Emotional prosodic processing in auditory hallucinations

T.L. Shea a,b,*, A.A. Sergejew a,b, D. Burnham c, C. Jones c,d, S.L. Rossell b, D.L. Copolov b, G.F. Egan a,e

a Centre for Neuroscience, University of Melbourne, Parkville Vic 3052, Australia
b Mental Health Research Institute of Victoria, Locked Bag 11, Parkville Vic 3052, Australia
c MARCS Auditory Laboratories, University of Western Sydney, Bullecourt Avenue, Milperra NSW 2214, Australia
d School of Education, University of New South Wales, Sydney NSW 2052, Australia
e Howard Florey Institute, University of Melbourne, Royal Parade, Parkville Vic 3052, Australia

Received 29 May 2006; received in revised form 29 September 2006; accepted 29 September 2006
Available online 14 November 2006

Abstract

Deficits in emotional prosodic processing, the expression of emotions in voice, have been widely reported in patients with schizophrenia, not only in comprehending emotional prosody but also expressing it. Given that prosodic cues are important in memory for voice and speaker identity, Cutting has proposed that prosodic deficits may contribute to the misattribution that appears to occur in auditory hallucinations in psychosis. The present study compared hallucinating patients with schizophrenia, non-hallucinating patients and normal controls on an emotional prosodic processing task. It was hypothesised that hallucinators would demonstrate greater deficits in emotional prosodic processing than non-hallucinators and normal controls. Participants were 67 patients with a diagnosis of schizophrenia or schizoaffective disorder (hallucinating = 38, non-hallucinating = 29) and 31 normal controls. The prosodic processing task used in this study comprised a series of semantically neutral sentences expressed in happy, sad and neutral voices which were rated on a 7-point Likert scale from sad (−3) through neutral (0) to happy (+3). Significant deficits in the prosodic processing tasks were found in hallucinating patients compared to non-hallucinating patients and normal controls. No significant differences were observed between non-hallucinating patients and normal controls. In the present study, patients experiencing auditory hallucinations were not as successful in recognising and using prosodic cues as the non-hallucinating patients. These results are consistent with Cutting’s hypothesis, that prosodic dysfunction may mediate the misattribution of auditory hallucinations.

© 2006 Published by Elsevier B.V.

Keywords: Emotional prosody; Auditory hallucinations; Schizophrenia

1. Introduction

Auditory hallucinations are a common and frequently distressing symptom of schizophrenia. It is generally accepted that auditory hallucinations are internal events such as inner speech or thoughts that have been misattributed to an external source. While the mechanisms of this misattribution process have yet to be explained, it has been proposed that a deficit in source monitoring (e.g. Johnson et al., 1993; Keefe et al., 1999) or reality monitoring (Bentall, 1990) might lead to the misattribution of auditory hallucinations. Source monitoring is defined as the ability of individuals to determine
the origins of their thoughts, memories and beliefs (Johnson et al.) while reality monitoring is a specific form of source monitoring that enables an individual to discriminate between memories for perceived events and memories for thoughts and imaginations (Johnson and Raye, 1981).

Bentall (1990) has proposed that auditory hallucinations are experienced as external events because patients fail to recognise internal experiences as belonging to the self. Accordingly, some studies investigating source monitoring in schizophrenia have reported that patients who hallucinate are less accurate on source monitoring tasks than patients who do not hallucinate (Brebion et al., 1997; Morrison and Haddock, 1997). However, Seal et al. (2004) have argued that a simple deficit in monitoring the self cannot entirely explain why patients misattribute their voices to an external source. Prior studies have investigated the possible contribution of other processes to this misattribution, such as cognitive dissonance, a propensity for the misattribution of emotional material (Morrison and Haddock, 1997) or response bias (Bentall, 1990). On the other hand, Cutting (1990) proposed an untested mechanism for the misattribution of auditory hallucinations. He assumed that hallucinatory voices are derived from a form of inner speech, and hypothesised that misattribution of inner speech may arise as a consequence of a dysfunction of prosodic processing in patients experiencing auditory hallucinations.

Prosody is the rhythm and intonation of speech and the prosodic components of speech are conveyed by variations in pitch, loudness and duration. Prosody serves two para-linguistic functions: firstly, to provide unambiguous meaning to speech and secondly, an affective function that provides information regarding the speaker’s affective state and the speaker’s message. Within the source monitoring framework, emotion is an important cue that contributes to the memory of an event (Johnson et al., 1993). In the case of hallucinatory voices, vocal emotion may function as a cue for determining the origin of an event. Such a cue may not only provide context for the origin of an event but also information on the individuality of a voice. Prosodic impairment may therefore lead to the misattribution of thought or inner speech.

Patients with schizophrenia have been reported to have difficulty in the comprehension and expression of emotional prosody (Borod et al., 1989; Murphy and Cutting, 1990), that is, they have difficulty understanding the vocal emotion of other people as well as expressing their own vocal emotion. In comparison to facial affect processing, there are relatively few investigations of emotional prosodic processing in schizophrenia. Experiments investigating vocal emotion have used a variety of measures, ranging from simple demonstrations of emotion by the experimenter, which would naturally vary between testing sessions (Fricchione et al., 1986), to the use of tasks such as that of Haskins et al. (1995) that have undergone extensive testing and validation. Stimuli used in previous studies include words (Billingberg and Jonsson, 1965), tones (Jonsson and Sjostedt, 1973), sentences (Leentjens et al., 1998; Murphy and Cutting, 1990; Turner, 1964), nonsense syllables (Kerr and Neale, 1993; Ross et al., 2001) and sounds such as screams, sighs and laughter (Billingberg and Jonsson, 1965; Ross et al., 2001; Rossell and Boundy, 2005). Generally, prior investigations have reported significant deficits in emotional prosodic processing in patients with schizophrenia compared to control participants. Only two studies failed to report differences between patient and control groups (Chernigovskaya et al., 2004; Whittaker et al., 1994).

Some studies have reported that patients with schizophrenia demonstrate greater deficits in the recognition of sad stimuli compared to controls (Bozikas et al., 2004, 2006; Edwards et al., 2001; Murphy and Cutting, 1990; Rossell and Boundy, 2005). Fear was recognised less accurately in patient groups compared to controls in some studies (Edwards et al.; Rossell and Boundy) but not all (Bozikas et al.). Deficits in anger and surprise were also reported for patient groups compared to controls (Bozikas et al.; Rossell and Boundy) again in some studies but not all (Edwards et al.).

Nearly all studies that have investigated emotional prosodic processing in schizophrenia have focused on differences between patient and control groups. There have only been two studies that have compared specific patient groups within schizophrenia (Turner, 1964; Rossell and Boundy, 2005). Turner compared paranoid and non-paranoid participants and reported that there were no significant differences between these patient groups in their recognition of emotional prosody. Rossell and Boundy (2005) compared the performance of hallucinating and non-hallucinating patients on two vocal emotion tasks using a standard sentence prosody task and a non-linguistic sounds task. This study reported that participants experiencing hallucinations were significantly less accurate than non-hallucinating patients in recognising emotion expressed by non-linguistic sounds but not when identifying emotion expressed in neutral sentences.

The findings from previous research in emotional prosody appear to have several limitations, including the use of non-standardised stimuli (Fricchione et al., 1986);
use of a small number of stimuli per affect (Bozikas et al., 2004, 2006; Fricchione et al., 1986; Leentjens et al., 1998; Murphy and Cutting, 1990; Ross et al., 2001) and the use of complex emotion categories such as condemnation and bewilderment (Chemigovskaya et al., 2004). In addition, only Bozikas et al., 2004, 2006 have explicitly stated using a practice task prior to testing in order to clarify the purpose of the task and familiarise participants with the stimuli.

The present study compared emotional prosodic processing between hallucinating and non-hallucinating schizophrenia patients. The study used a validated task based on emotions along a simple continuum from happy through neutral to sad. Our reasons for this approach were threefold. Firstly, to reduce the potential for error based on increased choice of emotions. Secondly, in choosing the happy–sad continuum we were able to avoid the potentially confounding issue of using emotional tones with similar acoustical properties. Finally, using this simple dichotomy of positive and negative emotions has ecological validity when explaining the misattribution of auditory hallucinations, which tend to be discussed in terms of positive and negative valence rather than specific emotions. On the basis of Cutting’s (1990) hypothesis, we hypothesised that emotional prosodic processing deficits would be greater in hallucinating patients compared to non-hallucinating patients. Specifically, we hypothesised that hallucinating patients would be significantly more impaired at identifying the emotional tone of the happy and sad sentences compared to the neutral sentence condition.

2. Method

2.1. Participants

Participants comprised of 67 patients with a DSM-IV diagnosis of schizophrenia or schizoaffective disorder, and 31 normal controls. Patients were classified into two groups: 29 patients with no current history of auditory hallucinations (not in the past year); and 38 patients with current auditory hallucinations (within in the past week). All were outpatients and were recruited from an existing patient database at the Mental Health Research Institute of Victoria and from the Mid-West Area Mental Health Service (Melbourne, Australia). The 31 normal controls were recruited by advertising in local newspapers.

All participants were aged between 21 and 62 years, spoke English as a first language, and had normal hearing. Exclusion criteria included IQ<70 or history of major head injury. Patients were excluded if they had a co-morbid diagnosis of another axis 1 disorder. Control participants were excluded if they had a personal or familial history of psychiatric disorder (e.g. mood disorders or substance abuse disorder) or neurological disorder.

After informed consent was obtained, participants completed a questionnaire of basic demographic details and the National Adult Reading Test (NART: Nelson, 1982). An assessment of current mood was performed using the Positive and Negative Affect Schedule (PANAS State: Watson et al., 1988).

There was no significant difference between hallucinating and non-hallucinating patients on measures of age, gender, the NART pre-morbid IQ, years of education, PANAS State, medication and age at onset of illness. However, there was a significant difference for duration of illness [F(1,61)=3.8, p=.05]. The mean duration of illness for non-hallucinating patients was 20 years compared to 14 years for hallucinating patients. Hallucinating patients differed significantly from control participants in years in education (p<.01), the NART pre-morbid IQ (p<.05) and PANAS State scores for negative affect (p<.01). Non-hallucinating patients differed significantly from control participants in age (p<.01), the NART measure of pre-morbid IQ (p<.05) and PANAS State scores for negative affect (p<.05). These demographic details are summarised in Table 1.

Patient symptoms were assessed using the Scales for Assessment of Positive and Negative Symptoms (SAPS and SANS: Andreasen, 1983, 1984), Table 2 summarises SAPS and SANS scores. The two patient groups were matched for age at onset and current dosage of medication. With the exception of the SAPS global hallucinations score (p<.001), patient groups were matched on positive and negative symptoms. Following

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of demographic data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
</tr>
<tr>
<td>N=31</td>
<td>N=29</td>
</tr>
<tr>
<td>Age†</td>
<td>37 (11)</td>
</tr>
<tr>
<td>Gender</td>
<td>20M/11F</td>
</tr>
<tr>
<td>Education (years)§</td>
<td>15 (3)</td>
</tr>
<tr>
<td>Predict FSIQ (NART)</td>
<td>115 (10)</td>
</tr>
<tr>
<td>PANAS-state positive</td>
<td>29(7)</td>
</tr>
<tr>
<td>PANAS-state negative§</td>
<td>12 (3)</td>
</tr>
<tr>
<td>Age at onset</td>
<td>NA</td>
</tr>
<tr>
<td>Duration of illness*</td>
<td>NA</td>
</tr>
<tr>
<td>Daily CPZe</td>
<td>NA</td>
</tr>
</tbody>
</table>

†p<.01, *p<.05.
these preliminary measures, all participants completed the prosodic processing task described below.

2.2. Prosodic processing task

The prosodic processing task used in this study was developed in collaboration with MARCS Auditory Laboratories at the University of Western Sydney (UWS). From a review of the literature on emotional prosody, all published sentences were included in a pilot trial of this task, along with a further 16 sentences written by the authors. A total of 57 sentences were recorded in happy, sad and neutral tones by seven actors, and rated at the UWS by 35 undergraduate students. The final selection of sentences was chosen on the basis of inter-rater agreement. A sentence was chosen if all three versions of the sentence by the same actor (i.e. happy, sad and neutral) were reliably assigned to their respective affective condition.

The task used in the present study comprised two sections, a short practice section and the test section. The practice section was based on three neutral sentences in each of the three affective conditions (happy, sad and neutral). The test items were 15 neutral sentences (e.g. You wrote it last night) which were expressed in happy, sad and neutral tones, making a total of 45 sentences. The sentences used in this study were recorded by two female actors and two male actors. This task was conducted in a quiet room and took approximately 10 min to complete. These sentences were presented to the participant’s free-field via a desktop cassette recorder (Sony Model A4022). Participant ability to hear the stimuli comfortably was established during the practice session. It was not necessary to alter the volume of the audio cassette between participants, and this was therefore kept at a constant level across all testing sessions. After each of the stimuli were presented there was a 5-s interval in which participants were able to make their response, and no warning tones were used to indicate the onset of the next stimulus. Participants were instructed to listen carefully to each sentence, ignoring the semantic content, and to listen only to the tone of the voice. After the presentation of each sentence, participants rated the emotionality of the sentence on a 7-point Likert scale from very sad (−3) through neutral (0) to very happy (+3). They were also informed that if any of the stimuli did not sound happy, sad or neutral (e.g. if it sounds angry), they could circle ‘other’.

The ability of participants to identify emotional prosody correctly was defined by calculating the number of correct identifications to the 45 sentences. Correct scores for happy were defined as ratings from 1 (slightly happy) to 3 (very happy). Correct scores for sad were defined as ratings from −1 (slightly sad) to −3 (very sad). Correct scores for neutral stimuli were defined as ratings from −1 through 0 to 1. Accuracy scores, in terms of percentage of correct responses, are reported for each of the three affective conditions (happy, sad and neutral). In addition misattribution errors were calculated as percentages of incorrect responses within each affect. Accuracy scores were analysed using repeated measures ANOVA with three groups (controls, non-hallucinating, hallucinating) and three emotions (happy, sad and neutral). Group differences were analysed using Tukey’s HSD.

3. Results

3.1. Prosodic processing task

Significant group differences were found in recognition accuracy of emotion in voice [\(F(2,95)=10.6, p<.001\)]. Post-hoc testing indicated that the hallucinating patient group was significantly less accurate (74%) in identifying the emotional prosody of sentences than both non-hallucinating patients (82%) and normal

---

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Non-hallucinator N=29</th>
<th>Hallucinator N=38</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hallucinations</td>
<td>0.5 (0.7)</td>
<td>3.0 (1.0)*</td>
</tr>
<tr>
<td>Delusions</td>
<td>1.4 (1.0)</td>
<td>1.7 (0.9)</td>
</tr>
<tr>
<td>Bizarre behaviour</td>
<td>0.4 (0.6)</td>
<td>0.6 (0.8)</td>
</tr>
<tr>
<td>Thought disorder</td>
<td>0.4 (0.7)</td>
<td>0.7 (0.9)</td>
</tr>
<tr>
<td>Affective Flattening</td>
<td>0.5 (0.8)</td>
<td>0.5 (0.8)</td>
</tr>
<tr>
<td>Alogia</td>
<td>0.3 (0.5)</td>
<td>0.6 (0.8)</td>
</tr>
<tr>
<td>Avolition</td>
<td>0.2 (0.5)</td>
<td>0.4 (0.7)</td>
</tr>
<tr>
<td>Anhedonia</td>
<td>0.8 (1.0)</td>
<td>0.9 (0.9)</td>
</tr>
<tr>
<td>Attention</td>
<td>0.2 (0.5)</td>
<td>0.5 (0.9)</td>
</tr>
<tr>
<td>Total mean SAPS</td>
<td>0.7 (0.5)</td>
<td>1.5 (0.5)*</td>
</tr>
<tr>
<td>Total mean SANS</td>
<td>0.4 (0.4)</td>
<td>0.6 (0.6)</td>
</tr>
<tr>
<td>Global mean symptoms</td>
<td>0.5 (0.4)</td>
<td>1.0 (0.5)*</td>
</tr>
</tbody>
</table>

* \(p<.01\).

---

Table 3

<table>
<thead>
<tr>
<th>Prosodic processing task — percentage of correct identifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls N=31</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Happy</td>
</tr>
<tr>
<td>Sad</td>
</tr>
<tr>
<td>Neutral</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

* \(p<.01\).
controls (86%). There were significant group differences in accuracy between normal control participants and hallucinating patients \( (p=.001) \), and between non-hallucinating patients and hallucinating patients \( (p=.006) \), but no significant difference between normal control participants and non-hallucinating patients was observed \( (p=.515) \). Table 3 summarises participant scores on this task. There was no interaction between group and emotion, and accuracy rates within each group were similar across emotion categories.

Given the significant demographic difference on duration of illness between non-hallucinating patients (20 years) and hallucinating patients (14 years), a post-hoc ANCOVA was conducted with duration of illness as a covariate. This indicated no main effect for duration of illness \( (p=.21) \), however group differences for total prosody scores remained significant \( [F(1,60)=9.6, p<.01] \). The hallucinating patient group was significantly less accurate (73%) in identifying the emotional prosody of the sentences than the non-hallucinating patients (83%). A post-hoc ANCOVA was also conducted using the PANAS State (negative) as a covariate. There was no main effect for the PANAS \( (p=.25) \) but group differences for prosodic recognition remained significant \( [F(1,91)=6.2, p<.01] \).

3.2. Misclassification of affective stimuli

The misclassification errors on the prosody task were analysed independently for each affective condition using repeated measures ANOVA (Greenhouse–Geisser corrected) with three groups (controls, non-hallucinating, hallucinating) and three emotion classifications. Misclassification errors are summarized in Table 4. For the misclassification of the happy stimuli there was a significant main effect for classified emotion \( [F(1,4,15.4)=31.0, p<.001] \). Planned contrasts revealed that on average, participants were more likely to misclassify happy to neutral than either the sad \( (p<.05) \) or “other” categories \( (p<.001) \). There was no interaction of group and emotion.

For the misclassification of sad stimuli there was a significant main effect for classified emotion \( [F(1,4,102.1)=83.4, p<.001] \). Planned contrasts revealed that on average, participants were more likely to misclassify sad stimuli to the neutral category than they were to the happy \( (p<.001) \) or “other” categories \( (p<.001) \). There was a significant interaction of group and classified emotion \( [F(2,7,102.1)=4.2, p<.01] \). Univariate analysis of variance revealed that hallucinating patients were less likely than control participants \( (p<.01) \) and non-hallucinating patients \( (p=.06) \) to misclassify sad stimuli to the neutral category \( [F(2,77)=5.5, p<.01] \). Hallucinating patients also misclassified sad stimuli to the happy \( (p<.05) \) and “other” \( (p=.08) \) categories to a greater extent than the control participants.

For the misclassification of neutral stimuli there was a significant main effect for classified emotion \( [F(1,8,145.9)=31.0, p<.001] \). Planned contrasts revealed that on average, participants were more likely to misclassify neutral stimuli to the sad category than they were to either the happy \( (p<.001) \) or the “other” \( (p<.001) \) categories. There was no interaction of group and classified emotion.

4. Discussion

The present study compared differences in emotional prosodic processing between hallucinating and non-hallucinating patients with psychosis. These results
reveal, for the first time, significant differences between these patient groups in differentiating emotional tone for spoken sentences. As predicted by Cutting (1990), the hallucinating patient group showed significant deficits compared to the non-hallucinating patient group. No significant difference was observed between non-hallucinating patients and normal controls.

In prior studies, emotional prosodic deficits were implicitly assumed to apply to all patients with psychosis. However, these studies have not accounted for the effects of sub-groups in this population. The present study indicates that emotional prosodic deficits may be specifically associated with the experience of auditory hallucinations rather than psychosis in general. We also investigated whether the participant groups would misclassify the intended affective conditions in a systematic way. All participant groups were significantly more likely to misclassify both happy and sad items to neutral before choosing the other emotion categories, however, this difference was less pronounced for the hallucinating patient group. Furthermore, all groups were significantly more likely to misclassify neutral items to sad than the other emotion categories. Again, this was less pronounced in the hallucinating patient group.

Prior research has reported that patients with psychosis appear to show greater deficits for some negative emotions compared to positive. For example, Kohler et al. (2003) reported deficits in recognition of facial emotion for fear, disgust and neutral but not for happy, sad or angry. In studies of emotional prosody in schizophrenia, deficits for sad stimuli specifically have been reported in patients groups compared to controls (Bozikas et al., 2006; Edwards et al., 2001; Murphy and Cutting, 1990; Rossell and Boundy, 2005). In the present study, there was no interaction of group and emotion, performance in emotion recognition was similar across the participant groups regardless of valence, hallucinating patients were less accurate overall than non-hallucinating patients and controls. Contrary to prior studies suggesting specific deficits in specific types of emotion, the present results suggest a non-specific or general deficit of emotional prosodic processing in patients with psychosis.

The conflicting results from may be due to the use of different prosodic processing tasks. Prior studies have generally used a smaller number of stimuli per affect, possibly reducing the power of those studies (Bozikas et al.; Murphy and Cutting). Most of these studies have also used only one actor (Bozikas et al.; Murphy and Cutting; Rossell and Boundy). Actors can vary not only in their ability to convey different affective conditions, but also the style in which they convey them. If this issue is not controlled by using more than one actor, then the ability of an emotional prosody task to measure all affective conditions equally well is questionable. Finally, the validation of a task is very important, a task that has had limited (Rossell and Boundy) or possibly no validation may not measure a psychological construct as reliably as a task that has had more extensive validation, as is the case in the present study.

Another important question in schizophrenia research is whether patient deficits in experimental tasks are related to not only specific symptoms such as auditory hallucinations but whether other symptoms or cognitive factors might contribute to poor performance. Poole et al. (1997) reported that deficits in recognition of emotional prosody were correlated with disorganized features, positive symptoms in general and attention. Bozikas et al. (2004) also noted correlations between emotional prosodic recognition and executive functions, fluency and attention. However, in this study, the presence of delusions was not significantly different between patient groups in the present study, suggesting that emotional prosodic deficits are associated with the presence of auditory hallucinations rather than delusions, or possibly other positive symptom as measured by the SAPS. Furthermore, patient groups were matched on other cognitive and demographic variables such as attention, estimated IQ, current mood and medication, indicating that these issues did not play a critical role in patient group differences in this study.

Due to the length of time taken by the study design, no attempt was made to record specific details regarding the patient’s hallucinatory experience on the day of testing. Future studies may wish to compare groups who were hallucinating on the day of testing against those who had hallucinated only in the past week or month. Another area of interest for future studies is the potential effect of specific psychotropic medications on emotional prosodic performance.

Cutting’s (1990) hypothesis predicted that prosodic processing deficits would be greater in patients experiencing auditory hallucinations compared to those who do not experience this phenomenon. While the hallucinating patient group performed above chance level in this study, indicating some ability in the identification of emotion, they appear to be less able to utilise prosodic cues to make finer distinctions compared to the other groups. Memory for voice and speaker identity is believed to rely on the ability to use prosodic cues. We suggest that deficits in emotional prosodic processing may contribute to errors of source monitoring via an inability to use these emotional cues to help ascertain the origin of an auditory event, in
this case inner speech or thought. The present study has established group differences in emotional prosodic processing of spoken sentences. An important next step is to test participants source monitoring and prosodic processing abilities together within a single study.

Acknowledgements

The authors wish to thank Dr. Kate Stevens of MARCS Auditory Laboratories at the University of Western Sydney for her assistance in the development and validation of the prosodic processing task used in this study, and Assoc. Prof. Deirdre Russell-Bowie, Mr Brett McLaren and Mr Mark Lemmon for recording the stimuli used in this task. We also thank Professor Norman Moore of Sunshine Hospital for his invaluable assistance in the recruitment of participants. This work was supported by the National Health and Medical Research Council Australia (Grant 236025).

References

Andreasen, N.C., 1983. The Scale for the Assessment of Negative Symptoms (SANS). The University of Iowa, Iowa City.


