



Published in final edited form as:

*Am J Psychiatry*. 2012 April ; 169(4): . doi:10.1176/appi.ajp.2011.11081230.

## Auditory emotion recognition impairments in Schizophrenia: Relationship to acoustic features and cognition

Rinat Gold, Ph.D.<sup>1</sup>, Pamela Butler, Ph.D.<sup>2,3</sup>, Nadine Revheim, Ph.D.<sup>1</sup>, David Leitman, Ph.D.<sup>2,4</sup>, John A. Hansen, Ph.D.<sup>4</sup>, Ruben Gur, Ph.D.<sup>4</sup>, Joshua T. Kantrowitz, M.D.<sup>1,3</sup>, Petri Laukka, Ph.D.<sup>5</sup>, Patrik N. Juslin, Ph.D.<sup>6</sup>, Gail S. Silipo, M.A.<sup>1</sup>, and Daniel C. Javitt, M.D., Ph.D.<sup>1,3</sup>

<sup>1</sup>Center for Translational Schizophrenia Research, Nathan Kline Institute, Orangeburg, NY, United States

<sup>2</sup>Psychiatry, New York University, New York, NY, United States

<sup>3</sup>Psychiatry, Columbia University, New York, NY, United States

<sup>4</sup>Neuropsychiatry, University of Pennsylvania, Philadelphia, PA, United States

<sup>5</sup>Psychology, Stockholm University, Stockholm, Sweden

<sup>6</sup>Psychology, Uppsala University, Uppsala, Sweden

### Abstract

**Objective**—Schizophrenia is associated with deficits in ability to perceive emotion based upon tone of voice. The basis for this deficit, however, remains unclear and assessment batteries remain limited. We evaluated performance in schizophrenia on a novel voice emotion recognition battery with well characterized physical features, relative to impairments in more general emotional and cognitive function.

**Methods**—We studied in a primary sample of 92 patients relative to 73 controls. Stimuli were characterized according to both intended emotion and physical features (e.g., pitch, intensity) that contributed to the emotional percept. Parallel measures of visual emotion recognition, pitch perception, general cognition, and overall outcome were obtained. More limited measures were obtained in an independent replication sample of 36 patients, 31 age-matched controls, and 188 general comparison subjects.

---

\*Contact info: Daniel C. Javitt, M.D., Ph.D., Director, Center for Translational Schizophrenia Research, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Rd Orangeburg, NY 10962, Phone: (845) 398-6534/6546, Fax: (845) 398-6545, Javitt@nki.rfmh.org.

#### Disclosures:

Gold: None

Butler: None

Revheim: None

Leitman: None

Hansen: None

Gur: None

Kantrowitz: Within the past year, Dr. Kantrowitz has conducted clinical research supported by Roche, Sunovion, Novartis, Pfizer and GlaxoSmithKline, and served as a paid consultant to Agency Rx and RTI Health Solutions. He and/or his spouse own a small number of shares of common stock in GlaxoSmithKline.

Laukka: None

Juslin: None

Silipo: None

Javitt: Dr. Javitt holds intellectual property rights for use of NMDA agonists, including glycine, D-serine, and glycine transport inhibitors in treatment of schizophrenia. Dr. Javitt is a major shareholder in Glytech, Inc. and Amino Acids Solutions, Inc. Within the past year, Dr. Javitt has served as a paid consultant to Sepracor, AstraZeneca, Pfizer, Cypress, Merck, Sunovion, Eli Lilly, and BMS

**Results**—Patients showed significant, large effect size deficits in voice emotion recognition ( $F=25.4$ ,  $p<.00001$ ,  $d=1.1$ ), and were preferentially impaired in recognition of emotion based upon pitch-, but not intensity-features (group X feature interaction:  $F=7.79$ ,  $p=.006$ ). Emotion recognition deficits were significantly correlated with pitch perception impairments both across ( $r=.56$ ,  $p<.0001$ ) and within ( $r=.47$ ,  $p<.0001$ ) group. Path analysis showed both sensory-specific and general cognitive contributions to auditory emotion recognition deficits in schizophrenia. Similar patterns of results were observed in the replication sample.

**Conclusions**—The present study demonstrates impairments in auditory emotion recognition in schizophrenia relative to acoustic features of underlying stimuli. Furthermore, it provides tools and highlights the need for greater attention to physical features of stimuli used for study of social cognition in neuropsychiatric disorders.

## Introduction

During human interaction, individuals communicate information not only verbally but also through tone of voice. Specifically, individuals modulate their voices differently during different emotional states, for example, speaking quietly and with little animation when sad, speaking loudly and with great animation when happy, or shouting when angry (1, 2). Detection of this non-verbal information allows listeners to adjust their behavior accordingly and thus to perform adequately in social situations. Accordingly, impairments in auditory emotion recognition (AER) and social cognition contribute strongly to poor psychosocial outcome in schizophrenia (3–5). Nevertheless, the pattern and basis of AER deficits remains an area of active research, as does the relationship between AER deficits in schizophrenia and deficits in more global forms of cognition dysfunction. In the visual modality, the ability to recognize emotion from faces has been operationalized using well-validated face recognition tests such as the Emotion recognition (ER-40) (6) or Ekman face (7) tests. In the auditory modality, however, batteries for assessment of emotion recognition remain relatively underdeveloped, limiting opportunities for clinical assessment and research.

To date multiple batteries have been used for study of AER impairments in schizophrenia with no standardization of the psychoacoustic properties of stimuli across studies. Potentially as a result of this, different patterns of auditory emotion recognition deficits have been reported (8–10), with some studies suggesting a generalized pattern of deficit (11), and others hemisphere- (12), emotion- (13) or valence-(14) specific patterns. In the present study we validate the use of a novel psychoacoustically well-characterized auditory battery in a large group of schizophrenia patients as a means of investigating not only pattern but also underlying basis of social cognition impairments in schizophrenia.

In auditory communication, overlapping but distinct patterns of psychoacoustic features contribute to communication of discreet auditory emotional percepts (see (2), for a comprehensive review). For example, specific pitch-related vocal features such as mean pitch (F0M), pitch variability (F0SD), and pitch contour (F0contour) are critical for communicating emotions such as happiness, sadness and fear, whereas specific intensity (loudness)-related aspects, such as mean voice intensity (VoInt), intensity variability (VoIntSD) and voice quality as reflected in the amount of high-frequency energy over 500 Hz (HF500) are particularly crucial for communicating the percept of anger.

Furthermore, several emotions may be communicated in more than one fashion (15). For example, anger may be communicated either by increased voice intensity (i.e. shouting, “hot” anger), or by reduction in the mean base pitch of the voice without increasing intensity (irritation, “cold” anger) (1). Similarly, conveyance of strong happiness (“elation”) may depend upon somewhat different cues than conveyance of weaker forms of happiness (1) (see on-line Supplemental files for examples). To date, auditory emotion batteries used in

schizophrenia research have not distinguished the precise tonal features used by different actors in portraying specific emotions, potentially contributing to heterogeneity of findings across studies.

Over recent years, we (16, 17) and others (18, 19) have documented deficits in basic pitch perception (i.e. tone matching) in schizophrenia related to structural (20) and functional (21, 22) impairments at the level of primary auditory cortex. In a recent prior emotion study we evaluated performance of schizophrenia patients on a novel emotion recognition battery using psychoacoustically characterized stimuli, and observed deficits in pitch- but not intensity-based emotion processing (23). A similar pattern of deficit was observed more recently for emotion conveyed by frequency-modulated tones (24). The present study extends these findings to a larger patient sample and validates a brief version of the test for widespread use.

Along with the specific sensory-level contributions to auditory emotion recognition, other potential contributors include emotion-level dysfunction, as well as general cognitive impairments. In the present study, we evaluated visual emotion recognition using the ER-40 test (25), which includes similar emotions to those in our auditory emotion battery. We also included the Processing Speed Index (PSI) of the Wechsler adult intelligence scale (WAIS-III), which contains tests such as the digit symbol that are thought to be particularly sensitive to generalized cognitive impairment in schizophrenia (26, 27).

We hypothesized that patients would show emotion recognition deficits for stimuli in which emotion was conveyed primarily by pitch-, but not intensity-based measures, and that correlations between basic pitch processing and auditory emotion recognition would remain significant even following co-variation for deficits in non-auditory emotion recognition and general cognitive function. In addition, we evaluated replicability of findings across two separate performance sites and thus applicability to general schizophrenia and neuropsychiatric populations.

## Methods

### Participants

In the primary sample, 92 chronically ill schizophrenia or schizoaffective patients participated as determined by Structured Clinical Interview for DSM-IV diagnosis (Table 1). Patients were drawn from chronic inpatient units and residential care facilities associated with the Nathan S. Kline Institute. All were receiving typical or atypical antipsychotics at the time of testing. Volunteers were staff or responded to local advertisement. Groups did not differ on mean parental socioeconomic status (SES), which reflects level of education and employment on a scale of 0–100 (ref) ( $p=.6$ ).

As expected, patients had lower individual SES relative to both controls ( $p<.001$ ) and parents ( $p<.0001$ ).

Clinical assessments included ratings on the Positive and Negative Syndrome Scale (PANSS) (28) and the Independent Living Scales-Problem Solving Scale (ILS-PB), (29). All procedures conducted received approval from the Institutional Review Board at Nathan S. Kline Institute. All subjects had the procedure explained to them verbally before giving their written informed consent.

In the replication sample (Table 2), subjects were drawn from psychiatric and both age-matched and more general normative populations association with the University of

Pennsylvania. As no significant differences were observed between the two control groups on any of the dependent measures, the two groups were combined in statistical analyses.

## Materials

For the full version of the task, stimuli consisted of 88 audio recordings of native British English speakers conveying 5 emotions—anger, disgust, fear, happiness, and sadness—or no emotion as previously described (2, 23). Acoustic features for these stimuli were measured using PRAAT speech analysis software (30). Sample stimuli are attached as supplemental information. Full stimuli and features are available for download at (<http://www.psyk.uu.se/research/researchgroups/musicpsychology/downloads/>).

For the brief version, a subset of 32 stimuli were selected incorporating all emotions other than disgust, which was eliminated in order to decrease number of choices and therefore number of stimuli. Pitch based stimuli (N=17) were selected based upon a prior study which showed that these stimuli were well recognized as expressing the intended emotion (i.e., % correct scores >60%). (23). In addition, physical characteristics of the stimuli such as base pitch and pitch variability were close to the mean value for the intended emotion (Figure 1).

Stimuli selected as intensity-based were confined to anger and happiness portrayals (N=9), and differed from pitch-based stimuli of the same emotion on physical intensity measures such as overall intensity or high frequency energy. Intensity-based anger portrayals would all be recognized as loud based upon overall intensity (Figure 1), and thus represent “hot” anger, as opposed to “cold” pitch-based anger. Pitch- vs. intensity-based happiness differed in sound quality (high frequency energy), with pitch-based stimuli showing features most characteristic of “happiness” and intensity-based happiness showing characteristics of “elation” as described by Banse and Scherer (1).

Participants were tested on either the full original 88 items (67 Sz/32 Ctl) or the 32-item brief version (40 Sz/53 Ctl). Stimuli representing different emotions were intermixed and presented in consistent order across subjects. After each stimulus, subjects were asked to identify the emotion (6 potential alternatives in the full version; 5 in the brief), as well as the intensity of portrayal on a scale of 1–10. A limited number of subjects received both the full and abbreviated versions of the task (15 Sz/12 Ctl). In this subgroup, no significant group X version interaction ( $F=.13$ ,  $df=1,25$ ,  $p=.7$ ) was observed. For subsequent analyses, these subjects were counted only once, with data from the full battery being used for statistical analysis.

Tone matching was assessed using pairs of 100-ms tones in series, with 500-ms intertone interval. Within each pair, tones (50% each) were either identical or differed in frequency by a specified amount in each block (2.5%, 5%, 10%, 20%, or 50%). Subjects indicated by keypress whether the pitch was the same or different. Three base frequencies (500, 1000, and 2000 Hz) were used within each block to avoid learning effects. In all, the test consisted of 5 sets of 26 pairs of tones.

Recognition of visual emotion was assessed using The Penn Emotion Recognition Task (ER40) (31–33). Global cognitive function was assessed using the WAIS-III Processing speed index (PSI), which includes the widely studied digit symbol coding (27) and the symbol search subtests.

## Data Analysis

Accuracy of auditory emotion recognition was assessed using repeated measures analysis of variance (ANOVA) with either emotion (happy, sad, anger, fear, disgust, no emotion) or feature (pitch, intensity) as within-subject factors, and diagnostic group as a between-subject

factor. The relationship between emotion identification and specific predictors was assessed using analysis of covariance (ANCOVA) with TMT, ER40, PSI included as potential covariates. Task version (full/brief) was included in the analysis as well to remove variance associated with these factors.

Structural equation modeling was used to further investigate the pattern of correlation observed with regression analysis, and to query both 1) directionality of relationship, and 2) covariance among measures within the context of the overall correlation pattern. Alternate models were accepted if they led to a significant reduction in variance as measured using the  $\chi^2$  goodness-of-fit parameter. Effect size measures were interpreted according to the convention of Cohen (34). All statistical tests were 2-tailed, with preset  $\alpha$  of  $p < .05$ , and computed using SPSS 18.0 (Chicago, IL).

## Results

### Full version

Patients showed highly significant, large effect-size impairment across stimuli ( $F=25.4$ ,  $df=1,97$ ,  $p < .00001$ ,  $d=1.1$ ) (Figure 2). In contrast, the group X emotion interaction ( $F=2.09$ ,  $df=5,93$ ,  $p=.07$ ) was significant at trend level only, suggesting statistically similar deficit across emotions. Despite their impairments, patients performed well above chance levels (16.7% correct) for all emotions.

When stimuli were divided according to underlying feature (pitch/intensity), a highly significant group effect was again observed ( $F=14.4$ ,  $df=1,97$ ,  $p=.0003$ ), as well as a highly significant group X feature interaction ( $F=7.79$ ,  $df=1,97$ ,  $p=.006$ ). Follow-up t-tests showed a highly significant difference in detection of pitch-based emotion ( $t=4.51$ ,  $df=97$ ,  $p < .0001$ ) but no significant difference in detection of intensity-based emotion ( $t=1.44$ ,  $df=97$ ,  $p=.15$ ) (Figure 3A). Mean performance levels across groups was not significantly different for pitch vs. intensity stimuli, suggesting that group interactions were not due to floor/ceiling effects but that controls were better able to discriminate emotions when emotion-relevant pitch information was present, whereas patients were not.

### Brief version

A second group received the 32-item brief version of the task. In the brief version, the main effect of group ( $F=18.3$ ,  $df=1,91$ ,  $p < .0001$ ) and group X feature interaction ( $F=9.61$ ,  $df=1,91$ ,  $p=.003$ ) were again significant, with significant between-group differences for pitch- ( $t=5.32$ ,  $df=91$ ,  $p < .0001$ ,  $d=1.1$ ), but not intensity- ( $t=1.59$ ,  $df=91$ ,  $p=.11$ ,  $d=.33$ ), based stimuli (Figure 3B). In the brief battery, as in the full battery, there was no significant group X emotion interaction ( $F=1.73$ ,  $df=4,88$ ,  $p=.15$ ).

### Relative contribution of sensory vs. general cognitive dysfunction

In addition to deficits in auditory emotion recognition, patients showed highly significant deficits in TMT, ER40 and PSI (Figure 4). As predicted, the correlation between tone matching and auditory emotion recognition was highly significant both across groups ( $r=.56$ ,  $n=164$ ,  $p < .0001$ ) (Figure 5A) and within patients ( $r=.47$ ,  $n=91$ ,  $p < .0001$ ) and controls ( $r=.34$ ,  $n=73$ ,  $p=.004$ ) independently.

In order to evaluate relative contribution of these measures, an ANCOVA was conducted incorporating group as a between-subject factor and TMT, ER40 and PSI as potential covariates. Both TMT ( $F=8.72$ ,  $df=1,117$ ,  $p=.004$ ) and PSI ( $F=12.9$ ,  $df=1,117$ ,  $p < .0001$ ) correlated significantly with AER performance, whereas the correlation with ER40 was non-

significant ( $F=2.24$ ,  $df=1,117$ ,  $p=.14$ ). Once accounting for effects of TMT and PSI, the main effect of group was no longer significant ( $F=.9$ ,  $df=1,117$ ,  $p=.35$ ).

Finally, inclusion of these factors into a path analysis yielded a strong model confirming both TMT and PSI as mediators of the group effect on auditory emotion recognition, and showing as well an interrelationship between TMT and PSI. In the path analysis, a significant relationship between auditory and visual emotion recognition was observed, with AER predicting performance on ER40 (Figure 5B).

### Validation of pitch vs. intensity dichotomy

Tone matching (TMT) measures were also used to validate the psychoacoustic dichotomization of stimuli into pitch- vs. intensity- based. An ANCOVA conducted across groups with TMT as covariate showed not only a significant effect of TMT ( $F=18.4$ ,  $df=1,159$ ,  $p<.0001$ ) but also a significant TMT X feature interaction ( $F=4.25$ ,  $df=1,159$ ,  $p=.041$ ) reflecting a significantly stronger relationship between TMT performance and accuracy in identifying pitch-based stimuli ( $F=30.8$ ,  $df=1,161$ ,  $p<.0001$ ) than between TMT and accuracy in identifying intensity-base stimuli ( $F=8.43$ ,  $df=1,161$ ,  $p=.002$ ). If analyses were restricted to happy stimuli alone, an even stronger dissociation was observed with a highly significant TMT X feature interaction ( $F=10.2$ ,  $df=1,157$ ,  $p=.002$ ) and a significant relationship between TMT and performance only for pitch-based ( $F=19.6$ ,  $df=1,159$ ,  $p<.0001$ ) but not intensity based ( $F=1.56$ ,  $df=1,159$ ,  $p=.2$ ) stimuli. Within patients alone, significant correlations were observed between TMT and ability to detect pitch-based happiness ( $r=.38$ ,  $df=90$ ,  $p<.0001$ ) and anger ( $r=.30$ ,  $df=65$ ,  $p=.017$ ), but not intensity-based emotions ( $p>.15$ ).

### “Cold” vs. “Hot” Anger

Pitch vs. intensity analyses were also conducted separately for both anger and happiness, both of which may be conveyed by either pitch- or intensity-modulation (Figure 1). Patients showed significant deficits in detection of anger conveyed by pitch modulation (“cold anger”, irritation) ( $t=2.51$ ,  $p=.014$ ), but not by intensity (“hot” anger) ( $t=.6$ ,  $p=.5$ ), although the group X feature interaction was significant at trend level only ( $F=3.38$ ,  $p=.07$ ). Similarly, patients showed significant deficits in detection of happiness conveyed primarily by pitch ( $t=2.57$ ,  $p=.011$ ) but not intensity (“elation”) ( $t=.38$ ,  $p=.7$ ), modulation (Suppl. Table 1).

### Auditory vs. visual emotion recognition

In the ER40 (Suppl. Table 2), patients showed significant impairments in detection of Sadness ( $p=.003$ ), Fear ( $p<.001$ ), and No emotion ( $p=.003$ ), with trends also for deficits in Happiness ( $p=.07$ ) and Anger ( $p=.06$ ). When correlations between AER and ER40 were conducted for individual emotions within patients (Suppl. Table 3), strongest correlations were found within emotion (mean  $r=.33$ ,  $p<.01$ ), with lower correlation across emotion (mean  $r=.12$ , NS).

### Correlation with symptoms and outcome

Deficits in AER correlated significantly with the cognitive factor of the PANSS ( $r=-.33$ ,  $p=.003$ ), but not with other PANSS factors. Deficits in emotion processing also correlated with standardized ILS-PB scores ( $r=.26$ ,  $p=.017$ ). Correlations with medication dose, as assessed using CPZ equivalents, were non-significant across all emotions ( $r=-.20$ ,  $p=.2$ ).

### Replication sample

In the replication sample (Table 2), as in the primary group, there was a highly significant mean effect of group ( $F=42.4$ ,  $df=1,253$ ,  $p<.0001$ ,  $d=1.49$ ) along with a significant group X



feature (pitch/intensity) interaction ( $F=6.35$ ,  $df=1,253$ ,  $p=.012$ ). In addition, TMT performance significantly predicted AER performance over and above effect of group ( $F=24.2$ ,  $df=1,249$ ,  $p<.0001$ ). In contrast, as in the primary sample, the group X emotion interaction was not significant ( $F=4.0$ ,  $df=4,250$ ,  $p=.18$ ) (Suppl. Table 4). Reliability of the measures across samples based on intraclass correlation (ICC) was .97 for patients and .96 for controls

## Discussion

Impairments in social cognition are among the greatest contributors to social disability in schizophrenia (25, 32, 35, 36). Operationally, these deficits are defined based upon inability to infer emotion from both facial expression and auditory perception. Although well validated batteries have been developed to assess visual aspects of social cognition (31, 37), auditory batteries remain highly variable, with limited standardization across studies (e.g., 9). Moreover, the relative contributions of specific sensory features vs. more generalized cognitive performance remains relatively unknown.

The present study reports on auditory emotion recognition (AER) deficits in two independent samples of patients and controls using a novel, well-characterized battery in which the physical features of the stimuli have been analyzed, and in which stimuli have been divided *a priori* according to physical stimulus features that contribute most strongly to the emotional percept. In addition to strongly confirming the AER deficit in schizophrenia that we have observed previously, this study provides the first demonstration of a specific sensory contribution to impaired AER that remains significant even when more general emotional and cognitive deficits are considered. Finally, the study provides both a general and a brief AER battery for study across neuropsychiatric disorders.

In the present battery, angry and happy stimuli were divided *a priori* into pitch- vs. intensity-based exemplars based upon physical stimulus features. As we have previously observed both with these stimuli (23) and with synthesized frequency modulated tones designed to reproduce the key physical characteristics of emotional prosody (24), patients show greater deficit in emotion recognition when emotional information is conveyed by modulations in pitch rather than intensity. Significant group X stimulus feature interactions were found for both the full and brief versions of the battery and in both the primary and replication sample. The present battery thus provides a replicable method both to characterize sensory contributions to AER impairments in schizophrenia, and to compare specific patterns of dysfunction across neuropsychiatric illness.

In addition to differential analysis of deficits by pitch- vs. intensity-based characterization, the present study analyzed AER relative to tone matching (TMT), which provides an objective index of auditory sensory processing ability, and both face emotion recognition (ER40), and WAIS-III processing speed index (PSI), which provide measures of visual emotion and general cognitive dysfunction in schizophrenia, respectively (27, 38). Relative contributions of these measures to AER deficits were assessed using both multivariate regression and path analysis.

All 3 sets of measures (TMT, ER40, PSI) showed highly significant independent correlations to AER function across groups, with no further difference in AER function between schizophrenia and control subjects once these factors were taken into account. Approximately equal contributions were found for TMT and PSI (Figure 5B), with AER deficits, in turn, predicting impairments in ER40. In addition, when correlations were analyzed between auditory and visual emotion recognition batteries, correlations were strongest within rather than across emotions, suggesting some shared emotional processing

disturbance in addition to contributions of specific sensory deficits. Similar findings were obtained in the replication sample, in which group membership, tone matching and ER40 performance all contributed significantly and independently to AER performance.

Finally, deficits in AER also correlated with score on the ILS-PB, a proxy measure for functional capacity (39, 40). Remediation of deficits in basic auditory processing has recently been found to induce improvement as well in global cognitive performance as measured using the MATRICS consensus cognitive battery (41). The present study suggests that sensory-based remediation, along with specific emotion-based remediation, may be most useful for addressing social cognitive impairments in schizophrenia.

Based on present findings, we propose that greater attention should be given to the physical characteristics of stimuli used for assessment of social cognition deficits not only in schizophrenia but also across neuropsychiatric disorders. Thus, for example, autism spectrum disorders (ASD) are associated with AER deficits using batteries similar to those used in schizophrenia (e.g. 42). However, the specific pattern of deficit may differ from that in schizophrenia. ASD patients are reported to show most pronounced deficits in vocal perception of anger, fear and disgust, with relatively spared perception of sadness (42). Our present study suggests that dissociation across emotion in schizophrenia is not observed once the physical nature of the stimuli are considered. Comparison across populations, however, would be facilitated by use of a consistent battery with well described physical features, such as proposed here, in order to allow identification of relative determinants of social cognition deficit across conditions.

Although the present battery represents a significant advance over prior batteries, some limitations remain. First, actors were not coached to emphasize specific features when portraying an emotion, so critical stimulus parameters had to be deduced post-hoc. Future batteries in which actors purposely try to convey emotion by modulation of specific tonal or intensity-based features would permit even further ability to evaluate differential mechanisms of emotion recognition dysfunction across diagnostic groups. Second, the present study used primarily a chronic, medicated subject population. Future studies with prodromal or first-episode subjects are needed to further delineate the temporal course of emotion recognition function relative to more basic impairments in tone matching ability. Third, although the pattern of results in the present study is similar to that we have observed previously with this battery (23), formal psychometric properties of the battery, such as test-retest reliability and sensitivity to change following intervention remain to be determined. Fourth, actors included in this battery spoke with British accent, which may have influenced results. Future studies using actors speaking native accent would therefore be desirable. Finally, other components of interpersonal interaction may also communicate emotion, such as body movement, context, or verbal content of language. These were not tested in the present study.

In summary, deficits in social cognition are now well recognized in schizophrenia, although underlying mechanisms are yet to be determined. The present study highlights substantial deficits of schizophrenia patients in the ability to decode specific stimulus features, such as pitch modulations, in interpreting emotion, leading to overall impairments in AER. Deficits correlated with more basic impairments in sensory processing even when general cognitive and non-auditory emotion deficits were taken into account. These findings highlight the importance of sensory impairments, along with more general cognitive measures, as a basis for social disability in schizophrenia. In the short term, such deficits must be considered during interactions with patients, and both clinicians and care-givers should be aware that patients may simply be unable to perceive the acoustic features in speech that permit normal



social interaction. In the long-term, such deficits represent appropriate targets for both behavioral and pharmacological intervention.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Joanna DiCostanza, Rachel Ziwich, and Jonathan Lehrfeld for their critical contributions to patient recruitment, assessment and data management and Tracey Keel for their administrative support. We also thank the faculty and staff of the Clinical Research and Evaluation Facility (CREF) and Outpatient Research Service (ORS).

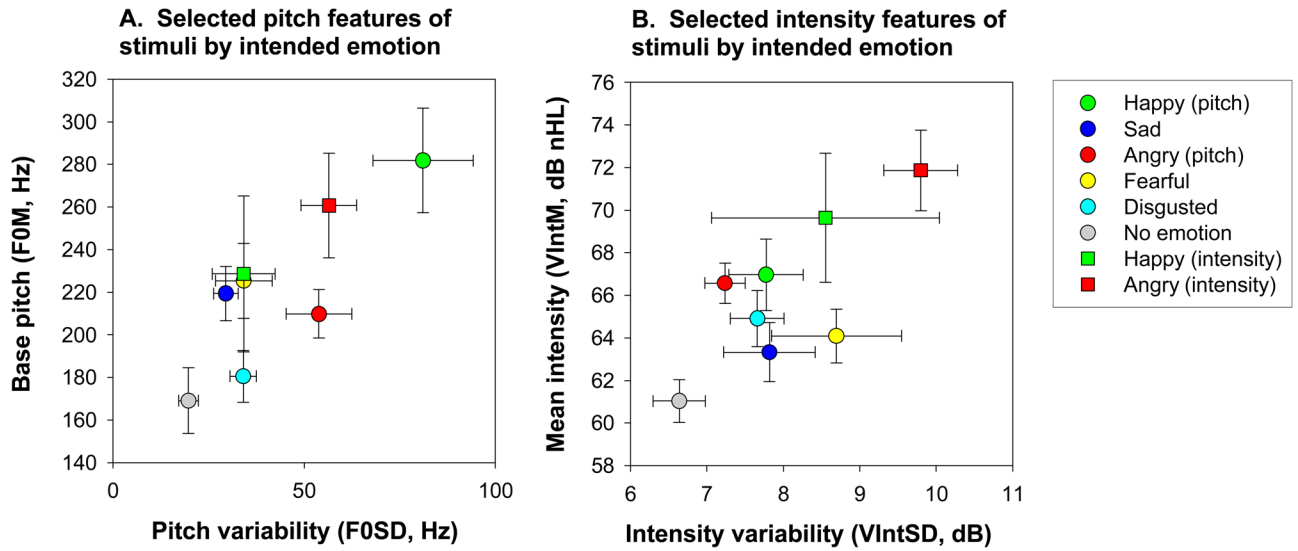
**Support:** This work was supported in part by NIMH grants R37 MH049334, ARRA supplement R37 MH49334 S1, P50 MH086385 and P50 MH086385 S1 (DCJ); R01 MH084848 (PDB); MH084856 and MH060722 (RCG) and a grant from NARSAD (DIL)

## References

1. Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol.* 1996; 70(3):614–36. [PubMed: 8851745]
2. Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological bulletin.* 2003; 129(5):770–814. [PubMed: 12956543]
3. Kee KS, Green MF, Mintz J, Brekke JS. Is emotion processing a predictor of functional outcome in schizophrenia? *Schizophr Bull.* 2003; 29(3):487–97. [PubMed: 14609242]
4. Brekke J, Kay DD, Lee KS, Green MF. Biosocial pathways to functional outcome in schizophrenia. *Schizophr Res.* 2005; 80(2–3):213–25. [PubMed: 16137859]
5. Harvey PO, Bodnar M, Sergerie K, Armony J, Lepage M. Relation between emotional face memory and social anhedonia in schizophrenia. *J Psychiatry Neurosci.* 2009; 34(2):102–10. [PubMed: 19270760]
6. Silver H, Shlomo N, Turner T, Gur RC. Perception of happy and sad facial expressions in chronic schizophrenia: evidence for two evaluative systems. *Schizophr Res.* 2002; 55(1–2):171–7. [PubMed: 11955976]
7. Sparks A, McDonald S, Lino B, O'Donnell M, Green MJ. Social cognition, empathy and functional outcome in schizophrenia. *Schizophr Res.* 2010; 122(1–3):172–8. [PubMed: 20609567]
8. Edwards J, Pattison PE, Jackson HJ, Wales RJ. Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophr Res.* 2001; 48(2–3):235–53. [PubMed: 11295377]
9. Edwards J, Jackson HJ, Pattison PE. Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review. *Clin Psychol Rev.* 2002; 22(6):789–832. [PubMed: 12214327]
10. Hoekert M, Kahn R, Pijnenborg M, Aleman A. Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophrenia Research.* 2007; 96(1–3): 135–45. [PubMed: 17766089]
11. Chapman LJ, Chapman JP. The measurement of differential deficit. *Journal of Psychiatric Research.* 1978; 14:303–11. [PubMed: 722633]
12. Ross ED, Orbelo DM, Cartwright J, Hansel S, Burgard M, Testa JA, et al. Affective-prosodic deficits in schizophrenia: comparison to patients with brain damage and relation to schizophrenic symptoms [corrected]. *J Neurol Neurosurg Psychiatry.* 2001; 70(5):597–604. [PubMed: 11309452]
13. Murphy D, Cutting J. Prosodic comprehension and expression in schizophrenia. *J Neurol Neurosurg Psychiatry.* 1990; 53(9):727–30. [PubMed: 2246653]
14. Bozikas VP, Kosmidis MH, Anezoulaki D, Giannakou M, Andreou C, Karavatos A. Impaired perception of affective prosody in schizophrenia. *J Neuropsychiatry Clin Neurosci.* 2006; 18(1): 81–5. [PubMed: 16525074]
15. Juslin PN, Laukka P. Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion.* 2001; 1(4):381–412. [PubMed: 12901399]

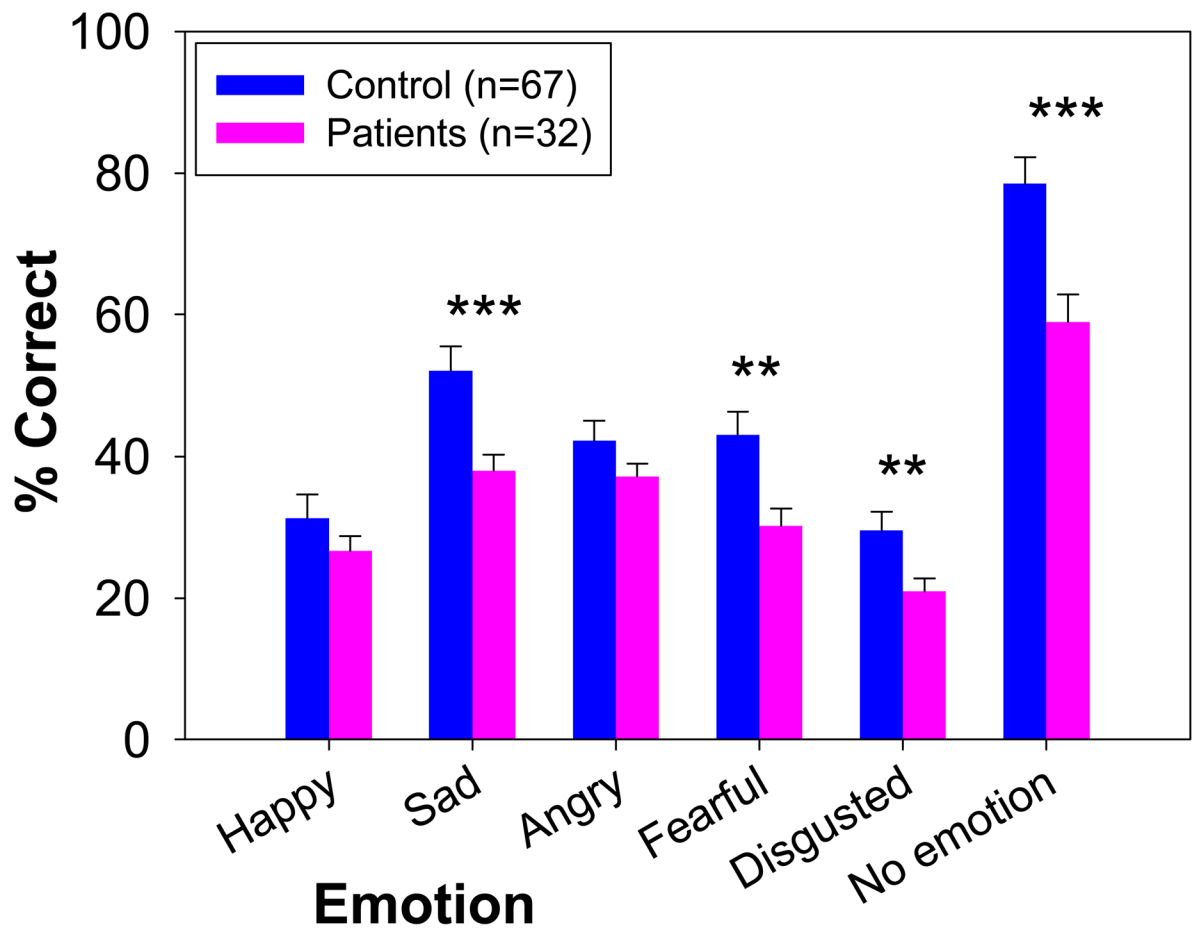
16. Strous RD, Grochowski S, Cowan N, Javitt DC. Dysfunctional encoding of auditory information in schizophrenia. *Schizophrenia Research*. 1995; 15:135.
17. Rabinowicz EF, Silipo G, Goldman R, Javitt DC. Auditory sensory dysfunction in schizophrenia: imprecision or distractibility? *Arch Gen Psychiatry*. 2000; 57(12):1149–55. [PubMed: 11115328]
18. Holcomb HH, Ritzl EK, Medoff DR, Nevitt J, Gordon B, Tamminga CA. Tone discrimination performance in schizophrenic patients and normal volunteers: impact of stimulus presentation levels and frequency differences. *Psychiatry Res*. 1995; 57(1):75–82. [PubMed: 7568562]
19. Wexler BE, Stevens AA, Bowers AA, Sernyak MJ, Goldman-Rakic PS. Word and tone working memory deficits in schizophrenia. *Arch Gen Psychiatry*. 1998; 55(12):1093–6. [PubMed: 9862552]
20. Leitman DI, Hoptman MJ, Foxe JJ, Saccente E, Wylie GR, Nierenberg J, et al. The neural substrates of impaired prosodic detection in schizophrenia and its sensorial antecedents. *Am J Psychiatry*. 2007; 164(3):474–82. [PubMed: 17329473]
21. Leitman DI, Wolf DH, Ragland JD, Laukka P, Loughhead J, Valdez JN, et al. “It’s Not What You Say, But How You Say it”: A Reciprocal Temporo-frontal Network for Affective Prosody. *Frontiers in human neuroscience*. 2010; 4:19. [PubMed: 20204074]
22. Leitman DI, Wolf DH, Laukka P, Ragland JD, Valdez JN, Turetsky BI, et al. Not pitch perfect: Sensory contributions to affective communication impairment in schizophrenia. *Biol Psychiatry*. 2011 in press.
23. Leitman DI, Laukka P, Juslin PN, Saccente E, Butler P, Javitt DC. Getting the cue: sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophr Bull*. 2010; 36(3):545–56. [PubMed: 18791077]
24. Kantrowitz JT, Leitman DI, Lehrfeld JM, Laukka P, Juslin PN, Butler PD, et al. Reduction in Tonal Discriminations Predicts Receptive Emotion Processing Deficits in Schizophrenia and Schizoaffective Disorder. *Schizophr Bull*. 2011
25. Pinkham AE, Gur RE, Gur RC. Affect recognition deficits in schizophrenia: neural substrates and psychopharmacological implications. Expert review of neurotherapeutics. 2007; 7(7):807–16. [PubMed: 17610388]
26. Allen DN, Huegel SG, Seaton BE, Goldstein G, Gurklis JA Jr, van Kammen DP. Confirmatory factor analysis of the WAIS-R in patients with schizophrenia. *Schizophr Res*. 1998; 34(1–2):87–94. [PubMed: 9824880]
27. Dickinson D, Ramsey ME, Gold JM. Overlooking the obvious: a meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Arch Gen Psychiatry*. 2007; 64(5):532–42. [PubMed: 17485605]
28. Kay SR, Sevy S. Pyramidal model of schizophrenia. *Schizophr Bull*. 1990; 16:537–45. [PubMed: 2287938]
29. Revheim N, Schechter I, Kim D, Silipo G, Allingham B, Butler P, et al. Neurocognitive and symptom correlates of daily problem-solving skills in schizophrenia. *Schizophr Res*. 2006; 83(2–3):237–45. [PubMed: 16443347]
30. Boersma P, Weenink D. Praat: Doing phonetics by computer [Computer program]. 2008
31. Silver H, Shlomo N. Perception of facial emotions in chronic schizophrenia does not correlate with negative symptoms but correlates with cognitive and motor dysfunction. *Schizophr Res*. 2001; 52(3):265–73. [PubMed: 11705720]
32. Carter CS, Barch DM, Gur R, Pinkham A, Ochsner K. CNTRICS final task selection: social cognitive and affective neuroscience-based measures. *Schizophr Bull*. 2009; 35(1):153–62. [PubMed: 19011231]
33. Butler PD, Abeles IY, Weiskopf NG, Tambini A, Jalbrzikowski M, Legatt ME, et al. Sensory contributions to impaired emotion processing in schizophrenia. *Schizophr Bull*. 2009; 35(6):1095–107. [PubMed: 19793797]
34. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. 2. Hillsdale, NJ: Lawrence Erlbaum Assoc; 1988.
35. Green MF, Leitman DI. Social cognition in schizophrenia. *Schizophr Bull*. 2008; 34(4):670–2. [PubMed: 18495642]

36. Harvey PD, Penn D. Social cognition: the key factor predicting social outcome in people with schizophrenia? *Psychiatry (Edgmont)*. 2010; 7(2):41–4. [PubMed: 20376275]
37. Eack SM, Greeno CG, Pogue-Geile MF, Newhill CE, Hogarty GE, Keshavan MS. Assessing social-cognitive deficits in schizophrenia with the Mayer-Salovey-Caruso Emotional Intelligence Test. *Schizophr Bull*. 2010; 36(2):370–80. [PubMed: 18648021]
38. Wilk CM, Gold JM, McMahon RP, Humber K, Iannone VN, Buchanan RW. No, it is not possible to be schizophrenic yet neuropsychologically normal. *Neuropsychology*. 2005; 19(6):778–86. [PubMed: 16351353]
39. Revheim N, Medalia A. The independent living scales as a measure of functional outcome for schizophrenia. *Psychiatr Serv*. 2004; 55(9):1052–4. [PubMed: 15345767]
40. Green MF, Schooler NR, Kern RS, Frese FJ, Granberry W, Harvey PD, et al. Evaluation of functionally meaningful measures for clinical trials of cognition enhancement in schizophrenia. *Am J Psychiatry*. 2011; 168(4):400–7. [PubMed: 21285142]
41. Fisher M, Holland C, Merzenich MM, Vinogradov S. Using neuroplasticity-based auditory training to improve verbal memory in schizophrenia. *Am J Psychiatry*. 2009; 166(7):805–11. [PubMed: 19448187]
42. Philip RC, Whalley HC, Stanfield AC, Sprengelmeyer R, Santos IM, Young AW, et al. Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychol Med*. 2010; 40(11):1919–29. [PubMed: 20102666]



**Figure 1. Map of stimuli included in the full battery on mean base pitch (F0M) and pitch variability (F0SD) (A) and mean intensity (VIntM) and intensity variability (VIntSD) (B)**

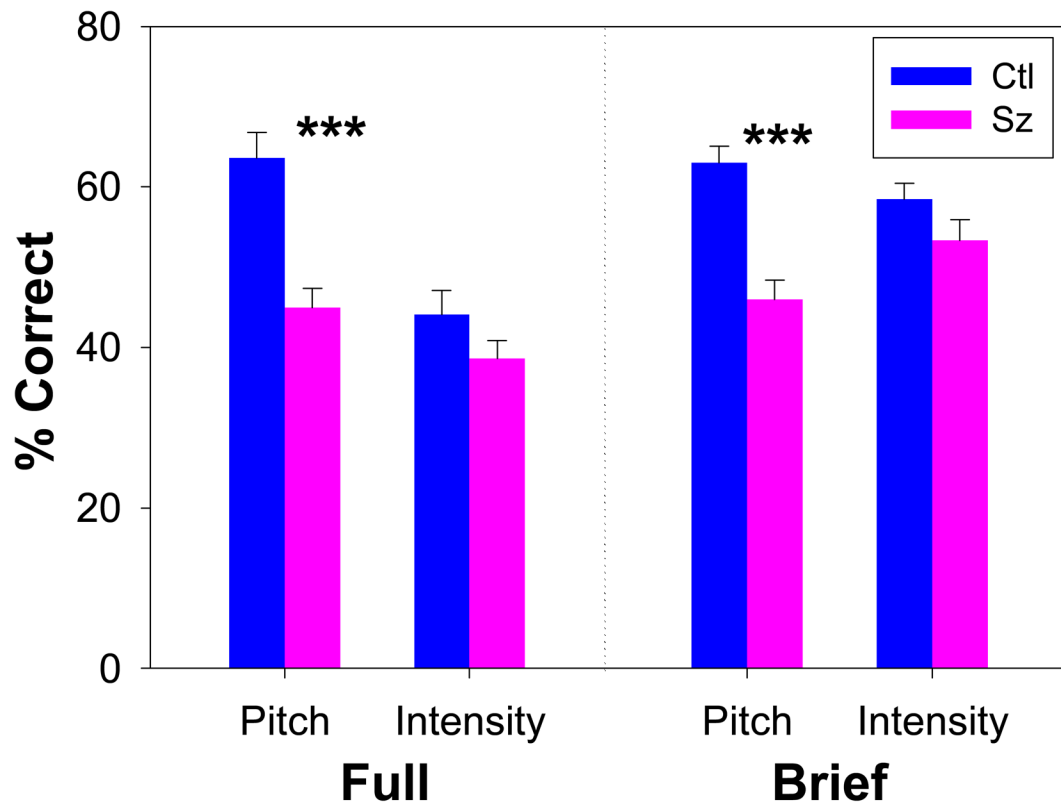
Variability in feature by pitch-based stimuli was determined by one-way ANOVAs across emotions. Among stimuli that were considered to be pitch-based, there was significantly variability in mean base pitch (F0M,  $p < .0001$ ) and pitch variability (F0SD,  $p < .0001$ ) but not mean voice intensity (VIntM,  $p = .13$ ) or intensity variability (VIntSD,  $p = .4$ ) (shown). Other variables (not shown) that showed significant variability across emotions were the floor frequency of the base pitch (F0floor, HF500 ( $p = .01$ ), mean pitch of  $p = .04$ ), maximum frequency of the base pitch (F0max,  $p < .0001$ ), pitch contour (F0contour,  $p = .02$ ) and mean pitch of the first formant (F1M,  $p = .006$ ). A discriminant function analysis with pairwise comparison demonstrated significant contribution of several pitch variables, including F0SD, F0Max, maximum frequency of the first formant (F1Max), and F0Contour differentiation of emotional stimuli. Neither VIntM or VIntSD contributed significantly to this discriminant function. When intensity-based stimuli as a group were compared to pitch-based stimuli, VIntM ( $p = .001$ ), VIntSD ( $p = .004$ ) (shown), and HF500 ( $p < .0001$ ), and mean bandwidth of the first formant (F1BW) ( $p = .011$ ) (not shown) were significantly different across stimuli. In contrast, pitch-based measures including F0M ( $p = .008$ ) and F0SD ( $p = .27$ ) were not different. A discriminant function showed significant contribution only of VIntM to differentiation of intensity- vs. pitch-based stimuli, with no further contribution of other intensity- or pitch-based variables.



**Figure 2.**

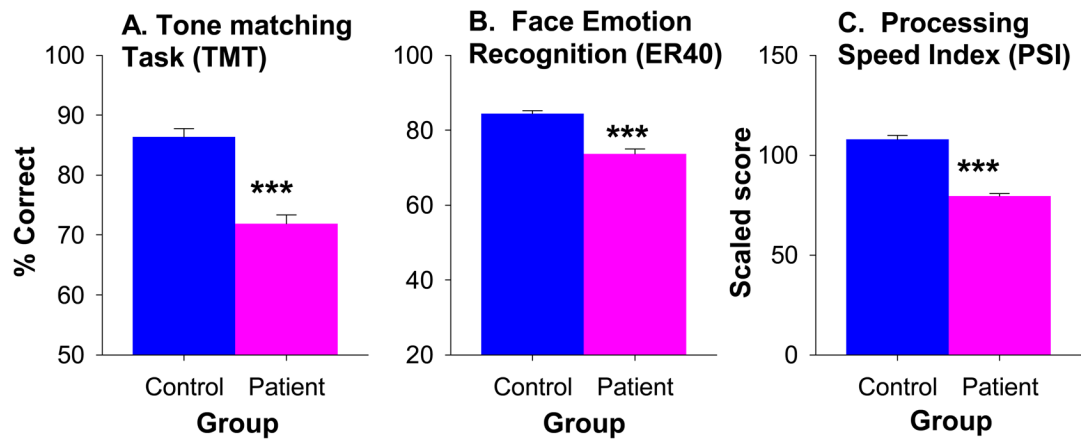
Relative between-group performance (mean  $\pm$  sem) on the full version of the auditory emotion recognition (AER) task.

\*\* $p < .01$  vs. control. \*\*\*  $p < .001$  vs. control

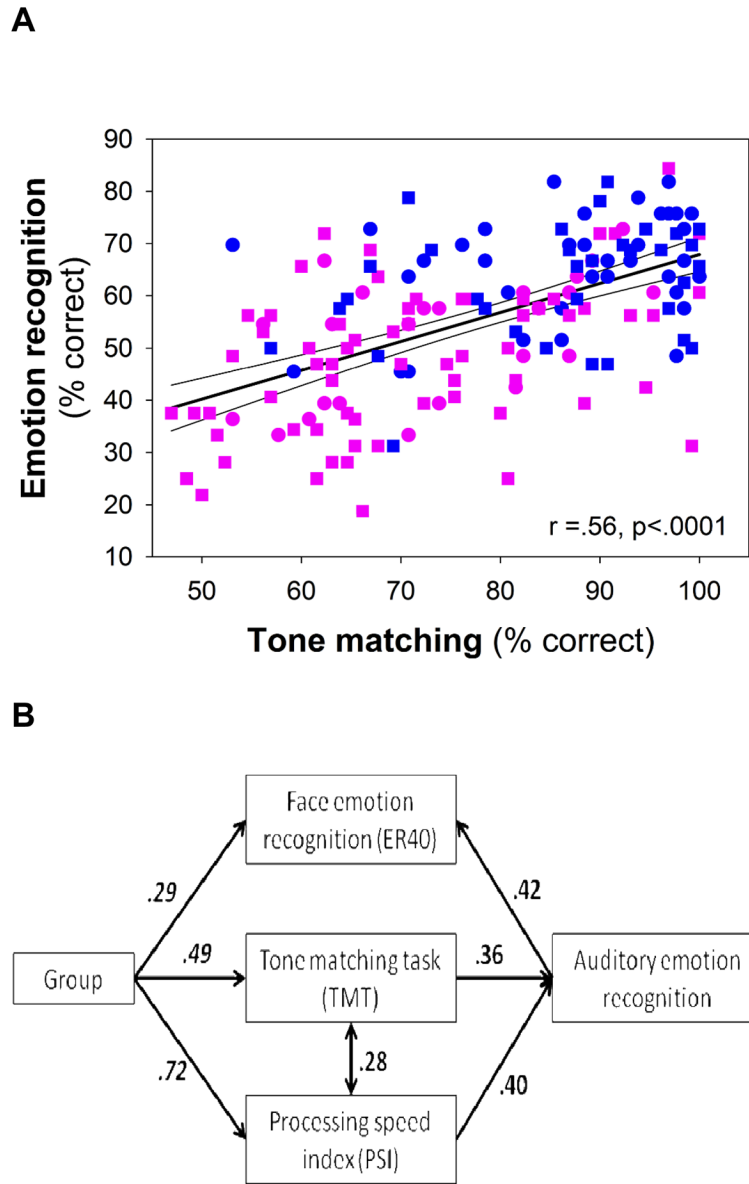


**Figure 3.** Relative between-group performance (mean  $\pm$  sem) to pitch- vs. intensity-based stimuli extracted from the full auditory emotion recognition battery (left) and from a brief replication battery (right), showing deficits to pitch- vs. intensity-based emotion recognition \*\*\* $p < .001$  vs. control





**Figure 4.** Relative between-group performance (mean  $\pm$  sem) in tone matching test (TMT), face emotion recognition (ER40) and WAIS-III Processing Speed Index (PSI)  
\*\*\* p<.001



**Figure 5.**  
 A. Correlation between tone matching and auditory emotion recognition (AER) performance across patients (pink) and controls (blue). Correlation was significant both across groups ( $r = .56$ ,  $n = 98$ ,  $p < .0001$ ) and within patients ( $r = .42$ ,  $n = 66$ ,  $p < .0001$ ) and controls ( $r = .49$ ,  $n = 32$ ,  $p = .004$ ) alone. Furthermore, correlations in both patients ( $p = .03$ ) and controls ( $p = .002$ ) remained significant even following co-variation for general cognitive dysfunction (PSI).  
 B: Path analysis demonstrating both sensory-specific (TMT) and general cognitive (PSI) contributions to impaired auditory emotion recognition in schizophrenia. Numbers represent standardized regression weights between indicated variables. Model fit parameters including residual Chi-square over degrees of freedom (CMIN/DF) = .91, Root mean square error of approximation (RMSEA) = 0 and Hoelter statistic (.05) = 560, suggest strong statistical model. Additional paths did not lead to further statistical improvement of the model fit.

**Table 1**

Demographic and clinical features, primary sample

Group	Control		Patient	
	Mean	SD	Mean	SD
Age	35.0	12.9	37.8	10.4
Gender (M/F)	45/28		79/13	
Handedness (R/L)	63/10		86/6	
Parental SES <sup>1</sup>	43.6	13.0	45.0	21.4
Individual SES <sup>1</sup>	44.6	9.3	26.6	11.9
PANSS scores <sup>2</sup>	---	---		
Total	---	---	71.4	13.6
Positive	---	---	18.1	5.5
Negative	---	---	18.0	4.5
General	---	---	35.4	13.6
Independent Living Scale – Problem Solving (29)	---	---	38.9	12.1
Antipsychotic dose	---	---	877	748

<sup>1</sup>Hollingshead socioeconomic status (ref)<sup>2</sup>Positive and negative symptom scale (28)

Demographic features and mean (sd) percent correct task performance on Brief Emotion Recognition (BER) battery, Tone Matching Test (TMT) and Visual Emotion Recognition (ER40) battery in the University of Pennsylvania replication sample

**Table 2**

Variable	Schizophrenia			Age-matched controls			General control		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Age	36	37.7	10.3	31	34.2	12.3	188	21.3	2.5
Gender (M/F)		21/15		17/14				91/97	
BER Overall	36	50.7	12.3	31	66.0	9.8	188	65.9	9.8
BER Intensity	36	47.2	16.5	31	57.1	14.2	188	57.6	15.5
BER Pitch	36	48.1	15.7	31	63.6	13.7	188	65.2	13.9
TMT	35	65.5	9.9	29	74.0	9.3	188	76.0	7.8
ER40	27	79.1	10.6	25	88.0	6.9	187	87.6	6.0