J Psycholinguist Res (2007) 36:159–173
DOI 10.1007/s10936-006-9037-4

ORIGINAL ARTICLE

Intonation and Emotion in Autistic Spectrum Disorders

Kathleen Hubbard · Doris A. Trauner

Published online: 29 November 2006

Abstract The classic picture of an autistic individual includes an impoverished ability to interpret or express emotion. The prosody of spoken language in autistic children is thought to lack emotional content. In this study, the verbal intonation of children with autism was examined and compared to that of children with Asperger Syndrome (AS) and normal controls (ctrl). Utterances elicited by repetition and by spontaneous story completion were analyzed by quantifying phonetic features (pitch, amplitude, and length) and comparing them to subjective ratings of produced emotion (happy, sad or angry). Since the most consistent phonetic correlate of these emotional targets has been demonstrated to be pitch range, speakers with autistic spectrum disorders were expected to have decreased pitch range; however in the repetition task, autistic subjects actually had a larger pitch range than the other groups. Other measures of intonation including amplitude, duration, and location of pitch peak revealed defects that are more complex than predicted. In spontaneous speech, autistic subjects performed more poorly on both phonetic targets and subjective ratings than ctrls, and AS subjects fell between autistics and normals.

Keywords Autism · Asperger syndrome · Prosody

Introduction

Autism (Au) is a developmental disorder that manifests before the age of 3 years and is characterized by language and speech impairments, abnormal interpersonal

K. Hubbard
Department of Psychiatry,
University of California San Diego School of Medicine,
La Jolla, CA, USA

D. A. Trauner (✉)
Department of Neurosciences and Pediatrics, UCSD School of Medicine,
9500 Gilman Drive, La Jolla, CA 92093-0935, USA
e-mail: dtrauner@ucsd.edu

interactions, social withdrawal, and typically cognitive deficits as well. Asperger Syndrome (AS) is considered to be an autistic spectrum disorder; children with AS tend to have normal intelligence and verbal abilities and a higher degree of social awareness than autistic children, but they have difficulty with social interaction and are often perceived as eccentric. In both disorders, individuals are often described as having "monotonous" speech or inappropriate tone of voice.

This study measured concrete parameters to test the subjective impression that the speech of children with autistic spectrum disorders differs from that of normal controls (ctrl) in its prosody, specifically in the correlation of prosody with emotional content. Few investigations of this question have been reported in the literature (Shriberg, 2001), and they have relied primarily on subjective ratings, not phonetic measurement of speech production. However, the phonetic features of emotional intonation have been examined in normal adult speech (Williams & Stevens, 1972; Streeter, MacDonald, Apple, Krause, & Galotti, 1983, Murray & Arnott, 1993, Pell, 2001), so results for the present subject groups can be evaluated and interpreted similarly.

The phonetic parameters in question are pitch, intensity/loudness, and length of speech sounds, i.e., the perceptual correlates of fundamental frequency (F0), amplitude, and duration, respectively. Together these features comprise what is perceived as intonation, which can be considered at the level of the word, phrase, or utterance (Frick, 1985). It is important to note that in English, the same components that are used to convey affective content are also employed in the realization of lexical stress (distinguishing words such as "PROject" and "proJECT") and sentential focus (e.g., contrastive "MARY gave me the book" versus "Mary gave me the BOOK"). Consideration of the interaction between linguistic and affective uses of intonation is beyond the scope of this paper (see, e.g., Bolinger, 1955, Ladd, 1996), however it has been claimed that the emotional overlay can be quantified essentially independently (McRoberts, Studdert-Kennedy, & Shankweiler, 1995) and this is the intent of the two experiments reported here. Data on sentential focus have been collected from the same subjects and will be examined separately.

In this study, subjects performed two speech tasks, one involving repetition of a stimulus phrase and the other eliciting spontaneous speech. Three target emotions were used: happy (H), sad (S), and angry (A). Subjective ratings of each utterance were performed to identify which emotion the speaker was conveying.

Methods

Participants

The 28 subjects in this study included nine children with a diagnosis of Au, nine with AS, and ten normal ctrls. Those in the experimental groups met DSM-IV (2000) diagnostic criteria for Au or AS, respectively. Ctrl subjects had no known speech or hearing deficits. Children with Au and AS were recruited from pediatric neurology clinics and from referrals from psychologists. Ctrls were recruited through advertisements in local parent magazines and through the UCSD subject pool. Most of the children were participating in a larger study of language and learning in children. Subjects ranged in age from 6 to 21 years, with a mean age of 14.5. It was not possible to balance the groups for gender because the incidence of autistic spectrum disorders, especially AS, is much higher in boys (Tanguay, 2000, Willemsen-Swinkels, 2002).

The Au group included six boys and three girls, the Asperger group had nine boys, and the ctrl group contained nine boys and one girl. Informed consent was obtained for each child's participation, as approved by the U.C. San Diego Human Research Protections Program.

Data collection

Experiment 1

In this repetition task, the child listened to a tape recording of an actress speaking a neutral-content phrase in one of three emotional intonations, and was instructed to repeat that phrase. For example, the phrase "it's almost finished" was modeled once each in happy (H), sad (S), and angry (A) intonation. Audiocassette recordings of the child's repetitions were made. Five phrases were presented three times, for a total of 15 tokens per subject. The order of phrases was constructed such that a given phrase did not appear twice in a row, nor did a given emotional intonation. Text of the stimuli for Experiment 1 is given in Appendix A. All of the participants were able to repeat the target items verbatim.

Experiment 2

In this free-response task, the child listened to a short story read by the experimenter and was instructed to complete the story in a single sentence while pretending to be one of the characters. Ten vignettes were presented, after each of which a free-response sentence was produced by the subject. Responses were tape-recorded. Each story was intended to evoke a clear emotional response of happy, sad or angry; for example, a character tells a friend he/she is being taken to Disneyland (happy). There were three stories designed to elicit a happy (H) intonation, three angry (A) and four sad (S). The expected responses for Experiment 2 are given in Appendix B. The order of experiments was randomly assigned such that half the subjects did Experiment 1 first, and half did Experiment 2 first (interspersed with other verbal tasks which are not examined in this study).

Because the audio recordings were made at different times and under variable conditions, background noise, and distance from subject to microphone were not constant. As a result some tokens were unusable and were excluded from measurement and analysis. In addition, four of the autistic subjects were unable to complete Experiment 2 because they did not understand the free-response task. These subjects were included only in the analysis of Experiment 1.

Phonetic measurement

The author was not present for taping of subjects, and remained blinded to the group identity of subjects and the intended emotional content of utterances throughout the process of phonetic measurement. Each utterance was 16-bit digitized at a sampling rate of 44.1 kHz using CoolEdit 2000 software and a standard PC. Frequency, amplitude, and duration measurements were recorded using Pitchworks 5 software. Pitch extraction was performed using an autocorrelation algorithm. Since subjects varied in age and gender, so did their inherent pitch; thus autocorrelation parameters

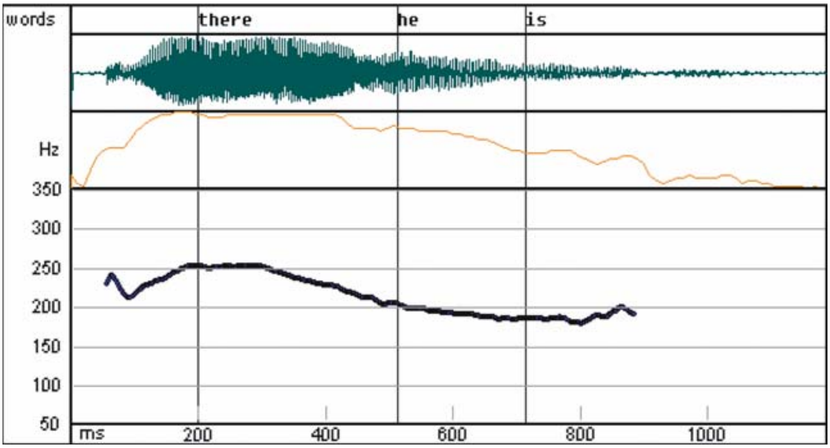


Fig. 1 Acoustic display of token LC514

(frequency deviation, frequency threshold, and calculation range) were adjusted for each subject's voice.

Experiment 1

Vowel F0, amplitude, and duration values were recorded for each syllable in each utterance. If a syllable had uniform pitch, one F0 value was recorded in hertz (Hz), according to the usual conventions of avoiding consonant transitions at the beginnings and ends of vowels (Bachorowski & Owren, 1995). If a pitch contour was present within a syllable and was judged to be a phonological contour (rather than a prolonged transition between vowels or the effect of utterance-final lowering), multiple values were recorded at the frequency maxima and minima. Amplitude values were recorded in decibels (dB) at the same time points as the F0 values. Total vowel duration was logged in milliseconds (ms). An example of phonetic labeling of an utterance is shown in Fig. 1.

An important feature of intonation in English is pitch *range*, i.e., how much variation a speaker employs in a given utterance (Lieberman & Michaels, 1962, Sobin & Alpert, 1999). Less consistent but also relevant is intensity or amplitude range. Thus raw F0 and amplitude ranges were calculated for each utterance, for comparison within subjects. In addition, because of the variation in inherent pitch and intensity across subjects, each F0, and amplitude value was normalized using the standard formula:

$$Z = \frac{X - \mu}{\sigma}.$$

Utterance ranges as well as total range for each subject were calculated using these normalized values for between-subjects comparison. For the third component of intonation, vowel duration, it is necessary to compensate for variation in speech rate as well as inherent differences between vowels (i.e., [a] takes longer to articulate than [i]). Therefore the longest value for each utterance was re-calculated as a percentage of total vowel duration for the sentence. Finally, the syllable locations of pitch peak (PPk), amplitude peak (APk), and duration maximum (DPk) were recorded and compared to the target utterances in the experimental stimuli.

Intended emotional content (IE) of happy (H), sad (S) or angry (A) was coded for each token, and group identity of Au, As or normal Ctrl was coded for each subject. Within-subject ANOVAs were performed for pitch and amplitude ranges (PRng, ARng) and maximum duration percentage (DPkPct), to test significance of differences by IE. Between-subjects ANOVAs were performed for normalized pitch and amplitude ranges (NPRng, NARng), and for DPkPct, to determine whether there were significant interactions of group and IE. In addition, location of pitch (PPkLoc), intensity (APkLoc), and duration maxima (DPkLoc) were scored as one if they matched the stimulus target and zero if they did not; accuracy scores were calculated for each subject and compared across groups.

Experiment 2

Because responses varied so widely in content in the spontaneous-speech task, it was not possible to make syllable-by-syllable comparisons as in Experiment 1; pitch range was the only phonetic variable measured. F0 measurements were recorded for each syllable using the same methods as above, and raw and normalized ranges were calculated for each utterance as well as total range for each subject. Expected emotional content and group identity were coded, and within- and between-subjects comparisons of pitch range were performed as in Experiment 1.

Subjective rating

Each utterance in both experiments was assessed for emotional content by three independent raters, who were blinded to the group identity of the subject and the target emotion of the sentence. Raters were trained to rate a large, representative set of model items (using the actual tapes of the actress) and had to achieve 98% accuracy before they were permitted to do actual subject rating. Utterances were rated as happy (H), sad (S), angry (A) or ambiguous/ indeterminate (Q). Where there was disagreement among the three listeners, the majority rating was recorded. If the recorded rating matched the target emotion, the item was scored correct; if it was judged to be a different emotion or equivocal, the item was scored incorrect. Total scores for each experiment were recorded for each subject, expressed as a percent correct. These subjective ratings were compared to determine correlation between intended emotional content and listener-assessed intonation among the two experimental groups and the ctrl group.

Results

Experiment 1

Pitch

A statistically significant correlation between pitch range and intended emotion was found in all three groups (Table 1). (Note that the means in Table 1 are unitless because F0 values were normalized in order to be comparable across subjects.)

Table 1 Mean normalized pitch range of tokens by IE and group (with s.e.)

	H	S	A	Total	N
Au	2.431 (0.124)	1.098 (0.126)	1.931 (0.121)	1.828 (0.090)	130
As	2.755 (0.132)	1.048 (0.142)	1.872 (0.126)	1.928 (0.097)	113
Ctrl	2.676 (0.116)	1.019 (0.130)	1.699 (0.118)	1.856 (0.088)	130
Total	2.618 (0.071)	1.056 (0.076)	1.830 (0.070)		
N	130	114	135		397

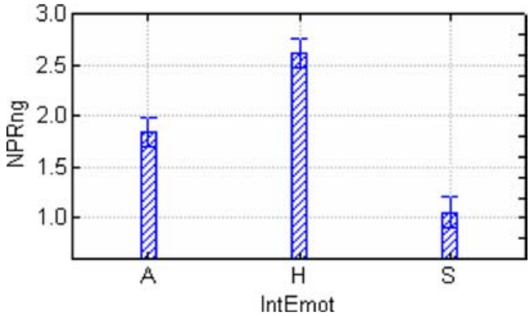


Fig. 2 Normalized pitch range by IE

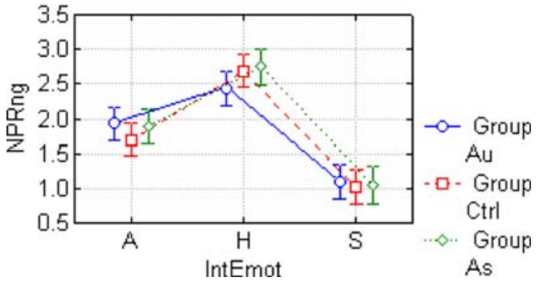


Fig. 3 Normalized pitch range by IE and group

The pitch range differences by emotion are highly statistically significant (ANOVA NPRngXIE $F(2, 376) = 111.81$, $p = 0.0000$, post hoc Tukey HSD significant for all pairwise comparisons). All groups show the same pattern of mean pitch range, namely H>A>S (which is consistent with what has been shown for normal adult speakers) (Fig. 2).

Two-way ANOVA shows no significant effect of experimental group on this pattern (Fig. 3). This result disconfirms the hypothesis that pitch range would be diminished in autistic subjects, and less markedly diminished in AS subjects. In fact, when total pitch range across all tokens is considered, autistic subjects have by far the greatest range(Fig. 4) (Table 2).

Within-subjects analysis shows that three of nine autistic subjects did not have a correlation of pitch range with emotion, and another showed a significant difference but with a different pattern from the norm (A>H=S, rather than H>A>S). One of nine AS and one of nine ctrl subjects also failed to reach statistically significant separation of pitch range by emotion. The remaining subjects, all of whom had significant variance of pitch by emotion, also had the canonical ranking (H>A>S)—thus with

Table 2 Autistic subjects with high accuracy on Experiment 1 ratings

subject	Experiment 1 score (% correct)	Experiment 2 score (% correct)
A	80	–
B	93	–
C	93	60

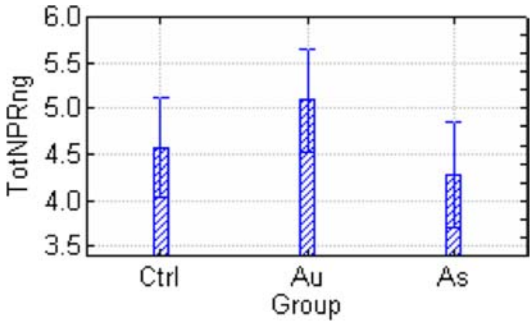


Fig. 4 Total normalized pitch range by group

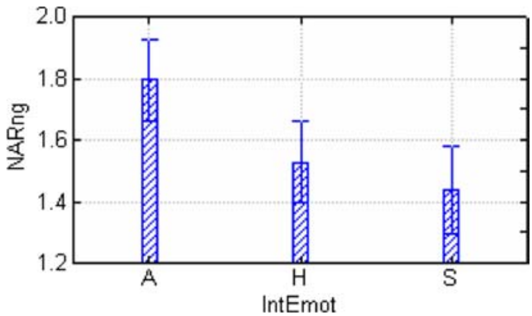


Fig. 5 Normalized amplitude range by emotion

the exception of one autistic speaker, when subjects made emotional distinctions in pitch they made the same ones.

Amplitude

The role of amplitude in marking emotion in this task is distinct from that of pitch: there is a difference in amplitude range between emotions, but it is not consistent across groups. Overall the effect of amplitude range is significant ($F(2, 376) = 7.3821$, $p = 0.00072$), but pairwise this only holds for A versus H and A versus S (Fig. 5).

This is because while the control group significantly distinguishes A>H=S (exactly as in the target stimuli), and the autistic group has a similar trend which does not reach statistical significance, the AS group has a significant difference of A=H>S (Fig. 6).

There is no significant interaction between IE and group. This result shows that, just as for pitch, the experimental groups (Au versus As) do not separate distinctly in how amplitude corresponds to emotion. It suggests, however, that normal speakers do use intensity as a component of encoding anger, and that speakers with autistic spectrum disorders do not do so reliably.

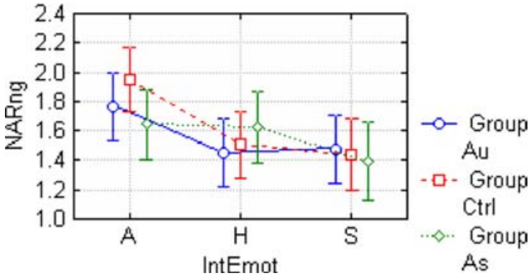


Fig. 6 Normalized amplitude range by IE and group

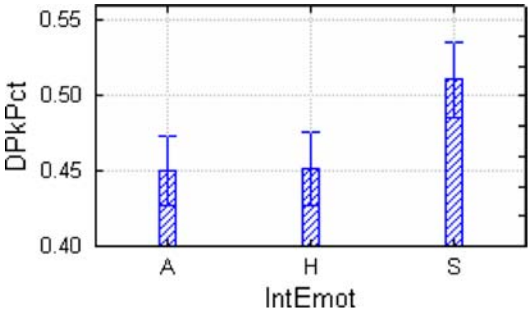


Fig. 7 Duration peak percentage by IE

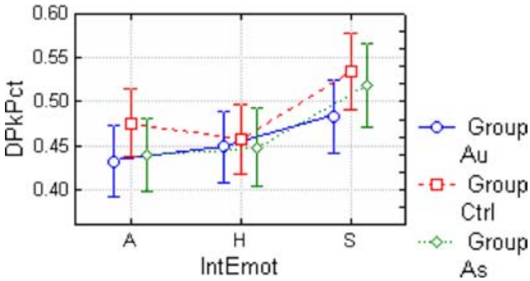


Fig. 8 Duration peak percentage by IE and group

Duration

Overall, there is a significant correlation of maximum syllable duration with emotion ($F(2, 377) = 7.4597$, $p = 0.00066$), which distinguishes Sad as the emotion with the longest maximum vowel length (Fig. 7). This ranking of S>H=A is concordant with what has been found for normal adult speakers, and is said to be a matter of slower speech rate in sad utterances. Autistic subjects, however, do not show the duration effect as reliably as the other two groups (Fig. 8).

While the ctrl and AS groups have a statistically significant difference in S>H=A duration, autistic subjects do not. In this case, AS subjects appear to group with ctrls in using duration as an additional emotion cue; autistic subjects do not do so consistently.

Peak locations

Since an important aspect of intonation is the placement of stress (a combination of pitch, intensity, and duration), it is useful to consider not only the magnitude of each phonetic component but also its location. Syllable location of each of the three

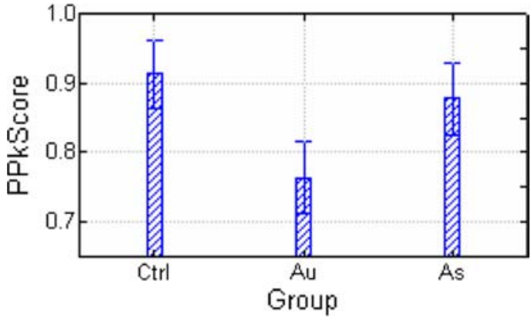


Fig. 9 Pitch peak location accuracy by group

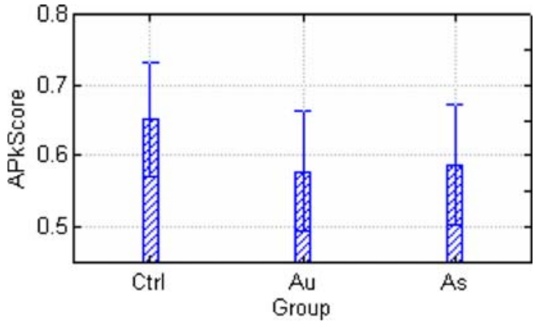


Fig. 10 Amplitude peak location accuracy by group

phonetic maxima was recorded and compared with the recorded stimuli. Subjects were scored for accuracy as percent correct averaged over all tokens. As before, pitch emerges as the most revealing (Fig. 9).

The overall effect of group on PPkLoc accuracy is significant ($F(2, 25) = 9.7347$, $p = 0.00075$), with pairwise comparison showing the autistic group significantly different from AS and ctrl.

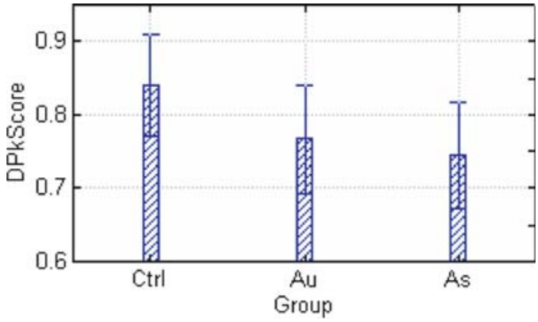
Interestingly, APKLoc distinguishes the groups differently from pitch: though the group effect does not reach statistical significance, the trend indicates that both autistic and AS subjects failed to match intensity peaks to the target utterances as reliably as ctrls did (Fig. 10).

The DPKLoc is also not significantly different by group, although the trend again separates the two experimental groups from the ctrl group (Fig. 11). What this indicates is a tendency for normal speakers to correctly emulate the target stimulus in prolonging a particular syllable depending on emotion, while the two groups of experimental subjects are less consistent in doing so.

Subjective ratings

For Experiment 1, subjective ratings of emotional content of utterances show autistic subjects to be less successful at encoding target emotion (59.1% correct) than AS (85.1%) or ctrls (76.7%). This difference is significant ($F(2, 24) = 4.0600$, $p = 0.03029$) for autistic versus other; AS and ctrl are not significantly different. Interestingly, three of nine autistic subjects scored very high on accurate production of target emotional intonation in Experiment 1; of these two were unable to perform Experiment 2, and the third received a low-rating score approaching chance on Experiment 2.

Fig. 11 Duration peak location accuracy by group



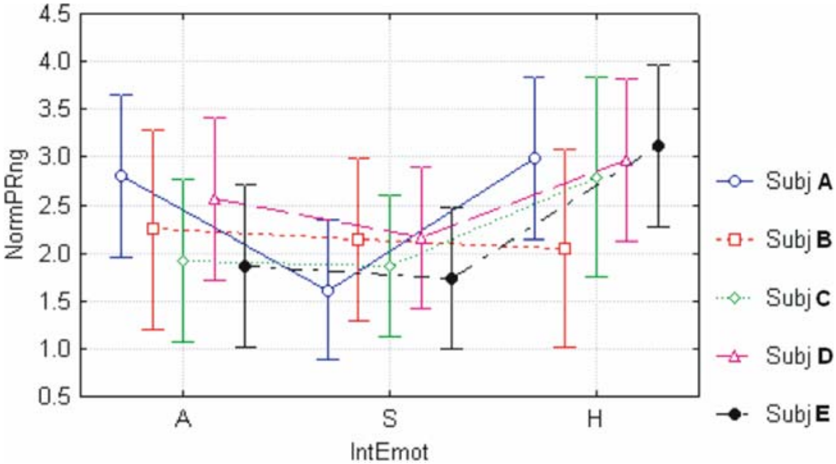


Fig. 12 Normalized pitch range by IE and subject, autistic group

It is possible that these children are particularly good at mimicking speech sounds but have difficulty producing spontaneous speech (or for subject *C*, at least with canonical intonation).

Experiment 2

Pitch range

The performance of subjects on the free-response task contrasts sharply with that of the repetition task. Of note, four of nine autistic subjects were unable to complete the task at all: in attempts at the first few tokens, these subjects either perseverated on the content of the story (or even on a token from an earlier task such as Experiment 1), failed to use first-person statements, or responded with a single word rather than a sentence. Of those who did successfully complete the task, none had a significant correlation of pitch range with emotion (Fig. 12).

As can be seen in Fig. 12, one subject had the same pitch range for all emotions, two had a larger range only for H, and the other two had H=A>S—no consistent pattern at all.

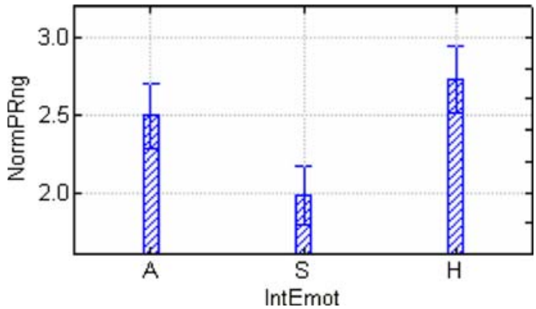


Fig. 13 Normalized pitch range by IE

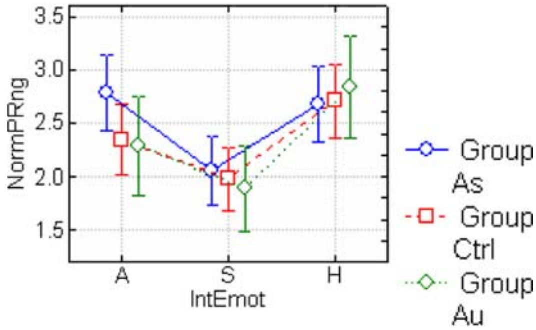


Fig. 14 Normalized pitch range by IE and group

Responses from the AS and ctrl groups were also, as expected in a spontaneous speech task, less consistent than in Experiment 1. Just one of eight AS subjects had a significant correlation of pitch range with emotion, with a ranking of A>H>S (rather than the H>A>S produced by the group and by that individual in Experiment 1). Two of nine ctrl subjects showed statistically significant effects of emotion on pitch range, and both had the expected H>A>S pattern. The majority of individuals in all groups did not display as robust a correlation of pitch range with emotion as they did in Experiment 1. Still, the overall effect across subjects was significant ($F(2, 200) = 13.316$, $p = 0.00000$), with the same H>A>S pattern (Fig. 13).

As before, there was no significant interaction of group and emotion (Fig. 14).

Subjective ratings

Subjective ratings are correlated with group here as in Experiment 1, with autistic subjects scoring lowest (55.0% correct), then AS (75.6%), and ctrls highest (86.7%). The overall differences are statistically significant ($F(2, 21) = 5.0823$, $p = 0.01584$), though pairwise only for autistic versus ctrl.

Summary results for the two experiments are shown in Tables 3 and 4.

Discussion

The results of the two experiments presented here do not support a characterization of prosody in autistic spectrum disorders as "flat" or "monotone." On the contrary, autistic and AS speakers in this study produced a full range of prosodic features,

Table 3 Summary results of Experiment 1

	Pitch range (tokens)	Total pitch range	Amplitude range	Duration maximum	Pitch accuracy	Amplitude accuracy	Duration accuracy	Subjective accuracy
Standout group	none	Autistic	Asperger	Autistic	Autistic	Control	Control	Autistic
Group pattern	Au = As = Ctrl	Au > As = Ctrl	As ≠ Au = Ctrl	Au ≠ As = Ctrl	Au < As = Ctrl	Ctrl > As = Au	Ctrl > As = Au	Au < As = Ctrl
Emotion pattern	H > A > S		A > H = S	S > H = A				

Au = Autistic, As = Asperger Syndrome, Ctrl = Control
H = Happy, A = Angry, S = Sad

Table 4 Summary results of Experiment 2

	Pitch range (tokens)	Subjective accuracy
Standout group	none	None
Group pattern	Au = As = Ctrl	Au < As < Ctrl
Emotion pattern	H > A > S	

Au = Autistic, As = Asperger Syndrome, Ctrl = Control
H = Happy, A = Angry, S = Sad

particularly pitch which has been implicated in subjective descriptions of the speech abnormalities of these children. In fact, the hypothesized difference in pitch realization of emotional targets between experimental and ctrl groups did not appear in this dataset; instead, all groups in both settings showed a strong distinction between Happy, Sad and Angry which was not statistically influenced by group identity. However, when other features are considered, notable patterns emerge.

As seen in Tables 3 and 4, the various parameters tested sort the experimental and ctrl groups differently: in Experiment 1 the most salient feature, pitch range of tokens, does not separate the three groups at all. Total pitch range, on the other hand, shows the autistic group to be distinct from the others (in an unexpected direction, with much greater range). The location of pitch peaks also separates out the autistic subjects, with a much poorer correlation with target stimuli. Amplitude shows trends, not robust differences, but the trends are suggestive: the ctrl group has significantly different amplitude range by emotion, the AS group also varies range by emotion but with a non-canonical pattern, and the autistic group shows no correlation at all. Placement of amplitude peaks shows the ctrl subjects to be highly accurate in imitating the targets, while the other two groups are not. Duration maxima separate the autistic group (poorer correlation with emotion) from the other two; placement of duration peaks is successfully matched only by ctrls and not the experimental subjects. In spontaneous speech, pitch range singles out autistics as having very poor correlation with emotion. Subjective ratings, interestingly, pick out the autistic group both in the repetition task, and spontaneous speech, but also reveal a trend for AS subjects to score worse than ctrls in spontaneous speech only. In short, each measure suggests differential abilities across the groups, but sometimes AS subjects group with autistics, and sometimes with ctrls.

The fact that autistic subjects had a significantly larger pitch range than the other two groups in the repetition task suggests that they do process pitch as a salient indicator of affective differences; what is anomalous about their production in this setting is that they *overshoot* the intonational target. However, the data also suggest that they miss other cues: amplitude and duration, which both vary with intended emotion in ctrl subjects and in the experimental stimuli, are not consistently employed by autistic subjects in this task. Perhaps most noticeable to listeners, autistic subjects tend to misplace the pitch peak in the sentence. So although in imitating a speech stimulus the autistic speakers clearly do not have "monotonous" intonation, they do have flatter amplitude and duration profiles as well as anomalous location of maximum pitch, which may account for the subjective impression of faulty prosody. The ratings scores support this interpretation: despite their large pitch ranges and significant differences by IE, autistic subjects were rated just 59% correct in reaching emotion targets in Experiment 1, significantly different from AS and ctrls.

It is intriguing that some autistic subjects outperform normal controls on subjective ratings of emotion in the repetition task, yet all score very low on ratings of spontaneous speech. This raises the question of whether the high-scoring subjects are truly processing the stimuli in Experiment 1 as speech or rather as sounds. Among autistic children who are more severely language-impaired than those in this study, echolalia is not uncommon; in these children it often appears that even when articulation is accurate, there is little or no comprehension of the linguistic content of these utterances. It may be that the exaggerated pitch range of autistic subjects in the repetition task is more a matter of sound mimicry than emotional expression.

The weaker effect of IE on pitch range in Experiment 2 than Experiment 1, which is assumed to be a matter of higher variability in the free-response task, may also occur because subjects assigned a different emotional content to a given short story than was intended by the experimental design. For example, the expected response for item 6, "Kathy won't let me look at the book," was intended to have Angry intonation but was produced by some subjects with what sounded like Sad intonation. (This item alone is not responsible for the statistical results, however; excluding it does not substantially change the outcome.) In any event, it is not possible to determine with certainty whether a subject's emotional targets were the same as intended, and therefore whether it is their prosodic realization that is faulty or their interpretation of the stimulus story.

Interestingly, AS subjects sometimes removed this uncertainty by making explicit statements about how they (as the story character) felt. For example, where most subjects produced something like "stop kicking sand in my face," several AS subjects added "that makes me really angry" to the beginning or end of the sentence. Similarly, "I love pizza" became "I'm so happy we're having pizza for dinner." These children also had a tendency to produce longer utterances than their counterparts in other groups, with or without the emotional commentary. One can only speculate about the reasons for this; it is possible that some individuals with AS are aware of the difficulties they have communicating with others, and attempt to compensate by adding more content or explication to their speech. Quantification of utterance length in spontaneous speech tasks in AS versus ctrls may be a fruitful avenue for future research.

The expected result of this study was that some profile of phonetic features of affective intonation would correspond to subjective rating of emotion in distinguishing the experimental groups from ctrls and from each other. While subjective ratings do differ significantly by group, there is no direct correlation of ratings with any acoustic measure. Some phonetic features show clear effects, but they differ as to whether they group autistic and AS subjects together, or AS and ctrls. In Experiment 1 the measures that most closely approach the subjective rating result are total pitch range—which may reflect a non-linguistic phenomenon, as mentioned above—and pitch peak location accuracy, which is more likely to be a reproducible result. Other features have marked trends but in this dataset are not statistically significant. This outcome is not unprecedented; other studies have found for normal speakers that subjective rating is more reliable in identifying affective content than acoustic analysis (Banse & Scherer, 1996). What this suggests is that we have not yet identified the multifactorial interactions that go into producing the acoustic complex that counts as emotional intonation. The data here argue for examining not only the presence or absence of a given feature, but how it is used by speakers, including timing, consistency, and magnitude. This more detailed type of analysis may reveal more precisely the nature of prosodic impairment among children with autistic spectrum disorders.

Appendix A. Experiment 1 stimuli (H=happy, S=sad, A=angry)

1. It's all over. (A)
2. There he is. (H)
3. I know. (S)
4. I could do that. (A)
5. There he is. (S)
6. It's all over. (H)
7. I could do that. (S)
8. It's almost finished. (A)
9. It's all over. (S)
10. I know. (H)
11. It's almost finished. (S)
12. I know. (A)
13. It's almost finished. (H)
14. There he is. (A)
15. I could do that. (H)

Appendix B. Experiment 2 expected responses and emotional content

1. Give me back my cheese. (A)
2. I lost my favorite teddy bear. (S)
3. I'm going to Disneyland tomorrow. (H)
4. Mother says I can't keep you. (S)
5. I found a quarter. (H)
6. Kathy won't let me look at the book. (A)
7. Stop kicking sand in my face. (A)
8. I can't find my shiny new car. (S)
9. I love pizza. (H)
10. I can't go to the movies. (S)

References

Bachorowski, J., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science, 6*, 219–224.

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Society of Psychology, 70*, 614–636.

Bolinger, D. L. (1955). Intersections of stress and intonation. *Word, 11*, 195–203.

Diagnostic and Statistical Manual of Mental Disorders—Text Revision (DSM-IV—TR) (2000). Arlington, VA: American Psychiatric Association.

Frick, R. W. (1985). Testing visual short-term memory: Simultaneous versus sequential presentations. *Memory & Cognition, 13*(4), 346–356.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustical of Society of America, 34*, 922–927.

McRoberts, G. W., Studdert-Kennedy, M., & Shankweiler, D. P. (1995). The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation. *Perception and Psychophysics, 57*, 159–174.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of Acoustical Society of America, 93*(2), 1097–1108.

Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *Journal fo Acoustical Society of America, 109*(4), 1668–1680.

Shriberg, L. D. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech Language and Hearing Research, 44*(5), 1097–1115.

Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research, 28*(4), 347–365.

Streeter, L., MacDonald, N., Apple, W., Krause, R., & Galotti, K. (1983). Acoustic and emotional indicators of emotional stress. *Journal of the Acoustical Society America, 73*, 1354–1360.

Tanguay, P. E. (2000). Pervasive developmental disorders: A 10-year review. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(9), 1079–1095.

Willemsen-Swinkels, S. H. (2002). The autistic spectrum: subgroups, boundaries, and treatment. *Psychiatric Clinics of North America, 25*(4), 811–836.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of Acoustical Society America, 52*(4), 1238–1250.