Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at *Acoustics* **2019**, *1*, 52; doi:10.3390/acoustics1040052

Review

Are Cry Studies Replicable? An Analysis of Participants, Setups, and Methods Adopted and Reported in Cry Studies

Giulio Gabrieli [1], Giulia Scapin[2], Marc H. Bornstein[3] and Gianluca Esposito [1,2*]

[1] Psychology Program, Nanyang Technological University, Singapore
[2] Department of Psychology and Cognitive Science, University of Trento, Italy,
[3] Child and Family Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland, USA
* Correspondence: gianluca.esposito@e.ntu.edu.sg

Abstract: Infant cry is evolutionarily, psychologically, and clinically significant. During the last 60 years, several researchers and clinicians assessed the possibility of investigating the acoustical properties of cry for medical purposes. However, there is a lack of standardization in conducting and reporting cry-based studies. In this work, methodologies and procedures employed in infant cry analysis are reviewed, and best practices for reporting studies are provided. First, available literature on vocal and audio acoustic analysis have been examined to identify critical aspects of participant information, data collection, methods, and data analysis. Then, 180 peer-reviewed research articles have been assessed to certify the presence of identified critical information. Results show a general lack of critical description. Researchers in the field of infant cry need to agree on a standard set of criteria to report experimental studies, to better demonstrate the validity of the methods and obtained results.

Keywords: infant cry; acoustic analysis; cry analysis

---

1. Introduction

Cry is one of the first forms of communication newborns can use to interact with their caregivers. Cry vocalizations are produced by the vibration of vocal folds that are controlled by the Central Nervous System (CNS). Thus, researchers and clinicians investigated the possibility of relying on the acoustic analysis of infant cry to assess, in a non-invasive way, the integrity and developmental status of the CNS. Moreover, this analysis has been proven to be effective in identifying the insurgency of Autism Spectrum Disorder (ASD) [1–3], Sudden Infant Death Syndrome (SIDS) [4] and a variety of auditory-related problems during the early stages of development [5–9].

Historically, the first attempts at investigating the acoustical properties of cry were conducted in the late 1960s. Since then, there has been an exponentially increasing interest within the field, as demonstrated by the number of articles that are published each year on the topic (Figure 1). Unfortunately, within the rich body of literature published on the acoustical properties of cry and its relationship with developmental pathology, it is not rare to find conflicting results. For examples, differences in the acoustical properties of cry vocalizations of children at risk for ASD were found in Sheinkopf *et al.* [10], but not in Unwin *et al.* [11]. Likewise, the possibility of achieving a high level of accuracy in identifying deaf children by their cry vocalization was shown in Garcia and Garcia [12], while no differences between cry properties of deaf and normal-hearing infants were found in Várallyay [13]. Differences in obtained results are not only due to the high inter-individual differences of cry vocalization but, as Etz and colleagues pointed out ([14], there is a general lack of standardization within the field that does not allow researchers to correctly replicate a study. More specifically, there are no standardized guidelines researchers can use to report information about investigated participants, employed methods, collected data, and analysis procedures, nor standardized datasets that can be used to compare the performances of different statistical models. Furthermore, according to Wermke and Mende, this lack of standardization seems generally accepted by the community of cry investigators [15].

In a typical cry study, samples are collected from multiple infants who are induced to cry using a *trigger*, such as painful stimulus, while acoustic samples are recorded using one or more microphones and then stored in an analogical or digital drive.

Acoustic analysis of infant cry is usually not based on the signal as it is, but on features extracted from the collected recordings. These are distinctive characteristics contained in a sample, such as the intensity of a specific frequency band. Among the most used features are the fundamental frequency (also defined as F_0) and its formants, the intensity of a frequency band, and duration and frequency of cry episodes.

Since the publication of the first monograph on cry analysis in 1968 [16], progress has been made in the hardware used for data collection, instrumentation employed for acoustic analysis, features of interest, and algorithms and methods used for data modeling. At first, researchers relied on visual inspection of the spectrograms to extract meaningful information from cry samples. With the advent of more powerful computing technologies, the interest shifted toward methods that allow the estimation of acoustic features with higher accuracy.

Unfortunately, results obtained by analysis of cry samples are tightly bonded to the pool of participants, instrumentation used, and methods employed for feature extraction. In this article, we investigate how studies published within the field of infant cry have been conducted and reported, highlighting the difficulties of replicating cry-based studies. Finally, we propose a list of elements researchers should consider when designing, conducting, and reporting their studies.

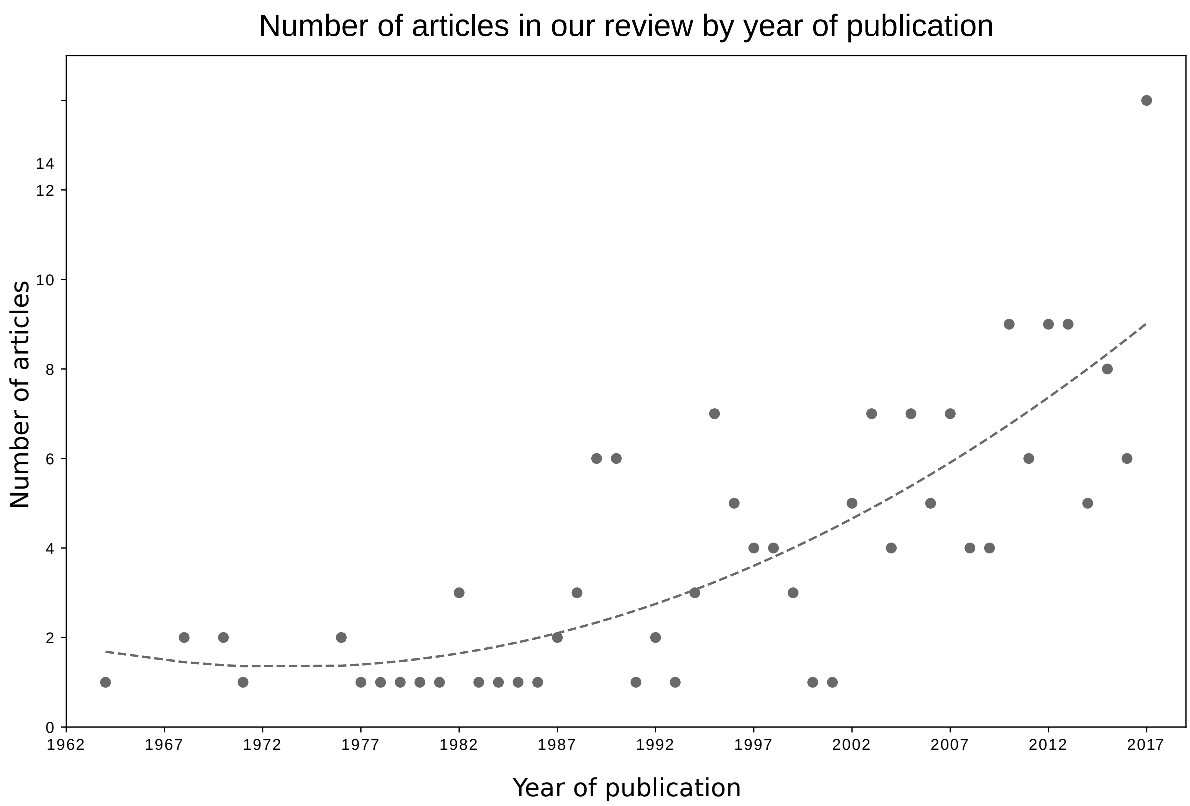

Figure 1. Number of articles per year of publication that succeeded our screening and met our inclusion criteria. Regression was evaluated using a 2nd degree least square polynomial fit.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

2. Materials and Method

2.1. Variable definition

In a typical study on cry samples, three different categories of variables can be distinguished: (1) Participant Information, (2) Information about Data Collection Processes, and (3) Information about Methods and Data Analysis. Participation information concerns itself with details about (a) number of participants (Part), (b) number of cry samples (Sam), (c) age of the infants (Age), (d) sex of the infants (Sex), (e) trigger (Tri), (f) position of the infants during the recording (Pos) and (g) health status of the infants (Hea). Regarding data collection, (a) the type of microphone used for data collection (Mic), (b) Microphone-to-Mouth distance (MtM), (c) recording environment (Env), (d) sampling rate of recorded signal (SR), and (e) file format used for storage (FF) were coded. Finally, with regard to the methods and data analysis stages, the (a) preprocessing procedure (PP), (b) software and hardware used (SwHw), (c) method of feature extraction (FE), (d) analyzed frequency range (FR), (e) analyzed features (AF) and (f) window size of the signal during features extraction (Ww) have been investigated. Detailed information about each variable is reported in Appendix A.

2.2. Search methods and results

To verify how information about participants, data collection, methods, and analysis have been reported in published cry studies, we selected a set of articles following the preferred reporting items for systematic reviews and meta-analyses (PRIMA) [17]) guidelines.
An electronic database search on Elseviers' Scopus was conducted in March 2018, using the query *"Acoustic* Cry"*. As a result of our search query, Scopus returned 391 results. After automatic duplicate removal, 383 articles remained and were manually inspected for inclusion in the current study. One of the authors (G.G.) screened article titles and abstracts to exclude irrelevant articles. A total of 133 articles were excluded as irrelevant to our research question. The remaining 250 articles were then examined by two independent coders to determine which articles met all four inclusion criteria. 77 articles were excluded because they did not meet our four inclusion criteria. A detailed representation of articles screening procedure is given in Figure 2. Therefore, the final dataset consists of 180 articles. Two independent coders (average between-coder percentage agreement = 84%, evaluated on a subset of 20 randomly selected articles), blind to the aim of the experiment, analyzed the included articles and coded for the presence of the 18 variables.

3. Results

Percentage of presence of each variable is presented in Table 1. Results for each article can be found in supplementary material, while detailed notes for each manuscript are available in the online version of the dataset [18].

3.1. Participants

Overall, a very high proportion of investigated articles reported the (a) number of participants of the experiment (N = 161, 89%), while the (b) number of cry samples analyzed in a study was reported in a slightly lower number of articles (N = 122, 68%). Even though (c) the age of the participants was reported in the majority of investigated studies (N = 133, 74%), researchers were not that concerned in reporting the (d) gender of their participants (N = 71, 39%). The reason that induced a crying occurrence in infants, or (e) Trigger, driven by the experimenter or spontaneous, was reported in many studies (N = 121, 67%). The (f) posture of the babies during the recording was reported in only a minority of investigated studies (N = 55, 30%). Finally, studies that investigated children suffering from pathological conditions almost all reported, when necessary, details about the (g) health status of their pool of participants (N = 73 out of 77 studies in which it was deemed necessary, 95%).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

	Variable	Abbr.	N	%
Participants	Number of Participants	Part	161	89%
	Number of Samples	Sam	122	68%
	Participants' Age	Age	133	74%
	Participants' Gender	Sex	71	39%
	Cry Trigger	Tri	121	67%
	Participants' Position	Pos	55	30%
	Participants' Health Status	Hea	73 (77)	95%
Data Collection	Microphones' Model	Mic	112	62%
	Mouth-To-Microphone Distance	MTM	102	57%
	Recording Environment	Env	106	59%
	Sampling Rate	SR	115	64%
	File Format	FF	69	38%
Methods and Analysis	Preprocessing Procedure	PP	98	54%
	Software / Hardware	SwHw	140	78%
	Feature Extraction Methods	FE	150	83%
	Frequency Range	FR	31	17%
	Analyzed Features	AF	161	89%
	Windows Size	Ww	53	29%

Table 1. (N = 180, where not differently stated)

There is high variability in the number of participants examined in analyzed studies[1]. Ranging from single-case studies (N = 5) to studies employing up to 1388 participants, with an average of 68.3 ± 150.7 participants per study. Similarly, the number of samples per study ranges from single-sample analysis (N = 1) to studies with up to 31400 samples. The average number of samples per study is 729.1 ± 2861.5, with an average of 36.2 ± 221.4 samples per participant (evaluated on single study manuscripts where both the number of participants and number of samples have been reported, N = 104). Ages of babies under examination in reviewed papers range from 1 to 4 years of life, with the average participants aged 133.5 ± 181.1 days old. More balanced is the situation for what concerns the number of males and females babies who participated in examined studies, with a total of 1709 females and 1663 males examined. An interesting view arises when it comes to the trigger used to induce a crying vocalization. Only in 70 (39%) studies, the researchers relied on non-spontaneous vocalizations induced with a single trigger. Overall, majority of the studies adopted painful vocalization (N = 76, e.g. Heel prick), followed by spontaneous cries (N = 30), discomfort calls (N = 29), hunger-induced calls (N = 26), fear-induced calls (N = 30) and fussiness calls (N = 30). More homogeneous is the adoption of the position in which the babies were placed during the recordings, with the majority of the studies reporting that recordings were obtained with infants lying on their back (supine position, N = 42), with just a minority adopting a seating position (N = 14) and only one article in which infants where placed on their front (prone position, N = 1). Regarding the health status of examined babies, fourteen (N = 14) studies had a group composed by Premature babies, eleven (N = 11) deaf children, ten (N = 10) children diagnosed with or at risk of developing Autism Spectrum Disorder, and nine (N = 9) studies had a group of children suffering from Asphyxia.

3.2. Data collection

In describing their data collection procedure, more than half of the research articles reported (a) the model of the microphone employed during the recordings (N = 112, 62%) and (b) the mouth-to-microphone distance (N = 102, 57%). The (c) recording environment was clearly stated in almost six papers out of ten (N = 106, 59%), slightly less then (d) the sampling rate of recorded signals

[1] One article may contain more than one study

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

or of the speed of the recording tape (N = 115, 64%), while only to a lower extent researchers indicated (e) the file format used for storage (N = 69, 38%), with it being a tape, disk, or digital format.

When it comes to the model of microphone, researchers relied on many different models and producers (for a complete list, please refer to [18]), while much more consistency is present for what concerns the most adopted mouth-to-microphone distance, where the 0.15m distance have been used in thirty-one (N = 35) studies, followed by the 0.20m (N = 13) and 0.30m (N = 11). A fairly unbalanced situation is presented when looking at the recording location, where studies employing recording collected in clinical situation (N = 75) are almost twice the number of studies in which recordings were performed in a non-clinical setting (N = 40). Only a minority of studies (N = 8) employing cries collected in both a clinical and non-clinical setting. As for the sample rate of recorded signals, a vast majority of collected samples were sampled at 44'100 Hz (N = 30), 10'000 Hz (N = 18), 16'000 Hz (N = 12) and 48'000 Hz (N = 13), with signals mostly recorded on magnetic tapes (N = 52) and in WAV format (N = 15).

3.3. Methods and data analysis

Amongst investigated studies, about half (N = 98, 54%) of all studies clearly reported the employed (a) preprocessing procedure, when one was applied. A majority of examined articles provided information about the (b) software or hardware used for the analysis (N = 140, 78%) and (c) feature extraction methods (N = 150, 83%). However, only a small number clearly stated the (d) region of interest within the spectrum (N = 31, 17%). (e) Studied features were clearly listed in almost all the investigated articles (N = 161, 89%), while the (f) window size used during the analysis was reported in only almost a third of examined works (N = 53, 29%).

Overall, a manual or automatic segmentation step of collected signals is usually adopted (N = 54), preceded or followed by a filtering phase (N = 44). In various papers, only a subset of extracted segments (for example, only the first three cry vocalizations of each participant [19]) was used in subsequent analysis (N = 32). As for the software used for the analysis, KAY's Computer Speech Lab was the most used tool (N = 24) until recently. It has since been slowly replaced by Praat (N = 36). On the other hand, researchers' favorite scripting language remains as Matlab (N = 31). Despite the effort in reporting the used software and instrumentation, the software version or model of the hardware were only reported in less than half of the cases (N = 107 out of 246, 43%). This is also reflected in feature extraction methods, in which methods based on the analysis of the spectrum of an audio sample dominate above the others (FFT N = 48, MFCC N = 22, LTAS N = 15). Concerning the most investigated frequency range, the majority of the studies focused on frequencies below 10kHz (N = 32), with a vast interest in frequencies up to 5kHZ (N = 21). The majority of the studies analyzed the fundamental frequency (F0, N = 124) of cry samples, the duration of cry vocalizations (N = 82), and the energy conveyed by the signal (N = 60) as some of the most investigated acoustic features. As for the window sizes, employed windows ranged between 5ms and 290ms, with 25ms (N = 20) and 50ms (N = 14) being the most adopted window sizes.

4. Discussion

Results of our analysis confirm Etz and colleagues' conclusion about the general lack of standardization in reporting cry-based studies.

When it comes to the information about the participants whose vocalizations have been reported and analyzed, researchers usually stress the importance of the number of babies that took part in the experiment, as well as the number of recorded samples, age, trigger, and health status. Little stress is instead posed on the importance of reporting participants' gender and position assumed during the recording. Taken together, this raises implications about the reproducibility issue of cry studies. If it is true that the gender of a babies affects the properties of his or her vocalizations [20–22], when this information is not reported, researchers will not be able to assume that acoustic properties of infants'

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

vocalizations are normally distributed, or that deviations from normality are not present because of an unbalanced number of male and females babies in the groups being compared.

Similarly, when researchers report the experimental setup they employed for their studies, in just about six out of ten articles the model of microphone used to record signals, the distance between the babies' mouth and the microphone, the sampling rate of recorded signal, or the recording location have been indicated, while the file format used for storing recorded signals have been reported in even fewer cases. The problem with such missing information is that even in cases where the same microphone is used across different studies, researchers are unable to ensure that recorded signals are comparable if the microphones have been placed at different distances, recorded in different locations, or with different sampling rates. Moreover, especially when it comes to digital signals, the format used for storage play an important role in the preservation of recorded frequency, as different compression algorithms may alter the frequency information conveyed in infants' vocalizations (additional details on file compression are reported in A.2.5).

Finally, regarding the methods and analysis sections of their papers, researchers are consistent in reporting the software and hardware used in their work (even though the software version or hardware version was reported only in approximately half of the cases studied), the feature extraction procedure, and the list of used features. Less agreement is found in indicating the preprocessing procedure the signal had undergone prior to the analysis, the region of interest within the spectrum, and the window sizes used to study the investigated features. However, it must be noted that knowledge about software version is crucial, especially when it comes to the software used for feature estimation, and the parameters used to process the signals. For example, among the articles we investigated, although Praat was widely adopted by different research groups, only 17 cases reported the software version used (out of the 36 in which it was employed), even though more than four hundred different releases of Praat are available[2]. Additionally, different preprocessing procedures may reflect differences in results obtained by different studies, and therefore a clear indication of the steps adopted in a study should always be reported in the final manuscripts. At the same time, we acknowledge that current methods of reporting analysis may change as more and more scientists are embracing open practices, such as data, script and software sharing.

One of the limitations of infant cry studies is, in fact, the absence of standardized datasets that can be used to compare novel approaches with traditionally employed techniques. It is not uncommon in investigated studies to see that researchers collected data for their works that were not published alongside the manuscript or in external repositories, or that they have employed outdated datasets, such as the Baby Chillanto Infant Dataset whose web page is no longer available [23].

With this work, we hope to enable a constructive discussion on how to standardize current and future investigations among cry researchers in order to enhance the reproducibility of those works and facilitate the adoption of cry based technologies for clinical purposes. Future manuscripts should contain all the information needed to correctly replicate the study in the future, and to allow a critical interpretation of obtained results based on participants' demographic information, data collection procedures, and analysis methods. Moreover, researchers should, whenever possible, share their original recordings. Availability of an accessible dataset may positively impact quality and quantity of published research through the development of a new and improved methodology for feature extraction, which can be used in a clinical environment for early diagnosis of developmental atypicality.

5. Conclusions

This study investigates how cry research has been conducted and reported in the last 60 years. In 1995, Robb wrote that the absence of acoustic validation studies was surprising for two reasons: firstly, there are researchers who state that acoustical analysis of cry is diagnostically significant and secondly,

---

[2] https://github.com/praat/praat/releases, 416 releases as per November 5, 2019

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

the unique configuration of an infant's vocal tract is difficult to ignore. Unfortunately, after more than 20 years, the presence of one or more standardized datasets or guidelines for analysis and publication is still missing.

In this article, 180 research articles were reviewed and the presence of variables that can be used to replicate, compare, and make assumptions about the relationship between acoustical features of infant cry and infant developmental status was annotated. We found a pervasive lack of critical description regarding various aspects of samples and their properties, the process of data collection, and in the methods and data analysis. This shortage of information is accentuated by the low number of freely available datasets that can be used to test and compare feature extraction methods. Similarly, to the best of our knowledge, there are no guidelines for reporting variables to effectively explain results obtained in cry analysis. Researchers in the field should agree on standardised ways to report their experimental studies to ensure the validity of their methods and results. We hope that this discussion will inspire a self-evaluation of the many points that have to be thoughtfully considered during participant selection, data collection, and subsequent analysis. Availability of reliable results affects the ability of pediatricians to recognize pathology and developmental problems in their early stages using non-invasive techniques, as well as generates discussion among researchers for the best methodologies that can be applied to cry analysis.

Supplementary Materials: The following are available online at http://www.mdpi.com/2624-599X/xx/1/5/ https://doi.org/10.21979/N9/UDQBEK, Script S1: Analysis Notebook used for this publication, R1 Raw Scopus Dump: , R2: Raw coders' annotation, R3: Standardized coders' annotations, P1: Summary Table

Author Contributions: Conceptualization, G.G., G.S., M.H.B and G.E.; methodology, G.G., G.S. and G.E.; formal analysis, G.G.; data curation, G.G. and G.E.; writing–original draft preparation, G.G.; writing–review and editing, G.G., G.S., M.H.B. and G.E.; visualization, G.G.; supervision, G.E.;

Funding: This research was supported by Nanyang Technological University (Singapore) under the NAP-SUG grant.

Acknowledgments: the authors are thankful to Anais Ang Xin Hui, Chiara Iannaccone, Giulia Garbin, Mengyu Lim, and Pamela Goh Pei Lin for for their help in this work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNS	Central Nervous System
ASD	Autism Spectrum Disorder
SIDS	Sudden Infant Death Syndrome
F_0	Fundamental Frequency
F_n	n^{th} Formant
Part	Number of Participants
Sam	Number of Samples
Age	Participants’ Age
Sex	Participants’ Gender
Tri	Cry Trigger
Pos	Participants’ Position
Hea	Participants’ Health Status
Mic	Microphones’ Model
MTM	Mouth-To-Microphone Distance
Env	Recording Environment
SR	Sampling Rate
FF	File Format
PP	Preprocessing Procedure
SwHw	Software / Hardware
FE	Feature Extraction Methods
FR	Frequency Range
AF	Analyzed Features
Ww	Windows Size

Appendix A. Supplementary Material - Variable descriptions

Appendix A.1. Participant information

Key variables associated with participant information include the number of participants (Part), number of cry samples (Sam), age (Age) and sex (Sex) of the infants, trigger (Tri), the position of the infants during recording (Pos), and infant health status (Hea).

Appendix A.1.1. Number of participants (Part)

The number of participants investigated in a cry study is an important parameter to consider for several reasons. Cry vocalizations are highly variable both within samples collected from the same subject and between the subjects. Due to between-subject variability, results based on small numbers of participants may induce errors in interpretation [24,25]. For example, the application of predictive models based on training done on a small number of participants results in models that are not suitable for generalization to a wider population [26].

Appendix A.1.2. Number of cry samples (Sam)

Because of the high within-subject variability [14], the number of different cry samples collected in a study can be used to assess the extent to which obtained results are influenced by a single individual's cry recordings. This shortcoming holds especially for studies with limited numbers of participants and few cry samples for each. In these cases, the properties of a single sample can severely influence the average values of the whole dataset. Studies with large numbers of samples collected from a small pool of participants may be affected by overfitting, with statistical models becoming too sensitive to the investigated individuals.

Appendix A.1.3. Age of the infants (Age)

As cry is produced by the vibration of the vocal folds that grow and change with an infants' age, acoustical properties of cry are influenced by the age of investigated participants [27]. Some authors suggest reporting gestational age at birth as well, together with infants' weight both at birth and during data collection [20,28].

Appendix A.1.4. Sex of the infants (Sex)

Controversial results on cry have been reported regarding the sex of infants. Multiple studies (e.g.: [20–22]) found significant differences between the acoustical properties of cry vocalizations of girls and boys, while in other studies (e.g. [29]) no differences were found or reported.

Appendix A.1.5. Trigger (Tri)

Acoustical properties of cry reflect the reason that induced the cry [30,31]. Vocalizations obtained from different functional roles (e.g., pain or hunger) are not comparable because cries from different trigger categories convey different frequency information. Researchers should report the trigger used to induce babies to cry and compare only cries obtained using the same trigger.

Appendix A.1.6. Position of the Infant During Recording (Pos)

Infants' body position during cry recording has been found to influence the acoustic properties of cry. This influence is separate from the developmental status of the infant. For example, Goberman et al. [32], identified differences in cry acoustics of infants recorded in a supine versus prone position even in response to the same pain stimulus [33,34].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

Appendix A.1.7. Health status of the infants (Hea)

Infants' health and developmental statuses are reflected in the acoustic properties of cry. This peculiarity is the basis for research on early screening of pathological problems. Specific differences in cry features are associated with different pathologies. For example, a higher fundamental frequency in cry utterances is associated with a higher risk of autism spectrum disorder diagnosis [27,35–37], while hearing-impaired infants produce longer vocalizations with lower second formant (F_2) and less energy in the higher frequency bands [38,39]. Knowledge about the health status of investigated infants is necessary to correctly evaluate the obtained results.

Appendix A.2. Data collection

During data collection, several aspects of the experimental setup influence the quality and properties of recorded signals. We identified six key variables: characteristics of the microphone used for data collection (Mic), the microphone-to-mouth distance (MtM), recording environment (Env), the sampling rate of recorded signal (SR), file format used for storage (FF), and number of channels employed for data recording (NC).

Appendix A.2.1. Microphone used for data collection (Mic)

The type and model of microphone used are necessary to ensure that data collected using the described experimental setup is suitable for the analysis of investigated frequencies. Different microphones respond optimally to specific frequency ranges, according to the type of technology employed and directionality of the microphones [40]. The directionality of a microphone, expressed as polar pattern, indicates its sensitivity to sounds coming from different directions. Omnidirectional microphones respond in the same way to sound waves coming from different directions; cardioid microphones are sensitive to sound waves coming from a specific direction, limiting the external noise coming from other directions [41, Chapter 3]. Microphone sensitivity is represented using a frequency response chart, which graphically represents microphone response (in dB) to each frequency with the source at a specific distance from the microphone. An example of a frequency response chart appears in Figure A1. Microphones employed in cry studies should have a homogeneous response to the investigated frequencies, that are based on the features of interests and methodology employed for extraction. For example, for a direct estimation of the fundamental frequency (F_0), sensitivity should be homogeneous in frequencies below 1 kHz, while for indirect estimation of F_0 from the first four formant peaks, the response should be homogeneous from 700 Hz and 3 kHz.

Appendix A.2.2. Microphone-to-mouth distance (MtM)

Microphones have specific frequency responses and polarity. Microphone distance to the infant's mouth influences the intensity of recorded signals as well as the intensity of external noises, such as wind [41, Chapter 10]. In free space, a sound is propagated uniformly in every direction. According to the inverse square law, the intensity of sound in a free field is inversely proportional to the square of the distance from the source [42, Chapter 1]. Different microphones have different optimal distances, at which the response to a specific frequency ranges is flat. For example, vocal microphones are designed to produce a flat frequency response at 5 to 10 cm from the mouth, while noise-canceling microphones are most effective when positioned next to the mouth [41, Chapter 17]. To maximize the intensity of cry and minimize external noise, and to obtain a flat response in the frequency range of interest, microphone-to-mouth distance has to be carefully evaluated.

Appendix A.2.3. Recording environment (Env)

Laboratory studies have samples recorded within the same environment, removing possible variance in the recordings arising from the different acoustics of different rooms. Acoustic characteristics of recordings obtained in different places, such as at different infants' homes, are

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

influenced by non-identical wave reflection, which produces shifts in recorded frequencies [42, Chapter 7]. These reflections are modulated by the distances between microphones, infants' mouths, and walls, as well as by the walls' absorption coefficient [43, Chapter 2]. The cry trigger also modulates the acoustical properties of recorded signals. Ross et al. [44] investigated differences between distress situations in 12- to 18-month-old infants at home or in laboratory conditions. When tested in non-familiar environments, children cried almost three times as long as when they have been tested at home. Furthermore, in laboratory studies, the presence of others has to be taken into account, especially where soundproofed rooms are not available. Infants cry when hearing cries of other infants [45–47]. In hospital situations, where many infants are present, tested infants may be able to hear cries of many other infants or patients. To avoid empathic distress responses, it is important to prevent tested infants from hearing the cry vocalizations of other individuals.

Appendix A.2.4. Sampling rate of recorded signal (SR)

The sampling rate (SR) is the number of samples per second contained in a digitalized continuous signal [48]. A signal sampling rate influences the highest and lowest detectable frequencies, as well as the accuracy of extracted features. The maximum detectable frequency, or Nyquist frequency, corresponds to half of the sampling frequency [49,50]. The extraction of acoustic properties in the frequency domain is influenced by the signal's resolution, which is the smallest detectable change within subsequent points of a signal. For audio signals, the frequency resolution after applications of a Discrete Fourier Transform (DFT) is given by the ratio between the sampling rate of the signal and the number of points used in the DFT. During data processing and feature extraction, variation in the sampling rate results in distortion of the original signal, and therefore in the lowest and highest detectable frequencies, resolution and accuracy of extracted frequency information.

Appendix A.2.5. File format used for storage (FF)

Cry data can be stored as analog or digital forms. Today, the majority of researchers opt for direct storage on digital devices or analog-to-digital conversion. When stored on a digital drive, digital files are encoded in a specific file format (FF). Different file formats store files according to different acoustic properties, and therefore the original signal is adapted when stored. There are three different types of audio file formats: uncompressed, lossless compressed, and lossy compressed [51, Ch. 5, p. 157ff].

- Uncompressed files store the signal as it is, applying no content compression, resulting in files taking more space on digital drives. Tx
- The lossless compressed format encodes in a way that reduces the size of an input file by creating a copy with the same acoustical properties that may have a smaller size, usually in the ratio 2:1 [52].
- To achieve a greater reduction in file space, a lossy compression algorithm can be used. Lossy compression achieves a higher compression ratio, usually around the ratio of 10:1, by reducing the audio quality of the signal. Although quality loss is almost imperceptible to human ears, modification of original signal influences the quality and accuracy of acoustical features estimated from it, such as F_0. The most popular lossy file format is the MP3 format, which is widely used for music compression, but it is also employed in the research environment.

Appendix A.3. Methods and data analysis information

We identified six key variables that influence accuracy and precision of preprocessing techniques and data analysis results: the preprocessing procedure (PP), software and hardware used (SwHw), feature extraction method (FE), analyzed frequency range (FR), analyzed features (AF), and window size of the signal during feature extraction (Ww).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

Appendix A.3.1. Preprocessing procedure (PP)

Once collected, if data contain external noise or artifacts, there may be a manual or automatic screening of "good" versus "bad" samples, with the latter removed before the analysis. Before the data analysis, recorded signals may undergo a series of treatments to increase the signal to noise ratio, and enhance the accuracy and robustness of extracted features. Preprocessing can modify signals in different ways. For example, signals can be parsed for noise removal and segmented into smaller fragments or downsampled. Downsampling a signal to reduce computational time or to save space on hard drives results in a reduction of information stored in a signal. This reduction leads to shifts in frequency information [53]. Similarly, the application of digital filters to increase the signal-to-noise ratio may alter the properties of investigated frequencies (FR) [54, Chapter 1]. For this reason, preprocessing procedures (Pp) have to be clearly reported by providing all the information required to correctly evaluate or replicate the methodology, stating all the possible alterations of a signal.

Appendix A.3.2. Software and hardware (SwHw)

With the advent of computer-based methodologies, more objective and accurate digital analyses of quantitative acoustic parameters are now available. [55]. In all investigations, the original cry signal underwent modification, because of the hardware used for data collection and methodologies applied during preprocessing and data analysis. During preprocessing and subsequent analysis, researchers use different software and instrumentation (SwHw). Software name, version, and parameters need to be specified as new releases often correct previous bugs within the code that may have generated incorrect analyses. Furthermore, changes in customizable parameters in different software lead to differences in the accuracy and precision of extracted features. To estimate the fundamental frequency of cry samples, researchers often use Praat, a open-source software designed for voice analysis. Praat's source code repository (a metadata container) received more than 2,400 commits (a change in one or more files) and 390 different complete releases of the software (as of April 6th, 2018). Significantly, Praat's default frequency range, from 75 to 500Hz, is not suitable for an accurate analysis of infant cry because healthy infant cries vary over a frequency range 300 Hz to 600 Hz or higher for infants with developmental pathologies, such as ASD [56–58].

Appendix A.3.3. Features extraction method (FE)

Several methodologies can be employed to estimate the value of F_0 from a cry sample. For example, it can be done by direct estimation from peaks in the investigated frequency range, between about 200 to 700 Hz, or by regression from formant peaks, averaging the ratios between formant peak frequencies and their order [59,60]. Those methodologies require low levels of computations, but on the downside, their robustness to noise, especially in the frequency bands of interests, is very low. Another class of methodologies widely accepted by researchers is the estimation of F_0 and its formants using the cepstrum approach. The cepstrum is defined as the inverse discrete Fourier transformation (DFT) of the logarithmic magnitude of the DFT of a signal, causing a compression of the dynamic range and reducing amplitude differences in the formants. Algorithms based on cepstrum can separate coefficients associated with the glottal excitation and the vocal tract, and proved to be suitable for the analysis of both the adult voice and the infant cry [61]. To analyze recorded data using algorithms based on the cepstrum, preprocessing on the signal is necessary to ensure the analysis of a clean signal, and therefore a preprocessing stage is required. Since the 1960s, methodologies employed in feature extraction evolved with the development of new technologies. While initially features were estimated manually by reading the spectrogram of a signal, the advent of computers enabled the development of automatic and semi-automatic methods for feature extraction. Computer-based methodologies chain together several stages, performed one after the other. During those stages, the original signal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

undergoes a series of modifications that are reflected in obtained results (see for example Boersma [58], where all steps of an algorithm are provided and explained in detail).

Appendix A.3.4. Analysed frequency range (FR)

During data analysis, researchers focus on a specific frequency range (FR) of the spectrum, for example by digitally filtering the signals or by selecting only a subset of frequency bins after passage to the frequency domain. The selection of a specific frequency range avoids low-frequency noise and interaction between higher frequencies. In Praat, for example, it is possible to specify a frequency range in which to search for F_0. As introduced in the paragraphs above, Praat's default settings are aimed at F_0 estimation in the field of voice analysis, and therefore not suitable for its estimation on infant cry samples [58]. Details about specified parameters of commercial software or self-developed tools give a better understanding of the accuracy of extracted features.

Appendix A.3.5. Analyzed features (AF)

Analysis of infant cry can be done by using different features, both in the time and frequency domains. Tahon and Devillers [62], investigated acoustic features for emotion recognition and identified 174 different features, in time and frequency domains. Providing a list of analyzed features ensures that others can evaluate -based on investigated sample, experimental setup and methodologies employed- the accuracy of obtained results.

Appendix A.3.6. Window size of the signal during features extraction (Ww)

During feature extraction, algorithms are applied to smaller portions of the original signal, called a window. As signals are composed of waves at different frequencies, feature extraction works by investigating the repeating patterns within those windows. The size of the windows (Ww) affects the resolution of estimated features. Translation of the signal from the time domain to the frequency domain is done by dividing the spectrum into frequency bins. The width of the bins is given by the ratio between the sampling frequency of the signal and the number of time points used in the Fast Fourier Transform [63, Chapter 3]. The finer a window is, the higher is the number of possible consecutive analyzable time points, but the resolution of the signal is reduced.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

Appendix B. Supplementary Material - Checklist

Participants' information

- ☐ Number of participants: expressed as total number of participants of the study and with clear indication the the number of participants per group (if more than one group is present).
- ☐ Number of samples: expressed as total number of samples recorded and with clear indication the the number of samples per group (if more than one group is present).
- ☐ Age of the participants: statistics (mean, std, min, max) age of the participants of the study, for the whole set of participants and for the subset of participants per group (if more than one group is present). If possible, researchers should also indicate the gestational age at birth. Whenever possible, the weight of the participants should be reported as well.
- ☐ Gender of the participants: total number of male and female participants and reported per group (if more than one group is present).
- ☐ Cry Trigger: information about the trigger that has been used to induce a crying vocalizations in babies.
- ☐ Posture during the recording: information about the position of the babies during the recordings (Supine, Prone, Seated).

Data Collection

- ☐ Microphone model: the model of the microphone(s) used for recording.
- ☐ Mouth-to-Microphone Distance: distance between the infants' mouths and the microphone.
- ☐ Recording Environment: environment in which the data have been recorded (clinical or non-clinical). Additional information (e.g. was the baby familiar with the environment? Was the room soundproof and or silent?) should be reported to clarify where data have been collected.
- ☐ Sampling rate: Sampling rate of recorded signal (and resolution in bit).
- ☐ File Format: format in which the file have been saved.

Data Collection

- ☐ Preprocessing Procedure: detailed information about the preprocessing steps should be reported, included settings and parameters of employed tools and software.
- ☐ Software & Hardware: information about the software (with versions) and hardware (with model) employed in the research.
- ☐ Feature Extraction Procedure: procedures has been used to estimated analyzed features (if necessary).
- ☐ Region of Interest: regions of interest of the signals that have been processed (e.g. between 100 and 4000Hz).
- ☐ Investigated Features: list of features that have been analyzed.
- ☐ Window size: size of the windows, if any, employed in the study.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

References

1. Esposito, G.; del Carmen Rostagno, M.; Venuti, P.; Haltigan, J.D.; Messinger, D.S. Brief Report: Atypical expression of distress during the separation phase of the strange situation procedure in infant siblings at high risk for ASD. *Journal of Autism and Developmental Disorders* **2014**, *44*, 975–980.
2. Esposito, G.; Hiroi, N.; Scattoni, M.L. Cry, baby, cry: Expression of distress as a biomarker and modulator in autism spectrum disorder. *International Journal of Neuropsychopharmacology* **2017**, *20*, 498–503.
3. Esposito, G.; Venuti, P. Understanding early communication signals in autism: a study of the perception of infants' cry. *Journal of Intellectual Disability Research* **2010**, *54*, 216–223.
4. Colton, R.; Steinschneider, A. The cry characteristics of an infant who died of the sudden infant death syndrome. *Journal of Speech and Hearing Disorders* **1981**, *46*, 359–363.
5. Hariharan, M.; Yaacob, S.; Awang, S.A. Pathological infant cry analysis using wavelet packet transform and probabilistic neural network. *Expert Systems with Applications* **2011**, *38*, 15377–15382.
6. LaGasse, L.L.; Neal, A.R.; Lester, B.M. Assessment of infant cry: acoustic cry analysis and parental perception. *Mental retardation and developmental disabilities research reviews* **2005**, *11*, 83–93.
7. Lester, B.M.; Corwin, M.; Golub, H. Early detection of the infant at risk through cry analysis. In *The physiological control of mammalian vocalization*; Springer, 1988; pp. 395–411.
8. Manfredi, C.; Bocchi, L.; Orlandi, S.; Spaccaterra, L.; Donzelli, G. High-resolution cry analysis in preterm newborn infants. *Medical engineering & physics* **2009**, *31*, 528–532.
9. Michelsson, K.; Michelsson, O. Phonation in the newborn, infant cry. *International journal of pediatric otorhinolaryngology* **1999**, *49*, S297–S301.
10. Sheinkopf, S.J.; Iverson, J.M.; Rinaldi, M.L.; Lester, B.M. Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. *Autism Research* **2012**, *5*, 331–339.
11. Unwin, L.M.; Bruz, I.; Maybery, M.T.; Reynolds, V.; Ciccone, N.; Dissanayake, C.; Hickey, M.; Whitehouse, A.J. Acoustic properties of cries in 12-month old infants at high-risk of autism spectrum disorder. *Journal of autism and developmental disorders* **2017**, *47*, 2108–2119.
12. Garcia, J.O.; Garcia, C.R. Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neural networks. Proceedings of the International Joint Conference on Neural Networks, 2003. IEEE, 2003, Vol. 4, pp. 3140–3145.
13. Várallyay, G. Future prospects of the application of the infant cry in the medicine. *Periodica Polytechnica Electrical Engineering* **2006**, *50*, 47–62.
14. Etz, T.; Reetz, H.; Wegener, C.; Bahlmann, F. Infant cry reliability: acoustic homogeneity of spontaneous cries and pain-induced cries. *Speech Communication* **2014**, *58*, 91–100.
15. Wermke, K.; Mende, W.; Borschberg, H.; Ruppert, R. Changes of voice parameters and melody patterns during the first year of life in human twins. *The Journal of the Acoustical Society of America* **1999**, *105*, 1303–1304.
16. Wasz-Hockert, O. The infant cry: A spectrographic and auditory analysis. *Clinics in developmental medicine* **1968**, pp. 1–42.
17. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* **2009**, *151*, 264–269.
18. Esposito, G.; Gabrieli, G. Replication Data for: Are cry studies replicable? An analysis of participants, setups, and methods adopted and reported in cry studies, 2019. doi:10.21979/N9/UDQBEK.
19. Robb, M.P.; Crowell, D.H.; Dunn-Rankin, P. Sudden infant death syndrome: cry characteristics. *International journal of pediatric otorhinolaryngology* **2013**, *77*, 1263–1267.
20. Sahin, M.; Sahin, S.; Sari, F.N.; Tatar, E.C.; Uras, N.; Oguz, S.S.; Korkmaz, M.H. Utilizing Infant Cry Acoustics to Determine Gestational Age. *Journal of Voice* **2017**, *31*, 506–e1.
21. Borysiak, A.; Hesse, V.; Wermke, P.; Hain, J.; Robb, M.; Wermke, K. Fundamental frequency of crying in two-month-old boys and girls: do sex hormones during mini-puberty mediate differences? *Journal of Voice* **2017**, *31*, 128–e21.
22. Goberman, A.M.; Whitfield, J.A. Acoustics of Infant Pain Cries: Fundamental Frequency as a Measure of Arousal. *Perspectives on Speech Science and Orofacial Disorders, 23 (1)* **2013**, pp. 18–26.
23. Reyes-Galaviz, O.F.; Reyes-Garcia, C.A. A system for the processing of infant cry to recognize pathologies in recently born babies with neural networks. 9th Conference Speech and Computer, 2004.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1
Peer-reviewed version available at *Acoustics* **2019**, *1*, 52; doi:10.3390/acoustics1040052

24. Hawkins, D.M. The problem of overfitting. *Journal of chemical information and computer sciences* **2004**, *44*, 1–12.
25. Hurvich, C.M.; Tsai, C.L. Regression and time series model selection in small samples. *Biometrika* **1989**, *76*, 297–307.
26. Babyak, M.A. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine* **2004**, *66*, 411–421.
27. Esposito, G.; Venuti, P. Developmental changes in the fundamental frequency (f0) of infants' cries: A study of children with Autism Spectrum Disorder. *Early Child Development and Care* **2010**, *180*, 1093–1102.
28. Cacace, A.T.; Robb, M.P.; Saxman, J.H.; Risemberg, H.; Koltai, P. Acoustic features of normal-hearing pre-term infant cry. *International journal of pediatric otorhinolaryngology* **1995**, *33*, 213–224.
29. Fuller, B.F.; Horii, Y. Spectral energy distribution in four types of infant vocalizations. *Journal of communication disorders* **1988**, *21*, 251–261.
30. Sharma, S.; Asthana, S.; Mittal, V.K. A database of infant cry sounds to study the likely cause of cry. Proceedings of the 12th International Conference on Natural Language Processing, 2015, pp. 112–117.
31. Parga, J.J.; Lewin, S.; Lewis, J.; Montoya-Williams, D.; Alwan, A.; Shaul, B.; Han, C.; Bookheimer, S.Y.; Eyer, S.; Dapretto, M.; others. Defining and distinguishing infant behavioral states using acoustic cry analysis: is colic painful? *Pediatric research* **2019**, pp. 1–6.
32. Goberman, A.M.; Johnson, S.; Cannizzaro, M.S.; Robb, M.P. The effect of positioning on infant cries: Implications for sudden infant death syndrome. *International journal of pediatric otorhinolaryngology* **2008**, *72*, 153–165.
33. Goto, K.; Maeda, T.; Mirmiran, M.; Ariagno, R. Effects of prone and supine position on sleep characteristics in preterm infants. *Psychiatry and clinical neurosciences* **1999**, *53*, 315–317.
34. Chang, Y.J.; Anderson, G.C.; Lin, C.H. Effects of prone and supine positions on sleep state and stress responses in mechanically ventilated preterm infants during the first postnatal week. *Journal of advanced nursing* **2002**, *40*, 161–169.
35. Esposito, G.; Venuti, P. Comparative analysis of crying in children with autism, developmental delays, and typical development. *Focus on Autism and Other Developmental Disabilities* **2009**, *24*, 240–247.
36. Woods, J.J.; Wetherby, A.M. Early identification of and intervention for infants and toddlers who are at risk for autism spectrum disorder. *Language, Speech, and Hearing Services in Schools* **2003**, *34*, 180–193.
37. Sheinkopf, S.J.; Mundy, P.; Oller, D.K.; Steffens, M. Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders* **2000**, *30*, 345–354.
38. Möller, S.; Schönweiler, R. Analysis of infant cries for the early detection of hearing impairment1. *Speech Communication* **1999**, *28*, 175–193.
39. Etz, T.; Reetz, H.; Wegener, C. A classification model for infant cries with hearing impairment and unilateral cleft lip and palate. *Folia Phoniatrica et Logopaedica* **2012**, *64*, 254–261.
40. Beranek, L.L.; Beranek, L.L.; Beranek, L.L.; Beranek, L.L. *Acoustical measurements*; Acoustical Society of America Melville, NY, 1988.
41. Eargle, J. *The Microphone Book: From mono to stereo to surround-a guide to microphone design and application*; CRC Press, 2012.
42. Everest, F.A. Master handbook of acoustics, 2001.
43. Kuttruff, H. *Room acoustics*; Crc Press, 2014.
44. Ross, G.; Kagan, J.; Zelazo, P.; Kotelchuck, M. Separation protest in infants in home and laboratory. *Developmental Psychology* **1975**, *11*, 256.
45. Simner, M.L. Newborn's response to the cry of another infant. *Developmental psychology* **1971**, *5*, 136.
46. Sagi, A.; Hoffman, M.L. Empathic distress in the newborn. *Developmental Psychology* **1976**, *12*, 175.
47. Martin, G.B.; Clark, R.D. Distress crying in neonates: Species and peer specificity. *Developmental psychology* **1982**, *18*, 3.
48. Warner, R.M. *Spectral analysis of time-series data*; Guilford Press, 1998.
49. Cook, P.R. *Real sound synthesis for interactive applications*; AK Peters/CRC Press, 2002.
50. Robin, M.; Gutjahr, A.; Sudicky, E.; Wilson, J. Cross-correlated random field generation with the direct Fourier transform method. *Water Resources Research* **1993**, *29*, 2385–2397.
51. Watkinson, J. *Introduction to digital audio*; Focal Press, 2013.

52. Hans, M.; Schafer, R.W. Lossless compression of digital audio. *IEEE Signal processing magazine* **2001**, *18*, 21–32.
53. Laroche, J.; Dolson, M. New phase-vocoder techniques are real-time pitch shifting, chorusing, harmonizing, and other exotic audio modifications. *Journal of the Audio Engineering Society* **1999**, *47*, 928–936.
54. Smith, J.O. *Introduction to digital filters: with audio applications*; Vol. 2, Julius Smith, 2007.
55. Corwin, M.J.; Lester, B.M.; Golub, H.L. The infant cry: what can it tell us? *Current Problems in Pediatrics* **1996**, *26*, 313–334.
56. Gabrieli, G.; Leck, W.Q.; Bizzego, A.; Esposito, G. Are Praat's default settings optimal for Infant cry analysis? Proceedings of the Linux Audio Conference, 2019. LAC 2009. CCRMA, 2019, pp. 83–88.
57. Bornstein, M.; Costlow, K.; Truzzi, A.; Esposito, G. Categorizing the cries of infants with ASD versus typically developing infants: a study of adult accuracy and reaction time. *Research in autism spectrum disorders* **2016**, *31*, 66–72.
58. Boersma, P.; others. Praat, a system for doing phonetics by computer. *Glot international* **2002**, *5*.
59. Várallyay, G. SSM–A Novel Method to Recognize the Fundamental Frequency in Voice Signals. Iberoamerican Congress on Pattern Recognition. Springer, 2007, pp. 88–95.
60. Chu, W.; Alwan, A. SAFE: A statistical approach to F0 estimation under clean and noisy conditions. *IEEE Transactions on Audio, Speech, and Language Processing* **2012**, *20*, 933–944.
61. Zabidi, A.; Mansor, W.; Khuan, L.Y.; Sahak, R.; Rahman, F.Y.A. Mel-frequency cepstrum coefficient analysis of infant cry with hypothyroidism. Signal Processing & Its Applications, 2009. CSPA 2009. 5th International Colloquium on. IEEE, 2009, pp. 204–208.
62. Tahon, M.; Devillers, L. Towards a small set of robust acoustic features for emotion recognition: challenges. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* **2016**, *24*, 16–28.
63. Harris, F. Windows, harmonic analysis, and the discrete Fourier transform, Rep. *NUC TP532, Nay. Undersea Center, San Diego, Calif* **1969**.

Sample Availability: The dataset generated for this publication is available on the Data Repository of the Nanyang Technological University https://doi.org/10.21979/N9/UDQBEK[18]

doi:10.20944/preprints201911.0176.v1

Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

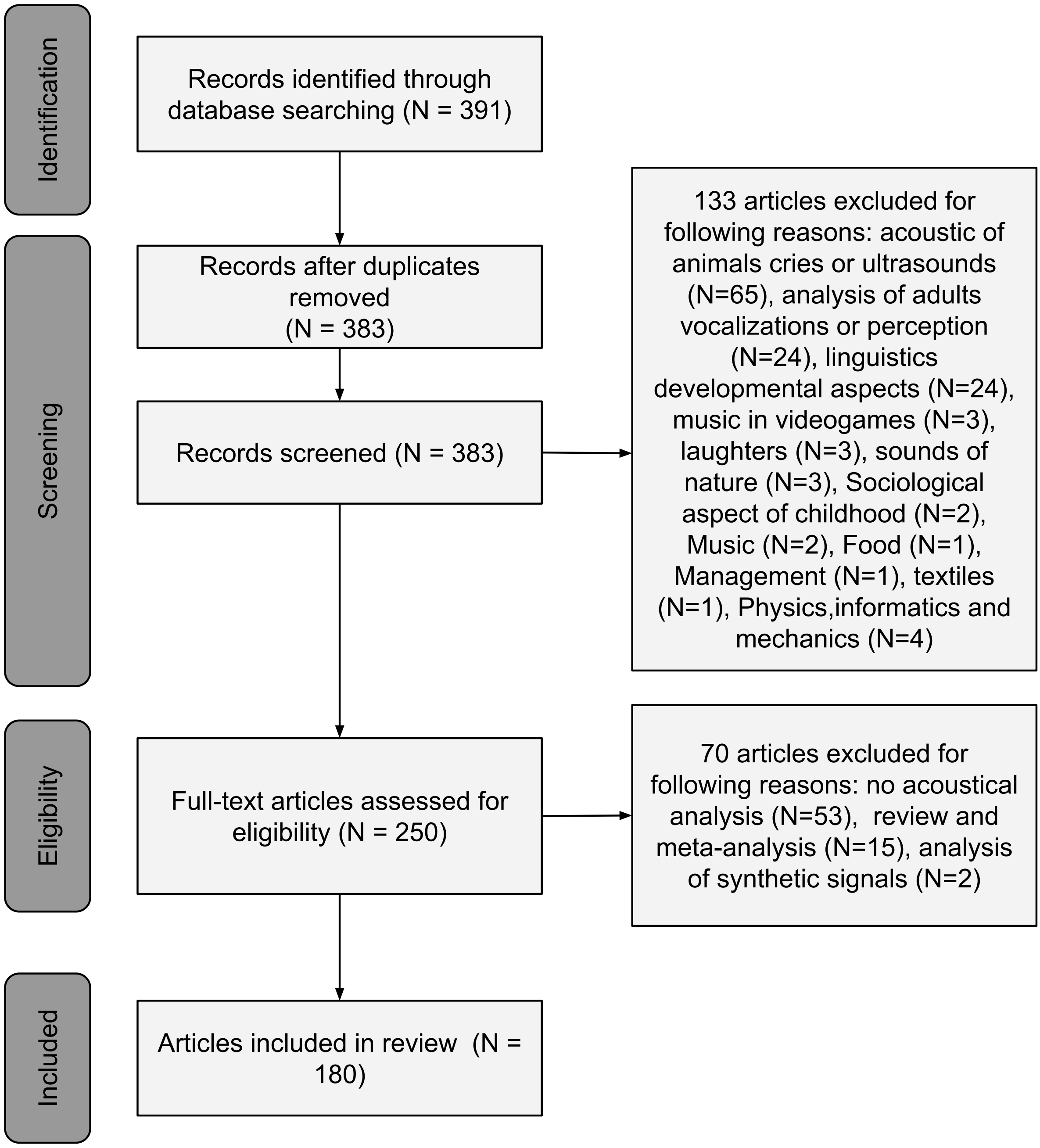

Figure 2. Article inclusion flow diagram (adapted from PRISMA[17])

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2019 doi:10.20944/preprints201911.0176.v1
Peer-reviewed version available at Acoustics 2019, 1, 52; doi:10.3390/acoustics1040052

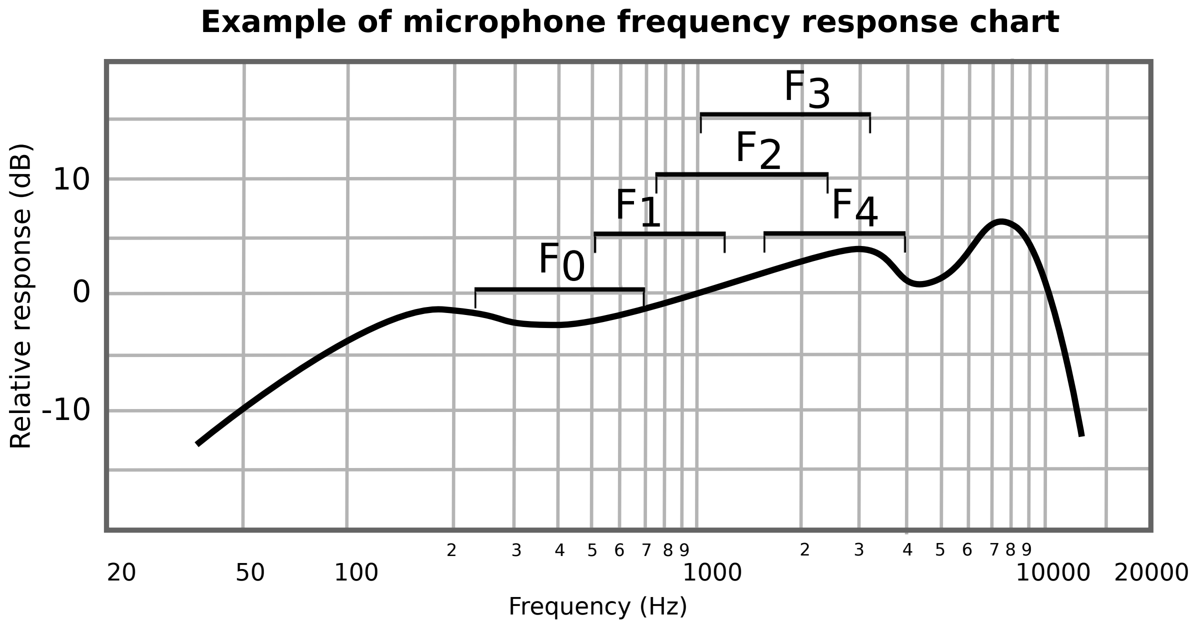

Figure A1. Example of microphone response chart. Infant cry region of interests for extraction of the fundamentals and the first four formants are highlighted.