Journal of Sociolinguistics
Payame Noor University

فصلنامهٔ زبان‌شناسی اجتماعی
شمارهٔ اول ـ سال اول- پاییز ۱۳۹۵ (صفحات ۱۲۱-۱۰۷)

تاثیر جنسیت بر درک نوای عاطفی گفتار در زبان فارسی

نیلوفر کشتیاری*
دکتری تخصصی زبان‌شناسی شناختی، گروه زبان‌های احساس دانشگاه آزاد برلین
دریافت: ۱۳۹۵/۰٤/۱٤ پذیرش:۱۳۹۵/۰۸/۱۷

Gender Effects on the Recognition of Emotional Prosody: Evidence from Persian Language

Niloofar Keshtiari
PhD in Cognitive Linguistics, Freie Universitaet Berlin, Germany
Received:2016/07/04 Accepted:2016/11/07

چکیده
لازمهٔ تعامل اجتماعی مؤثر پردازش و درک صحیح نوای عاطفی گفتار گوینده است. گفتار عاطفی، که حاوی اطلاعات زیادی دربارهٔ احساسات و مقاصد اجتماعی گوینده است، خود از تفاوت‌هایی فردی-جامعه‌شناختی همچون سن و جنس تاثیر می‌پذیرد که از این میان نقش جنسیت به اندازهٔ کافی مطالعه نشده و هنوز محل بحث است. همچنین معدود مطالعات انجام‌شده، تنها عملکرد گویشوران زبان‌های دارای وزن تکیه‌ای (همچون آلمانی و انگلیسی) را بررسی کرده‌اند. اما از آنجا که نوای عاطفی گفتار، خود، متاثر از شاخصه‌های زبان-ویژه، مانند تقسیم وزنیِ زمان، است و این شاخصه زبان‌های دارای وزن تکیه‌ای را از زبان‌های دارای وزن هجایی تفکیک می‌کند نگارنده بر آن شد تا قابلیت کاربرد نتایج این تحقیقات را در زبان‌هایی که تقسیم زمانی متفاوتی دارد بررسی کند. بنابراین، پژوهش حاضر نقش جنسیت را در تشخیص و درک نوای عاطفی گفتار در زبان فارسی، به عنوان زبانی که وزن هجایی دارد و تقسیم وزنی-زمانیِ آن متفاوت است، مورد مطالعه قرار می‌دهد. بدین منظور در یک آزمایش رفتاری، از ۶۶ گویشور فارسی‌زبان خواسته شد تا نوای عاطفی مجموعه‌ای از متون صوتی نمایشنامه‌های معتبر را تشخیص‌دهند. این متون حاوی مجموعه جملاتی بود که واژگان موجود در آن‌ها به لحاظ محتوایی خنثی بودند و در پنج حالت عاطفی خشم، تنفر، ترس، شادی و غم توسط یک بازیگر خانم و یک بازیگر آقا اجرا شده بودند. داده‌های این آزمون با مدل رگرسیون لجستیک بررسی شدند و به طور کلی، نتایج نشان می‌دهند که: ۱) میزان تشخیص نوای عاطفی گفتار به صورت معناداری با جنسیت شنونده ارتباط دارد؛ بدین معنا که آزمودنی‌های مؤنث در مقایسه با آزمودنی‌های مذکر توانمندی بیشتری در تشخیص تمامی مقوله‌های عاطفی گفتار دارند؛ ۲)تقسیم وزنی زمان (یعنی وزن تکیه‌ای در مقابل وزن هجایی) بر تشخیص نوای عاطفی گفتار تاثیری ندارد.

کلیدواژه‌ها: نوای عاطفی، ارتباط کلامی، جنسیت، تشخیص احساس، زبان فارسی.

Abstract
Proper understanding and processing of a speaker's emotional prosody 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 has not been sufficiently studied and is still under debate. Moreover, the existing literature on gender and emotional prosody has targeted speakers of stress-timed languages (i.e., German and English). However, since prosody is affected by language-specific features such as the rhythmic division of time (i.e., stressed- vs. syllable-timed), the author 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. To do this, 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. The data of this test were examined by availing from logistic regression. In general, the results established that the 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.

Key Words: Emotional Prosody, Verbal Communication, Gender, Emotion Recognition, Persian.

***نویسندهٔ مسئول ـ رایانامه:** n.keshtiari@gmail.com

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 (ie, 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 (eg, 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.

1.1. 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 non-verbal expressions (eg, laughter, screams, and sighs) (Lima, Castro, & Scott, 2013). Emotional prosody is regarded as one of the most basic features of language (Besson, Magne & Schön, 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 cuesis 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.

1.2. 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;

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

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*[2] (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 non-verbal 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, and 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

2. 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.

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 (eg, 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 (eg, 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 (Brody & Hall, 2008; Lima et al., 2013). Besides the methodological procedures (ie, 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. Cultural difference may be the other explanation of these inconsistent findings (Keshtiari et al., 2016). It has been noted that due to the motivational and the adaptive communicative functions of emotions, sociocultural variables mediate, and moderate gender differences in emotional functioning (Brody & Hall, 2008). A number of studies (Markus & Kitayama, 1991; Matsumoto, 1989; Safdar et al., 2009) have revealed that sensitivity to vocal emotional cues vary among the members of collectivistic and individualistic culture (see Keshtiari et al., (2016) for more explanations).

1.3. 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)[3]. 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 stark contrast [4] between stressed-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 detailed description). This distinction in the rhythmic division of time (ie, 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 (eg, Bonebright et al., 1996[5]) 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 (ie, English/German). Besides these two studies, in the meta-analyses

3. 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.

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

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

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; Dauer, 1983).

1.4. 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 the 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.

2. Method

2.1. Participants

Sixty six native speakers of Persian (33 females and 33 males) 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.

2.2. 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, 2015). 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, 2015 for extended details of this database).

The stimulus material selected for this experiment consisted of 12 syntactically similar sentences (ie, 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 speakers. The duration of each utterance was about 4.2 seconds (M = 4170 ms; SD = 232). See Example (1) for a sample sentence.

2.3. 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 (eg, Paulmann et al., 2008; Sauter et al., 2013). Response options corresponded to anger, disgust, fear, happiness, sadness, and an extra response option (ie, neutral). A single practice run including six items was given prior to the start of the experiment. To limit 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. Results

3.1. 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).

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 (ie, 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.

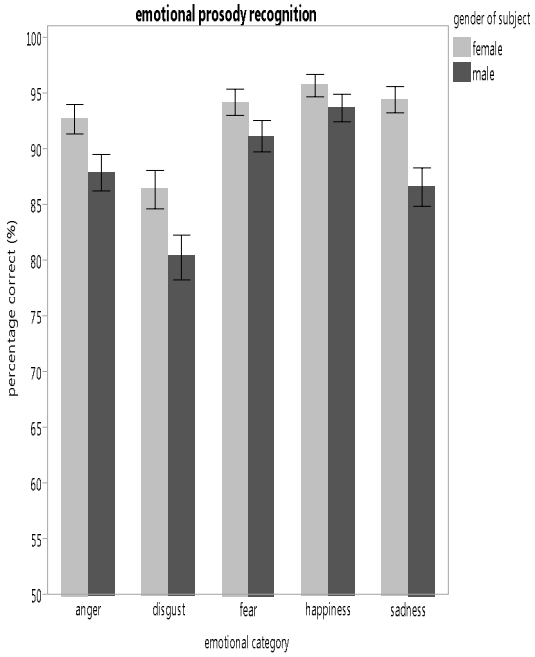

Fig. 1 Accuracy (in %) of emotional prosody recognition according to the gender of the participants (N = 66). Error bars represent the standard errors.

3.2 Nominal logistic regression

Recognition data is often analyzed with methods that require an interval scaled level of measurement and homoscedasticity of data (eg, 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, David, Lemeshow, & Sturdivant, 2013), to obtain more accurate results. Using the software JMP 11 pro, we conducted a logistic nominal regression where 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).

4. Discussion

The current study investigated the relation between gender and recognition of emotional prosody in an underinvestigated 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 (eg, 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.

4.1 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 (eg, Briton & Hall, 1995; Hall et al., 2000; Schirmer & Kotz, 2003; Thompson & Voyer 2014) and with particular findings on the recognition of emotional prosody (eg, 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 non-verbal emotional cues (for a review see Hall, 1978; Hall et al., 2000). It is likewise consistent with 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 ie, females performed better than males in recognizing all of the five intended emotions. Our findings are, thus, in line with those of Hall et al., (2000).

4.2. 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 ie, English (eg, Bonebright et al., 1996) and German speakers (eg, 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 (ie, syllable-timed language) and earlier findings from German and English (ie, stress-timed languages), it seems that the observed gender differences may be relatively independent of rhythmic division of time (ie, stressed- vs. syllable-timed).

4.3. Emotional prosody recognition

Previous research on non-verbal vocalizations and speech prosody has shown that acoustic cues (e g., pitch, intensity, duration) predict subjective emotion judgments (eg, 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 (ie., pitch, intensity, duration), a previous study (Keshtiari et al., 2015) 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., 2015 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 rate (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).

5. 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.

The present study has a number of limitations. The first concerns the 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 makes 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 (eg, 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 (ie, neutral). On the 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, eg, 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 (eg, 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 (ie, 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 underinvestigated syllable-timed language (ie, Persian) as opposed to stress-timed languages (eg, 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 (ie, 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).

6.Acknowledgement

The author would like to express her appreciation to Prof. Gisela Klan-Delius, Michael Kuhlmann, Dr. Silke Paulmann, Dr. Francesca Citron, Dr. Evgeniya Kirilina, Martin Schultze, Dr. James Murphy, Hoseyn Namvar Arefi, Trudi Falamaki-Zehnder, and Pooya Keshtiari for their help with data collection and organization, and manuscript preparation. A special thank you goes to Prof. Hassan Ashayeri, Prof. Yahya Modarresi Tehrani, and Prof. Reza Nilipour without their help the **bureaucratic arrangements** for the data collection in Iran would have been very difficult. This research project was supported by the cluster of excellence 'Languages of Emotion' at Freie Universität Berlin, and funded by the German Research Foundation (DFG).

References

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

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

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

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

Banse, R., & Scherer, K. R. (1996).Acoustic profiles in vocal emotion expression. *Journal ofpersonality and social psychology*, *70*, 614-636.Press.

Bertinetto, P. M. (1989). Reflections on the dichotomy 'stress' vs.'syllable-timing'. *Revue de phonétiqueappliquée*, *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.

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.

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.

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.

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.

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.

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

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

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. (1972). *Universals and cultural differences in facial expressions of emotion. In Nebraska Symposium on Motivation,* J. Cole (ed.), 207–283. Lincoln: University of Nebraska Press.

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

Fraisse, P. (1982). Rhythm and timing. In D. Deutsch (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.

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

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

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

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. *ActaPsychologicaSinica*, *41*(11), 1081-1090.

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.

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

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.

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.

Keshtiari, N., & Kuhlmann, M. (2016). The Effects of Culture and Gender on the Recognition of Emotional Speech: Evidence from Persian Speakers Living in a Collectivist Society. *International Journal of Society, Culture & Language*, 1-16.

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.

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

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

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

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.

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

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

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.

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

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

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.

Oatley, K. (1992). *Best laid schemes: The psychology of the emotions*. Cambridge, UK: Cambridge University 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 PhoniatricaetLogopaedica*, *56*(1), 7-13.

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.

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

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.

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

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*(3), 262-269.

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. *The Journal of the Acoustical Society of America*, *109*(4), 1668-1680.

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

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.

Pittam, J., & Scherer, K. R. (1993).Vocal expression and communication of emotion. *Handbook of emotions*, 185-197.

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

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

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

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.

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

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*(6), 1161.

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 Behavioral Science/Revue Canadienne des Sciences du Comportement*, 41(1), 1-10.

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.

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.

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

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research.Psychological Bulletin, 99, 143–165.

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

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.

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

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

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

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

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.

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.

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

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

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.

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.

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

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 University Press.

Windfuhr, G. L. (1979). *Persian grammar: History and state of its study* (Vol. 12). Walter de Gruyter.

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*(10), 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.