional prosody). During vocal communications, in conjunction with linguistic decoding; listeners attend to changes in not only in pitch, loudness and voice quality, but also to changes

[1] As Schluter (2005) suggests, the distinction between the different rhythmic types is based on the fact that all human speech is split into temporal intervals that have a tendency to be of similar length.

[2] Besides the two very common stress-timed and syllable-timed classifications, a few number of languages (e.g., Japanese, ancient Greek, Hawaiian, Gilbertese, and Vedik Sanskrit) are classified as mora-timed (Clark, Yallop, & Fletcher, 2007), in which the duration of every mora is equal. A *mora* is a unit of phonological quantity based on both vowel quantity and the occurrence of consonants within a given syllable (Clark, Yallop, & Fletcher, 2007).

[3] One such contrast is durational variability in consonant sequences (Ramus, Nespor & Mehler, 1999).

in rhythm, to form an impression about the speaker's emotional state (Wilson & Wharton 2006). Considering all the above mentioned points, it is important to take into account the distinction of rhythmic division of time while studying recognition of emotional prosody. Contrary to research on the facial expressions of emotion, researchers interested in vocal expressions of emotion have not presented a sufficiently detailed table of vocal elements representing the vocal attributes of an emotion (Pell *et al*., 2009). Therefore, the probable effect of the rhythmic division of time (i.e., stress- vs. syllable-timed) on the recognition of vocal expressions of emotion (emotional prosody) is still an open question.

1.4 A short overview of Persian language (Farsi)

Persian (also known as Farsi) is an Indo-Iranian language, a sub-branch of the Indo-European language family (Anvari & Givi, 1996). It is spoken by almost 110 million people around the world, and has official-language status in Iran, Tajikistan, and Afghanistan (Sims-Williams & Bailey, 2002). As for the number of speakers from among the estimated 6,000 languages spoken in the world today, Persian is ranked sixteen. In Iran, Persian is the language of government and public instruction, the mother tongue of almost half of the Iranian population, and a second language for a large proportion of the rest of the Iranians (Moseley & Asher, 1994).

In Persian, like other natural languages, sounds are produced with a pulmonic egressive airstream mechanism. Modern Persian includes a total of about 30 phonemes including 23 consonants, 6 vowels and one diphthong (Mahootian, 1997). These vowels are divided into two groups of lax (/æ/, /o/, /ɛ/) and tense vowels (/ɒ/, /i/, /u/) vowels. There is another phoneme, /ow/, which although is a diphthong, behaves similar to vowels. As for syllables structure[1], according to the most popular categorization there are three syllables in Farsi namely CV, CVC, and CVCC (Comrie 1987; Samareh, 1977).

1.5 The concept of sex and gender

Prior to any investigations it is essential to have a clear understanding of the concept of gender and sex. The importance of differentiating between the terms *gender* and *sex* has been challenged along

[1] Persian syllables structure is very briefly mentioned here, as in the dissertation we focus on Persian as a syllables-timed language.

various lines (e.g., Rahman & Jackson 2010; Nobelius, 2004, Haig, 2004; Baden & Goetz, 1998; Udry, 1994; White, 1993). The term *sex* is considered to reflect the biologically determined categories of men and women according to certain identifiable physical features such as reproductive organs and functions assigned by chromosomal complement (Rahman & Jackson 2010; Haig, 2004), whereas the term *gender* refers to biologically rooted, socio-culturally constructed roles that a given society considers appropriate for men and women (Pardue & Wizemann, 2001). As Jackson (2006, p. 106) states, "Gender … encompasses the division or distinction between women and men, female and male, these binary categories themselves and the content of those categories – the characteristics and identities embodied through membership of them". Gender is thus a social division and a cultural distinction, given meaning and substance in the everyday actions, interactions and subjective interpretations through which it is lived. Considering the mentioned points, the use of the term *gender* in this dissertation is intentional. Likewise, in this thesis the terms women and men, female and male are used to refer to the concept of gender[1].

1.6 Gender and emotion

The idea of gender differences in emotional responding forms one of the most robust gender stereotypes there is (Timmers, Fischer, & Manstead, 2003). In popular eastern and western culture, common wisdom holds that women are the more emotional gender, experiencing and expressing emotions in general more than do men (Shields, 2003). Stereotypes of women being more acquainted with their emotions are often acknowledged by both men and women (Hess *et al*., 2000; Belk & Snell, 1986). These beliefs permit society members to regulate behavior, permit particular actions, restrict or punish others, and determine what is accepted or censured in a society (Fairclough, 2003). Over the past decades to examine this widespread consensus empirical studies have increasingly focused on gender differences in expression and recognition of emotion. With the exception of anger, the mentioned gender stereotypical belief which regards females as the more emotional gender, is particularly pronounced for the behavioral expression of discrete emotions (e.g., fear, disgust, happiness, and sadness) (Grossman & Wood, 1993; Shields, 2003). Thus, as far as expression of emotions is concerned, the belief that women

[1] It should be noted that although often the concept of gender is regarded as a dichotomy with two distinct categories (i.e., male/female; man/woman), various researchers argue that gender is a linear spectrum ranging from 100% female to 100% male including various states in between (Bockting, 2008). Yet as a point of departure, in this dissertation we focus on the binary classification of gender.

are more emotional than men is strongly held across individuals and across different emotions (McRae *et al.*, 2008).

In the domain of emotion recognition, in spite of this expanded attention, conflicting findings have left the exact nature of these presumed differences an open question. What we know about gender differences in emotion recognition has been summarized in four meta-analyses so far (Hall, 1978, 1984; McClure, 2000; Thompson & Voyer, 2014). Hall's (1978) review was the first to summarize gender differences in non-verbal (auditory, visual, audio-visual) displays of emotions. She summarized the results from 75 studies with adults and children, showing that females consistently outperform males on the recognition of vocal and facial emotional expressions. In a second independent meta-analysis in 1984, Hall reported similar results suggesting that females significantly outperformed males. In this second review, she reported that female advantage was more conspicuous in the more recently published experiments, arguing that more accurate instruments and more powerful statistical analyses were the likely reason (Hall, 1984). In a third meta-analysis McClure (2000) investigated gender effects on the recognition of facial expressions in three age groups (infants, children and adolescents). She examined the findings of 60 adolescent and child studies and 6 infant studies. Overall, her findings indicated that there was a gender effect in favor of females in recognizing non-verbal emotional displays. Recently, in a multilevel meta-analysis of 215 studies, Thompson and Voyer (2014) examined the existence of gender differences in the recognition of non-verbal (auditory, visual, audio-visual) displays of emotion. They reported a small overall advantage in favor of females on emotion recognition tasks. Nevertheless, they claimed that the magnitude of the gender difference was moderated by various factors including specific emotion category, emotion type (positive, negative), sensory modality (auditory, visual, audio-visual), gender of the actor, and age of the participants. Nevertheless, they argue that the potential effect of the sensory modality on the recognition of emotions is still an open question and more studies are required on gender effects on the recognition of emotions (Thompson & Voyer, 2014).

Taken together, these studies support the broadly held notion that compared to men, women are the more emotional gender. Although, as documented by Eid and Diener (2001), Mesquita (2001) and Matsumoto (2001), for a comprehensive understanding of the issue and to achieve more accurate results, these findings should be replicated with regard to other influential factors, such as language (Pell, 2001) and culture (Matsumoto, 2001).

1.7 Culture and emotion regulation

It is difficult to define the term *culture* as it has been used in different ways within and across disciplines. Roughly speaking, culture designates shared commonalities within groups of people (Matsumoto, 2007; Triandis, 1994). In the context of this dissertation, the term *culture* refers to shared elements (e.g., rituals and habits, belief and value system, social institutions, and behavior patterns) within any interactive collectivity of people (Scherer & Brosch, 2009) as well as the basis of the individual preferences in daily life (Matsumoto, Grissom & Dinnel, 2001). In cross-cultural research (e.g., Scherer, Banse, & Wallbott, 2001) culture is often depicted by country (Matsumoto, Takeuchi, Andayani, Kouznetsova, & Krupp, 1998).

Investigating the relationship between emotions and culture dates back to 1872 when Darwin claimed emotions and the expression of emotions are universal (Darwin, 1989). Since then, researchers following the evolutionary perspective, acknowledge that the ability to recognize and express emotions is an initial human potential. A number of cross-cultural endeavors concerning recognition of facial and vocal expression of emotions have empirically supported theories of universality of human emotional capability, specifically concerning the so-called *basic emotions* such as anger, fear, happiness and sadness (e.g., Pell *et al*., 2009; Pell & Skorup, 2008; Thompson & Balkwill, 2006; Scherer & Wallbott, 1994; Ekman, 1973; Ekman, Friesen, & Ellsworth, 1972).

Although emotions are indispensable in our lives, we do not experience emotions in the same way. From the view point of cross-cultural psychology and anthropology, the exclusively biological perspective is not adequate for a comprehensive understanding of the perception and expression of emotions (Matsumoto & Hwang, 2011; Eid & Diener, 2001; Mesquita, 2001; Matsumoto, 2001). Recent psychological studies suggest that emotions and emotional experience are largely influenced by culture (Matsumoto & Hwang, 2011; Tsai, Knutson, & Fung, 2006; Mesquita, 2003; Matsumoto, 2001; Wierzbicka, 1994; Markus & Kitayama, 1991; Matsumoto, 1989). Namely, how emotions are perceived, expressed, and regulated differs as a function of cultural norms of each society (Tanaka *et al*., 2010; Tsai *et al*., 2006; Matsumoto, 1989). Further research (Matsumoto, 1992; Ekman, 1972) has documented that cultural norms regulate the display of emotion in cases in which display of specific emotions might be potentially disruptive to social interaction, as well as the decoding of emotions when recognizing theses emotions would disrupt social interaction. Accordingly, cultural variation in the accuracy of emotion recognition is attributed to the norms that cultures impose on their members as for recognition and expression of

specific emotions. In the same way, Mesquita (2003) suggests that emotional experiences vary greatly across cultures and that cultural models are essential to understand and predict these variations. Accordingly, for a full understanding of emotional processes we should consider the cultural context where emotions are perceived, expressed, and experienced[1] (e.g., Richeson & Boyd, 2005; Mesquita & Ellsworth, 2001; Scherer, 1997; Kitayama & Markus, 1995; Scherer & Wallbott, 1994, Markus & Kitayama, 1991). As Consedine, Magai and Bonanno (2002, p.217) note, "Through socialization, culture determines the kinds and frequencies of events to which children are exposed; the kinds of reactions that receive scaffolding and support or, conversely, punishment and containment; and the kinds of social expectations for comportment". Likewise Mesquita (2001) believes that emotion and culture are intermingled with each other. On the one hand, emotions are shaped in a way comparable to ideas of the cultures in which they develop; On the other hand, emotions reinforce the cultural themes of a given society. As such, in any given cultural context, behaviors and emotional experiences that fit with and enhance the goals of that specific culture are reinforced (Fischer & Manstead, 2000). In support of this perspective, various studies have reported cultural differences in how emotional expressions are decoded (e.g., Mesquita & Markus, 2004; Fischer & Manstead, 2000; Matsumoto, Kasri & Kooken, 1999; Matsumoto & Kudoh, 1993; Matsumoto,1992, 1989), in when and how emotions are expressed (e.g., Ekman, 1972; Matsumoto, 1993; Stephan, Stephan, & de Vargas, 1996; Tsai & Chentsova-Dutton, 2003), in the events that elicit emotions (see Mesquita & Frijda, 1992 for a review), in the frequency and intensity of emotional experiences (e.g., Scherer, Matsumoto, Wallbott, & Kudoh, 1988; Soto, Levenson, & Ebling, 2005), and in the way by which emotion-eliciting events are assessed (e.g., Matsumoto, Kudoh, Scherer, & Wallbot, 1988; Roseman *et al.,* 1995; Scherer, 1997).

1.7.1 Emotional display rules

Ekman and Friesen (1969) argue that cultural variation in social interactions is manifested in *norms* for decoding and displaying emotions (Ekman & Friesen, 1969). Matsumoto and Hwang (2011) argue that cultures create norms concerning the regulation of emotion to facilitate social coordination. They further mention this is due to the point that emotions are initial motivators of behavior. These norms are called *emotional display rules*. The concept of emotional display rules was introduced for the

[1] Research (Lu & Yang, 2006) has documented that cultures are changing structures and can be largely influenced by shifts in political and economic organizations. Nevertheless, it is essential to consider different cultural contexts (e.g., collectivist vs. individualist) as they affect the expression and perception of emotion (Matsumoto, & Hwang, 2011; Tsai *et al.,* 2006; Eid & Diener, 2001).

first time by Ekman and Friesen (1969) as a term in a study on culture and emotion to explain the observed differences between Japanese and Americans. Emotional display rules are cultural norms observed as an important part of any culture; they refer to culturally prescribed rules learnt early in life via socialization (Koopmann-Holm & Matsumoto, 2011). These rules affect the emotional expression of people from any culture depending on what that particular culture has featured as an appropriate or inappropriate expression of emotion (Ishii, Reyes, & Kitayama, 2003; Matsumoto, Kasri, & Kooken, 1999). These culturally shared norms and unwritten codes dictate *when*, *how* and *to whom* individuals should express their emotional experiences. (Matsumoto 1990). Emotional display rules may magnify, deamplify, neutralize, qualify, or mask an emotional response (Ekman & Friesen, 1969). Further research (Matsumoto, 1990; Koopmann-Holm & Matsumoto, 2011) has documented that display rules not only encompass the dimension of expression appropriateness but also appropriateness of recognition and evaluation of a behavior.

Cultural differences via these norms regarded as *display rules* are well established in psychology (Koopmann-Holm & Matsumoto, 2011; Thompson & Balkwill, 2006; Ishii *et al.*, 2003; Matsumoto *et al.,* 1998; Matsumoto, 1990; Malatesta & Haviland, 1982; Ekman & Friesen, 1975). In fact Cultures encourage adherence to norms. Violation of any of these rules by displaying (Ekman, 1972) or recognizing (Matsumoto, 1992) an emotion which does not fit the cultural context, can threaten the degree of harmony and disrupt social interactions. Therefore, regulating emotions via norms, cultures assure that behaviors follow culturally prescribed scripts, enhancing social coordination and diminishing social chaos (Matsumoto & Hwang, 2011).

So far within- and cross-cultural studies on emotions have produced a wealth of information regarding cultural differences and similarities with regard to display rules in the communication of emotions (see Koopmann-Holm & Matsumoto, 2011, and Matsumoto, 1990 for detailed explanations and examples). For instance researchers (Ishii *et al*., 2003; Tanaka *et al*., 2010; Hall, 1976) argue that to express emotions in vocal communication, members of Asian cultures express less explicit verbal information and rely more on contextual cues to interpret the communication and understand the feelings of others. Likewise, Markus and Kitayama (1991) state that collectivist cultures (i.e., Asian cultures) are more context-dependent than individualistic cultures. Furthermore, due to the very restrictive emotional display rules within Asian cultures, the contextual cues provided by the interaction partners in these cultures are very subtle (Tanaka *et al*., 2010; Matsumoto, Kasri, & Kooken, 1999; Matsumoto, 1990). Therefore, to avoid breakdown in interpersonal communication and maintain interpersonal harmony in such cultural contexts great sensitivity is needed for the perception and

interpretation of the implicitly conveyed emotional cues (Safdar *et al.,* 2009; Matsumoto *et al.,* 1999). A number of researchers (Tanaka *et al*., 2010; Schirmer *et al.,* 2006; Ishii *et al*., 2003) suggest that this high sensitivity in the perception of implicit emotional vocal cues may be absent in members of Individualist (i.e., Western cultures). If true, this would suggest that culture can modulate the recognition of emotional speech.

1.7.2 Hofstede's model of cultural dimensions

To study emotions within cultural context, Matsumoto (1989) argues that it is essential to make use of stable and meaningful dimensions of cultural variability proposed by Hofstede (1980). *Cultural dimensions theory* developed by Hofstede (1991, 1983, 1980) is a framework for describing cross-cultural communication. Using a structure emanated from factor analysis, the theory explains the influence of a society's culture on the values of its members, and that how these values relate to their behavior. In this context *dimension* refers to an aspect of a culture that can be measured relative to other cultures. Dimensions of cultural variability include a) Individualism-Collectivism, b) Power Distance, c) Uncertainty Avoidance, d) Masculinity-Femininity, and e) Long Term Orientation-Short Term Normative Orientation, (Hofstede, 2001)[1]. The strength of cultural dimensions theory is that, it provides the researchers with the possibility to compare two cultures in the same areas objectively. As a result, cultural dimensions theory is widely used in various fields (e.g., cross-cultural psychology) as a paradigm for research (Matsumoto, 1990).

Matsumoto (1990) and Matsumoto *et al.* (1998) further note that from among the dimensions of cultural variability, *individualism-collectivism* which is conceived as two poles of one dimension, is a meaningful predictor of cultural variations in emotional norms (i.e., display rules). Similarly, studying cultural differences, several researchers have recognized *individualism-collectivism* as a stable dimension of cultural variability (e.g., Hofstede, 2001, 1980, 1983; Triandis, 1972, 1994). Likewise, Safdar *et al.* (2009) and Greenfield *et al.* (2003) assert that, the Hofstede ranking on individualism-collectivism (Hofstede,

[1] Hofstede (2001) explains the five cultural dimensions as follows: a) Individualism-Collectivism (IDV): stands for the degree the society reinforces individual or collective achievement and interpersonal relationships." b) Power Distance (PDI): stands for the degree of equality, or inequality, between people in the country's society." c) Masculinity-Feminity (MAS): stands for the degree the society reinforces, or does not reinforce, the traditional masculine work role model of male achievement, control, and power." d) Long Term Orientation (LTO): stands for the degree the society embraces, or does not embrace long-term devotion to traditional, forward thinking" and e) Uncertainty Avoidance (UAI): stands for the level of tolerance for uncertainty and ambiguity within the society - i.e. unstructured situations."

2001) represents extensive distillations of various socio-psychological phenomena (encompassing social beliefs, values, norms, personality and behavioral characteristics) that pertain to ways in which human-environment relations are assessed.

1.7.3 Individualism

Individualism is characterized by independent relationships where the priority is on the self (Triandis, 1994). In such cultures the concept of self is concerned with independence, where the independent individual is believed to contain unique internal features and is expected to behave in accordance with these features (Markus & Kitayama, 1994). Individualists tend to emphasize and enhance their personal goals, interests and values over the society they belong to (Markus & Kitayama, 1991; Triandis, 1995; Oyserman, Coon, & Kemmelmeier, 2002). Self-enhancement, the need for individual autonomy, detachment from others (Scherer & Brosch, 2009; Andersen, Reznik, & Chen 1997) and functioning based on personal choices (Walsh & Banaji, 1997) are main features of Individualistic cultures. Considering these features, individualist cultures regard emotions as essential personal experiences whose expression is any individual's right. In these cultural contexts, individuals tend to consider emotions as inner states vented spontaneously (Matsumoto, *et al.,* 1998). Individualistic cultures reinforce outward displays of emotion[1] that exaggerate the strength of the feeling (Matsumoto, *et al.,* 1998). Similarly, Mesquita (2001) claims that in individualist cultures emotions are assumed to amplify and underline a subjective self.

1.7.4 Collectivism

In contrast to individualism, collectivism is featured by interdependent relationships which focus on social cohesion and a group[2]'s harmony (Triandis, 1994). In a collectivist culture the concept of self is bound to surrounding social context in which the self cannot be split from others and the goal is to keep oneself tied to others (Markus & Kitayama, 1994). Collectivists identify themselves as members of a group they belong to, therefore they emphasize and enhance group goals, interests and values over those of each

[1] Although collectivism and individualism characterize cultural groups, not every individual in a given cultural context engages in the same ideas and practices, nor does he/she engage in them in exactly similar manners (Markus, Mullally, & Kitayama, 1997). However, across individuals within a collectivist or individualist context, emotions are patterned in discernible ways (Mesquita, 2001).

[2] For instance: family, tribe, and nation.

individual member of the group (Markus & Kitayama, 1991; Triandis, 1995; Oyserman *et al.,* 2002). Self-transcendence, the need for harmony among the in-group (Scherer & Brosch, 2009; Oyserman *et al.,* 2002; Markus & Kitayama, 1991), attachment to others (Andersen *et al.,* 1997) and functioning based on group priorities (Walsh & Banaji 1997) are main features of collectivist cultures. In these cultures emotions are regarded as interactive experiences which mirror the social context rather than the individual's inner self. As Mesquita (2001) argues, in such cultural contexts expression of emotion is controlled, due to the point that it is grounded in evaluation of the relationship between the self and the others. In other words, in collectivist cultures emotions are expected to emphasize and reproduce the self in relation to others and the world (Mesquita, 2001).This issue implies that emotions are conceived as situation-specific clues about the relationship between the individuals. By and large, collectivist cultures reinforce the perpetuation of cohesion within the group and accordingly, expression of emotions gains high priority (Potter, 1988).

1.8 The interplay of culture and gender

Considering that gender refers to biologically rooted, socio-culturally constructed roles that a given society considers appropriate for men and women (Pardue & Wizemann, 2001), and that socio-cultural variables mediate and moderate gender differences in emotional functioning (Brody & Hall, 2008), it is clearly revealed that culture plays an important role in shaping the concept of gender. Gender role is a cultural orientation conditioned by a traditional social system in which men are supposed to behave as men i.e., masculine, and women are expected to behave and think as women i.e., feminine (Mungai & Ogot, 2012). Gender identity is developed through socialization where children at early ages are taught appropriate gender roles and what it means to be a male or a female (Mungai & Ogot, 2012). The concept of gender role is a critical aspect of culture because it shapes the way daily life is lived not only in the family, but also in the wider community (Schalkwyk, 2000). The concept of *masculinity and femininity* refers to gender roles in a society (Stets & Burke, 2000). Many cultures[1] were and still are of masculine nature. In such cultures men are privileged in terms of social position, career, income and authority (Alvesson, 2002). In the context of these cultures, women are seen as the more emotional gender experiencing and expressing emotions in general more than do

[1] In cross-cultural research, culture is typically depicted by country (Matsumoto *et al*., 1998).

men (Ember & Ember, 2004). In cultures with masculine nature, very few women can get higher-level and better-paying jobs (Mungai & Ogot, 2012).

Considering that women and girls are very often discriminated against in health, education, labor market, and the like (Hausmann, Tyson, & Zahidi, 2006), United Nations set-out to create a measurement index for measuring the extent of gender inequality across the globe's countries (Klasen, 2006). In doing so, the Gender Empowerment Measure (GEM), was introduced and meant to measure "whether women and men are able to actively participate in economic and political life and take part in decision-making" (UNDP[1], 1995, p. 73). Using three basic indicators[2], GEM aimed at gaining international recognition for the importance of eliminating gender inequality (Klasen, 2006). The GEM was thought to be a valuable instrument because it allowed certain dimensions that were previously difficult to compare between countries to come into international comparison. However, due to conceptual and methodological limitations, GEM faced much criticism[3] (Permanyer, 2011; Bardhan & Klasen, 1999). In an attempt to improve the GEM, the UNDP introduced the Gender Inequality Index (GII) in 2010 (Ferrant, 2010). The new index is a composite measure which, as reported by the UNDP, captures the loss of achievement due to gender inequality via using three important aspects of human development (i.e., reproductive health, empowerment, and labor market participation)[4]. These dimensions are captured in one synthetic index, as to account for joint significance (Klugman, Kovacevic, Twigg, & Zambrano, 2010). This new index does not include income levels as a component, which was reported to be one of the most controversial factors of the GEM. The GII also does not allow for high achievements in one dimension to compensate for low achievement in another (Ferrant, 2010). The GII sheds new light on the position of women in over 150 countries and cultural contexts. This index yields insights in gender gaps in major areas of human development (Permanyer, 2013). The GII is interpreted as a percentage and indicates the percentage of potential human development lost due to gender inequality. Based on the latest[5] Human Development Report[6] the GII shows that from

[1] United Nations Development Program

[2] 1) Proportion of seats held by women in national parliaments, 2) percentage of women in economic decision making positions (incl. administrative, managerial, professional and technical occupations), and 3) female share of income (earned incomes of males vs. females).

[3] Please see Klasen and Schüler (2011) for an overview of these criticisms.

[4] According to (Klugman, Kovacevic, Twigg, & Zambrano, 2010) reproductive health is measured by maternal mortality ratio and adolescent birth rates. Empowerment is calculated by proportion of parliamentary seats occupied by females and proportion of adult females and males aged 25 years and older with at least some secondary education. Economic status expressed as labor market participation is measured by labor force participation rate of female and male populations aged 15 years and older.

[5] For the year 2013

[6] Available at: http://hdr.undp.org/en/content/table-4-gender-inequality-index

among the 148 countries, Slovenia (in Europe) is ranked 1st (the lowest gender inequality level), and Mali (in Africa) is ranked 148 (the highest gender inequality level). In the same year the GII ranking for Germany is 3, while Iran is ranked 109. According to this GII ranking, women and girls living in Iran are far more discriminated compared to their German counterparts[1].

1.8.1 Gender discrimination and coping strategies

Gender discrimination which can arise in different contexts such as domestic, educational, occupational, political, and the like, act as stressors and have several effects and consequences such as psychiatric and physical problems (Feder & Brougher, 2013; Kreiger, 2000). Research (Piko, 2001; Ro & Choi; Folkman & Lazarus, 1980) has documented that as individuals are encountered with stressors, they appraise the potential threat and therefore employ *coping strategies*[2] to mitigate the emotional and physiological imbalances caused by the stressors. In other words, coping strategies help individuals to deal more efficiently with stressors and persistent problems and challenges, and can eventually improve the quality of their lives (Braun-Lewensohn, Sagy, & Roth, 2011). Researchers mention that while encountering stressors, individuals may employ two types of coping strategies: a) *emotion focused* which involves trying to reduce the negative emotional responses associated with stress (e.g., embarrassment, fear, anxiety, depression, excitement and frustration), and b) *problem focused* which targets the causes of stress in practical ways such that tackles the problem or stressful situation that is causing stress, as a consequence directly reducing the stress (Piko, 2001; Lazarus & Folkman, 1984). In general, psychologists (e.g., Thoits, 2013; Seiffge-Krenke, 2013; Webb, Miles, & Sheeran, 2012; Piko, 2001) propose that, while both coping strategies can be useful in reducing the stress, problem-focused coping strategy is regarded as a rather more effective solution, as it removes the stressor by dealing with the root cause of the problem, providing a long term solution.

1.8.2 Gender discrimination in Iran

While we aim to investigate gender differences in the recognition of vocal emotions in Iranian male and females, it is essential to have an overview of gender discrimination in this society. As

[1] The complete list of the latest GII is available at: http://hdr.undp.org/en/content/table-4-gender-inequality-index

[2] *Coping* is a very complex process, which varies according to various variables (e.g., the situation, the evaluation of the situation, and the resources available). In this context *coping strategies* refer to the thoughts and actions we use to deal with a threatening situation (Thoits, 2013).

indicated by the GII, women and girls living in Iran are way more discriminated compared to females living in many other countries (see the latest GII report for exact details). Researchers believe that gender discrimination in Iran has a long history and is not regarded as a new phenomenon (Moghadam, 2010; Mehran, 2003; Najmabadi, 1998). In Iran, where the culture is of masculine nature (Kar, 2002), women are usually under pressure due to gender inequality in relation to different individuals (i.e., their husband, male members of the extended family, and male members of the society[1]) where their status is perceived as being lower than their male counterparts (Keyvanara & Haghshenas, 2010; Mehran, 2009; Najmabadi, 2005). In the same way, the expectations and responsibilities assumed to women by the cultural norms of the country leaves women with less control over their own lives compared with men, and limits their choice (Keyvanara & Haghshenas, 2010; Moghadam, 2010).

Social studies have revealed that gender discrimination in Iran has led to negative effects and consequences such as psychiatric and physical problems for Iranian women (see Bagherian, 2007 and Kar, 2002 for detailed explanations). However, negative consequences are not the only outcome of the discriminatory situation (Moghissi, 2008). Despite of the high frequency of these stressors, Iranian women's aspirations for less social limitation, greater freedom, higher education, better work opportunities, and generally less gender inequality, have led them to employ effective coping strategies (Mehran, 2009; Najmabadi, 2005, 1998). Moghissi (2008) and Kar (2002) point out that, gender discrimination in Iran has to some extent resulted in vigilance in women, and a tendency to be willing to take risks. Investigating various aspects of gender inequality in Iran, Moghissi (2008), Kar (2002, 1996) and Roshanfar (2003) mention that by using active ways (i.e., problem focused strategies) to tackle the situation that causes the stress and pressure due to gender inequality, many Iranian women and girls have effectively managed to deal with the gender inequality to some extent. Referring to the high number of Iranian women in academia, Mozaffarian and Jamali (2008) point out that despite the masculine culture, a large number of Iranian women have successfully managed to deal with cultural barriers. In the same way, reporting some actual examples of Iranian women (from various cities and different social classes) who to some extent overcome gender discriminative situations, Moghadam (2010) and Sullivan (2010) claim that developing effective *interpersonal skills*[2]

[1] Kar (2002) believes that gender discrimination in Iran may also appear as female against female.

[2] Interpersonal skills are regarded as life skills that we use to communicate and interact with others individually and in groups and include: a) Verbal Communication, b) Non-Verbal Communication, c) Listening Skills, d) Negotiation, e) Problem Solving, f) Decision Making, and g) Assertiveness (Koprowska, 2014).

can be counted as the key to success in these Iranian women. Likewise, Kar (2002) has pointed to effective Interpersonal skills as a probable peaceful solution of confronting gender discrimination in Iran. Pointing to a number of empirical studies (Mokhtari, 2013; Rezaienia, 2012), state that interpersonal skills and gender are significantly correlated among Iranian college students. They further argue that Iranian female college students reveal significantly higher interpersonal skills compared to their male counterparts. Although these studies (Mokhtari, 2013; Rezaienia, 2012) reveal higher interpersonal skills among Iranian female college students, to gain a more comprehensive view of Iranian women and girls coping strategies against gender discrimination, more empirical studies on Iranian women from various social classes and age groups are to be conducted.

1.9 Overview of the present dissertation

This research is interdisciplinary in character, incorporating interests, theories, and methods from the disciplines of linguistics, psychology and anthropology. This dissertation serves three purposes. First, we will generate the first comprehensive emotional speech database for Persian. To date, numerous databases of vocal expressions of the basic emotions have been established in several languages including English (Cowie & Cornelius, 2003; Petrushin, 1999), German (Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005), Chinese (Liu & Pell, 2012; Yu, Chang, Xu, & Shum, 2001), Japanese (Niimi, Kasamatsu, Nishinoto, & Araki, 2001), Russian (Makarova & Petrushin; 2002), as well as many other languages (for reviews, see Douglas-Cowie, Campbell, Cowie, & Roach, 2003; Juslin & Laukka, 2003; Ververidis & Kotropoulos, 2003). However, this has not been achieved in Persian. Therefore, this database will provide a useful language recourse for conducting basic research on a range of vocal emotions in Persian, as well as for conducting cross-cultural/cross-linguistic studies of vocal emotion communication. Moreover, the stimuli in this database can be used as a reliable language recourse (lexical and vocal) for the assessment and rehabilitation of communication skills in patients with brain injuries.

Despite the large amount of research on the recognition of vocal emotions in stress-timed languages (e.g., German, and English), no such study has been performed on a syllable-timed language (e.g., Persian, Turkish). Pell et al. (2009) claim that, vocal expressions of emotion manifested in emotional prosody, are inherently dynamic and integrated with linguistic properties of the language under study. Respectively, Pell (2001) mentions that, linguistic-specific features of a language (e.g., rhythm) affect vocal expressions of emotion (i.e., emotional prosody). Therefore, there is a gap in research on gender effects

on the recognition of emotional prosody in languages with a different rhythmic division of time. As a result, the second objective of this dissertation is to address this paucity of research by making the first attempt to investigate the potential role of gender in the recognition of emotional prosody in a syllable-timed language, namely Persian as spoken in Iran.

So far, the only existing study on gender and emotional speech recognition was carried out on German speakers living Leipzig, Germany (Paulmann *et al.,* 2008). According to the Hofstede Model of Cultural Dimensions[1] (Hofstede, 2001), Germany with a relatively high score of 67, ranks 15 among the 53 countries[2] and is considered as an individualistic society (see Hofstede, 2001 & The Hofstede Center, n.d. for more details). Considering that 1) emotional processes are largely influenced by culture (Markus & Kitayama, 1991), 2) socio-cultural variables mediate and moderate gender differences in emotional functioning (Brody & Hall, 2008), and that 3) due to the interplay of language and emotion in vocal communication, socio-cultural influences on emotion recognition are specifically conspicuous in the emotional speech (Pell *et al.,* 2009), there is a gap in the literature of gender effects in the recognition of emotional speech in speakers of different linguistic and cultural background. Therefore to fulfill the third objective of the present dissertation, I will seek to address the aforementioned gap by making the first attempt to investigate the potential role of gender in the recognition of emotional speech in a different and under investigated cultural context (i.e., Persian speakers living in a collectivist society).

The present dissertation consists of three separate studies presented in separate chapters. Chapter 2 pertains to the establishment and validation of an emotional speech database for Persian. Chapter 3 explores gender effects on the recognition of emotional prosody (i.e., neutral linguistic content intoned in emotional voice) in Persian as an instance of a syllable-timed language, while Chapter 4 investigates the effects of gender on the recognition of emotional speech (i.e., emotional linguistic content intoned in congruent emotional voice) in a collectivist culture. Chapter 5 finally concludes this thesis with a general discussion as well as implications and suggestions for future research. In what follows, I will present the particular aims and methods for each study in more details.

[1] With regard to analyzing a country's culture, the Hofstede Model of Cultural Dimensions is a widely accepted model (Matsumoto, Yoo & Fontaine, 2008).

[2] These data on culture dimension came from Hofstede's (2001) longitudinal study of cultural dimensions from 50 countries and three regions.

Study One: Recognizing emotional speech in Persian: A validated database of Persian emotional speech (Persian ESD)

This study reports the process of designing, compiling, and evaluating a comprehensive emotional speech database for colloquial Persian.

According to Pell (2001), emotional prosody can be affected by the linguistic features of a language. Besides, as Liu and Pell (2012) have claimed, it is essential for researchers to generate valid emotional stimuli that is suitable for the linguistic background of the participants of a study. Considering this issue, recording samples of all the basic emotions are required for a comprehensive study on emotional speech for the language under study. Therefore, by generating a robust set of validated stimuli, I aimed to fill this gap for Persian language. Such a stimuli can minimize the influence of individual bias, and can avoid subjectivity in stimulus selection in future studies of Persian emotional speech.

To separate the impact of lexical content from prosody on the processing of emotional speech, a set of validated sentences is required (Ben-David *et al*., 2011). Therefore, three experiments were conducted to generate a set of validated sentences (lexical material) in colloquial Persian. To make sure that each generated sentence only conveys one specific emotion or no emotion at all (neutral), I performed two perceptual studies to validate the 252 sentences. As the next step, the validated sentences were articulated by two Persian speaking, semi-professional actors (one male, one female) in three conditions: (1) congruent (emotional lexical content articulated in a congruent emotional voice), (2) incongruent (neutral sentences articulated in an emotional voice), and (3) baseline (all emotional and neutral sentences articulated in neutral voice).

The validity of the database (i.e., 472 spoken utterances) was evaluated by a group of 34 native speakers in a perception test. Utterances recognized better than five times chance performance (71.4 %) were regarded as valid portrayals of the target emotions. Acoustic analyses were performed to determine whether the vocal portrayals would show obvious differences in acoustic parameters that might help participants to distinguish the intended emotions correctly. The analyses were based on three critical parameters (i.e., mean pitch, mean intensity, and duration) that have been reported to differentiate well among vocal emotion categories and perceptual terms (Juslin & Laukka, 2003). A total of 468 vocal utterances (all of the validated emotional and neutral portrayals), encompassing the congruent, incongruent, and baseline conditions, were included in this analysis. These vocal utterances were analyzed using the Praat speech analysis software (Boersma & Weenink, 2006). Results of the acoustic analysis revealed differences in pitch, intensity, and duration.Afterwards, a discriminant function analysis for all

three conditions (i.e., congruent, incongruent, and baseline) was performed to determine how well the six categories (five emotions plus neutral) could be classified on the basis of the intended acoustic measures (the mean pitch, mean intensity, and duration). Analyzing the results of the discriminant function analysis revealed expected variations among the five emotion categories.

The database of Persian Emotional Speech (Persian ESD) is designed to be used as a reliable material source (for both text and speech) in future cross-cultural or cross-linguistic studies of emotional speech, and it is available for academic research purposes free of charge.

Study Two: Gender effects on the recognition of emotional prosody: Evidence from Persian Language (Farsi)

Vocal emotional expressions, which carry a rich source of information about a speaker's emotions and his/her social intentions, are influenced by individual differences (e.g. gender, age). On one hand, among these individual difference factors the role of gender is not sufficiently studied and is still under debate. On the other hand, the existing literature on gender and emotional prosody has targeted speakers of two specific languages, i.e., German and English, representing stress-timed languages. However, since prosody is affected by language-specific features such as the rhythmic division of time (i.e., stress- vs. syllable-timed), the applicability of the existing results to languages with a different rhythmic division of time is under debate. Therefore in the second study, we aimed at investigating gender effects on the recognition of emotional prosody in an under-investigated linguistic context (Persian, a syllable-timed language). On the basis of evidence from the existing literature[1], we hypothesized that gender would affect the recognition of emotional prosody in Persian, and in particular that there would be a female advantage in the recognition of emotional prosody.

In a behavioral experiment, 66 native speakers of Persian were asked to recognize the emotional prosody of a set of validated vocal portrayals. These vocal portrayals were a set of sentences with emotionally neutral lexical content, intoned in one of the five intended emotions (anger, disgust, fear, happiness, and sadness) by a male and a female actor. Participants were instructed to listen to the utterances, to focus on the voice of the speaker and to identify the emotional prosody of each utterance based on a forced choice response format. Obtained data were analyzed via nominal logistic regression.

[1] Detailed explanation of the related literature is provided in chapter three, part 3.1.3 Gender and emotion recognition.

Taken together, findings of this study revealed that recognition accuracy for the emotional prosody of vocal portrayals differed significantly as a function of gender of decoders (i.e., participants). Specifically, there was an overall recognition advantage for female participants over male participants in our task. One reason for female advantage might be that females are more socially oriented than males and are more likely to define themselves in relational terms. Additionally, biological (i.e., difference in specific hormones such as estrogen and oxytocin) as well as social factors (i.e., socio-culturally constructed roles that a given society considers appropriate for men and women) can be the probable reasons behind female advantage in the recognition of vocal emotions. Furthermore, our results with Persian speakers were in line with the results of German and English speakers. Since Persian is a syllable-timed language, our findings suggest that the nature of the rhythmic division of time in a language (i.e., stress- vs. syllable-timed) does not influence the decoding of emotional prosody.

Study Three: Gender effects on the recognition of emotional speech: Evidence from a collectivist culture

It is generally acknowledged that the ability to recognize and express emotions is an initial human potential (Matsumoto, 1989); cross-cultural endeavors concerning recognition of facial and vocal expression of emotions have strengthened theories of universality of human emotional capability specifically concerning the so-called *basic emotions* such as anger, fear, happiness and sadness (e.g., Pell, Monetta, Paulmann, & Kotz, 2009; Pell & Skorup, 2008; Thompson & Balkwill, 2006; Scherer & Wallbott, 1994; Ekman, 1973; Ekman, Friesen, & Ellsworth, 1972). Nevertheless, from the view point of cross-cultural psychology and anthropology, the biological perspective is not adequate for a comprehensive understanding of the perception and expression of emotions (Eid & Diener, 2001). As emotion and culture are intermingled with each other (Mesquita, 2001) emotional processes are largely influenced by culture (Matsumoto, 2001; Markus & Kitayama, 1991). Accordingly, for a full understanding of emotional processes we should consider the cultural context where emotions are perceived, expressed, and experienced (e.g., Mesquita & Ellsworth, 2001; Scherer, 1997; Kitayama & Markus, 1995; Scherer & Wallbott, 1994, Markus & Kitayama, 1991).

So far cross-cultural studies on emotions have produced a wealth of information regarding cultural differences and similarities in the communication of emotions. However, there still remains a number of gaps in our knowledge. For instance, the only existing study on gender and emotional speech recognition was carried out on German speakers living Leipzig, Germany (Paulmann *et al.,* 2008).

According to the Hofstede Model of Cultural Dimensions (Hofstede, 2001), Germany is considered as an individualistic society (see Hofstede, 2001 & The Hofstede Center, n.d. for more details). Therefore, as culture plays a significant role in the recognition of vocal emotions, results of the existing study might not apply to members of other cultural groups (i.e., collectivist societies). Therefore in the third study, we addressed this gap for the first time, and aimed to examine the role of gender in the recognition of emotional speech in a different cultural context (i.e., a collectivist society). According to the state of the art[1] we hypothesized that 1) gender affects the recognition of emotional speech, and in particular there will be a female advantage in the recognition of emotional prosody and, 2) compared to the members of an individualist society, members of a collectivist culture show higher sensitivity to vocal emotional cues.

Seventy native speakers of Persian (35 females) were instructed to listen to a set of spoken utterances, to focus on the voice of the speaker and to identify the emotional prosody of each vocal utterance based on a six-choice response format corresponding to anger, disgust, fear, happiness, sadness, and neutral. The stimulus material selected for this experiment consisted of 72 vocal utterances (i.e., 12 sentences in each of the five intended emotional category plus 12 sentences in the neutral category). These vocal utterances were generated based on a syntactically similar lexical content (i.e., subject + object + prepositional phrase + verb) intoned by a male and a female native speaker of Persian. These vocal utterances were previously matched based on the emotional intensity of their lexical content.

Obtained data were analyzed via nominal logistic regression. Findings of the study revealed a significant effect of gender in favor of females in the recognition of emotional speech. One reason females might be more sensitive to emotional speech than males, is that females are more socially oriented than males and are more likely to define themselves in relational terms (Kashima *et al.*, 2004). Besides this, in the context of social accounts of gender differences, these findings may be regarded as evidence of the point that gender differences in social orientation are biologically rooted (Kashima et al., 1995), for instance by hormones such as estrogen and oxytocin (Babcock & Laschever, 2009; Wood & Eagly, 2002; Taylor *et al*., 2000). In line with previous studies with regard to the great sensitivity of the members of collectivist cultures in perception of the emotional cues, findings of the current study revealed very high overall accuracy rates 95% (i.e., almost close to six times chance performance) by members of a collectivist culture. In a previous emotion recognition study performed on members of an individualist culture, an overall accuracy rate of 70% (i.e., almost close to five times chance performance) was

[1] For detailed explanation of the related literature please see chapter four of this dissertation.

reported. A content based comparison between the findings of these two studies indicated that members of a collectivist culture show higher sensitivity to vocal emotional cues. Research has documented that due to the very restrictive emotional display rules within collectivist cultures, the contextual cues provided by the interaction partners in these cultures are very subtle. As a result, to avoid breakdown in interpersonal communication and maintain interpersonal harmony in such cultural contexts great sensitivity is needed for in the perception and interpretation of the implicitly conveyed emotional cues. Accordingly, restrictive emotional display rules may be the cause of the high sensitivity to emotional vocal cues. In summary, our findings support the assumption that cultural background influences perception of emotional speech. It is important to take this point into consideration in scenarios where culture is an important factor in human-human or human-computer communication.

Chapter 2: STUDY ONE

Recognizing emotional speech in Persian: A validated database of Persian emotional speech (Persian ESD)

due to copyright reasons, this paper is not included in the online version.

Keshtiari, N., Kuhlmann, M., Eslami, M., & Klann-Delius, G. (2015). Recognizing emotional speech in Persian: A validated database of Persian emotional speech (Persian ESD). Behavior research methods, 47(1), 275-294. http://doi: 10.3758/s13428-014-0467-x.

Chapter 3: STUDY TWO

Gender effects on the recognition of emotional prosody: Evidence from Persian Language (Farsi)

Keshtiari, N., Kuhlmann, M., & Klann-Delius, G. (Submitted). Gender effects on the recognition of emotional prosody: Evidence from Persian Language (Farsi).

Abstract

Proper understanding and processing of a speaker's emotional state is critical to effective social interaction. Vocal emotional expressions which carry a rich source of information about a speaker's emotions and his/her social intentions are influenced by individual differences (e.g. gender, age); among those, the role of gender is not sufficiently studied and is still under debate. The existing literature on gender and emotional prosody has targeted speakers of two specific languages, i.e., German and English, representing stress-timed languages. However, since prosody is affected by language-specific features such as the rhythmic division of time (i.e., stressed- vs. syllable-timed), the authors questioned the applicability of the results of the existing studies to languages with a different rhythmic division of time. Therefore, the present study examined the role of gender in the recognition of emotional prosody in native speakers of Persian, a syllable-timed language.

In a behavioral experiment, 66 native speakers of Persian were asked to recognize the emotional prosody of a set of validated vocal portrayals. These vocal portrayals were a set of sentences with emotionally neutral lexical content, intoned in one of 5 intended emotions (anger, disgust, fear, happiness, and sadness) by a male and a female actor.

In general, our results established that recognition of emotional prosody significantly differed as a function of gender, that is, female participants outperformed their male counterparts in recognizing all the intended emotional categories. Additionally, our results suggest that rhythmic division of time (i.e., stressed- vs. syllable-timed) does not influence the emotion recognition performance.

Keywords Emotion, prosody, gender, recognition, Persian

3.1 Introduction

Communication of emotions is an essential aspect of human social life (Ekman, 1992). In fact, to avoid breakdowns in social communication, humans must encode and decode vocal emotional expressions efficiently (Lima, Alves, Scott & Castro, 2014). As Surcinelli, Codispoti, Montebarocci, Rossi, and Baldaro (2006) argue difficulties in non-verbal communication of emotions can result in problematic relationships and may finally lead to the development of psychopathology. Therefore, it is of considerable social relevance to determine which factors affect the encoding and decoding of emotions in interpersonal communication. As a contribution to this literature, the current study investigated the recognition of the five intended emotions (i.e., anger, disgust, fear, happiness, and sadness) from vocal expression as a function of gender. Within a discrete emotion framework, anger, disgust, fear, happiness and sadness are frequently regarded as *"Basic emotions"*. Having a distinct biological basis, basic emotions are a small number of innate, discrete, and universal emotional categories shared across languages (Ekman, 1999) and cultures (Sauter, Eisner, Ekman, & Scott, 2010; Ekman, 1999). All other emotional categories may be derived from the basic emotion (e.g., Ekman, 1992, 1999; Izard, 1992). Each of the basic emotions can be defined in terms of an appraisal of goal-relevant events that have recurred during evolution (see Power & Dalgleish, 2007). Oatley (1992) gives instances of such appraisals: happiness (sub goals being achieved), anger (active plan frustrated), sadness (failure of major plan or loss of active goal), fear (self-preservation goal threatened or goal conflict), and disgust (gustatory goal violated). Additionally, as mentioned by Banse and Scherer (1996) these set of emotions have specific prosodic profiles.

Vocal communication and Emotional prosody

Vocal expression is the modality that is often regarded as an effective means of emotional communication (Scherer, 1986). In this modality, emotions can be conveyed through three channels: (1) lexical content (the emotional meaning of the words and sentences), (2) emotional prosody[1] (the emotional melody of speech) (Szymanowski, Kotz, Schröder, Rotte, & Dengler, 2007; Lima & Castro, 2011), and (3) purely nonverbal expressions (e.g., laughter, screams, and sighs) (Lima, Castro, & Scott, 2013). Emotional prosody is regarded as one of the most basic features of language (Besson, Magne & Schön,

[1] The term prosody was first introduced by Monrad-Kohen (1947, p.23), who defined it as "that faculty of speech which conveys different shades of meaning through of variations in stress and pitch-irrespective of the word and grammatical construction".

2002). It is defined as the ability to express emotions through variations of speech parameters, comprising intensity, pitch, tempo, rhythm, and timbre (Schirmer & Kotz, 2006). It encompasses non-verbal aspects of human language and carries a rich source of information about a speaker's emotions and social intentions (Wilson & Wharton, 2006; Banse & Scherer, 1996; Scherer, 1986). In fact, emotional prosody plays a crucial role in various aspects of human social communication (Plutchik, 2003). Arndt and Janney (1991:521) suggest that "the ability to encode and decode vocal emotional cues is a prerequisite for negotiating claims to power, respect, or equality, defining degrees of intimacy, showing affiliation or non-affiliation, avoiding threat, repairing interpersonal misunderstandings and so forth". Emotional prosody is, therefore, a crucial element of verbal communication, and an integral part of social interaction.

Gender and emotion recognition

Gender as a biologically rooted social mechanism (Udry, 1994, Nobelius, 2004, Haig, 2004) is one of various individual difference factors which have been found to affect cognitive performance (Wood & Eagly, 2002; Herlitz & Lovén, 2009). To date considerable amount of research on Gender differences in *emotion recognition* has been conducted. The findings of these studies have been summarized in four different meta-analyses by Hall (1978, 1984), McClure (2000) and Thompson and Voyer (2014). Hall's (1978) review was the first to summarize gender differences in non-verbal (audio, visual, audio-visual) displays of emotions. She summarized the results from 75 studies with adults and children. Her results showed that females consistently outperform males on the recognition of vocal and facial emotional expressions. In a second independent meta-analysis in 1984, Hall reported similar results suggesting that females significantly outperformed males. In her second review (Hall, 1984) she reported that female advantage was more conspicuous in the more recently published experiments. She argued that more accurate instruments and more powerful statistical analyses can be the likely reason. In a third meta-analysis McClure (2000) investigated gender effects on the recognition of facial expressions in three age groups (infants, children and adolescents). She examined the findings of 60 adolescent and child studies and 6 infant studies. Overall, her findings indicated that there was a gender effect in favor of females in recognizing non-verbal emotional displays. Recently in a multilevel meta-analyses of 215 studies, Thompson and Voyer (2014) examined the existence of gender differences in the recognition of non-verbal (audio, visual, audio-visual) displays of emotion. They reported a small overall advantage in favor of females on emotion recognition tasks. However, they claimed that the magnitude of the gender difference was moderated by various factors including specific emotion category, emotion type (positive,

negative), sensory modality (audio, visual, audio-visual), gender of the actor, and age of the participants. Still they mention that the potential effect of the sensory modality on the recognition of emotions is still an open question and more studies are required on gender effects on the recognition of *vocal emotions*[1] (Thompson & Voyer, 2014).

These results are in accordance with proposals of gender differences with regard to social orientation (Broverman et al., 1972). Based on these proposals, a) females define themselves in relational terms more than do males (Kashima et al., 2004; Dollinger et al., 1996; Clancy & Dollinger, 1993), and b) females are more sensitive to nonverbal cues (Schirmer, Striano & Friederici, 2005; Briton & Hall, 1995; Robinson & Kinnier, 1985).

Although the findings of the above mentioned meta-analyses indicated overall advantage in favor of females, still studies investigating only the recognition of vocal emotions have reported conflicting results. For instance Bonebright and colleagues (1996) asked a group of male and female participants to listen to vocal emotional portrayals of three male and three female actors articulating stories with non-emotional lexical content, and to identify the four basic emotions of anger, fear, happiness, sadness. The conclusion they reached was that female participants out-performed their male counterparts in the recognition of vocal portrayals of fear, happiness, and sadness but not anger. Schirmer, Kotz and Friederici (2002) in an event-related potential (ERP) experiment found gender differences in the sensitivity to emotional prosody. They reported that during word processing, women integrated words into emotional prosody at an earlier point in time. In another study Schirmer and Kotz (2003) examined the influence of emotional prosody and word valance on both genders in an emotion judgment task. They reported that only females combined emotional prosody and lexical content when emotional prosody was task-irrelevant. In addition to the investigations performed on healthy individuals, recent evidence from bipolar patients (Van Rheenen & Rossell, 2013) revealed the presence of a female advantage when it comes to the processing of emotional prosody.

In contrast to the above mentioned studies reporting female advantage, there exists a number of studies reporting no difference between the two genders in the recognition of vocal emotions (e.g., Paulmann et al., 2008; Fecteau, Armony, Joanette, & Belin, 2005; Orbelo, Grim, Talbott, & Ross, 2005; Raithel & Hielscher-Fastabend, 2004) as well as nonlinguistic vocalizations of emotions (e.g., Hawk, Van Kleef, Fischer, & Van der Schalk, 2009; Sauter, Panattoni, & Happé, 2013; Lima et al., 2014). In fact, studies examining gender effects on the recognition of vocal emotions have often revealed inconsistent results

[1] In the present study we refer to "Vocal emotions" as neutral yet meaningful lexical content intoned in emotional prosody. It should not be mistaken with "nonlinguistic emotional vocalizations" such as laughter, sobs, or sighs.

(Brody & Hall, 2008; Lima et al., 2013). Besides the methodological procedures (i e., varying methodologies, different task demands, and measurements) another potential explanation of these inconsistencies could be difference in linguistic background of the participants (Pell, 2001). Pell (2001) believes that emotional prosody is embedded in language and it is affected by linguistic features of a specific language. Therefore, specific linguistic features of the languages under investigation may also be the cause of the inconsistent results.

Language, prosody and rhythmic division of time

Linguistic "rhythm" defined as the repetition of patterned sequences of elements, which often vary in prominence (Fraisse, 1982), is an important aspect of prosody (Wells, 2006). It pertains to the way languages are organized in time (Patel & Daniele, 2003)[1]. The postulated classification of stress-timed rhythm and syllable-timed rhythm (Abercrombie 1967) is the foundation of phonetic research on rhythm (Abercrombie, 1967; Lehiste, 1977; Bertinetto, 1989). While this distinction is not totally uncontroversial (see Jassem & Witten, 1984; Vihman, Nakai & De Paolis, 2006, for details), it is regularly used to make a prosody-wise division between languages, which are said to be (primarily) stress-timed, or (primarily) syllable-timed (Roach 1982).

According to Windfuhr (1979) there is a is in stark contrast[2] between stressed-timed languages where the amount of time to say a sentence depends on the number of stressed syllables and syllable-timed language where the amount of time to say a sentence depends on the total number of the syllables. (see Patel & Daniele, 2003 for detailed description).This distinction in the rhythmic division of time (i.e., stressed- vs. syllable-timed) is important for our study because emotional prosody is affected by the linguistic features of the specific language under study (Pell, 2001).

So far the role of gender on the recognition of vocal emotions has been explored with native speakers of English (e.g., Bonebright et al., 1996[3]) and German (Paulmann et al., 2008). In these two separate studies participants were asked to recognize the emotional tone of a set of meaningful sentences uttered in their mother tongue (i.e., English/German). Besides these two studies, in the meta-

[1] As Schluter (2005) suggests, the distinction between the different rhythmic types is based on the fact that all human speech is split into temporal intervals that have a tendency to be of similar length.

[2] One such contrast is durational variability in consonant sequences (Ramus, Nespor & Mehler, 1999).

[3] Participants of this study were all native speakers of North American English studying at the University of Nebraska—Lincoln.

analyses performed by Hall (1978) and Thompson and Voyer (2014) English and German are the only languages under investigation. English and German are widely acknowledged by linguists to have similar rhythmic divisions, representing "stress-timed" languages (Windfuhr, 1979). It is to be considered that there can be varying levels of stress-timing within the different standards of a language. British and North American English are considered as prototypical members of the stress-timed languages (Roach 1982; Daur, 1983).

The present study

Despite the large amount of research on the recognition of vocal and facial emotions, no such study was performed on a syllable-timed language. This reveals a gap in research in gender effects on the recognition of emotional prosody in languages with a different rhythmic division of time. The present study, therefore, seeks to address this paucity of research by making the first attempt to investigate the potential role of gender on the recognition of emotional prosody in a syllable-timed language namely Persian as spoken in Iran.

Persian (also known as Farsi) is an Indo-Iranian language, a sub-branch of Indo-European family (Anvari & Givi, 1996) spoken by almost 110 million people around the world, while holding official status in Iran, Tajikistan, and Afghanistan (Sims-Williams & Bailey, 2002). Unlike English and German, Persian is categorized as a syllable-timed language (Windfuhr, 1979).

On the basis of evidence from the existing literature, we hypothesized that gender affects the recognition of emotional prosody in Persian language, and in particular there will be a female advantage in the recognition of emotional prosody. Using vocal emotion recognition task, we asked male and female participants to recognize the emotional prosody of a set of Persian emotional vocal portrayals. This study was conducted in Tehran in November and December 2012. The dialect examined in this study is Modern Conversational Persian as spoken in Tehran, Iran.

3.2 Method

Participants

Sixty six native speakers of Persian (33 females) ranging from 18 to 30 years participated in the experiment. Participants were divided into two gender- and age-controlled groups. Mean age of the

participants separated by gender was female: 24.5 (*SD* 4.9); male 24.9 (*SD* 4.7). Mean age of the whole group was 24.7 years (*SD* 4.8). Participants were university students recruited from the University of Tehran and roughly equivalent in years of formal education (15.6 ±1.2). All participants displayed normal or corrected-to-normal vision as verified by the examiner. They reported good hearing and did not have any psychopathological conditions, had no history of neurological problems, and took no psychoactive medication, as assessed by a detailed questionnaire. The study was conducted according to the ethical guidelines of the Declaration of Helsinki and participants gave their written informed consent. Participants received the Iranian Rial equivalent of eight Euro per hour as financial compensation.

Stimulus Material

In studying vocal emotions, researchers have often prepared their own experiment-specific vocal stimuli. Considering that well-prepared and validated stimuli are prerequisite to study vocal emotions (Castro & Lima, 2010), the stimuli used in this experiment were chosen from the Persian Emotional Speech Database (Persian ESD), an inventory of validated vocal stimuli (Keshtiari, Kuhlman, Eslami, & Klann-Delius, 2014). Persian ESD is the first validated emotional speech database for Persian comprising about 470 vocal utterances. This database is designed to be used as a reliable material source (for both text and speech) to examine the effect of prosody and lexical content on the identification of emotions in speech in behavioral, neuroimaging and clinical studies.

To establish Persian ESD, first in a series of experiments a set of sentences (lexical content) was generated and then validated. Two native Persian speakers (a male and a female speaker semi-professional in acting) then articulated these sentences in a series of basic emotions through their tone of voice “prosody”. The emotional categories included, anger, disgust, fear, happiness, sadness, and neutral. These vocal utterances were recorded on digital tapes under identical conditions, using a high-quality fixed microphone (Senheiser MKH 20 P48). The recordings were digitalized at a 16-bit/ 44.1 kHz sampling rate. The sound files were recorded on digital tapes (TASCAM DA-20 MK II), digitally transferred to a computer and edited to mark the onset and offset of each sentence. Following Pell and Skorup (2008), amplitudes were normalized to a peak intensity of 70dB (with *Adobe Audition* version 1.5) to control for unavoidable differences in the sound level of the source recordings across actors. These processes were performed in a professional recording studio in Berlin, Germany.

Validity of these vocal utterances was established perceptually by 34 (17 male, 17 female) decoders in a pilot study. Decoders were asked to identify the emotional category of the vocal utterances in a seven-choice emotion recognition task (choices: anger, disgust, fear, happiness, sadness, neutral, and

none of the above). Those utterances recognized five times above chance performance (71.42%) were selected as valid vocal portrayals. Additionally, acoustic analysis of the valid utterances showed obvious differences in pitch, intensity and tempo which may help listeners to correctly classify the intended emotion. These background data resulted in a controlled selection of utterances effectively conveying the intended emotions (see Keshtiari, Kuhlman, Eslami, & Klann-Delius, 2014 for extended details of this database).

The stimulus material selected for this experiment consisted of 12 syntactically similar sentences (i.e., subject + object + prepositional phrase + verb), neutral in lexical content intoned in the five intended emotions (anger, disgust, fear, happiness, sadness). Seventy two vocal utterances were used in total, 12 tokens per emotional category, plus 12 neutral tokens (which were treated as filler items); half of the utterances were uttered by the male and the other half by the female speaker. The duration of each utterance was about 4.2 seconds (*M* = 4170 ms; *SD* = 232). See Example (1) for a sample sentence and Appendix C for the compelete list of sentences.

(1) خانم خونه‌دار روسری گلدارش رو بعداز شام اتو کرد[1].

Xānom-e	xunedār	rusari-ye	goldār-eš	ro	baʔdaz	šām	ʔotu	kard.
Lady-Ez	housewife	scarf-Ez	damask-CL-3SG	DOM	after	dinner	iron	did-3SG[2]

(1) Ms. housewife ironed her damask scarf after the dinner.

Procedure

Each participant was tested individually in a quiet and dim lit room. Participants were seated in a chair in front of a laptop computer with a six-button answer pad before them. They were instructed to listen to the utterances, to focus on the voice of the speaker and to identify the emotional prosody of each utterance as accurately as possible. Consistent with most of the previous literature, the current study employed a forced-choice paradigm (e.g., Paulmann et al., 2008; Sauter et al., 2013). Response options corresponded to anger, disgust, fear, happiness, sadness, and an extra response option (i.e., neutral). A single practice run including six items was given prior to the start of the experiment. To limit

[1] Persian is written from right to left.

[2] Abbreviations used are as follows: Ez: ezafe particle; CL.3SG third person singular clitic; DOM: direct object marker; 3SG: third person singular.

fatigue and probable inattention, the stimulus set was presented in 2 blocks in a fully randomized design. Both blocks contained an equal number of trials, posed by an equal ratio of female and male speakers. The experiment was run as follows: vocal stimuli were presented from a laptop computer, controlled by E-Prime experimental presentation software (Schneider, Eschman, & Zuccolotto, 2002). The stimuli were played through high quality stereo headphones (Sennheiser HD600) with manual volume adjustment. Each trial sequence consisted of: a centrally-displayed fixation cross for 200 ms; a 200 ms pause; the vocal stimulus; a question mark indicating that emotion judgment decision should be made; and a 2000 ms pause.

At the end of the experiment participants were asked about the strategy they had employed in cases in which they had no clue as to the intended emotion of a vocal portrayal. Participants' comments were collected by the examiner.

3.3 Results

Descriptive statistics

Averaging across emotions, an overall accuracy rate of 90 % was perceived. This rate is approximately five times higher than chance level (having six response option, chance level was 16.6%). Fig. 1 presents the mean recognition rates for male and female participants separately. Mean accuracy rates for happiness (94.7%) and fear (92.7%) were highest, followed by anger (90.3%) and sadness (90.5%). Expressions of disgust (83.4%) were recognized at least accuracy overall. Comparable difficulties in recognizing the vocal portrayals of disgust have been reported in the literature, by Scherer et al. (1991).

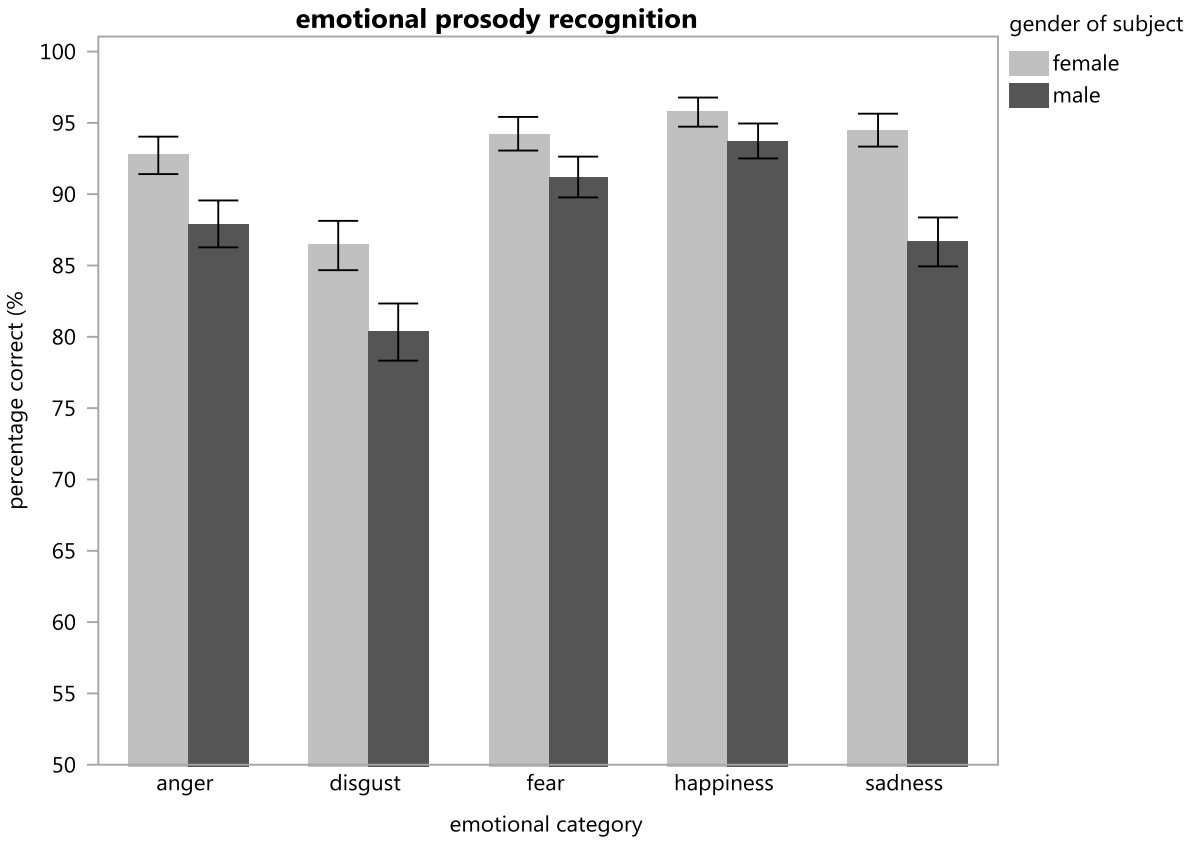

Fig. 3.1 Accuracy (in %) of emotional prosody recognition according to the gender of the participants (N = 66). Error bars represent the standard errors.

At the end of the experiment participants were asked about the strategy they had employed in cases in which they had no clue as to the intended emotion. Analysis of the participants' comments revealed that when they had difficulty recognizing the intended emotion, most of the participants (i.e., 58 out of 66) chose "Neutral" as the intended category. This may explain cases in which the classification "neutral" was given, while other emotion categories were intended.

Nominal logistic regression

Recognition data is often analyzed with methods that require an interval scaled level of measurement and homoscedasticity of data (e.g., Paulmann et al., 2008). Recognition Rates however are basically ordinal scaled and can only approximate an interval scaled level of measurement. Therefore we opted for a method to analyze ordinal data, *logistic nominal regression* (Hosmer, J., David W., Lemeshow, S., & Sturdivant, R. X. (2013), to obtain more accurate results. Using the software JMP 11 pro, we conducted a logistic nominal regression were gender (male, female) and emotion categories (anger,

disgust, fear, happiness, and sadness) were taken as predictor variables and error rates were considered as outcome variable. The regression converged in gradient after 6 interactions with an R^2 (U) = 0.04. As estimate of effect size we report Nagelkerke's R^2 = 0.96. The effect likelihood ratio test showed significant main effect for gender χ^2 = 28.45, *df* = 1, $p < .001$; and for emotion χ^2 = 61.82, *df* = 4, $p < .001$, revealing that across all emotion categories female participants had significantly performed better. However there was no interaction between gender and emotion categories χ^2 = 3.21, *df* = 4, *p* (n.s).

3.4 Discussion

The current study investigated the relation between gender and recognition of emotional prosody in an under investigated linguistic context (Persian as spoken in Iran). Taken together, our findings revealed that recognition accuracy rates differ significantly as a function of gender while listening to emotional vocal portrayals. Namely, there was an overall recognition advantage for female participants over male participants in our task. Averaging across emotions, an overall accuracy rate of 90 % was perceived. This rate is approximately five times higher than chance level (16.6%) in our task and in line with previous studies (e.g., Paulmann *et al*., 2008). So far, studies examining the recognition of emotional prosody demonstrate that recognition rates are almost four times above chance level (Pittam & Scherer, 1993). In addition, based on the rhythmic division of languages (stress- vs. syllable-timed), we drew an analogy between our results with that of English and German. Finally, our results showed recognition accuracy rates do not differ significantly as a function of emotional category. We will elaborate on these findings in the following paragraphs.

Influence of gender on emotional prosody recognition

Our investigations revealed a significant effect of gender of the decoders in favor of females in the recognition of emotional prosody. The direction of this effect is in line with general findings on emotion recognition (e.g., Briton & Hall, 1995; Hall *et al*., 2000; Schirmer & Kotz, 2003; Thompson & Voyer 2014) and with particular findings on the recognition of emotional prosody (e.g., Bonebright *et al*., 1996; Schirmer *et al*., 2002; Schirmer *et al*., 2005). This tentative conclusion therefore coincides with a number of behavioral studies reporting a higher sensitivity in females than in males in a range of conceptually comparable tasks of decoding nonverbal emotional cues (for a review see Hall, 1978; Hall *et al*., 2000). It is likewise consistent *w*ith gender role stereotypes (Broverman *et al*., 1972). One reason females might be more sensitive to emotional prosody is that females are more socially oriented and are more likely to define themselves in relational terms (Kashima *et al*., 2004). Besides this social accounts of gender

differences such as social institution of task-oriented female and male behavior (Kashima *et al*., 1995), our findings may be regarded as evidence of the point that gender differences in social orientation are biologically rooted, for instance by hormones such as estrogen and oxytocin (Babcock & Laschever, 2009; Wood & Eagly, 2002; Taylor *et al*., 2000). Additionally, our results further revealed that recognition accuracy rates did not differ significantly as a function of emotional category i.e., females performed better than males in recognizing all of the five intended emotions. Our findings are, thus, in line with those of Hall colleges (2000).

Rhythmic division of time and emotional prosody recognition

So far the role of gender on the recognition of vocal emotions has been explored only with stress-timed languages i.e., English (e.g., Bonebright *et al*., 1996) and German speakers (e.g., Schirmer *et al*., 2003, 2005). Our results with Persian speakers are in line with the results of German (Schirmer *et al*., 2003, 2005) and English speakers (Bonebright *et al*., 1996). Given the similarities between findings from Persian (i.e., syllable-timed language) and earlier findings from German and English (i.e., stress-timed languages), it seems that the observed gender differences may be relatively independent of rhythmic division of time (i.e., stressed- vs. syllable-timed).

Emotional prosody recognition

Previous research on nonverbal vocalizations and speech prosody have shown that acoustic cues (e g., pitch, intensity, duration) predict subjective emotion judgments (e.g., Banse & Scherer, 1996; Lima & Castro, 2011). Accordingly to determine how well the emotional utterances could be classified on the basis of the intended acoustic measures (i e., pitch, intensity, duration), in a previous study (Keshtiari *et al*., 2014) we conducted a discriminant function analysis on the intended measures. Results of this analysis indicated that there were obvious differences in the intended acoustic measures which could help participants to correctly classify the intended emotion (see Keshtiari *et al*., 2014 for details).

Results of the emotion recognition task indicated that participants categorized all five intended emotions (presented in the form of neutral lexical content intoned in emotional prosody) at a high rates (with a 90% overall accuracy rate). However our findings reveal that some emotions are decoded better than others: anger, fear, happiness, and sadness were recognized better than disgust (review Fig. 1). Differences in emotion recognition rates could be due to biological factors. As Williams and Mattingley (2006) report, detecting anger is essential in perceiving potential dangers. They further add that anger is

the emotion which helped primitive humans to survive the harshness of scarcity. Recognizing fear was likewise essential to survival (Olsson & Phelps, 2007).

Conclusions

The results of our study indicate that recognition accuracy rates differ significantly as a function of gender while listening to emotional vocal portrayals. Namely, there is an overall recognition advantage for female participants over male participants in our task. Furthermore, our results with Persian speakers are in line with the results of German (Schirmer et al., 2003, 2005) and English speakers (Bonebright et al., 1996). Considering that Persian is a syllable-timed language, our findings suggest that the rhythmic division of time (stressed- vs. syllable-timed languages) does not affect the direction of the results.

At the end of the experiment participants were asked what strategy they had employed in cases they had no clue as to the intended emotion of a vocal portrayal. Analysis of the participants' comments revealed that when they had difficulty recognizing the intended emotion, very often (i.e., 58 out of 66 participants) they chose "neutral" as the intended category. This may explain cases in which "neutral" was selected, while other emotion categories were intended

The present study has a number of limitations. The first concerns number of speakers. Vocal portrayals of the current study were intoned only by two speakers (a male and a female). A larger number of speakers will exclude the probable effects of speaker-specific idiosyncrasies and likely artifacts (Pell, 2002). Moreover, having more speakers of both genders also make it possible to investigate the relationship between the gender of the speakers and that of the participants in the recognition of emotional prosody. The second limitation concerns the use of forced-choice response format. Following previous studies (e.g., Paulmann et al., 2008; Sauter et al., 2013), we asked participants to respond based on a forced-choice task corresponding to anger, disgust, fear, happiness, sadness, and an extra response option (i.e., neutral). On one hand, forcing participants to choose an option from a short list of emotions may inflate agreement scores and produce artifacts (for detailed discussions on limitations of forced-choice response paradigms see, e.g., Russell, 1993, 1994; Wagner, 1993; Scherer, 2003). On the other hand, providing participants with more options or allowing them to label the emotions freely would result in very high variability (Banse & Scherer, 1996; Russell, 1994). However, if participants are provided with an extra response option (e.g., none) together with a discrete number of emotion choices, some of the artifacts can be avoided (Frank & Stennett, 2001). Therefore, in future studies it is optimal to use this extra response option to elude probable artificial high recognition rates. The last limitation concerns generalization of the findings. Only one syllable-timed language (i.e., Persian) was examined in the current

study. Accordingly only replication of the present study in other syllable-timed languages would allow generalization of the findings.

In conclusion, the present study is the first of its kind to examine the recognition of emotional prosody in an under investigated syllable-timed language (i.e., Persian) as opposed to stress-timed languages (e.g., English and German). The results obtained established that females outperform males in the recognition of emotional prosody. The female advantage was found for all the five emotional categories examined. Our findings also indicated that male and female listeners can successfully identify vocal portrayals of the five intended emotions only based on prosodic cues (i.e., participants were provided with no semantic emotional cues). These results add to the growing body of literature examining the link between gender and recognition of emotional prosody. In particular findings of the present study may also be useful for the development of interactive computer systems that recognize and synthesize vocalizations that vary with emotional state and gender (see Rao, Koolagudi, & Vempada, 2013; Koolagudi, & Rao, 2012, Gobl & Ní Chasaide, 2003 for detailed explanation of these systems).

Chapter 4: STUDY THREE

Gender effects on the recognition of emotional speech: Evidence from a collectivist culture

Keshtiari, N., Kuhlmann, M., & Klann-Delius, G. (Submitted). Gender effects on the recognition of emotional speech: Evidence from a collectivist culture.

Abstract

So far cross-cultural studies on emotions have produced a wealth of information regarding cultural differences and similarities in the communication of emotions. Still there remain several gaps in our knowledge. Emotional speech which carries a rich source of information about a speaker's emotions is influenced by individual differences (e.g. gender, age); among those, the role of gender is still under debate. Moreover, cultural variables mediate and moderate gender differences in emotional functioning. The existing literature on gender and emotional speech has only targeted speakers of a specific cultural background (i.e., individualist). However, as culture plays a significant role in the recognition of vocal emotions, results of the existing study might not apply to members of other cultural groups. Therefore, the present study examined the role of gender in the recognition of emotional speech in a different cultural context (i.e., a collectivist society).

In a behavioral experiment, native speakers of Persian were asked to recognize the emotional prosody of a set of validated emotional vocal portrayals (including anger, disgust, fear, happiness, and sadness) articulated by a male and a female actor.

In general, our results established that both biological and cultural factors modulate the recognition of emotional speech. More specifically, females outperformed males in the recognition of emotional speech. Additionally, it was revealed that members of a collectivist culture show higher sensitivity to vocal emotional cues compared to their individualist counterparts.

Keywords: gender, culture, emotional speech, collectivist, individualist

4.1 Introduction

Communication of emotions is a fundamental aspect of social life (Ekman, 1992). Taking into account that accurate processing of emotions is crucial in comprehending the social world (Van Rheenen & Rossell, 2013), there is an increasing interest in studying this domain. Humans must encode and decode vocal emotional expressions of others efficiently or risk a breakdown in social communication (Surcinelli, Codispoti, Montebarocci, Rossi & Baldaro, 2006). Therefore, it is of considerable social relevance to determine which factors affect the encoding and decoding of emotions in interpersonal communication. As a contribution to this literature, the current study investigated the recognition of the five intended emotions[1] (i.e., anger, disgust, fear, happiness, and sadness) from vocal expression as a function of gender and culture.

Vocal expression is a modality that is considered as an effective means of emotional communication (Scherer, 1986). In this modality, emotions are transferred via two channels: (a) lexical content (the emotional meaning of the words and sentences), and (b) emotional prosody (the emotional melody of speech) (Szymanowski, Kotz, Schröder, Rotte, & Dengler, 2007; Lima & Castro, 2011). Emotional prosody, one of the most basic features of language (Besson, Magne & Schön, 2002), is defined as the ability to express emotions through variations of speech parameters, including intensity, pitch, tempo, and rhythm (Schirmer & Kotz, 2006). By carrying a rich source of information about a speaker's emotions and social intentions (Wilson & Wharton, 2006; Banse & Scherer, 1996; Scherer, 1986), emotional prosody plays a crucial role in various aspects of human social communication (Plutchik, 2003).

Gender is one of the individual difference factors which affects cognitive performance (Herlitz & Lovén, 2009). Gender differences in specific cognitive functions are well documented (see Donges, Kersting, & Suslow, 2012 for a review). One such example is male dominance in mental rotation and female advantage in object location and verbal fluency (Kimura, 1999). However, studies investigating gender differences in other social cognitive functions such as emotion recognition are still few in number (Bonebright, Thompson & Leger, 1996; Paulmann, Pell & Kotz, 2008). So far, most researchers investigating the role of gender in emotion recognition have focused on the domain of emotional prosody, where neutral lexical content is articulated in specific emotional voice (e.g., Schirmer, Kotz & Friederici,

[1] These specific set of emotions were selected because they are known to be recognizable across cultures (Sauter, Eisner, Ekman, & Scott, 2010) while having specific prosodic profiles (Banse & Scherer, 1996).

2002; Schirmer & Kotz, 2003; Raithel & Hielscher-Fastabend, 2004; Fecteau, Armony, Joanette, & Belin, 2005; Toivanen, Väyrynen & Seppänen, 2005). For instance, Schirmer *et al.,* (2002) in an event-related potential (ERP) experiment found gender differences in the sensitivity to emotional prosody. They reported, that during word processing female participants integrated words into emotional prosody at an earlier point in time. In another study Schirmer and Kotz (2003) examined the influence of emotional prosody and word valance on both genders in an emotion judgment task. Their results revealed that only females combined emotional prosody and lexical content when emotional prosody was task-irrelevant. Toivanen, Väyrynen and Seppänen (2005) asked a group of male and female participants to recognize the emotional vocal portrayals (neutral lexical content intoned in emotional voice) of 14 professional speakers (six women, eight men) and found out that females outperform males in the recognition of emotional prosody. In a recent study Keshtiari, Kuhlmann, and Klann-Delius (submitted) investigated gender effects on the recognition of emotional prosody where participants were asked to identify the emotional prosody of a set of vocal portrayals with neutral lexical content. Their results illustrated a significant female advantage. In addition to the investigations performed on healthy individuals, recent evidence from bipolar patients (Van Rheenen & Rossell, 2013) illustrated the presence of a female advantage in the processing of emotional prosody.

Compared to the literature on the recognition of emotional prosody (i.e., non-emotional lexical content intoned in emotional voice), very little is known about the role of gender in the recognition of emotional speech, where emotional lexical content is articulated in a congruent emotional voice (Paulmann *et al.,* 2008). So far only Paulmann and colleagues (2008) have exclusively tackled this question in a behavioral study. In an emotion recognition task they asked a group of young (18 -30 years old) and middle aged (38- 50 years old) participants (native German speakers) to recognize the emotional category of a set of emotional vocal portrayals (emotional lexical content articulated in a congruent emotional voice by two male and two female German speakers). Their stimulus material included 350 syntactically similar sentences, grouped into one of the six emotional categories: anger, disgust, fear, happiness, sadness, and pleasant surprise, or semantically neutral category. Averaging across emotions they reported an overall accuracy rate of 70% (almost close to five times higher than chance level). They reported an overall recognition advantage for young participants compared to the middle aged. But they found no significant gender difference in the recognition of emotional speech. However, due to the interplay of language and emotion in vocal communication, the results achieved by Paulmann *et al.,* (2008) may be specific for the German culture and/or language. Therefore, as no study has further examined this issue within other cultural background, it is still unclear whether these findings apply to members of other cultural context.

Further to possible gender effects in the recognition of vocal emotions, it has been noted that due to the motivational and the adaptive communicative functions of emotions, socio-cultural variables mediate and moderate gender differences in emotional functioning (Brody & Hall, 2008). Considering that expectations, conventions, and the rules to with each speaker is supposed to adhere are idiosyncratic to each particular society with its specific cultural context, Tanaka (2015) notes that the relation between language and gender has to be analyzed within the given society where the language is communicated. Additionally, Pell, Monetta, Paulmann, and Kotz (2009) argue that as a result of the interplay of language and emotion in vocal communication, socio-cultural influences on emotion recognition are specifically conspicuous in the emotional speech.

It is generally acknowledged that the ability to recognize and express emotions is an initial human potential (Matsumoto, 1989); cross-cultural endeavors concerning recognition of facial and vocal expression of emotions have strengthened theories of universality of human emotional capability specifically concerning the so-called 'basic emotions' such as anger, fear, happiness and sadness (e.g., Pell, Monetta, Paulmann, & Kotz, 2009; Pell & Skorup, 2008; Thompson & Balkwill, 2006; Scherer & Wallbott, 1994; Ekman, 1973; Ekman, Friesen, & Ellsworth, 1972). Nevertheless, from the view point of cross-cultural psychology and anthropology, the biological perspective is not adequate for a comprehensive understanding of the perception and expression of emotions (Eid & Diener, 2001). As emotions and emotional processes are largely influenced by culture[1] (Matsumoto, 2001; Markus & Kitayama, 1991), for a full understanding of emotional processes we should consider the cultural context where emotions are perceived, expressed, and experienced (e.g., Mesquita & Ellsworth, 2001; Scherer, 1997; Kitayama & Markus, 1995; Scherer & Wallbott, 1994, Markus & Kitayama, 1991). As Consedine, Magai and Bonanno (2002) note, "Through socialization, culture determines the kinds and frequencies of events to which children are exposed; the kinds of reactions that receive scaffolding and support or, conversely, punishment and containment; and the kinds of social expectations for comportment" (p.217). Likewise Mesquita (2000) believes that emotion and culture are intermingled with each other. On the one hand emotions are shaped in a way comparable to ideas of the cultures in which they develop. On the other hand emotions reinforce the cultural themes of a given society. As such, in any given cultural context,

[1] It is difficult to define the term *culture* as it has been used in different ways within and across disciplines. Roughly speaking, culture designates shared commonalities within groups of people (Matsumoto, 2007; Triandis, 1994). In the context of this paper, the term "culture" refers to shared elements (e.g., rituals and habits, belief and value system, social institutions, language, and behavior patterns) within any interactive collectivity of people (Scherer & Brosch, 2009) as well as the basis of the individual preferences in daily life (Matsumoto, Grissom & Dinnel, 2001). In cross-cultural research, culture is typically depicted by country (Matsumoto, Takeuchi, Andayani, Kouznetsova, & Krupp, 1998).

behaviors and emotional experiences that fit with and enhance the goals of that specific culture are reinforced (Fischer & Manstead, 2000). In support of this perspective, various studies have reported cultural differences in how emotional expressions are decoded (e.g., Mesquita, & Markus, 2004; Fischer & Manstead, 2000; Matsumoto, Kasri & Kooken, 1999; Matsumoto & Kudoh, 1993; Matsumoto,1992; 1989), in when and how emotions are expressed (e.g., Ekman, 1972; Matsumoto, 1993; Stephan, Stephan, & de Vargas, 1996; Tsai & Chentsova-Dutton, 2003), in the events that elicit emotions (see Mesquita & Frijda, 1992 for a review), in the frequency and intensity of emotional experiences (e.g., Scherer, Matsumoto, Wallbott, & Kudoh, 1988; Soto, Levenson, & Ebling, 2005), and in the way by which emotion-eliciting events are assessed (e.g., Matsumoto, Kudoh, Scherer, & Wallbot, 1988; Roseman *et al.,* 1995; Scherer, 1997).

Ekman and Friesen (1969) argue that cultural variation in the perceived importance of social interactions is manifested in *norms* for decoding and displaying emotions. Cultural differences via these norms regarded as *display rules*[1] are well established in psychology (Koopmann-Holm & Matsumoto, 2011; Thompson & Balkwill, 2006; Matsumoto *et al.,* 1998; Matsumoto, 1990; Malatesta & Haviland, 1982; Ekman & Friesen, 1975). Violation of any of these rules by displaying (Ekman, 1972) or recognizing (Matsumoto, 1992) an emotion which does not fit the cultural context, can threaten the degree of harmony and disrupt social interactions. So far within- and cross-cultural studies on emotions have produced a wealth of information regarding cultural differences and similarities with regard to display rules in the communication of emotions (see Koopmann-Holm & Matsumoto, 2011 and Matsumoto, 1990 for detailed explanations and examples). For instance Hall (1976) argues that to express emotions in vocal communication, members of Asian cultures express less explicit verbal information and rely more on contextual cues to interpret the communication. Likewise, Markus and Kitayama (1991) note that, collectivistic cultures (i.e., Asian cultures) are more context-dependent than individualistic cultures. Furthermore, due to the very restrictive emotional display rules within Asian cultures, the contextual cues provided by the interaction partners in these cultures are very subtle (Matsumoto, Kasri, & Kooken, 1999;

[1] The concept of emotional "display rules" was introduced for the first time by Ekman and Friesen (1969) as a hypothetical notion in a study on culture and emotion to explain the observed differences between Japanese and Americans. Emotional display rules are cultural norms observed as an important part of any culture; they refer to culturally prescribed rules learnt early in life via socialization (Koopmann-Holm & Matsumoto, 2011). These rules affect the emotional expression of people from any culture depending on what that particular culture has featured as an appropriate or inappropriate expression of emotion (Matsumoto, Kasri, & Kooken, 1999). These culturally shared norms and unwritten codes dictate when, how and to whom individuals should express their emotional experiences. (Matsumoto 1990). Matsumoto (1990) and Koopmann-Holm and Matsumoto (2011) further note that display rules not only encompass the dimension of expression appropriateness but also appropriateness of recognition and evaluation of a behavior.

Matsumoto, 1990). To avoid breakdown in interpersonal communication and maintain interpersonal harmony in such cultural contexts great sensitivity is needed for in the perception and interpretation of the implicitly conveyed emotional cues (Safdar *et al.,* 2009; Matsumoto *et al.,* 1999). Schirmer *et al.,* (2006) argue that this high sensitivity in the perception of implicit emotional vocal cues may be absent in members of Western cultures. If true, this would suggest that culture modulates the recognition of emotional speech.

To study emotions within cultural context, Matsumoto (1989) argues that it is essential to make use of stable and meaningful dimensions of cultural variability proposed by Hofstede (1980). Matsumoto (1990) and Matsumoto *et al*. (1998) further note that from among the dimensions[1] of cultural variability, *individualism-collectivism* which is conceived as two poles of one dimension, is a meaningful predictor of cultural variations in emotional norms (i.e., display rules). Similarly in an experimental study Mesquita (2000) examines the concept emotion in two different cultures (Dutch participants as members of individualist society, and Turkish participants as members of a collectivist culture) and claims that *individualism-collectivism* is not causal origins of emotions but rather the features of the signs of which emotions are a part. Likewise, studying cultural differences, several researchers have recognized *individualism-collectivism* as a stable dimension of cultural variability (e.g., Hofstede, 2001, 1980, 1983; Triandis, 1972, 1994).

Individualism is characterized by independent relationships where the priority is on the self (Triandis, 1994). In such cultures the concept of self is concerned with independence, where the independent individual is believed to contain unique internal features and is expected to behave in accordance with these features (Markus & Kitayama, 1994). Individualists tend to emphasize and enhance their personal goals, interests and values over the society they belong to (Markus & Kitayama, 1991; Triandis, 1995; Oyserman, Coon, & Kemmelmeier, 2002). Self-enhancement, the need for individual autonomy, detachment from others (Scherer & Brosch, 2009; Andersen, Reznik, & Chen 1997) and functioning based on personal choices (Walsh & Banaji, 1997) are main features of Individualistic cultures. Considering these features, individualist cultures regard emotions as essential personal experiences

[1]*Cultural dimensions theory* developed by Hofstede (1980) is a framework for cross-cultural communication. Using a structure *emanated from* factor analysis, the theory explains the influence of a society's culture on the values of its members, and that how these values relate to their behavior. In this context *dimension* refers to an aspect of a culture that can be measured relative to other cultures. Cultural dimensions theory is widely used in various fields (e.g., cross-cultural psychology) as a paradigm for research (Matsumoto, 1990). For more explanation on Cultural dimensions theory see Hofstede (1983) and (1991).

[2] Dimensions of cultural variability include a) Individualism-Collectivism, b) Power Distance, c) Uncertainty Avoidance, d) Masculinity-Femininity, e) Long Term Orientation-Short Term Normative Orientation, and f) Indulgence-Restraint (Hofstede, 2001).

whose expression is any individual's right. In these cultural contexts, individuals tend to consider emotions as inner states vented spontaneously (Matsumoto, *et al*., 1998). Individualist cultures reinforce outward displays of emotion[1] that exaggerate the strength of the feeling (Matsumoto, *et al.,* 1998). Similarly, Mesquita (2000) claims that in individualist cultures emotions are assumed to amplify and underline a subjective self.

In contrast to individualism, collectivism is featured by interdependent relationships which focus on social cohesion and a group[2]'s harmony (Triandis, 1994). In a collectivist culture the concept of self is bound to surrounding social context in which the self cannot be split from others and the goal is to keep oneself tied to others (Markus & Kitayama, 1994). Collectivists identify themselves as members of a group they belong to, therefore they emphasize and enhance group goals, interests and values over those of each individual member of the group (Markus & Kitayama, 1991; Triandis, 1995; Oyserman, *et al.,* 2002). Self- transcendence, the need for harmony among the in-group (Scherer & Brosch, 2009; Oyserman et. al. 2002; Markus & Kitayama, 1991), attachment to others (Andersen, *et al.,* 1997) and functioning based on group priorities (Walsh & Banaji 1997) are main features of collectivist cultures. In these cultures emotions are regarded as interactive experiences which mirror the social context rather than the individual's inner self. As Mesquita (2000) argues, in such cultural contexts expression of emotion is controlled, due to the point that it is grounded in evaluation of the relationship between the self and the others. In other words in collectivist cultures emotions are expected to emphasize and reproduce the self in relation to others and the world (Mesquita, 2000).This issue implies that emotions are conceived as situation-specific clues about the relationship between the individuals. By and large, collectivist cultures reinforce the perpetuation of cohesion within the group and accordingly, control of emotions gains high priority (Potter, 1988).

In a recent meta-analysis of 83 studies, Oyserman, *et al.,* (2002) identified seven major domains featuring individualism (i.e., independence, direct communication, goals, uniqueness, competition, privacy, and self-knowing) and eight major domains pertaining to collectivism (i.e., group orientation, harmony, relatedness, belonging, duty, context dependency, advice seeking, and hierarchy).

So far, the only existing study on gender and emotional speech recognition was carried out on German speakers living Leipzig, Germany (Paulmann *et al.,* 2008). According to the Hofstede Model of

[1] Although collectivism and individualism characterize cultural groups, not everybody in a given cultural context engages in the same ideas and practices, nor do they try them in identical manners (Markus, Mullally, & Kitayama, 1997). However, across individuals within a collectivist or individualist context, emotions are patterned in discernable ways (Mesquita, 2000).

[2] For instance: family, tribe, and nation.

Cultural Dimensions[1] (Hofstede, 2001), Germany with a relatively high score of 67 ranks 15 among the 53 countries[2] and is considered as an individualistic society (see Hofstede, 2001 & The Hofstede Center, n.d. for more details). According to Safdar *et al*. (2009) and Greenfield *et al*. (2003) the Hofstede ranking on individualism-collectivism (Hofstede, 2001) represents extensive distillations of various socio-psychological phenomena (encompassing social beliefs, values, norms, personality and behavioral characteristics) that pertain to ways in which human-environment relations are assessed.

Considering that (1) emotional processes are largely influenced by culture (Markus & Kitayama, 1991), (2) socio-cultural variables mediate and moderate gender differences in emotional functioning (Brody & Hall, 2008), and that (3) due to the interplay of language and emotion in vocal communication, socio-cultural influences on emotion recognition are specifically conspicuous in the emotional speech (Pell *et al.,* 2009), the current gap in the literature of gender effects in the recognition of emotional speech will be filled with further experimental studies on speakers of different linguistic[3] and cultural background. The present study, therefore, seeks to address this paucity of research by making the first attempt to investigate the potential role of gender in the recognition of emotional prosody in a different and under investigated cultural context (i.e., Persian speakers living in Tehran, Iran).

Persian (also known as Farsi) is an Indo-Iranian language, a sub-branch of Indo-European family (Anvari & Givi, 1996) spoken by almost 110 million people around the world, while holding official status in Iran, Tajikistan, and Afghanistan (Sims-Williams & Bailey, 2002). Modern Iran, as a developing Asian country with rich culture and old history, is an exceptional sociolinguistic laboratory for researchers (Modarressi-Tehrani, 2001). As Beeman (1986) argues, personal relations among Iranians are comparable to an art which requires sophisticated skills. Additionally, Ahmadi and Ahmadi (1998) contrast the Iranian *way of thinking*[4] with the Western ways of thinking and argue that in a broad sense Iranian way of thinking is very different from that of Westerners. They add that for Iranians an individual's value is tied to its relationship with "other selves". They further note that in Iran there is a tight link between the good of oneself and that of others, implying that people are highly related to each other. As relatedness between

[1] With regard to analyzing a country's culture, the Hofstede Model of Cultural Dimensions is a widely accepted model (Matsumoto, Yoo & Fontaine, 2008).

[2] These data on culture dimension came from Hofstede's (2001) longitudinal study of cultural dimensions from 50 countries and three regions.

[3] In a recent study Keshtiari *et al.,* (submitted) have addressed how linguistic differences between German and Persian affect the recognition of vocal emotions.

[4] The phrase "way of thinking" refers to any individual's thinking in which the characteristic feature of the thinking habits of the culture to which he belongs are revealed (Rosenthal, 1977).

people and saving one's face within a group is highly promoted in Iran (Beeman, 1986), relationships (e.g., employer/employee, teacher/student, parents/children) are regarded as moral terms similar to family link and a major concern for those living in such cultures is being perceived as qualified for relationships (Mesquita, 2000). Taken together, these features make Iranians highly attentive to others' emotions and needs (Ahmadi & Ahmadi, 1998; Beeman, 1976). In Hofstede's (2001) analysis, out of the 53 countries studied, Iran with the score of 41 ranked 24 on the *individualism-collectivism* dimension, marking it a collectivist society.

On the basis of evidence from the existing literature, we suggest that culture influences the recognition of emotion speech. Therefore we hypothesized that 1) gender affects the recognition of emotional speech, and in particular there will be a female advantage in the recognition of emotional prosody and, 2) compared to the members of an individualist society, members of a collectivist culture show higher sensitivity to vocal emotional cues.

In a vocal emotion recognition task, we asked male and female participants to recognize the emotional prosody of a set of vocal portrayals (sentences with emotional lexical content articulated in a congruent emotional voice). This study was conducted in Tehran in November and December 2012. The dialect examined in this study is Modern Conversational Persian as spoken in Tehran, Iran.

4.2 Method

Participants

Seventy native speakers of Persian (35 females) participated in the experiment. The data for four participants were excluded from the analysis due to their excessively high error rates (i.e., above 36%). Thus the data from 66 participants (33 females) were analyzed. The distribution of age was controlled with regard to the gender of the participants. The mean age of the participants was 25.4 years, *SD* =4.2, ranging from 18 to 30 years. The mean age of the participants separated by gender was: female: 25.2, *SD* 4.3; male 25.6, *SD* 4.1. Participants were roughly equivalent in years of formal education (14.9 ±1.4). All participants displayed good hearing and normal or corrected-to-normal vision as verified by the examiner. They did not suffer from any psychopathological conditions, had no history of neurological problems, and took no psychoactive medication, as assessed by a detailed questionnaire. At the beginning of the experiment participants were informed of the conditions of the experiment and gave written informed

consent. The study was conducted according to the ethical guidelines of the Declaration of Helsinki and participants gave their written informed consent. Participants received the Iranian Rial equivalent of eight Euro per hour as financial compensation.

Stimulus Material

In studying emotional speech, researchers have often prepared their own experiment-specific vocal stimuli. Since well-prepared and validated stimuli are prerequisite to study emotional speech (Castro & Lima, 2010), the stimuli used in this experiment were chosen from the Persian Emotional Speech Database (Persian ESD), an inventory of validated vocal stimuli. Persian ESD is the first validated emotional speech database for Persian comprising about 470 vocal utterances.

To establish Persian ESD, first in a series of experiments a set of sentences (lexical content) was generated and then validated. The intensity of the emotion conveyed via lexical content of each sentence may affect the participants' recognition of the intended emotions (Keshtiari, *et al.,* 2014). Therefore, the emotional intensity of each of the sentences were then identified in a separate pilot study. Two native Persian speakers (a male and a female speaker semi-professional in acting) then articulated these sentences in a series of basic emotions through their tone of voice "prosody". The emotional categories included, anger, disgust, fear, happiness, sadness, and neutral. These vocal utterances were recorded on digital tapes under identical conditions, using a high-quality fixed microphone (Senheiser MKH 20 P48). The recordings were digitalized at a 16-bit/ 44.1 kHz sampling rate. The sound files were recorded on digital tapes (TASCAM DA-20 MK II), digitally transferred to a computer and edited to mark the onset and offset of each sentence. Following Pell and Skorup (2008), amplitudes were normalized to a peak intensity of 70dB (with *Adobe Audition* version 1.5) to control for unavoidable differences in the sound level of the source recordings across actors. These processes were performed in a professional recording studio in Berlin, Germany. In a pilot study all the vocal utterances were perceptually validated by 34 (17 male, 17 female) decoders. Decoders were asked to identify the emotional category of the vocal utterances in a seven-choice emotion recognition task (choices: anger, disgust, fear, happiness, sadness, neutral, and none of the above). Those utterances recognized above five times chance performance (71.42%) were selected as valid vocal portrayals. Additionally, acoustic analysis of the valid utterances showed obvious differences in pitch, intensity and tempo which may help listeners to correctly classify the intended emotion. These background data resulted in a controlled selection of utterances effectively conveying the intended emotions (see Keshtiari, Kuhlman, Eslami, & Klann-Delius, 2014 for extended details of this database).

The stimulus material selected for this experiment consisted of 72 vocal utterances (i.e., 12 sentences in each of the five intended emotional category plus 12 sentences in the neutral category). These vocal utterances were generated based on a syntactically similar lexical content (i.e., subject + object + prepositional phrase + verb) intoned by a male and a female native speaker of Persian. These vocal utterances were previously matched based on the emotional intensity of their lexical content via *Match*[1], a program which assists in matching the conditions of factual experiments (Van Casteren & Davis, 2007). Therefore, the numerical values obtained in the previously mentioned pilot study were analyzed and for each emotional category 12 items were selected which were matched for the intensity of lexical emotional meaning (see Appendix D and E for the list of sentences along with their English translation).

Procedure

Participants were tested individually in a quiet and dim lit room. Each participant was seated in a chair in front of a laptop computer with a six-button answer pad before him/her. They were instructed to listen to the utterances, to focus on the voice of the speaker and to identify the emotional prosody of each vocal utterance based on a six-choice response format corresponding to anger, disgust, fear, happiness, sadness, and neutral. A single practice run including six items (one example for each of the five intended emotions plus one for the neutral mode) was given prior to the start of the experiment. The stimulus set was presented in 2 blocks in a semi- randomized design to limit fatigue and probable inattention. Both blocks contained an equal number of trials, posed by an equal ratio of female and male speakers. The experiment was run as follows: vocal stimuli were presented from a laptop computer, controlled by E-Prime experimental presentation software (Schneider, Eschman, & Zuccolotto, 2002). The stimuli were played through high quality stereo headphones (Sennheiser HD600) with manual volume adjustment. Each trial sequence consisted of: (1) a centrally-displayed fixation cross for 200 ms, (2) a blank screen for 200 ms, (3) an image of a loudspeaker with audio presentation of an item for the duration of the item, (4) a question mark indicating that emotion judgment decision should be made presented until response, and (5) a blank screen for 2000 ms. Figure 4. 1 represents the schematic illustration of the procedure.

[1] Before using *Match*, in a separate study the emotional intensity of each of the sentences (lexical content) were calculated and presented in the form of numerical values (see Keshtiari *et al*., 2014 for the details).

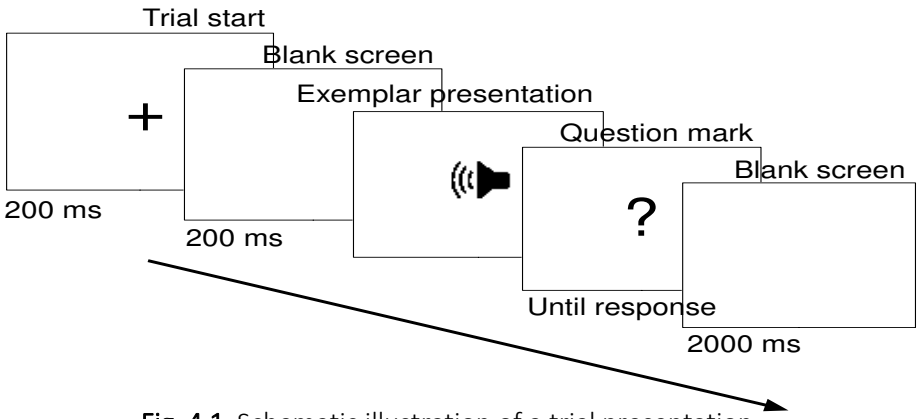

Fig. 4.1 Schematic illustration of a trial presentation

4.3 Results

Descriptive statistics

Overall, emotional speech recognition rate was well above chance level (having six response option, chance level was 16.6 %). Table 4.1 and Fig. 4.2 present the mean recognition rates for male and female participants separately. As can be seen in Table 4.1 mean accuracy rates for sadness (97.8 %) and happiness (97.5 %) were highest, followed by disgust (95.6 %). Expressions of anger (94.1 %) and fear (94%) were recognized with lower accuracy overall.

Table 4.1

Accuracy rates of emotional speech recognition according to gender of participants

Emotion	Female	Male	Average
Anger	96.2	91.9	94.1
Disgust	97.2	94	95.6
Fear	96.2	91.7	94
Happiness	99.5	95.5	97.5
Sadness	99	96.5	97.8

Values are the mean correct answers (in %)

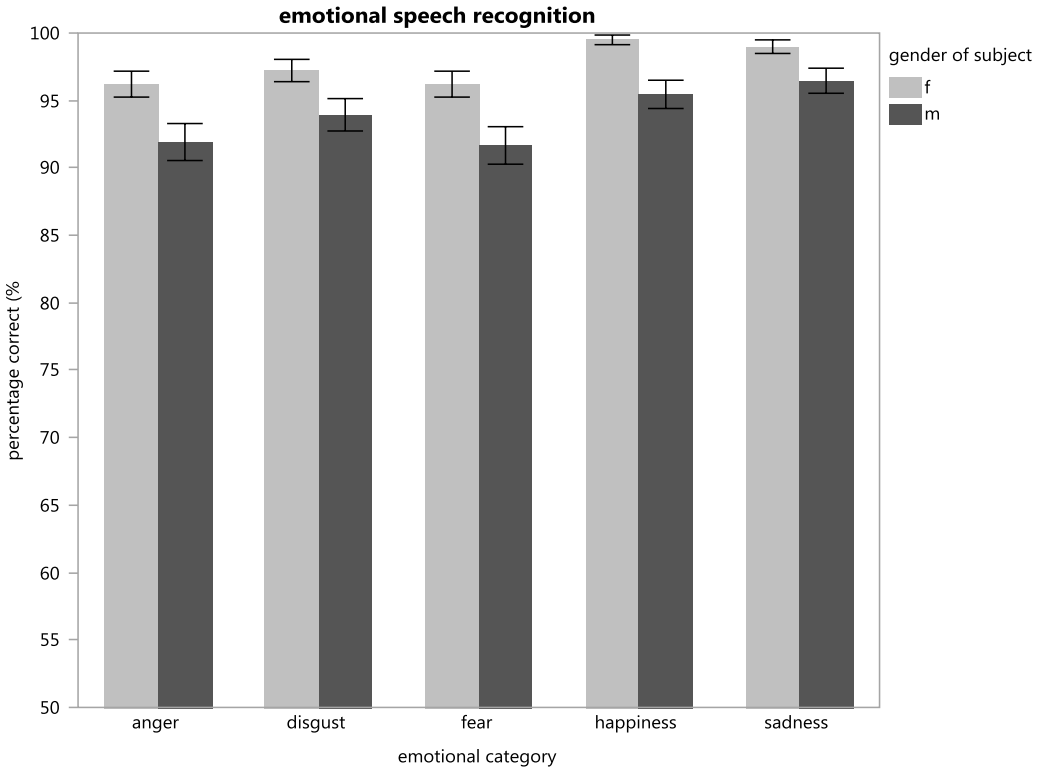

Fig. 4.2 Accuracy (in %) of emotional speech recognition according to the gender of the participants. Error bars represent the standard errors.

Nominal logistic regression

We conducted a logistic nominal regression using the software JMP 11 pro. The regression converged in gradient after 8 interactions with an R square (U) = 0.0480 the effect likelihood ratio test showed highly significant main effect for gender chi-square (1) = 38.56, $p < .001$; and for emotion chi-square (4) = 30.23, $p < .001$. However there was no interaction between gender and emotion categories chi-square (4) = 4.86, p (n.s).

Descriptive statistics of errors

Error patterns are reported in Table 4.2 for male and female participants separately. The very low rate of errors (< 0.5) implies the use of the validated high quality stimuli.

Table 4.2 Error Analysis: the matrix shows the confusion pattern for errors made in the emotional recognition task split by participants' gender

Group	Emotion	Intended emotion					
		Anger	Disgust	Fear	Happiness	Sadness	Neutral
Female participants	Anger		0	1.8	0.3	0.3	1.5
	Disgust	1		0.5	0.8	0.3	0.3
	Fear	0.8	0		0	3	0
	Happiness	0	0	0		0.5	0
	Sadness	0	0	1	0		0
Male participants	Anger		0.3	2.8	1.5	0.3	3.3
	Disgust	0.5		1	2	1	1.5
	Fear	2.8	0.5		0.5	4.9	0
	Happiness	0	0	0		3.8	0.3
	Sadness	0	0	4.5	0		0

Values are the mean incorrect answers (in%)

4.4 Discussion

The current study investigated the effect of gender on the recognition of emotional speech in an under investigated cultural context (Persian speakers living in a collectivist society: i.e., Iran). Taken together, our findings revealed that recognition accuracy rates differ significantly as a function of gender while listening to emotional speech where both the lexical content and the prosody portray a same emotional meaning. Namely, there was an overall recognition advantage for female participants over male participants in our task. Averaging across emotions, an overall accuracy rate of 95 % was perceived. Finally, to compare the recognition rates within two different cultures and to have an over view of the cultural effects, we conducted a content-based comparison with the study of Paulmann *et al.*, (2008). We will elaborate on these findings in the following paragraphs.

Emotional speech recognition

In a previous study (Keshtiari *et al.,* 2014), acoustic analysis of the emotional vocal portrayals (used as the stimuli of the current study) showed that there were obvious differences in pitch, intensity and tempo which might have helped participants to correctly classify the intended emotion. In the present study, participants categorized all five intended emotions (anger, disgust, fear, happiness, sadness) at a very high rate. Averaging across emotions, our results revealed an overall accuracy rate of 95% (which is very close to six times chance performance). As the accuracy recognition rates were very high, error rate were very low. Besides the use of validated stimuli and comprehensive task, this could be due to the high sensitivity of the participants.

Influence of gender on emotional speech recognition

Supporting hypothesis 1, our investigations revealed a highly significant effect of gender in favor of females in the recognition of emotional speech. The direction of this effect is in line with general findings on emotion recognition (e.g., Briton & Hall, 1995; Hall *et al.,* 2000; Schirmer & Kotz, 2003) and with particular findings on recognition of emotional prosody (e.g., Bonebright *et al*., 1996; Schirmer *et al*., 2002, 2006; Schirmer & Kotz 2003). Our findings are thus consistent with Hall and colleagues' (2000) assumption that females are reliably more accurate in a range of conceptually comparable tasks of decoding nonverbal emotional cues. This implies that biologically rooted social mechanisms (i.e., gender) may underlie the observed results. In addition, our results further revealed that recognition accuracy rates did not differ significantly as a function of emotional category i.e., females performed better than males in recognizing all of the five intended emotions.

Influence of culture on emotional speech recognition

Our investigations further revealed very high overall accuracy rate of 95% (almost close to six times chance performance). In General these results are in line with Safdar *et al.* (2009) and Matsumoto *et al.* (1999) findings with regard to the great sensitivity of the members of collectivist cultures as for the perception and interpretation of the emotional cues. To maintain interpersonal harmony and to avoid breakdown in communication this great sensitivity is a vital part of vocal communication in collectivist cultures (Safdar *et al.*, 2009). In particular, these results support Ahmadi and Ahmadi (1998) and Beeman (1976) arguments regarding Iranians' high sensitivity to others' emotions. Considering that Paulmann *et al.* (2008) have reported an overall accuracy rate of 70% (almost close to five times chance performance) in a similar emotion recognition task, our results indicate a higher sensitivity to emotional vocal emotional

cues in the members of a collectivist culture. These findings are, therefore, consistent with Schirmer *et al.,* (2006) argument regarding the lower sensitivity in the perception of emotional vocal cues in members of individualist cultures. Accordingly, our second hypothesis concerning the higher sensitivity to vocal emotional cues by the members of a collectivist culture is supported.

In the present study the vocal portrayals were intoned by two speakers (a male and a female). A larger number of speakers will exclude the probable effects of speaker-specific idiosyncrasies and likely artifacts (Pell, 2002). Moreover, having more speakers of both genders also make it possible to investigate the relationship between the gender of the speakers and that of the participants in the recognition of emotional speech. In addition, replication of the present study in other collectivist and individualist societies would also allow generalization of the findings.

In conclusion, results of the present study which is the first of its kind provides an understanding of human perception in the context of the recognition of emotional speech from the simultaneous presentation of information in prosodic and lexical channels in the context of a collectivist culture. In particular our results established that both, gender as a biologically rooted social mechanism and cultural factors modulate the recognition of emotional speech. More specifically, our results support the view that with regard to vocal emotions, females are more sensitive compared to males. This female advantage was found for all the five emotional categories examined. Additionally, it was revealed that members of a collectivist culture show higher sensitivity to vocal emotional cues compared to their individualist counterparts. This might indicate that cultures that center on group harmony (i.e., collectivist cultures), may thus promote higher default levels of emotional sensitivity. Taken together, our study enlighten our knowledge regarding how cultures exploit the biological differences to different degree and in a different way. Comprehending how emotion judgments are directed by cultural dimensions (i.e., individualism-collectivism) has important implications for cross-cultural communication in various domains, such as education, business and conflict resolution (Thompson & Balkwill, 2006).

Chapter 5: General Discussion

5.1 Introduction

The purpose of the present dissertation was threefold. First, we generated the first comprehensive emotional speech database for Persian (Persian ESD). To date, numerous databases of vocal expressions of the basic emotions have been established in several languages including English, German, and many other languages. However, our attempt was the first of this kind for Persian. Second, we investigated the role of gender[1] in the recognition of emotional prosody in a syllable-timed language, namely Persian as spoken in Iran. During vocal communications, in conjunction with linguistic decoding, listeners attend to changes in not only in pitch, loudness and voice quality, but also to changes in rhythm, to form an impression about the speaker's emotional state (Wilson & Wharton 2006). Despite the large amount of research on the recognition of emotional prosody in stress-timed languages (e.g., German, and English), so far no such study had been performed on a syllable-timed language (e.g., Persian, Turkish). Therefore, the probable effect of the rhythmic division of time (i.e., stress- vs. syllable-timed) on the recognition of vocal expressions of emotion (emotional prosody) was still an open question. By examining the role of gender in recognition of emotional prosody in a syllable-timed language (i. e., Persian), we addressed this gap in the literature. Third, we examined the potential role of gender in the recognition of emotional speech in an under investigated cultural context (i.e., a collectivist society). So far, the only existing study on gender and emotional speech recognition was carried out on German speakers (i.e., members of an individualistic society). Given that emotional processes are largely influenced by culture, and that socio-cultural variables mediate and moderate gender differences in emotional functioning, there was a gap in the literature of gender and emotional speech. Thus, in the third study we addressed this paucity of research and investigated the potential role of gender in the recognition of emotional speech in an under investigated cultural context.

5.2 Summary of the results

In chapter 2, we have reported the process of designing, compiling, and evaluating Persian ESD, a comprehensive emotional speech database for colloquial Persian. To date, numerous databases of vocal

[1] It should be noted that although often the concept of gender is regarded as a dichotomy with two distinct categories (i.e., male/female; man/woman), various researchers argue that gender is a linear spectrum ranging from 100% female to 100% male including various states in between (Bockting, 2008). Yet as a point of departure, in this dissertation we focused on the binary classification of gender.

expressions of the basic emotions have been established in several languages including English (Cowie & Cornelius, 2003; Petrushin, 1999), German (Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005), Chinese (Liu & Pell, 2012; Yu, Chang, Xu, & Shum, 2001), Japanese (Niimi, Kasamatsu, Nishinoto, & Araki, 2001), Russian (Makarova & Petrushin; 2002), as well as many other languages (for reviews, see Douglas-Cowie, Campbell, Cowie, & Roach, 2003; Juslin & Laukka, 2003; Ververidis & Kotropoulos, 2003). However, this has not been achieved in Persian. Generating Persian ESD, we established a useful language resource for conducting basic research on a range of vocal emotions in Persian, as well as for conducting cross-cultural/cross-linguistic studies of vocal emotion communication in various fields including gender studies, clinical and neuroimaging studies, as well as speech synthesis research. Moreover, the stimuli in this database can be used as a reliable language resource (lexical and vocal) for the assessment and rehabilitation of communication skills in patients with brain injuries.

The existing literature on emotional speech has usually emphasized the role of prosody and neglected the role of lexical content (Ben-David *et al*., 2011). In studying emotional speech, various researchers have often prepared their own, study-specific lists of sentences without validating the emotional lexical content (see, e.g., Luo, Fu, & Galvin, 2007; Maurage, Joassin, Philippot, & Campanella, 2007). However, to conduct a study on emotional speech, a set of validated sentences is required to separate the impact of lexical content from prosody on the processing of emotional speech (Ben-David *et al.,* 2011). In this study, we performed three experiments to generate a set of validated sentences (lexical material) in colloquial Persian. The outcome of these three experiments is a set of 90 validated novel Persian sentences classified in five basic emotional categories (anger, disgust, fear, happiness, and sadness), as well as a neutral category.

One of the greatest challenges in emotion and speech research is obtaining authentic vocal data. Researchers have developed a number of strategies to obtain recordings of emotional speech, each with their own merits and shortcomings (Campbell, 2000). Generating the audio files, we employed one of the oldest, and still most frequently used, approaches for obtaining emotional speech data: *acted emotions* (Banse & Scherer, 1996). The advantages of this approach are: control over the verbal and prosodic content (i.e., all of the intended emotional categories can be produced using the same lexical content), a number of speakers could be employed to utter the same set of verbal content in all intended emotions, and high-quality recordings can be produced in an anechoic chamber. This production strategy allows direct comparison of acoustic and prosodic realizations for the various intended emotions portrayed. Critics of the acted emotions approach question the authenticity of the actors' portrayals. However, this drawback was minimized by employing the *Stanislavski method*. The validated sentences were then

articulated by two semi-professional Persian speaking actors (one male, one female) in three conditions: (1) congruent (emotional lexical content articulated in a congruent emotional voice), (2) incongruent (neutral sentences articulated in an emotional voice), and (3) baseline (all emotional and neutral sentences articulated in neutral voice). Having these three conditions allow researchers to separately investigate the role of lexicon and prosody in the recognition of vocal emotions.

Next, the validity of the produced speech materials (i.e., 472 vocal sentences) was evaluated by a group of 34 native speakers in a perceptual study. Previous work on emotion recognition (Scherer, Banse, Wallbott, & Goldbeck, 1991) suggest that emotional portrayals are recognized approximately at four times chance performance. Accordingly, to develop the best possible exemplars, a minimum of five times chance performance in emotion recognition task—that is, 71.4 %—was set as the cutoff level in this study. Utterances recognized better than five times chance performance were regarded as valid portrayals of the target emotions. Finally, acoustic analyses were performed to determine whether the vocal portrayals would show obvious differences in acoustic parameters that might help listeners to distinguish the intended emotions correctly. A total of 468 vocal utterances (all of the validated emotional and neutral portrayals), encompassing the congruent, incongruent, and baseline conditions, were included in this analysis. These vocal utterances were analyzed using the Praat speech analysis software (Boersma & Weenink, 2006). Results of the acoustic analysis revealed differences in pitch, intensity, and duration, attributes that may help listeners to correctly classify the intended emotion.

In sum, Persian emotional speech database (Persian ESD) encompasses a meaningful set of validated lexical (90 items) and vocal (468 utterances) stimuli, conveying five emotional meanings. Since the database covers the three conditions of (a) congruent, (b) incongruent, and (c) baseline, it provides the unique possibility to separately identify the effect of prosody and lexical content on the identification of emotions in speech. This database is generated to be used as a reliable material source (for both text and speech) in future cross-cultural or cross-linguistic studies of emotional speech. Researchers from several scientific fields may find our database useful, but it is primarily aimed at psychologists, neuroscientists, linguists, and computer scientists.

In Chapter 3, we tackled the question of gender effects on the recognition of emotional prosody in an under-investigated language. So far, the existing literature on gender and emotional prosody has targeted speakers of two specific languages, i.e., German and English, representing stress-timed languages. However, since prosody is affected by language-specific features such as the rhythmic division of time (i.e., stress- vs. syllable-timed), we questioned the applicability of the results of the existing studies to languages with a different rhythmic division of time i.e., Persian, representing syllable-timed languages.

To date a considerable amount of research on gender differences in recognition of non-verbal (auditory, visual, audio-visual) displays of emotions has been conducted. The findings of these studies have been summarized in four different meta-analyses by Hall (1978, 1984), McClure (2000) and Thompson and Voyer (2014). Overall, the results of these meta-analyses indicates that there is a small but consistent female superiority in recognizing non-verbal emotional displays. Nevertheless, studies investigating only the recognition of vocal emotions have often reported conflicting results (Brody & Hall, 2008; Lima *et al.,* 2013). Taking into account that emotional prosody is embedded in language and is affected by the linguistic features of the language under investigation (Pell, 2001), specific linguistic features of the language under investigation could be among the causes of the inconsistent results. So far, the role of gender in the recognition of vocal emotions has been only explored with native speakers of two stress-timed languages (e.g., English and German). But, no such study has been performed on a syllable-timed language (e.g., Persian, Turkish, and French). Thus, we addressed this paucity of research and investigated the potential role of gender in the recognition of emotional prosody in a syllable-timed language, namely Persian as spoken in Iran. On the basis of evidence from the existing literature, we hypothesized that gender would affect the recognition of emotional prosody in Persian, and in particular that there would be a female advantage in the recognition of emotional prosody.

Using a vocal emotion recognition task, male and female participants were asked to recognize the emotional prosody of a set of Persian emotional vocal portrayals. The stimulus material consisted of 12 syntactically similar sentences which were neutral in lexical content intoned in the five intended emotions (anger, disgust, fear, happiness, sadness). Seventy-two utterances were used in total: 12 tokens per emotional category, plus 12 neutral tokens (which were treated as filler items). Consistent with most of the previous literature (e.g., Paulmann *et al.,* 2008; Sauter *et al*., 2013), we employed a forced-choice response paradigm.

Averaging across emotions, an overall accuracy rate of 90 % was found. This rate is approximately five times higher than chance level (with six response options, chance level was 16.6%). As predicted, the effect likelihood ratio test showed a significant main effect of participant gender (i.e., in favor of females) $\chi^2 = 28.45$, $df = 1$, $p < .001$, and of emotion category, $\chi^2 = 61.82$, $df = 4$, $p < .001$. The direction of this effect is in line with general findings on emotion recognition (e.g., Briton & Hall, 1995; Hall *et al*., 2000; Schirmer & Kotz, 2003; Thompson & Voyer 2014) and in particular, with findings on the recognition of emotional prosody (e.g., Bonebright *et al.,* 1996; Schirmer *et al.,* 2002; Schirmer, Zysset, Kotz, & von Cramon, 2004; Schirmer *et al.,* 2005). This conclusion, therefore, coincides with a number of behavioral studies reporting a higher sensitivity in females than in males in a range of conceptually comparable tasks relating to the

decoding of nonverbal emotional cues (for a review see Thompson & Voyer 2014; Hall, 1978; Hall *et al.*, 2000). It is likewise consistent with gender role stereotypes (Broverman *et al.,* 1972). Given these reports, it seems that the present finding of female advantage in the recognition of vocal emotions was not a chance observation. One reason females might be more sensitive to emotional prosody than males is that females are more socially oriented than males and are more likely to define themselves in relational terms (Kashima *et al.*, 2004). Besides this, in the context of social accounts of gender differences, these findings may be regarded as evidence of the point that gender differences in social orientation are biologically rooted (Kashima *et al.,* 1995), for instance by hormones such as estrogen and oxytocin (Babcock & Laschever, 2009; Wood & Eagly, 2002; Taylor *et al*., 2000). Additionally, these results further revealed that recognition accuracy rates did not differ significantly between the two genders as a function of emotional category i.e., females performed better than males in recognizing all of the five intended emotions. These findings are, thus, in line with those of Hall *et al*. (2000) and Thompson and Voyer (2014). Furthermore, results of this study with Persian (i.e., a syllable-timed language) were in line with those from English and German, indicating that that the observed gender differences are relatively independent of rhythmic division of time (i.e., stress- vs. syllable-timed).

In Chapter 4, we reported our investigations on Gender effects on the recognition of emotional speech in an under-investigated culture. Despite the point that emotions are indispensable in our lives, we do not experience emotions in the same way. From the view point of cross-cultural psychology and anthropology, the exclusively biological perspective is not adequate for a comprehensive understanding of the perception and expression of emotions (Matsumoto & Hwang, 2011; Eid & Diener, 2001; Mesquita, 2001; Matsumoto, 2001). Psychological studies suggest that emotions and emotional experiences are largely influenced by culture (Matsumoto & Hwang, 2011; Tsai, Knutson, & Fung, 2006; Mesquita, 2003; Matsumoto, 2001; Wierzbicka, 1994; Markus & Kitayama, 1991; Matsumoto, 1989). Namely, how emotions are perceived, expressed, and regulated differs as a function of cultural norms of each society (Tanaka *et al*., 2010; Tsai *et al*., 2006; Matsumoto, 1989). Further research (Matsumoto, 1992; Ekman, 1972) has documented that cultural norms regulate the display of emotion in cases in which display of specific emotions might be potentially disruptive to social interaction, as well as the decoding of emotions when recognizing theses emotions would disrupt social interaction. Accordingly, cultural variation in the accuracy of emotion recognition is attributed to the norms that cultures impose on their members as for recognition and expression of specific emotions.

To study emotions within cultural context, Matsumoto (1989) argues that it is essential to make use of stable and meaningful dimensions of cultural variability[1] proposed by Hofstede (1980). Matsumoto (1990) and Matsumoto *et al.* (1998) further note that from among the dimensions[2] of cultural variability, *individualism-collectivism* which is conceived as two poles of one dimension, is a meaningful predictor of cultural variations in emotional norms (i.e., display rules). Individualist cultures regard emotions as essential personal experiences whose expression is any individual's right. In these cultural contexts, individuals tend to consider emotions as inner states vented spontaneously (Matsumoto, *et al.,* 1998). According to Mesquita (2001) in individualist cultures emotions are assumed to amplify and underline a subjective self, therefore these cultures reinforce outward displays of emotion (Matsumoto, *et al.,* 1998). In collectivist cultures, however, emotions are regarded as interactive experiences which mirror the social context rather than the individual's inner self. As Mesquita (2001) argues, in such cultural contexts expression of emotion is controlled, due to the point that it is grounded in evaluation of the relationship between the self and the others. In other words, in collectivist cultures emotions are expected to emphasize and reproduce the self in relation to others and the world (Mesquita, 2001). This issue implies that emotions are conceived as situation-specific clues about the relationship between the individuals. By and large, collectivist cultures reinforce the perpetuation of cohesion within the group and accordingly, control of emotions gains high priority (Potter, 1988). Similarly, Hall (1976) argues that to express emotions in vocal communication, members of Asian cultures (i.e., collectivist cultures) express less explicit verbal information and rely more on contextual cues to interpret the emotional state of the speaker. Likewise, Markus and Kitayama (1991) note that, collectivistic cultures (i.e., Asian cultures) are more context-dependent than individualistic cultures. Furthermore, due to the very restrictive emotional display rules within Asian cultures, the contextual cues provided by the interaction partners in these cultures are very subtle (Matsumoto, Kasri, & Kooken, 1999; Matsumoto, 1990). Consequently, to avoid a breakdown in interpersonal communication and maintain interpersonal harmony in such cultural contexts great sensitivity is needed for in the perception and interpretation of the implicitly conveyed emotional cues (Safdar *et al.,* 2009; Matsumoto *et al.,* 1999). Schirmer *et al.,* (2006) believe that this high sensitivity

[1] *Cultural dimensions theory* developed by Hofstede (1980) is a framework for cross-cultural communication. Using a structure *emanated from* factor analysis, the theory explains the influence of a society's culture on the values of its members, and that how these values relate to their behavior. In this context *dimension* refers to an aspect of a culture that can be measured relative to other cultures. Cultural dimensions theory is widely used in various fields (e.g., cross-cultural psychology) as a paradigm for research (Matsumoto, 1990). For more explanation on Cultural dimensions theory, see Hofstede (1983; 1991).

[2] Dimensions of cultural variability include a) Individualism-Collectivism, b) Power Distance, c) Uncertainty Avoidance, d) Masculinity-Femininity, and e) Long Term Orientation-Short Term Normative Orientation (Hofstede, 2001).

in the perception of implicit emotional vocal cues may be absent in members of Western cultures. If true, this would suggest that culture also modulates the accuracy rates in the recognition of emotional speech.

Most researchers investigating the role of gender in emotion recognition have focused on the domain of emotional prosody (e.g., Fecteau, Armony, Joanette, & Belin, 2005; Toivanen, Väyrynen & Seppänen, 2005). Compared to the literature on the recognition of emotional prosody (i.e., non-emotional lexical content intoned in emotional voice), very little is known about the role of gender in the recognition of emotional speech, where emotional lexical content is articulated in a congruent emotional voice (Paulmann *et al.,* 2008). Paulmann and colleagues (2008) were the only researchers who exclusively tackled this question in a behavioral study. As a result, the existing literature on gender and emotional speech (i.e., Paulmann *et al*., 2008) has only targeted speakers of a specific cultural background (i.e., German speakers living Leipzig, Germany). According to the Hofstede Model of Cultural Dimensions (Hofstede, 2001), Germany is considered as an individualistic society (see Hofstede, 2001 & The Hofstede Center, n.d. for more details). Research (Safdar *et al.*,2009; Greenfield *et al.*, 2003) has documented that the Hofstede ranking on individualism-collectivism (Hofstede, 2001) represents extensive distillations of various socio-psychological phenomena (encompassing social beliefs, values, norms, personality and behavioral characteristics) that pertain to ways in which human-environment relations are assessed. Considering that emotional processes are largely influenced by culture (Markus & Kitayama, 1991), and that socio-cultural variables mediate and moderate gender differences in emotional functioning (Brody & Hall, 2008), for a full understanding of emotional processes we should consider the cultural context where emotions are perceived, expressed, and experienced (Richeson & Boyd, 2005; Mesquita & Ellsworth, 2001). Therefore, results of the existing study on a sample from an individualist culture (i.e., German participants living in Leipzig) might not apply to members of collectivist cultures. Hence, in the third study we addressed this gap in the literature by investigating gender effects in the recognition of emotional speech in a collectivist society.

On the basis of evidence from the existing literature, we suggested that both gender and culture influence the recognition of emotional speech. Therefore, we hypothesized that 1) gender affects the recognition of emotional speech, and in particular there will be a female advantage in the recognition of emotional speech and, 2) compared to the members of an individualist society, members of a collectivist culture show higher sensitivity to vocal emotional cues. In a vocal emotion recognition task, 70 male and female speakers of Persian (regarded as members of a collectivist culture) were asked to recognize the emotional prosody of a set of vocal portrayals (sentences with emotional lexical content articulated in a congruent emotional voice). The stimulus material selected for this experiment consisted of 72 vocal

utterances (i.e., 12 sentences in each of the five intended emotional category plus 12 sentences in the neutral category). These vocal utterances were generated based on a syntactically similar lexical content (i.e., subject + object + prepositional phrase + verb) intoned by a male and a female Persian speaking actors. These vocal utterances were previously matched based on the emotional intensity of their lexical content.

Overall, emotional speech recognition rate was well above chance level (having six response option, chance level was 16.6 %). As predicted, the effect likelihood ratio test showed significant main effect for gender of the decoders (i.e., chi-square (1) = 38.56, $p < .001$) in favor of females in the recognition of emotional speech. The direction of this effect is in line with general findings on emotion recognition (e.g., Briton & Hall, 1995; Hall *et al.,* 2000; Schirmer & Kotz, 2003) and with particular findings on recognition of emotional prosody (e.g., Bonebright *et al*., 1996; Schirmer *et al*., 2002, 2006; Schirmer & Kotz 2003). These findings are thus consistent with Hall and colleagues' (2000) assumption participants might be partly due to the coping strategies employed by Iranian women to deal with the gender discrimination. In Iran, where the culture is of masculine nature (Kar, 2002), women are usually under pressure due to gender inequality in relation to different individuals (i.e., their husband, male members of the extended family, and male members of the society) where their status is perceived as being lower than their male counterparts (Keyvanara & Haghshenas, 2010; Mehran, 2009; Najmabadi, 2005). Likewise, the expectations and responsibilities assumed to women by the cultural norms of the country leaves women with less control over their own lives compared with men, and limits their choice (Keyvanara & Haghshenas, 2010; Moghadam, 2010). Despite of the high frequency of these stressors (i.e. gender discriminatory behaviors originated from the masculine nature of the culture) Iranian women's aspirations for less social limitation, greater freedom, higher education, better work opportunities, and generally less gender inequality, have led them to employ effective coping strategies (Mehran, 2009; Najmabadi that females are reliably more accurate in a range of conceptually comparable tasks of decoding nonverbal emotional cues.

Besides the biological reasons behind the mentioned gender difference, and the social factors (i.e., socio-culturally constructed roles that a given society considers appropriate for men and women), the female advantage in Iranian participants might be partly due to the coping strategies employed by Iranian women to deal with the gender discrimination. In Iran, where the culture is of masculine nature (Kar, 2002), women are usually under pressure due to gender inequality in relation to different individuals (i.e., their husband, male members of the extended family, and male members of the society) where their status is perceived as being lower than their male counterparts (Keyvanara &

Haghshenas, 2010; Mehran, 2009; Najmabadi, 2005). Likewise, the expectations and responsibilities assumed to women by the cultural norms of the country leaves women with less control over their own lives compared with men, and limits their choice (Keyvanara & Haghshenas, 2010; Moghadam, 2010). Despite of the high frequency of these stressors (i.e. gender discriminatory behaviors originated from the masculine nature of the culture) Iranian women's aspirations for less social limitation, greater freedom, higher education, better work opportunities, and generally less gender inequality, have led them to employ effective coping strategies (Mehran, 2009; Najmabadi, 2005, 1998) to deal with cultural barriers (Mozaffarian & Jamali, 2008). Among these coping strategies, developing *interpersonal skills* (i.e., verbal communication, non-verbal communication, listening skills, negotiation, problem solving, decision making, and assertiveness) is reported to be one of the most effective ways Iranian women have employed to manage the gender inequality (Moghadam, 2010; Sullivan, 2010; Kar, 2002). In fact recent investigations have revealed that Iranian women show higher interpersonal skills compared to Iranian men. Therefore, the female advantage among the Iranian participants in the vocal emotion recognition task might be partly due to their well-developed interpersonal skills (mainly verbal and non-verbal communication skills). These points together imply that, gender as a biologically rooted social mechanism may underlie the observed results. In addition, results of this study further revealed that recognition accuracy rates did not differ significantly as a function of emotional category i.e., females performed better than males in recognizing all of the five intended emotions.

Averaging across emotions, an overall accuracy rate of 95 % was found (i.e., very close to six times chance performance). In General these results are in line with Safdar *et al.* (2009) and Matsumoto *et al.* (1999) findings with regard to the great sensitivity of the members of collectivist cultures as for the perception and interpretation of the emotional cues. To maintain interpersonal harmony and to avoid breakdown in communication, this great sensitivity is a vital part of vocal communication in collectivist cultures (Safdar *et al.*, 2009). In particular, these results support Ahmadi and Ahmadi (1998) and Beeman (1976) arguments regarding Iranians' high sensitivity to others' emotions. In Iran (like in other collectivist societies) there is a tight link between the good of oneself and that of others, implying that people are highly related to each other (Ahmadi & Ahmadi, 1998). As relatedness between people and saving one's face within a group is highly promoted in Iran (Beeman, 1986), relationships (e.g., employer/employee, teacher/student, parents/children) are regarded as moral terms similar to family link and a major concern for those living in such cultures is being perceived as qualified for relationships (Mesquita, 2000). Taken together, these features make Iranians highly attentive to others' emotions and attitudes (Ahmadi & Ahmadi, 1998; Beeman, 1976). Considering that in a similar emotion recognition task, Paulmann *et al.*

(2008) examining German speakers living in an individualistic society, have reported an overall accuracy rate of 70% (almost close to five times chance performance), results of the current study indicated a higher sensitivity to emotional vocal emotional cues in the members of a collectivist culture. These findings are, therefore, consistent with Schirmer *et al.,* (2006) argument regarding the lower sensitivity in the perception of emotional vocal cues in members of individualist cultures.

Accordingly, the second hypothesis concerning the higher sensitivity to vocal emotional cues by the members of a collectivist culture was supported. This factor is important to consider in situations where culture plays an important role in human-human or computer-human interaction. Comprehending how emotion judgments are directed by cultural dimensions (i.e., individualism-collectivism) can prove useful for efficient collaboration in a multicultural environment in various domains, such as education, business and conflict resolution (Thompson & Balkwill, 2006).

5.3 Limitations

Although the findings reported here broaden our understanding of gender and culture effects on the recognition of vocal emotions, there are a number of limitations to the studies of this dissertation that should be acknowledged. Regarding study one, having only two speakers (one male, 50 years old; one female 49 years old) as encoders makes it difficult to gauge the extent of inter-speaker variability (Douglas-Cowie *et al*., 2003). This might have led to greater variability in acoustic and perceptual measures. Two reasons account for having only two speakers in this study: time and financial constraints. It already took a long time to find and prepare these two Persian speaking, semi-professional actors in Germany. Prior to recording, each speaker had four practice sessions (each session took almost three hours), starting with a review and discussion of the literal and figurative meanings of a given emotion, it ranges, and the ways it could be portrayed in speech, continuing with going through the standardized emotion portrayal instructions which were prepared based on a scenario approach. Besides, each of these speakers were separately recorded in six sessions in a professional recording studio (one emotion per session). Moreover, the costs of having recording sessions in a professional recording studio under the supervision of an acoustic engineer are very high. Another potential problem of the first study is that a small number of decoders (i.e., 34) were recruited for the perceptual validation study, aiming to select the best possible exemplars. A larger number of decoders would likely improve the reliability of the results. Again, time and financial constraints were the reasons behind this limitation. Finally, following prior research (Paulmann, *et al*., 2008) only a small number of acoustic parameters (i.e., pitch, intensity, and duration) were taken into account in the acoustic analysis. Having a larger set of acoustic parameters

(e.g., timber and energy) would provide more information about the acoustic features of Persian emotional speech.

For the second and third study, one can also argue that having only two speakers (one speaker of each gender) will result in the probable effects of speaker-specific idiosyncrasies and likely artifacts. A larger number of speakers will exclude these probable artifacts (Pell, 2002). Moreover, having more speakers of both genders also make it possible to investigate the relationship between the gender of the speakers and that of the participants in the recognition of vocal emotions. As explained above, this was due to time and financial constraints. Another limitation in study two and three concerns the use of the forced-choice response format. Following previous studies (e.g., Paulmann *et al.*, 2008; Sauter *et al.*, 2013), we asked participants to make a forced choice between anger, disgust, fear, happiness, sadness, and an extra response option (i.e., neutral) when determining the emotional prosody of an utterance. On one hand, forcing participants to choose an option from a short list of emotions may inflate agreement scores and produce artifacts (for detailed discussions on limitations of forced-choice response paradigms see, e.g., Russell, 1993, 1994; Wagner, 1993; Scherer, 2003). On the other hand, providing participants with more options or allowing them to label the emotions freely would result in very high variability (Banse & Scherer, 1996; Russell, 1994). However, if participants are provided with an extra response option (e.g., none) together with a discrete number of emotion choices, some of the artifacts can be avoided (Frank & Stennett, 2001). Therefore, in future studies it is optimal to use this extra response option to elude probable artificial high recognition rates.

An additional point to mention about both study two and three, concerns generalization of the findings: Only one syllable-timed language (i.e., Persian) in study two, and a sample group from one collectivist society in study three were examined. Accordingly, only replication of the present studies in other syllable-timed languages (e.g., Turkish, Spanish, and French)/collectivist societies would allow us to draw more far-reaching conclusions.

Finally, although collectivism and individualism characterize cultural groups (Markus, Mullally, & Kitayama, 1997), and within a collectivist or individualist context, emotions are patterned in discernible ways across individuals (Mesquita, 2001) not every individual in a given cultural context engages in the same ideas and practices, nor does he/she engage in them in exactly similar manners (Markus, Mullally, & Kitayama, 1997). Therefore, the point that in the third study we did not check the cultural orientation of the participants one by one can be considered as a limitation. Time and financial constraints were the reasons behind this limitation.

5.4 Implications of the findings and future directions

Our work has laid solid foundation for further research in several directions. First, with the establishment of the first validated emotional speech database for Persian, we have provided researchers with the unique possibility to separately identify the effect of prosody and lexical content on the identification of emotions in speech in Persian. The database can prove to be a valuable language resource for future investigations of emotion recognition in the fields of linguistics, cognitive psychology, as well as neuroscience. Additionally, this database can open up new opportunities for future investigations in artificial intelligence, speech synthesis, as well as in gender studies.

Another fascinating direction of further research would be development of interactive computer systems that recognize and synthesize vocalizations that vary with emotional state and gender. Findings of the second and third study can be useful for the development of such interactive computer systems. In recent years, the role of the vocal expression of emotions is gaining increasing importance in the computer speech research community, for instance in the context of automatic classification of emotional content from speech (ten Bosch, 2003). In such research context, human emotion recognition performance level (e.g., study 2 and 3 of this dissertation) is of great importance while it is regarded as the base-line data (Toivanen, Väyrynen, & Seppänen, 2005).

Obviously, within this dissertation, not all aspects of emotional prosody recognition could be investigated. Nevertheless, the current findings have shown that it is a worth wile endeavor to try to disentangle the potential factors that might influence emotional prosody processing. Future studies will hopefully investigate other factors (e.g., intensity) which might contribute to the recognition of emotional prosody.

The current dissertation investigated the influence of Persian speaking participant's gender on the recognition of emotional prosody and emotional speech. When group differences in emotional perception are observed, it is often unclear whether they arise from personality traits, cultural or other factors (Mesquita & Karasawa, 2002). Therefore, to gain a more precise conclusion, it will be interesting to repeat these investigations while checking each participant's personality traits (using validated scales such as the one introduced by Gosling, Rentfrow, & Swann, 2003) as well as cultural and religious orientation (using cultural and religiosity scales).

Besides, it would be interesting to design a similar study with its stimulus set in two different languages (e.g., German and Persian) where each participant attends the stimuli of his/her mother

tongue, and then compare the results of the male and female participants within each language as well as across the two languages under investigation.

Finally, results of the third study enlightened the role of cultural dimensions (i.e., individualism-collectivism) in recognition of emotional speech. Comprehending how emotion judgments are directed by cultural dimensions has important implications for cross-cultural communication in various domains, such as education, business and conflict resolution (Thompson & Balkwill, 2006).

5.5 Final remarks

As communication of emotions is vital for human social interaction (Ekman, 1992), scientific study of emotion has been a long-standing subject of research in many fields of science. Psychologists, linguists, anthropologists and neuroscientists have started to devote increasing attention to study of emotions and their effects on human behavior and society (de Gelder *et al.,* 2010). Accordingly, researchers investigating emotions from various perspectives (e.g., Lima *et al.*, 2014), call for further investigations on the existing gaps on emotion literature. Therefore, expanding the existing perspectives on emotion research by focusing on gender effects on the recognition of vocal emotions within an under investigated cultural and linguistic context was the specific aim of this dissertation.

To this end, theories and methods from the fields of psychology, linguistics and anthropology were combined to illuminate gender effects on the recognition of vocal emotions. A number of researchers (e.g., Van Rheenen & Rossell, 2013; Schirmer, Striano & Friederici, 2005; Schirmer & Kotz, 2003; Briton & Hall, 1995) suggest that gender can modulate emotions. While previous research has yielded a wealth of information regarding the role of gender in emotional processes, a number of essential factors influencing these processes have been neglected. *Language* and *culture* were instances of these overlooked factors.

Language is a crucial factor in transmission of emotions. Vocal expressions of emotion manifested in emotional prosody, are inherently dynamic and are integrated with linguistic properties of the language under study. Additionally, linguistic-specific features of a language (e.g., rhythm) affect vocal expressions of emotion (i.e., emotional prosody). Despite the large amount of research on gender and the recognition of vocal emotions in stress-timed languages (e.g., English, German), no such study has been performed on a syllable-timed language (e.g., Persian, Turkish and French).

Besides language, culture is another influential factor in emotional processes. In fact, culture and emotion are intermingled with each other. On the one hand, emotions and emotional experiences are

largely influenced by culture. Namely, how emotions are perceived, expressed, and regulated differ as a function of cultural norms of each society. On the other hand, emotions reinforce the cultural themes of a given society. As such, in any given cultural context, behaviors and emotional experiences that fit with and enhance the goals of that specific culture are reinforced. Despite the influential role of culture in emotion processes, the existing studies on vocal emotions have only targeted members of individualist cultures, but no study has investigated the effect of gender on recognition of vocal emotions in a collectivist culture.

To add to the body of knowledge in the field of gender and emotional processes, in study two and three of this dissertation we considered these two neglected factors (i.e., language and culture) and examined the role of gender in the recognition of vocal emotions in a syllable-timed language (i.e. Persian) spoken in a collectivist society (i.e. Iran).

Notably, results of the second study revealed a significant effect of gender of the participants in favor of females in the recognition of emotional prosody. The direction of this effect is in line with general findings on emotion recognition (e.g., Briton & Hall, 1995; Hall *et al.,* 2000; Schirmer & Kotz, 2003; Thompson & Voyer 2014) and in particular, with findings on the recognition of emotional prosody (e.g., Bonebright *et al.,* 1996; Schirmer *et al.,* 2002; Schirmer *et al.,* 2005). Additionally, results of the second study with Persian (i.e., a syllable-timed language) are in line with those from English and German, indicating that that the observed gender differences are relatively independent of rhythmic division of time (i.e., stress- vs. syllable-timed).

Findings of the third study indicated that recognition accuracy rates differ significantly as a function of gender while listening to emotional speech where both the lexical content and the prosody portray a same emotional meaning. Namely, there was an overall recognition advantage for female participants over male participants. Besides, in line with Safdar *et al.* (2009) and Matsumoto *et al.* (1999) arguments regarding the great sensitivity of the members of collectivist cultures, our investigations further revealed very high overall accuracy rate of 95% (almost close to six times chance performance) in the recognition of emotional speech in Persian speakers living in a collectivist society (i.e., Iran). Considering that Paulmann *et al.* (2008) testing German speakers living in an individualist culture, have reported an overall accuracy rate of 70% (almost close to five times chance performance), our results indicated a higher sensitivity to emotional vocal emotional cues in the members of a collectivist culture. These findings, therefore, empirically support Schirmer *et al.,* (2006) argument regarding the lower sensitivity in the perception of emotional vocal cues in members of individualist cultures.

In sum, findings of the studies of this dissertation has provided concrete evidence for a) females higher sensitivity to vocal emotions (both emotional prosody and emotional speech) compared to males, b) independency of gender differences in recognition of vocal emotions to the nature of the rhythmic division of time in a language (i.e., stress- vs. syllable-timed), and c) a higher sensitivity to emotional vocal emotional cues in the members of a collectivist culture.

Obviously, within this dissertation not all aspects of vocal emotion recognition could be investigated. Nevertheless, the current findings have revealed that it is a worthwhile endeavor to disentangle the potential factors that might influence the recognition of vocal emotions. Particularly generating Persian Emotional Speech Database, a certain number of acoustic parameters (i.e., pitch, intensity and duration) were analyzed. In future studies it would be interesting to analyze the vocal utterances in terms of other acoustic features (e.g., timber and energy) as to what are the crucial properties of emotional speech in Persian that allow a listener to recognize a particular emotional category. This will help us to attain even a deeper understanding of the vocal expressions of emotions. Moreover, investigating various speaker identities (e.g., gender, age) can provide further insight into recognition of vocal emotions by listeners. Therefore, future research will hopefully continue on this issue. The current dissertation primarily investigated the influence of participant gender in an under investigated cultural context on the recognition of vocal emotions at sentence level, although, there are other factors that might contribute to vocal emotion recognition. For instance, participants' age, speaker identity, and context (e.g., single sentence as the stimuli vs. stories as stimuli) may affect vocal emotion recognition and need closer examination in future studies. Additionally, in real-life social interaction an emotional utterance is usually accompanied by multiple sensory inputs (e.g., visual and auditory cues) in different contexts, and these cues are essential for recognizing a speaker's emotional state (Brosch, Grandjean, Sander, & Scherer, 2009). Therefore, it is interesting to see how future studies can investigate the potential effect of these sensory inputs on vocal emotion recognition.

To conclude, the current research has further expanded our knowledge of the role of gender in the recognition of vocal emotions in a syllable-timed language and within a collectivist society. Meanwhile, these findings give rise to various questions awaiting to be addressed in future studies.

References

Matsumoto, D. (2006). Culture and nonverbal behavior. In *Handbook of nonverbal communication*. Thousand Oaks, CA: Sage Publications Inc.

Abercrombie, D. (1967). *Elements of general phonetics* (Vol. 203). Edinburgh: Edinburgh University Press.

Ahmadi, N. & Ahmadi, F. (1998). *Iranian Islam; The Concept of the Individual*. London: MacMillan.

Andersen, S. M., Reznik, I., & Chen, S. (1997). The self in relation to others: Cognitive and motivational underpinnings. In J. G. Snodgrass & R. L. Thompson (Eds.), *The self across psychology* (pp. 233-275). New York: Academy of Sciences.

Anvari, H., & Givi, H. (1996). *Persian grammar* (2 Volumes). Tehran, Iran: Fatemi.

Arndt, H., & Janney, R.W. (1991). Verbal, prosodic, and kinesic emotive contrasts in speech. *Journal of Pragmatics*, *15*, 521–549.

Babchuk, W. A., Hames, R. B., & Thompson, R. A. (1985). Sex differences in the recognition of infant facial expressions of emotion: The primary caretaker hypothesis. *Ethology and Sociobiology*, *6*(2), 89-101.

Babcock, L., & Laschever, S. (2009). *Women don't ask: Negotiation and the gender divide*. Princeton, NJ: Princeton University Press.

Baden, S., & Goetz, A. M. (1998). Who needs [sex] when you can have [gender]. *Feminist visions of development: gender analysis and policy*, 19-38.

Bagherian, M. (2007). Women employment and unemployment from the development perspective (In Persian). *Journal of Women studies*. 2, 45-53.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 614–636. doi:10.1037/0022-3514.70.3.614

Bardhan, K., & Klasen, S. (1999). *UNDP's Gender-Related Indices: A Critical Review, World Development.* 27:6, pp.985-1010.

Barrett, L. F., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74,* 967–984.

Braun-Lewensohn, O., Sagy, S., & Roth, G. (2011). Coping strategies as mediators of the relationship between sense of coherence and stress reactions: Israeli adolescents under missile attacks. *Anxiety, Stress, & Coping*, 24(3), 327-341.

Beeman, W.O. (1986). *Language, Status, and Power in Iran*. Bloomington, Indiana: Indiana University Press.

Beeman, W. O. (1976). Status, style and strategy in Iranian interaction. *Anthropological Linguistics*, 305-322.

Belk, S. S., & Snell, W. E. (1986). Beliefs about Women: Components and Correlates. *Personality and Social Psychology Bulletin*, *12*(4), 403-413.

Ben-David, B. M., van Lieshout, P.H., & Leszcz, T. (2011). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury*, 25,206–220.

Bertinetto, P. M. (1989). Reflections on the dichotomy 'stress' vs. 'syllable-timing'. *Revue de phonétique appliquée*, *91*(93), 99-130.

Besson, M., Magne, C., & Schön, D. (2002). Emotional prosody: sex differences in sensitivity to speech melody. *Trends in cognitive sciences*, *6*(10), 405-407.

Boersma, P., &Weenink, D. (2006). Praat: Doing phonetics by computer (Version 4.4.11) [Computer program]. Retrieved February 26, 2010, from www.praat.org

Bockting, W. O. (2008). Psychotherapy and the real-life experience: From gender dichotomy to gender diversity. *Sexologies*, *17*(4), 211-224.

Bonebright, T. L., Thompson, J. L., & Leger, D. W. (1996). Gender stereotypes in the expression and perception of vocal affect. *Sex Roles*, *34*(5-6), 429-445.

Breedlove, S. M., Watson, N. V., & Rosenzweig, M. R. (2010). *Biological psychology: An introduction to behavioral, cognitive, and clinical neuroscience* (6th ed.). Sunderland, MA: Sinauer Associates, Inc. Publishers.

Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, *32*,79–90.

Brody, L. R., & Hall, J. A. (2008). Gender and emotion in context. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 395-408). Cambridge, England: Cambridge University Press.

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2009). Cross-modal emotional attention: emotional voices modulate early stages of visual processing. *Journal of cognitive neuroscience*, *21*(9), 1670-1679.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-Role Stereotypes: A Current Appraisal1. *Journal of Social issues*, *28*(2), 59-78.

Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, *8*(1-2), 1-2.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005, September). *A database of*

German emotional speech. Paper presented at the 9th European Conference on Speech Communication and Technology, Lisbon, Portugal. Calder, J. (1998). Survey research methods. Medical Education, 32, 636– 652.

Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S. J. (2009). Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience*, *21*(9), 1736-1750. doi:10.1162/jocn.2009.21121

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 173–191). New York: Guilford Press.

Campbell, N. (2000, September). *Databases of emotional speech*. Paper presented at the ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, Newcastle, Northern Ireland, UK.

Carmen, R. A., Geher, G., Glass, D. J., Guitar, A. E., Grandis, T. L., Johnsen, L., Philip, M. M., Newmark, R. L., Trouton, G. T., & Tauber, B. R. (2013). Evolution integrated across all islands of the human behavioral archipelago: All psychology as Evolutionary Psychology. *EvoS Journal: The Journal of the Evolutionary Studies Consortium*, *5*(1), 108-126.

Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*, *42*, 74-81.

Ciompi, L. & Panksepp, J. (2005). Energetic effects of emotions on cognitions—complementary psychobiological and psychosocial findings. in *Consciousness & Emotion Book Series*, Vol 1, ed. R.D. Ellis & N. Newton, pp. 23–56. Amsterdam: John Benjamins.

Clancy, S. M., & Dollinger, S. J. (1993). Photographic depictions of the self: Gender and age differences in social connectedness. *Sex Roles*, *29*(7-8), 477-495.

Clark, J. E., Yallop, C., & Fletcher, J. (2007). Introduction to phonetics and phonology. Oxford: Blackwell.

Comrie, B. (ed.), (1987). *The world's Major Languages.* Kent: Croom Helm.

Consedine, N. S., Magai, C., & Bonanno, G. A. (2002). Moderators of the emotion inhibition-health relationship: A review and research agenda. *Review of General Psychology, 6*, 204-228.

Cornelius, R. R. (1996). *The science of emotion: Research and tradition in the psychology of emotions*. Prentice-Hall, Inc. Upper Saddle River.

Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication*, 40, 5–32.

Crystal, D. (1995). *The Cambridge Encyclopedia of the English Language*. Cambridge, UK: Cambridge University Press.

Darwin, C. (1872). *The expression of the emotions in man and animals*. Chicago: University of Chicago Press.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of phonetics*, 11, 51–62.

Davidson, R. J., Scherer, K. R., & Goldsmith, H. (2003). *Handbook of affective sciences*. Oxford: Oxford University Press.

Davidson, R. J., Scherer, K. R., & Goldsmith, H. (2003). *Handbook of affective sciences*. New York: Oxford University Press.

de Gelder, B., Van den Stock, J., Meeren, H. K., Sinke, C., Kret, M. E., & Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neuroscience & Biobehavioral Reviews*, *34*(4), 513-527.

Demir, E., Desmet, P., & Hekkert, P. (2009). Appraisal Patterns of Emotions in Human-Product Interaction. *International Journal of Design*, 3(2), 41-51.

Dollinger, S. J., Preston, L. A., O'Brien, S. P., & DiLalla, D. L. (1996). Individuality and relatedness of the self: An autophotographic study. *Journal of Personality and Social Psychology*, *71*(6), 1268-1278.

Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech Communication*, 40, 33–60.

Eid, M., & Diener, E. (2001). Norms for experiencing emotions in different cultures: inter-and intranational differences. *Journal of personality and social psychology*, *81*(5), 869.

Ekman, P. (Ed.). (2008). *Emotional awareness: Overcoming the obstacles to psychological balance and compassion*. Macmillan.

Ekman, P. (2007). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. Macmillan.

Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. London: Weidenfeld and Nicolson.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & T. Power (Eds.), *The handbook of cognition and emotion* (pp. 45–60). Hove, UK: Wiley.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169-200.

Ekman, P. (1973). Cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression* (pp. 169-222). New York: Academic Press.

Ekman, P. (1972). *Universals and cultural differences in facial expression of emotion*. In: Cole, J.R. (Ed.), Nebraska Symposium on Motivation. University of Nebraska Press, Lincoln, pp. 207–283.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364-370.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, 54, 414–420. doi:10.1037/0022-3514.54.3.414

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial cues*. Englewood Cliffs, NJ: Prentice-Hall.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1, 49-98.

Ember, C. R., & Ember, M. (Eds.). (2004). *Encyclopedia of Sex and Gender: Men and Women in the World's Cultures Topics and Cultures* AK-Volume 1; Cultures LZ. Springer Science & Business Media.

Fairclough, N. (2003). *Analyzing discourse: Textual analysis for social research*. London and New York: Routledge.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2005). Judgment of emotional nonlinguistic vocalizations: Age-related differences. *Applied Neuropsychology*, 12(1), 40–48.

Feder, J., & Brougher, C. (2013). Sexual Orientation and Gender Identity Discrimination in Employment: A Legal Analysis of the Employment Non-Discrimination Act (ENDA). Available (online): www.fas.org/sgp/crs/misc/R40934.pdf

Feldman, R. S., & Rimé, B. (Eds.). (1991). *Fundamentals of nonverbal behavior*. New York: Cambridge University Press, New York.

Folkman, S., & Lazarus, R. S. (1980). An analysis of coping in a middle-aged community sample. *Journal of health and social behavior*. 219-239.

Fougeron, C. E. C., & Jun, S. A. (1998). Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics*, *26*(1), 45-69.

Fraisse, P. (1982). Rhythm and timing. In Deutsch, D. (Ed.), *The psychology of music* (pp. 149-180). New York: Academic Press.

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology*, 80, 75–85. doi:10.1037/0022-3514.80.1.75

Frijda, N. H. (2007). *The laws of emotion.* Mahwah, NJ: Lawrence Erlbaum.

Frijda, N. H. (1986). *The emotions*. Cambridge, United Kingdom: Cambridge University Press.

Frijda, N. H., and Parrott, W. G. (2011). Basic emotions or ur-emotions? *Emotion Review.* 3, 406–415.

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American psychologist*, *40*(3), 266.

Gerrards-Hesse, A., Spies, K., & Hesse, F. W. (1994). Experimental inductions of emotional states and

their effectiveness: *A review. British Journal of Psychology,* 85, 55–78.

Gharavian, D., Sheikhan, M., Nazerieh, A., & Garoucy, S. (2012). Speech emotion recognition using FCBF feature selection method and GA optimized fuzzy ARTMAP neural network. *Neural Computing and Applications*, 21, 2115–2126.

Gharavian, D., & Sheikhan,M. (2010). Emotion recognition and emotion spotting improvement using formant-related features. *Majlesi Journal of Electrical Engineering*, 4(4).

Gharavian, D., & Ahadi, S. M. (2009). Emotional speech recognition and emotion identification in Farsi language. *Modares Technical and Engineering*, 34(13), 2.

Ghayoomi,M.,Momtazi, S., & Bijankhan, M. (2010). A study of corpus development for Persian. *International Journal on Asian Language Processing*, 20, 17–33.

Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech communication*, *40*(1), 189-212.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, *37*(6), 504-528.

Gray, J.A. (1990). Brain systems that mediate both emotion and cognition. *Cognition and Emotion,* 4 (3), pp. 269–88.

Greenfield, P., Keller, H., Fuligni, A., & Maynard, A. (2003). Cultural pathways through universal development. *Annual Review of Psychology, 54,* 461–490.

Grossman, M., & Wood, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology*, 65, 1010–1022.

Haig, D. (2004). The inexorable rise of gender and the decline of sex: Social change in academic titles, 1945–2001. *Archives of sexual behavior*, *33*(2), 87-96.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: Johns Hopkins University Press.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological bulletin*, *85*(4), 845-857.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201.

Hall, J. A., Carter, J. D., & Horgan, T. G. (2000). Gender differences in nonverbal communication of emotion. *Gender and emotion: Social psychological perspectives*, 97-117.

Hargreaves, W. A., Starkweather, J. A., & Blacker, K. H. (1965). Voice quality in depression. *Journal of Abnormal Psychology*, *70*(3), 218–220.

Harris, P. L., Olthof, T., Terwogt, M. M., & Hardman, C. E. (1987). Children's knowledge of the situations that provoke emotion. *International Journal of Behavioral Development*, *10*(3), 319-343. doi:10.1177/016502548701000304

Harper, R. G., Wiens, A. N., & Matarazzo, J. D. (1978). *Nonverbal communication: The state of the art*. New York: John Wiley & Sons.

Harris, P. L. (1989). *Children and emotion.* Oxford, England: Blackwell.

Hausmann, R., Tyson, L.D., Zahidi, S., (2006). The Global Gender Gap Report. World Economic Forum, Geneva, Switzerland.

Hawk, S. T., Van Kleef, G. A., Fischer, A. H., & Van der Schalk, J. (2009). Worth a thousand words: absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, *9*(3), 293.

Herlitz, A., & Lovén, J. (2009). Sex differences in cognitive functions. *Acta Psychologica Sinica*, *41*(11), 1081-1090.

Hess, U.; Thibault, P. (2009). "Darwin and Emotion Expression". *American Psychologist* 64 (2): 120–8. doi:10.1037/a0013386

Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., & Kleck, R. E. (2000). Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cognition & Emotion*, 14, 609–642.

Herzog, H. (1933). Stimme und Persönlichkeit [Voice and personality]. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane.* 130, 300–379.

Hoffman, M. L. (2001). *Empathy and moral development: Implications for caring and justice*. Cambridge: Cambridge University Press.

Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.), Thousand Oaks, CA: Sage.

Hofstede, G. (1991). *Cultures and Organizations: Software of the Mind.* London: McGraw Hill.

Hofstede. G. (1983). Dimensions of natural cultures in fifty countries and three regions. In Deregowski, J.B., Dziurawiec, S. and Annis, R.C. (Eds.), *Expiscations in cross-cultural psychology*, pp. 335-355. Lisse, The Netherlands: Swets & Zeitlinger.

Hofstede, G. (1980). *Culture's Consequences*. Beverly Hills: CA. Sage.

Hofstede, G., Hofstede, G., J., Minkov, M. (2010). Cultures and organizations: Software of the mind. *McGraw-Hill, New York*.

Hosmer, J., David W., Lemeshow, S., & Sturdivant, R. X. (2013). Introduction to the Logistic Regression Model. In *Applied Logistic Regression* (S. 1–33). John Wiley & Sons, Inc.

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone differences among three cultures. *Psychological Science*, *14*(1), 39-46.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science, 2(3), 260-280.*

Izard, C. E. (1993). Organizational and motivational functions of discrete emotions. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 631–641). New York: Guilford Press.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review. 99*, 561-565.

Izard, C.E. (1971). *The Face of Emotion*. New York: Appleton-Century-Crofts.

Jassem, W., Hill, D. R., & Witten, I. H. (1984). Isochrony in English speech: its statistical validity and linguistic relevance. *Intonation, Accent and Rhythm*, *8*, 203-225.

Jackson, S. (2006). Gender, Sexuality and Heterosexuality - The Complexity (and limits) of Heternormativity. *Feminist Theory* 7(2): 105-121.

Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry*, 43, 280–283. doi:10.1001/archpsyc.1986.01800030098011

Johnstone, T., & Scherer, K. R. (1999, August). The effects of emotions on voice quality. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 2029–2032). San Francisco, CA: University of California, Berkeley.

Joseph, S., & Najmabadi, A. (Eds.). (2003). *Encyclopedia of Women and Islamic Cultures: Family, Law and Politics*. Leiden, Netherlands: Brill Academic Publishers.

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

Kar. M. (2002). *Violence against women in Iran* (In Persian). Tehran, Iran: Roshangaran publications.

Kar. M. (1996). *Women in job market in Iran* (In Persian). Tehran, Iran: Roshangaran publications.

Karpf, A. (2006). The human voice. London: Bloomsbury.

Kashima, Y., Kokubo, T., Kashima, E. S., Boxall, D., Yamaguchi, S., & Macrae, K. (2004). Culture and self: Are there within-culture differences in self between metropolitan areas and regional cities? *Personality and Social Psychology Bulletin*, *30*(7), 816-823.

Kashima, Y., Yamaguchi, S., Kim, U., Choi, S. C., Gelfand, M. J., & Yuki, M. (1995). Culture, gender, and self: a perspective from individualism-collectivism research. *Journal of personality and social psychology*, *69*(5), 925-937.

Keltner, D., Oatley, K., & Jenkins, J. M. (2013). *Understanding emotions*. Wiley Global Education.

Keshtiari, N., Kuhlmann, M., Eslami, M., & Klann-Delius, G. (2014). Recognizing emotional speech in

Persian: A validated database of Persian emotional speech (Persian ESD). *Behavior research methods*, 1-20. doi: 10.3758/s13428-014-0467-x

Keyvanara, M., & Haghshenas, A. (2010). The sociocultural contexts of attempting suicide among women in Iran. Health care for women international, 31(9), 771-783.

Khoshnood, K., Hashemian, F., Moshtagh, N., Eftekahri, M., & Setayesh, S. (2008). T03-O-08 Social stigma, homosexuality and transsexuality in Iran. *Sexologies*, *17*, S69.

Kimura, D. (1999). *Sex and cognition*. Cambridge: MIT Press.

Kitayama, S., & Markus, H. R. (Eds.) (1995). *Emotion and culture: Empirical studies of mutual influence* (pp. 23-50). Washington, DC: American Psychological Association.

Klasen, S. (2006). UNDP's gender-related measures: some conceptual problems and possible solutions. *Journal of Human Development*, 7(2), 243-274. doi: 10.1080/14649880600768595

Klasen, S., & Schüler, D. (2011). Reforming the gender-related development index and the gender empowerment measure: Implementing some specific proposals. *Feminist Economics,* 17(1), 1-30.

Kleinginna Jr, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion*, *5*(4), 345-379.

Klugman, J., Kovacevic, M., Twigg, S., & Zambrano, E. (2010). Measuring key disparities in human development: The gender inequality index. *United Nations Development Programme.*

Knight, D.C., Smith, C.N., Cheng, D.T., Stein, E.A., & Helmstetter, F.J. (2004). Amygdala and hippocampal activity during acquisition and extinction of human fear conditioning. *Cognitive, Affective & Behavioral Neuroscience,* 4 (3), pp. 317–25.

Koolagudi, S. G., & Rao, K. S. (2012). Emotion recognition from speech: a review. *International Journal of Speech Technology*, *15*(2), 99-117.

Koopmann-Holm, B., & Matsumoto, D. (2011). Values and display rules for specific emotions. *Journal of Cross-Cultural Psychology*, *42*(3), 355-371.

Koprowska, J. (2014). *Communication and interpersonal skills in social work*. Learning Matters.

Krieger, N. (2000). *Discrimination and health*. In L. F. Berkman (Ed.). Social epidemiology. New York: Oxford University Press.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23*, 155–184.

LeDoux, J. E., & Phelps, E. A. (2008). Emotional networks in the brain. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 159-179). New York: The

Guilford Press.

LeDoux, J. E., & Phelps, E. A. (2008). Emotional networks in the brain. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), Handbook of emotions (3rd ed., pp. 159-179). New York: The Guilford.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of phonetics*. 5, 253–263.

Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychological Science, 3,* 23–27.

Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *The Journal of the Acoustical Society of America*, *34*(7), 922-927.

Likert, R. (1936). A method for measuring the sales influence of a radio program. *Journal of Applied Psychology*, 20, 175–182.

Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (2014). In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. *Emotion*, *14*(1), 145-160.

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: a corpus of nonverbal vocalizations for research on emotion processing. *Behavior research methods*, *45*(4), 1234-1245.

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021-1031.

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods*, 44, 1042–1051. doi:10.3758/s13428-012-0203-3

Lu, L. & Yang, K.-S. (2006). Emergence and composition of the traditional modern bicultural self of people in contemporary Taiwanese societies. *Asian Journal of Social Psychology,* 9(3):167–175.

Luo, X., Fu, Q. J., & Galvin, J. J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification*, 11, 301–315.

Mahl, G.F., & Schulze, G. (1964). Psychological research in the extra linguistic area. In: Sebeok, T.A., Hayes, A.S., Bateson, M.C. (Eds.), *Approaches to Semiotics*. The Hague: Mouton. pp. 51–124.

Mahootian, S. (1997) *Persian*. London: Routledge.

Makarova, V., & Petrushin, V. A. (2002, September). *RUSLANA: A database of Russian emotional utterances.* Paper presented at the International Conference of Spoken Language Processing, Colorado, USA.

Malatesta, C. Z., & Haviland, J. M. (1982). Learning display rules: The socialization of emotion expression in infancy. *Child development*, 991-1003.

Malti-Douglas, F. (2007). Encyclopedia of Sex and Gender: AC (Vol. 1). Macmillan Reference USA.

Markus, H. R., Mullally, P. R., & Kitayama, S. (1997). Selfways: Diversity in modes of cultural participation. In U. Neisser & D. Jopling (Eds.), *The conceptual self in context* (pp. 13-61). New York: Cambridge University Press.

Markus, H. R. E., & Kitayama, S. E. (1994). The cultural construction of self and emotion: Implications for social behavior. In Kitayama, S. E., & Markus, H. R. E. (Eds). *Emotion and culture: Empirical studies of mutual influence*. American Psychological Association.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224-253.

Matsumoto, D. (2007). Culture, context, and behavior. *Journal of personality*, *75*(6), 1285-1320.

Matsumoto, D. (2006). Culture and nonverbal behavior. In *Handbook of nonverbal communication*. Thousand Oaks, CA: Sage Publications Inc.

Matsumoto, D. E. (2001). *The handbook of culture and psychology*. Oxford: Oxford University Press.

Matsumoto, D. (1993). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation & Emotion, 17*, 107-123.

Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology*, *23*(1), 72-84.

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and emotion*, *14*(3), 195-214.

Matsumoto, D. (1989). Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*, *20*(1), 92-105.

Matsumoto, D., & Hwang, H. S. (2011). Culture and Emotion: The Integration of Biological and Cultural Contributions. *Journal of Cross-Cultural Psychology,* 43(1):91–118.

Matsumoto, D., Yoo, S. H., & Fontaine, J. (2008). Mapping expressive differences around the world the relationship between emotional display rules and individualism versus collectivism. *Journal of cross-cultural psychology*,*39* (1), 55-74.

Matsumoto, D., Grissom R. J., & Dinnel, D. L. (2001). Do between-culture differences really mean that people are different? A look at some measures of cultural effect size. *Journal of Cross-Cultural Psychology*, 32(4), 478-490.

Matsumoto, D., Kasri, F., & Kooken, K. (1999). American-Japanese cultural differences in judgments of expression intensity and subjective experience. *Cognition and Emotion, 13*, 201-218.

Matsumoto, D., Takeuchi, S., Andayani, S., Kouznetsova, N., & Krupp, D. (1998). The Contribution of Individualism vs. Collectivism to Cross-national Differences in Display Rules. *Asian Journal of Social Psychology*, *1*(2), 147-165.

Matsumoto, D., Kudoh, T., Scherer, K., & Wallbott, H. (1988). Antecedents of and reactions to emotions in the United States and Japan. *Journal of Cross-Cultural Psychology, 19*, 267-286.

Maurage, P., Joassin, F., Philippot, P., & Campanella, S. (2007). A validated battery of vocal emotional expressions. *Neuropsychological Trends*, 2, 63–74.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*, 424–453.

McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J., & Gross, J. J. (2008). Gender differences in emotion regulation: An fMRI study of cognitive reappraisal. *Group Processes & Intergroup Relations*, 11(2), 143-162.

Mehran, G. (2009). Doing and undoing gender: Female higher education in the Islamic Republic of Iran. *International Review of Education*, *55*(5-6), 541-559.

Mehran, G. (2003). Gender and education in Iran. *Paper commissioned for the EFA Global Monitoring Report*, *4*.

Mesquita, B. (2003). Emotions as dynamic cultural phenomena. In R. J. Davidson, K. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 871-890). New York: Oxford University Press.

Mesquita, B. (2001). Emotions in collectivist and individualist contexts. *Journal of personality and social psychology*, *80,* 68–74.

Mesquita, B., & Markus, H. R., (2004). Culture and emotion: models of agency as sources of cultural variation in emotion. In N. H. Frijda, A. S. R. Manstead, & A. H. Fischer (Eds.), *Feelings and emotions: The Amsterdam symposium*, Cambridge MA: Cambridge University Press.

Mesquita, B., & Karasawa, M. (2002). Different emotional lives. *Cognition & Emotion*, *16*(1), 127-141.

Mesquita, B., & Ellsworth, P. (2001). The role of culture in appraisal. In K.R. Scherer & A. Schorr (Eds.). *Appraisal processes in emotion: theory, methods, research* (pp.233-248). New York: Oxford University Press.

Mesquita, B. and Frijda, N. H. (1992). *Cultural variations in emotions: A review*. Psychological Bulletin 112, 197–204.

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2004). Neural response to emotional prosody in schizophrenia and in bipolar affective disorder. *British Journal of Psychiatry*, 184, 223–230.

Modarressi-Tehrani, Y. (2001). Aspects of sociolinguistics in Iran. *International journal of the Sociology of Language*, 148: 1-3.

Moghadam, A., V. (2010). Women and Employment in Iran (in Farsi). Retrieved January 16, 2015 from http://www.feministschool.org/spip.php?article5480After

Moghadam, V. M. (2003). *Modernizing women: Gender and social change in the Middle East.* Lynne

Rienner Publishers.

Moghissi, H. (2008). Islamic cultural nationalism and gender politics in Iran. *Third World Quarterly*, *29*(3), 541-554.

Mokhtari, P. (2013) Analysis of interpersonal skills among Iranian college students (In Persian). Master thesis. Islamic Azad University, Tehran Central Branch. Tehran, Iran.

Monrad-Kohn G. H (1947). The prosodic quality of speech and its disorders (a brief survey from a neurologist's point of view). *Acta Psychiatrica Scandinavica*, 22(3-4), 255–269.

Moors, A. (2013). On the causal role of appraisals in emotion. *Emotion review,* 5 (2) 132-140.

Moseley, C., & Asher, R. E. (Eds.). (1994). *Atlas of the world's languages.* London: Routledge.

Mozaffarian, M., & Jamali, H. R. (2008, September). Iranian women in science: a gender study of scientific productivity in an Islamic country. In *Aslib Proceedings* (Vol. 60, No. 5, pp. 463-473). Emerald Group Publishing Limited.

Mungai, E. N., & Ogot, M. (2012). Gender, culture and entrepreneurship in Kenya. *International Business Research,* 5 (5), 175-183.

Najmabadi, A. (2005). *Women with mustaches and Men without beards. Gender and Sexual Anxieties of Iranian Modernity*. University of California Press.

Najmabadi, A. (1998). Crafting an educated housewife in Iran. *Remaking women: Feminism and modernity in the Middle East*, 91-125.

Niimi, Y., Kasamatsu, M., Nishinoto, T., & Araki, M. (2001, August). *Synthesis of emotional speech using prosodically balanced VCV segments. Paper presented at the 4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis, Perthshire, Scotland.*

Nobelius, A. M. (2004). What is the difference between sex and gender? Retrieved October 16, 2014 from <*http://www. med. monash. edu. au/gendermed/sexandgender. html*> (Created 23 June 2004 Monash University).

O' Brien, N. (2011). *Stanislavski in practice: Exercises for students*. New York, NY: Routledge.

Oatley, K., & Jenkins, J. M. (1996). *Understanding emotions.* Oxford, England: Blackwell.

Oatley, K. (1992). *Best laid schemes: The psychology of the emotions*. Cambridge, UK: Cambridge University Press.

Oatley, K., & Johnson-Laird, P. N. (2014). Cognitive approaches to emotions. *Trends in cognitive sciences*, *18*(3), 134-140.

Oatley, K., Parrott, W. G., Smith, C., & Watts, F. (2011). Cognition and Emotion over twenty-five years. *Cognition & emotion*, *25*(8), 1341-1348.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and emotion*,

1(1), 29-50.

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature neuroscience*, *10*(9), 1095-1102.

Orbelo, D. M., Grim, M. A., Talbott, R. E., & Ross, E. D. (2005). Impaired comprehension of affective prosody in elderly subjects is not predicted by age-related hearing loss or age-related cognitive decline. *Journal of Geriatric Psychiatry and Neurology*, *18*(1), 25-32.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? .*Psychological review*, *97*(3), 315.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*(1), 3-72.

Pakosz, M. (1983). Attitudinal judgments in intonation: Some evidence for a theory. *Journal of Psycholinguistic Research*, 12, 311–326.

Panksepp, J. (2008). The affective brain and core consciousness: How does neural activity generate emotional feelings? In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 47-67). New York: The Guilford Press.

Panksepp, J. (2007a). Emotional feelings originate below the neocortex: Toward a neurobiology of the soul. *Behavioral and Brain Sciences,* 30, 101-103.

Panksepp, J. (2007b). Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives on Psychological Science*, 2(3), 281-296.

Panksepp, J. (2005a). Affective consciousness. Core emotional feelings in animals and humans. Consciousness and Cognition 12, p. 30-80.

Panksepp, J. (2005b). On the embodied neural nature of core emotional affects. *Journal of consciousness studies*, *12*(8-10), 158-184.

Panksepp, J. (2000). Emotions as natural kinds within the mammalian brain. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 137–156). New York: Guilford Press.

Panksepp, J. (1998). Affective neuroscience. *The foundations of human and animal emotions.* Oxford, New York: Oxford University Press.

Panksepp, J. (1992). A critical role for affective neuroscience in resolving what is basic about basic emotions. *Psychological Review, 99,* 554 –560.

Panksepp, J. (1990). Can “mind” and behavior be understood without understanding the brain? A response to Bunge. *New Ideas in Psychology*, *8*, 139–149.

Panksepp, J., & Biven, L. (2011). *The archaeology of mind: Neuroevolutionary origins of human emotions*. New York, NY: Norton.

Panksepp, J., Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion review,* 3 (4) 387-396.

Panksepp, J., & Panksepp, J. B. (2000). The seven sins of evolutionary psychology. *Evolution and cognition*, *6*(2), 108-131.

Pardue, M. L., & Wizemann, T. M. (Eds.). (2001). *Exploring the Biological Contributions to Human Health: Does Sex Matter?* Washington, D.C.: National Academy Press.

Parkinson, B. (1996). Emotions are social. *British journal of psychology*, *87*(4), 663-683.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), 35-45.

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language*, 104, 262– 269. doi:10.1016/j.bandl.2007.03.002

Pell, M. D. (2002). Evaluation of nonverbal emotion in face and voice: Some preliminary findings on a new battery of tests. *Brain and Cognition*, 48, 499–514.

Pell,M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *Journal of the Acoustical Society of America,* 109, 1668–1680. doi:10.1121/1.1352088

Pell, M. D., Jaywant, A., Monetta, L., & Kotz, S. A. (2011). Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cognition and Emotion*, 25, 834–853. doi:10.1080/02699931.2010.516915

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, 6, e27252. doi:10.1371/journal.pone.0016505

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107-120.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37, 417–435.

Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication*, 50, 519–530.

Permanyer, I. (2013). A critical assessment of the UNDP's Gender Inequality Index. *Feminist Economics*, *19*(2), 1-32.

Petrushin, V. (1999, November). *Emotion in speech: Recognition and application to call centers*. Paper presented at the Conference on Artificial Neural Networks in Engineering, St. Louis, USA.

Piko, B. (2001). Gender differences and similarities on adolescents' ways of coping. *Psychological Record*, *51,* 223–235.

Pittam, J., & Scherer, K. R. (1993). Vocal expression and communication of emotion. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 185-198). New York: Guilford Press.

Planalp, S. (1999). *Communicating emotion: Social, moral, and cultural processes*. Cambridge: Cambridge University Press.

Plutchik, R. (2003). *Emotions and life: Perspectives from psychology, biology, and evolution*. American Psychological Association.

Plutchik, R. (1994). *The psychology and biology of emotion.* New York: Harper-Collins.

Plutchik, R. (1980). *Emotion: A psychoevolutionary synthesis*. New York: Haper and Row.

Plutchik, R. (1962). *The emotions: Facts, theories, and a new model.* New York: Random House.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European journal of developmental psychology*, *1*(2), 127-152.

Potter, S. H. (1988). The cultural construction of emotion in rural Chinese social life. *Ethos, 16,* 181–208.

Power, M., & Dalgleish, T. (2007). *Cognition and emotion: From order to disorder*. Hove: Psychology press.

Rahman, M., & Jackson, S. (2010). *Gender and Sexuality: Sociological Approaches*. Cambridge: Polity Press.

Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Sex, sexual orientation, and identification of positive and negative facial affect. *Brain and cognition*, *54*(3), 179-185.

Raithel, V., & Hielscher-Fastabend, M. (2004). Emotional and linguistic perception of prosody. *Folia Phoniatrica et Logopaedica*, *56*(1), 7-13.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292.

Rao, K. S., Koolagudi, S. G., & Vempada, R. R. (2013). Emotion recognition from speech using global and local prosodic features. *International Journal of Speech Technology*, *16*(2), 143-160.

Rezaienia, A. (2012). Interpersonal skills among Iranian youth (In Persian). Master thesis. Tarbiat Modarres University. Tehran, Iran.

Richeson, P. J. & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago, IL: University of Chicago Press.

Roach, P. (2000, September). *Techniques for the phonetic description of emotional speech*. Paper presented at the ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, Newcastle, Northern Ireland, UK.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable timed' languages. *Linguistic controversies*, 73-79.

Roach, P., Stibbard, R., Osborne, J., Arnfield, S., & Setter, J. (1998). Transcription of prosodic and paralinguistic features of emotional speech. *Journal of the International Phonetic Association*, 28, 83–94.

Robinson, S. E., & Kinnier, R. T. (1985). Sex and Order of Counseling Effects on Paired Role-Play Practice. *Counselor Education and Supervision*, *25*(2), 143-148.

Roseman, I. J. (2013). Author Reply: On the Frontiers of Appraisal Theory. *Emotion Review*, *5*(2), 187-188.

Roseman, I. J., Dhawan, N., Rettek, S. I., Nadidu, R. K., & Thapa, K. (1995). Cultural differences and cross-cultural similarities in appraisals and emotional responses. *Journal of Cross-Cultural Psychology, 26*, 23-48.

Rosenthal, F. (1977). I am You—Individual Piety and Society in Islam. In Banani, A. and Vryonis, S. (Eds.), *Individualism and conformity in classical Islam.* Wiesbaden: Otto Harrassowitz.

Roshanfar, M. (2003). *Woman, Employment and ownership (In Persian).* Tehran, Iran: Barge Zeytoon publications.

Ross, E. D. (1993). Nonverbal aspects of language. *Neurologic Clinics*, *11*, 9–23

Ross, I. (2010). *Autogenic Dynamics, Stress, Affect Regulation, and Autogenic Therapy*. North Berwick: Tantallon Press.

Russ, J. B., Gur, R. C., & Bilker,W. B. (2008). Validation of affective and neutral sentence content for prosodic testing. *Behavior Research Methods*, 40, 935–939. doi:10.3758/BRM.40.4.935

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110*(1), 145-172.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141. doi:10.1037/0033-2909.115.1.102

Russell, J. A. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion*, *17*(1), 41-51.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178.

Safdar, S., Friedlmeier, W., Matsumoto, D., Yoo, S. H., Kwantes, C. T., Kakai, H., & Shigemasu, E. (2009). Variations of emotional display rules within and across cultures: A comparison between Canada, USA, and Japan. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *41*(1), 1-10.

Samareh, Y. (1977). *The arrangement of segmental phonemes in Farsi*. University of Tehran Press.

Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, *31*(1), 97-113.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, USA, 107,* 2408–2412.

Schalkwyk, J. (2000). Culture, gender equality and development cooperation. Note prepared for the Canadian International Development Agency (Cida), Quebec. Online: www. oecd. org/dataoecd/2/9/1896320. pdf (downloaded Feb. 14, 2015).

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, *23*, 1307– 1351. doi: 10.1080/02699930902928969

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 693–727.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication, 40, 227–256.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92– 120). New York: Oxford University Press.

Scherer, K. R. (2000). Psychological models of emotion. In J. Borod (Ed.). *The neuropsychology of emotion* (pp. 137–162). Oxford/New York: Oxford University Press.

Scherer, K. R. (1997). The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology*, 73(5), 902–922.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143–165. doi:10.1037/0033-2909.99.2.143

Scherer, K.R. (Ed.). (1982). *Vokale Kommunikation: Nonverbale Aspekte des Sprachverhaltens.* [Vocal Communication: Nonverbal Aspects of Speech]. Beltz, Weinheim.

Scherer, K., Clark-Polner, E., Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology* 46 (6), 401-435.

Scherer, K. R., & Brosch, T. (2009). Culture-specific appraisal biases contribute to emotion dispositions. *European Journal of Personality*, *23*(3), 265-288.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural psychology*, *32*(1), 76-92.

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). Appraisal processes in emotion: Theory, research,

application. New York: Oxford University Press.

Scherer, K. R., & Wallbott H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology,* 66(2), 310-328.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15, 123-148.

Scherer, K. R., Matsumoto, D., Wallbott, H., & Kudoh, T. (1988). Emotional experience in cultural context: A comparison between Europe, Japan, and the USA. In K. Scherer (Ed.), *Facets of emotion: Recent research* (pp. 5-30). Hillsdale, NJ: Erlbaum.

Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America,* 76, 1346–1356. doi:10.1121/1.391450

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in cognitive sciences*, *10*(1), 24-30.

Schirmer, A., Lui, M., Maess, B., Escoffier, N., Chan, M., & Penney, T. B. (2006). Task and sex modulate the brain response to emotional incongruity in Asian listeners. *Emotion*, *6*(3), 406.

Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport*, *16*(6), 635-639.

Schirmer, A., Zysset, S., Kotz, S. A., & von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, *21*(3), 1114-1123.

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of cognitive neuroscience*, *15*(8), 1135-1148.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, 14, 228–233.

Schlüter, J. (2005). *Rhythmic grammar: The influence of rhythm on grammatical variation and change in English* (Vol. 46). Walter de Gruyter.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, 2, 419–436.

Schneider,W., Eschman, A.,& Zuccolotto, A. (2002). E-Prime 1.0 user's guide. Pittsburgh, PA: Psychological Software Tools.

Seiffge-Krenke, I. (2013). *Stress, coping, and relationships in adolescence*. Psychology Press.

Scott, J. P. (1980). The function of emotions in behavioral systems: A systems theory analysis. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 35–56). New York: Academic Press.

Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for?. *Current Directions in Psychological*

Science, *20*(6), 395-399. doi:10.1177/0963721411424739

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further explorations of a prototype approach. *Journal of Personality and Social Psychology, 52,* 1061–1086.

Shields, S. A. (2003). *Speaking from the heart: Gender and the social meaning of emotion*. New York: Cambridge University Press.

Shih, C., & Kochanski, G. (2002). *Prosody and Prosodic Modeling.* (Three hour tutorial at The Seventh International Conference on Spoken Language Processing. Denver CO. http://prosodies.org/tutorial2002

Sims-Williams, N., & Bailey, H. W. (Eds.). (2002). *Indo-Iranian languages and peoples*. Oxford, UK: Oxford University Press.

Skinner, E. R. (1935). A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness; and a determination of the pitch and force of the subjective concepts of ordinary, soft, and loud tones. *Communications Monographs*, *2*(1), 81-137.

Soto, J. A., Levenson, R. W., & Ebling, R. (2005). Cultures of moderation and expression: Emotional experience, behavior, and physiology in Chinese Americans and Mexican Americans. *Emotion, 5*, 154-165.

Stephan, W. G., Stephan, C. W., & de Vargas, M. (1996). Emotional expression in Costa 89 Rica and the United States. *Journal of Cross-Cultural Psychology, 27*, 147-160.

Stets, J.E., & Burke, P.J. (2000). Femininity/masculinity. In Borgatta, E. F, & Montgomery, R.J.V. (Eds.). *Encyclopedia of sociology*. New York: Macmillan.

Stevens, K. N., & Williams, C. E. (1969). On determining the emotional state of pilots during flight- An exploratory study (Pilot emotional state during stressful situations from tape recorded vocal utterances of air to ground radio communications using spectrographic analysis). *Aerospace Medicine*, *40*, 1369-1372.

Sullivan, Z. (2010). *Exiled memories: Stories of Iranian diaspora*. Philadelphia: Temple University Press.

Surcinelli, P., Codispoti, M., Montebarocci, O., Rossi, N., & Baldaro, B. (2006). Facial emotion recognition in trait anxiety. *Journal of Anxiety Disorders*, *20*, 110–117.

Susskind, J.M., Lee, D.H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A.K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, 11, 843–850.

Szymanowski, F., Szymanowski, F., Kotz, S. A., Schröder, C., Rotte, M., & Dengler, R. (2007). Gender differences in processing emotional prosody. *Clinical Neurophysiology*, *118*(4), 102-103.

Tanaka, K. (2015). Language, Gender and Culture. In F. Sharifian (Ed.). *The Routledge Handbook of Language and Culture.* Pp. 150–165. London, New York: Routledge.

Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I feel your voice

cultural differences in the multisensory perception of emotion. *Psychological science*.

Tanenhaus, M. K., & Brown-Schmidt, S. (2007). Language processing in the natural world. *Philosophical Transactions of the Royal Society B, 363*, 1105–1122. doi:10.1098/rstb.2007.2162

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: tend-and-befriend, not fight-or flight. *Psychological review*, *107*(3), 411–429.

ten Bosch, L. (2003). Emotions, speech and the ASR framework. *Speech Communication*, *40*(1), 213-225.

The Hofstede Center (n.d). Cultural tools: country comparison. Retrieved August 12, 2014, from http://geert-hofstede.com/germany.html

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognize non-verbal displays of emotion: A meta-analysis. *Cognition & emotion*, 28(7), 1164-1195.

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica*, (158), 407-424.

Thoits, P. A. (2013). Self, identity, stress, and mental health. In *Handbook of the sociology of mental health* (pp. 357-377). Springer Netherlands.

Timmers, M., Fischer, A., & Manstead, A. (2003). Ability versus vulnerability: Beliefs about men's and women's emotional behaviour. *Cognition & Emotion*, 17, 41–63.

Toivanen, J., Väyrynen, E., & Seppänen, T. (2005). Gender differences in the ability to recognize emotional content from speech. In *Proceedings of Fonetik* (Pp. 119-122).

Tomkins, S. S. (1962). *Affect, imagery, consciousness: The positive affects* (Vol. 1). London: Tavistock.

Triandis, H. C. (1995). *Individualism and collectivism.* Boulder, CO: Westview.

Triandis, H. C. (1994). Major cultural syndromes and emotion. In Kitayama, S. E., & Markus, H. R. E. (Eds). *Emotion and culture: Empirical studies of mutual influence*. American Psychological Association.

Triandis, H. C. (1972). *The analysis of subjective culture.* New York: Wiley.

Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of personality and social psychology*, *90*(2), 288.

Tsai, J.L., Chentsova-Dutton, Y. (2003). Variation among European Americans in emotional facial expression. *Journal of Cross-Cultural Psychology, 34*, 650-657.

Turner, T.J., & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review,* 99, 566–571.

Udry, J. R. (1994). The nature of gender. *Demography*, *31*(4), 561-573.

Van Bezooijen, R., Otto, S., & Heenan, T. (1983). Recognition of vocal expressions of emotion: A threenation study to identify universal characteristics. *Journal of Cross-Cultural Psychology*, 14(4), 387–406.

Van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. Behavior Research Methods, Instruments, Computers. 39 (4), 973-978

Van Lancker, D., & Sidtis, J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: all errors are not created equal. *Journal of Speech and Hearing Research*, *35*(5), 963-70.

Van Rheenen, T. E., & Rossell, S. L. (2013). Auditory-prosodic processing in bipolar disorder; from sensory perception to emotion. *Journal of affective disorders*, *151*(3), 1102-1107.

Ververidis, D., & Kotropoulos, C. (2003, October). *A state of the art review on emotional speech databases.* Paper presented at the 1st Richmedia Conference, Lausanne, Switzerland.

Vihman, M. M., Nakai, S., & DePaolis, R. A. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. *Laboratory phonology*, *8*, 341-366.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3-28.

Wallbott, H. G., & Scherer, K. R. (1986). How universal and specific is emotional experience? Evidence from 27 countries on five continents. *Social Science Information*, 25, 763–795.

Walsh, W. A., & Banaji, M. R. (1997). The Collective Selfa. *Annals of the New York Academy of Sciences*, *818*(1), 193-214.

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: a meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological bulletin*, 138(4), 775–808. doi: 10.1037/a0027600

Weitz, S. (1976). Sex differences in nonverbal communication. *Sex Roles*, *2*(2), 175-184.

Wells, J. C. (2006). *English Intonation PB and Audio CD: An Introduction*. Cambridge: Cambridge University Press.

Wierzbicka, A. (1994). Emotion, language, and cultural scripts. In S. Kitayama & H. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 133–196). Washington, DC: American Psychological Association.

Williams, M. A., & Mattingley, J. B. (2006). Do angry men get noticed? *Current Biology*, 16, R402–R404.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, 38, 1559–1579.

Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and men:

implications for the origins of sex differences. *Psychological bulletin*, *128*(5), 699.

Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion*, *11*(4), 705–731. doi:10.1037/a0023980

Yu, F., Chang, E., Xu, Y., & Shum, H. Y. (2001, October). *Emotion detection from speech to enrich multimedia content.* Paper presented at the 2nd IEEE Pacific Rim Conference on Multimedia, London, United Kingdom.

Appendix A

Sample of the Persian sentences included in the database, along with their transliteration, glosses, and English translation.[1]

عصبانیت: آقای مهندس همه رو با بوق گوشخراش ماشینش بیدار کرد.

ʔāqā-ye mohandes hame ro bā buq-e gušxarāš-e māšin-eš bidār kard.

Mrister--Ez engineer everybody DOM with horn-Ez ear-irritating-Ez car-CL-3SG awake did-3SG

Anger: Mr. Engineer woke up everybody with the awful sound of his car's horn.

چندش: آقای آشپز چند تا کرم سفید تو آش رشته پیدا کرد .

ʔāqā-ye ʔāšpaz čand-tā kerm-e sefid tu ʔāš-e rešte peydā kard.

Mrister-Ez cook some-CL worm-Ez white in soup-Ez noodle find did-3SG

Disgust: Mr. Cook found some white worms in the noodle soup.

[1] Abbreviations used are as follows: Ez: ezafe particle; CL: clitic ; CL.3SG third person singular clitic; DOM: direct object marker; 3SG: third person singular.

ترس: دختر شیرنی‌فروش یه رتیل یه وجبی تو تختش پیدا کرد.

Doxtar-e širinifruš ye roteyl-e ye vajab-i tu taxt-eš peydā kard.

Daughte-Ez confectioner one spider-Ez one hand size-CL in bed-CL-3SG obvious did-3SG

Fear: The Confectioner's daughter found a hand-size spider in her bed.

شادی : پسر معدنچی هر دو چشمش رو با موفقیت معالجه کرد.

Pesar-e maʔdanči har do češm-eš ro bā movafaqiyat moʔāleje kard.

Son-Ez miner each two eye-CL-3SG DOM with success treatment did-3SG

Happiness: The Miner's son had treated both of his eyes successfully.

غم : دختر صاحبخونه بچه‌اش رو به‌خاطر سرطان سقط کرد.

Dokhtar-e sāhebxune bačče-aš ro bexāter-e saratān seqt kard.

Daughter-Ez landlord child-CL-3SG DOM for-Ez cancer miscarriage did-3SG

Sadness: The Landlord’s daughter suffered a miscarriage due to cancer.

خنثی : خانم خونه‌دار روسری گلدارش رو بعداز شام اتو کرد.

Xānom-e	xunedār	rusari-ye	goldār-eš	ro	baʔdaz	šām	ʔotu	kard.
Lady-Ez	housewife	scarf-Ez	damask-CL-3SG	DOM	after	dinner	iron	did-3SG

Neutral: Ms. Housewife ironed her damask scarf after the dinner.

Appendix B

List of scenarios

Anger: The director is late for the rehearsal again and we have to work until late at night. Once again I have to cancel an important date.

Disgust: I have a summer job in a restaurant. Today I have to clean the toilets which are incredibly filthy and smell very strongly.

Fear: While I am on a tour bus, the driver loses control of the bus while trying to avoid another car. The bus comes to a standstill at the edge of a precipice, threatening to fall over.

Happiness: I am acting in a new play. From the start, I get along extremely well with my colleagues who even throw a party for me.

Sadness: I get a call to tell me that my best friend died suddenly.

Appendix C

List of the Stimuli used in study two along with the English translation

12 Persian syntactically similar sentences (i.e., subject + object + prepositional phrase + verb), neutral in lexical content intoned in the five intended emotions (anger, disgust, fear, happiness, sadness) by a male and a female speaker.

List of the sentences
Mr. Notary officer arranged the deeds based on their dates. آقای محضردار سندها رو از روی تایخشون دسته بندی کرد.
Mr. Manager cleaned his glasses with tissue paper. آقای مدیرعینکش رو با یه دستمال کاغذی سفید تمیز کرد.
Mr. Greengrocer lifted the paperboard boxes from the ground. آقای سبزی فروش جعبه های مقوایی رو از روی زمین بلند کرد.
Mr. Reporter switched on his old radio before dinner. آقای گزارشگر رادیوی قدیمیش رو قبل از شام روشن کرد.
Ms. Clerk took out all the papers from the drawer. خانم دفتردار همه کاغذ ها رو از تو کشو بیرون آورد.
Mr. Lawyer took his suit from the suitcase. آقای وکیل کت شلوارش رو از تو چمدون بیرون آورد.
The young lady cleaner hung the office keys on the key holder. دختر نظافتچی کلید در شرکت رو به جا کلیدی آویزون کرد.
The young male athlete dried his hair with a cotton towel. پسر ورزشکار موهاش رو با یه حوله نخی خشک کرد.
The young waitress chopped the vegetables in the kitchen. دختر پیشخدمت سبزی ها رو تو آشپزخونه خورد کرد.
Ms. Housewife ironed her damask scarf after the dinner. خانم خونه دار روسری گلدارش رو بعد از شام اتو کرد .
The Café owner's daughter peeled the apples with a small knife. دختر کافه چی سیب ها را با یه چاقوی کوچیک پوست کند.
The young male visitor distributed the flyer to the people. پسر بازاریاب آگهی ها ی تبلیغاتی رو بین مردم پخش کرد.

Appendix D

English translation of the Stimuli used in study three.

List of the sentences	*Speaker Gender*	*Intended Emotional prosody*		
Mr. Engineer awaked everybody with the awful sound of his car horn.	M	Anger	*Anger*	*Intended emotion of the lexical content*
The young repairman crashed a customer's new car with a key.	M	Anger		
Ms. Secretary listened to her colleagues' private phone calls without permission.	M	Anger		
The daughter of the teacher torn her friend's new backpack with a metal fork.	M	Anger		
The motor cyclist guy snapped a young lady's wallet from her hands.	M	Anger		
The daughter of the principal threw her classmate's leather bag into the river	M	Anger		
The neighbor's daughter ripped her sister's wedding dress with scissors.	F	Anger		
Mr. School master belted all the fifth-grade school boys.	F	Anger		
The son of the manager crushed the new cell phone of one of the personnel with a big stone.	F	Anger		
The gatekeeper's wife made the neighbors tired and sick with her screams.	F	Anger		
Mr. Landlord smashed his tenant's car's light with his crow bar.	F	Anger		
The daughter of the boss opened the staffs' personal letters without permission	F	Anger		
The daughter of the director filled her glass from the lavatory.	M	Disgust	*Disgust*	
Mr. Neighbor spread his son's excrement with his feet.	M	Disgust		
Ms. Confectioner found a lot of mouse excrement in the flour bags.	M	Disgust		
The son of the landlord plunged his hand into the toilet.	M	Disgust		
The Tea man found a dead cockroach in the sugar bowl.	M	Disgust		
Ms. Licensed left a dirty diaper on the dining table.	M	Disgust		
Mr. Cook found some white worms in the noodle soup.	F	Disgust		
The chef's daughter minced a toad in the meat grinder.	F	Disgust		
Ms. Waitress plunged her dirty socks into her juice.	F	Disgust		
Mr. Judge spat his phlegm in his cup of tea.	F	Disgust		
The son of the pilot crushed a live cockroach to death in his fist.	F	Disgust		
Ms. Consultant wiped her son's nose on her own sleeve.	F	Disgust		

List of the sentences	*Speaker Gender*	*Intended Emotional prosody*		
The gang-leader's son placed the dead body in the trunk of the car.	M	Fear	*Fear*	*Intended emotion of the lexical content*
The young male huckster threatened the bank manager with a scalpel.	M	Fear		
Ms. Servant found a living scorpion in her sneakers.	M	Fear		
The Confectioner's daughter found a hand-size hairy spider in her bed.	M	Fear		
The lady mountain climber lost her fellow travelers in a foggy forest.	M	Fear		
The son of the drug dealer disabled his hostage with anesthetic.	M	Fear		
The drug dealer's daughter opened the door of a forlorn house with apprehension.	F	Fear		
The young male driver threatened her passenger with pepper gas.	F	Fear		
Lady Kidnapper took her pistol out from under her head scarf.	F	Fear		
The lady photographer lost her traveling companions in a dark valley.	F	Fear		
The son of the café owner pointed his pistol towards the guests.	F	Fear		
The son of the butcher hid the dead body of his victim in the basement of the fort.	F	Fear		
Ms. Teacher encouraged the top student in front of all other students.	M	Happiness	*Happiness*	
The young Huckster girl found her father after years.	M	Happiness		
The miner's son had both of his eyes treated successfully.	M	Happiness		
Mr. Teacher provided the money he needed in three days.	M	Happiness		
Ms. Physician tested the new medicine for diabetes successfully.	M	Happiness		
Mr. Janitor won a Peugeot in the bank lottery.	M	Happiness		
The daughter of the gardener decorated her wedding flower bouquet with a golden ribbon.	F	Happiness		
After searching for a while, Ms. Hairdresser found both her money and her ring.	F	Happiness		
Mr. Film Director won a golden statue in the film festival.	F	Happiness		
The tailor's daughter decorated her brother's wedding table with roses.	F	Happiness		
The son of the gate-keeper received his PhD with honours.	F	Happiness		
The son of the farmer won first place in the final matches.	F	Happiness		

List of the sentences	*Speaker Gender*	*Intended Emotional prosody*		
Mr. Gardener lost his daughter and his son in the war.	M	Sadness	*Sadness*	*Intended emotion of the lexical content*
Ms. Host lost her parents in childhood.	M	Sadness		
The young lady Florist abandoned her one-month infant due to poverty.	M	Sadness		
Mr. Minister buried his teenage son in the cemetery.	M	Sadness		
Ms. Engineer hospitalized her father in the cancer section.	M	Sadness		
The son of the baker lost all his family in last year's earthquake.	M	Sadness		
Mr. Driver lost one of his legs due to diabetes.	F	Sadness		
The Gatekeeper's daughter left school due to poverty.	F	Sadness		
Ms. Actress destroyed her life by her addiction.	F	Sadness		
The young male peddler left his family out of necessity.	F	Sadness		
Mr. Writer sold all his possessions to pay for his daughter's treatment.	F	Sadness		
The daughter of the landlord suffered a miscarriage due to blood cancer.	F	Sadness		
The Café owner's daughter peeled the apples with a small knife.	M	Neutral	*Neutral*	
Mr. Manager cleaned his glasses with tissue paper.	M	Neutral		
Mr. Greengrocer lifted the paperboard boxes from the ground.	M	Neutral		
Ms. Clerk took out the papers from the drawer.	M	Neutral		
Mr. Lawyer took his suit from the suitcase.	M	Neutral		
The young lady cleaner hung the office keys on the key holder.	M	Neutral		
Ms. Housewife ironed her damask scarf after the dinner.	F	Neutral		
The young male visitor distributed the flyer to the people.	F	Neutral		
Ms. Neighbor collected the white bed sheets from the line.	F	Neutral		
Mr. Notary officer arranged the deeds based on their dates.	F	Neutral		
Mr. Reporter switched on his old radio before dinner.	F	Neutral		
The young waitress chopped the vegetables in the kitchen.	F	Neutral		

Appendix E

List of the Persian Stimuli used in study three.

Intended emotion of the lexical content		*Intended Emotional prosody*	*Speaker Gender*	*List of the sentences*
	عصبانیت	Anger	M	آقای مهندس همه رو با بوق گوشخراش ماشینش بیدار کرد .
		Anger	M	پسر تعمیرکار ماشین نو یه مشتری رو با کلید خط انداخت.
		Anger	M	خانم منشی تلفن های خصوصی همکارهاش رو بدون اجازه گوش داد .
		Anger	M	دخترمعلم کوله پشتی نو دوستش رو با یه چنگال فلزی پاره کرد.
		Anger	M	پسرموتورسوار کیف پول یه خانم جوان رو از دستش قاپ زد.
		Anger	M	دختر مدیر کیف چرمی همکلاسیش رو تو رودخونه پرت کرد .
		Anger	F	دختر همسایه لباس عروسی خواهرش روبا قیچی پاره کرد .
		Anger	F	آقای ناظم همه شاگردهای کلاس پنجمی رو با کمربند کتک زد .
		Anger	F	پسر رییس موبایل یکی از کارمند ها رو با یه قلوه سنگ خورد کرد.
		Anger	F	خانم سرایدار همسایه ها رو با جیغ و دادش به ستوه آورد .
		Anger	F	آقای صاحبخونه چراغ ماشین مستاجرش رو با قفل فرمونش خورد کرد.
		Anger	F	دختر رییس نامه های خصوصی کارمند ها رو بدون اجازه باز کرد.
	چندش	Disgust	M	دختر کارگردان لیوانش رو از آب چاه توالت پر کرد.
		Disgust	M	آقای همسایه مدفوع پسرش رو با پاهاش پخش کرد .
		Disgust	M	خانم قناد کلی فضله موش تو کیسه های آرد پیدا کرد.
		Disgust	M	پسر صاحبخونه دستش رو تو چاه توالت فرو کرد.
		Disgust	M	آقای آبدارچی یه سوسک مرده تو قندون پیدا کرد .
		Disgust	M	خانم بهیار یه پوشک کثیف رو روی میز نهارخوری جا گذاشت.
		Disgust	F	آقای آشپز چند تا کرم سفید تو آش رشته پیدا کرد.
		Disgust	F	دختر سرآشپز یه وزغ رو تو چرخ گوشت چرخ کرد .
		Disgust	F	خانم پیشخدمت جوراب کثیفش رو تولیوان شربتش فرو کرد .
		Disgust	F	آقای قاضی خلطش رو توی فنجون چایش تف کرد.
		Disgust	F	پسر خلبان یه سوسک زنده رو تو مشتش له کرد.
		Disgust	F	خانم مشاور آبِ دماغ پسرش رو با لبه آستینش پاک کرد.

Intended emotion of the lexical content		Intended Emotional prosody	Speaker Gender	List of the sentences
	ترس	Fear	M	پسررییس باند جسد مقتول رو تو صندوق عقب جاسازی کرد.
		Fear	M	پسر دوره گرد رییس بانک رو با تیغ جراحی تهدید کرد.
		Fear	M	خانم خدمتکار یه عقرب زنده تو کفش کتونیش پیدا کرد.
		Fear	M	دختر شیرنی فروش یه رتیل یه وجبی تو تختش پیدا کرد.
		Fear	M	دختر کوهنورد راهش رو تو یه جنگل مه گرفته گم کرد .
		Fear	M	پسر قاچاقچی گروگانش رو با داروی خواب آور بیهوش کرد .
		Fear	F	دختر موادفروش در یه خونه متروکه رو با دلهره باز کرد.
		Fear	F	پسر راننده مسافرش رو با اسپری خردل تهدید کرد .
		Fear	F	دختر گروگان گیر اسلحه اش رو از زیر مقنعه اش بیرون آورد .
		Fear	F	خانم عکاس همه همسفرهاش رو توی یه دره تاریک گم کرد .
		Fear	F	پسر کافه چی اسلحه اش رو به طرف مسافرها نشونه رفت .
		Fear	F	پسر قصاب جسد قربانیش رو تو زیرزمین قلعه مخفی کرد .
	شادی	Happiness	M	خانم معلم شاگرد اول کلاس رو سر صف تشویق کرد.
		Happiness	M	دختر دوره گرد پدرش رو بعد از سالها بی خبری پیدا کرد .
		Happiness	M	پسر معدنچی هر دو چشمش رو با موفقیت معالجه کرد .
		Happiness	M	آقای معلم همه پول مورد نیازش رو درعرض سه روز تهیه کرد.
		Happiness	M	خانم دکتر داروی جدید بیماری قند رو با موفقیت امتحان کرد .
		Happiness	M	آقای رفتگر یه ماشین پژو تو قرعه کشی بانک برنده شد .
		Happiness	F	دختربا‌غبون دسته گل عروسیش رو با یه روبان طلایی تزیین کرد .
		Happiness	F	خانم آرایشگر پولها و انگشترش رو بعد از کلی گشتن پیدا کرد.
		Happiness	F	آقای کارگردان یه تندیس طلایی تو جشنواره فیلم بدست آورد.
		Happiness	F	دختر خیاط سفره عقد برادرش رو با گلای رز تزیین کرد .
		Happiness	F	پسر سرایدار مدرک دکتراش رو با درجه عالی دریافت کرد .
		Happiness	F	پسر کشاورز مقام اول رو تو مسابقات نهایی بدست آورد .

List of the sentences	Speaker Gender	Intended Emotional prosody		
آقای باغبون دختر و پسرش رو تو جنگ ازدست داد .	M	Sadness	غم	Intended emotion of the lexical content
خانم مجری پدر و مادرش رو تو بچگی از دست داد .	M	Sadness		
دختر گلفروش نوزاد یک ماهه اش رو به خاطر فقر رها کرد .	M	Sadness		
آقای وزیر پسر نوجوانش رو تو بهشت زهرا به خاک سپرد.	M	Sadness		
خانم مهندس پدرش رو تو بخش بیماران سرطانی بستری کرد.	M	Sadness		
پسر نانوا همه خونواده اش رو تو زلزله سال پیش گم کرد .	M	Sadness		
آقای راننده یه پاش رو به خاطر بیماری قند از دست داد.	F	Sadness		
دختر سرایدار مدرسه رو به خاطر بی پولی ترک کرد .	F	Sadness		
خانم بازیگر همه ی زندگیش رو به خاطر اعتیادش تباه کرد .	F	Sadness		
پسر دستفروش خانواده اش رو از روی ناچاری ترک کرد .	F	Sadness		
آقای نویسنده دار و ندارش رو برای معالجه دخترش حراج کرد.	F	Sadness		
دختر صاحبخونه بچه اش رو به خاطر سرطان سقط کرد .	F	Sadness		
دختر کافه چی سیب ها را با یه چاقوی کوچیک پوست کند .	M	Neutral	خنثی	
آقای مدیر عینکش رو با یه دستمال کاغذی سفید تمیز کرد .	M	Neutral		
آقای سبزی فروش جعبه های مقوایی رو از رو زمین بلند کرد .	M	Neutral		
خانم دفتردار کاغذ ها رو از تو کشو بیرون آورد .	M	Neutral		
آقای وکیل کت شلوارش رو از تو چمدون بیرون آورد .	M	Neutral		
دخترنظافتچی کلید در شرکت رو به جاکلیدی آویزون کرد .	M	Neutral		
خانم خونه دار روسری گلدارش رو بعد از شام اتو کرد .	F	Neutral		
پسر بازاریاب آگهی ها ی تبلیغاتی رو بین مردم پخش کرد .	F	Neutral		
خانم همسایه ملحفه های سفید رو از رو طناب رخت جمع کرد .	F	Neutral		
آقای محضردار سند ها رو از رو تاریخشون دسته بندی کرد.	F	Neutral		
آقای گزارشگر رادیوی قدیمیش رو قبل از شام روشن کرد .	F	Neutral		
دختر پیشخدمت سبزی ها رو تو آشپز خونه خورد کرد .	F	Neutral		

List of Tables

List of Figures

curriculum vitae

"For reasons of data protection, the curriculum vitae is not published in the electronic version"

Affidavit

I hereby swear in lieu of an oath that I have independently prepared this thesis and without using other aids than those stated. The data and concepts taken over from other sources or taken over indirectly are indicated citing the source. The thesis was not submitted so far either in Germany or in another country in the same or a similar form in a procedure for obtaining an academic title.

Niloofar Keshtiari
Berlin, 2015