Screaming, Yelling, Whining, and Crying: Categorical and Intensity Differences in Vocal Expressions of Anger and Sadness in Children's Tantrums

James A. Green and Pamela G. Whitney
University of Connecticut

Michael Potegal
University of Minnesota Medical Center

Young children's temper tantrums offer a unique window into the expression and regulation of strong emotions. Previous work, largely based on parental report, suggests that two emotions, anger and sadness, have different behavioral manifestations and different time courses within tantrums. Individual motor and vocal behaviors, reported by parents, have been interpreted as representing different levels of intensity within each emotion category. The present study used high-fidelity audio recordings to capture the acoustic features of children's vocalizations during tantrums. Results indicated that perceptually categorized screaming, yelling, crying, whining, and fussing each have distinct acoustic features. Screaming and yelling form a group with similar acoustic features while crying, whining, and fussing form a second acoustically related group. Within these groups, screaming may reflect a higher intensity of anger than yelling while fussing, whining, and crying may reflect an increasing intensity of sadness.

Keywords: temper tantrums, vocalizations, acoustic features, anger, sadness

Temper tantrums are both prevalent and frequent in young children, and they often present a serious management problem for their parents. Tantrums are challenging to prevent and even more difficult to interrupt once underway. Their emotional momentum speaks to the intensity of the emotions involved. Furthermore, tantrums that are excessive in early childhood and/or persist into later childhood can predict future maladjustment and even psychopathology (Caspi, Elder, & Bem, 1987; Stevenson & Goodman, 2001; Stoolmiller, 2001). Because tantrums offer a window onto the early expression and (dys)regulation of strong emotions that are otherwise difficult to observe, they are a compelling phenomenon for scientific study (Potegal & Davidson, 2003).

Goodenough's (1931) pioneering monograph *Anger in Young Children* was an early landmark that drew from parental diaries to chronicle outbursts of 45 children between 6 months and 8 years of age. The most frequent motor acts during tantrums were kicking, stamping, jumping up and down, hitting, and throwing the self on floor, which occurred in somewhere between 3% and 28% of tantrums. The most common vocal behaviors were classified as cry, scream, fuss, whine, and snarl, occurring in from 25% to 85% of tantrums, depending on age. Analyzing a new set of narratives written more than 60 years later by the parents of 335 children 18

months to 5 years old, Potegal and Davidson (2003) found that crying was the most frequent vocal expression, occurring in 86% of tantrums. Screaming and shouting (hereafter termed yelling) occurred in about 40% of tantrums, and whining in about 13%.

The latter prevalence estimates are remarkably similar to those of Goodenough (1931.) In our view, however, anger is only half the story of tantrums. Factor analyses by Potegal and Davison (2003) indicated that these motor and vocal acts form groups that represent varying degrees of two different emotions, anger and "distress" (sadness and comfort-seeking). With regard to anger, Potegal and Davidson (2003) found three factors that could be ranked as different in levels of intensity. *Scream* loaded with kick, hit, and stiffen on a factor labeled *High Anger*. *Yell* loaded with throw on a second factor labeled *Intermediate Anger*. The lowest intensity anger factor, defined by stamping, contained no vocal component. Multidimensional scaling showed that all three factors had a similar temporal pattern of peaking near the onset of the tantrum and declining thereafter suggesting that they were indeed expressions of the same emotion (Potegal, Kosorok, & Davidson, 2003). The specific behavioral components of the respective factors provided face validity for their identification as different intensities of anger. This identification was further supported by the progressively higher correlations of Low, Intermediate, and High Anger with tantrum duration and autonomic activation, as well as with parental judgments of overall tantrum intensity.

"Distress" in these publications referred to expressions of sadness and co-occurring comfort-seeking. The vocalizations *whine* and *cry* loaded together on a factor representing sadness. These vocalizations were more evenly distributed across the tantrum in a temporal pattern distinctly different from the anger-related behaviors. Anger and sadness factors were also differentially correlated with likelihood of parental intervention during the tantrum. Because the present report does not address comfort-seeking, we will refer to the second tantrum emotion simply as sadness. However,

the use of *anger* and *sadness* as descriptors of emotions displayed during tantrum behaviors should not be interpreted as implying continuity with adult expressions of anger and sadness.[1]

The model of tantrums as composed of groups of behaviors reflecting anger and sadness at different levels of intensity (Potegal & Davidson, 2003) has been extended and confirmed by the finding that the on-ward "rages" of older child psychiatry inpatients have a structure quite similar to the tantrums of younger, nonpsychiatrically disturbed children (Potegal, Carlson, Margulies, Gutkovitch, & Wall, 2009). It is noteworthy that sadness appeared at two levels of intensity in the older children's data.

The observation that different types of vocalizations are important in distinguishing anger versus sadness during children's temper tantrums suggests that a more detailed analysis of their acoustic characteristics might further clarify differences between tantrum emotion expressions, as well as between different intensities of the same emotion. In adult speech, several studies support the acoustic differentiation of different emotions. Banse and Scherer (1996) reported that adult actors' portrayals of anger (both mild and severe) versus sadness (both mild and severe) differed on several acoustic features (see Banse & Scherer, 1996, Table 6). Sadness was associated with longer vocalizations and vocalizations with lower concentrations of energy in the 0–1 kHz range. Sad speech also had less energy overall, as well as somewhat lower fundamental frequencies. Hot anger, in contrast, had a higher fundamental frequency, more energy overall, and a greater concentration of energy above 1,000 Hz. Juslin and Laukka (2003, Table 8), using similar methods, reported acoustic characteristics of angry and sad speech that were consistent with these characterizations. Indeed, studies of infant crying have examined similar acoustic features and found that fundamental frequency, duration, and energy distribution are related to how aroused the infant is presumed to be; that is, how much hunger or pain is being experienced (see Green, Gustafson, & McGhie, 1998; Porter, Porges, & Marshall, 1988).

However, the studies on adult vocal expressions have all used actors to portray the emotions, and there is a compelling need for naturalistic samples of vocal expressions of emotion (see review by Green, Gustafson, & Whitney, 2010). As noted earlier, children's tantrums offer abundant examples of spontaneous emotion-related vocalizations. However, previous coding of tantrum vocalizations has been by parents at home, or staff of a psychiatric inpatient facility, and their judgments were made while trying to manage the tantrums as well as to record them. These concerns raise a series of methodological questions. Can well-trained observers perceptually classify vocalizations reliably using the above mentioned categories (i.e., scream, yell, cry, whine)? If they can, what are the acoustic properties that enable them to do so (i.e., what are the bases of their perceptual classifications)? Do the acoustic properties that distinguish anger from sadness vocalizations during children's temper tantrums support the distinction between these emotions and/or provide clues about how they differ? Finally, do different types of vocalizations suggest different intensities of anger or sadness?

To answer these questions, we used new technology to collect the high-quality audio recordings necessary for acoustic analysis. To the best of our knowledge, this is the first study to analyze high-fidelity audio recordings of emotion-related vocalizations in a naturalistic situation. Our goal was to define the acoustic charac-

teristics of tantrum vocalizations and to test the following specific hypotheses:

1. The majority of tantrum vocalizations can be classified as scream, yell, cry, and whine by trained observers.

2. There is a basis for these perceptual classifications in specific acoustic characteristics that are similar to those used in previous studies of infants' and adults' vocal expressions of emotions.

3. Vocal expressions of anger and sadness can be further divided into vocalizations representing different emotion intensities. This proposal can be sorted into four subhypotheses:

(A) Among the vocalization types, scream and yell are acoustically most similar to each other

(B) Cry and whine are acoustically most similar to each other

(C) Referring to previous characterization of adult emotional speech, yell, and scream are more anger-like, whine and cry are more sadness-like

(D) Within their respective emotion categories, scream represents a higher intensity of anger than yell, cry represents a higher intensity of sadness than whine.

Method

Participants

Participants were thirteen 2- and 3-year-old children ($M_{age} = 31$ months, $SD = 6$ months) whose families lived in the Minneapolis-St. Paul, MN, area. Seven were boys. Two children were Hispanic, and the others were Caucasian. More than 80% of fathers and 90% of mothers had Bachelors or Masters degrees. Median family income was $75,000 to $85,000. Children were recruited from the Infant Participant Pool of volunteers maintained by University of Minnesota's Institute for Child Development by a 3-step, institutional review board-approved consent process: (a) After being contacted by a member of the research team, parents who consented to a telephone interview reported the frequency and average duration of their child's tantrums, (b) those with frequencies ≥3–6/week and durations ≥1–2 min were recontacted by one of the authors (MP), who explained the purpose of the research, manipulations to be performed, measures to be taken, and compensation, and (c) interested parents were sent a written consent form. Those who consented were enrolled. Exclusions were children with any major developmental, language, or physical health disorder.

Recordings

Audio recordings were made by using professional equipment specially chosen for the project (equipment selection, assembly and testing by Orfield Laboratories, Minneapolis, MN). A Countryman EMW microphone and Lectrosonics MM400A transmitter

---

[1] It is important to note that we are not claiming that the motor and vocal behaviors labeled *anger* and *sadness* during children's temper tantrums are isomorphic with adults' expression of anger and sadness. For example, whining is not a behavior that would typically be considered indicative of sadness in adults. Further, the children's behaviors are manifest during the context of temper tantrums, which have been described as "explosive" episodes (Caspi, Elider, & Bem, 1987) and may be characterized as periods of dysregulation. In the discussion of the data, some comparisons of vocal expressions of anger and sadness in children and adults will be offered.

were contained in a pouch sewn to the front of a toddler sized "onesie." A Lectronics R400A receiver and Marantz PMD-670 digital audio recorder were housed in a cabinet, and the system was calibrated in the laboratory. Because the PMD-670 omits the range-compression typical of consumer products, the system was found to have a flat frequency response up to 4 kHz in the range of 35 to 115 dBA. (The sampling rate was 48kHz, with 16-bit encoding.)

The audio system (and an accompanying video system) were set-up in a room in the child's home during a 2-hr installation and parent training session. On subsequent day(s) and time(s) of her choosing, the parent inserted a battery in the transmitter, dressed the child in the onesie, and activated the system. Recording then took place continuously for up to 4 hr at a time (limited by life of the transmitter battery.). When a tantrum occurred, the parent turned off the system at its end; a research assistant then returned to the home to retrieve the recording.

Coding

The digital files representing each tantrum were coded using Praat, a program designed for acoustic analysis of speech (Boersma, 2001). The basic unit of coding was the child's vocalization, defined as any audible activity of the vocal tract emitted during the course of a single respiratory expiration. (The intake of air during the inspiration phase was typically audible on the recordings and made segmentation of each vocalization relatively simple.) The first phase of the acoustic analysis, then, involved marking the beginning and end of each vocalization on the digital file. There were a total of 2,543 vocalizations during the 24 tantrums.

The second phase of the analysis involved coding each vocalization into one of 5 categories (Table 1), based on the tantrum studies reviewed previously and also in part on previous studies of crying during the first 18 months of life (e.g., Green et al., 1988). The previous tantrum work categorized vocalizations as scream, yell (or shout), cry, or whine, whereas the previous infant cry work distinguished cries and fusses. Careful analysis of the audiotaped tantrums suggested a class of vocalizations different from whining and crying and similar to infant fusses, so this fifth category was added to the intended 4-way classification. Intercoder agreement at this stage was based on two coders independently scoring 5 tan-

Table 1
Definitions of Vocal Expressions Coded During Temper Tantrums

Scream	Typically shrill, loud, and with no verbal content. Usually short and flat melody.
Yell	Typically short in duration, command-like and usually containing some verbal content. Loud, but not as shrill as a scream.
Cry	Relatively loud and effortful, typically with and up and down melody. Breath may be interrupted, as in sobbing. Similar to an infant's cry.
Whine	Typically contains some verbal content with an up and down melody. May also include relatively shrill, monotonous nonverbal vocalization.
Fuss	Typically short, flat or falling melody, relatively quiet and low pitched.

trums for a total of 349 vocalizations. Agreement was high, with 93% agreement and $\kappa = .91$. Digital spectrograms of exemplars of each type are shown in Figure 1.

Of the 2,543 vocalizations, 1,022 co-occurred with parent speech or other ambient noise. Of the remaining high quality vocalizations of the child alone, about 7% were not clearly emotion-related sounds or were otherwise uncodable, and about 7% did not fall clearly into one of the five categories, typically because they exhibited characteristics of more than 1 of the 5 categories. The final corpus of sounds used in the acoustic analysis totaled 1,299 individual vocalizations. The frequency distribution of these sounds across children and tantrum episodes is given in Table 2. It should be noted that Hypothesis 1 was clearly supported; of the vocalizations these children emitted during tantrums, almost 85% could be classified into the five categories in Table 1.

Each vocalization was then analyzed for 11 different acoustic parameters (Table 3). These parameters were chosen because they are similar to those used in analyzing adults' vocal expressions of anger (see review in Green et al., 2010), as well as those used in characterizing infants' cry sounds (e.g., Gustafson, Green, & Tomic, 1984; Green, Jones, & Gustafson, 1987). The fundamental frequency of a vocalization has been especially prominent in both literatures, with higher fundamental frequencies (generally perceived as higher pitch) associated with higher distress, but overall duration, manner of phonation (related to harmonicity), and distribution of energy in different frequency bands have been important in perception of distress as well (Gustafson & Green, 1989).[2]

Analyses

Because the design of the study involved nesting of vocalizations within tantrums, which were nested within participants, multilevel modeling was employed to compare the vocalization categories (i.e., fusses, whines, cries, yells, and screams) on each of the 11 acoustic features. In addition, discriminant function analysis was used to describe the optimal linear combination of acoustic features for separating the vocalization categories.

Results

Differences Among Vocalization Types: Preliminary Comparisons

Simple one-factor analyses of variance (ANOVAs) indicated that the five perceptually based categories of tantrum vocaliza-

---

[2] As fundamental frequency is a challenging measure to obtain for children's vocalizations, F_0 was calculated by Praat in two different ways and the results compared. First, an autocorrelation method was used to obtain a pitch plot for the entire sound. The pitch floor and ceiling were set to 200 and 700 Hz, respectively, with 15 possible candidates allowed (see algorithm description in Boersma, 1993). Second, 10 equally spaced points in each cry were selected and a similar algorithm was used to calculate the F_0 for a short window surrounding each point, again with 200 and 700 Hz set as pitch floors and ceilings. The average of these 10 F_0 values was then computed. The resulting mean F_0 values (one from the F_0 plot and one from the 10 equally spaced points) for each sound was highly correlated ($r = .98$), although there was more missing data for the second method. Occasional comparison of automatically extracted F_0 values with spectrograms and short-term FFT plots indicated that the algorithms were giving reasonable values.
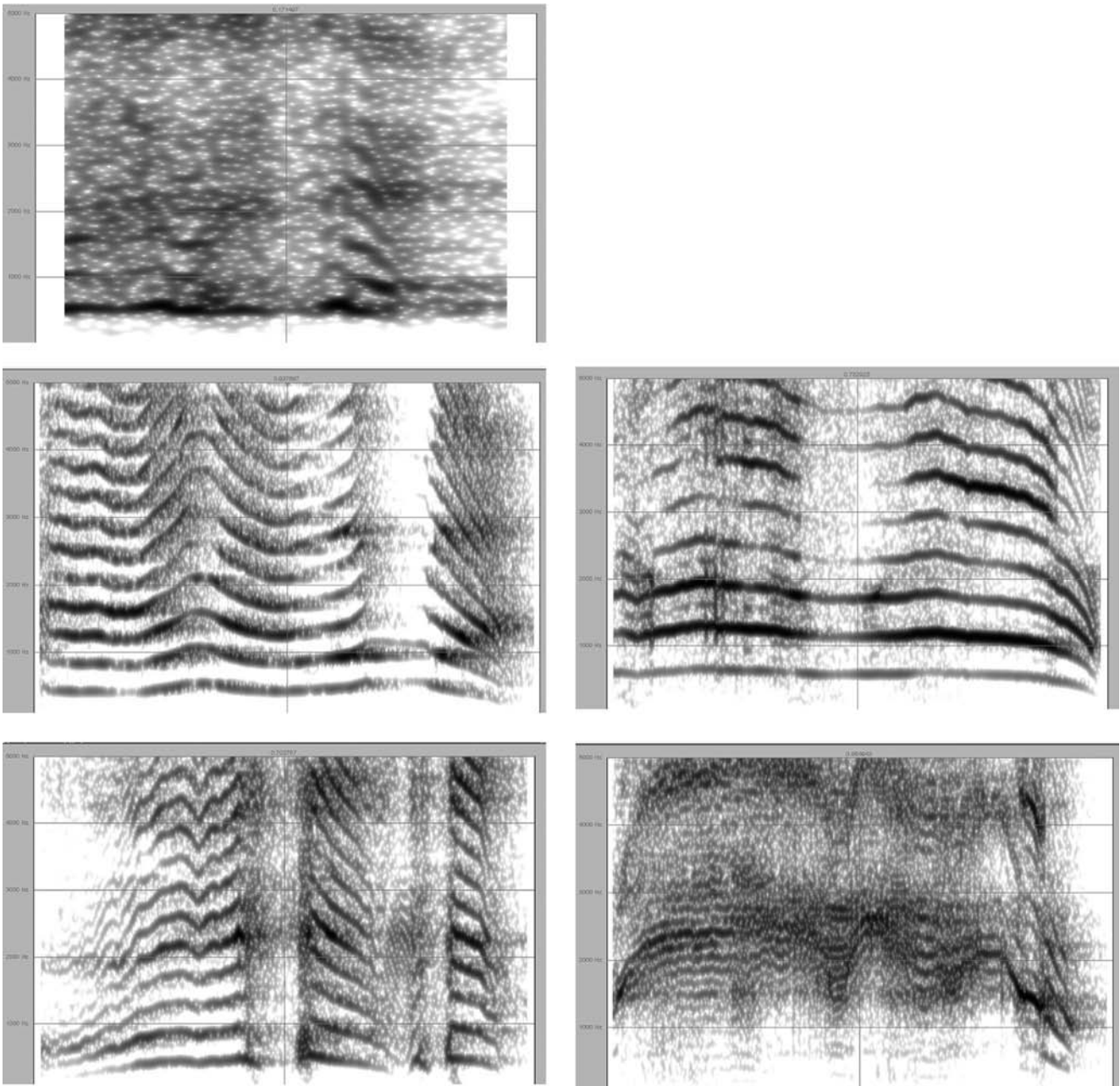
Figure 1. Sample spectrograms of fuss, whine, and cry, (left column, reading down) and yell and scream (right column, reading down). Duration is on the X axis and frequency is on the Y-axis (from 0 to 5,000 Hz with grid lines every 1,000 Hz). (Duration has been equated in these 5 plots but differs for each type of sound.) Darkness of lines indicated more energy in that frequency range. See text for description of acoustic characteristics of each type of vocalization.

tions were significantly different in mean values for every one of the acoustic features. Fusses were shortest in duration (Table 4), with the least overall energy, lowest fundamental frequency and peak frequency, and least energy above 1,000 Hz (frequency bands 3, 4, and 5). The distribution of energy for whines was also concentrated below 1,000 Hz, but whines were longer in duration with slightly higher fundamental frequencies and peak frequencies. Cries were the longest sounds ($M = 1.87$ s),

with greater structure in the harmonics, high fundamental frequencies ($M = 397.43$) and more energy in the 500- to 1,500-Hz range. Screams contained the most energy, the highest peak frequency ($M = 1,297$ Hz), the largest variation in fundamental frequency, and the most energy in the 1,500- to 2,000-Hz frequency band (band 4).

Clearly, the perceptual categories of tantrum vocalizations have many acoustic bases, supporting Hypothesis 2.

Multilevel Modeling

A more appropriate analysis of differences across the vocalization categories, however, takes into account the nested structure of the data (Table 5). Multilevel modeling proceeded by fitting two models for each of the 11 acoustic features. First, a base model was fit that predicted the acoustic feature of each vocalization (level 1) from parameters reflecting the mean for each participant (the fully independent unit in the study, level 3) and for each tantrum nested within participant (level 2). A second model was then fit that added a level one predictor (i.e., vocalization category) to the base model. The significance of the change in fit was evaluated for each acoustic feature and was significant in every case (Table 5). The intraclass correlations for the acoustic features ranged widely, from .03 to .21, indicating that most acoustic features showed a moderate proportion of variance attributable to infants. Overall, there was striking agreement between the standard ANOVAs (Table 4) and the more appropriate multilevel modeling of acoustic features (Table 5).

Multivariate Analyses

To supplement the univariate analyses above, take into account covariance there was among acoustic parameters, and evaluate Hypothesis 3, two discriminant function analyses were employed to determine the number and nature of combinations of acoustic features that separated vocalization categories. For these analyses, harmonicity was dropped because this it was undefined for 157 of the vocalizations Two sounds were missing the measure of variability of the F_0, and these were eliminated from the discriminant analysis, leaving a total of 1,297 vocalizations.

First, the presumptive anger and sadness vocalizations were combined to perform a 2-group discriminant analysis. Scream and yell were combined and compared with cry, whine, and fuss. In this analysis, the single discriminant function was highly significant, canonical $r = .73$, and Wilks' $\lambda = .466$, $p < .001$. The highest loading variable on the standardized discriminant function was .725 for *energy*, and the other high loadings in absolute value ($-.354, -.443$) were for the *proportion of energy in the two lowest frequency bands*, 100 to 500 Hz, and 500 to 1,000 Hz. Thus, yells and screams had much more energy overall and the fusses, whines and cries had proportionately more energy in the lower frequency ranges. Figure 2 shows discriminant scores for the high versus low intensity sounds, and it is evident that there was very little overlap.

The prediction of group membership was excellent, with 92% of the sounds correctly classified as one or the other group (98% of

Table 2
Frequency Distribution of Vocalization Categories Used for Acoustic Analysis

Tantrum number	Fuss	Whine	Cry	Yell	Scream
1	159	357	115	15	9
2	64	233	71	99	63
3	22	25	64	00	03
Total	245	615	250	114	75

Note. Seven children were recorded for two tantrums each, four for one tantrum, and two children contributed three tantrums.

Table 3
Definitions of Acoustic Measures

Duration	Time from beginning to end of each expiratory segment (in ms).
Fundamental frequency (F_0)	Mean and *SD* of the extracted pitch trace for the entire vocalization using an autocorrelation estimation method. Pitch extraction range set at 200–700 Hz.
Energy	Sum of the squared amplitudes of the time sampled vocalization.
Harmonicity	The degree of acoustic periodicity, expressed in dB.
Peak frequency	The frequency in the Long Term Average Spectrum with the greatest amplitude (in Hz).
Spectral energy	The energy in selected frequency bands from the overall spectrum of the vocalizations. Expressed as a proportion of the total energy. Bands: 100–500 Hz, 500–1,000 Hz, 1,000–1,500 Hz, 1,500–2,000 Hz, and 2,000–2,500 Hz.

Note. Further description of these measures is available from the authors or from the manual available at www.praat.org.

fuss, whine, and cry sounds were correctly classified, along with 65% of yell and scream sounds). All other combinations of vocalizations into 2 groups (e.g., scream and cry vs. yell, whine and fuss) yielded substantially fewer correct classifications on discriminant analysis, ranging from 64% to 78%. This comparison suggests that yell and scream versus fuss, whine and cry is the most appropriate categorical grouping.[3]

A second discriminant analysis was performed using all 5 vocalization categories. Here, all four discriminant functions were statistically significant and accounted for 81%, 13%, 4%, and 2% of the discriminating power, respectively (Table 6). The canonical correlations for the functions were .814, .488, .312, and .227. Examination of the standardized coefficients showed that vocalizations that were relatively low in total energy (that is, were quiet) and had their energy concentrated in bands below 1,000 Hz had high scores on Function 1. Indeed, fusses and whines generally fit this profile and have high group means on Function 1 (Figure 3).

Function 2 separated vocalizations long in duration and with relatively high energy in the 500- to 1,000-Hz band from the others. The clear separation here was of Scream and Fuss (i.e., relatively short, little energy in the 500- to 1,000-Hz range) versus Cry and Yell. Function 3 contrasted sounds with relatively large amounts of energy less than 1500 Hz (i.e., frequency bands 1, 2, and 3) from other sounds. Yells scored highly on this third function. However, the 5-group discriminant analysis (which used prior probabilities determined by the relative frequency of the sound types) correctly classified only 806 of the 1297 sounds (62%) according to their acoustic characteristics. This result would be expected if differences *within* the presumptive categorical classi-

---

[3] One of reviewers suggested a comparison of screams versus cries because of the potential confounding of verbal content (often found in yells and whines) and acoustic features. Screams ($n = 75$) and Cries ($n = 250$) were easily discriminated by these 11 acoustic features, canonical $r = .845$, and 96% classification accuracy. Energy, mean F0, and proportion of energy from 500 to 1,000 Hz were the primary contributors to the standardized canonical discriminant function.

Table 4

Means, SDs, and Analysis of Variance (ANOVA) Results for Five Sound Types

Sound type	Fuss	Whine	Cry	Yell	Scream	$F(4, 1294)$	η_p^2
Duration	0.46^{wcys} (.39)	1.38^{fcys} (1.00)	1.87^{fw} (1.22)	1.64^{fw} (.84)	1.66^{fw} (.65)	83.17***	0.21
Energy	0.000^{cys} (.001)	0.002^{cys} (.004)	0.006^{fwys} (.007)	0.017^{fwcs} (.015)	0.034^{fwcy} (.021)	355.28***	0.52
Harmonicity	5.89^{wcys} (9.04)	12.50^{fcys} (10.01)	13.44^{fws} (6.83)	14.51^{fws} (4.39)	6.66^{fwcy} (4.16)	39.66***	0.22
Peak Freq	609.59^{wcys} (220.19)	693.74^{fcys} (233.70)	858.00^{fwys} (291.40)	1028.95^{fwcs} (325.46)	1296.67^{fwcy} (449.12)	139.39***	0.30
Mean F$_0$	361.92^{cy} (86.59)	376.10^{cy} (55.28)	397.43^{fws} (51.97)	426.22^{fwcs} (51.23)	377.60^{cy} (65.49)	26.18***	0.07
SD of F$_0$	50.75^{wcys} (35.23)	61.40^{fcys} (28.91)	72.44^{fws} (26.42)	79.11^{fws} (27.33)	102.16^{fwcy} (30.57)	55.91***	0.15
Band1	0.35^{wcys} (.28)	0.28^{fcys} (.27)	0.12^{fw} (.13)	0.08^{fw} (.09)	0.02^{fwc} (.03)	66.90***	0.17
Band2	0.56^{ys} (.26)	0.57^{ys} (.26)	0.61^{ys} (.23)	0.40^{fwcs} (.24)	0.22^{fwcy} (.22)	48.10***	0.13
Band3	0.05^{wcys} (.09)	0.09^{fcys} (.14)	0.13^{fwys} (.16)	0.35^{fwc} (.22)	0.35^{fwc} (.18)	141.20***	0.30
Band4	0.02^{cys} (.04)	0.03^{ys} (.05)	0.05^{fys} (.07)	0.08^{fwcs} (.10)	0.20^{fwcy} (.17)	116.55***	0.27
Band5	0.02^{cys} (.03)	0.02^{cys} (.03)	0.06^{fws} (.07)	0.06^{fws} (.06)	0.11^{fwcy} (.10)	78.60***	0.20

Note. See Table 3 for basic definitions of these acoustic features. Superscripts next to a mean indicate which other groups are significantly different from the given mean.
*** $p < .001$.

fications noted above were matters of intensity, in which there might well be notable overlaps. Figure 4 shows exactly this situation of substantial overlap among fuss, whine and cry in the discriminant plane.

Taken together, the 2 discriminant analyses support hypotheses 3A and B: That is, there is better than 90% correct prediction of the presumptive anger (scream, yell) versus presumptive sadness (cry, whine, and fuss) categories. With regard to Hypothesis 3D, the five vocalizations are distributed across Function 1 in the order scream > yell > cry > whine > fuss, meaning that these vocalizations had progressively less energy overall and less energy above 1,000 Hz, suggesting that within-category intensity may be related to energy at less than 1,500 Hz.

A final statistical analysis compared the "melody" or change in fundamental frequency over time for each vocalization type (Figure 5). To examine relative change in F$_0$, the mean F$_0$ was subtracted from the F$_0$ value at 10 equally spaced segments within each vocalization. In this transformation, shown in Figure 4, yell and scream each had an early, sharp peak while cry and whine

showed later, much more rounded peaks. A 5 (vocalization category) × 10 (position in the expiratory segment) mixed ANOVA was computed, with a special contrast to compare scream plus yell with the average of cry, whine, and fuss. (The contrast coefficients were .5, .5, −.333, −.333, and −.333 for scream, yell, cry, whine, and fuss, respectively.) The interaction of vocalization category and time into sound was significant, $F(36, 10566) = 3.582$, $p < .001$, and the contrast of (scream + yell) versus (cry + whine + fuss) was significant, $p < .001$. As in the 5 group discriminant analysis, then, scream and yell were similar to each other but different from cry, whine, and fuss. The qualitative similarities and differences in the shapes of the curves and the quantitative results of the ANOVA support the putative categorical distinction between yell and scream versus whine and cry.

Finally, to place these sounds in the context of other negative vocalizations, per Hypothesis 3C, we compared F$_0$ values to published data on infant crying. Many publications exist on the acoustic features of newborn cry sounds, but there are relatively few studies of acoustic features from later ages. Wermke, Mende,

Table 5

Multilevel Model Results

Estimated marginal means	Fuss	Whine	Cry	Yell	Scream	-2LL Change	ICC	F
Duration	0.57^{wcys}	1.31^{fcys}	2.04^{fw}	1.98^{fw}	2.00^{fw}	337.56***	0.19	108.82***
Energy	0.00^{cys}	0.00^{cys}	0.01^{fwys}	0.01^{fwcs}	0.03^{fwcy}	642.50***	0.08	177.93***
Harm	6.95^{wcy}	12.85^{fs}	14.56^{fs}	15.60^{fs}	7.68^{wcy}	137.07***	0.05	116.22***
Peak frequency[a]	613.95^{wcys}	701.24^{fcys}	849.43^{fwys}	989.60^{fwcs}	1252.60^{fwcy}	338.19***	0.03	549.86***
Mean F0	386.38^{cy}	393.17^{cy}	413.64^{fws}	428.00^{fws}	379.17^{cy}	68.46***	0.21	712.23***
SD F0[a]	55.05^{wcys}	61.94^{fcys}	74.43^{fws}	76.99^{fws}	99.17^{fwcy}	147.94***	0.06	186.20***
Band1[a]	0.36^{wcys}	0.29^{fcys}	0.14^{fw}	0.13^{fw}	0.08^{fw}	173.54***	0.06	73.99***
Band2[a]	0.54^{ys}	0.53cys	0.59^{wys}	0.36^{fwcs}	0.18^{fwcy}	161.36***	0.10	128.08***
Band3	0.05^{wcys}	0.10^{fcys}	0.14f^{wys}	0.29^{fwc}	0.30^{fwc}	244.38***	0.12	70.11***
Band4	0.03^{wcys}	0.04^{fys}	0.05^{fys}	0.10^{fwcs}	0.21^{fwcy}	368.64***	0.15	94.11***
Band5	0.02^{cys}	0.02^{cys}	0.05^{fws}	0.06^{fws}	0.11^{fwcy}	199.26***	0.12	51.52***

Note. -2LL = -2loglikelihood or deviance; ICC = Intraclass Correlation Coefficient.
[a] Because of small sample sizes at levels 2 and 3 and distributional properties of these variables, only 2 tantrums could be nested at level 2 under the level 3 individual participants.
Superscripts next to a mean indicate which other groups are significantly different from the given mean.
*** $p < .001$.

Table 6
Standardized Canonical Discriminant Function Coefficients

	1	2	3	4
Duration	.050	**.823**	−.259	**.451**
Total energy	**−.741**	−.338	.088	−.100
Peak frequency	−.061	.041	.034	.214
F_0 M	.170	.331	.421	−.266
F_0 SD	−.235	.086	−.278	.100
Energy band1	**1.091**	.101	**1.993**	**.901**
Energy band2	**1.199**	**.449**	**2.059**	.395
Energy band3	.453	.372	**1.920**	**.502**
Energy band4	−.059	−.160	.308	**.525**
Energy band5	.135	.272	.279	**−.610**

Note. Entries in bold are the coefficients used to interpret the 4 discriminant functions (see text for details). These 4 combinations of acoustic features were able to correctly classify 38% of fusses, 87% of whines, 26% of cries, 46% of yells, and 79% of screams.

Manfredi, and Bruscaglioni (2002) reported longitudinal data on 6 infants showing little change in F_0 during the first 6 months, with values ranging from 377 to 588 Hz. Baeck & Nogueira de Souza (2007) reported a range of 385 to 410 Hz from 30 infants in the first 6 months of life.

Perhaps the most relevant data come from Rothgänger (2003), who reported an average F_0 of 463 to 480 Hz for the cry sounds of 15 infants across the entire first 12 months, which was considerably higher than the 342 Hz for babbling sounds during the same time period. However, the present value of almost 400 Hz for the F_0 of the cry sounds (Table 4) is similar to the values for cry sounds reported by all 3 studies.

Discussion

Previous empirical research on young children's temper tantrums has yielded a model proposing two sets of motor and vocal behaviors, one that expresses anger and the other, sadness (Potegal et al., 2009; Potegal & Davidson, 2003). Within this model, four distinct types of tantrum vocalizations were identified for younger children, with whine and cry associated with a factor labeled sadness while scream and yell (previously termed shout) were associated with a factor labeled anger. However, the previous research on tantrum vocalizations was based on parent report, not on actual audio recordings. In the present work, we evaluated the previously proposed distinctions among tantrum vocalizations using novel data collection methods in naturalistic situations. We found that the majority of tantrum vocalizations could be reliably perceptually classified as scream, yell, cry, and whine. We also identified a fifth, low-energy type that we called fuss based on previous research on infant crying. Both the perceptual and the statistical reliability of these identifications were high. Together, the five types accounted for almost 85% of tantrum vocalizations.

Furthermore, young children's tantrum vocalizations identified perceptually are readily separable by their acoustic features. These acoustic features overlap considerably with those previously identified in analyses of both adult and infant expressions of anger and sadness. Banse and Scherer (1996), and others, have reported that adults' angry speech (both "hot" and "cold" angry speech) is relatively short, has more energy overall, higher fundamental frequencies, and more energy in high frequencies. Results of our analysis of tantrum vocalizations are in substantial agreement with the findings on adult emotion speech. Screams and yells had more overall energy, higher fundamental frequency (especially yells), and more energy above 1,000 Hz. The only contrary finding is in the duration of the sound; yell and screams were relatively longer compared with whines, although a bit shorter than cries.

One difficulty in making comparisons between the vocalization in the present study and those used in adult studies is the source of the sounds. Adult studies have invariably used staged emotion speech, sometimes even the same phrase spoken in anger versus
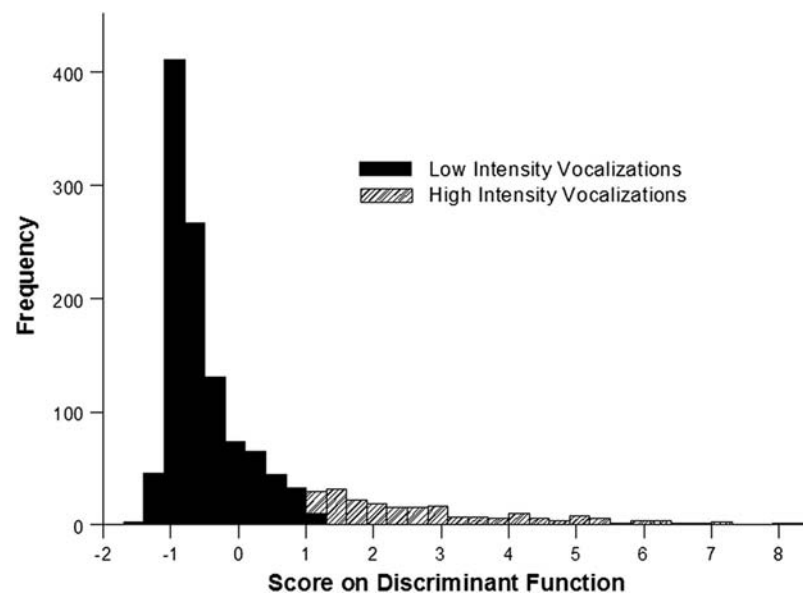

Figure 2. Distribution of scores for low intensity (fuss and whine) and high intensity sounds (cry, yell, and scream) on the discriminant function. Note that there is almost no overlap in scores, and the classification accuracy is 93%. See text for a description of the variables loading on the discriminant function.
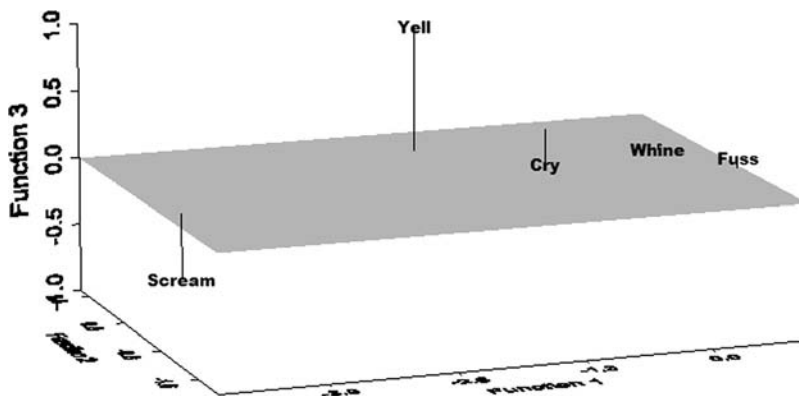
Figure 3. Group centroids for fuss, whine, cry, yell, and scream in discriminant 3-space. See text for descriptions of the 3 discriminant functions.

sadness. Thus, comparison with the present, naturalistically recorded vocalizations, especially on duration measures, is problematic. In addition to the staged speech, one other factor in our naturalistic recordings it that expressions of sadness during preschoolers' temper tantrums are influenced by juxtaposed expressions of anger. This potential carryover effect would not be present in adults' portrayals of sadness in the experimental literature on vocal expressions of emotions.

Fuss sounds had the least overall energy, the lowest peak frequency, and the least energy above 1 kHz. These low-energy, low-frequency distress sounds have been studied extensively during the period of infancy as part of the literature on crying. There, fusses are recognized as briefer, less intense negative sounds than full blown wails (Gustafson et al., 1984). Given that fusses are located near to whines in discriminant space (Figures 3 and 4), it is perhaps understandable that neither sound is especially salient to parents during tantrums. In fact, although Potegal and Davidson

(2003) previously coded whining as an indicator of distress, they did not find justification for using it alone as a marker for the onset of a temper tantrum. Whether future studies should continue to separate whine and fuss would depend on the purpose of the study, because both seem to be indicators of low-intensity distress.

One of our hypotheses was that a pattern of similarities and differences among acoustic characteristics would emerge such that the several types of tantrum vocalizations could be effectively classified into two groups; furthermore, these groups would be consistent with the emotions of anger and sadness. We found that the best discrimination in the classification analysis, amounting to correct identification of over 90% of cases, grouped scream and yell together and cry, whine, and fuss together. Complementing this result was a pattern of qualitative similarities and differences in "melody." Yell and scream both had early, sharp peaks in their fundamental frequency within vocalizations over time while cry and whine showed later, much more rounded peaks (Figure 5.) The
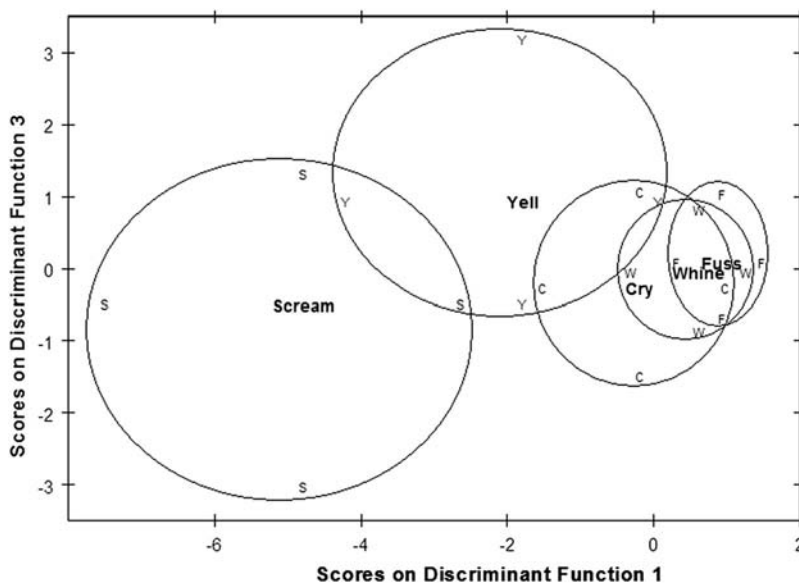

Figure 4. Distributions of fuss, whine, cry, yell, and scream (10th to 90th percentiles encircled) in the discriminant plane of Function 1 and Function 3.
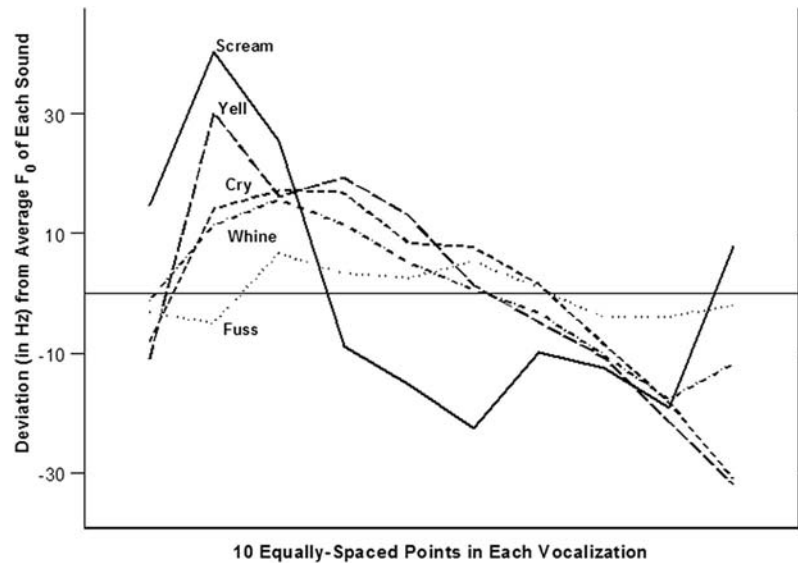
Figure 5. Melody contours for each type of vocalization. F_0 was measured at 10 equally spaced points in each sound, and each value in each sound was subtracted from the mean F_0 for that sound.

acoustic distinction between yell/scream versus fuss/whine/cry parallels the distinction between the prosody of angry versus sad utterances by adults.

There was also converging evidence for graded acoustic differences within emotion groupings. Both in terms of energy distributions across frequency bands and temporal contours of F_0 (Figure 5), screams could be viewed as more intense than yells. Similarly, fuss, whine and cry appeared to be ordered by increasing intensity according to both energy distributions across frequency bands and F_0 temporal contours. The gradation of intensity within a class of vocalizations is not unlike what has been proposed for the production of infant cries (Gustafson, Wood, & Green, 2000); in that literature, cry sounds are generally perceived as more intense than fuss sounds but both are classified as "negative" vocalizations.

One theoretical controversy in the area of emotion expressions centers on whether emotional expressions represent discrete or continuous processes. Some argue vigorously for separable physiological states and behavioral (especially facial) manifestations of happiness, sadness, anger, disgust, pride, and so on. Others argue that two underlying dimensions of arousal and valence are responsible for the various expressions of emotion (Barrett, 2006; Russell, Bachorowski, & Fernández-Dols, 2003).

Although we have presented arguments favoring the categorical, discrete emotions view, our empirical base is *continuous* acoustic features that were used to describe *categories* of sounds reliably coded by trained listeners. It is worth noting here that categorical perception is, in fact, a rather ubiquitous process across stimulus domains (including spoken language) and indeed across species (see review by Hauser, 2001). More relevant data for the discrete versus continuous arguments in the emotion expression literature might have to come from combining perceptual, behavioral, and functional neural architecture data.

Of course, there are several limitations of this research. First, the sample is a high socioeconomic status group, which was prese-

lected for higher tantrum frequency and duration. However, the vocalizations and behaviors during these tantrums are similar to those reported as far back as Goodenough (1931), so it seems unlikely that these tantrums are different in quality. Second, the sample size is relatively small. This limitation is because of the dense, naturalistic behavior samples required for acoustic analysis. Finally, the acoustic features themselves are a limited set. Although the features are similar to those studies in adult emotion expressions as well as in infant distress studies, they do not contain many features used in, for example, language processing. Future studies might expand this set and focus on formant transitions or other features adults use in processing language.

Although the present study did not analyze the behaviors of the children, the co-occurrence of vocal expressions with behavioral indicators, such as kicking, hitting, and dropping down, should be one of the next steps in this research. Whether behavioral indicators, such as hit or kick, and vocal indicators, such as scream, can be substituted for one another as manifestations of high anger, or whether these expressions necessarily co-occur or synergize each other, is unknown. These questions are presently being studied in our laboratories, and the answers will bear on the important question of how young children express and regulate intense emotions. Temper tantrums are ideal, naturally occurring, episodes to examine issues of emotion reactivity and regulation.

References

Baeck, H. E., & Nogueira de Souza, M. (2007). Longitudinal study of the fundamental frequency of hunger cries along the first 6 months of healthy babies. *Journal of Voice, 21,* 551–559.

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

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science, 1,* 28–58.

Boersma, P. (1993). Accurate short-term analysis of the fundamental

frequency and the harmonics-to-noise ratio of a sampled sounds. *Proceedings of the Institute of Phonetic Sciences, 17,* 97–110.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Caspi, A., Elder G. H., Jr., & Bem D. J. (1987). Moving against the world: Life course patterns of explosive children. *Developmental Psychology, 23,* 308–313.

Goodenough, F. L. (1931). *Anger in young children.* Minneapolis, MN: University of Minnesota Press.

Green, J. A., Gustafson, G. E., & McGhie, A. C. (1998). Changes in infants' cries as a function of time in a cry bout. *Child Development, 69,* 271–279.

Green, J. A., Gustafson, G. E., & Whitney, P. (2010). Vocal expressions of anger. In M. Potegal, G. Stimmler, & C. Spielberger (Eds.), *International handbook of anger.* Secaucus, NJ: Springer.

Green, J. A., Jones, L. E., & Gustafson, G. E. (1987). Perception of cries by parents and nonparents: Relationship to cry acoustics. *Developmental Psychology, 23,* 370–382.

Gustafson, G. E., & Green, J. A. (1989). On the importance of fundamental frequency and other acoustic features in cry perception and infant development. *Child Development, 60,* 772–780.

Gustafson, G. E., Green, J. A., & Tomic, T. (1984). Acoustic correlates of individuality in the cries of human infants. *Developmental Psychobiology, 17,* 311–324.

Gustafson, G. E., Wood, R. M., & Green, J. A. (2000). Can parents "tell" what their infants' cries mean? In R. G. Barr, B. Hopkins, & J. A. Green (Eds.), *Crying as a sign, a symptom, and a signal: Clinical, emotional, and developmental aspects of infant and toddler crying* (pp. 8–22). London, England: MacKeith Press.

Hauser, M. D. (2001). What's so special about speech? In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of J. Mehler* (pp. 417–434). Cambridge, MA: MIT Press.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814.

Porter, F. L., Porges, S. W., & Marshall, R. E. (1988). Newborn pain cries and vagal tone: Parallel changes in response to circumcision. *Child Development, 59,* 495–505.

Potegal, M., Carlson, G., Margulies, D., Gutkovitch, Z., & Wall, M. (2009). Rages or temper tantrums? The behavioral organization, temporal characteristics, and clinical significance of angry-agitated outbursts in child psychiatry inpatients. *Child Psychiatry and Human Development, 40,* 621–636.

Potegal, M., & Davidson, R. J. (2003). Temper tantrums in young children: 1. behavioral composition. *Developmental and Behavioral Pediatrics, 24,* 140–147.

Potegal, M., Kosorok, M. R., & Davidson, R. J. (2003). Temper tantrums in young children: 2. Tantrum duration and temporal organization. *Developmental and Behavioral Pediatrics, 24,* 148–154.

Rothgänger, H. (2003). Analysis of the sounds of the child in the first year of age and a comparison to the language. *Early Human Development, 75,* 55–69.

Russell, J. A., Bachorowski, J., & Fernández-Dols, J. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54,* 329–349.

Stevenson, J., & Goodman, R. (2001). Association between behaviour at age 3 years and adult criminality. *British Journal of Psychiatry, 179,* 197–202.

Stoolmiller, M. (2001). Synergistic interaction of child manageability problems and parent-discipline tactics in predicting future growth in externalizing behavior for boys. *Developmental Psychology, 37,* 814–825.

Wermke, K., Mende, W., Manfredi, C., & Bruscaglioni, P. (2002). Developmental aspects of infant's cry melody and formants. *Medical Engineering and Physics, 24,* 501–514.