APPENDIX:

Background:

Infant crying activates distinct brain regions in both mothers[1, 2] and fathers[3], showing that parents are wired to be responsive to a baby's needs. The ChatterBaby project was originally designed as a baby monitor for Deaf parents, using machine learning and signal processing algorithms to identify whether and why a baby is crying. These algorithms are embedded into free smart-phone apps to translate a baby's cry into three primary states described below, provided for parents and caretakers (ChatterBaby.org, Figure 1). The app also functions as a data collection device, which will map cry features to developmental disabilities such as autism spectrum disorder which affects up to 1 in 25 children who are Deaf. [4]

Once subjects download the app, they consent to be in the IRB-approved research study and are followed for six years. Developmental questionnaires and autism screening instruments are administered by email each year on the child's birthday, using validated instruments such as the CDC 12-month Milestones [5], the Quantitative Checklist for Autism in Toddlers (Q-CHAT) [6], and the Childhood Asperger Syndrome Test or the Social and Communication Development Questionnaire (CAST) [7]. The CAST is a screening questionnaire for autism spectrum conditions in children ages 4-11 [7]. RedCap is used for administering all screening instruments, which permits automation of all processes. Beginning at age 2, risk scores for ASD are calculated using the Q-CHAT and CAST scoring systems, and these scores are provided to the participant.

To collect infant cry data for later research, ChatterBaby performs a free service to parents by predicting the reason of a baby's cry, using data previously labeled by a "mom-panel" or obtained during painful procedures (vaccinations, ear-piercings). Most estimates of inter-rater reliability of infant scales are high [8-10] with some divergence [11]. While some studies showed agreement[12], other studies showed poor agreement across these scales in measurements [13, 14], suggesting that both clinical factors and the choice of scale may strongly influence the magnitude and the reliability of these pain measurements. An objective clinical marker may provide a benchmark for comparing the results of these instruments, with acoustic analyses providing small-sample studies demonstrating feasibility of such a vocal pain metric.

Supplemental Methods:

This study was conducted according to and approved by the UCLA Institutional Review Board. Because audio was recorded in the home environment, infants were in a variety of settings while being recorded, with naturally occurring background noise (adult voices, etc.) using different recording devices (e.g., cell-phones). All recordings were converted into 5 second raw pulse-code modulation (PCM), sampling rate 16000 Hz, quantization of 16 bits, and were part of a larger infant vocalization database discussed further in the Appendix. All cry audio recordings were 5 seconds long with the exception of 21 separation anxiety cries which were <3 seconds in length collected by the Oxford Vocal (OxVoc) Sounds database[15].

Modeling:

Supra-segmental (utterance-level) acoustic features were extracted from 5 second cry clips [16-18] using IS13_ComParE.conf in OpenSmile [19]. This feature set has been used previously to assess pathological vocal patterns in neurological disorders such as Parkinson's Disease [20] and Amyotrophic Lateral Sclerosis [21], and contains 6,373 static features computed from various functionals over low-level descriptor (LLD) contours. The pre-defined acoustic features computed various statistics of jitter, shimmer, MFCCs, and spectral features described in more depth in the documentation [19]. A probabilistic random forests classifier was used to predict the category of a cry (fussy, hungry, pain) given its acoustic features using default parameter settings in R (500 trees, 1/3 of features sampled with replacement as possible predictors to construct individual trees). Parameter tuning was avoided to obtain an unbiased estimate of testing accuracy. The random forests model (package randomForest in R) uses resampling of both features and data to create decision tree models, where voting over many decision trees determines an observation's final class [22]. The random forests out-of-sample classification accuracy, analogous to the cross-validation error, was computed to estimate the testing accuracy of the algorithm on new data. Secondary analyses using parameter tuning showed marginal improvement in the random Forests algorithm, and similar results were realized using other untuned machine learning algorithms (support vector machines) using traditional cross-validation approaches.

External Validation on Secondary Cries:

The random forests algorithm provides out-of-sample predictions for observations (out-of-bag estimates), analogous to a cross-validation procedure in other machine learning algorithms. This provides an estimate of how, for example, a new fussy cry performs when using the 3-way classification algorithm. These out-of-sample fussy and hungry states provided control conditions for colic, to assess whether pain levels are increased for infants who are described by caregivers as having colic.

The feature importance values of the algorithm provide a measure of the predictive ability of each feature, and decreased sharply after roughly 100 features (Figure 2). Using only the 200 most predictive features, the algorithm was retrained on the primary cries and tested on the colic cries, with roughly 6 colic cries obtained from the same cry episode per child. The average pain probability from colic cries were compared with the out-of-sample pain predictions from fussy, hungry, and pain to test the hypothesis that colic cries were more closely associated with pain than the hungry or fussy states.

Analyses of Reduced Feature Sets: the value of extensive feature extraction

Pain cry acoustics literature has relied heavily on a limited number of acoustic descriptors such as pitch and energy. We assessed the value of the additional acoustic features by extracting a reduced feature set containing 62 acoustic features in the OpenSmile GeMAPSv01a.conf file by Eyben and colleagues[23] . The F0 (pitch) values, extracted with GeMaps in OpenSmile, were tested to assess whether the pitch of colic and pain were similar, and whether the pitch of these states were different from the fussy and hungry states.

Supplemental Results:

The primary cry algorithm achieved overall accuracy in classifying among the three states as 71.5%, with the confusion matrix shown in the Appendix, Table 3. The primary cry algorithm, trained as a multivariate classifier, was then treated as a binary classifier for obtaining Pain accuracy rates by pooling the Fussy and Hungry predictions as a "No Pain" category. The predictive accuracies for painful cries are shown in Table 1: sensitivity/recall of .91 (95% CI = .876, .937), specificity = .68 (95% CI = .628, .727), positive predictive value = .75, negative predictive value = .87. The prevalence of Pain was .51, with the algorithm performing significantly above chance (p<0.001). The AUC = .88 is shown in the ROC curve in Figure 3.

Within the longitudinal vaccination recordings from a single child who was not used to create the algorithm, the Fussy/Hungry/Pain algorithm predicted similarly and consistently that the baby was experiencing pain for all six trials (average pain probability = .63, sd =.04). This suggests that the algorithm was not sensitive to aging effects within the age range evaluated (Figure 4, Supplemental Figure 4 for Spectrograms of vaccine cry across age). Five-second audio clips from this child's vaccinations over an 18 month period are available at https://www.youtube.com/watch?v=eu332YZFTkA

Although the random forests algorithm had access to over 6,000 features, many of these features had low importance values and were not useful to discriminate among the different cry states, as shown in Figure 5. The top 10 acoustic features most useful for classification are listed in Table 4 Appendix. The top 200 features were used to train the reduced algorithm for validation on the secondary cry states.

When testing the algorithm on Colic cries, the algorithm demonstrated that the probability of pain was significantly different across the three states (p<0.0001, ANOVA), and the pain levels in colic were significantly greater than the pain levels seen in fussy and hungry states (p<0.001, 2-sample t-test, Bonferroni corrected). As shown in Figure 5, the average pain rating in colic was 0.73 (sd = 0.21), while the average out-of-sample predicted pain rating for fussy was 0.30 (sd = 0.18), hungry = .38 (sd = 0.19), and pain = 0.67 (sd = 0.20).

Our comparative analyses of the OpenSmile GeMaps feature set found a roughly 1% increase in classification accuracy using the full IS13_ComParE.conf features compared to the reduced feature set, with features unique to the IS13_ComParE feature set outranking those in the restricted feature set. This suggests that the higher-order features may be aiding in the classification accuracy observed.

Previous literature demonstrated an increased pitch (fundamental frequency) in both pain and colic [24] cries compared to fussy and hungry states, which we confirmed here (p<0.05; 2-sample t-test with Bonferroni correction, Supplemental Table 5). The pitch did not significantly differ between colic and pain (p>0.05, 2-sample t-test with Bonferroni correction). However, elevated pitch was not specific to either pain or colic, and was also seen in baby cries from scared and separation anxiety states when compared to samples from older children. Moreover, pitch was not one of the 200 most important features in the predictive algorithm; the most important predictive feature in the algorithm was a summary statistic of spectral flux, measuring how different one audio frame

was from the next, suggesting that the regularity of a cry provided much of the classification power as to the type of cry. This suggests that the acoustic features of pain extend far beyond pitch. Measures of energy and silence did show important variability across cry profiles (Figure 6).

Cry detection algorithm:

The Why-Cry algorithm in the ChatterBaby app is embedded within a primary Cry-detection algorithm trained using n=2,346 audio samples, comparing 1,192 Cries to 1,154 distractor states which were recommended by our Deaf focus group. In addition to fussy, hungry, colic, and pain, cry samples attributable to secondary stimuli were also collected taken during a single episode of crying: scared (n=161, 89 babies) and separation anxiety (n=36, 23 babies). Scared, colic, and separation anxiety labels were provided by the parent or primary caretaker; examples of Scared stimuli included "Daddy singing," "bear," and "loud noise" as labeled by the participant. A total of new 15 separation anxiety cries were collected and combined with 21 separation anxiety cries provided in Parsons et al. [15], which were acquired when the infants were separated from their primary caregiver (mean age = 6.7 months, *sd* = 0.9 months). The Cry states pooled all six cry conditions described above: Hungry, Fussy, Pain, Separation Anxiety, Scared, Colic.

Our Deaf focus group had reported that conventional cry detectors were triggered by non-cry sounds, so the algorithm trained on Cry vs. Distractor States. The Distractor states included baby voices (80%) as well as nuisance noises which included animal sounds (cats, dogs) [15], adult males and females laughing [15], adult females and males speaking, three year old children speaking, doorbells, phones ringing, and electronic baby toys collected manually as shown in Table 1. Approximately 80% of Distractor states were from infant noises (laughing, neutral). Of the baby laugh/neutral/cry sounds collected, 79.1% of these were from unique speakers.

Similar to before, the random forests algorithm predicted "Cry" vs Distractor States. This algorithm achieved overall 85% accuracy, with AUC = .92 as shown in Figure 2. The sensitivity = .82 (95% CI = .80, .84) with specificity = .88 (95% CI = .86, .90), as shown in Table 2 and Figure 7.

Source	n
3 year old female speaking	16
3 year old male speaking	7
adult female laughing	15
adult female speaker	26
adult male laughing	15
adult male speaker	25
Baby neutral (babbling, cooing)	728

Baby laughing	209
Baby toys	72
Cat Meow	15
Dog Bark	15
Doorbell	6
Phone ringing	5

Table 1: Distractor states (n=1154) for the Cry Detection algorithm were collected from a variety of household ambient noise sources which were reported by our Deaf focus group to have generated false positives in conventional cry-detection devices. The Distractor states were compared to 1,192 baby cries to create a Cry Detection algorithm.

Calculated Diagnostic Accuracy Parameters:	
Sample size =	2346
Prevalence =	0.508
Sensitivity =	0.821
Specificity =	0.880
PPV =	0.876
NPV =	0.827
LR + result*=	6.819
LR - result*=	0.203

Table 2: Accuracy values for Cry Detection algorithm. This algorithm predicted baby cries from a variety of distractor states: baby laughing/neutral, animal noises (dog, cat), adults speaking and laughing, 3-year-old speech, and household noises (phones, doorbells). 80% of all distractors were from baby sounds, while the remaining 20% were from the noise sources recommended by our Deaf focus group.

|	Fussy	Hungry	Pain	class.error
Fussy	101	24	46	0.41
Hungry	31	73	63	0.56
Pain	16	17	320	0.09

Table 3: Classification accuracy from 3-way classifier.

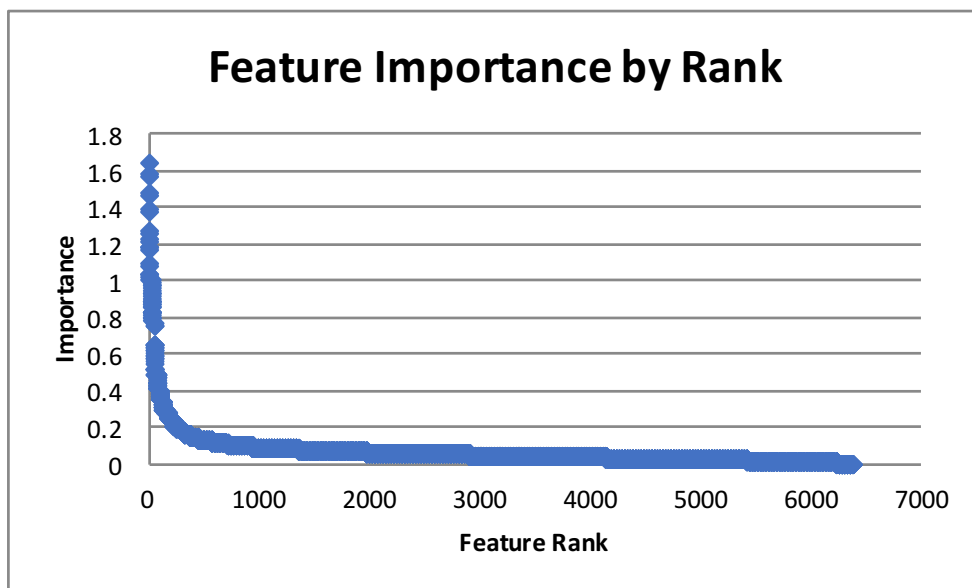| Feature |
| --- |
| pcm_fftMag_spectralFlux_sma_de_quartile3 |
| pcm_fftMag_fband1000.4000_sma_de_quartile3 |
| pcm_fftMag_fband1000.4000_sma_de_iqr1.2 |
| pcm_fftMag_fband1000.4000_sma_de_iqr1.3 |
| pcm_fftMag_spectralFlux_sma_quartile2 |
| pcm_fftMag_fband1000.4000_sma_quartile1 |
| pcm_fftMag_spectralFlux_sma_de_quartile1 |
| pcm_fftMag_spectralHarmonicity_sma_iqr1.2 |
| logHNR_sma_leftctime |
| pcm_fftMag_spectralFlux_sma_quartile1 |

Table 4: Feature Importance of WhyCry algorithm: 10 most predictive acoustic features

Variable	Hungry		Fussy		Colic		Pain	
	mean	sd	mean	sd	mean	sd	mean	sd
F0semitoneFrom27.5Hz_sma3nz_amean	47.37	2.88	46.25	5.81	47.99	3.22	47.67	4.00
F0semitoneFrom27.5Hz_sma3nz_stddevNorm	0.10	0.05	0.14	0.11	0.12	0.07	0.14	0.07
F0semitoneFrom27.5Hz_sma3nz_percentile20.0	44.42	4.22	42.18	8.70	44.33	4.76	43.11	6.54
F0semitoneFrom27.5Hz_sma3nz_percentile50.0	47.84	2.77	46.74	6.70	48.45	3.38	48.56	3.95
F0semitoneFrom27.5Hz_sma3nz_percentile80.0	50.55	3.63	50.72	5.09	52.35	3.39	52.65	3.66
F0semitoneFrom27.5Hz_sma3nz_pctlrange0.2	6.13	4.62	8.55	7.22	8.01	4.68	9.54	5.94
F0semitoneFrom27.5Hz_sma3nz_meanRisingSlope	190.74	157.21	188.92	159.66	207.74	148.56	195.33	147.96
F0semitoneFrom27.5Hz_sma3nz_stddevRisingSlope	218.48	255.36	178.61	200.85	226.30	236.20	202.54	224.72
F0semitoneFrom27.5Hz_sma3nz_meanFallingSlope	113.45	94.22	128.85	116.32	132.14	80.67	118.46	77.98
F0semitoneFrom27.5Hz_sma3nz_stddevFallingSlope	141.86	166.58	147.45	189.43	166.49	156.16	133.93	140.77
loudness_sma3_amean	0.66	0.33	0.57	0.33	1.56	0.54	0.98	0.41
loudness_sma3_stddevNorm	0.73	0.24	0.79	0.27	0.58	0.17	0.54	0.19
loudness_sma3_percentile20.0	0.20	0.22	0.19	0.18	0.65	0.38	0.49	0.35
loudness_sma3_percentile50.0	0.64	0.41	0.49	0.40	1.53	0.59	0.98	0.48
loudness_sma3_percentile80.0	1.08	0.48	0.96	0.56	2.40	0.90	1.44	0.58
loudness_sma3_pctlrange0.2	0.88	0.40	0.77	0.51	1.75	0.83	0.95	0.48
loudness_sma3_meanRisingSlope	4.86	2.74	4.71	3.21	12.55	6.74	7.03	3.97
loudness_sma3_stddevRisingSlope	3.34	2.52	3.62	3.07	8.93	5.91	4.97	3.32
loudness_sma3_meanFallingSlope	3.21	1.85	3.16	2.12	7.27	3.96	4.50	2.90
loudness_sma3_stddevFallingSlope	2.21	1.60	2.32	1.95	4.93	3.49	2.89	2.21
jitterLocal_sma3nz_amean	0.03	0.01	0.04	0.02	0.04	0.01	0.04	0.02
jitterLocal_sma3nz_stddevNorm	1.54	0.49	1.49	0.53	1.17	0.44	1.37	0.41
shimmerLocaldB_sma3nz_amean	1.11	0.40	1.18	0.42	1.45	0.41	1.30	0.38
shimmerLocaldB_sma3nz_stddevNorm	0.82	0.23	0.75	0.23	0.62	0.16	0.60	0.17
HNRdBACF_sma3nz_amean	6.76	3.79	6.21	4.07	3.96	2.91	4.31	3.42
HNRdBACF_sma3nz_stddevNorm	1.16	2.67	0.99	3.28	-2.25	45.62	1.56	6.01
logRelF0.H1.H2_sma3nz_amean	2.20	5.81	3.42	6.55	0.62	6.08	1.35	5.77
logRelF0.H1.H2_sma3nz_stddevNorm	1.58	28.07	-15.00	137.51	0.52	30.24	4.34	64.47
logRelF0.H1.A3_sma3nz_amean	7.14	7.68	8.20	8.80	2.08	8.09	3.53	7.30
logRelF0.H1.A3_sma3nz_stddevNorm	2.11	27.78	0.09	10.79	3.80	61.50	1.00	10.92
F1frequency_sma3nz_amean	929.49	153.47	930.02	176.25	974.94	137.29	943.93	169.76
F1frequency_sma3nz_stddevNorm	0.30	0.09	0.30	0.08	0.30	0.08	0.31	0.08
F1bandwidth_sma3nz_amean	970.76	175.43	980.14	191.29	918.24	152.65	877.87	173.85
F1bandwidth_sma3nz_stddevNorm	0.32	0.13	0.31	0.12	0.30	0.12	0.38	0.14
F1amplitudeLogRelF0_sma3nz_amean	-81.11	38.70	-93.46	47.52	-52.19	34.85	-51.50	37.52
F1amplitudeLogRelF0_sma3nz_stddevNorm	-1.54	1.83	-0.96	7.85	-2.41	2.39	-2.66	11.14
F2frequency_sma3nz_amean	1978.08	157.47	1959.24	172.09	2053.19	131.15	2001.82	169.34
F2frequency_sma3nz_stddevNorm	0.14	0.04	0.13	0.04	0.14	0.03	0.14	0.03
F2amplitudeLogRelF0_sma3nz_amean	-67.84	33.79	-81.00	44.73	-41.18	27.81	-37.18	27.46
F2amplitudeLogRelF0_sma3nz_stddevNorm	-0.16	15.96	-1.49	2.32	-4.15	24.21	-2.79	7.79
F3frequency_sma3nz_amean	2979.21	172.59	2996.92	190.33	3018.29	132.11	2996.06	176.32
F3frequency_sma3nz_stddevNorm	0.10	0.03	0.11	0.03	0.10	0.02	0.10	0.02
F3amplitudeLogRelF0_sma3nz_amean	-69.14	33.78	-82.52	44.05	-41.40	27.32	-38.06	27.03
F3amplitudeLogRelF0_sma3nz_stddevNorm	-1.67	3.09	7.05	126.68	-3.34	8.14	-2.59	9.47
alphaRatioV_sma3nz_amean	3.64	7.93	1.63	9.32	7.10	7.29	6.57	7.66
alphaRatioV_sma3nz_stddevNorm	-0.38	22.81	-1.72	14.33	-0.86	18.01	-0.73	46.33
hammarbergIndexV_sma3nz_amean	9.07	6.72	10.30	7.98	4.64	7.29	6.48	6.42
hammarbergIndexV_sma3nz_stddevNorm	-2.62	47.64	1.43	6.41	-1.17	28.84	1.66	9.85
slopeV0.500_sma3nz_amean	0.10	0.05	0.09	0.05	0.10	0.04	0.07	0.05
slopeV0.500_sma3nz_stddevNorm	0.47	1.29	0.32	2.01	-0.36	13.34	0.07	9.85
slopeV500.1500_sma3nz_amean	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01
slopeV500.1500_sma3nz_stddevNorm	1.68	7.00	1.53	16.49	2.44	22.84	-9.30	180.39
alphaRatioUV_sma3nz_amean	-1.19	5.09	-2.17	5.52	1.78	4.09	0.34	5.48
hammarbergIndexUV_sma3nz_amean	9.13	5.64	9.82	6.21	5.45	4.95	7.70	6.14
slopeUV0.500_sma3nz_amean	0.07	0.05	0.07	0.05	0.06	0.04	0.06	0.05
slopeUV500.1500_sma3nz_amean	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01
loudnessPeaksPerSec	2.05	0.71	1.97	0.82	2.65	0.94	2.70	1.02
VoicedSegmentsPerSec	1.45	0.69	1.43	0.83	1.67	0.67	1.48	0.73
MeanVoicedSegmentLengthSec	0.59	0.44	0.58	0.58	0.60	0.52	0.78	0.75
StddevVoicedSegmentLengthSec	0.39	0.25	0.43	0.30	0.46	0.24	0.59	0.37
MeanUnvoicedSegmentLength	0.27	0.17	0.33	0.23	0.14	0.08	0.15	0.12
StddevUnvoicedSegmentLength	0.18	0.14	0.26	0.21	0.10	0.08	0.11	0.10

**Table 5:** Acoustic summary statistics by cry type using GeMaps feature set
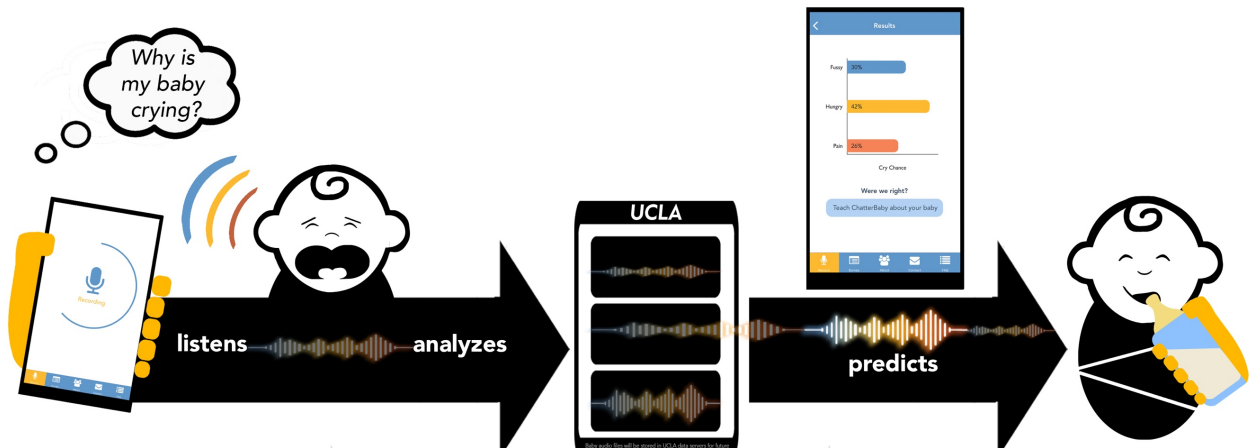
Supplemental Figure S1: The free ChatterBaby app predicts the most likely reason for a baby's cry using machine learning. Although originally designed for Deaf parents, this app has shown tremendous international uptake and is used to harvest autism research data from children around the world. Future work on ChatterBaby will assess whether the combination of acoustical biomarkers, phenotypic, and behavioral information may help improve autism screening in children.


Supplemental Figure S2: Although over 6000 features were used for classification, many of these had low importance in the random forests machine learning algorithm.

ROC Curve for Identifying Painful Infant Cries
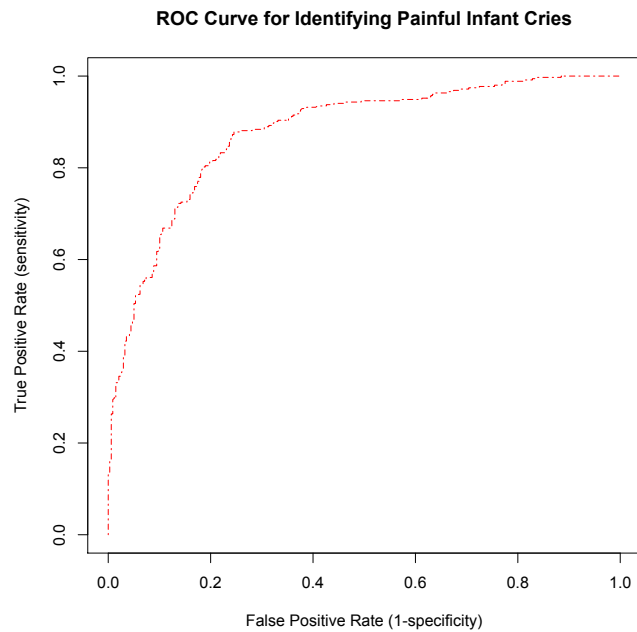
Supplemental Figure S3: Predictive accuracy of algorithm to identify painful infant cries. AUC= .877. Accuracy was defined with respect to the out-of-bag testing accuracy in the random forests algorithm, which is analogous to the cross-validation error. The curve represents the change in the true and false positive rate over a range of probabilistic thresholds used to define a cry as painful.

Correlation of Predictions
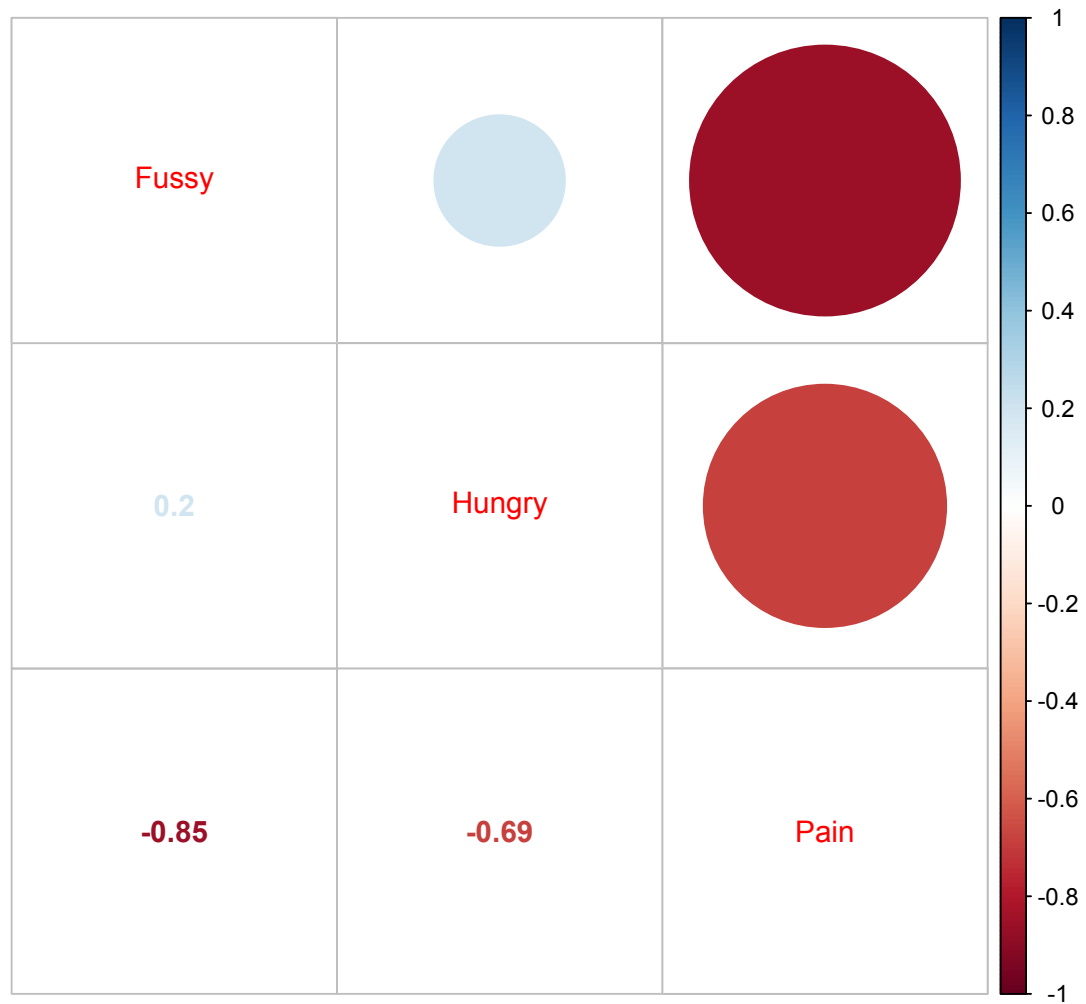
Supplemental Figure S4: As the probability of Pain in a cry increased, the probability of Fussy decreased.
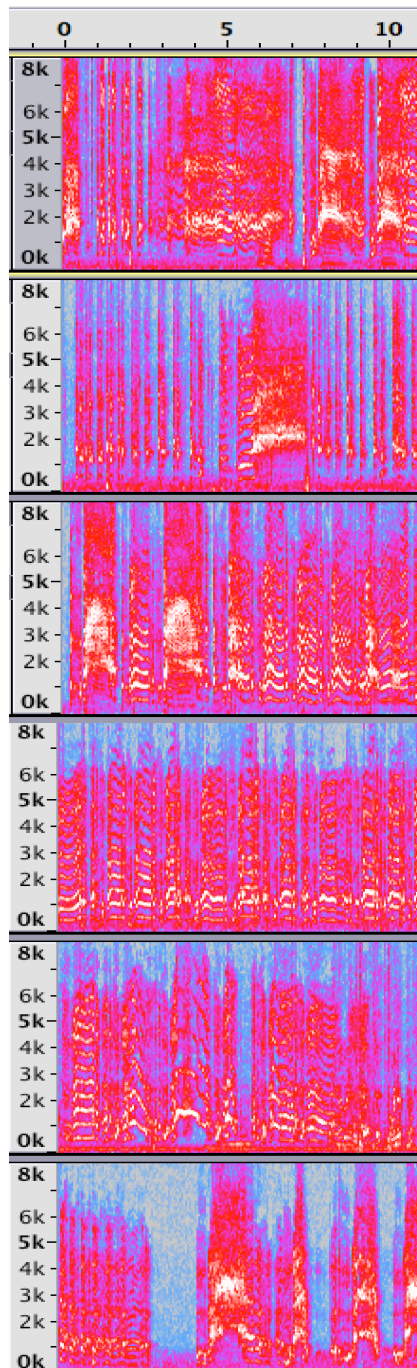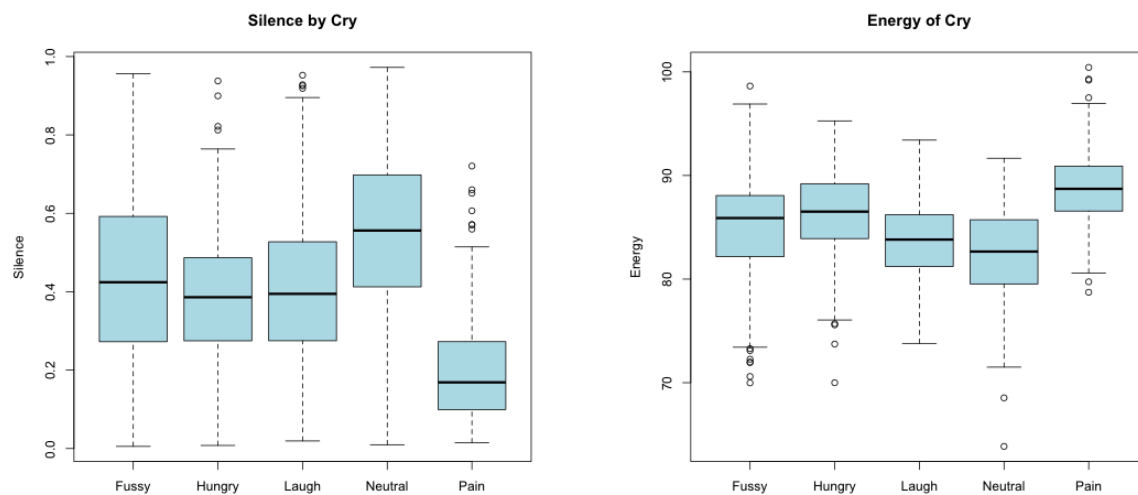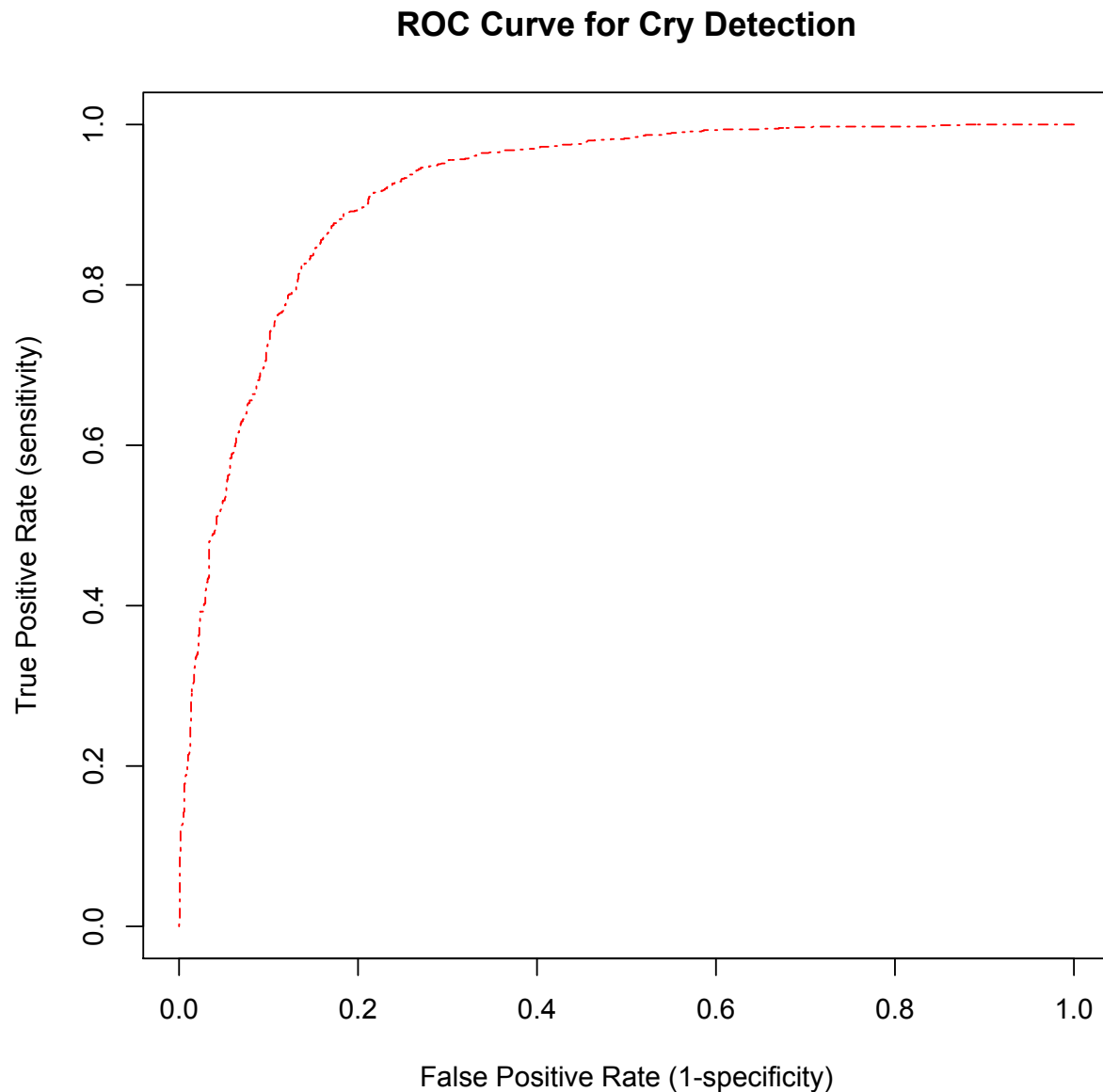
Supplemental Figure S5: 11 second Spectrogram of vaccination cries for a single male child aged 87-618 days. Five-second audio clips from this child's vaccinations over an 18 month period are available at https://www.youtube.com/watch?v=eu332YZFTkA

Figure 6: Cries from babies in pain had more energy and fewer periods of silence.

Figure 7: ROC Curve for cry detection. AUC = .92. The 1,152 Distractor states were compared to 1,192 baby cries to create a Cry Detection algorithm.

1.   Bornstein, M.H., et al., *Neurobiology of culturally common maternal responses to infant cry.* Proceedings of the National Academy of Sciences, 2017. **114**(45): p. E9465-E9473.
2.   Kim, P., et al., *Breastfeeding, brain activation to own infant cry, and maternal sensitivity.* Journal of child psychology and psychiatry, 2011. **52**(8): p. 907-915.
3.   Truzzi, A., et al., *Infant communicative signals elicit differential brain dynamics in fathers and non-fathers.* Early Child Development and Care, 2018: p. 1-9.
4.   Hitoglou, M., et al., *Childhood autism and auditory system abnormalities.* Pediatric Neurology, 2010. **42**(5): p. 309-314.

5. Control, C.f.D. and Prevention, *Learn the signs. Act early: Developmental milestones.* 2011.

6. Allison, C., et al., *The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): a normally distributed quantitative measure of autistic traits at 18–24 months of age: preliminary report.* Journal of autism and developmental disorders, 2008. **38**(8): p. 1414-1425.

7. Scott, F.J., et al., *The CAST (Childhood Asperger Syndrome Test) Preliminary development of a UK screen for mainstream primary-school-age children.* Autism, 2002. **6**(1): p. 9-31.

8. Jonsdottir, R.B. and G. Kristjansdottir, *The sensitivity of the premature infant pain profile–PIPP to measure pain in hospitalized neonates.* Journal of evaluation in clinical practice, 2005. **11**(6): p. 598-605.

9. Lawrence, J., et al., *The development of a tool to assess neonatal pain.* Neonatal network: NN, 1993. **12**(6): p. 59-66.

10. Taddio, A., et al., *Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections.* Vaccine, 2011. **29**(7): p. 1390-1394.

11. Bellieni, C.V., et al., *Inter-observer reliability of two pain scales for newborns.* Early human development, 2007. **83**(8): p. 549-552.

12. Ge, X., et al., *Bayesian estimation on diagnostic performance of Face, Legs, Activity, Cry, and Consolability and Neonatal Infant Pain Scale for infant pain assessment in the absence of a gold standard.* Pediatric Anesthesia, 2015. **25**(8): p. 834-839.

13. Ahn, Y., H. Kang, and E. Shin, *Pain assessment using CRIES, FLACC and PIPP in high-risk infants.* Journal of Korean Academy of Nursing, 2005. **35**(7): p. 1401-1409.

14. Suraseranivongse, S., et al., *A comparison of postoperative pain scales in neonates.* BJA: British Journal of Anaesthesia, 2006. **97**(4): p. 540-544.

15. Parsons, C.E., et al., *Introducing the Oxford Vocal (OxVoc) Sounds database: a validated set of non-acted affective sounds from human infants, adults, and domestic animals.* Frontiers in psychology, 2014. **5**: p. 562.

16. Schuller, B., et al., *The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism.* 2013.

17. Schuller, B.W., et al. *The INTERSPEECH 2014 computational paralinguistics challenge: cognitive & physical load.* in *INTERSPEECH.* 2014.

18. Schuller, B., et al. *The INTERSPEECH 2015 Computational Paralinguistics Challenge: Nativeness, Parkinson's & Eating Condition.* in *Sixteenth Annual Conference of the International Speech Communication Association.* 2015.

19. Eyben, F., M. Wöllmer, and B. Schuller. *Opensmile: the munich versatile and fast open-source audio feature extractor.* in *Proceedings of the 18th ACM international conference on Multimedia.* 2010. ACM.

20. Grósz, T., et al. *Assessing the degree of nativeness and Parkinson's condition using Gaussian processes and deep rectifier neural networks.* in *Sixteenth Annual Conference of the International Speech Communication Association.* 2015.

21. Wang, J., et al. *Predicting intelligible speaking rate in individuals with amyotrophic lateral sclerosis from a small number of speech acoustic and articulatory samples.* in *Workshop on Speech and Language Processing for Assistive Technologies.* 2016. NIH Public Access.

22.	Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.
23.	Eyben, F., et al., *The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing.* IEEE Transactions on Affective Computing, 2016. **7**(2): p. 190-202.
24.	Barr, R.G., et al., *Prolonged and unsoothable crying bouts in infants with and without colic.* Journal of Developmental & Behavioral Pediatrics, 2005. **26**(1): p. 14-23.