Reduction in Tonal Discriminations Predicts Receptive Emotion Processing Deficits in Schizophrenia and Schizoaffective Disorder

Joshua T. Kantrowitz*1,2, David I. Leitman3, Jonathan M. Lehrfeld1, Petri Laukka4, Patrik N. Juslin5, Pamela D. Butler1,6, Gail Silipo1, and Daniel C. Javitt1,6

1Schizophrenia Research Center, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962; 2Department of Psychiatry, Columbia University, New York, NY; 3Department of Neuropsychiatry, University of Pennsylvania, Philadelphia, PA; 4Department of Psychology, Stockholm University, Stockholm, Sweden; 5Department of Psychology, Uppsala University, Uppsala, Sweden; 6Department of Psychiatry, New York University, New York, NY

*To whom correspondence should be addressed; tel: 845-398-5503, fax: 845-398-6545, e-mail: jkantrowitz@nki.rfmh.org

Introduction: Schizophrenia patients show decreased ability to identify emotion based upon tone of voice (voice emotion recognition), along with deficits in basic auditory processing. Interrelationship among these measures is poorly understood. Methods: Forty-one patients with schizophrenia/schizoaffective disorder and 41 controls were asked to identify the emotional valence (happy, sad, angry, fear, or neutral) of 38 synthesized frequency-modulated (FM) tones designed to mimic key acoustic features of human vocal expressions. The mean (F0M) and variability (F0SD) of fundamental frequency (pitch) and absence or presence of high frequency energy (HF500) of the tones were independently manipulated to assess contributions on emotion identification. Forty patients and 39 controls also completed tone-matching and voice emotion recognition tasks. Results: Both groups showed a nonrandom response pattern (P < .0001). Stimuli with highest and lowest F0M/F0SD were preferentially identified as happy and sad, respectively. Stimuli with low F0M and midrange F0SD values were identified as angry. Addition of HF500 increased rates of angry and decreased rates of sad identifications. Patients showed less differentiation of response across frequency changes, leading to a highly significant between-group difference in response pattern to maximally identifiable stimuli (d = 1.4). The differential identification pattern for FM tones correlated with deficits in basic tone-matching ability (P = .01), voice emotion recognition (P < .001), and negative symptoms (P < .001). Conclusions: Specific FM tones conveyed reliable emotional percepts in both patients and controls and correlated highly with deficits in ability to recognize information based upon tone of voice, suggesting significant bottom-up contributions to social cognition and negative symptom impairments in schizophrenia.

Key words: affective prosody/early sensory processing/social cognition

Introduction

In most languages, emotion is conveyed by variation of tone of voice, leading to the process of receptive (vocal or voice) emotion recognition.1,2 Interpretation of tone of voice, in turn, allows individuals to infer a speaker’s true emotional state even when the verbal content of speech is itself neutral. Voice emotion recognition (also referred to as “auditory emotion recognition” or “receptive affective prosody”) is thus a critical feature of human interaction and a key contributor to the construct of social cognition.3 Individuals with schizophrenia show severe impairments in voice emotion recognition,4–9 along with broader deficits in social cognition.10–13 The present study investigates determinants of voice emotion recognition in schizophrenia using a novel behavioral task designed to probe the underlying features important for communicating emotion.

A complexity of voice emotion recognition is that no single aspect of speech, by itself, conveys any single emotion.1,2 Instead, emotion is conveyed by a complex constellation of acoustic features that are decoded simultaneously. These features result in turn from the complex physical changes that occur in the vocal apparatus in response to emotional events (see ref.1,2 for review). Thus, constriction of the throat during sadness may lead to both a lowering of the mean fundamental frequency of the voice (base pitch or F0M) as well as reduction in the variability of pitch over the course of an utterance (F0SD). In contrast, opening of the throat during happiness may lead to increases in both F0M and F0SD, resulting in a higher pitch.
Anger is most typically expressed through increased volume (intensity) of the voice (ie, shouting), which is also often accompanied by an increase in the relative proportion of high frequency noise above, vs below, 500 Hz (HF500), reflecting straining of the vocal cords ("hot" anger) and making the voice quality sound "sharper." However, anger may also be expressed through modulation of acoustic features in the absence of increase in intensity or voice quality. This form of anger, termed "cold," irritated or seething anger, is expressed primarily by reduction in F0M in absence of lowering F0SD and may represent an active attempt to control the manifestations of hot anger.\(^1\)\(^4\)\(^,\)\(^5\)\(^,\)\(^6\) Emotions are also conveyed by more complex alterations in acoustic features, such as changes in rate, tonal contours, or harmonic patterns.\(^1\)\(^,\)\(^2\)

In a prior study,\(^1\)\(^6\) we observed that schizophrenia patients showed reduced sensitivity to happiness and sadness and to "cold" but not "hot" anger, leading to the suggestion that voice emotion recognition deficits may reflect reduced ability to process pitch-based prosodic features, such as F0M or F0SD. Furthermore, deficits in emotion recognition correlated with impairments in the ability to match the pitch of 2 consecutively played "simple" tones.\(^1\)\(^7\)

However, because stimuli consisted of speech (ie, actors portrayals of emotion), it was only possible to determine acoustic features of the stimuli post hoc rather than to specify them in advance. Moreover, the range of features available was limited by the range of features that actors chose to include in their portrayals. Finally, individual features covary extensively in speech, making it difficult to isolate the contribution of any specific individual feature. The present study utilizes synthesized frequency modulated (FM) tones to assess contributions of specific acoustic features to voice emotion recognition deficits in schizophrenia. The tones were designed to mimic the key acoustic features that differentiated controls and patients in our earlier study, but with specific features manipulated independently.

We varied 2 pitch features of these synthesized FM tones—mean pitch (F0M), pitch variability (F0SD) and 1 voice quality (sharpness) feature—high frequency energy (HF500). Based upon our earlier studies, we hypothesized that stimuli with highest pitch would be recognizable as happy despite lack of linguistic structure, whereas those with lowest pitch would be recognizable as sad and those with low mean pitch but moderate pitch variability would be recognizable as angry. We further hypothesized that increasing voice sharpness by adding HF500 would increase the percept of anger. Based upon our study with human speech, we hypothesized that patients with schizophrenia would show reduced ability to utilize the pitch-based features (F0M and F0SD), whereas their utilization of the HF500 feature would be relatively intact.

In addition to analyzing the overall pattern of response across groups using categorical analyses (ie, \(\chi^2\)), we further analyzed the percentage of happy, sad, and angry responses to the best exemplars of each individual emotion for both between-group and within-group correlational analyses. Using this approach, we hypothesized that reduced ability to differentiate emotions for synthesized FM tones would correlate with well-replicated deficits in both basic tone matching and with deficits in more complex emotion recognition based on tone of voice.

**Methods**

**Subjects**

Subjects consisted of 41 medicated patients recruited from chronic inpatient (\(n = 22\)) and supervised residential sites (\(n = 19\)) associated with the Nathan Kline Institute (NKI) and 41 controls recruited from the healthy volunteer pool at NKI. There were no statistical differences (all \(P > .2\)) on auditory tasks between in/outpatients. Two controls and 1 patient did not participate in the ancillary emotional prosody and tone matching tasks. All subjects signed informed consent.

All patients met *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)* criteria for either schizophrenia (\(n = 33\)) or schizoaffective disorder (\(n = 8\)). We excluded controls with a history of an Axis I psychiatric disorder, as defined by the Structured Clinical Interview for DSM-IV.\(^1\)\(^8\) Patients and controls were excluded if they had any neurological or auditory disorders noted on medical history or in prior records, or for alcohol, or substance dependence within the last 6 months and/or abuse within the last month.\(^1\)\(^8\) A subsample was interviewed using semi-structured clinical interviews (Positive and Negative Symptom Scale [PANSS])\(^1\)\(^9\) and the processing speed index [PSI] of the Wechsler Adult Intelligence Scale-3\(^2\)\(^0\) by trained raters (intraclass correlation coefficient = 0.8 \(j\)). All data in text are mean \(\pm\) SD.

Patients had total, positive, negative, and general PANSS scores (\(n = 32\)) of 70.6 \(\pm 14.5\), 17.8 \(\pm 6.0\), 17.5 \(\pm 4.0\), 35.4 \(\pm 8.0\), respectively, suggesting mild to moderate illness severity. Demographics are presented in table 1. Although there were a higher proportion of males in the patient group, results remained statistically significant when gender was included as a covariate.

**Auditory Tasks**

All auditory tasks were presented on a CD player at a sound level that was comfortable for each listener in a sound-attenuated room. Responses were verbal without time limit.

**FM Tone Task.** Complex FM tones were synthesized (Adobe Audition) to be 500 ms in length with
Table 1. Study Demographics

<table>
<thead>
<tr>
<th></th>
<th>Control (41)</th>
<th>Patient (41)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>39.1 ± 13.7</td>
<td>36.5 ± 10.9</td>
</tr>
<tr>
<td>Male (%)*</td>
<td>80</td>
<td>53</td>
</tr>
<tr>
<td>Highest education*</td>
<td>14.9 ± 2.1</td>
<td>12.3 ± 2.3</td>
</tr>
<tr>
<td>High school graduate (%)*</td>
<td>100</td>
<td>73</td>
</tr>
<tr>
<td>Parental SESb*</td>
<td>43.8 ± 14.8 (38)</td>
<td>37.6 ± 17.1 (30)</td>
</tr>
<tr>
<td>Participant SESb*</td>
<td>45.1 ± 9.3 (40)</td>
<td>28.7 ± 12.8 (40)</td>
</tr>
<tr>
<td>Processing speed</td>
<td>109.7 ± 16.7 (39)</td>
<td>81.1 ± 12.5 (39)</td>
</tr>
<tr>
<td>PSI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of illness</td>
<td>NA</td>
<td>16.0 ± 10.0</td>
</tr>
<tr>
<td>Atypical (%)</td>
<td>NA</td>
<td>88</td>
</tr>
<tr>
<td>Frequency-modulated tone task (%)</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>Voice emotion recognition and tone matching (%)</td>
<td>39</td>
<td>40</td>
</tr>
</tbody>
</table>

Note: NA, Not applicable.
*Available n.
bSES, socioeconomic status, as measured by the 4-factor Hollingshead Scale.
P < 0.05 on Mann-Whitney for categorical values and independent sample t-tests for continuous values.

Table 2. Stimuli Used for Frequency-Modulated Tone Composite Score

<table>
<thead>
<tr>
<th></th>
<th>Base Tone (%)</th>
<th>With High Frequency Energy (HF500) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>225/150 (55)a</td>
<td>225/175 (63)</td>
</tr>
<tr>
<td></td>
<td>378/80 (46)</td>
<td>378/125 (54)</td>
</tr>
<tr>
<td></td>
<td>378/125 (54)</td>
<td>378/150 (61)</td>
</tr>
<tr>
<td></td>
<td>378/150 (49)</td>
<td>378/175 (60)</td>
</tr>
<tr>
<td></td>
<td>378/175 (59)</td>
<td></td>
</tr>
<tr>
<td>Sad</td>
<td>125/20 (61)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>125/40 (49)</td>
<td></td>
</tr>
</tbody>
</table>

aAll stimuli used for calculation of composite scores were identified with >45% consistency in both groups vs chance performance of 20%. Across-group % identification scores are shown in parentheses.

Voice Emotion Recognition. Voice emotion recognition was assessed using 32 stimuli from Juslin and Laukka’s emotional prosody task, as described previously.16 The sentences were scored based on the speaker’s intended emotion (happy, sad, angry, fear, or neutral). The sentences were semantically neutral and consisted of both statements and questions (ie, “It is 11 o’clock”, “Is it 11 o’clock?”). Correct percent responses were analyzed across groups.

Statistical Analyses
Performance on the FM tone task was compared with chance performance (ie, 20% for each emotion) using \( \chi^2 \) for proportions across subjects, followed-up by separate repeated measure ANOVA (group by stimulus-type [eg, the 19 stimuli in online supplementary table 2]) for each of the individual emotions. Primary analyses were conducted on tones presented with no added HF500 (base tones). Subsequent analyses evaluated the additional effect of added HF500.

Between-group comparisons were performed using independent-sample t-tests for continuous values and the nonparametric Mann-Whitney for categorical values. Secondary repeated measure ANOVA evaluated alteration in response both as a function of F0SD while F0M was constant across stimuli and for F0SD when F0M was constant across stimuli (eg, group by F0M/ F0SD). Analysis was restricted to F0M/F0SD values associated with clear emotional characterization in our previous study (eg, 378 F0M and 125 F0SD for happy, 125 F0M and 20 F0SD for sad, and 125 F0M for angry).

Relationship among measures was determined by Pearson correlations and multivariate linear regression,

a modulation frequency of 3 Hz (1½ cycles per tone). Stimuli differed in F0M (125, 225, or 378 Hz), F0SD (20, 40, 60, 80, 125, 150, and 175), and the presence or absence of added high frequency noise (HF500) and approximated the range of mean fundamental frequencies (F0M) and frequency variability (F0SD) associated with vocal affective prosody in previous studies.16,21 HF500 was added as regularly distributed noise. Examples of stimuli and spectra are provided as online supplementary table 1 and online supplementary figure 1.

Stimuli were presented sequentially in a randomly generated standard order. Subjects were asked to identify the emotional percept corresponding to each stimulus by choosing 1 of 5 alternatives (happy, sad, anger, fear, or neutral).16 In order to create a single dependent measure for the FM tone task, percent response measures (see online supplementary tables 2 and 3) were averaged across emotions for stimuli that elicited the most consistent predominant responses (operationalized post hoc as >45% across groups: table 2). The outcome measure for each individual subject was the percentage of responses that agreed with the predominant response across these selected stimuli (composite score).

Simple Tone Matching. Pitch processing was obtained using a simple tone-matching task.16 This task consists of pairs of 100-ms tones in series, with 500-ms intertone interval. Within each pair, tones are either identical or differ in frequency by specified amounts in each block (2.5%, 5%, 10%, 20%, or 50%). In each block, 12 of the tones are identical and 14 are dissimilar. Tones are derived from 3 reference frequencies (500, 1000, and 2000 Hz) to avoid learning effects. In all, the test consisted of 5 blocks of 26 pairs of tones.

Voice Emotion Recognition. Voice emotion recognition was performed on a simple tone-matching task.16 This task consisted of 5 blocks of 26 pairs of tones.

Notes:
- NA, Not applicable.
- Available n.
- SES, socioeconomic status, as measured by the 4-factor Hollingshead Scale.
- P < 0.05 on Mann-Whitney for categorical values and independent sample t-tests for continuous values.

Study Demographics

<table>
<thead>
<tr>
<th></th>
<th>Control (41)</th>
<th>Patient (41)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>39.1 ± 13.7</td>
<td>36.5 ± 10.9</td>
</tr>
<tr>
<td>Male (%)*</td>
<td>80</td>
<td>53</td>
</tr>
<tr>
<td>Highest education*</td>
<td>14.9 ± 2.1</td>
<td>12.3 ± 2.3</td>
</tr>
<tr>
<td>High school graduate (%)*</td>
<td>100</td>
<td>73</td>
</tr>
<tr>
<td>Parental SESb*</td>
<td>43.8 ± 14.8 (38)</td>
<td>37.6 ± 17.1 (30)</td>
</tr>
<tr>
<td>Participant SESb*</td>
<td>45.1 ± 9.3 (40)</td>
<td>28.7 ± 12.8 (40)</td>
</tr>
<tr>
<td>Processing speed</td>
<td>109.7 ± 16.7 (39)</td>
<td>81.1 ± 12.5 (39)</td>
</tr>
<tr>
<td>PSI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of illness</td>
<td>NA</td>
<td>16.0 ± 10.0</td>
</tr>
<tr>
<td>Atypical (%)</td>
<td>NA</td>
<td>88</td>
</tr>
<tr>
<td>Frequency-modulated tone task (%)</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>Voice emotion recognition and tone matching (%)</td>
<td>39</td>
<td>40</td>
</tr>
</tbody>
</table>

Note: NA, Not applicable.
*Available n.
bSES, socioeconomic status, as measured by the 4-factor Hollingshead Scale.
P < 0.05 on Mann-Whitney for categorical values and independent sample t-tests for continuous values.

Stimuli Used for Frequency-Modulated Tone Composite Score

<table>
<thead>
<tr>
<th></th>
<th>Base Tone (%)</th>
<th>With High Frequency Energy (HF500) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>225/150 (55)a</td>
<td>225/175 (63)</td>
</tr>
<tr>
<td></td>
<td>378/80 (46)</td>
<td>378/125 (54)</td>
</tr>
<tr>
<td></td>
<td>378/125 (54)</td>
<td>378/150 (61)</td>
</tr>
<tr>
<td></td>
<td>378/150 (49)</td>
<td>378/175 (60)</td>
</tr>
<tr>
<td></td>
<td>378/175 (59)</td>
<td></td>
</tr>
<tr>
<td>Sad</td>
<td>125/20 (61)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>125/40 (49)</td>
<td></td>
</tr>
</tbody>
</table>

aAll stimuli used for calculation of composite scores were identified with >45% consistency in both groups vs chance performance of 20%. Across-group % identification scores are shown in parentheses.

Voice Emotion Recognition. Voice emotion recognition was assessed using 32 stimuli from Juslin and Laukka’s emotional prosody task, as described previously.16 The sentences were scored based on the speaker’s intended emotion (happy, sad, angry, fear, or neutral). The sentences were semantically neutral and consisted of both statements and questions (ie, “It is 11 o’clock”, “Is it 11 o’clock?”). Correct percent responses were analyzed across groups.

Statistical Analyses
Performance on the FM tone task was compared with chance performance (ie, 20% for each emotion) using \( \chi^2 \) for proportions across subjects, followed-up by separate repeated measure ANOVA (group by stimulus-type [eg, the 19 stimuli in online supplementary table 2]) for each of the individual emotions. Primary analyses were conducted on tones presented with no added HF500 (base tones). Subsequent analyses evaluated the additional effect of added HF500.

Between-group comparisons were performed using independent-sample t-tests for continuous values and the nonparametric Mann-Whitney for categorical values. Secondary repeated measure ANOVA evaluated alteration in response both as a function of F0SD while F0M was constant across stimuli and for F0SD when F0M was constant across stimuli (eg, group by F0M/ F0SD). Analysis was restricted to F0M/F0SD values associated with clear emotional characterization in our previous study (eg, 378 F0M and 125 F0SD for happy, 125 F0M and 20 F0SD for sad, and 125 F0M for angry).

Relationship among measures was determined by Pearson correlations and multivariate linear regression,
as indicated. Within linear regression, partial correlations were used to assess significance of association. Two-tailed statistics are used throughout with a significance level of $P < .05$. Between group effect sizes (Cohen’s $d$) were calculated.\textsuperscript{22}

Results

**FM Tone Task**

Primary analyses were conducted on FM tones that were synthesized to vary systematically as a function of F0M and F0SD (base tones). Visualization of response pattern was accomplished using contour plots (figure 1), with the full data set presented in online supplementary table 2. As expected based upon studies with speech,\textsuperscript{1,2,16} tones with the highest levels F0M and F0SD were identifiable as happy in both groups (figure 1A and B, pink), whereas those with lowest F0M/F0SD levels were identifiable as sad (figures 1A and B, blue). In addition, several stimuli were consistently identified as angry in controls, but not patients (figure 1A, red). No regions were present in the schizophrenia map that were not also present in the control map.

These visual impressions were supported statistically using $\chi^2$ and repeated-measures ANOVA analyses. In the $\chi^2$ analyses, we found a significantly nonrandom pattern of response in both controls ($\chi^2 = 1236; df = 72, P < .0001$) and patients ($\chi^2 = 396; df = 72, P < .0001$), indicating that different FM tones reliably convey different emotions in both groups. To assess for between group differences in patterns for each individual emotion, we then used repeated measure ANOVA across all 19 stimuli (see online supplementary table 2), which varied across both F0M and F0SD.

---

**Fig. 1.** (A–B): Contour maps showing the pattern of responses for emotions across changes in mean fundamental frequency (F0M) (X-axis) and variability (F0SD) (Y-axis) for controls (A) and patients (B) for frequency-modulated tone stimuli (without HF500). For illustration purposes, a cutoff of 30% is used to delineate regions relative to the chance performance level (20%). The numbers on the colored lines indicate the percent response associated with that area. Neutral is not shown for clarity. Full data set found in online supplementary table 2. (C–G): Bar graphs of the percent of subjects rating a stimulus a particular emotion across a given change in mean fundamental frequency (F0M) and variability (F0SD), and dashed line passing through chance level performance (20%). Happy = pink, sad = blue, angry = red, fear = green, and neutral = gray. Controls are on left in lighter shade. $^*P < .05$, $^{**}P < .01$, $^{***}P < .001$ patients vs controls (Mann-Whitney).
A significant group by stimulus-type interaction was found for ratings of happy ($F_{18,63} = 2.7; P = .002$) and sad ($F_{18,63} = 2.3; P = .008$), with a trend toward significance for ratings of angry ($F_{18,63} = 1.7; P = .056$). As expected, there were no significant group by stimulus-type interactions for ratings of fear ($F_{18,63} = 1.3; P = .212$) or neutral ($F_{18,63} = 0.9; P = .628$).

Analysis by Pitch-Based Parameters (F0M and F0SD)

Given the overall significant group by stimulus-type effects, secondary analyses evaluated the specific contributions of variation in the pitch-based parameters F0M and F0SD to the interaction. Based on a priori hypothesis as well as lack of group by stimulus-type differences for fear and neutral ratings, secondary analyses were restricted to happy, sad, and angry.

To isolate effects of specific F0M and F0SD levels, analyses by 1 parameter (F0M or F0SD) were conducted across fixed levels of the other. Thus, across stimuli with an F0M level of 378 Hz (figure 1C), likelihood that stimuli would be recognized as happy increased progressively as the 7 levels of F0SD increased across both groups (main effect of F0SD = $F_{6,75} = 9.5; P < .001$). However, the degree of change was significantly less for patients leading to a significant group by F0SD level interaction ($F_{6,75} = 2.9; P = .014$). Similarly, when only stimuli with F0SD of 125 Hz were considered (figure 1D), there was a significant change in likelihood of happy ratings with increasing F0M level across groups (main effect of F0M = $F_{2,79} = 21.5; P < .001$), with patients again showing significantly less modulation across stimuli ($F_{2,79} = 4.0; P = .022$).

Likewise, sad and angry ratings were also related to changes in F0M and F0SD. Across an F0M of 125 Hz, there was a significant decrease in likelihood of sad ratings with increasing F0SD level across groups (main effect of F0SD = $F_{4,77} = 16.4; P < .001$, figure 1E), with patients again showing significantly less modulation across stimuli ($F_{4,77} = 4.8; P = .006$). A main effect was also seen across groups for sad responses across an F0SD level of 20 Hz ($F_{2,79} = 7.1; P = .001$, figure 1F) and for angry responses across a F0M level of 125 Hz ($F_{4,77} = 5.0; P = .001$, figure 1G). Again, patients showed significantly less modulation of response for sad ($F_{2,79} = 4.8; P = .01$) and angry responses ($F_{4,77} = 4.3; P = .004$), respectively.

Effect of High Frequency Energy (HF500)

In human speech, high levels of high frequency (>500 Hz) energy contribute to the percept of anger. A second set of stimuli were therefore intermixed with the base tones which only differed from the primary set by the addition of HF500 noise. Response patterns to each tone by group are shown in online supplementary table 3. Across groups, addition of HF500 to the FM tones significantly reduced the percept of sadness (figure 2, main effect of HF500: $F_{1,80} = 15.2; P < .001$) and increased the percept of anger (main effect of HF500: $F_{1,80} = 9.2; P = .003$). There was no group by response interaction on repeated-measures-ANOVA (figure 3, $F_{4,77} = 0.6; P = .65$), suggesting that perceptual patterns in both groups were similarly affected by addition of HF500 to the stimuli.

Comparison Across Tasks

FM Tone Composite Score. In order to permit correlational analyses across tasks, a composite score for emotion identification across FM tones was constructed for each subject by averaging scores on the 11 most identifiable stimuli (table 2). All stimuli included in the composite score had absolute rates of identification of >45% and significantly differed from the chance 20% response rate in both patients and controls independently (all $P < .001$). Large effect size between group differences were seen for this composite score (control = 68% ± 15%, patient = 42% ± 22%, $t_{60} = 6.1; P < .0001; d = 1.4$ SD) (figure 3A).

Comparison Tasks. Consistent with prior reports,16 patients also showed deficits in voice emotion recognition ($t_{77} = 5.7; P < .001$, figure 3B) and in simple tone matching ($t_{77} = 4.1; P < .001$, figure 3C). As predicted, there was a highly significant correlation between reduced FM tone composite score on the one hand and both voice emotion recognition ($r = .50; P < .001$) (figure 3D) and tone matching ($r = .42; P < .001$, figure 3E) on the other. Significant correlations with both voice emotion recognition ($P = .02$) and tone matching ($P = .055$) remained even following covariation for PSI. As we have reported previously,16 tone matching itself also correlated significantly with voice emotion recognition ($r = .43; P < .001$).

Finally, when both tone matching and composite score of the FM tone task were entered simultaneously into a multivariate regression vs voice emotion recognition,
a strong overall correlation was observed ($r = .56$; $P < .001$) with significant independent contributions of tone matching (partial $r = .29$; $P = .01$) and the FM tone composite score (partial $r = .40$; $P < .001$).

**Symptoms.** Within the schizophrenia group, PANSS negative symptom subscale scores significantly correlated with the FM tone composite score ($r = -.49$; $P < .001$). Thus, worse performance on the task correlated with higher levels of negative symptoms. No significant correlations were observed with other symptom domains. Finally, no correlations with any task were observed for medication dose.

**Discussion**

Individuals with schizophrenia show deficits in the ability to recognize emotion from tone of voice, leading to impairments in social cognition. Although studies showing correlations between emotion processing deficits and impaired sensory processing were first published in the 1970s, few follow-up studies were conducted until recently. The present study expands on our prior finding that deficits in detecting emotional prosody in spoken language are related to deficits in detecting acoustic features such as mean fundamental frequency (F0M) or pitch variability (F0SD). By using synthetic tones instead of human speech, we were able to evaluate the specific effects of such features independent of the more complex acoustic information present in speech.

Our findings are 3-fold: first, that simply varying the pitch-based parameters F0M and F0SD is sufficient to lead to reliable, differentiated patterns of emotion categorization in both normal volunteers and schizophrenia patients; second, that patients show reduced sensitivity to alterations in such acoustic features, and finally, that reduced sensitivity to these features correlates significantly with deficits in both simple tone matching and with more complex impairments in voice emotion recognition and negative symptoms. Taken together, our results suggest that deficits in low-level acoustic processing contribute significantly to impairments in voice emotion recognition in schizophrenia, which, in turn, is a key component of complex social interaction.

The present study builds upon our recent prior study in which we first evaluated emotion processing in schizophrenia using the same voice emotion recognition task used in the present study. Although deficits in voice emotion recognition are well established in schizophrenia, the battery used in our recent study is the first in which the acoustic features of the stimuli were available and thus the first in which the contributions of specific acoustic features could be evaluated. In this study, we confirm our earlier report of impaired voice emotion recognition.
recognition in schizophrenia using this battery, along with the correlation between tone matching deficits and impaired voice recognition deficits on this task. In addition, we extend these findings by demonstrating that patients have impaired emotion recognition even for FM tones containing only those features implicated in voice emotion recognition deficits in our prior study.

In our prior study, patients were particularly impaired in their ability to utilize pitch features, such as F0M and F0SD, while ability to utilize other features, such as HF500, was relatively intact. In the present study, FM tones were constructed with F0M and F0SD features similar to those found in well-recognized happy and sad stimuli in our prior study. Both patients and controls recognized stimuli with high F0M/F0SD as happy, and those with low F0M/F0SD as sad. However, as with speech, patients showed less differentiated responses to these FM tones, which suggest a reduced ability to utilize this acoustic information.

As in our recent study, we also observed 2 different ways in which anger may be portrayed and differential impairment in schizophrenia to the 2 patterns. First, for controls, the anger percept was reliably produced by tones with low F0M and moderate F0SD, approximating the features of “cold” anger. Patients did not show a similar tendency to identify these stimuli as angry, consistent with our prior observation that patients were also relatively insensitive to cold anger in speech. By contrast, adding HF500 to stimuli also increased the percept of anger in both groups. Patients showed as much change in response as controls, supporting the concept that ability to utilize voice quality and/or intensity cues is relatively intact, despite impaired pitch processing ability. A limitation of this analysis, however, is that we evaluated effects of only a single level of added HF500 energy. If a different level of HF500 energy had been used, it is possible that a different pattern of result might have been obtained.

This study has few direct parallels in the literature. Ross et al9 used stimuli that were passed through a 70–300 Hz band-pass filter that preserved prosodic information while markedly reducing phonetic information. Relative to controls, patients appeared to be worse at this than at a traditional emotional prosody task. Bach et al24 used a stimulus set consisting of meaningless sentences comprised of phonemes from several Indo-European languages2 and found no evidence for a relative deficit of a particular emotion. Neither study, however, evaluated the acoustic features of the stimuli, and moreover, acoustic features were not independently manipulated. The present study suggests that reduced ability to utilize specific acoustic features may underlie specific patterns of auditory emotion recognition impairment in schizophrenia.

A potential limitation is that the inclusion criteria for the specific stimuli used for the composite score of the FM tone task was defined post hoc and therefore needs to be replicated a priori in an independent sample. Moreover, for this initial battery, we varied only 3 (F0M, F0SD and HF500) of the many acoustic features that potentially contribute to emotion recognition. The features that we varied proved sufficient to evoke reliable percepts of happy, sad, and to a lesser extent, anger but not other emotional percepts such as fear or neutral expression. Future batteries incorporating more features,1,2 such as frequency contour, which may be important for fear, or intensity/intensity variability, which may be important for anger, may permit evaluation of a wider range of emotional categorizations. We also note that all patients were on antipsychotics and that only the PSI was used to measure cognition, so we could not assess relationship to other cognitive domains.

An additional strength of the present study is its development of a battery that is language free and so could potentially be used in transnational studies. At present, it remains relatively unknown how specific acoustic features of stimuli are used to determine emotions cross-culturally. The present battery would facilitate such research. To the extent that the acoustic features conveying emotion are based upon physical changes in the vocal apparatus that accompany the subjective emotion state (eg, throat constriction during sadness), patterns may be relatively stable cross-culturally. If so, use of FM tones may permit cross-cultural studies of emotion recognition in schizophrenia.

In the present study, we found strong correlations between ability to match simple tones following brief delay, a function that localizes to primary auditory cortex, and ability to differentiate emotional content of FM tones. In addition, both tone matching and emotional differentiation of FM tones correlated with ability to recognize intended emotion in human speech. This pattern of correlation can be interpreted most easily as indicating bottom-up contributions to impaired voice emotion recognition. It can, however, be argued that the deficit may lie not in sensory processing itself but in ability to assign meaning to specific tones. Future studies using neurophysiological or neuroimaging-based techniques may be needed, therefore, to confirm localization of these deficits to sensory brain regions. In addition, the present study provides further support for use of sensory-based cognitive remediation paradigms.

In conclusion, deficits in social cognition are increasingly being recognized as a major contributor to impaired functional outcome in schizophrenia. Impaired voice emotion recognition is a key component of the social cognition construct.2 The present study demonstrates that patients are as impaired in their ability to detect emotion in FM tones because they are in detecting emotion in speech, with specific interrelationship among measures. Much like simple pitch processing, FM modulation is also decoded at the level of auditory sensory cortex,26 supporting sensory-based theories of higher cortical impairments in schizophrenia.
Funding
National Institutes of Mental Health grants (R37 MH49334 to D.C.J., R01MH084848 to D.C.J./P.D.B.); New York University Conte Center for Schizophrenia Research (P50MH86385).

Supplementary Material
Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

Acknowledgments
We thank Christina Garidis for her help in this study. The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References