

Auditory Emotion Recognition Impairments in Schizophrenia: Relationship to Acoustic Features and Cognition

Rinat Gold, Ph.D.

Pamela Butler, Ph.D.

Nadine Revheim, Ph.D.

David I. Leitman, Ph.D.

John A. Hansen, Ph.D.

Ruben C. Gur, Ph.D.

Joshua T. Kantrowitz, M.D.

Petri Laukka, Ph.D.

Patrik N. Juslin, Ph.D.

Gail S. Silipo, M.A.

Daniel C. Javitt, M.D., Ph.D.

Objective: Schizophrenia is associated with deficits in the ability to perceive emotion based on tone of voice. The basis for this deficit remains unclear, however, and relevant assessment batteries remain limited. The authors 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 functioning.

Method: The authors studied a primary sample of 92 patients and 73 comparison subjects. Stimuli were characterized according to both intended emotion and acoustic 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 comparison subjects, and 188 general comparison subjects.

Results: Patients showed statistically significant large-effect-size deficits in voice emotion recognition ($d=1.1$) and were preferentially impaired in recognition of emotion based on pitch features but not intensity features. Emotion recognition deficits were significantly correlated with pitch perception impairments both across ($r=0.56$) and within ($r=0.47$) groups. 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 results demonstrate that patients with schizophrenia show a significant deficit in the ability to recognize emotion based on tone of voice and that this deficit is related to impairment in detecting the underlying acoustic features, such as change in pitch, required for auditory emotion recognition. This study provides tools for, and highlights the need for, greater attention to physical features of stimuli used in studying social cognition in neuropsychiatric disorders.

(*Am J Psychiatry* 2012; 169:424–432)

During human interaction, individuals communicate information not only verbally but also through tone of voice. 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, and shouting when angry (1, 2). Detection of this nonverbal information allows listeners to adjust their behavior accordingly and thus to perform adequately in social situations. Thus, impairments in auditory emotion recognition (AER) and social cognition contribute strongly to poor psychosocial outcome in schizophrenia (3–5). The pattern and basis of AER deficits are an area of active research, as is 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 Penn Emotion Recognition Task (ER-40) (6) and the Ekman 60 Faces Test (7). In the audito-

ry 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 studying AER impairments in schizophrenia, with no standardization of the psychoacoustic properties of stimuli across studies. Potentially as a result of this variation, different patterns of AER deficits have been reported (8–10), with some studies suggesting a generalized pattern of deficit (11) and others hemisphere- (12), emotion- (13), or valence-specific (14) patterns. We sought in this study to validate the use of a novel, psychoacoustically well-characterized auditory battery in a large group of schizophrenia patients as a means of investigating both the pattern and the underlying basis of social cognition impairments in schizophrenia.

In auditory communication, overlapping but distinct patterns of psychoacoustic features contribute to communication of discrete auditory emotional percepts (see

This article is featured in this month's AJP **Audio** and is discussed in an **Editorial** by Keefe and Kraus (p. 354)

reference 2 for a comprehensive review). For example, specific pitch-related vocal features such as mean pitch, pitch variability, and pitch contour are critical for communicating emotions such as happiness, sadness, and fear, whereas specific intensity-related aspects (i.e., loudness-related), such as mean voice intensity, intensity variability, and voice quality as reflected in the amount of high-frequency energy over 500 Hz, 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 by increased voice intensity (i.e., shouting, or “hot” anger) or by reduction in the mean base pitch of the voice without increasing intensity (irritation, or “cold” anger) (1). Similarly, conveyance of strong happiness (“elation”) may depend on somewhat different cues than conveyance of weaker forms of happiness (1) (for examples, see the data supplement that accompanies the online edition of this article). 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 a heterogeneity of findings across studies.

In 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 the primary auditory cortex. In a recent emotion study (23), we evaluated the performance of schizophrenia patients on a novel emotion recognition battery using psychoacoustically characterized stimuli and observed deficits in pitch-based but not intensity-based emotion processing. A similar pattern of deficit was observed more recently for emotion conveyed by frequency-modulated tones (24). In the present study, we extend these findings to a larger patient sample and validate a brief version of the test for widespread use.

Along with the specific sensory-level contributions to AER, other potential contributors include emotion-level dysfunction and general cognitive impairments. In this study, we evaluated visual emotion recognition using the ER-40 test (25), which includes emotions similar to those presented in our auditory emotion battery. We also included the processing speed index (PSI) from the WAIS-III, which contains tests that are thought to be particularly sensitive to generalized cognitive impairment in schizophrenia, such as the digit symbol subtest (26, 27).

We hypothesized that patients would show emotion recognition deficits for stimuli in which emotion was conveyed primarily by pitch-based but not intensity-based measures and that correlations between basic pitch processing and AER would remain significant even after covariation for deficits in nonauditory emotion recognition and general cognitive function. We also evaluated the replicability of the findings across two separate performance sites and thus applicability to general schizophrenia and neuropsychiatric populations.

Method

Participants

The primary sample included 92 chronically ill patients with schizophrenia or schizoaffective disorder (Table 1). Patients were drawn from chronic inpatient units and residential care facilities associated with the Nathan S. Kline Institute in New York. All were receiving conventional or atypical antipsychotics at the time of testing. Diagnoses were determined by the Structured Clinical Interview for DSM-IV. Comparison subjects (N=73) were volunteers who were on the staff or who responded to local advertisements. The groups did not differ on mean parental Hollingshead socioeconomic status (28), which reflects level of education and employment on a scale of 0–100. As expected, the patient group had a lower socioeconomic status on average than the comparison group ($p<0.001$) and parents ($p<0.0001$).

Clinical assessments included ratings on the Positive and Negative Syndrome Scale (PANSS) (29) and the problem-solving subscale of the Independent Living Scales (30). All study procedures received approval from the Institutional Review Board at the Nathan S. Kline Institute. All participants had the procedure explained to them verbally before giving written informed consent.

In the replication sample (Table 2), participants were drawn from psychiatric populations and both age-matched (N=31) and more general normative populations (N=188) associated with the University of Pennsylvania in Philadelphia. As no significant differences were observed between the two comparison groups on any of the dependent measures, the two groups were combined in statistical analyses.

Procedure

For the full version of the task, stimuli consisted of 88 audio recordings of native British English speakers conveying five emotions—anger, disgust, fear, happiness, and sadness—or no emotion, as previously described (2, 23). Acoustic features for these stimuli were measured using the Praat speech analysis software program (www.fon.hum.uva.nl/praat). (Sample stimuli are included in the online data supplement.)

For the brief version, a subset of 32 stimuli were selected incorporating all emotions except disgust, which was eliminated to decrease the number of choices and therefore the number of stimuli. Pitch-based stimuli (N=17) were selected based on a previous study showing that these stimuli were well recognized as expressing the intended emotion (i.e., >60% correct scores) (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 portrayals of anger and happiness (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 on overall intensity (Figure 1) and thus represent “hot” anger as opposed to “cold” pitch-based anger. Pitch-based and 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 battery with the original 88 items (67 schizophrenia patients and 32 comparison subjects) or the 32-item brief version (40 schizophrenia patients and 53 comparison subjects). Stimuli representing different emotions were intermixed and presented in consistent order across subjects. After each stimulus, subjects were asked to identify the emotion (six possible alternatives in the full version; five in the brief version) as well as the intensity of portrayal on a scale of 1

TABLE 1. Demographic and Clinical Characteristics of the Primary Sample in a Study of Auditory Emotion Recognition

Variable	Schizophrenia Group (N=92)		Comparison Group (N=73)	
	N	%	N	%
Male	79	85.9	45	61.6
Right-handed	86	93.5	63	86.3
	Mean	SD	Mean	SD
Age	37.8	10.4	35.0	12.9
Parental socioeconomic status ^a	45.0	21.4	43.6	13.0
Individual socioeconomic status ^a	26.6	11.9	44.6	9.3
Positive and Negative Syndrome Scale				
Total score	71.4	13.6		
Positive score	18.1	5.5		
Negative score	18.0	4.5		
General score	35.4	13.6		
Independent Living Scales, problem-solving subscale	38.9	12.1		
Antipsychotic dosage (mg/day chlorpromazine equivalents)	877	748		

^a Hollingshead scale, which reflects level of education and employment on a scale of 0–100.

TABLE 2. Task Performance (Percent Correct) on Auditory Emotion Recognition, Facial Emotion Recognition, and Tone Matching in the University of Pennsylvania Replication Sample^a

Measure	Schizophrenia Group (N=36)		Age-Matched Comparison Group (N=33)		General Comparison Group (N=188)	
	Mean	SD	Mean	SD	Mean	SD
Brief Auditory Emotion Recognition Battery						
Overall	50.7	12.3	66.0	9.8	65.9	9.8
Intensity	47.2	16.5	57.1	14.2	57.6	15.5
Pitch	48.1	15.7	63.6	13.7	65.2	13.9
Penn Emotion Recognition Task	79.1	10.6	88.0	6.9	87.6	6.0
Tone matching test	65.5	9.9	74.0	9.3	76.0	7.8

^a The schizophrenia group was 58% male, with a mean age of 37.7 years (SD=10.3); the age-matched comparison group was 55% male, with a mean age of 34.2 years (SD=12.3); and the general comparison group was 48% male, with a mean age of 21.3 (SD=2.5). For the tone matching test, data were available for 35 members of the schizophrenia group, 29 members of the age-matched comparison group, and all members of the general comparison group. For the Penn Emotion Recognition Task, data were available for 27 members of the schizophrenia group, 25 members of the age-matched comparison group, and 187 members of the general comparison group.

to 10. A limited number of participants received both the full and abbreviated versions of the task (15 schizophrenia patients and 12 comparison subjects). In this subgroup, no significant group-by-version interaction was observed. For subsequent analyses, these participants were counted only once, with data from the full battery being used for statistical analysis.

Tone matching was assessed using pairs of 100-msec tones in series, with a 500-msec intertone interval. Within each pair, tones (50% each) either were identical or differed in frequency by a specified amount in each block (2.5%, 5%, 10%, 20%, or 50%). Participants 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 five sets of 26 pairs of tones.

Recognition of visual emotion was assessed using the ER-40 (31–33). Global cognitive functioning was assessed using the WAIS-III PSI, which includes the widely studied digit symbol coding subtest (27) and the symbol search subtest.

Data Analysis

The accuracy of AER was assessed using repeated-measures analysis of variance with either emotion (happy, sad, anger, fear, disgust, no emotion) or feature (pitch, intensity) as within-sub-

ject factors and group (schizophrenia group, comparison group) as a between-subject factor. The relationship between emotion identification and specific predictors was assessed using analysis of covariance (ANCOVA) with tone matching, ER-40, and PSI included as potential covariates. Task version (full or brief) was also included in the analysis 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 directionality of relationship and 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-square goodness-of-fit parameter. Effect size measures were interpreted according to the convention established by Cohen (34). All statistical tests were two-tailed, with alpha set at 0.05, and were computed using SPSS, version 18.0 (SPSS, Chicago).

Results

Full Version

On the full battery, patients showed highly significant large-effect-size impairment across stimuli ($F=25.4$, $df=1$,

97, $p < 0.00001$; $d = 1.1$) (Figure 2). In contrast, the group-by-emotion interaction ($F = 2.09$, $df = 5$, 93, $p = 0.07$) fell short of statistical significance, suggesting statistically similar deficits across emotions. Despite their impairments, patients performed well above chance levels for all emotions.

When stimuli were divided according to underlying feature (pitch or intensity), a highly significant group effect was again observed ($F = 14.4$, $df = 1$, 97, $p = 0.0003$), as well as a highly significant group-by-feature interaction ($F = 7.79$, $df = 1$, 97, $p = 0.006$). Follow-up t tests showed a highly significant difference in detection of pitch-based emotion ($t = 4.51$, $df = 97$, $p < 0.0001$) but no significant difference in detection of intensity-based emotion ($t = 1.44$, $df = 97$, $p = 0.15$) (Figure 3). Mean performance levels across groups were not significantly different for pitch versus intensity stimuli, suggesting that group interactions were not due to floor or ceiling effects but that comparison subjects 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, $p < 0.0001$) and the group-by-feature interaction ($F = 9.61$, $df = 1$, 91, $p = 0.003$) were again significant, with significant between-group differences for pitch-based ($t = 5.32$, $df = 91$, $p < 0.0001$, $d = 1.1$) but not intensity-based ($t = 1.59$, $df = 91$, $p = 0.11$, $d = 0.33$) stimuli (Figure 3). In the brief battery, as in the full battery, there was no significant group-by-emotion interaction.

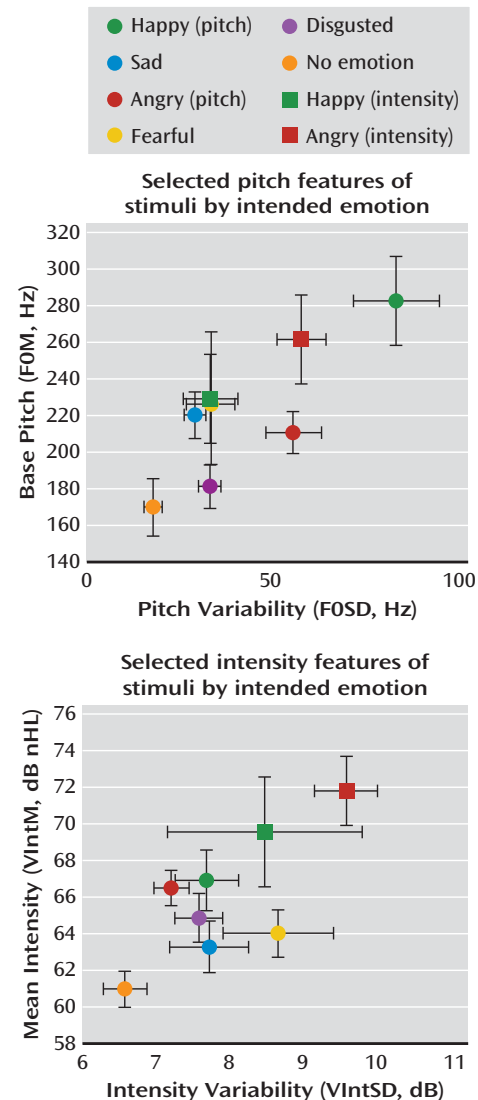
Relative Contributions of Sensory and General Cognitive Dysfunction

In addition to deficits in AER, patients showed highly significant deficits in the tone matching test, the ER-40, and the PSI (Figure 4). As predicted, the correlation between tone matching and AER was highly significant both across groups ($r = 0.56$, $N = 164$, $p < 0.0001$) (Figure 5A) and within patients ($r = 0.47$, $N = 91$, $p < 0.0001$) and comparison subjects ($r = 0.34$, $N = 73$, $p = 0.004$) independently.

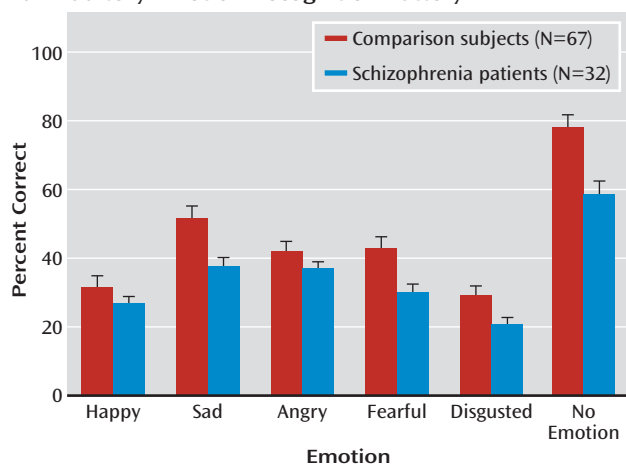
To evaluate the relative contribution of these measures, an ANCOVA was conducted incorporating group as a between-subject factor and the tone matching test, ER-40, and PSI as potential covariates. Both tone matching performance ($F = 8.72$, $df = 1$, 117, $p = 0.004$) and PSI ($F = 12.9$, $df = 1$, 117, $p < 0.0001$) correlated significantly with AER performance, whereas the correlation with ER-40 was nonsignificant. After effects of tone matching performance and PSI were accounted for, the main effect of group was no longer significant.

Finally, inclusion of these factors into a path analysis yielded a strong model confirming both tone matching performance and PSI as mediators of the group effect on AER and showing an interrelationship between tone matching performance and PSI. In the path analysis, a significant relationship between auditory and visual emo-

FIGURE 1. Pitch and Intensity Map of Stimuli Included in the Full Auditory Emotion Recognition Battery^a



^a Variability in feature by pitch-based stimuli was determined by one-way analyses of variance across emotions. dB nHL = decibels relative to normal hearing level. Among stimuli that were considered to be pitch based, there was significant variability in mean base pitch (F0M, $p < 0.0001$) and pitch variability (F0SD, $p < 0.0001$), but not mean intensity (VIntM, $p = 0.13$) or intensity variability (VIntSD, $p = 0.4$) (shown). Other variables (not shown) that showed significant variability across emotions were the floor frequency of the base pitch (F0Floor, $p = 0.037$), mean frequency of the base pitch (F0M, $p < 0.0001$), high frequency energy > 500 Hz (HF500, $p = 0.01$) maximum frequency of the base pitch (F0Max, $p < 0.0001$), pitch contour (F0Contour, $p = 0.02$), and mean pitch of the first formant (F1M, $p = 0.006$). A discriminant function analysis with pairwise comparison demonstrated significant contributions from several pitch variables, including F0SD, F0Max, maximum frequency of the first formant (F1Max), and F0Contour differentiation of emotional stimuli. Neither VIntM nor VIntSD contributed significantly to this discriminant function. When intensity-based stimuli as a group were compared to pitch-based stimuli, VIntM ($p = 0.001$), VIntSD ($p = 0.004$) (shown), HF500 ($p < 0.0001$), and mean bandwidth of the first formant (F1BW) ($p = 0.011$) (not shown) were significantly different across stimuli. In contrast, pitch-based measures including F0M ($p = 0.008$) and F0SD ($p = 0.27$) did not differ. A discriminant function showed a significant contribution only of VIntM to differentiation of intensity- versus pitch-based stimuli, with no further contribution from other intensity- or pitch-based variables.

FIGURE 2. Relative Between-Group Performance on the Full Auditory Emotion Recognition Battery^a

^a Significant difference between groups at $p < 0.01$ for fearful and disgusted and at $p < 0.001$ for sad and no emotion. Error bars indicate standard error of the mean.

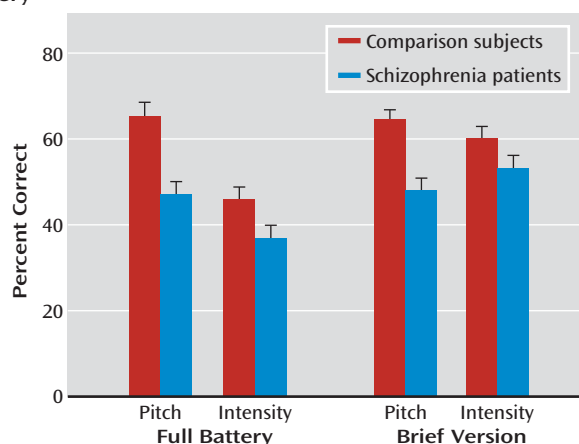
tion recognition was observed, with AER predicting performance on ER-40 (Figure 5B).

Validation of Pitch Versus Intensity Dichotomy

Tone matching measures were also used to validate the psychoacoustic dichotomization of stimuli into pitch based versus intensity based. An ANCOVA conducted across groups with tone matching performance as a covariate showed not only a significant effect of tone matching ($F=18.4$, $df=1$, 159 , $p < 0.0001$) but also a significant tone matching-by-feature interaction ($F=4.25$, $df=1$, 159 , $p=0.041$) reflecting a significantly stronger relationship between tone matching performance and accuracy in identifying pitch-based stimuli ($F=30.8$, $df=1$, 161 , $p < 0.0001$) than between tone matching performance and accuracy in identifying intensity based stimuli ($F=8.43$, $df=1$, 161 , $p=0.002$). When analyses were restricted to happy stimuli alone, an even stronger dissociation was observed, with a significant tone matching-by-feature interaction ($F=10.2$, $df=1$, 157 , $p=0.002$) and a significant relationship between tone matching performance and performance for pitch-based ($F=19.6$, $df=1$, 159 , $p < 0.0001$) but not intensity-based stimuli. Within patients alone, significant correlations were observed between tone matching performance and ability to detect pitch-based happiness ($r=0.38$, $df=90$, $p < 0.0001$) and anger ($r=0.30$, $df=65$, $p=0.017$), but not intensity-based emotions.

“Cold” Versus “Hot” Anger

Pitch versus 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) ($t=2.51$, $p=0.014$), but not by intensity (“hot” anger), although the group-by-feature interaction only ap-

FIGURE 3. Relative Between-Group Performance to Pitch Versus Intensity-Based Stimuli From the Full Auditory Emotion Recognition Battery and From a Brief Replication Battery^a

^a The results show deficits in pitch- versus intensity-based emotion recognition ($p < 0.001$). Error bars indicate standard error of the mean.

proached significance ($F=3.38$, $p=0.07$). Similarly, patients showed significant deficits in detection of happiness conveyed primarily by pitch ($t=2.57$, $p=0.011$) but not intensity (“elation”) modulation (see Table S1 in the online data supplement).

Auditory Versus Visual Emotion Recognition

On the ER-40 (see Table S2 in the online data supplement), patients showed significant impairments in detection of sadness ($p=0.003$), fear ($p < 0.001$), and no emotion ($p=0.003$), with deficits in detecting happiness ($p=0.07$) and anger ($p=0.06$) approaching significance. When correlations between AER and ER-40 were conducted for individual emotions within patients (see Table S3 in the online data supplement), the strongest correlations were found within emotion (mean $r=0.33$, $p < 0.01$), with lower correlations across emotion (mean $r=0.12$, n.s.).

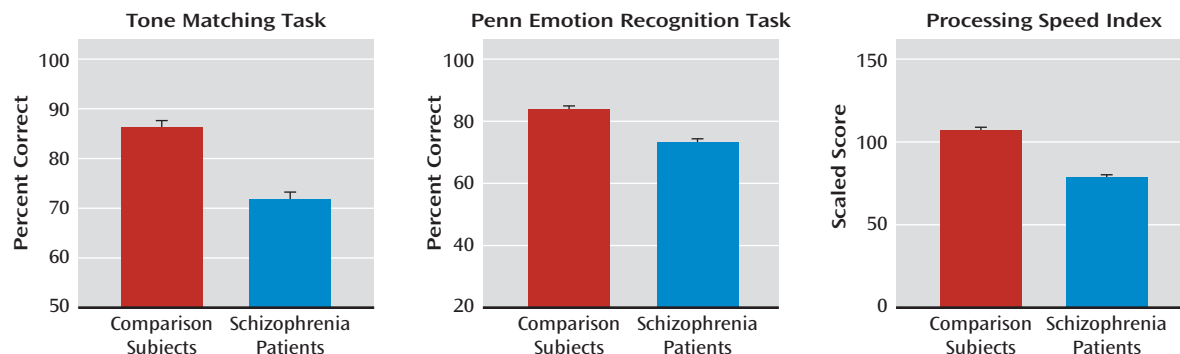
Correlation With Symptoms and Outcome

Deficits in AER correlated significantly with the cognitive factor of the PANSS ($r=-0.33$, $p=0.003$) but not with other PANSS factors. Deficits in emotion processing also correlated with standardized scores on the problem-solving subscale of the Independent Living Scales ($r=0.26$, $p=0.017$). Correlations with medication dosage, as assessed using chlorpromazine equivalents, were nonsignificant across all emotions.

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 , $p < 0.0001$; $d=1.49$), along with a significant group-by-feature interaction ($F=6.35$, $df=1$, 253 , $p=0.012$). In addition, tone matching performance significantly predicted AER performance over and above the ef-

FIGURE 4. Relative Between-Group Performance in Tone Matching, the Penn Emotion Recognition Task, and the WAIS-III Processing Speed Index^a



^a Significant difference between groups at $p < 0.001$ on all three measures. Error bars indicate standard error of the mean.

fect of group ($F=24.2$, $df=1$, 249 , $p<0.0001$). In contrast, as in the primary sample, the group-by-emotion interaction was not significant (see Table S4 in the online data supplement). The reliability of the measures across samples based on intraclass correlation was 0.97 for patients and 0.96 for comparison subjects.

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 on 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 (9). Moreover, the relative contributions of specific sensory features and more generalized cognitive performance remain largely unknown.

We assessed AER deficits in two independent samples of patients and comparison subjects using a novel, well-characterized battery in which the physical features of the stimuli were analyzed and in which stimuli were 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 observed previously (23), 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, we provide both a general and a brief AER battery for study across neuropsychiatric disorders.

In the battery, angry and happy stimuli were divided a priori into pitch- versus intensity-based exemplars based on 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

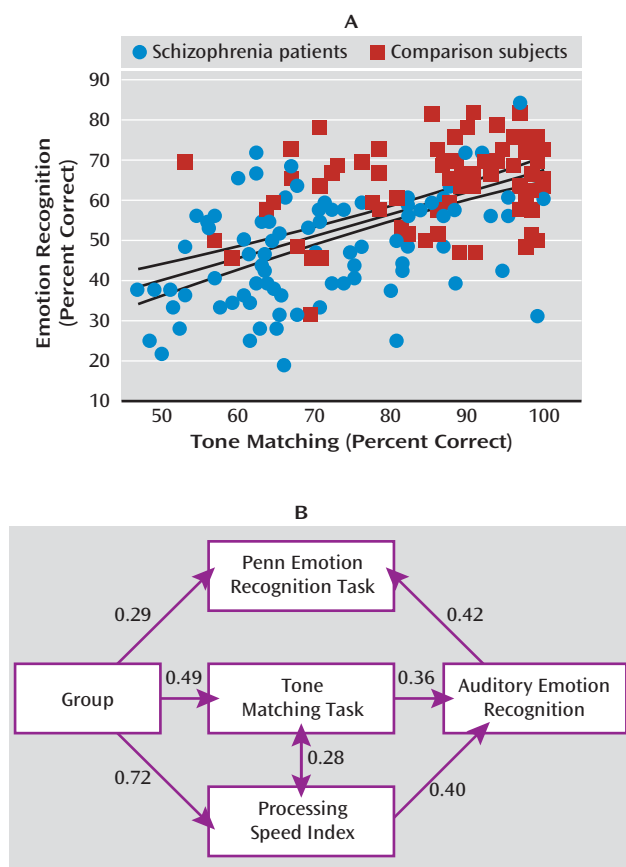
emotional information is conveyed by modulations in pitch rather than intensity. Significant group-by-stimulus feature interactions were found for both the full and brief versions of the battery and in both the primary and replication samples. The battery thus provides a replicable method both for characterizing sensory contributions to AER impairments in schizophrenia and for comparing specific patterns of dysfunction across neuropsychiatric illnesses.

In addition to differential analysis of deficits by pitch-based versus intensity-based characterization, we analyzed AER relative to tone matching performance, which provides an objective index of auditory sensory processing ability, and relative to both face emotion recognition (ER-40) and WAIS-III PSI, which provide measures of visual emotion and general cognitive dysfunction in schizophrenia, respectively (27, 38). The relative contributions of these measures to AER deficits were assessed using both multivariate regression and path analysis.

All three sets of measures (tone matching, ER-40, PSI) showed highly significant independent correlations to AER function across groups, with no further difference observed in AER function between schizophrenia patients and comparison subjects once these factors were taken into account. Approximately equal contributions were found for tone matching and PSI (Figure 5B), with AER deficits in turn predicting impairments in ER-40. 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 ER-40 performance all contributed significantly and independently to AER performance.

Finally, deficits in AER also correlated with score on the problem-solving subscale of the Independent Living Scales, a proxy measure for functional capacity (39, 40). Remediation of deficits in basic auditory processing has

FIGURE 5. Correlation Between Tone Matching and Auditory Emotion Recognition Performance Across Patients and Comparison Subjects and Path Analysis of Contributions to Impaired Auditory Emotion Recognition^a



^a In panel A, the correlation was significant both across groups ($r=0.56$, $N=98$, $p<0.0001$) and within patients ($r=0.42$, $N=66$, $p<0.0001$) and comparison subjects ($r=0.49$, $N=32$, $p=0.004$) alone. Furthermore, correlations in both patients ($p=0.03$) and comparison subjects ($p=0.002$) remained significant even following covariation for general cognitive dysfunction (processing speed index). In panel B, the path analysis demonstrates both sensory-specific (tone matching) and general cognitive (processing speed index) contributions to impaired auditory emotion recognition in schizophrenia. The numbers represent standardized regression weights between indicated variables. Model fit parameters (including residual chi-square over degrees of freedom [CMIN/DF=0.91], root mean square error of approximation [RMSEA=0], and the Hoelter 0.05 statistic [$N=560$]) suggest a strong statistical model. Additional paths did not lead to further statistical improvement of the model fit.

recently been found to induce improvement as well in global cognitive performance as measured using the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Consensus Cognitive Battery (41). Our results suggest that sensory-based remediation, along with specific emotion-based remediation, may be most useful for addressing social cognitive impairments in schizophrenia.

Based on our findings in this study, 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 are associated with AER deficits as indicated by performance on batteries similar to those used in schizophrenia (42). However, the specific pattern of deficit may differ from that in schizophrenia. Autism spectrum patients are reported to show most pronounced deficits in vocal perception of anger, fear, and disgust, with relatively spared perception of sadness (42). Our study suggests that dissociation across emotion in schizophrenia is not observed once the physical nature of the stimuli is considered. Comparison across populations, however, would be facilitated by use of a consistent battery with well-described physical features, such as the one we used in this study, in order to allow identification of the relative determinants of social cognition deficits across conditions.

Although our battery represents a significant advance over previous 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. Batteries in which actors purposely try to convey emotion by modulation of specific tonal or intensity-based features would give us even more ability to evaluate the differential mechanisms of emotion recognition dysfunction across diagnostic groups. Second, we used primarily a chronic, medicated patient population. Studies with prodromal or first-episode patients 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 this 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, the actors included in this battery spoke with a British accent, which may have influenced the results. Studies using actors speaking in a local accent would be desirable. Finally, other components of interpersonal interaction may also communicate emotion, such as body movement, context, and the 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. This study highlights substantial deficits in the ability of schizophrenia patients to decode specific stimulus features, such as pitch modulations, in interpreting emotion, leading to overall impairments in auditory emotion recognition. These deficits correlated with more basic impairments in sensory processing even when general cognitive and nonauditory 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 caregivers should be aware that pa-

tients 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.

Received Aug. 14, 2011; revision received Oct. 16, 2011; accepted Nov. 7, 2011 (doi: 10.1176/appi.ajp.2011.11081230). From the Center for Translational Schizophrenia Research, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, N.Y.; Department of Psychiatry, New York University, New York; Department of Psychiatry, Columbia University, New York; Department of Neuropsychiatry, University of Pennsylvania, Philadelphia; Department of Psychology, Stockholm University, Stockholm; Department of Psychology, Uppsala University, Uppsala, Sweden. Address correspondence to Dr. Javitt (Javitt@nki.rfmh.org).

Dr. Kantrowitz has conducted clinical research supported by Roche, Lilly, Sunovion, Novartis, and Pfizer and has served as a consultant to Agency Rx, Quadrant Health, and RTI Health Solutions; he and his spouse own shares of stock in GlaxoSmithKline. Dr. Javitt has received research grants from Jazz Pharmaceuticals, Pfizer, and Roche; has served as a consultant to AstraZeneca, Bristol-Myers Squibb, Cypress, Lilly, Lundbeck, Merck, NPS, Pfizer, Sanofi, Schering-Plough, Sepracor, Solvay, Takeda, and Sunovion; serves on the advisory board of Pro-mentis Pharmaceuticals; and has equity in Glytech, Inc. The other authors report no financial relationships with commercial interests.

The authors thank Joanna DiCostanza, Rachel Ziwich, and Jonathan Lehrfeld for their critical contributions to patient recruitment, assessment, and data management and Tracey Keel for administrative support. They also thank the faculty and staff of the Clinical Research and Evaluation Facility and the Outpatient Research Service at the Nathan S. Kline Institute for Psychiatric Research.

Supported in part by NIMH grants R37 MH049334, ARRA supplement R37 MH49334 S1, P50 MH086385, and P50 MH086385 S1 (to Dr. Javitt); R01 MH084848 (to Dr. Butler); MH084856 and MH060722 (to Dr. Gur); and by a NARSAD grant (to Dr. Leitman).

References

- Banase R, Scherer KR: Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* 1996; 70:614–636
- Juslin PN, Laukka P: Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol Bull* 2003; 129:770–814
- Kee KS, Green MF, Mintz J, Brekke JS: Is emotion processing a predictor of functional outcome in schizophrenia? *Schizophr Bull* 2003; 29:487–497
- Brekke J, Kay DD, Lee KS, Green MF: Biosocial pathways to functional outcome in schizophrenia. *Schizophr Res* 2005; 80:213–225
- Harvey PO, Bodnar M, Sergerie K, Armony J, Lepage M: Relation between emotional face memory and social anhedonia in schizophrenia. *J Psychiatry Neurosci* 2009; 34:102–110
- 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; 55:171–177
- Sparks A, McDonald S, Lino B, O'Donnell M, Green MJ: Social cognition, empathy, and functional outcome in schizophrenia. *Schizophr Res* 2010; 122:172–178
- Edwards J, Pattison PE, Jackson HJ, Wales RJ: Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophr Res* 2001; 48:235–253
- Edwards J, Jackson HJ, Pattison PE: Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review. *Clin Psychol Rev* 2002; 22:789–832
- Hoekert M, Kahn R, Pijnenborg M, Aleman A: Impaired recognition and expression of emotional prosody in schizophrenia: review and meta-analysis. *Schizophrenia Res* 2007; 96:135–145
- Chapman LJ, Chapman JP: The measurement of differential deficit. *J Psychiatr Res* 1978; 14:303–311
- Ross ED, Orbelo DM, Cartwright J, Hansel S, Burgard M, Testa JA, Buck R: Affective-prosodic deficits in schizophrenia: comparison to patients with brain damage and relation to schizophrenic symptoms. *J Neurol Neurosurg Psychiatry* 2001; 70:597–604
- Murphy D, Cutting J: Prosodic comprehension and expression in schizophrenia. *J Neurol Neurosurg Psychiatry* 1990; 53:727–730
- Bozikas VP, Kosmidis MH, Anezoulaki D, Giannakou M, Andreou C, Karavatos A: Impaired perception of affective prosody in schizophrenia. *J Neuropsychiatry Clin Neurosci* 2006; 18:81–85
- Juslin PN, Laukka P: Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion* 2001; 1:381–412
- Strous RD, Grochowski S, Cowan N, Javitt DC: Dysfunctional encoding of auditory information in schizophrenia. *Schizophr Res* 1995; 15:135
- Rabinowicz EF, Silipo G, Goldman R, Javitt DC: Auditory sensory dysfunction in schizophrenia: imprecision or distractibility? *Arch Gen Psychiatry* 2000; 57:1149–1155
- Holcomb HH, Ritzel 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; 57:75–82
- Wexler BE, Stevens AA, Bowers AA, Sernyak MJ, Goldman-Rakic PS: Word and tone working memory deficits in schizophrenia. *Arch Gen Psychiatry* 1998; 55:1093–1096
- Leitman DI, Hoptman MJ, Foxe JJ, Saccante E, Wylie GR, Nierenberg J, Jalbrzikowski M, Lim KO, Javitt DC: The neural substrates of impaired prosodic detection in schizophrenia and its sensorial antecedents. *Am J Psychiatry* 2007; 164:474–482
- Leitman DI, Wolf DH, Ragland JD, Laukka P, Loughhead J, Valdez JN, Javitt DC, Turetsky BI, Gur RC: "It's Not What You Say, But How You Say It": a reciprocal temporo-frontal network for affective prosody. *Front Hum Neurosci* 2010; 4:19
- Leitman DI, Wolf DH, Laukka P, Ragland JD, Valdez JN, Turetsky BI, Gur RE, Gur RC: Not pitch perfect: sensory contributions to affective communication impairment in schizophrenia. *Biol Psychiatry* 2011; 70:611–618
- Leitman DI, Laukka P, Juslin PN, Saccante E, Butler P, Javitt DC: Getting the cue: sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophr Bull* 2010; 36:545–556
- Kantrowitz JT, Leitman DI, Lehrfeld JM, Laukka P, Juslin PN, Butler PD, Silipo G, Javitt DC: Reduction in tonal discriminations predicts receptive emotion processing deficits in schizophrenia and schizoaffective disorder. *Schizophr Bull* (Epub ahead of print, Jul 1, 2011)
- Pinkham AE, Gur RE, Gur RC: Affect recognition deficits in schizophrenia: neural substrates and psychopharmacological implications. *Expert Rev Neurother* 2007; 7:807–816
- 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; 34:87–94
- 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; 64:532–542
- Hollingshead AB, Redlich FC: Schizophrenia and social structure. *Am J Psychiatry* 1954; 110:695–701
- Kay SR, Sevy S: Pyramidal model of schizophrenia. *Schizophr Bull* 1990; 16:537–545
- Revheim N, Schechter I, Kim D, Silipo G, Allingham B, Butler P, Javitt DC: Neurocognitive and symptom correlates of daily problem-solving skills in schizophrenia. *Schizophr Res* 2006; 83:237–245

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; 52:265–273
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; 35:153–162
33. Butler PD, Abeles IY, Weiskopf NG, Tambini A, Jalbrzikowski M, Legatt ME, Zemon V, Loughhead J, Gur RC, Javitt DC: Sensory contributions to impaired emotion processing in schizophrenia. *Schizophr Bull* 2009; 35:1095–1107
34. Cohen J: *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, Lawrence Erlbaum Associates, 1988
35. Green MF, Leitman DI: Social cognition in schizophrenia. *Schizophr Bull* 2008; 34:670–672
36. Harvey PD, Penn D: Social cognition: the key factor predicting social outcome in people with schizophrenia? *Psychiatry* (Edgmont) 2010; 7:41–44
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; 36:370–380
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; 19:778–786
39. Revheim N, Medalia A: The Independent Living Scales as a measure of functional outcome for schizophrenia. *Psychiatr Serv* 2004; 55:1052–1054
40. Green MF, Schooler NR, Kern RS, Frese FJ, Granberry W, Harvey PD, Karson CN, Peters N, Stewart M, Seidman LJ, Sonnenberg J, Stone WS, Walling D, Stover E, Marder SR: Evaluation of functionally meaningful measures for clinical trials of cognition enhancement in schizophrenia. *Am J Psychiatry* 2011; 168:400–407
41. Fisher M, Holland C, Merzenich MM, Vinogradov S: Using neuroplasticity-based auditory training to improve verbal memory in schizophrenia. *Am J Psychiatry* 2009; 166:805–811
42. Philip RC, Whalley HC, Stanfield AC, Sprengelmeyer R, Santos IM, Young AW, Atkinson AP, Calder AJ, Johnstone EC, Lawrie SM, Hall J: Deficits in facial, body movement, and vocal emotional processing in autism spectrum disorders. *Psychol Med* 2010; 40:1919–1929