Individuals with schizophrenia show reliable deficits in the ability to recognize emotions from vocal expressions. Here, we examined emotion recognition ability in 23 schizophrenia patients relative to 17 healthy controls using a stimulus battery with well-characterized acoustic features. We further evaluated performance deficits relative to ancillary assessments of underlying pitch perception abilities. As predicted, patients showed reduced emotion recognition ability across a range of emotions, which correlated with impaired basic tone matching abilities. Emotion identification deficits were strongly related to pitch-based acoustic cues such as mean and variability of fundamental frequency. Whereas healthy subjects’ performance varied as a function of the relative presence or absence of these cues, with higher cue levels leading to enhanced performance, schizophrenia patients showed significantly less variation in performance as a function of cue level. In contrast to pitch-based cues, both groups showed equivalent variation in performance as a function of intensity-based cues. Finally, patients were less able than controls to differentiate between expressions with high and low emotion intensity, and this deficit was also correlated with impaired tone matching ability. Both emotion identification and intensity rating deficits were unrelated to valence of intended emotions. Deficits in both auditory emotion identification and more basic perceptual abilities correlated with impaired functional outcome. Overall, these findings support the concept that auditory emotion identification deficits in schizophrenia reflect, at least in part, a relative inability to process critical acoustic characteristics of prosodic stimuli and that such deficits contribute to poor global outcome.

Key words: schizophrenia/social cognition/prosody/emotion/speech/multidimensional scaling

Introduction

Individuals with schizophrenia show well-documented disturbances in ability to identify emotions based upon tone of voice (prosody). These deficits—along with parallel deficits in visual emotion identification—have been associated with poor functional outcome. Yet, the nature and etiopathology of these social communication deficits remain obscure. Some argue that such failures may reflect specific emotional disabilities attributable to dysfunction of limbic circuitry and/or global right hemisphere dysfunction. Still others, beginning with Chapman and Chapman, have advocated a more generalized cognitive and perceptual deficit. It may also be that social communicatory deficits stem from an inability to recognize the physical/acoustical cues that characterize emotional distinctions. This would be akin to not being able to distinguish a smile from a frown and therefore not being able to discern happiness from anger. Within emotional prosody, such distinctions are represented by the configuration of a variety of acoustical cues including pitch, voice intensity, voice quality, and temporal cues. Prior studies indicate that prosodic deficits in schizophrenia correlate consistently with deficits in basic tonal processing ability. This suggests that patient dysprosodia could stem in part from elemental pitch processing deficits and that patients may simply not “pick up” on pitch changes which signal affective intent. However, to date, this hypothesis has not been directly tested. In order to examine whether specific cue processing impairments may drive emotional prosodic deficits in schizophrenia, the present study utilized a stimulus battery consisting of vocal expressions that have been extensively characterized in normative populations and for which acoustic features have been well described. This is the first study to utilize an acoustically...
characterized prosodic dataset for use in schizophrenia research.

Previous research has shown that there are relatively distinct patterns of acoustic cues that differentiate between specific emotions, as summarized in table 1. For example, anger, happiness, and fear are typically characterized by high mean pitch and mean voice intensity, whereas sadness is characterized by low mean pitch and mean voice intensity. Also, anger and happiness expressions typically have large pitch variability, whereas fear and sadness expressions instead have small pitch variability. Regarding voice quality, anger and happiness expressions typically have a large proportion of high-frequency energy (HF500) in the spectrum, whereas sadness has less high-frequency energy (as the proportion of high-frequency energy increases, the voice sounds sharper and less soft). Finally, happiness and fear expressions typically have fast speech rate, whereas sadness expressions are characterized by a slow speech rate.

<table>
<thead>
<tr>
<th>Acoustic Cue (Abbreviation)</th>
<th>Description</th>
<th>Subjective Experience</th>
<th>Emotional Distinctions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch cues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0M</td>
<td>Mean fundamental frequency</td>
<td>Pitch</td>
<td>High: anger, fear, happiness; low: sadness</td>
</tr>
<tr>
<td>F0SD</td>
<td>Fundamental frequency variability (SD)</td>
<td>Pitch</td>
<td>Large: anger, happiness; small: fear, sadness</td>
</tr>
<tr>
<td>F0_contour</td>
<td>Shape of fundamental frequency contour</td>
<td>Pitch contour</td>
<td>Up: anger, fear, happiness; down: sadness</td>
</tr>
<tr>
<td>Intensity cues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceint_M</td>
<td>Mean voice intensity</td>
<td>Loudness of speech</td>
<td>High: anger, happiness; low: sadness</td>
</tr>
<tr>
<td>Voiceint_SD</td>
<td>Voice intensity variability (SD)</td>
<td>Loudness of speech</td>
<td>Large: anger, fear, happiness; small: sadness</td>
</tr>
<tr>
<td>Attack</td>
<td>Rate of rise of amplitude</td>
<td>Rapidity of voice onsets</td>
<td>Fast: anger, happiness; slow: sadness</td>
</tr>
<tr>
<td>Voice quality cues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF500</td>
<td>Relative proportion of acoustic energy above, vs below, 500 Hz</td>
<td>Voice quality</td>
<td>Much: anger, happiness; little: sadness</td>
</tr>
<tr>
<td>F1BW</td>
<td>Bandwidth of the first formant</td>
<td>Voice quality</td>
<td>Narrow: anger; wide: sadness</td>
</tr>
<tr>
<td>Temporal cues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech rate</td>
<td>Velocity of speech</td>
<td>Velocity of speech</td>
<td>Fast: anger, fear, happiness; slow: sadness</td>
</tr>
<tr>
<td>Pause proportion</td>
<td>Amount of pauses in speech</td>
<td>Velocity of speech</td>
<td>Much: sadness; little: anger</td>
</tr>
</tbody>
</table>

Table 1. Simplified Summary of Selective Voice Cue Patterns Obtained for Specific Emotions in Previous Research (Adapted From Emotion 2001; pp.790–5 [extracted from tables 6,7] and Psychol Bull 2003;12916,17)

lead not only to an overall reduction in accuracy of identification but also potentially to a characteristic pattern of misidentification across emotions. To date, no study of auditory emotion processing in schizophrenia has analyzed such misidentification patterns. Based upon prior demonstrations of impaired pitch processing in schizophrenia, we hypothesized that patients with schizophrenia would show deficits in detection of emotional stimuli that relied primarily on pitch-based discriminations. We further hypothesized that such discrimination deficits may underlie differences in the misidentification patterns observed between healthy controls and schizophrenia patients.

Finally, while many studies have assessed the ability of patients to recognize emotions from vocal expressions, as well as identification accuracy for strongly posed vs weakly posed expressions, no studies to date have examined subjects’ ratings of the emotional intensity of such expressions. Emotional intensity assessment, like emotion identification itself, requires ability to detect and interpret alterations in the basic perceptual components of speech. Further, assessment of the intensity of another person’s emotional feelings may be as critical for social communication as assessment of his/her specific emotional state. In the battery utilized for this study, actors were asked to portray the same emotion both weakly and strongly, and subjects’ ability to both recognize emotions and to judge the emotional intensity of the expressions was assessed. Our study therefore uniquely assessed the ability of patients with schizophrenia to differentiate among intended strong and weak emotion portrayals relative to controls. As with pitch, we hypothesized that pitch-based contributions to emotion intensity detection would be impaired.
Methods

Participants

Twenty-three chronically ill (illness duration = 18.2 ± 9.2 years) patients (4 females, age = 39.4 ± 10.9 years) meeting Diagnostic and Statistical Manual of Mental Disorders (DSM), Fourth Edition, criteria for either schizophrenia (N = 20) or schizoaffective disorder (N = 3) as determined by Structured Clinical Interview for DSM diagnosis and 16 healthy volunteers (3 female, age = 34.3 ± 11.5 years) with no history of mental illness participated in the study. Patients were drawn from inpatient and outpatient units associated with the Nathan S. Kline Institute, and they were all receiving typical or atypical antipsychotics at the time of testing (mean dose = 1265 ± 830 CPZ equivalents). Volunteers were staff or responded to local advertisement. Groups did not differ on mean parental socioeconomic status (SES) (schizophrenic [sz]: 54.2 ± 28.4, control [ctl]: 49.4 ± 18.7, t = 0.67, P = .6), although, as expected, patients had lower individual SES (sz: 25.8 ± 10.6, ctl: 47.3 ± 14.6, t = 5.24, P < .001) and verbal IQ (sz: 95.9 ± 28.4, ctl: 109.5 ± 11.2, t = 3.99, P < .001) as measured by the Ammons and Ammons Quick Test than did controls.

Clinical assessments included ratings on the Positive and Negative Syndrome Scale (PANSS) and the Independent Living Scale-Problem Solving Scale (ILS-PB). which predicts ability of individuals with schizophrenia to live independently. Mean PANSS scores were 11.1 ± 3.6, 12.5 ± 5.6, and 10.9 ± 2.9 for positive, negative, and cognitive factors, respectively. Mean scale ILS-PB score was 34.8 ± 12.1.

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. All subjects save 2 (one schizophrenia) were right handed.

Materials

Recognition of emotional prosody was assessed using stimuli from the prosody task of Juslin and Laukka. The stimuli consisted of audio recordings of 2 male and 2 female actors portraying 5 emotions—anger, disgust, fear, happiness, and sadness—with 2 levels of emotion intensity (weak and strong) as well as utterances with no emotional expressions (“no expression”). The sentences spoken were semantically neutral and consisted of both statements and questions (ie, “It is eleven o’clock”, “Is it eleven o’clock?”). These parameters yielded 88 stimuli [4 speakers × 2 intensity levels × 5 emotions (plus 8 no-expression stimuli) × 2 forms (questions or statements)]. All speakers were native British English speakers. For this stimulus set, measurement of acoustic characteristics was conducted in PRAAT speech analysis software as described previously (see table 1 for a description of the acoustic features included in the present study). All vocal stimuli were presented in random order in the recognition experiments, and the subjects were asked first to identify the emotional expression of each utterance by choosing 1 of 6 alternatives (anger, disgust, fear, happiness, sadness, or no expression) and then to rate the utterance’s emotional intensity on a scale from 1 (very low intensity) to 10 (very high intensity).

Ancillary assessment of pitch processing was measured with a version of the Distorted Tunes Task (DTT) and a simple tone matching task. The DTT consists of 26 familiar tunes adapted for the United States ranging in length from 12 to 26 notes. Seventeen of the tunes are rendered melodically incorrect by changing the pitch of 2–9 notes within the tune. Subjects respond “yes” or “no” as to whether the melody is correct. Subject scores are calculated based on the percentage of correctly categorized melodies. The tone matching task consists of pairs of 100-ms tones in series, with 500-ms intertone interval. Within each pair, tones are either identical or differed in frequency by specified amounts in each block (2.5%, 5%, 10%, 20%, or 50%). In each block, half the tones are identical and half are dissimilar. Subjects respond by pressing 1 of 2 keys to answer whether the pitch was the same or different. Tones are derived from 3 reference base frequencies (500, 1000, and 2000 Hz) to avoid learning effects. In all, the test consisted of 5 sets of 26 pairs of tones.

All auditory tasks were presented on a CD player at a sound level that was comfortable for each listener in a sound-attenuated room.

Data Analysis

Between-groups effects for both identification and intensity ratings were assessed using multivariate analysis of variance (MANOVA) using a within-subject factor of emotion (6 levels) and a between-subject factor of diagnostic group.

A second stage of analysis examined contributions of specific acoustic properties to between-group performance. For this analysis, cues were selected if they differentiated a particular emotion from the mean of all others. For example, as shown in table 2, mean pitch variability (F0SD) was significantly higher for intended happy and anger stimuli and significantly lower for intended fear stimuli than for the remaining emotions. Similarly, mean pitch of the stimuli (F0M) was significantly higher for happy stimuli and lower for disgust and neutral (no expression) stimuli than for other emotions. Cues chosen for this analysis are shown in table 2, and values that are differentiated for each cue are bolded within the table. Other cues that have been related to emotion perception, such as the temporal cues of speech rate and pause proportion (table 1) nevertheless did not differentiate significantly between intended emotions and so were not included in this analysis.
In order to examine the contributions of these cues on performance between groups, stimuli were then divided into low, medium, and high levels based upon parameter strength relative to mean scores for each emotion. In general, divisions were made according to the following rule: for each feature, the mean value was calculated for the emotion of interest, as well as for all emotions with emotion of interest excluded. These mean values were used as cut points with stimuli falling either outside of these mean values (high or low, respectively, depending upon emotion) or between, reflecting medium values. For features relevant to multiple emotions, additional cut points gave rise to division into additional levels designated med1 and med2, although for each emotion parameters varied across only 3 levels. For example, for F0SD, 4 levels were used (low, med1, med2, high). However, because happy stimuli are associated with increased F0SD vs others, all happy stimuli were contained entirely within the levels of med1 and med2 (figure 2). Mean values for each feature and emotion are provided in table 2. Cutoff values are provided in the legends of figures 2 and 3.

For selected emotion/cue parings, between-group analyses were conducted using MANOVA with a within-subject factor for cue level and a between-subject factor of diagnostic group. Potential valence effects were assessed by between-group MANOVA as well, using positive (ie, happy) vs negative (ie, anger, disgust, fear, and sadness) valence as the between-group factor.

Misattribution patterns for emotion identification were examined using the ALLSCAL scaling algorithm (SPSS 15.0) to scale and calculate Euclidean distances between the dependent variables. R² and stress parameters were used to assess model fit. Features contributing to the misattribution scaling values were assessed using Pearson correlations across emotions. The relationship between emotion identification and pitch perception (tone matching and DTT) was assessed using linear regression across subjects.

All statistical tests were 2-tailed, with \( \alpha < .05 \), and computed using either JMP (SAS Institute Inc, Cary, NC, 2005) or SPSS 15.0 (Chicago, IL, 2006).

Results

Emotion Identification

Between-Groups Analysis. As in prior studies of acoustic emotion identification, patients were substantially less accurate than controls in identifying intended emotion (ctl: 44.8 ± 7.0; scz: 34.2 ± 10.0; \( F_{1,37} = 24.7, P < .0001, d = 1.2 \)). Performance in both groups was well above chance (16.6%). Both groups showed significant variation in identification accuracy across intended emotions (\( F_{5,33} = 13.4, P < .001 \)) with greatest accuracy for

### Table 2. Summary of Cue Values (Mean and SD) as a Function of Emotion for Acoustic Cues That Differentiated Between Emotions

<table>
<thead>
<tr>
<th>Cue</th>
<th>Happy (n = 16)</th>
<th>Sad (n = 16)</th>
<th>Anger (n = 16)</th>
<th>Fear (n = 16)</th>
<th>Disgust (n = 16)</th>
<th>No Expression (n = 8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch cues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0SD (Hz, log values)</td>
<td>1.78 (0.27)</td>
<td>1.45 (0.17)</td>
<td>1.71 (0.18)</td>
<td>1.38 (0.29)</td>
<td>1.53 (0.20)</td>
<td>1.29 (0.22)</td>
</tr>
<tr>
<td>F0M (Hz, log values)</td>
<td>2.42 (0.13)</td>
<td>2.33 (0.10)</td>
<td>2.37 (0.13)</td>
<td>2.33 (0.15)</td>
<td>2.26 (0.11)</td>
<td>2.19 (0.11)</td>
</tr>
<tr>
<td>F0contour</td>
<td>1.14 (0.56)</td>
<td><strong>0.59 (0.37)</strong></td>
<td>1.47 (1.10)</td>
<td><strong>1.62 (1.07)</strong></td>
<td>0.80 (0.62)</td>
<td>1.05 (0.74)</td>
</tr>
<tr>
<td>Intensity cues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VoiceintM (dB)</td>
<td>−32.53 (5.85)</td>
<td>−36.75 (5.21)</td>
<td>−29.79 (5.83)</td>
<td>−37.54 (5.51)</td>
<td>−34.63 (5.46)</td>
<td>−38.12 (2.40)</td>
</tr>
<tr>
<td>VoiceintSD (dB)</td>
<td>7.92 (1.86)</td>
<td>7.77 (2.25)</td>
<td><strong>9.00 (1.82)</strong></td>
<td>7.82 (3.04)</td>
<td>7.97 (1.89)</td>
<td>6.99 (0.76)</td>
</tr>
<tr>
<td>Voice quality cues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF500</td>
<td><strong>0.70 (0.08)</strong></td>
<td>0.61 (0.07)</td>
<td><strong>0.71 (0.10)</strong></td>
<td>0.62 (0.08)</td>
<td>0.63 (0.10)</td>
<td><strong>0.57 (0.06)</strong></td>
</tr>
</tbody>
</table>

Notes: Bolded values reflect properties that are different for indicated emotion vs mean of all remaining emotions. Cues with at least one such significant difference were used in a second stage of analysis that examined contributions of specific acoustic properties to between-group performance.

---

D. I. Leitman et al.
sadness and neutral and least for disgust (figure 1), suggesting that some emotions may be easier to portray and/or identify than others. The group $\times$ emotion interaction was not significant ($F_{5,33} = 0.6, P > .7$).

**Emotion Identification as a Function of Acoustic Cues Within Tokens.** In order to evaluate the degree to which deficits in utilization of specific cues might contribute to impaired ability to identify intended emotion, cue types were identified that differentiated most strongly between emotions, as shown in table 2. Four pitch-based measures ($F_{0SD}, F_{0M}, F_{0contour}, F_{0floor}$), 2 intensity-based measures ($voiceint_{M}, voiceint_{SD}$), and 1 voice quality measure ($HF_{500}$) were identified as having mean cue levels that significantly differentiated at least one emotion from all others. For $F_{0SD}$ and $F_{0M}$, log values were used to normalize values. Each measure contributed to one or more emotion identification types (table 2). For each measure, stimuli were grouped into levels depending upon mean values for specific emotions of interest.

**Pitch Cues**

Emotion identification performance as a function of pitch-based measures is shown in figure 2. Pitch variability ($F_{0SD}$) differed significantly across intended emotions, with intended happy stimuli showing on average greater levels of $F_{0SD}$ than those intended to portray other emotions and intended fear stimuli showing, on average, lower levels (table 2). $F_{0SD}$ was thus analyzed relative to accuracy in decoding happiness and fear. Both patients ($F_{2,21} = 9.4, P < .001$) and controls ($F_{2,14} = 82.2, P < .001$) showed greater accuracy in correctly identifying intended emotion for stimuli characterized by high pitch variability (high $F_{0SD}$) vs those who expressed lower degrees. However, patients showed less utilization of this cue than controls, leading to a significant group $\times$ level interaction ($F_{2,36} = 4.7, P < .05$), and highly significant ($P < .001$) between-group performance differences for correct identification of intended happy stimuli incorporating high, but not moderate or low, levels of $F_{0SD}$ (figure 2A).

Similarly, patients showed less variation in their fear identification performance as a function of pitch-based measures as a function of $F_{0SD}$ than controls, leading also to a significant group $\times$ level interaction ($F_{2,36} = 3.3, P < .05$) (figure 2B). Further, in this case, significant variation in performance as a function of $F_{0SD}$ was seen only for controls ($F_{2,14} = 4.7, P < .05$), whereas no significant variation was seen within the patient group ($F_{2,21} = 0.07, P > .9$). As a result,
between-group differences in accuracy of identifying fear were observed only at the lowest (most fearful) levels of F0SD.

F0SD also differentiated anger from other emotions as a group, with levels for intended anger stimuli being similar to those for happy stimuli and higher than those for other emotions. However, neither group showed significant variation in accuracy of response for anger based upon level of F0SD, suggesting that both groups used other anger-associated cues (e.g., intensity cues) as primary cues for anger determination.

The mean pitch of each stimulus (F0M) differentiated stimuli intended to portray either disgust or happiness from those portraying other emotions, with disgust being associated with low levels of F0M and happiness with high levels (table 2). For controls, accuracy of performance varied significantly by F0M cue level in identifying disgust (F2,14 = 4.2, P < .05), whereas for patients performance did not vary significantly (F2,21 = 0.19, P > .8) (figure 2B). Despite association of high F0M levels with happiness, neither group showed significant variation in accuracy across levels of F0M.

The slope or trajectory of the pitch contour across the duration of the stimulus (F0contour) differentiated sadness and fear from other emotions, with intended sadness being associated with F0contour levels below 1 (i.e., downward sloping pitch trajectory) and intended fear associated with values above 1 (i.e., upward sloping trajectory) (table 2). For controls, accuracy of performance varied significantly as a function of F0contour for both sadness (F2,14 = 4.5, P < .05) and fear (F2,14 = 5.2, P < .05), whereas for patients it did not (sadness: F2,21 = 0.57, P > .5; fear: F2,21 = 0.44, P > .6) (figure 2C). The minimum pitch levels (F0low) associated with each stimulus also differentiated intended disgust from all other emotions (table 2) and thus served as a potential cue. However, neither group showed significant variation in accuracy of performance as a function of this measure.

**Intensity Cues**

In contrast to pitch-based measures, which were particularly relevant to perception of happiness, sadness, and disgust, intensity-based measures significantly differentiated intended anger stimuli from all others. Two measures were particularly important: voiceintSD, which reflects the degree of variability of intensity over the course of a stimulus, and voiceintM, which reflects overall stimulus energy (table 2). Both patients (F2,21 = 17.78, P < .001) and controls (F2,14 = 21.6, P < .05) showed increased accuracy in identifying intended emotion as voiceintM increased, with no significant difference between groups in overall accuracy (F2,36 = 2.32, P = .14) and no group × feature level interaction (F2,36 = 0.22, P = .81) and (figure 3A). Similarly, both patients (F2,21 = 20.6, P < .0001) and controls (F2,14 = 12.1, P = .001) showed increased accuracy in identifying intended emotion as voiceintM increased (figure 3B). In this case, patients showed greater variation in performance across levels of voiceintM than did controls, leading to a significant group × feature interaction (F2,36 = 4.11, P = .025). Further, patients showed decreased accuracy compared with controls only at the lowest (least angry) levels of voiceintM, suggesting that patients utilized low voiceintM levels to exclude anger as an intended emotion to a greater extent than did controls. Anger exemplars with low voiceintM levels (all of which were intended as weak portrayals of anger) nevertheless had F0SD values (1.63 ± 0.19) significantly above the expected range for neutral stimuli (P < .03), suggesting that controls may have been able to utilize F0SD as a secondary cue for recognition of intended anger emotion for these stimuli whereas patients were not.

**Voice Quality Cues**

Of potential voice quality cues, only HF500—reflecting vocal timbre—distinguished between intended emotions, with angry stimuli showing greater mean levels of HF500 than others (table 2). As with intensity cues, patients and controls showed similar variation in performance across level of HF500. Thus, both patients and controls showed a similar tendency to identify intended angry stimuli correctly as levels of HF500 increased, leading to significant group × level interactions for both patients (F2,21 = 29.01, P < .001) and controls (F2,14 = 7.52, P < .05) and no group × level interaction (F2,36 = 0.69, P > .5) (figure 3C).

**Misidentification Analysis.** In order to further identify the basis for impaired emotion identification in...
schizophrenia, the pattern of misidentifications were also analyzed. In this analysis, both groups showed a nonhomogeneous pattern of misattributions, with some distinctions (eg, sad/fear) yielding higher rates of misidentification of intended emotion than others (eg, sad/disgust). Overall, patients showed a similar pattern of errors to controls (table 4). Nevertheless, patients were more likely than controls to misidentify happy stimuli as being neutral (P = .021), fearful stimuli as expressing either anger (P = .007) or disgust (P = .03), disgust stimuli as expressing sadness (P = .006), and no expression stimuli as expressing fear (P = .001). Some, but not all, of these distinctions were also difficult for controls, suggesting that the basis for emotional distinctions may differ between patients and controls.

Multidimensional scaling (MDS) was used to map apparent distances between emotions based upon the overall misattribution confusion matrix. MDS analysis has previously been used to map subjective affective representations in schizophrenia but has not been previously applied to analysis of physical determinants of prosodic identification.

Two-dimensional models accounted well for observed attribution pattern with good parametric properties for both schizophrenia patients (stress = 0.11, \( R^2 = 0.93 \)) and controls (stress = 0.09, \( R^2 = 0.95 \)) (figure 4A). For both controls and patients, scaling values along the first dimension of the MDS correlated most strongly with the voice quality cue F1_bw, with sadness and anger occupying opposite ends of the scale for controls and sadness and no emotion occupying opposite ends of the scale for patients (figure 4B). The strength of correlation, however, was higher for controls (r = 0.96, P = .003) than for patients (r = 0.72, P = .11). For controls, spacing along the second dimension correlated more strongly with the pitch cues F0_mean (r = 0.91, P = .03) and F0_SD (r = 0.79, P = .06) than with the voice intensity cue voiceint_SD (r = 0.68, P = .17) (figure 4C). In contrast, patients showed significant correlation with spacing along the second dimension and voiceint_SD (r = 0.88, P = .02) but no significant correlation with either F0Mean (r = 0.37, P = .48) or F0_SD (r = 0.48, P = .34), suggesting different usage of pitch vs intensity cues in differentiating among emotional percepts. For controls, happiness and no emotion occupied opposite ends of the second MDS dimension, consistent with primary use of pitch variables. In contrast, for patients, anger and no emotion occupied opposite ends of this dimension, consistent with primary differentiation of emotions by intensity.

Correlational Analyses. A final approach for assessment of the relationship between emotion processing deficits and basic perceptual abilities utilized linear regression analyses between the 2 sets of measures. Basic perceptual abilities were assessed using both tone matching, across a range of levels, and DTT score (table 3). A linear regression of emotion identification accuracy vs perceptual performance showed a significant overall relationship (adjusted \( R^2 = 0.51, P = .03 \)) with significant correlations between performance and tone matching accuracy at 2.5% (P = .02), 5% (P = .02), and 50% (P = .005) and DTT score (P = .012). Including generalized cognitive measures, such as digit symbol, in the regression did not significantly improve overall fit (adjusted \( R^2 = 0.44, P = .12 \)) vs fits obtained with perceptual measures alone. No similar relationship was observed for controls (adjusted \( R^2 = 0.26, P = .45 \)).

Finally, performance on emotion identification across emotions correlated significantly with the cognitive subscale of PANSS (r = −0.49, P = .021) and with ILS-PB score (r = 0.59, P < .01) but not with either positive (r = −0.3, P = .18) or negative (r = −0.05, P = .8) factor scores or medication dosage (CPZ) (r = 0.04, P = .83).

### Emotion-Intensity Rating

**Between-Groups Analysis.** In addition to portraying specific emotions, actors for this stimulus set had been asked to provide both weak and strong exemplars of each emotion. Sensitivity of subjects to the intended intensity of emotions, irrespective of identity, was assessed using MANOVA, with a within-subject factor of intended strength of portrayal (weak/strong) as well as

---

**Table 3. Comparison of Basic Perceptual Abilities in Patients and Controls**

<table>
<thead>
<tr>
<th>Basic Perceptual Ability Measure</th>
<th>Patients Mean Score, Mean % (SD)</th>
<th>Controls Mean Score, Mean % (SD)</th>
<th>t Value (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTT</td>
<td>71.91 (20.40)</td>
<td>89.18 (16.63)</td>
<td>-2.8 (.008)</td>
</tr>
<tr>
<td>TMT_2.5%</td>
<td>54.06 (17.72)</td>
<td>76.54 (21.33)</td>
<td>-3.0 (.006)</td>
</tr>
<tr>
<td>TMT_5%</td>
<td>65.17 (19.63)</td>
<td>88.46 (19.86)</td>
<td>-3.0 (.006)</td>
</tr>
<tr>
<td>TMT_10%</td>
<td>67.74 (20.91)</td>
<td>88.85 (17.53)</td>
<td>-2.7 (.012)</td>
</tr>
<tr>
<td>TMT_20%</td>
<td>77.99 (16.99)</td>
<td>91.15 (17.95)</td>
<td>-1.9 (.065)</td>
</tr>
<tr>
<td>TMT_50%</td>
<td>86.54 (13.21)</td>
<td>97.31 (4.81)</td>
<td>-3.1 (.005)</td>
</tr>
<tr>
<td>TMT_total</td>
<td>70.30 (14.07)</td>
<td>88.46 (14.51)</td>
<td>-3.2 (.003)</td>
</tr>
</tbody>
</table>

*Note: DTT = Distorted Tunes Task, TMT = tone matching task, 2.5%–50% = frequency difference levels (%Δf), for TMT significant t values are shown in bold.*
emotion and cohort. As expected, there was a significant main effect of intended strength of portrayal on rated intensity ($F_{1,37} = 82.2, P < .0001$), suggesting that across groups, subjects were able to successfully differentiate weak and strong portrayals (figure 5). Nevertheless, patients differentiated weak and strong portrayals to a significantly lesser degree than controls, as reflected in a highly significant group x intensity interaction ($F_{1,37} = 7.6, P < .009$). Patients also showed a significant tendency to rate all portrayals as being more intense than controls ($F_{1,37} = 6.2, P < .02$), with the differences being more profound for intended weak ($F_{1,37} = 8.58, P = .006$) than for intended strong ($F_{1,37} = 2.88, P = .1$) portrayals. As expected, there was no significant main effect of emotion ($F_{4,33} = 1.8, P = .15$) or group x emotion interaction ($F_{5,33} = 1.1, P > .37$). As with emotion identification, alterations in emotion intensity rating strength ($r = 0.51, P = .007$) and difference in perceived intensity for intended weak vs intended strong portrayals ($r = 0.59, P = .003$) correlated significantly with deficits in tone matching performance, such that individuals with the worst tone matching performance tended to rate intensity most strongly regardless of intended intensity and observed least difference in intensity between intended emotions.

### Discussion

During normal conversation, emotion is conveyed largely by modulation of tone and rhythm of voice. Decoding of such information in real time, therefore, is critical for normal social interaction. Rules for decoding of emotion are either learned implicitly $^{28}$ or may reflect an innate ability $^{17}$ because individuals are never taught exact rules for emotion identification. $^{28}$ As shown in this and previous studies, $^{16,29}$ identification of intended emotions by voice alone is an inexact science, with most individuals correctly identifying intended emotion only 50%–60% of the time. Given the complexity of the task and its dependence upon basic perceptual abilities, it is axiomatic that deficits in basic sensory abilities, such as the ability to detect tonal patterns over time, in schizophrenia would produce deficits in acoustic emotion identification ability. Nevertheless, the sensitivity of patients with schizophrenia to the specific features of speech that are used to convey emotional intent has not previously been evaluated.

In the present study, performance of patients was compared with that of controls as a function of a range of cues, including those involving pitch ($F_{0\text{M}}, F_{0\text{SD}}, F_{0\text{contour}}$), intensity (voiceintSD, voiceintM), and voice quality ($HF_{\text{500}}$). Analyses took advantage of the fact that some stimuli within the battery were relatively good exemplars of the intended emotion, as reflected in higher levels of correct identification, whereas others were less easily identified. Thus, for each emotion, it was possible to analyze responses across a range of acoustic values. As expected, $^{16}$ pitch cues were critical for detecting happiness, sadness, fear, and disgust in controls, with different pitch measures contributing differentially. In contrast, intensity and voice quality cues were critical for detection of anger. These results are consistent with prior findings $^{12,16}$ showing similar cue emotion interactions.

This study demonstrates for the first time that, when compared with controls, patients’ performance did not vary to the same degree as a function of the relative presence or absence of pitch cues but that such variability was

### Table 4. Misidentification Patterns for Patients and Controls.

<table>
<thead>
<tr>
<th>Intended Emotion</th>
<th>Rated Emotion</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Happy</td>
<td>Sad</td>
<td>Anger</td>
<td>Fear</td>
<td>Disgust</td>
<td>No Emotion</td>
</tr>
<tr>
<td>Happy</td>
<td>7.88 (8.04)</td>
<td>9.51 (7.27)</td>
<td>18.21 (11.13)</td>
<td>13.04 (10.13)</td>
<td>23.37 (23.17)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.64 (4.25)</td>
<td>10.94 (5.35)</td>
<td>22.27 (7.56)</td>
<td>9.77 (11.85)</td>
<td>10.55 (8.45)</td>
<td></td>
</tr>
<tr>
<td>Sad</td>
<td>11.88 (19.89)</td>
<td>3.19 (4.87)</td>
<td>15.94 (11.14)</td>
<td>6.38 (8.64)</td>
<td>21.74 (22.67)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.33 (5.96)</td>
<td>1.25 (2.69)</td>
<td>22.50 (11.89)</td>
<td>2.50 (4.13)</td>
<td>12.92 (11.01)</td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>8.97 (9.00)</td>
<td>4.35 (5.14)</td>
<td>10.87 (9.65)</td>
<td>21.20 (11.11)</td>
<td>14.67 (15.03)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.72 (9.92)</td>
<td>2.73 (3.20)</td>
<td>10.94 (7.03)</td>
<td>19.92 (11.91)</td>
<td>7.42 (7.30)</td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>10.56 (15.51)</td>
<td>28.57 (17.76)</td>
<td>8.39 (7.96)</td>
<td>8.07 (9.94)</td>
<td>16.15 (21.99)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.14 (9.76)</td>
<td>37.50 (15.54)</td>
<td>2.68 (4.42)</td>
<td>2.68 (4.42)</td>
<td>9.38 (9.30)</td>
<td></td>
</tr>
<tr>
<td>Disgust</td>
<td>8.97 (12.62)</td>
<td>7.34 (8.77)</td>
<td>29.89 (13.97)</td>
<td>5.71 (9.02)</td>
<td>25.82 (20.84)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.69 (8.39)</td>
<td>1.56 (2.80)</td>
<td>31.64 (14.15)</td>
<td>3.13 (3.95)</td>
<td>17.19 (12.40)</td>
<td></td>
</tr>
<tr>
<td>No emotion</td>
<td>18.48 (21.93)</td>
<td>5.98 (11.22)</td>
<td>6.52 (10.57)</td>
<td>9.78 (11.28)</td>
<td>2.72 (8.39)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.81 (14.34)</td>
<td>3.13 (7.22)</td>
<td>6.25 (7.91)</td>
<td>0.78 (3.13)</td>
<td>4.69 (7.74)</td>
<td></td>
</tr>
</tbody>
</table>
This finding suggests that patients were less able than their healthy counterparts to utilize pitch-based cues to identify emotion, whereas their ability to utilize intensity cues was relatively intact. For controls, stimuli with highest levels of pitch variability were 4-fold more identifiable than those with more moderate levels. For patients, the difference was only 2-fold, leading to a large effect size ($d = 1.34$) deficit in identifying happiness only for those stimuli with highest (ie, most happy) F0 SD levels. Similarly, low levels of F0SD serve as a primary cue for fear. For controls, stimuli with low F0SD levels are 5-fold more identifiable than those with more moderate levels. For patients, the difference was only 2-fold, leading to a large effect size ($d = 1.34$) deficit in identifying happiness only for those stimuli with highest (ie, most happy) F0 SD levels.

In contrast to their inability to utilize pitch cues, patients did appear to utilize intensity cues, such as voiceintSD and voiceint M,, to detect anger equivalently to controls. In this case, patients showed increased, rather than decreased, variation in accuracy of detection of anger responses as compared with controls, leading patients to incorrectly reject anger as intended emotion for those exemplars with lowest levels of voiceintM. Further, for patients, MDS distances correlated significantly with intensity measures voiceintSD and ATTACK, whereas for controls, correlations were nonsignificant (voiceintSD: $r^2 = 0.46$, ATTACK: $r^2 = 0.05$). In the MDS analysis, distances between emotions were significantly influenced by pitch-based measures for controls but not for patients (figure 4C), suggesting that the differential pattern of misidentification seen in patients vs controls also relates to relative inability to utilize pitch cues in discriminating emotions.

In contrast to their inability to utilize pitch cues, patients did appear to utilize intensity cues, such as voiceintSD and voiceint M,, to detect anger equivalently to controls. In this case, patients showed increased, rather than decreased, variation in accuracy of detection of anger responses as compared with controls, leading patients to incorrectly reject anger as intended emotion for those exemplars with lowest levels of voiceintM. Further, for patients, MDS distances correlated significantly with intensity measures voiceintSD and ATTACK, whereas for controls, correlations were nonsignificant (voiceintSD: $r^2 = 0.46$, ATTACK: $r^2 = 0.05$). In the MDS analysis, distances between emotions were significantly influenced by pitch-based measures for controls but not for patients (figure 4C), suggesting that the differential pattern of misidentification seen in patients vs controls also relates to relative inability to utilize pitch cues in discriminating emotions.

In contrast to their inability to utilize pitch cues, patients did appear to utilize intensity cues, such as voiceintSD and voiceint M,, to detect anger equivalently to controls. In this case, patients showed increased, rather than decreased, variation in accuracy of detection of anger responses as compared with controls, leading patients to incorrectly reject anger as intended emotion for those exemplars with lowest levels of voiceintM. Further, for patients, MDS distances correlated significantly with intensity measures voiceintSD and ATTACK, whereas for controls, correlations were nonsignificant (voiceintSD: $r^2 = 0.46$, ATTACK: $r^2 = 0.05$). In the MDS analysis, distances between emotions were significantly influenced by pitch-based measures for controls but not for patients (figure 4C), suggesting that the differential pattern of misidentification seen in patients vs controls also relates to relative inability to utilize pitch cues in discriminating emotions.

In contrast to their inability to utilize pitch cues, patients did appear to utilize intensity cues, such as voiceintSD and voiceint M,, to detect anger equivalently to controls. In this case, patients showed increased, rather than decreased, variation in accuracy of detection of anger responses as compared with controls, leading patients to incorrectly reject anger as intended emotion for those exemplars with lowest levels of voiceintM. Further, for patients, MDS distances correlated significantly with intensity measures voiceintSD and ATTACK, whereas for controls, correlations were nonsignificant (voiceintSD: $r^2 = 0.46$, ATTACK: $r^2 = 0.05$). In the MDS analysis, distances between emotions were significantly influenced by pitch-based measures for controls but not for patients (figure 4C), suggesting that the differential pattern of misidentification seen in patients vs controls also relates to relative inability to utilize pitch cues in discriminating emotions.

In contrast to their inability to utilize pitch cues, patients did appear to utilize intensity cues, such as voiceintSD and voiceint M,, to detect anger equivalently to controls. In this case, patients showed increased, rather than decreased, variation in accuracy of detection of anger responses as compared with controls, leading patients to incorrectly reject anger as intended emotion for those exemplars with lowest levels of voiceintM. Further, for patients, MDS distances correlated significantly with intensity measures voiceintSD and ATTACK, whereas for controls, correlations were nonsignificant (voiceintSD: $r^2 = 0.46$, ATTACK: $r^2 = 0.05$). In the MDS analysis, distances between emotions were significantly influenced by pitch-based measures for controls but not for patients (figure 4C), suggesting that the differential pattern of misidentification seen in patients vs controls also relates to relative inability to utilize pitch cues in discriminating emotions.
Finally, patients showed some ability to modulate response based upon voice quality cues such as HF500 and F1BW, although the exact pattern of use differed somewhat between groups. In the MDS analysis, both controls and schizophrenia patients appeared to use F1BW as a principal cue in differentiating emotions. F1BW, although not strongly predictive of any single emotion, nevertheless is thought to convey mood/attitude information that is superordinate to emotion, such as the degree to which an individual is relaxed vs stressed.30 For controls, the relaxed/stressed distinction is the voice feature that is most readily perceived.30 Although patients were able to utilize F1BW to differentiate emotions associated with high power (eg, anger, disgust) relative to those associated with low power (eg, sadness), nevertheless, they showed reduced spacing of emotions along this dimension than did controls.

As in prior studies of emotional identification, deficits in performance correlated significantly both with more basic deficits in pitch perception, such as tone matching and DTT score, as well as with deficits in global outcome, as reflected by the ILS-PB. Correlations with tone matching and DTT underscore the importance of treatment strategies aimed at reversing social communicatory disturbance, as well as the importance of correcting underlying deficits in pitch processing. In contrast, correlations with ILS-PB, as noted previously,2,31 underscore the relationship between poor acoustic emotion identification ability and poor functional outcome. Personality researchers have suggested that individuals with higher degrees of empathic ability and greater degrees of “social connectedness” perceive vocal emotional cues better than individuals who have lower degrees of these traits.32 The present findings suggest that the inverse may also be true and that one’s basic perceptual abilities may determine in large part one’s social experience.

At present, relatively little is known about the development of pitch perceptual abilities over the course of schizophrenia. Pitch detection deficits have been demonstrated consistently in chronic patients in schizophrenia, along with impaired generation of early auditory event-related potentials such as mismatch negativity (MMN), which reflects preattentive detection of stimulus deviation. Pitch processing was also studied in one sample of 15 first-episode (FE) patients, of whom 9 were felt to be stabilized, while 6 were persistently symptomatic. As a group, FE patients showed a moderate effect-size (d = 0.58) deficit in tone matching. Further, approximately one-third of FE patients showed tone-matching thresholds outside the control range (>10% difference in pitch), leading to a significant difference vs age-matched controls (P < .02).33

Deficits in MMN generation have also been reported early in the course of schizophrenia, with onset of deficit within 1.5 years of first hospitalization.34–36 Deficits in MMN generation have not been observed in patients during first hospitalization overall,35–37 although the number of subjects remains small. Further, even among patients studied within the first hospitalization, a significant correlation has been observed between MMN amplitude and the education item of the Premorbid Adjustment Scale, such that those patients who failed to complete high school showed significantly reduced MMN at first onset relative to both age-matched controls and FE patients with college education.35

Prosodic detection has, at present, been evaluated in only one study of FE patients of which we are aware. In that study, patients were evaluated within 6 months of discharge following initial hospitalization. Notably, mean years of education for this group was 10.8, with less than 30% of patients having any degree of college education. FE psychotic patients as a group showed a moderate (d = 0.69) decrement in affective prosodic detection relative to age-matched controls.

As with the present study, there was no significant difference across emotions. Patients nevertheless showed apparently greater deficits on detection of fear and sadness vs angry stimuli, similar to the pattern observed in the present study. Thus, while comparison across studies is difficult due to different cohort composition and different definitions of the term “FE,” both tone matching and prosodic detection deficits appear to be present early in the course of the illness and may be enriched with those subjects showing relatively low educational attainment. Effect sizes for both early auditory dysfunction and affective prosodic deficits appear to progress as a function of illness chronicity, although whether this reflects true degeneration within individual subjects, as opposed to distillation of poor-outcome subjects within chronic patient
coherent, remains to be determined. Overall, more study of both basic auditory processing ability and affective prosodic ability over the course of schizophrenia is required.

In the present study, patients also showed significantly reduced sensitivity to differential strength of emotional portrayals, in particular tending to overestimate intensity of intended weak emotions. As with categorization deficits, the failure to discriminate emotional intensity reflected an inability to utilize tonal, rather than intensity, information and correlated with reduced perceptual sensitivity. This is the first study to evaluate the ability of patients to discriminate auditory emotional intensity, along with emotion identity. As with deficits in emotion identification, the tendency to overestimate strength of weak emotional portrayals may also lead to significant misinterpretations in social communicatory situations.

The present study represents an initial attempt to go beyond simple correlational analyses of emotion processing accuracy vs pitch measures (eg, tone matching) and to develop instead a taxonomy of affective dysprosodia in schizophrenia. Development of such a taxonomy is crucial not only for achieving greater understanding of the neurophysiological bases of acoustic prosodic dysfunction in schizophrenia but also for the development of appropriate remediation or compensation techniques. This ecological approach is akin to attempts in both autism and schizophrenia to link abnormal gaze of facial features to facial affect deficits. Nevertheless, because it is the first study of this type, certain limitations must be acknowledged. First and foremost, because multivariate featural batteries of this type have not previously been used for between-group analyses, the statistical approach for between-group analyses was developed post hoc. Because of this, results must be considered exploratory and confirmed with additional stimulus batteries, additional patient samples, or both. Nevertheless, we feel that the observation that controls show strong variation in ability to identify intended emotion based upon pitch-based measures whereas patients do not strongly validates the utilized analysis approach.

Second, the study took advantage of an existing, naturalistic stimulus set rather than using synthetically constructed utterances with predetermined characteristics. As a result, numerous stimulus features were strongly intercorrelated, limiting the extent to which any single feature could be isolated. This battery has the advantage of having physical features that have previously been published. Nevertheless, future synthetic stimulus development is needed in which specific parameters are modulated independently across a continuum of levels. This stimulus set further consisted of posed portrayals of emotional speech rather than evoked prosody from natural discourse. Though it can reasonably be argued that posed expressions must be relatively similar to naturally occurring expressions in order for communication to be successful, it may nevertheless be the case that posed expressions are exaggerated and more intense than authentic expressions and that the acoustic properties between posed and authentic stimuli may differ. Similarly, the ability to generalize performance from tasks involving posed portrayals to normal discourse may have its limitations. Nevertheless, because speakers may have complex emotions during normal discourse, obtaining objective validation of intended emotion may be impossible under naturalistic circumstances. All patients were also receiving antipsychotic medication at the time of testing, raising the possibility of a medication effect. However, no correlation was observed between performance and medication dosage.

In conclusion, disturbed emotion identification ability represents a key determinant of social cognition and functional outcome in schizophrenia. This is the first study to evaluate contribution of specific underlying acoustic cues to emotion identification dysfunction, as well as to assess the perception of emotion intensity. Primary findings are that patients show intact ability to utilize intensity-based cues but reduced ability to identify emotions based upon critical pitch modulations. Such deficits contribute to disturbances not only in identification of intended emotions but also in discrimination between intended weak and strong emotional portrayals. Further, as in prior studies, deficits correlated highly with more basic deficits in auditory processing, as well as with global outcome measures. These findings indicate the need for cue-, as well as emotion-, based assessment of prosodic dysfunction in schizophrenia and for ecological approaches to conceptualization and remediation of social communicatory impairments in schizophrenia. Furthermore, a cue-based approach may provide a method to discriminate schizophrenia dysprosodia from dysprosodias found in other illnesses such as Parkinsonism and autism.

Funding

National Institute of Mental Health (NRSA F1-MH067339 to D.L.L., K02 MH01439 to D.C.J., R01 MH49334 to D.C.J.).

References


