The influence of alexithymia and sex in the recognition of emotions from visual, auditory, and bimodal cues

Diana S. Cortes

Handledare: Håkan Fischer
MASTERUPPSATS I PSYKOLOGI, 30 HÖGSKOLEPOÄNG, 2013

STOCKHOLMS UNIVERSITET
PSYKOLOGISKA INSTITUTIONEN
THE INFLUENCE OF ALEXITHYMIA AND SEX IN THE RECOGNITION OF EMOTIONS FROM VISUAL, AUDITORY AND BIMODAL CUES

Diana S. Cortes

Alexithymia is a personality trait associated with impairments in emotional processing. This study investigated the influence of alexithymia and sex in the ability to recognize emotional expressions presented in faces, voices, and their combination. Alexithymia was assessed by the Toronto Alexithymia Scale (TAS-20) and participants ($n = 122$) judged 12 emotions displayed unimodally or bimodally in two sensory modalities as measured by the Geneva Multimodal Emotion Portrayals Core Set (GEMEP-CS). According to their scores, participants were grouped into low, average, and high alexithymia. The results showed that sex did not moderate the relationship between alexithymia and emotional recognition. The low alexithymia group recognized emotions more accurately than the other two subgroups, at least in the visual modality. No group differences were found in the voice and the bimodal tasks. These findings illustrate the importance of accounting for how different modalities influence the presentation of emotional cues, as well as suggesting the use of dynamic instruments such as GEMEP-CS that increment ecological validity and are more sensitive in detecting individual differences, over posed techniques such as still pictures.

Social interaction is essential for humans and the ability to accurately perceive and interpret feelings, intentions, and motivation of others is central for the success of this interaction. This ability is also central for properly responding and predicting behaviors of others, leading to the adjustment of our own behavior (Ebner, Johnson & Fischer, 2012). For example, emotional cues are often displayed in nonverbal behaviors like facial and vocal expressions, considered as primary sources that provide specific cues for recognizing the affective states of others (Scherer & Scherer, 2011). Correctly reading emotional reactions of others is fundamental for professions which involve face-to-face interactions (Scherer & Scherer, 2011), such as health professionals, policemen, teachers, etc. The inability in some individuals in perceiving, understanding, describing, and evaluating emotional expressions has been named as alexithymia (Sifneos, 1973). Research regarding facial expressions and alexithymia is well documented, whereas research with vocal cues and the combination of facial and vocal cues is still emerging. The focus of the present study is to investigate alexithymia in relation to emotional recognition by way of separate cues (unimodal) and combining facial and vocal cues (bimodal).

**Alexithymia**

The word alexithymia was introduced in the early 1970’s by the psychotherapist Peter Sifneos and derives from the Greek words *lexis* and *thumos*, meaning “no words for feelings” (Sifneos, 1973). Alexithymia encompasses difficulties in the following aspects: a) identifying and describing own subjective feelings which can also lead to a lack of understanding feelings of others; b) distinguishing between feelings and the bodily sensations of emotional arousal; c) limited imagination capacities (lack of fantasies); and d) an externally oriented cognitive
style meaning a too concrete, realistic and rational thinking (lack of affective thinking) (Taylor, Bagby & Parker, 1997; Taylor, 2000). Emotion regulation is thus affected due to deficits in cognitive processing of emotional experiences and failure to communicate affective states to others. These can lead to overwhelming attention towards body distress and physical illness, increasing the reaction and responses of the autonomic nervous system and neuroendocrine system, which in turn might end in somatic complaints (Simonsson-Sarnecki, 2001; Taylor, 2000). For example, it could represent a challenge for individuals with alexithymia to define whether they are tired, hungry, sad, happy or ill (Taylor et al., 1997).

The concept of alexithymia was first observed in psychosomatic patients and other psychiatric disorders; however, it is not considered a mental disorder. Alexithymia is rather regarded as a multidimensional personality construct that differs in each individual and which can be observed also in nonclinical populations (Grynberg, et al., 2012). There is a 10 percent prevalence of alexithymia in healthy individuals (Grynberg, et al., 2012; Zhang, et al., 2012); the prevalence for men is 9-17 percent compared to a 5-10 percent for women (Lane, Sechrest & Riedel, 1998; Mattila, Aholac, Honkonenc, Salminend, Huhtalaav & Joukamaav, 2007). Specifically, alexithymia is highly represented in males, older age, lower socioeconomic status, poor education, and poorer perceived health (Lane, et al., 1998; Mattila, et al., 2007; Sonnby-Borgström, 2009).

In clinical populations, there are five disorders commonly associated with high alexithymia scores: eating disorders, autism spectrum disorders, substance abuse, panic disorders, and somatoform disorders (Grynberg, et al., 2012). Due to the comorbidity of these disorders with anxiety and depression, it has been questioned if alexithymia is a stable trait or if it merely acts as a parallel reaction to any disorder (Taylor, Bagby & Luminet, 2000). Both in clinical and nonclinical samples, alexithymia can be considered a stable factor which does not depend on a specific medical or psychiatric disorder or psychological distress (Grynberg et al., 2012; Montebarocci, Surcinelli, Rossi & Baldaro, 2011; Taylor, et al., 2000). On the other hand, due to the significant associations between alexithymia and mood disorders, it has been recommended to statistically control for affect, depression, and anxiety when studying alexithymia (Grynberg, et al., 2012; Honkalampi, Hintikka, Saarinen, Lehtonen & Viinamaki, 2000; Mattila, et al., 2007).

Emotional recognition

Faces and voices with varying emotional expressions have extensively been used as cues to examine the ability to recognize another person’s affective states. Elfenbein and Ambady (2002) conducted a meta-analysis examining if emotional expressions can be regarded as universal or as culture-specific, and concluded that facial emotions and other nonverbal cues, such as the voice, are universally recognized well above chance levels. Elfenbein and Ambady (2002) found that some emotions are better understood depending on the modality in which emotions are presented (whether there were facial or vocal emotional expressions). Happiness was the most accurately recognized emotion in the face, but the least accurately recognized emotion in the voice, and anger showed the opposite direction since it was the best recognized in voice but less well understood as facial expression (Elfenbein & Ambady, 2002). Further research regarding the differences and similarities displayed in facial and vocal expressions, is then needed.

Individuals with alexithymia are less able to recognize affective states from both facial and vocal expressions (Goerlich, Aleman & Martens, 2012; Goerlich, Witteman, Aleman & Martens, 2011; Grynberg et al., 2012). For example, Lane and colleagues (1996) compared
participants with high scores in alexithymia and those with low scores, in matching verbal with non-verbal emotional cues. Individuals with higher scores in alexithymia were less accurate on recognizing all basic emotions in all the tasks, meaning that the deficits in emotional processing are independent of using language to describe emotions (Lane, Sechrest, Riedel, Weldon, Kaszniak, & Schwartz, 1996). On the other hand, Montebarocci and colleagues (2011) demonstrated that after taking into account anxiety and depression, individuals with higher scores in alexithymia tended to perform worse on labeling facial emotional expressions. However, when controlling for verbal abilities, there was no difference in emotion recognition between high and low scores in alexithymia (Montebarocci et al., 2011).

Diminished perception in negative emotions like sadness, anger, and fear has also been investigated. For example, Prkarchin, Casey and Prkarchin (2009) suggested that individuals with high scores in alexithymia are worse at detecting sadness, anger, and fear than at detecting happiness, surprise, and disgust. A similar pattern of results was observed by Parker, Prkachin and Prkachin (2005) who found that differences in emotional recognition between high versus low scores in alexithymia are more prone when emotional cues are presented during 1 second or less.

In contrast to the visual modality (facial cues), the auditory modality (vocal cues) has rarely been investigated in relation to alexithymia and results have been inconsistent. Swart, Kortekaas and Aleman (2009) compared emotional processing in subjects with high and low scores in alexithymia in both the visual and the auditory modality. Participants were tested in a prosody task where they had to label which emotion was expressed in a 20 seconds statement, either by the tone of the voice or by its content, finding no group differences or deficiencies in the processing of emotional language. In line with this reasoning, Goerlich and colleagues (2011) found no significant results when participants listened to happy and sad pseudo-words for 600ms. Conversely, in a more recent study Goerlich and colleagues (2012) employed an auditory oddball paradigm with neutral, happy, sad, angry, and disgust tones in brief vocal cues also presented during 600ms. Small significant differences were found between women scoring high and low in alexithymia, mainly for disgust.

There are very few studies that have investigated the combination of how emotions are perceived both by the auditory and the visual modalities, not just taking into account one of these sensory modalities. Furthermore, no studies regarding alexithymia and bimodality were found.

De Gelder and Vroomen (2000) explored the bimodal recognition of emotions by exposing participants to still, black and white morphed photographs of faces depicting the extremes on the continuum between sad and happy emotions. The photographs were also paired together with the recording of a sentence pronounced either as a happy or as a sad tone. Participants were tested in video-only, audio-only, and bimodal condition (video+audio). The results demonstrated the influence of bimodality: participants were faster at recognizing the emotions when combining both visual and auditory modalities. To further explore the effects of bimodality, De Gelder and Vroomen (2000) conducted a second and a third trial where they instructed the participants to ignore either the facial cue or the vocal cue (in the second trial the participants unattended the vocal cue and only judged the facial cue while in the third trial, participants had to attend only the vocal cue and ignore the facial one). The results suggest that the participants’ perceptions were influenced by the vocal/facial cue even though they had unattended one of such modalities and that a multimodal integration still occurs.
A more recent study by Paulmann and Pell (2011) came to similar findings suggesting that individuals recognize emotions more accurately when using two modalities. In addition, Paulmann and Pell (2011) tested a multimodal condition which involved the presentation of facial cues, prosody and semantics at the same time, and the results indicate towards increased recognition of emotional expressions.

*Emotional recognition and sex*

Research regarding sex differences shows a female advantage when it comes to recognition of facial cues (Kret & De Gelder, 2012). Nevertheless, there is discrepancy around the size of the advantage and whether it is the same for all emotions. For instance, Hoffman and colleagues (2010) emphasized that women tend to perceive subtle emotional cues more accurately than men. Once the intensity of the cue is incremented, there are no longer significant sex differences (Hoffman, Kessler, Eppel, Rukavina & Traue, 2010). Similarly, a meta-analysis found that women appear to recognize emotions better than men (Kret & De Gelder, 2012); however men displayed greater responses to threatening signals.

It has also been hypothesized that sex differences on emotional processing are moderated by personality factors and their severity levels (Campanella, et al., 2012), for example, higher prevalence of alexithymia in men and twice prevalence of depression and anxiety in women than men. In a study by Campanella and colleagues (2012), participants were matched so that there were no significant differences between sex groups on age, anxiety, depression, and alexithymia scores. The results show no main or interaction effect of sex and once the personality variables were controlled for, sex differences in the detection of happy and fearful faces disappeared. Campanella and colleagues (2012) also found that alexithymia and depression predicted respectively, N2 and P3b event-related potential latencies (while sex and anxiety did not).

*Dynamic cues*

Most of the research in emotion recognition has used posed photographs as facial cues which can be perceived as unchallenging, unnatural, lacking movement and thus, leaving aside the dynamic aspects of emotions (Krumhuber, Kappas & Manstead, 2013). In contrast, dynamic cues are assumed to be greater ecologically valid, reflecting real-life and daily situations. Dynamic cues are often seen as an extension of still images and the framework consists in identifying facial expressions in short clips in contrast to using posed images of faces. It has also been suggested that employing dynamic clips increase the coherence in the perception of affect, give advantage to discriminate between genuine and fake expressed emotions and consequently, are more susceptible to detect individual differences (Aviezer, Trope & Todorov, 2012; Bänziger, Mortillaro & Scherer, 2012; Bänziger & Scherer, 2010; Krumhuber, et al., 2013; Paulmann & Pell, 2011). It has also been hypothesized that angry and happy static pictures recruit different neural pathways than those of a dynamic framework (Kilts, Egan, Gideon, Ely & Hoffman, 2003). Aware of the importance of dynamic cues, Bänziger and colleagues (2012) developed the *Geneva Multimodal Emotion Portrayals Core Set*, an instrument that fulfills such features by including all three modalities: audio-only, video-only (dynamic clips) and video-audio (see the method section for further details). This investigation group is now additionally developing a whole body modality too.

Research about the ability to recognize emotions is well documented, especially the visual modality as assessed by static facial expressions. Only a few studies have investigated the auditory modality and research about the combination of these two modalities is scarce, above all in alexithymia, where no studies using an interaction of facial and vocal cues were found.
Therefore, the main aim of the present study was to examine to what extent individuals with alexithymia benefit from presenting bimodal emotional cues (combining the visual and the auditory modalities) using a dynamic framework. In addition to shed light to this aspect, sex differences were highlighted. A more specific aim of the current study was to explore the effects of alexithymia and sex in the ability to recognize emotional expressions as depicted in short video-clips either combining facial and vocal cues or presenting them apart. Based on the presented theoretical reasoning, the following questions guided the present work:

1) Is there a difference in emotional recognition for subjects with low, average and high alexithymia scores, and if so, are these differences similar or different for unimodal and bimodal emotional recognition?

2) What is the impact of sex and alexithymia on emotional recognition? Does sex moderate the relationship between alexithymia and emotional recognition?

Method

Participants
One hundred eighty seven young individuals (116 women, 71 men) were recruited through posters advertised either at different colleges and universities in the Stockholm area or through the website www.studentkaninen.se. The participants had a mean age of 23.31 (sd = 3.36; range 18-34) years. The mean age for the men was 23.72 (sd = 3.31) years and for the women the mean age was 23.06 (sd = 3.37) years. All participants were right-handed, healthy, and fluent in Swedish and reported neither previous/current substance abuse nor psychiatric disease. Six participants were excluded due to incomplete data in the alexithymia scale.

Participants were divided into three groups according to their total scores in the Toronto Alexithymia Scale (TAS; Bagby, Parker & Taylor, 1994; Swedish version: Simonsson-Sarnecki et al., 2000) from the original sample. This division into groups was made following the cut-off scores from other studies (Goerlich et al., 2012) where individuals with average scores were additionally included in order to ensure a broad and continuous spectrum of alexithymia scores. The low alexithymia group (TAS ≤ 35; n = 19 women, 12 men), the average alexithymia group (scores 40 - 48; n = 35 women, 25 men) and the high alexithymia group (TAS ≥ 57; n = 22 women, 9 men). Affectivity, depression and anxiety were also assessed in order to control participants’ differences in these variables.

In exchange for participating, three movie vouchers or course credits were given. The study was approved and conducted in accordance with guidelines established by the regional ethics committee, and written consent was obtained from all participants before starting the study.

Materials and Stimuli

Toronto Alexithymia Scale (TAS-20).
Participants completed the Swedish version of the Toronto Alexithymia Scale (TAS-20; Simonsson-Sarnecki et al., 2000), a 20-item self-report instrument assessing the severity of alexithymia. The items are rated on a five-point Likert scale, where 1 means “totally disagree” and 5 indicates “totally agree”. The total scores range from 20 to 100, higher scores reflecting greater alexithymia. The TAS-20 items load onto three factors: (1) Difficulty Identifying
Feelings (DI) which includes seven items, i.e., “I am often confused about what emotion I am feeling”; (2) Difficulty Describing Feelings (DD) which contains five statements, i.e., “It is difficult for me to find the right words for my feelings”; and (3) External-Oriented Thinking (EOT) which contains eight items, i.e., “I prefer to analyze problems rather than just describe them” (Bagby, et al., 1994).

The internal consistency (Cronbach’s alpha) of the total scale has been reported to be 0.81 and the three factors reliabilities are DI = 0.78, DD = 0.75 and EOT = 0.66, while the test-retest reliability is 0.77 at p < 0.01 (Bagby, et al., 1994). The Swedish translation also shows good internal reliability (Cronbach’s alpha = .83) and confirmatory factor analysis (in a student sample, n = 161) has revealed the same three factor structure (Simonsson-Sarnecki et al., 2000; Sonnby-Borgström, 2009). In the current study, the Cronbach alpha coefficient was .81.

**Positive Affect and Negative Affect Schedule (PANAS).**
Affectivity was assessed by the Positive Affect and Negative Affect Schedule (PANAS; Watson, Clark & Tellegren, 1988) which is a 20-item self-report measure; 10 items for positive affect and the other 10 depicting negative affect. High scores in positive affect reflect pleasure engagement, disposition, enthusiasm, full concentration, and high energy. High scores in negative affect are often related to sadness, subjective distress, nervousness, lethargy, and unpleasurable engagement. Participants are asked to rate the extent to which they have experienced each particular emotion within a specified time period, with reference to a five-point Likert scale where 1 means “not at all” and 5 indicates “very much”. A number of different time-frames have been used with the PANAS, but in the current study the adopted time-frame was “right now, at the present moment”.

**Hospital Anxiety and Depression Scale (HADS).**
Participants completed the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983). It consists of 14 items, seven of which measure depression, and the other seven anxiety. Each item on the questionnaire is scored from 0-3 and the total scores range from 0 to 21, higher scores reflecting either greater depression, anxiety or both of them.

**Geneva Multimodal Emotion Portrayals Core Set (GEMEP-CS).**
Emotion recognition was assessed by the Geneva Multimodal Emotion Portrayals Core Set (GEMEP-CS; Bänziger, et al., 2012). The GEMEP-CS is a dynamic multimodal expression corpus that contains 145 audio and colored video clips performed by ten professional theater actors (5 women, 5 men) depicting different emotions, all of which are presented in different modalities (audio, video, and audio-video). The meaning of each emotion is explained to the actors and three possible scenarios on how to display each emotion are given to the actors a couple of weeks before recording.

To record the cues, the actors performed the emotional expressions in collaboration with a director. This interactive setting between actor and director is preferred over prototype posing techniques since the emotions are dynamically displayed. Facial expressions from the actors were zoomed in with the help of digital cameras, with the head oriented towards the expressers; only the face and the upper body are visible (720x576 at 25 fps). Sound was recorded with professional microphones placed in the cameras and in the left ear of the actors (Bänziger, et al., 2012). The vocal expressions consisted of two meaningless sentences and one brief utterance. One of the pseudo-sentences (nekal ibam soud molen!) was expected to express an exclamation meaning “I cannot believe it!” and the other pseudo-sentence (koun se mina lod belam?) which was articulated as a question meant to express “is it really your
opinion?” A professional phonetician suggested the mentioned combination of phonemes because they can be pronounced in a similar way in a number of Western languages. The brief utterance involved a sustained vowel (/aaa/) without an established length (Bänziger, et al., 2012, p. 1168).

A total of 72 audio, video and audio-video clips were selected from the GEMEP-CS database for the current study; thus each modality comprised 24 cues plus one example of each. The portrayed emotions were grouped into positive valence: pride, joy, pleasure, relief, interest and negative valence: rage, fear, despair, irritation, anxiety, sadness, disgust. Each emotion was portrayed six occasions by different actors. After the brief presentation of the cues (2 seconds approx.), a list containing the 12 emotions were presented and participants must choose the emotion (only one is allowed) that best represents the expressed emotion in the cue. Test duration was approximately 15 to 20 minutes (Bänziger et al., 2012).

Procedure
All data was recollected in a laboratory at Stockholm’s University which was equipped with four desktop computers. Participants were contacted by e-mail and asked to choose the day and hour for performing the tests; sometimes participants came in groups of four while on other occasions there was only one single participant at a time. Participants were first required to sign a written consent, along with a description of the study. After that, demographic data was obtained.

The trial began with the GEMEP-CS. The first screen of the GEMEP-CS showed the instructions for the test, followed by three examples (one audio, one video and one audio-video). Afterwards, a message saying that the examples were over and that the test was beginning was displayed. Images were presented on 22 inch LED computer screens and the audio clips were presented with a pair of AKG 618 professional headphones connected to the same computer. The use of headphones was mandatory and the volume was set at 67 percent. The cues were presented in the following order: first the 24 video clips (only image) followed by the 24 audio clips (only audio) and the 24 audio-video clips (image+sound), making a total of 72 recordings. After each clip, participants were asked to choose the emotion that best represented what the actor had just expressed. There was no time limit to choose an answer.

The next part of the study consisted of administering the self-report measures. The questionnaires were computer-based and completed via LimeSurvey. Leaving on the headphones while answering the self-report measures was optional. Participants completed first the PANAS, then the TAS-20 and lastly the HADS. Movie vouchers or course credits were given as compensation for participating in the study. All collected data was anonymized, treated confidentially and used only for research purposes.

Results
A 3x2x3 multivariate analysis of variance (MANOVA) was performed in order to explore differences between the low, average and high scores in alexithymia in the ability to recognize emotions. Sex and scores in alexithymia (low, average and high) were included as predictor variables. The scores of emotional recognition in each modality were introduced as criterion

---

1 Bänziger and colleagues (2012) did not specify the duration of the cue, the average of 2 seconds comes from an article by Yang and Bhanu (2012).
variables (i.e. visual, auditory and bimodal). One-way analysis of variance (ANOVA) and hierarchical multiple regressions were also performed in order to further investigate the results. An alpha level of .05 was used in the analyses and missing values were handled by using case-wise deletion. Effect sizes were denoted by partial eta squared ($\eta^2_p$) or eta squared ($\eta^2$) for the ANOVAs. All the analyses were conducted following the criteria and recommendations proposed by Pallant (2010). Means and standard deviations for performance in alexithymia, mood as well as the accuracy scores for emotion recognition are presented in Table 1.

<table>
<thead>
<tr>
<th>TAS Groups</th>
<th>Low (n=31)</th>
<th>Average (n=60)</th>
<th>High (n=31)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>TAS</td>
<td>30.81</td>
<td>3.67</td>
<td>44.03</td>
</tr>
<tr>
<td>PA</td>
<td>35.32</td>
<td>5.70</td>
<td>32.22</td>
</tr>
<tr>
<td>NA</td>
<td>13.61</td>
<td>3.37</td>
<td>13.80</td>
</tr>
<tr>
<td>HADS Anxiety</td>
<td>18.65</td>
<td>2.02</td>
<td>18.68</td>
</tr>
<tr>
<td>HADS Depression</td>
<td>15.84</td>
<td>1.34</td>
<td>15.73</td>
</tr>
<tr>
<td>Visual</td>
<td>13.81</td>
<td>3.70</td>
<td>12.72</td>
</tr>
<tr>
<td>Auditory</td>
<td>11.50</td>
<td>3.15</td>
<td>11.81</td>
</tr>
<tr>
<td>Bimodal</td>
<td>15.71</td>
<td>4.74</td>
<td>14.80</td>
</tr>
</tbody>
</table>

*Note.* SD: standard deviation; TAS: Toronto Alexithymia Scale; PA: Positive Affect; NA: Negative Affect; HADS: Hospital Anxiety and Depression Scales.

The data fulfilled all the assumption validations needed in a MANOVA, that is, no serious violations of normality, linearity, outliers, homogeneity of variance-covariance matrices, and multicollinearity were found. There were no statistically significant results for sex in neither the unimodal nor the bimodal conditions. On the other hand, there was a statistically significant multivariate difference between the alexithymia groups $F(6, 222) = 3.81, p = .001$; Wilks’ Lambda = .82; $\eta^2_p = .093$. The univariate results showed that the only difference to reach statistical significance was the visual modality, $F(2, 113) = 7.22, p = .001, \eta^2_p = .113$. Looking closer at the differences between the mean scores, these indicated that low scores in alexithymia present higher accuracy levels of emotional recognition in the visual modality ($m = 14.27, sd = 2.71$) than average group ($m = 13.16, sd = 2.95$) and high scores in alexithymia ($m = 11.26, sd = 2.88$) (see Figure 1).

To further assess the effect of alexithymia on the visual modality, one-way ANOVA was conducted introducing the visual scores as criterion variable and alexithymia (i.e. low, average, and high) as predictor variable. There was a statistically significant difference at the $p < .05$ level in the visual modality for the three alexithymia groups: $F(2,119) = 4.05, p = .02$. The effect size was $\eta^2 = .06$, which can be considered as medium. Comparisons indicated a significant difference between the low alexithymia group and the high alexithymia group; $t_{119} = 2.83, p = .005$. The average group did not differ significantly from neither the low ($t_{119} = 1.39, p = .167$) nor the high ($t_{119} = 1.86, p = .065$) alexithymia groups. In summary, these data shows that the participants in the low alexithymia group were more prone to recognize emotional expressions only when presented in the unimodal visual modality.
Figure 1. Means of the recognition accuracy scores for the visual ($m = 12.94$, $sd = 3.06$), auditory ($m = 11.50$, $sd = 3.05$), and bimodal ($m = 15.60$, $sd = 3.28$) conditions as a function of alexithymia scores. A significant difference was found between the low alexithymia and the high alexithymia group (* $p < .05$) for the unimodal visual condition. Error bars +/- 2 SE.

A two stage hierarchical multiple regression analysis was conducted to further assess the predictive value of alexithymia in the visual modality, after controlling for the influence of sex, affectivity, anxiety and depression. Positive and negative affect as well as anxiety and depression were entered at step 1. Alexithymia and sex were entered at step 2. The results revealed that at step 1, positive affect, negative affect, anxiety, and depression accounted for 1.6% of the variance in the visual modality, $F(4, 116) = .46$, $p < .05$. Introducing sex and alexithymia at step 2 explained an additional 9.5% of variation in visual modality scores and this change in $R^2$ was significant $F(2, 114) = 6.08$, $p < .001$. The only statistically significant predictor of visual modality was alexithymia ($beta = -.289$, $p < .001$), which uniquely explained 7.2% of the variance in visual modality. Together, the six independent variables accounted for 11.1% of the variance in visual modality. In sum, after controlling for sex, positive affect, negative affect, anxiety, and depression, alexithymia remained a stable predictor of visual modality.

**Discussion**

The present study aimed to examine how alexithymia and sex influence the ability to recognize emotional expressions displayed in facial and audio clips, as well as in the combination of these two. Additionally, the main purpose was to explore how low, average and high alexithymia groups differ in each of the unimodal and the bimodal conditions.

The first research question of this study asked whether low, average and high alexithymia groups would differ in the ability to identify emotions and if such differences would appear in
all three modalities (the two unimodal and the bimodal one). According to the present results, no statistically significant differences were found between the average and the high alexithymia scores, and nor were there any differences between the average and the low groups. Yet, participants in the low alexithymia group were more prone to recognize emotional expressions in the visual modality than in the auditory or in the bimodal condition. A common finding on the visual modality and alexithymia is that there are more impairments for the high alexithymia group (Grynberg et al., 2012; Lane et al., 1996; Montebaroceci et al., 2011; Parker et al., 2009; Prkarchin et al., 2009). This holds true even when using different emotions, tasks, and thresholds, as well as after controlling for depression, anxiety, affectivity.

Since the majority of the studies reviewed here employed posed pictures of faces in contrast to the use of dynamic clips as in the present study, the findings obtained here attempt to augment a general understanding concerning the deficits of emotional recognition in alexithymia. Rather surprisingly, all of the research points to the same conclusion, meaning that even when using different instruments, techniques (posed versus dynamic), type of emotions, number of cues, sample size, and grouping participants in similar but not the same way (different alexithymia cutoff scores and including an average group), the results still show a disadvantage for visually displayed emotions for individuals with higher scores in alexithymia.

Regarding the auditory and the bimodal displayed emotions, the results showed that the alexithymia groups did not differ from each other, which is similar to previous findings (Goerlich et al., 2011; Swart et al., 2009). The explanations used in the visual modality can also be extended towards the auditory and bimodal conditions, which involve the duration of the cues. It has been hypothesized that participants who score high in alexithymia often need more time to understand and correctly identify emotions and that these impairments are more evident if the duration of the presented cue is 1 second or less. Unfamiliarity to emotions or being uninterested in them can be one of the motives behind why people with alexithymia need more time to accurately perceive emotions (Grynberg et al., 2012). The cue duration average in the instrument used in the present study, the GEMEP-CS, is 2 seconds, which might explain why alexithymia-related differences were not found.

This leads to the question whether emotional cues are better recognized in certain modalities. Paulmann and Pell (2011) suggested that information presented in the visual modality seems to be recognized more effectively than in the auditory modality and when cues are displayed in more than one modality, faster and more accurate recognition rates are expected. However, it should not be presumed that bimodal emotional cues will automatically lead to better recognition rates since it has been argued that the bimodal advantage is only present when compared with the auditory modality, and the advantage remains unclear as far as the visual modality is concerned (Bänziger, Grandjean & Scherer, 2009). This can be considered as an alternative interpretation of the present results where low, average and high alexithymia scorers did not differ in the auditory or bimodal conditions, but where the low alexithymia group did recognize emotional cues in the visual modality as compared to the high alexithymia group.

The second research question of this study concerned the effect of sex and alexithymia in emotional recognition. The present results indicated that sex do not moderate the relationship between alexithymia and the perception of emotions. Besides, no sex effect was found in any of the three modalities. The results of the present study concur with those found by
Campanella and colleagues (2012) who investigated the combined effects of sex and personality traits (alexithymia, depression and anxiety) on emotional processing. Their results indicated that after controlling for these personality variables, sex differences disappear, showing a significant effect of above all, alexithymia and depression. Campanella and colleagues (2012) reached these results using both behavioral and event-related potential data. Overall, their data support the notion that sex differences in the perception of emotions may be due to intrinsic female/male personality traits. Since the present study involved the recognition of twelve different emotions, the results obtained here attempt to lend further support to those of Campanella et al. (2012) – i.e. that alexithymia is a better predictor of emotional recognition than sex – despite the fact that in this study group membership was determined by alexithymia rather than by obtaining groups with no sex differences on personality traits.

The fact that there is no statistically significant sex effect in the visual, auditory or bimodal conditions is in line with previous findings (Borod et al., 2000) where sex was not correlated with different facial, prosodic and lexical emotional tasks. Borod and colleagues (2000) explained that their findings support the notion of a general emotional processor for the identification of emotional cues across modalities that are to a certain extent independent from cognitive processes and demographic variables such as sex. Borod and colleagues (2000) also suggested the importance of controlling for other variables like mood disorders that could possibly influence emotional perception. Previous research suggest that mood disorders and alexithymia might share a common variance even though they are different constructs (Grynberg et al., 2012). This was also the case in the present study where affectivity, anxiety and depression were controlled together with alexithymia and sex in order to assess the predictive value that these variables had on the visual modality (the variables were only controlled in this modality because alexithymia groups did not differ in the other modalities). The present results showed no statistically predictive value for affect, anxiety and depression which would suggest that alexithymia is linked to impairments in perceiving emotions independently of having certain characteristics of mood disorders. Verbal IQ has also been mentioned as another important confounding variable that could influence the relationship between alexithymia and emotional recognition (Montebarocci et al., 2011), however this study did not include any measure to assess verbal ability. Future studies should also consider controlling for verbal ability too.

The present results show that sex is not a confounding variable when it comes to alexithymia and the ability to perceive emotions. This is also partially in line with the study of Hoffman and colleagues (2010) which demonstrated that there tends to be no differences between men and women in the accuracy of recognition when facial cues are presented at a high intensity level (80-100%). Hoffman and colleagues (2010) used posed pictures of faces that were morphed from a neutral image throughout different intensities until reaching a full-blown face. In contrast, the present study employed dynamic videos showing expressions portrayed by professional actors which are assumed to be more like everyday situations and interactions, hence incrementing the ecological validity. This dynamic approach towards presenting multimodal emotion expressions is believed not only to increase the authenticity of the emotions (Bänziger et al., 2012), but is also likely to display major expressive variations rather than almost identical and rigid prototypical emotions (as in the case of posed techniques).

Subsequently, it can be argued that the lack of significant differences between men and women holds true when facial cues are presented at more defined levels, as shown by
Hoffman and colleagues’ (2010) study where they found that women are more accurate than men in perceiving emotions at lower intensities (40-70%). However, Motley and Camden (1988) revealed that facial affect expressions in daily interactions are spontaneous and vary from low to mid intensity. As mentioned above, the GEMEP-CS, is based on a dynamic approach purportedly displaying spontaneous expressions of emotions. Consequently, the results obtained here through dynamic frame-working can perhaps shed some light to the role of sex in the ability to recognize emotions.

Several methodological aspects of the present study should be kept in mind. For instance, the sample used here compromised only young healthy individuals who on average are highly functional and therefore the results may not be generalizable to clinical populations or to older individuals. Sample size is essential too; here the high alexithymia group involved only 9 men compared to 22 women, suggesting that future studies should also try to include a sample that contains a more or less equal number of males and females similarly distributed in the different alexithymia groups. Another noteworthy issue, true to several alexithymia researches, is the fact that samples are not grouped according to the same cutoff scores in the TAS scale. These types of studies have commonly used the cutoff recommended by Taylor and colleagues (1997) however this division leads to the dichotomization of alexithymia into low and high scores. Future studies should consider grouping alexithymia into three levels (low, average and high) in order to obtain more data to compare and to ensure a broad and continuous range of alexithymia scores or to use the alexithymia scores as a continuous variable in the analysis. Besides, no other studies analyzing alexithymia in bimodal conditions were found, meaning that future research in both this area and in the auditory modality is needed as well.

To some extent the shortcomings in the ability to recognize emotions that people with alexithymia experience has already been investigated, albeit without employing dynamic techniques or comparing alexithymia in three different, yet complementary conditions. Moreover, the GEMEP-CS database has a section in which emotions are expressed with complete body cues, which also would be interesting to investigate in the context of alexithymia. The present results were obtained when analyzing all 12 emotions; however, the question is whether the results hold true when looking at specific emotions (i.e. just fear, joy, surprise, disgust, etc.). Exploring specific emotions could also lead to determine if certain emotions are preferentially perceived through different modalities. Other aspects that future research needs to address are different duration of cues, taking into account the time rate, as well as the use of physiological measures, tracking eye movement and brain imaging.

In conclusion, the results of the present study support earlier findings about the role of alexithymia on emotional recognition (Campanella et al., 2012; Grynberg et al., 2012) and exemplify how different modalities might influence the presentation of emotional cues. Most importantly, the results reveal that people with low scores in alexithymia recognize emotions more accurately in the visual modality, than both people with average and high alexithymia. However, these differences were not seen in the auditory or the bimodal conditions. In addition, sex does not seem to moderate the relationship between alexithymia and the perception of emotions. No sex effect was found in the unimodal and the bimodal conditions. Future research should target the use of dynamic instruments such as GEMEP-CS that increment ecological validity and are more sensitive in detecting individual differences, over posed techniques such as still pictures.
References


