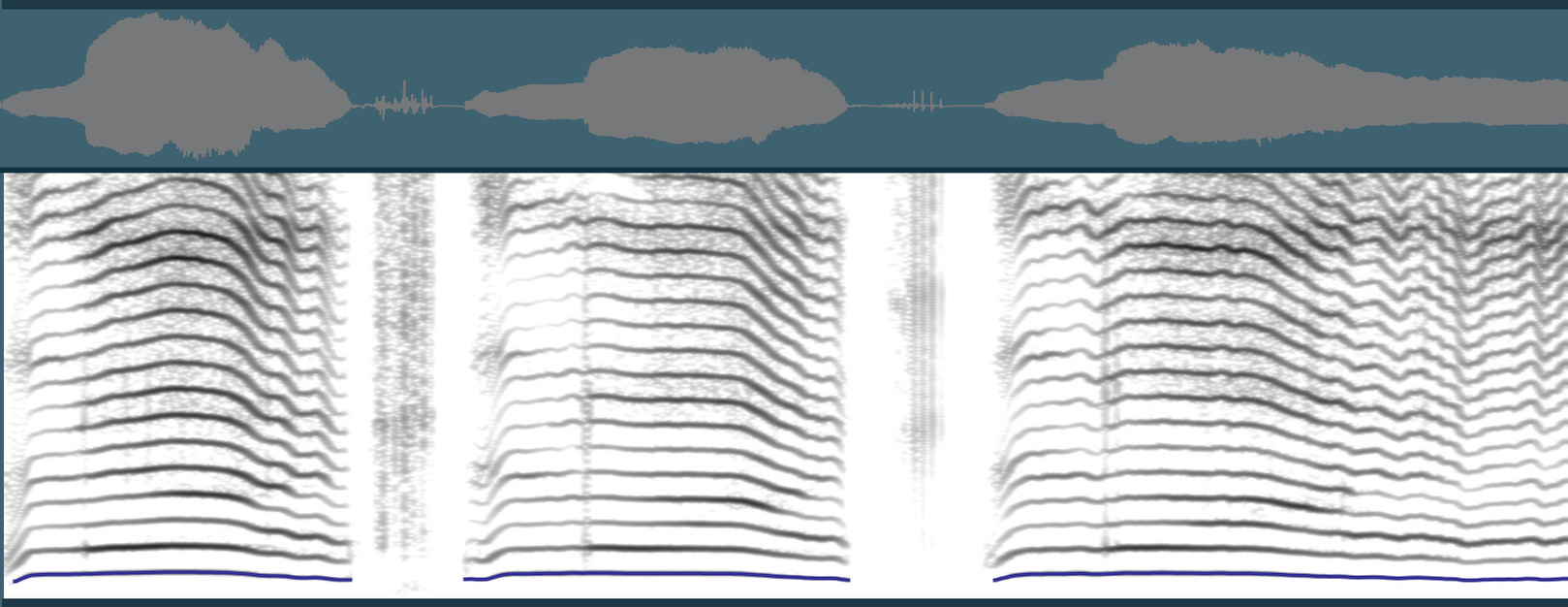


Exploring the ability of acoustic infant cry analysis for discriminating developmental pathologies

Tanja Fuhr

PhD Thesis

Johann Wolfgang Goethe-University,
Frankfurt am Main

2018

Exploring the ability of acoustic infant cry analysis for discriminating developmental pathologies

Inauguraldissertation

zur Erlangung des Grades einer Doktorin der Philosophie
im Fachbereich Sprach- und Kulturwissenschaften
der Johann Wolfgang Goethe-Universität
zu Frankfurt am Main

vorgelegt von Tanja Fuhr, geb. Etz
aus Wiesbaden

2017
(Einreichungsjahr)

2018
(Erscheinungsjahr)

Tag der Disputation: 24.07.2018

Gutachter:

Prof. Dr. H. Reetz, Goethe-Universität Frankfurt am Main
Prof. Dr. C. Wegener, Hochschule Fresenius, Idstein

Abstract

This thesis aims at exploring the ability of acoustic infant cry analysis for discriminating developmental pathologies. Cries of healthy infants as well as cries of infants suffering from cleft lip and palate, hearing impairment, laryngomalacia, asphyxia and brain damage were recorded and acoustically analyzed. The acoustic properties of the infant cries were identified and tested on their suitability to predict the health state of the infants in an reliable, valid and objective way.

To test the reliability of infant cry analysis, Krippendorff's Alpha coefficient was calculated to test how homogeneous cries of healthy infants as well as cries of infants suffering from various pathologies are.

To asses if valid methods exist for classifying infant cries, different approaches that can be used to differentiate between the groups and to predict the health state of the infants — e.g., analysis of variances, supervised-learning models and auditory discrimination by human listeners — were tested on their validity.

The objectivity of computer-based and human-based classification approaches was explored and techniques to enhance the objectivity for both approaches are proposed.

Computer-based approaches are more objective and reached higher sensitivity and specificity values in their classification to predict the health state of the infants. Especially C5.0 decision trees reached high and therefore promising classification results, even though infant cries have a great statistical spread and can be seen as very heterogeneous and are therefore not very reliable in general.

Acknowledgements

This thesis has become a noticeable part of my life. Now, after finishing this work, I'd like to thank all people who supported me during the PhD thesis. Especially, I'd like to thank and express my appreciation for my two supervisors, Prof. Dr. Henning Reetz and Prof. Dr. Carla Wegener. Without you both I would not have been able to reach this point and you excellently supported me to master this important period.

Henning, thank you for supervising this thesis. Your comments were always very helpful and I'd like to thank you for your patience and your profound knowledge in phonetics. Our meetings were always instructive and our stimulating discussions helped me a lot to widen my researches and to see them from various perspectives. Thank you for your encouragement all this time.

Carla, I would like to thank you for encouraging my researches and for allowing me to grow as a scientist. I would never have visited and participated in so many congresses without you. That helped me a lot to learn about performing and presenting scientific contents. Thank you for supporting me in recruiting subjects, especially those with diverse developmental disorders and for providing me with your wide connections to various institutions. You supported me in so many situations and your advice always helped me. Thank you for our close collaboration during this period. Without you, I would never have developed such an interest in infant cry analysis.

Further, I would like to thank all people who supported me during the recording process. Thank you, Dr. Franz Bahlmann, for supporting my work and allowing me to record infants at your institution. Thank you, Dr. Ulrike Wohlleben and Silvia Söhleman, that you enabled me to conduct my researches at your workshops. Furthermore, I'd like to thank Michaela Kreutz-Zimmermann and my colleagues for their encouragement in finding subjects for my study. You all had valuable ideas and kept your eyes peeled for potential participants for my studies.

I take this opportunity to express gratitude to all therapists, nurses, midwives, parents and students who spent time to participate in my study. Without their precious support, it would not have been possible to successfully conduct this thesis.

Last but not least, I would like to thank all infants for their crying and for the great moments I experienced through my data collections.

Contents

List of tables

List of figures

List of abbreviations

Acc	index that represents the accuracy of classification models
α	alpha
ANOVA	analysis of variances
APGAR	appearance, pulse, grimace, activity, respiration
AS	cry group “asphyxia”
BD	cry group “brain damage”
BERA	Brainstem Evoked Response Audiometry
CHAID	chi-squared automatic interaction detection trees
CLP	cry group “cleft lip and palate”
cm	centimeter
Conf	index that represents the conformability of classification models
CRT	classification and regression trees
dB	decibel
DC	direct current
DegVB	degree of voice breaks
Dur	duration
DYMP	algorithm to compute pitch marks using DYPSA with pitch-synchronous LPC coefficients for jitter estimation
DYPSA	dynamic programming projected phase-slope algorithm
F0	fundamental frequency
FFT	fast Fourier transformation
F1 – F6	formant and number of the formant
GLMM	generalized linear mixed model
g	gram
H2n	name of the recorder that was used for recoding the infant cries
He	cry group “healthy”
HI	cry group “hearing impaired”
HL	hearing level
HNR	harmonics-to-noise ratio
Hz	hertz
ICC	intra-class coefficient
Int	intensity

IQR	interquartile range
IRR	inter-rater reliability
Jitt	local jitter
KAlpha	Krippendorff's alpha coefficient
κ	Cohen's kappa coefficient
kHz	kilohertz
KNN	k-nearest neighbor
KN	cry group "known cries"
LA	cry group "laryngomalacia"
LocJitt	local jitter estimation based on the non-monotonic difference in period length
log	logarithm
LPC	linear predictive coding
max	maximum
min	minimum
m	meter
ND	cry group "non-distressed cries"
NoVB	number of voice breaks
No	number of
N	number of
$Ofit$	index that represents the overfitting of classification models
P10	10th percentile
P25	25th percentile
P75	75th percentile
P90	90th percentile
%	percent
pH	measure of the hydrogen ion concentration of a solution
PI w/o 1st.	cry group "pain-induced cries without the first cry occurrence"
PI	cry group "pain-induced cries"
PKU	phenylketonuria screening
ppq5	five-point period perturbation quotient
QUEST	quick, unbiased, efficient statistical trees
RAP	relative average perturbation
RQ	research question
R	the overall classification model's rating computed during the systematic model review
SD	standard deviation
sec.	second
Shim	local shimmer
sig.	significant
SPL	sound pressure level
SPSS	name of the statistic software that was used for computing the statistics
SP	cry group "spontaneous cries"

Std.	standard
STJE	short time jitter estimation
SVM	support vector machine
s	second
T	period
UCLP	unilateral cleft lip and palate
UKN	cry group “unknown cries”
wav	audio file format

Part I.

Introduction and research motivation

Chapter 1.

Introduction

1.1. Motivation

The cry of an infant is a form of vocalization and the first articulatory way to communicate. By crying an infant attracts attention of its parents, elicits care-giving behavior and communicates needs like hunger, discomfort or pain (Fischer, 2009; Fox, Kimmerly, & Schafer, 1991; Morsbach & Murphy, 1979; Morse, 1972; Vallotton, 2009). However, the infant cry is not only a behavior, but also an acoustic signal encoding enough information for the infant's environment to react adequately (Cecchini, Lai, & Langher, 2010; Howard, Lanphear, Lanphear, Eberly, & Lawrence, 2006; Laurent, Stevens, & Ablow, 2011; Zeskind et al., 2011). In addition, previous research indicates that pathological conditions influence the acoustic properties of infant cries (Goberman & Robb, 2005; Lester et al., 2002; Várallyay, Benyó, Illényi, Farkas, & Kovacs, 2004; Wermke, 2008; Zabidi, Khuan, & Mansor, 2012). This thesis explores the acoustic properties of infant cries in order to clarify if the acoustic information contained in the infant cries is suitable to predict the health status of an infant in an objective, reliable and valid way.

For many years, researchers have explored the infant cry from multiple perspectives: from a physiological perspective, examining how cries are produced, from a perceptive and behavioral perspective exploring how crying is perceived and how it influences the behavior of an infant's environment and from an acoustic perspective, analyzing the properties of cry signals. In all three areas of research, cries of healthy infants as well as cries of infants with various pathological health states were examined and differences between both groups were analyzed (Barr, Hopkins, & Green, 2000; de la Peña, 2007; Lester et al., 2002; Thoden & Koivisto, 1980).

From a *physiological perspective*, crying is a complex interaction of anatomic structures and physiological mechanisms which produce a crying sound. Based on a stimulus, the central nervous

system decides if crying is an appropriate answer to the stimulus and then triggers and controls a variety of muscles responsible for cry production (Bornstein & Esposito, 2014). According to the physioacoustic model of cry production of Golub and Corwin (2000), air from the lungs (subglottal system) is pressed through the glottal system during the expiratory phase. The vocal folds (glottal system) vibrate and produce an acoustic sound wave which is formed in its acoustic properties while traveling through the vocal tract (supraglottal system). In this cry production process, a variety of nerves and muscles are involved, each influencing the properties of the resulting cry sound (Verduzco-Mendoza, Arch-Tirado, Reyes-García, Leybon-Ibarra, & Licona-Bonilla, 2012).

The tension and length of the vocal folds are responsible for the pitch of the cry. Changes in the muscles' tension of the larynx (especially the m. vocalis and m. cricothyroideus) and the abdominal muscles being involved in the respiratory process, are thought to be responsible for variations in the pitch of the infant cry (LaGasse, Neal, & Lester, 2005). An atypical pitch, which can be found in infants with pathological health conditions, can also be caused by damages in the vagal nerve (Soltis, 2004). For variations in loudness, the rhythm of breathing (including the inspiratory and expiratory phase) and the coordination of breathing and crying as well as breath holding are assumed to be closely associated with the control of the lower brain stem as well as the phrenic and thoracic nerves (Lester & Boukydis, 1990). The cranial nerves VII and IX to XII (responsible for the innervation of the vocal tract, the larynx, the pharynx and the chest) also control the supraglottal systems which filter and form the cry sound (LaGasse et al., 2005).

Wermke (2008), Wermke et al. (2011), Wermke, Leising, and Stellzig–Eisenhauer (2007) found that the "melody contour" of a cry, i.e., the distribution of pitch over time, can be categorized by its shape into different patterns, ranging from simple to complex contours. Studies show that a higher age correlates with more complex melody contours (Wermke et al., 2011; Wermke, Hain, Oehler, Wermke, & Hesse, 2014; Wermke et al., 2016). The authors explain this observation with a maturation of the neuro-muscular system and therefore see crying as the preliminary stage of speech and language acquisition (Wermke, 2002; Wermke & Mende, 1992; Wermke, Mende, Manfredi, & Bruscaglioni, 2002).

From a *perceptive and behavioral perspective*, crying is seen as a psycho-acoustic effect that elicits care-giving behavior from the infant's environment. Infants' crying is an hereditary capability, essential for survival. In order to elicit different behavior from care-givers, infants are able to produce different types of cries and care-givers are able to differentiate them and react adequately (Soltis, 2004). As an example, infants use crying to establish and keep close contact to their parents (Soltis, 2004; Zeskind & Lester, 2001). Such basic behavior cannot only be observed for humans, but also for animals, especially vertebrate animals. Even these animals can produce

a special type of sounds, which indicates that the young animal is isolated or endangered to be easy prey (Newman, 2007). These so called "isolated calls" (Lester & Boukydis, 1985) elicit protecting behavior from the animal's parents to search for the young animal and protect it. Similar behavior can be observed for human infants (Newman, 2007). Other types of crying are necessary to communicate needs like hunger, pain or pleasure to their parents. It is essential for survival that care-givers are able to distinguish between different types of crying to react adequately (LaGasse et al., 2005; Schuetze & Zeskind, 2001; Schuetze, Zeskind, & Eiden, 2003).

The first studies that were concerned with classifying the infant cry into different cry types focused on the categories hunger, pain and pleasure cries (J. Lind, 1965; K. Michelsson, 1971; Rothgänger, Lüdge, & Grauel, 1990; Thoden & Koivisto, 1980; Wasz-Höckert, Lind, Vuorenkoski, Partanen, & Valanne, 1968; Wolff, 1969). Up to today, the analyses of hunger and pain cries are often conducted (Bellieni, Sisto, Cordelli, & Buonocore, 2004; de Pisapia et al., 2013; Gilbert & Robb, 1996; Mijovic et al., 2010; Runefors & Arnbjörnsson, 2005; Runefors, Arnbjörnsson, Elander, & Michelsson, 2000; M. Silva et al., 2010). K. Michelsson and Michelsson (1999) suggested to distinguish between only two types of crying: pain cries and non-pain cries, because they are clearly separable. To sum up, all types of crying which were not caused by pain, the so called spontaneous cries, were also often used as a self-contained group of crying (Shinya, Kawai, Niwa, & Myowa-Yamakoshi, 2014; Wermke, Hauser, Komposch, & Stellzig–Eisenhauer, 2002). A sub-group of the spontaneous cries, the so called non-distressed cry, is often used to analyze the contour of the fundamental frequency (Denner, 2007; Wermke, 2008; Wermke et al., 2011; Wermke et al., 2007; Wermke & Robb, 2010).

A large number of studies explored the perception of cries and showed that human listeners are able to recognize an infant's needs by its crying. Mothers seem to develop the capability to identity their own infant acoustically, when they hear infants' crying (Leerkes, Parade, & Burney, 2010; Truby & Lind, 1965). Illingworth (1955) examined this phenomenon and reported that mothers shortly after childbirth wake up when their own infant starts to cry but keep on sleeping when other infants start crying. Gustafson, Wood, and Green (2000) showed that mothers can distinguish between pain and hunger cries. Parsons et al. (2014) showed that infants' distress can be identified by their crying.

Other studies explored the effects of crying on the infant's environment. Long and excessive crying (e.g., caused by colic) leads to a deterioration of the interaction between the infant and the care-giving environment (Fairbrother, Barr, Pauwels, Brant, & Green, 2014; Frodi & Senchak, 1990; Howard et al., 2006; Raiha, Lehtonen, Huhtala, Saleva, & Korvenranta, 2002). Prolonged crying also leads to frustration of the carer and can be a cause of child abuse (Barr et al., 2014).

From an *acoustic perspective*, crying is a sound signal. It can be described by acoustic properties. Various acoustic parameters have been described for infant cries.

Manfredi et al. (2008), Scheiner, Hammerschmidt, Jürgens, and Zwirner (2002), Wermke, Mende, et al. (2002) explored the fundamental frequency (F0) of infant cries. The cry duration (Cacace, Robb, Saxman, Risemberg, & Koltai, 1995; Pinyerd, 1994) and intensity (Pinyerd, 1994), as well as formant frequencies (Fuller, 1991; Orlandi, Reyes-García, Bandini, Donzelli, & Manfredi, 2015; Robb & Cacace, 1995) were described and analyzed for infant cries. Branco, Fekete, Rugolo, and Rehder (2007) explored the harmonics-to-noise ratio (HNR), and micro variations of the vocal folds have also been considered (Lüdge & Gips, 1989; Protopapas & Eimas, 1997).

In addition to these basic parameters, acoustic properties of infant cries were described using acoustic models like linear predictive coding (LPC, (Hariharan, Chee, & Yaacob, 2010; Robb & Cacace, 1995)) or by complex coefficients like Mel-frequency cepstral coefficients (MFCC, (Galaviz & García, 2005; Reggiannini, Sheinkopf, Silverman, Li, & Lester, 2013; Zabidi, Mansor, Khuan, Yassin, & Sahak, 2010)).

Other studies explored how acoustic properties differ across various cry types. For example, pain cries have a high energy level (Thoden & Koivisto, 1980) whereas non-distressed cries have a low energy level (Wermke et al., 2011). Hunger cries can also reach high energy level (Baeck & de Souza, 2007) and are therefore comparable to pain cries.

Much research has been conducted to analyze the impact of various medical conditions on the infant cry. As described before, crying is a complex interaction of the central nervous system and many nerves and muscles; therefore, medical conditions influencing any of these structures may result in differences in the acoustic properties of the infant cry. As an example, Lederman (2010) found that the fundamental frequency as indicator for vibrations of the vocal folds is sensitive to neurological diseases in general. In many studies, the increase of the fundamental frequency was explained by disturbances of the vocal neuromuscular maturation (Corwin et al., 1992; K. Michelsson, Raes, Wasz-Höckert, & Thoden, 1981; Quick, Robb, & Woodward, 2009; Zeskind et al., 2014). Also, atypical development of nerves or damages in the innervation of the nerves (especially the vagal cranial nerve complex (LaGasse et al., 2005)), as well as the coordination among the brain regions of the brainstem and the midbrain may lead to changes in acoustic properties of an infant cry. Low birth weight (below 2500g) as well as premature birth (before the 37th gestational week) also influence acoustic parameters in terms of a higher fundamental frequency with more breaks, higher formant frequencies (especially the first formant) and shorter cry duration (Manfredi et al., 2008; Manfredi, Bocchi, Orlandi, Spaccaterra, & Donzelli, 2009; Orlandi et al., 2015; Rautava et al., 2007; Shinya et al., 2014).

Based on these assumptions, a lot of studies analyzed the cries of infants with various medical conditions. The first research in this field were performed in the 1960s by J. Lind et al. (1967), Truby and Lind (1965), Wasz-Höckert et al. (1968). They found higher fundamental frequencies for infants with medical conditions than for healthy infants. Infants suffering from asphyxia show higher F0 values and longer cry durations as well as an instability of F0 (K. Michelsson, 1971; K. Michelsson, Sirviö, & Wasz-Höckert, 1977; Partanen, Wasz-Höckert, Vuorenkoski, Valanne, & Lind, 1967). Infants with brain damage also show higher values and more instability with shifts and breaks of F0 (Accardo, 2013; Fisichelli et al., 1966; Sirviö & Michelsson, 1976; Wasz-Höckert et al., 1968). Infants with cri-du-chat syndrome (Bauer, 1968; K. Michelsson, Tuppurainen, & Aula, 1980; Sohner & Mitchell, 1991; Vassella et al., 1967) as well as infants with Hyperbilirubinemia (Koivisto, Wasz-Höckert, Vuorenkoski, Partanen, & Lind, 1970; Vohr et al., 1989; Wasz-Höckert, Koivisto, Vuorenkoski, Partanen, & Lind, 1971) also show differences in the acoustic characteristics compared to healthy infants. Cries of infants with hearing impairment (Jones, 1971; Möller & Schönweiler, 1999; Várallyay, 2007) and Krabbe's disease (Thoden & Michelsson, 1979) show higher F0 values. Studies analyzing the effect of drug exposure during pregnancy also find differences in acoustic parameters compared to healthy infants (Blinick, Tavolga, & Antopol, 1971; Corwin et al., 1992; Lester et al., 2002). The acoustic analysis of the cries of infants, being at risk to suffer from the sudden infant death syndrome revealed a longer cry duration and differences in the first and second formant, compared to healthy infants (Corwin et al., 1995; Robb, Crowell, & Dunn-Rankin, 2013). Infants with an at risk status for autism because of a familial disposition show a higher fundamental frequency and more variability in F0 (Esposito, Nakazawa, Venuti, & Bornstein, 2013; Sheinkopf, Iverson, Rinaldi, & Lester, 2012).

In addition to exploring the correlation between pathological development and acoustic parameters, studies also explored the relationship between acoustic properties and cry perception (LaGasse et al., 2005). Higher fundamental frequency values were perceived by parents as sick, urgent and aversive (Schuetze & Zeskind, 2001; Schuetze et al., 2003). More variability in the fundamental frequency was rated as sick, urgent and distressed (Protopapas & Eimas, 1997) and longer cry utterances were perceived as more distressed (Wood & Gustafson, 2001).

Summarizing, the infant cry has been analyzed from different perspectives and many studies have shown that cries from healthy infants and those from infants with medical conditions are different and can be distinguished by either acoustic analysis or by listening. For that reason, various studies suggested that the infant cry may be a powerful tool to identify pathological conditions by analyzing the acoustic properties of these cries (Golub & Corwin, 1982; Hariharan, Saraswathy, Sindhu, Khairunizam, & Yaacob, 2012; Reggiannini et al., 2013).

To use the infant cry for screening purposes, several steps are necessary: (1) an infant's crying must be recorded, (2) single cry utterances must be extracted from the recording, (3) acoustic parameters must be computed and, (4) applying statistical approaches on the acoustic parameters, the cries must be allocated to healthy or medical conditions. To implement such a screening instrument successfully, it must further fulfill three quality criteria: it must be objective, it must be reliable and it must be valid (Golden, Espe-Pfeifer, & Wachsler-Felder, 2002; Weiner, Freedheim, Graham, Schinka, & Velicer, 2003).

This thesis explored the infant cry's suitability to meet these three quality criteria.

1.2. Structure of the thesis

The remainder of the thesis is structured as described in the following.

Chapter 2 summarizes the current state of research, identifies research gaps and defines the research questions explored in this thesis. In addition, the main contributions of the thesis are highlighted and an overview about articles published in the context of the thesis is provided.

Part II introduces the theoretical background necessary for following this thesis. Chapter 3 describes the anatomy of the infantile vocal tract and the physiology of infant cry production. The influence of the infantile vocal tract on acoustic parameters is explained and various pathologies of the infants included in this thesis are specified. Chapter 4 defines a standardized procedure for recording infant cries in this thesis. Chapter 5 and chapter 6 explain how single cry utterances were extracted from the recordings and how these cries were analyzed acoustically. Finally, chapter 7 introduces statistical approaches that were used in this thesis and which go beyond the basics of statistical analysis.

The main contributions of this thesis are presented in part III. Chapter 8 analyzes the reliability of healthy infant cries and cries of infants with various pathologies. Chapter 9 uses various statistical approaches to classify infant cries according to the infants' health states and rates the validity of infant cry classification. Lastly, chapter 10 explores the objectivity of infant cry classification by humans and by computational models.

Part IV summarizes and discusses the overall findings of the thesis.

Chapter 2.

Research scope

2.1. Research gaps, research aim, objectives and questions

As described in section 1.1, the infant cry contains a lot of information. Previous research found the infant cry to be categorizable into different kinds of cry types which are auditorily distinguishable, in order to alert care-giving behavior and to contain different acoustic information indicating the health status of the infant. However, previous research did not analyze the infant cry from a screening instrument's perspective comprehensively. Only few researches tried to solve the question if the infant cry is capable of meeting three main quality criteria of screening instruments (Vidakovic, 2011): reliability, validity and objectivity.

The research scope of this thesis is defined by postulating a main *research aim*, i.e., the main topic the thesis deals with. *Research objectives* are defined as coarse-grained goals the thesis wants to achieve. For each research objective *research questions* are defined that are answered by this thesis in order to achieve the research objectives (Thomas & Hodges, 2010).

The research aim of this thesis is to explore the infant cry's potential to be used in screening instruments. To provide new insights for this aim, three research objectives were defined for this thesis:

1. To assess the reliability of infant cries.
2. To assess the validity of infant cries.
3. To assess the objectivity of infant cries.

For each research objective, the research gap in the current state of research is described and research questions that must be answered to achieve the research objective are identified in the following.

2.1.1. Infant cry reliability

In the context of infant cry analysis, *reliability* means that different cry utterances of one infant are rated equally. As infants produce many cries it is interesting to explore if these cries are similar to each other, or not. If the cries of one infant differ vastly from each other, it will be difficult to develop a screening instrument that is able to predict the same result for all cries. Therefore, the "reliability" of the infant cry directly influences the reliability of the screening instrument. In the past, different types of crying have been used in infant cry analysis, e.g., pain cries or spontaneous cries. However, no research was conducted to identify which type of crying is the most reliable one and is therefore suited best for developing a screening instrument. This research gap rises the following question:

Research Question 1

Which type of crying is suited best for the analysis of infant cries and is therefore the most reliable one?

This aspect is very important for screening instruments in order to get repeatable and reliable test results.

2.1.2. Infant cry validity

Validity can be defined in the context of infant cry analysis as the accuracy of the screening result. Here, two aspects are important: the specificity and the sensitivity of the ratings. The specificity defines how well healthy infant cries are recognized as healthy and the sensitivity defines how well pathological cries are recognized as pathological. In addition to classifying pathological cries as pathological, it is important to determine if the kind of medical condition can be identified correctly.

In this thesis, three approaches that have been used in infant cry research are analyzed regarding their validity: two statistical classification approaches using analysis of variances and supervised-learning models, as well as humans rating the health state of infants by listening to their crying.

Infant cry classification using analysis of variances

Many researchers previously identified differences in single acoustic parameters between cries of healthy infants and infants with various pathologies. Examples are Arch-Tirado et al. (2004), Barr et al. (2000), Boero, Weber, Vigone, and Lenti (2000), Jones (1971), Möller and Schönweiler (1999). In these studies, the analysis of variances was used in order to explore if single acoustic parameters are significantly different for healthy and pathological infants. Based on these differences, the authors often claimed that the infant cry is suited for developing screening instruments. However, acoustic parameters like the fundamental frequency were found to be influenced by different pathologies and even by non-pathological conditions like low birth weight (Shinya et al., 2014) in a very similar way.

Previous research did not explore if differences in single acoustic parameters between healthy and pathological cries are still significant when multiple pathologies are included in the analysis. For developing a screening instrument, it is important to clarify if differences in acoustic parameters between healthy and non-healthy cries are a specific indicator for the pathology, leading to the following research question:

Research Question 2

Are differences in single acoustic parameters found by analysis of variances specific for certain developmental pathologies?

Infant cry classification using supervised-learning models

Using multivariate techniques (Jambu, 1991) that consider more than one acoustic parameter at once when testing the differences between infant groups has been proposed in literature, too. Here, different approaches for classifying infant cries according to their health status have been explored, e.g., by Hariharan, Yaacob, and Awang (2011), Lederman, Zmora, Hauschildt, Stellzig–Eisenhauer, and Wermke (2008) and Saraswathy, Hariharan, Nadarajaw, Khairunizam, and Yaacob (2014).

All kinds of classifiers are trained on a training dataset for which the health status of the infants is known. The different research groups all used their own training dataset and therefore the different approaches are not comparable to each other. In addition, similar to the ANOVA approaches, most research did not consider multiple different pathologies when classifying the cries. For developing screening instruments, an objective comparison of the different approaches is necessary to answer the following question:

Research Question 3

Which classification technique is suited best for discriminating cries of healthy infants and cries of infants with different pathologies?

Infant cry classification by human listeners

In addition to classifying infant cries by using statistical approaches, the validity of the human intuition to rate the health state of infants by listening to their crying is explored in this thesis, too. Human listeners are taken into account because the mathematical approaches might not be able to identify all kinds of relationships or patterns encoded in the infant cry. The human brain is able to relate changes in acoustic parameters to the health status of an infant (Protopapas & Eimas, 1997; Schuetze & Zeskind, 2001; Schuetze et al., 2003) and may have the better ability to identify relationships between the infant cry and certain pathologies. For this reason, previous research also conducted listening experiments to analyze if human listeners are able to *hear* if a cry belongs to a healthy infant or an infant with a pathology (Möller & Schönweiler, 1999).

However, the previous literature did not explore if listeners are able to distinguish between healthy and non-healthy cries when cries with multiple different pathologies were presented to them. In addition, no research was yet conducted to explore if human listeners are able to identify different pathologies after a listening training.

For developing screening approaches based on human listeners, the following research questions must be answered:

Research Question 4

Are human listeners able to auditorily discriminate between healthy infant cries and non-healthy infant cries and are they able to differentiate between different pathologies?

2.1.3. Infant cry objectivity

For infant cry analysis, *objectivity* means, all steps in a screening process must be independent from the person conducting the screening: (a) The recording of the cry samples must be independent (implementation objectivity), (b) the acoustic analysis and the rating of the cries must be independent (application objectivity) and (c) the interpretations of the results must be independent (interpretation objectivity). Although the previous literature did not put their focus on objectivity in infant cry analysis, most studies achieved objectivity “by accident”. Implementation objectivity

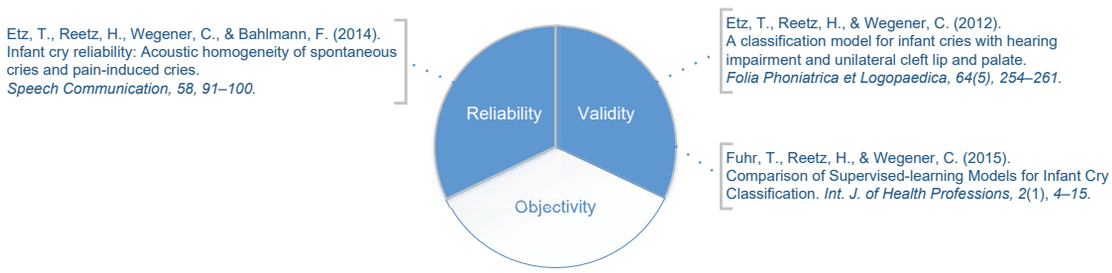


Figure 2.1.: Quality criteria for screening instruments and articles published as part of this thesis, associated to the criteria

was achieved by standardizing the recording of infant cries to ensure that all cry recordings were conducted under the same conditions, independent of the person taking the recordings. Application objectivity is naturally given when the acoustic analysis of infant cries is automated and is executed without human interaction. Interpretation objectivity can be achieved for computational models when using classification techniques that classify infant cries into disjoint groups. For classifying infant cries by "non-machine approaches", thus by human listeners, the interpretation objectivity must be considered. In contrast to automatic classification models, different human listeners may rate the same infant cries with varying results. For this reason, the following research question is formulated:

Research Question 5

How objective are human listeners' ratings of infant cries when classifying the cries by the infant's health status?

The research questions formulated in this section will be answered in part III.

2.2. Contribution of the thesis

In this section, the novelty and the contribution of the thesis for infant cry research are described. Articles that were published during the thesis are mentioned. Figure 2.1 provides an overview over the articles that were published and the research objectives they covered.

2.2.1. Contribution to assessing the infant cry reliability

The thesis examines how similar various cry types are and which cry type is the most reliable one and therefore suited best for infant cry analysis (chapter 8). Previous research often assumed that spontaneous cries are more inhomogeneous and have a greater variance than pain cries, because the cause of crying can have multiple reasons. The thesis suggests that not the same cause of crying provides the most reliable cry types (for example crying caused by a painful stimulus). Rather, a cry type that was used so far for analyzing the melody contour only, the non-distressed cry, was found to be the most reliable one by trend for healthy infants. Up to today, the non-distressed cry is rarely used for the analysis of acoustic parameters in order to distinguish between healthy and pathological cries.

Although the non-distressed cry was found to be the most reliable one by trend, the thesis also shows that infants suffering from certain pathologies do not all produce non-distressed cries. For this reason, spontaneous cries, which are the second-best type of crying regarding their reliability, are proposed to be used in future infant cry research and to be best-suited for developing screening instruments.

These findings have been published in

- Etz, T., Reetz, H., Wegener, C., & Bahlmann, F. (2014). Infant cry reliability: Acoustic homogeneity of spontaneous cries and pain-induced cries. *Speech Communication*, *58*, 91–100. doi:10.1016/j.specom.2013.11.006

2.2.2. Contribution to assessing the infant cry validity

This thesis provides new insights into the ways of classifying infant cries according to the infants' health states in a valid way (chapter 9).

One common method in infant cry research was to identify differences between acoustic parameters of healthy and pathological infant cries using the analysis of variances. This thesis shows that univariate ANOVA of single acoustic parameters cannot be used for developing a screening instrument. Differences in acoustic parameters that had been found in previous research were found to be not specific enough for certain pathologies in this thesis (section 9.1); percentile ranges of one acoustic parameter can overlap. Even if the medians of two cry groups are significantly different, the cries can be too similar to be clearly separated by one acoustic parameter only. Therefore, ANOVA is not suited for screening instruments.

Different multivariate approaches, especially various supervised-learning classification approaches like neural networks, have been used to classify infant cries. Mostly, one pathological cry group and a group of healthy cries were included to calculate a classification model. As described in section 9.2, this thesis applied different classification models that have been used in previous research to one reference dataset of infant cries (including healthy cries and five pathological cry groups). The thesis developed a rating scheme which allowed to objectively evaluate the performance of the classification models in infant cry classification. The results show that models that have often been used in previous research (e.g., neural networks) perform not very well when multiple pathological cry groups are included in the classification. In contrast, alternative classification models that have not widely been used in infant cry analysis achieved promising results indicating that these models should be paid more attention when developing screening instruments.

The findings regarding the validity of infant cry classification have been published in

- Etz, T., Reetz, H., & Wegener, C. (2012). A classification model for infant cries with hearing impairment and unilateral cleft lip and palate. *Folia Phoniatrica et Logopaedica*, *64*(5), 254–261. doi:10.1159/000343994
- Fuhr, T., Reetz, H., & Wegener, C. (2015). Comparison of supervised-learning models for infant cry classification. *International Journal of Health Professions*, *2*(1), 4–15. doi:10.1515/ijhp-2015-0005

2.2.3. Contribution to assessing the infant cry objectivity

Not much research has been conducted yet to analyze if human listeners are able to identify the health status of infants based on their crying. Previous studies only explored if listeners can distinguish between healthy and non-healthy cries, without having any previous training on how to distinguish cries by listening. In contrast, listeners in this thesis attended a listening training to learn how cries from healthy infants and cries from infants with five different pathologies sound (chapter 10). This training approach was used to enhance the objectivity of infant cry classification by human listeners. The listening experiment in this thesis analyzed how listeners perform in distinguishing healthy from non-healthy cries as well as if they are able to distinguish the five different pathologies from each other. The classification's objectivity was tested by analyzing how multiple listeners rate the same cries. The results show that healthy cries could be distinguished from non-healthy cries, even when multiple pathologies are presented in the listening experiment. However, listeners did not perform very well in distinguishing the pathologies from each other.

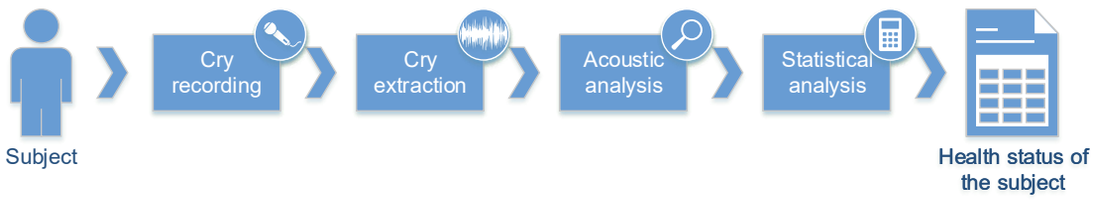


Figure 2.2.: Steps of a screening process based on infant cry analysis

Consequently, listening performance gives a hint that a cry is a pathological one. But, using listening performances, the kind of pathology cannot be identified in an objective way.

2.3. Steps of a screening process

A screening process to determine the health status of an infant by analyzing the cry consists of several steps as shown in figure 2.2. Cries of healthy infants as well as cries of infants with pathological health status are recorded. The single cry utterances are extracted from the recording and are analyzed to compute the acoustic parameters characterizing the cry signal. With statistical methods, the infant cries are classified into the healthy or into a pathological cry group.

In the following chapter, a brief theoretical overview for each step is provided, starting from the theory of infant crying up to the statistical analyses used in this thesis.

Part II.

Foundations

Chapter 3.

Theory of infant crying

When analyzing infant cries, several anatomical aspects differ from adults' anatomical structures and can influence the acoustic analysis process. They are described in the following section. Section 3.2 describes inclusion and exclusion criteria of healthy infants and infants with pathological disorders contained in this thesis. Section 3.3 introduces a physioacoustic model for infant cry production that is used as theoretical basis to select acoustic parameters for infant cry classification.

3.1. Infants' anatomical conditions influencing the cry analysis

When comparing the vocal tract of infants and adults, the infants' vocal tract cannot be seen as a smaller version of the adults' vocal tract (Jadcherla, Hogan, & Shaker, 2010). Single anatomic components differ in their size, their constitution and their position (Prakash & Johnny, 2015). This leads to differences in the acoustic properties of the vocal tract compared to the vocal tract of adults as explained in the following.

First, an important difference is the position of the larynx. The larynx of infants has a higher position (positioned at the second cervical vertebra) and the tip of the epiglottis lies at the first cervical vertebra and is close to the velum (Prakash & Johnny, 2015). This construction allows an infant to simultaneously breathe and drink (Prochnow, 2013).

The pharynx is, compared to the total length of the vocal tract, relatively short with a proportion of only one-third of the vocal tract (Sapienza, Ruddy, & Baker, 2004). The way of the air through the vocal tract also varies: the infantile airway has a more gradual bend instead of a right-angled bend in an adult's airway (Fischer, 2009). The tongue is proportionately larger and less maneuvrable.

The larynx and also the vocal folds show different structural characteristics. The cartilaginous percentage is about 50-75 % compared to the membranous percentage (Eckel et al., 2000). Adults have a cartilaginous percentage of about 30 % compared to the membranous part (Tucker, 1993). These factors influence the transfer of vibrations.

The histological structure of the vocal folds are not matured completely during infancy. The Lamina propria, the membranous part between the epithelium and the m. vocalis cannot be differentiated into three parts (upper, middle and lower part) as seen in the tissues of adults (Sato, Hirano, & Nagashima, 2001). These monomorph tissues directly influence the vibration properties of the vocal folds by not being as flexible as the vocal folds of adults (Sato et al., 2001).

Furthermore, the ribs of newborns are perpendicular to the spine. Hence, infants are not able to control the subglottal air pressure like adults do (Fischer, 2009).

Summing up, the infantile orofacial, pharyngeal and laryngeal systems differ in their physiological properties from adults' properties. For this reason, acoustic analysis techniques may not be interpreted in the same way as done for adults. In the following section, health states of the subjects recorded in this thesis as well as inclusion and exclusion criteria are described.

3.2. Health states of the recorded subjects

Cries of healthy infants as well as cries of infants suffering from hearing impairment, unilateral cleft lip and palate, laryngomalacia, asphyxia and brain damage were recorded and analyzed for this thesis. Inclusion and exclusion criteria that applied to selecting appropriate subjects for the studies are described in the following.

All the infants were up to 7 months of age. All the infant's parents were native speakers of German to ensure that all the infants cry with the same prosody. Wermke and Mende (1992) described that German infants cry with different melodic structures than, for example, French infants. German infants show more cries with a tendency of a falling intonation structure whereas French infants show more cries with a tendency of a rising intonation structure. To ensure that this will not influence the calculation of the acoustic parameters, only infants with native speaking parents of German were included.

For the infants suffering from one of the pathologies included in the studies, it was ensured that the infants did not suffer from additional pathologies, i.e., combined pathologies must not exist.

In the following sections, the various health states of infants included in this thesis are briefly introduced and inclusion and exclusion criteria that are specific for the groups are defined. In addition, reasons for including the pathological pictures in the studies are provided.

3.2.1. Healthy infants

To ensure that infants were healthy, no incidence of complication during birth was allowed. They were found to be healthy by paediatricians at postpartum examination. The gestational age as well as the birth weight were without pathological findings. Thus, infants had to be born between the 37th and 42nd week of gestation. A birth weight of 2500 g as minimum and 4500 g as maximum were seen as normal. APGAR scores (a rating scheme to asses the health condition of newborn infants in a standardized way (Apgar, 1953)) were without pathological findings (i.e., the APGAR score was 10 at each time directly after birth, after 5 minutes and 10 minutes). Infants must not have anomalies or adumbration of neurological diseases or any diagnosis that might influence normal development. Furthermore, the hearing function had to be normal, tested by otoacoustic emissions or evoked brainstem response.

3.2.2. Infants with hearing impairment

Infants included in this group suffered from conductive deafness (i.e., the sound is not conducted efficiently to the inner ear (Stenton, 2010)) or sensorineural hearing loss (damages of the inner ear (Editore, 2014)). All infants were examined with the *Brainstem Evoked Response Audiometry (BERA)* to confirm the hearing impairment and to identify the hearing level of the infants. The Brainstem Evoked Response Audiometry is an objective method of hearing assessment in infancy and detects electrical activity from the inner ear to the so called inferior colliculus (Arruda, Dell' Aringa, Dell' Aringa, Esteves, & Nardi, 2009), a region of the brain which receives information about the auditory pathway. The electrical response from the brainstem is detected by electrodes placed on the scalp of the infant.

The hearing impairment was also confirmed by phoniatrists in an audiologic hearing evaluation. All infants included in this study had a bilateral hearing impairment with a hearing threshold of minimum -60 dB. A hearing loss of an about 60 dB hearing level can be allocated to a severe impairment (Clark, 1981). The incidence of suffering from hearing impairment is indicated with about 1.8 - 2.0 occurrences per 1000 infants (Bielecki, Horbulewicz, & Wolan, 2012; Haghshenas et al., 2014).

Analyzing cries of infants with hearing impairment is interesting because infants suffering from hearing impairment have limitations of their own auditory feedback. Based on this assumption, cries of infants with hearing impairment are assumed to differ from the cries of healthy infants in acoustic parameters. Möller and Schönweiler (1999) as well as Jones (1971) confirmed this assumption and found a longer cry duration for infants suffering from hearing impairment compared to healthy infants.

Up to today, the hearing function of infants is screened during the Newborn Infant Hearing Screening in the first three days of life. Common screening techniques are mostly the measurement of the otoacoustic emissions (vibrations of the outer hair cells are measured, when the inner ear is stimulated by a sound (Mühler & Hoth, 2014)). When the measurement of the otoacoustic emissions failed, Brainstem Evoked Response Audiometry is often used to confirm the results of the otoacoustic emissions (White et al., 2005). Are the findings confirmed by the Brainstem Evoked Response Audiometry, additionally a phoniatrist examines the hearing function and also determines the hearing level. According to the American Speech-Language-Hearing-Association (2015), "failing the hearing screening does not necessarily mean that the baby has a hearing loss. Not all babies pass the hearing screening the first time. Infants who do not pass a screening are usually given a second screening to confirm the findings." Getting false positive results in screening the hearing function is thus possible. Considering the sensitivity of the two approaches, the sensitivity of the otoacoustic emissions range from 80 % to 98 % and for the Brainstem Evoked Response Audiometry the sensitivity range between 84 % to 90 % (Farhadi, Mahmoudian, Mohammad, & Daneshi, 2006).

When analyzing the cries of infants with hearing impairment, it will be of interest how sensitive cry analysis is for detecting hearing impairment.

3.2.3. Infants with cleft lip and palate

For the group of infants having a cleft lip and palate, only infants having a unilateral cleft lip and palate (UCLP) were included. A cleft lip and palate is a fissure of the oral-pharyngeal structures caused by a non-fusion of these structures antenatal (Wyszynski, 2002). A unilateral cleft lip and palate is a complete fissure with a cleft in the lip on only one side of the face. The hard and soft palate are also split. Different forms of the cleft lip and palate exist. Other forms like isolated lip palate are not considered in this thesis and therefore not explained in detail. Further information about the single types can be found in Berkowitz (2013). The incidence of a form of the cleft lip

and palate is about 1.94 per 1000 (Berkowitz, 2013). It is one of the most hereditary disorders in infancy.

For infants included in this thesis, no surgery had been done at the time of the recording to correct the physical defect caused by the cleft lip and palate. For the recording no feeding plate was inserted.

Only complete clefts, like the unilateral cleft lip and palate were included, because isolated gaps like cleft lip influence the vocal tract less than a complete cleft.

The cleft lip and palate was included in the study to analyze which acoustic parameters will be influenced by the pathology and therefore differ from healthy infants. It can be assumed, that formant frequencies are influenced by a unilateral cleft lip and palate, because the anatomical defects affect the vocal tract, where formant frequencies result from. Because the cleft lip and palate is the only developmental disorder which originates from malformation of the oral-pharyngeal system included in this thesis, the impact on the acoustic characteristics of this disorder might differ in its influence on the acoustic parameters compared to the other developmental disorders.

Prior research focuses on deviations of F0 in infants with cleft lip and palate like K. Michelsson, Sirviö, Koivisto, Sovijarvi, and Wasz-Höckert (1975). They analyzed 13 infants with cleft lip and palate and could not find differences concerning the mean fundamental frequency or the maximum pitch. They also considered the melody types and found 88 % agreement between the melody contour of healthy infants and infants with cleft lip and palate. Raes, Michelsson, Dehaen, and Despontin (1982) confirm these findings. Further, little research has been conducted about the cries of infants with cleft lip and palate. Much later, Wermke et al. (2011) confute these results because they found differences in the melody development compared to healthy infants. Wermke, Hauser, et al. (2002) also found differences in F0 of infants with cleft lip and palate compared to healthy infants. Because K. Michelsson et al. (1975) considered only a few acoustic parameters, acoustic parameters like formant frequency become interesting to analyze for infants with cleft lip and palate to identify acoustic parameters correlating with the vocal tract malformations.

3.2.4. Infants with laryngomalacia

Laryngomalacia is caused by a hereditary softening of the larynx (Koitschev & Sittel, 2012). The tissue of the larynx is softer in infants with laryngomalacia than in normal developed infants. Especially the supraglottic tissues of the larynx collapse during inspiration (Dobbie & White, 2013). This congenital disease is often recognizable by a stridor (an audible noise during inspiration)

(Reinhard & Sandu, 2014). These symptoms mostly disappear spontaneously during the first two years of life (Dobbie & White, 2013). The epiglottis is longer and the aryepiglottic folds are shorter compared to healthy infants (Ayari, Aubertin, Girschig, Van Den Abbeele, T., & Mondain, 2012). The severity of laryngomalacia can be divided into three groups: a mild, moderate and severe form. The real incidence of this clinical picture is difficult to determine because mild forms of laryngomalacia are often undetected and laryngomalacia can also be part of a syndrome, e.g, the down syndrome or charge syndrome (Ayari et al., 2012).

All infants in this thesis were classified having a moderate form of laryngomalacia (the arytenoids collapse during respiratory). The laryngomalacia was confirmed by pediatricians through laryngoscopy and the risk on paresis of the vocal folds were suspended.

Infants suffering from laryngomalacia were included in the study because the congenital diseases directly affect the larynx. Considering infants suffering from laryngomalacia is relevant because the voice signal is directly influenced by anatomical alterations in terms of the softening of the tissue of the larynx and not by deficits of the neuromuscular control of the larynx caused by cranial nerve damages. In this thesis, it is explored if laryngomalacia can be clearly separated from the cries of infants with neuromuscular deficits (like infants suffering from asphyxia or brain damage), which are assumed to have similar pathological deviations in acoustic parameters like instability and shifts in the fundamental frequency (Hariharan, Saraswathy, et al., 2012; Raes, Michelsson, & Despontin, 1980), but the deviations are caused by neurological and not by anatomical alterations.

3.2.5. Infants with asphyxia

In this group, infants suffering from asphyxia were included. Asphyxia is an oxygen deficiency and can lead to organ damage and hypoxic ischaemic encephalopathy (a brain injury due to asphyxia) (Radulova & Slancheva, 2014). Asphyxia can occur because of complications during pregnancy or childbirth with a lack of oxygen for the infant. A prediction of the outcome of asphyxia is difficult and infants suffering from asphyxia can develop neurological damage which can influence for example the psychomotoric development or may lead to an increased risk to develop epilepsy (Allemand et al., 2013). Birth asphyxia (originate prenatal and perinatal) causes about 22 %-24 % of all neonatal deaths (Ferriero, 2004; Grow & Barks, 2002; Perlman, 2004).

Infants included in this thesis had an acute neurological manifestation after birth, e.g., hypotonia or convulsions, and applied to all of the following criteria for asphyxia (Committee on Fetus and Newborn American Academy of Pediatrics, 1996): (1) a profound metabolic or respiratory

acidemia (pH < 7.00) on an umbilical arterial blood sample existed, (2) the infant had an APGAR score ≤ 3 for longer than five minutes and (3) evidence was found for a multiorganic dysfunction. No indication for further developmental pathologies, like syndromes, existed for this cry group.

Studies examining the cries of infants suffering from asphyxia found that they had abnormal cry characteristics compared to healthy infants (Hariharan, Saraswathy, et al., 2012; K. Michelsson, 1971; K. Michelsson et al., 1977; Verduzco-Mendoza et al., 2012). Infants suffering from asphyxia usually have high F0 values (K. Michelsson, 1971; K. Michelsson et al., 1977). Deviations in the cry characteristics go back to neurological damage. A considerable increase of the fundamental frequency is often assumed to be an indicator for a neurological impairment.

3.2.6. Infants with brain damage

The group of infants with brain damage includes two types of brain hemorrhage. One type is a subarachnoid hemorrhage and one type is an intraventricular hemorrhage. This cry group is more heterogeneous than the other groups because the hemorrhages did not occur in exact the same brain region. However, disorders like brain hemorrhage are always a heterogeneous disorder because the region where the hemorrhage occurs, as well as the size and the reason of the hemorrhage influence the degree of the dysfunction of brain regions.

The two infants included in this study had the same cause of their clinical picture: both developed a brain hemorrhage due to complications during the childbirth and were examined with the Magnetic Resonance Imaging (MRI), a medical imaging technique to visualize anatomical structures that allows to identify the localization of the brain hemorrhage. The severity of the hemorrhage was classified as III according to the classification scheme of brain hemorrhage proposed by Papile, Burstein, Burstein, and Koffler (1978).

3.2.7. Influence of age on the acoustic properties of infant cries

Infants included in this thesis and described above were up to 7 month of age. Because different opinions and research results about the influence of age on acoustic parameters exist, this chapter analyzes the existing studies about the influence of age on cry properties.

Regarding the influence of an infant's age on acoustic parameters, especially the fundamental frequency, studies came to contradicting results. Some studies found that the age of infants has

a significant impact on the fundamental frequency, others came to the conclusion that there is no impact on the fundamental frequency at all.

Acoustic analyses conducted in the sixties by Prescott (1975) found an increase in the fundamental frequency correlated to the growth of infants by analyzing hunger cries of infants with up to 9 months of age. He described increasing values of F0 from 556 Hz to 640 Hz. Also, hunger cries were analyzed by Gilbert and Robb (1996) within the 12 first months of age on 4 infants. They also reported an increasing of F0 and described values between 429 Hz and 527 Hz for the fundamental frequency. K. Lind and Wermke (2002) examined the cries of 1 infant over the first 3 months of life and reported an increase of F0 from 399 Hz up to 411 Hz. An increase of F0 from 317 Hz to 338 Hz was reported by Laufer and Horii (1977), analyzing 4 infants within the first 12 weeks of life. All studies interpreted the increasing of F0 with an age-related increase in control of the laryngeal system and the voice production.

In contrast, Colton, Steinschneider, Black, and Gleason (1985), Ray D. Kent (1976) as well as Sheppard and Lane (1968) described a decrease of F0 in infant cries during the first year of life on no more than 5 subjects. They explained their findings of a decreasing F0 value with an increasing age and by a general growth of the infant, especially a growth of the laryngeal system, including an increase of length and thickness of the vocal folds.

K. Michelsson, Eklund, Leppänen, and Lyytinen (2002) analyzed the cries of 172 infants that were 1 to 7 days old and found no significant differences in the mean of the fundamental frequency related to the age or gender of the infants.

A study from Baeck and de Souza (2007) analyzed the development of the fundamental frequency of healthy infant cries over the first 6 months of life. Altogether, cries of 30 infants were recorded in a biweekly interval and the correlation between the age and the values of F0 were analyzed. The fundamental frequency ranged from 380 Hz to 435 Hz. They calculated the Pearson correlation coefficient. The correlation between age and the fundamental frequency was 0.182. This can be interpreted as a slight agreement between the age and the fundamental frequency.

Comparing the above mentioned studies, most studies that found correlations between the fundamental frequency and the infants' age included only a very limited number of subjects. For example, K. Lind and Wermke (2002) analyzed only 1 infant in their study and Gilbert and Robb (1996), Prescott (1975) examined 4 and Gilbert and Robb (1996), Prescott (1975) scrutinized 5 infants. In contrast, the sample size in the study by Baeck and de Souza (2007) can be seen as more representative with a sample size of 30 infants. This study found no significant differences in F0 related to the age of infants within the first six months of life.

Considering the sample sizes of the studies, the assumption of Baeck and de Souza (2007) (the fundamental frequency does not vary significantly within the first six months of life) is followed in this thesis. For this reason, infants with an age between 0 and 7 months are included.

3.3. A physioacoustic model of the infant cry

To determine which acoustic features are suitable to model the specifics of infant cry characteristics, a physioacoustic model of Golub and Corwin (1982, 2000) shown in figure 3.1 is used to describe the anatomical conditions which influence acoustic features of the infant cry.

The model divides the process of initiating any muscle control into three levels of processing within the central nervous system, called the upper processor, middle processor and lower processor.

The *upper processor* describes cortical functions for processing extrinsic or intrinsic impulses, e.g., proprioceptive or auditory impulses. It is seen as a "control structure", receiving internal and external information, assessing them and initiating corresponding actions.

The *middle processor* controls the vegetative states and vegetative functions like respiration, swallowing or crying. Golub and Corwin (1982) describe this processor as a very autarkic system triggered by an infant's "survival pressure". When a vegetative state involves crying (as is the case for hunger or pain), the middle processor triggers the lower processor that is responsible for controlling the muscle groups involved in crying.

The *lower processor* is seen as a "coordinative structure" that controls groups of related muscles. The model assumes that the muscle groups are coordinated independently from each other and without the need of control by the upper and middle processors. For crying, three muscle systems are described by the model, the subglottal system, the glottal system and the supraglottal system.

The *subglottal system* contains the muscles of the lungs, e.g., the diaphragm, the intercostals or the abdominals. The *glottal system* represents the larynx and covers, amongst others, the vocalis and the cricothyroideus. The *supraglottal system* stands for the vocal tract and contains the velar, jaw, tongue, lips and others.

Following this model, an infant cry is seen as a reaction to intrinsic or extrinsic impulses, initiated by higher-order central nervous functions and put into execution by three independent muscular

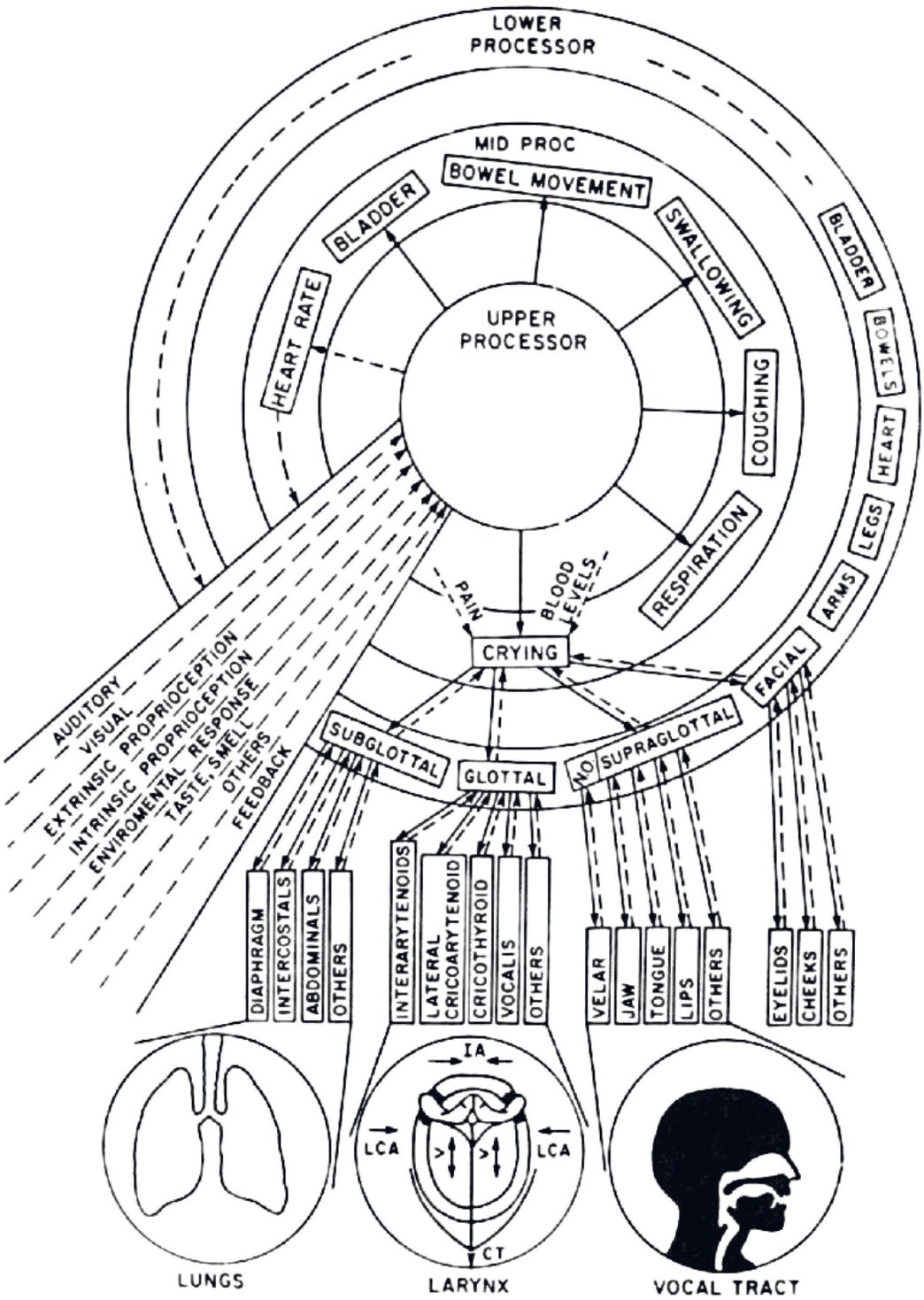


Figure 3.1.: Cry production model (Golub & Corwin, 1982)

systems, working together during cry production. Following Golub and Corwin (2000), abnormalities or pathological changes in the physical structures in any of the three systems should result in variations of acoustic properties of the systems.

The acoustic parameters analyzed in this thesis were chosen based on this physioacoustic cry model in order to identify a physiological or pathological state of each of the three systems. For example, the duration of a cry or the intensity can give evidence of a dysfunction of the subglottal system. Parameters like the fundamental frequency, the jitter and shimmer values or the harmonic-to-noise ratio can reveal dysfunctions in the glottal system. Finally, the analysis of formant frequencies can provide an acoustical image of the vocal tract, representing the supraglottal system.

In addition to selecting acoustic parameters according to the cry model, the pathological cry groups analyzed in this thesis are selected following the physioacoustic model, too.

Diseases like laryngomalacia are influencing the shape and the tissue of the vocal folds and are assumed to influence the glottal system as well as the subglottal system, caused by a lower blowing-pressure. Neurological diseases like asphyxia or brain damage are assumed to influence the glottal system, because the glottal system is said to be the "neurological high-resolution system" (Stevens et al., 2007) and a strong indicator for neurological diseases in infant cries (LaGasse et al., 2005). A cleft-lip and palate is assumed to influence the shape of the infants vocal tract representing the supraglottal system.

Another interesting aspect not considered in detail in the cry production model is the influence of the auditory system. The model explains that extrinsic impulses like auditory impulses trigger the upper processor to initiate corresponding actions like crying or stopping to cry. For this reason, and because the hearing function is known to influence speech production significantly (Blamey et al., 2001), hearing impairment is assumed to influence the cry production, too. For example, longer cry durations resulting from a missing auditory feedback may occur and would be measurable in acoustic analyses.

The following section describes the method of recording the infant cries. Chapter 5 explains how single cry utterances were extracted from the cry recordings and chapter 6 defines the acoustic parameters and analyses that were selected for infant cry classification.

3.4. Ethical clearing

All single studies and procedural methods which are used in this thesis are proven by the Ethic Committee of the Fresenius University of Applied Sciences. Overall, two ethical clearings were applied. The first ethical clearing (see appendix B) was given to ensure that the approach of how infants were recorded is without ethical compunction. This is taken into account by section 8.2, section 9.1 section 9.2 and section 9.3. The second ethical clearing (see appendix C) additionally proved the procedure regarding the recording of pain-induced cries. The pain stimulus was independent from the cry recording. The recording was conducted during the phenylketonuria screening. A blood sample was drawn by heel prick and the crying after the heel prick was recorded. The analysis of pain cries is conducted in section 8.1.

Chapter 4.

Cry recording

To standardize the recording approach, several aspects were specified: all infants were recorded in a supine position with a distance of 30 cm between the microphone and the infants' mouth. The cries were recorded on a Zoom H2n recorder with a built-in microphone. For recording, a sampling frequency of 48 kHz and 24 bit digital resolution was selected. To ensure that the recordings were made under almost the same condition, a pre-recording was conducted to ensure that no signal was heard or seen on the calibrated level meter of the H2n, which indicates background noise. An average setting of the recording volume control was selected. Because not all recordings were made in the same room, several criteria were defined to ensure that recordings were made under almost the same condition (table 4.1).

For this thesis, recordings were not made in soundproof cabins. Many of the infants included in the studies required medical treatment at the time the recordings were made, and most hospitals cannot provide soundproof cabins. In addition, when recording spontaneous infant cries, much time may be spent before infants start crying; "locking" parents and infants into soundproof cabins for a long time would be unethical.

Table 4.1.: Criteria for the recording environment

Inclusion criteria	**Exclusion criteria**
closed doors and windows	stone floor or flaggings
mobiles, laptop, tv and radio must be switched off, no standby modus	fixed neon lamps in the room
furnished room	existent background noise, identified on the calibrated level meter, which can't be identified or removed
	switched on aquarium
	ticking clocks

For each infant, one full episode of crying was recorded. Recordings were started with the first cry of the infant. The pre-recording functionality of the H2n was used to ensure that the whole first cry utterance was recorded. Recordings were stopped after the last cry of an infant when there was a 15s pause with no crying.

In order to guarantee sufficient quality of the recordings, only recordings with more than 30 dB intensity between the minimum intensity within an episode of crying (corresponding to the noise level) and the maximum intensity within the episode of crying were included in the study.

Chapter 5.

Cry extraction

All cry recordings contained multiple *cry utterances* of an infant, interrupted by phases of silence. Separating the single infant cries from the silent parts is necessary for applying acoustic analyses to the crying parts of the recording only. Here, two types of approaches exist: (1) computer algorithms can be used to automatically identify the crying parts and the noise parts in a cry recording or (2) cries can be extracted manually.

In this thesis, the single cry utterances were not detected automatically. Studies dealing with the automatic detection of inspiratory and expiratory phases found error rates of between 5 and 16 % for the different algorithms. For example, L. Abou-Abbas, Alaie, and Tadj (2015) compared different modern approaches and were able to identify inspiratory and expiratory phases automatically with an accuracy of 83.79 %. Reggiannini et al. (2013) were able to identify cries by means of fundamental frequency detection with an accuracy between 88 % and 95 %.

To avoid misclassification by algorithmic cry detection, in this thesis the inspiratory and expiratory utterances of a cry period were identified using auditory and visual discrimination.

To extract the cries, the waveform of the signal and its spectrogram were computed. Based on both visualizations, the beginning and ending of the cry utterance were identified and confirmed auditorily. Only the cries fulfilling the inclusion criteria shown in 5.1 were selected for the acoustic analysis.

Table 5.1.: Inclusion and exclusion criteria for infant cries

Inclusion criteria	**Exclusion criteria**
cry during the expiratory phase cry duration of at least 0.4 seconds	cry during the inspiratory phase

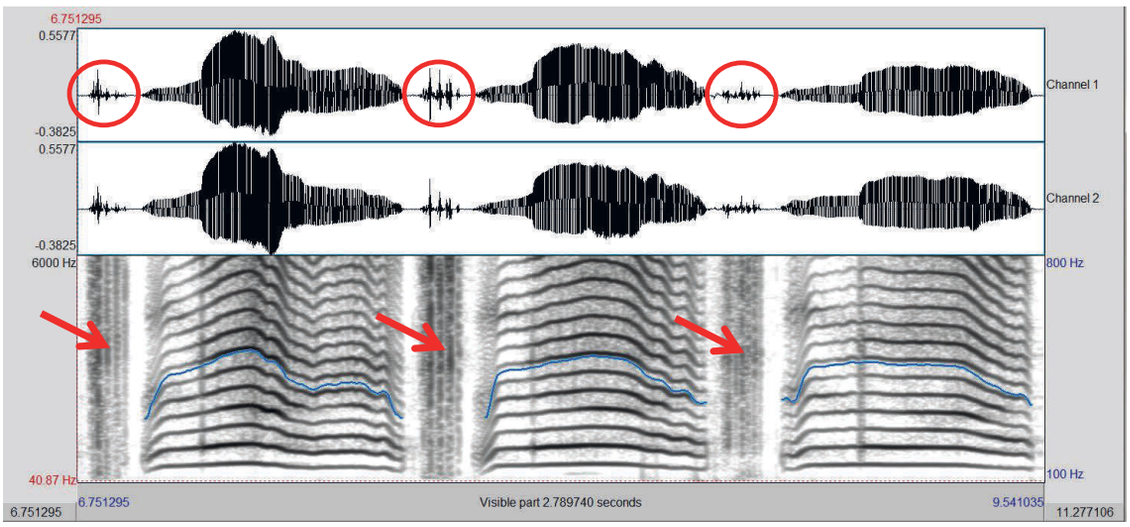


Figure 5.1.: Waveform and spectrographic visualization of inspiratory and expiratory cry utterances. Inspiratory cry utterances are marked in red color. Blue curves represent the fundamental frequency found in expiratory cry utterances.

Especially inspiratory cries (that often occur directly after the end of an expiratory cry) were discarded, because previous analyses showed significant differences in acoustic parameters between inspiratory and expiratory cries that could therefore distort the statistical analysis of the cries (Etz & Pörschmann, 2011).

Infants younger than four months of age can exhibit inspiratory cries. This phenomenon is caused by the anatomic structure of the infantile orofacial, pharyngeal and laryngeal tract. Up to the age of four months, the larynx lies in a superior position that is situated higher compared to the larynx' position of toddlers or adults (Boudewyns, Claes, & van de Heyning, 2010). In addition, the larynx, epiglottis and the aryepiglottic fold are not fully hardened and therefore, tend to collapse during inspiration caused by the Bernoulli effect (LaGasse et al., 2005). This leads to the inspiratory cries of infants and is a normal phenomenon existing in cry utterances of healthy infants.

Inspiratory cries were auditorily and visually identified. Auditorily, this cry type has different characteristics compared to expiratory cries. By auditory discrimination, an inspiratory cry is characterized by a stridor-similar noise and can be described as noisy breathing. Visually, the inspiratory cry parts are clearly recognizable in the spectrogram (figure 5.1). The inspiratory utterances can be visually identified as noise, having no periodic parts.

Altogether 544 cry samples from 72 infants (37 male, 35 female) could be extracted. The cry dataset used to answer the single research questions differs across the chapters of this thesis.

Not every research question required the analysis of the whole dataset, e.g. when only healthy infants were considered (section 8.1) or when only two pathologies were examined (section 9.1). In addition, the number of cries varies between the different chapters because additional infants were recorded and added to the dataset over the time this thesis was written. It was decided to include these recordings because it gave the possibility to analyze more infants with some very rare pathologies.

Chapter 6.

Acoustic analysis

The acoustic analysis of infant cries follows the same principles as the acoustic analysis of adult's speech signals. However, the anatomy of the infantile vocal tract differs significantly from the adult vocal tract (section 3.1). Therefore, infant cries are less complex then adult's speech signals which makes acoustic analysis techniques more reliable and which makes it easier to interpret analysis results. However, acoustic analysis techniques that make assumptions about the human speech production system or the vocal tract, may not be valid for analyzing infant cry signals and can therefore not be used or not be interpreted in the same way as for speech signals of adults.

In this thesis, the acoustic analysis of infant cries was used as input for computational classification techniques. These techniques work best when not too many input variables exist. In addition, the interpretation of the classification results is getting harder when too many parameters are involved. For this reason, not the full acoustic signal with all sampling values was used as input, but a set of acoustic parameters abstracting the acoustic signal was used to characterize the sound signal.

Spectral smoothing techniques that are well known for adult speech signals like smoothed FFTs, Cepstrums or LPC spectrums, were not used for signal abstraction for the following reasons:

1. The techniques remove acoustic information that may be relevant for identifying the health state of infants. In particular, the techniques try to filter out acoustic details about the vocal folds which may be relevant for various pathologies like laryngomalacia.
2. The techniques may not be valid for infant cries. As example, LPC defines a very simplified model of the human vocal tract that was only developed for the adult vocal tract. Studies are missing that prove LPC to be valid for correctly abstracting the infantile vocal tract.

Therefore, a custom set of acoustic parameters for abstracting an infant cry signal was defined. The selection of acoustic parameters and the algorithms used for computing them, is described in the remainder of this chapter.

6.1. Acoustic parameters for infant cry classification

To acoustically characterize cry utterances of healthy infants and infants with pathological disorders, acoustic parameters were selected following the physioacoustic cry production model that was introduced in section 3.3. The parameters were computed stationarily and no temporal development over the period of crying was captured. Stationary analysis is a common approach in infant cry research (Branco et al., 2007; Goberman & Robb, 2005; K. Michelsson et al., 2002; Rautava et al., 2007; Robb, Crowell, & Dunn-Rankin, 2007), as pathological disturbances are expected to change acoustic properties over the full period of crying.

In order to model an overall characteristic shape of the cry signal, most acoustic parameters were described by the median value of the parameter in the complete cry period representing the average value of this parameter. The 10th and 90th percentile values of the parameter were selected to represent lower and upper bounds within the cry period. The interquartile range (the range between the 75th and 25th percentile) was chosen to represent the dispersion of the parameter.

By this approach, the acoustic characteristics of the cry period were captured while the signal was abstracted enough to keep an overview about the parameters involved later in the statistical analyses.

All acoustic parameters were computed with the phonetic Software Praat 5.3.39 (Boersma & Weenink, 2013b). The algorithms and settings that were used are described in the following sections.

6.1.1. Parameters for the subglottal system

For acoustically representing the *subglottal system* of the physioacoustic cry production model, parameters were selected that refer to the process of building and keeping the air pressure for producing the cry. The intensity (*Int*) was used to model the energy which infants put into crying and the cry duration (*Dur*) represents the time a cry lasts. The number and degree of voice breaks (*NoVB*, *DegVB*) represents an infant's ability to keep crying without interruptions. For infant crying, voice breaks are mainly associated with problems in the subglottal system, because the

infants are not able to precisely adjust the vocal folds as shifting between registers (Kiese-Himmel, 2016). Hence, voice breaks originate from changes in the air flow.

Intensity

The intensity's median, the 10th and 90th percentile and the interquartile range were computed. Constant noise levels that might have been introduced by the microphone (DC offset) were subtracted by Praat's intensity algorithm automatically. In addition, the algorithm filtered pitch-synchronous intensity variations (Boersma & Weenink, 2013a). Praat's energy algorithm was used to approximate the sound pressure level (SPL) of the original cry signal:

$$Int[dB] = 10\log_{10}\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} 10^{\frac{x(t)}{10}}\,dt \tag{6.1}$$

where t_1 and t_2 are the start and end tine of the infant cry signal and $x(t)$ is the intensity of the signal as function of time (Boersma & Weenink, 2013a).

Cry duration

Infant cries were extracted manually from the episode of crying for each infant. Boundaries of cries were identified by spectrographic analysis, based on the intensity contour as well as the waveform and the spectrum. Then, cry duration (*Dur*) was computed as the duration of the extracted cry.

Number and degree of voice breaks

The number of voice breaks counts the distances of consecutive glottis pulses that are longer than 1.25 divided by the pitch floor, which is 100 Hz in this setting. The degree of voice breaks is the fraction of voice breaks compared to the number of all distances between consecutive glottis pulses. The number of voice breaks is calculated through the distance between the single glottal pulses of the signal.

The number and degree of voice breaks are assigned to the parameters of the subglottal system because it can be assumed that the incomplete maturation of nerves in infancy causes friction of a balanced controlling of the respiratory muscles; this causes voice breaks during the expiratory

phase. Also for adults, the interaction between the inspiratory and expiratory muscles is described as a very complex one (Borden & Harris, 1984; Raymond D. Kent, 1997; Reetz, 2003).

6.1.2. Parameters for the glottal system

To model the *glottal system*, the fundamental frequency ($F0$) was chosen to characterize the function of the vocal folds. To capture pathological modifications of the vocal folds, the local shimmer and jitter were computed ($Shim$, $Jitt$). In addition, the harmonics-to-noise ration (HNR) was computed to capture hoarseness within infant crying.

Fundamental frequency

The fundamental frequency ($F0$) was computed with Praat's autocorrelation algorithm (Boersma, 1993). The algorithm was parameterized to use a Gaussian window; the pitch floor was set to 100 Hz and the pitch ceiling to 1000 Hz. According to various studies, spontaneous cries as well as pain-induced cries range from 200 Hz to 600 Hz (Crowe & Zeskind, 1992; Furlow, 1997; K. Michelsson et al., 2002; Sirviö & Michelsson, 1976; Wolff, 1969). Porter, Miller, and Marshall (1986) described fundamental frequencies of pain-induced cries not exceeding 730 Hz. Cries that were auditorily perceived as high-pitched were analyzed in an oscillogram to validate that the fundamental frequency was below the pitch ceiling of 1000 Hz. The frame duration was left to be selected automatically by the autocorrelation algorithm ($\frac{6}{\text{pitch floor } [Hz]} = \frac{6}{100\ Hz} = 0.06\ s$).

For the fundamental frequency, the median was computed as robust parameter for the central tendency of F0. The 10th percentile (P_{10}) and the 90th percentile (P_{90}) were computed as lower and upper bounds of F0. In analyses of the spectrogram, the 10th and 90th percentile proved to be more robust against outliers in F0 than minimum and maximum. The interquartile range ($IQR = P_{75} - P_{25}$) was used as measure for the dispersion of F0.

Micro-variability of vocal folds

For analyzing the micro-variability of the vocal folds, Praat's waveform matching algorithm (Boersma, 2009) was used to estimate pitch marks and the local jitter and shimmer values were computed.

In Praat, the local jitter ($Jitt$) is defined as the mean absolute difference between the duration of succeeding periods, divided by the mean duration of periods:

$$Jitt = jitter_{local} = \frac{jitter_{absolute}}{meanPeriod} \tag{6.2}$$

with

$$jitter_{absolute} = \frac{\sum_{i=2}^{N} T_i - T_{i-1}}{N-1}$$

$$meanPeriod = \frac{\sum_{i=1}^{N} T_i}{N}$$

where T_i is the duration of the i-th period and N is the number of periods. A period is defined to be the smallest interval after a signal recurs.

Analogously, the local shimmer ($Shim$) is defined as the mean absolute difference between the amplitudes of succeeding periods divided by the mean amplitude of the cry. Formulas are similar to the local jitter; only the duration of a period T_i is replaced by the amplitude of a period A_i.

Even though not typical for infant cry research, jitter and shimmer were included in the set of parameters. According to Barr et al. (2000), jitter and shimmer are features of the infant cry worth to be analyzed. Two decisions were made in this thesis regarding shimmer and jitter computation:

1. Local jitter and shimmer were chosen instead of their absolute counterparts $jitter_{absolute}$ and $shimmer_{absolute}$ as they are comparable across cry samples with differences in their period length.
2. A waveform matching algorithm was chosen instead of peak picking, as Praat's waveform matching algorithm is more robust against additive noise (Boersma, 2009).

In addition, studies comparing the accuracy of different algorithms for computing pitch marks (DYPSA/DYMP, peak picking, waveform matching) jitter (local jitter, absolute jitter, STJE, LocJitt, RAP, ppq5) conducted by D. G. Silva, Oliveira, and Andrea (2009) and Teixeira and Gonçalves (2014) identified Praat's waveform matching algorithm combined with the local jitter as one of the most accurate algorithms and it was therefore chosen for infant cry analysis in this thesis.

Harmonics-to-noise ratio

For computing the harmonics-to-noise ratio (*HNR*), Praat uses a forward cross-correlation analysis (Boersma & Weenink, 2013a) and presents the HNR values in the decibel scale:

$$HNR[dB] = 10\log_{10}\left(\frac{harmonicpart[\%]}{noisepart[\%]}\right) \tag{6.3}$$

The HNR mean and the standard deviation were computed as Praat does not support computing the median or any percentile on harmonicity values.

6.1.3. Parameters for the supraglottal systen

The structure and acoustic properties of the *supraglottal system* were modeled by computing the first six formants ($F1$ – $F6$). In contrast to voice analyses for adults, the interpretation of the formants is not investigated very well, but the formant values were used rather as a spectral smoothing technique.

Formants

The first six formants ($F1$ – $F6$) as references to frequency ranges with high spectral intensities were computed with Praat's Burg algorithm (Andersen, 1974; Childers, 1978; Press, Flannery, Teukolsky, & Vetterling, 2002).

In this thesis, the formants are used as a *spectral smoothing technique* rather than a modeling of the infant's vocal tract. Therefore, more than the usual first two formants were computed to model key properties of the cry signal. Following experimental analyses, six formants were found to be appropriate for approximating the cry signal sufficiently.

A one-tube-model according to Fant (1971) was used to approximate the ceiling of Praat's formant search range. The one-tube-model was assumed suitable for modeling the vocal tract during infant crying as crying is characterized by a very open and straight form of the vocal tract. Following this model and assuming a minimum length of the infant's vocal tract of 7 cm (Vorperian et al., 2009), the approximate frequency of the 6th formant can be computed as

$$F_6 = \frac{6 \cdot 340\ m/s}{4 \cdot 0.07\ m} = 7286\ Hz \tag{6.4}$$

The ceiling of the formant search range was therefore set to 8000 Hz. For each formant, the median, 10th percentile, 90th percentile and interquartile range were measured to represent the characteristics of each formant within the cry sample.

6.2. Automation of acoustic analyses using Praat script

The acoustic parameters described above were computed for the cry samples using Praat software (Boersma & Weenink, 2013b). The computation was automated using a Praat script (appendix A). The Praat script contains a dialog for defining various settings of the acoustic analyses and it executes a series of commands for computing all acoustic parameters for a set of infant cries. The results of the analyses are stored as comma-separated values in a text file for further processing in the statistical analyses.

Chapter 7.

Computational infant cry classification

After having quantified the properties of infant cries using the acoustic analysis techniques described in the previous chapter, computer-based statistical approaches can be applied on these parameters to classify the infant cries according to the health state of the infants.

For such classification tasks, the algorithmic class of "supervised-learning models" exists in the scientific field of statistics.

This section briefly introduces the concepts of supervised-learning models and introduces established algorithms that are in principle suited for infant cry classification. More common statistical methods that were used for testing hypothesis and for answering the research questions (e.g., the Krippendorff's Alpha coefficient or generalized linear mixed models) are described in the method sections of the main part of this thesis. Standard statistical approaches, like Analysis of Variances (ANOVA) are not described in detail, but can be found in Bortz (2010).

7.1. The concepts of supervised-learning models

In statistics, two main classes of concepts exist that are able to identify structures in data: unsupervised-learning and supervised-learning models (Wang, Pedrycz, Chan, & He, 2014).

Unsupervised-learning models try to find structures in data that are unknown to researchers. They are based on some kind of similarity or dissimilarity measure to allocate similar parts of data to each other. Well-known algorithms of this kind are clustering algorithms that can be applied to data for which no structures are known before the analyses. The advantage of being able to find unknown structures comes at the cost that the structures that are identified may be statistically sound but cannot be interpreted from a scientific perspective.

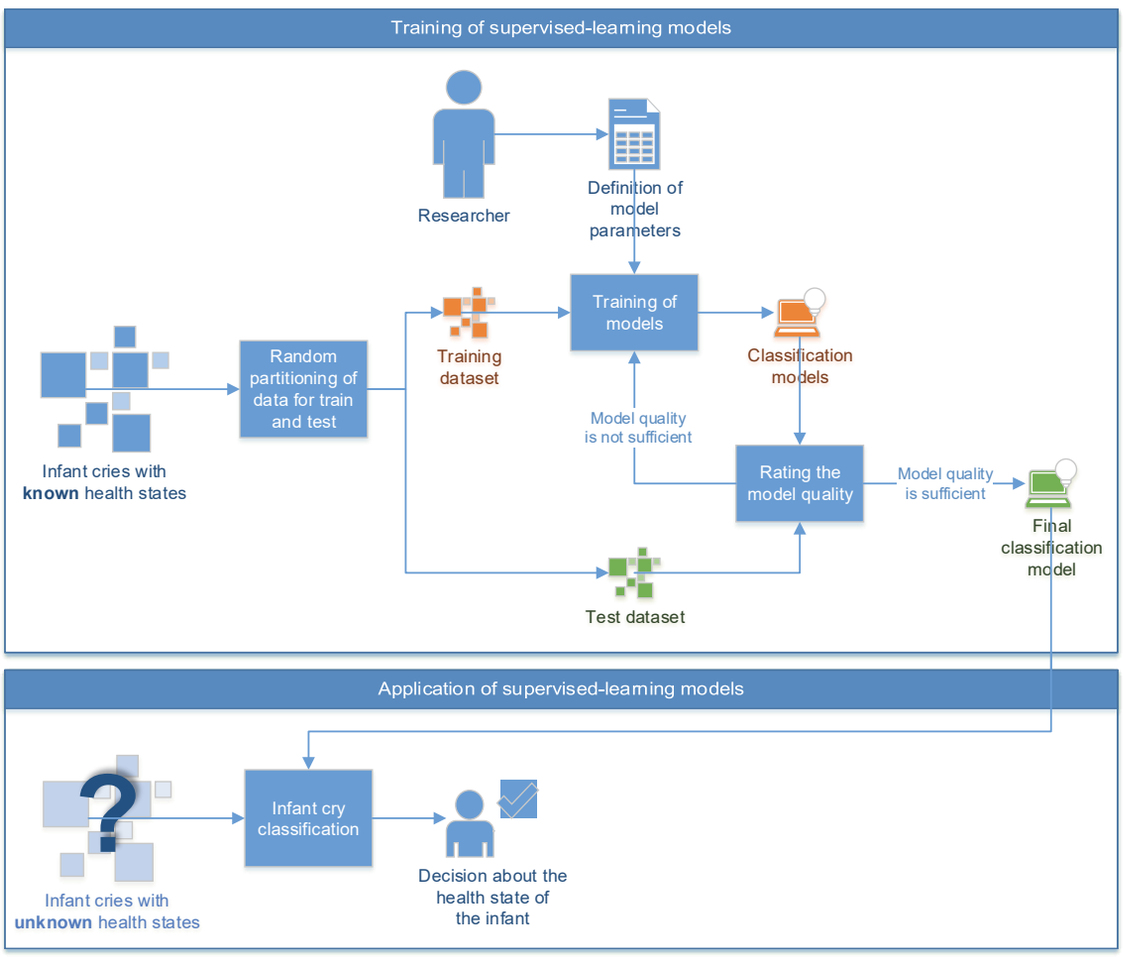


Figure 7.1.: Training and application of supervised-learning models

In contrast, *supervised-learning models* require that some structure is known for the data that shall be analyzed. Such structures are for example distinct categories or groups into which the data shall be classified. In order to classify data correctly, supervised-learning models are trained on a dataset for which the classification results are provided and can then be applied to unknown data (figure 7.1).

In this thesis, supervised-learning models were chosen for infant cry classification because there are natural categories into which infant cries shall be classified. Theses categories are the health states of infants which a screening instrument should detect; in this thesis, the following categories exist: healthy (HI), unilateral cleft-lip and palate (UCLP), hearing impairment (HI), laryngomalacia (LA), asphyxia (AS) and brain damage (BD). The training and application of supervised-learning models is explained in more detail in the following. The descriptions are based on

Cleophas and Zwinderman (2013).

7.1.1. Training of supervised-learning models

Before supervised-learning models can classify data, the algorithms must "learn" how to do so and how to identify the given structure in data for which the classification results are not known (the upper part of figure 7.1).

For learning, supervised-learning models require data for which the classification results are provided; in infant cry classification such datasets contain multiple cries, each data item consisting of the acoustic parameters of the cry as well as the health state of the infant that uttered the cry.

The dataset is randomly split into two subsets; one dataset for training the classification model and one dataset for estimating the validity and reliability of the model. This approach is called "train and test". Common proportions of these datasets are 70 % for the training dataset and 30 % for the test dataset. These proportions provide enough data, for both an accurate training of the models and for convincing approximations of the models' quality.

Researchers choose reasonable values for parameterizing the classification model algorithms and apply the algorithms to the training dataset in order to compute a classification model able to classify data.

The resulting classification models are then applied to the test dataset but are not provided with the results of these data. Instead, the results are used after classification to rate the accuracy of the model that was achieved when rating the test data. A separate test dataset is used here because testing the models on the dataset that was used for training often leads to distorted accuracy values and models then tend to fit too neatly to the training data to be applied to unknown data that do not exactly match the structure of the training data (this is called "overfitting").

Depending on the results of this test, researchers can accept the model when it is sufficiently accurate or can adapt the parameters and explore if the new models with the new parameters are more accurate.

As final result of the training phase, a classification model will be selected and a reasonable approximation for the validity and reliability of the model exists. This model can then be applied to unknown data for classification.

7.1.2. Application of supervised-learning models

After the training phase, a final classification model exists that is able to classify unknown data and no further parametrization of the model is required. For this reason, the application of supervised-learning models to unknown data is straight-forward. The unknown data are fed into the classification model and the model computes its classification result and presents it to a user.

7.2. Supervised-learning model algorithms

In this thesis, various supervised-learning models were trained and tested on infant cries and the models were compared to each other according to their quality. In the following, the different algorithms are introduced very briefly. Further information about the approaches can be found in e.g., Suthaharan (2016), Weiss and Kulikowski (1994).

Artificial neural networks encompass different machine learning approaches following functions of animal brains by simulating information flow through systems of interconnected "neurons". For cry analysis, various kinds of neural networks have been applied, e.g., multilayer perceptrons, radial basis function networks, self-organizing feature maps and others. The *Bayes classifier* is a probabilistic model based on Bayes' theorem describing classes by statistical processes. *Hidden Markov models* simulate Markov processes with hidden states and are widely used to identify patterns in temporal data like acoustic signals. *Linear discriminant analysis* identifies linear functions to separate groups in data. *Support vector machines* work similar to linear discriminant analysis, except they can be extended for non-linear discrimination between data sets. *Fuzzy logic* associates certainty values to data giving information about the possibility of the data item belonging to given groups. *Decision trees* cover different algorithms which compute hierarchical decision rules in order to decide in which group data items belong. *K-nearest neighbor* associates data to groups by analyzing the k-most-similar neighbor data items of an item to be classified. *Weighted rough set framework* targets the class imbalance problem (i.e., groups have different sizes) by providing a mathematical model based on lower and upper approximation of groups.

Classification using C5.0 decision trees

As result of the comparison of the supervised-learning approaches presented in section 9.2, one algorithm from the class of decision trees achieved better ratings than all other models, the C5.0

decision tree algorithm (figure 7.2). For this reason, this algorithm is described in more detail in the following.

The C5.0 decision tree algorithm computes so-called decision rules (Quinlan, 2003). These rules can be seen as clinical cut-off values for deciding into which group a cry should be classified. How the tree is used to classify an unknown cry will be shown in an example.

For example, an unknown cry is classified into one of the three groups healthy, unilateral cleft lip and palate or hearing impairment. For this example, consider an unknown cry with an F0 maximum of 434 Hz, an HNR mean of 16.9 dB, an intensity maximum at 75.1 dB and an F0 median of 412.3 Hz.

The tree would classify the cry in the following way: on top of the decision tree is the root node (the first box on top). Below the root node there is the first (and most important) decision criterion, i.e., the first acoustic parameter that is used for discrimination. In this tree it is the maximum of the fundamental frequency that is explored. On the edges between the root node and the two nodes below, the clinical cut-offs for the F0 maximum are noted. Because with 434 Hz the exemplary cry has an F0 maximum higher than 432.779 Hz, it is put into node 2. Next, the HNR mean of the cry is explored. Having an HNR mean below 19.664 dB, the exemplary cry is put into node 3. Because of an Intensity maximum higher than 74.553 dB, the exemplary infant cry is allocated to node 5. Finally, the F0 median is analyzed and since the cry has a lower F0 median value than 414.166 Hz, it is put into node 6. There are no more nodes below node 6. Hence, this node provides a decision about the classification of the exemplary infant cry. The cry belongs to the hearing impaired group with a probability of 92.86 %.

During the so called Boosting process, various decision trees are trained on the data set. At the end, the results of the single trees are put into a voting process and the prediction with the highest votes is returned as the final classification for that cry.

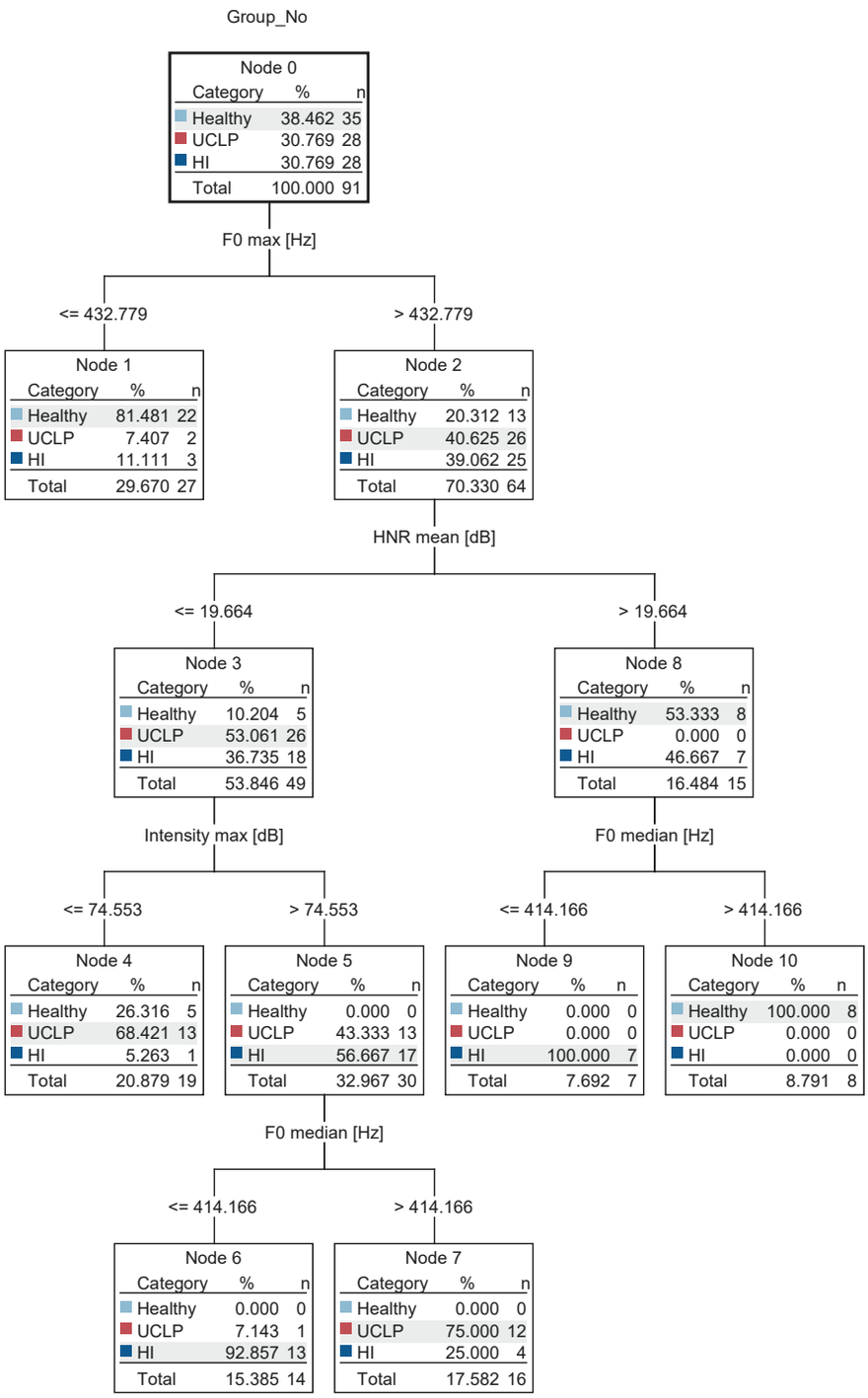


Figure 7.2.: C5.0 decision tree predicting the group membership based on acoustic parameters of an infant cry

Part III.

Main part

Chapter 8.

Reliability of infant cry analysis

While differences between healthy cries and those related to disorders have already been explored (Green, Gustafson, & McGhie, 1998; K. Lind & Wermke, 2002; Robb, Goberman, & Cacace, 1997), only little research about the *reliability* of the infant cry has been conducted.

For this study, an *episode of crying* was defined as the total period of continuous crying activity (Grau, Robb, & Cacace, 1995). In an episode of crying, an infant produces multiple *cries*. Cries are extracted from the episode of crying for analysis. In the context of infant cry analysis, "reliability" is defined as the homogeneity of all cries in an episode of an infant's crying. Therefore, reliability provides information about the reproducibility of acoustic analyses when analyzing multiple cries of an infant. This especially becomes important when infant cries shall be classified automatically, e.g., for diagnostic purposes. Here, all cries of an infant must be similar enough for a classification model to be able to predict the same diagnostic result for each cry. Otherwise, an infant's cries are classified ambiguously and the infant may not be allocated to one (diagnostic) group.

Research on infant cry reliability especially lacks studies about which *type of crying* is the most reliable one. Cry types that have often been used in infant cry research are spontaneous cries (e.g., by Manfredi et al., 2009; Wermke et al., 2011; Wermke, Mende, et al., 2002) and pain-induced cries (e.g., by Branco et al., 2007; Runefors & Arnbjörnsson, 2005; Runefors et al., 2000). For *spontaneous cries*, the researcher has to wait until the infant starts crying without intervention. Here, the cry can have various causes like hunger, mood, desires or indisposition. Often, it is not traceable which of those might be the actual cause of crying. For *pain-induced cries*, infants are recorded when they start crying due to a pain stimulus. For ethical reasons, pain stimuli are necessarily independent of the cry analysis. Vaccinations or blood withdrawals in the context of regular screenings are often used as pain stimulus. Here, the cause of the cry can clearly be related to the pain stimulus. Although pain-induced cries are assumed to be more standardized because of

their known cause, they are also said to be more biased due to their high energy of the cry (Thoden & Koivisto, 1980). In contrast, spontaneous cries seem to be less standardized because of their unknown reason. Which type of crying is suited best to be used in diagnostic instruments is not answered conclusively.

The remainder of this chapter describes the research conducted as part of the first research objective — to assess if acoustic parameters of healthy infants' cries and of infants with pathologies are reliable and which type of crying is the most reliable one.

Section 8.1 describes a study exploring the reliability of healthy infant cries and section 8.2 deals with the findings of the study concerning cries of pathological infants.

Section 8.3 summarizes and interprets the new insights, such as which acoustic parameters are consistent over multiple cries of an infant. In addition it answers which type of crying is the most reliable one and therefore might be suited best for infant cry research.

8.1. Reliability of healthy infant cries[1]

To answer research question 1, a study was conducted analyzing the reliability of healthy infant cries for spontaneous cries and for pain-induced cries as well as for two subgroups of these two groups. Multiple cries of each infant were tested on the homogeneity of their acoustic parameters within their cry group. Differences in the reliability between the cry groups were tested on significance.

8.1.1. Method

Subjects

In this study, 68 healthy infants were included. 268 spontaneous cries were recorded from 35 infants (14 female, 21 male). 236 pain-induced cries (after heel prick) were recorded from 33 infants (15 female and 18 male).

Inclusion and exclusion criteria for all infant groups are described in section 3.2.

[1] The results of the study described in this section were published in Etz, T., Reetz, H., Wegener, C., & Bahlmann, F. (2014). Infant cry reliability: Acoustic homogeneity of spontaneous cries and pain-induced cries. *Speech Communication*, *58*, 91–100. doi:10.1016/j.specom.2013.11.006

Table 8.1.: Statistical parameters for the subjects ($N = 68$)

Parameters	Mean	SD	Range
Birth weight [g]	3320.85	354.51	2710 – 4120
Gestational age [weeks]	39.25	1.24	37 – 42
Age [days]	2.01	0.77	1 – 3

Data acquisition

Infant cries were recorded as described in chapter 4 and single cry utterances were extracted from each period of crying as described in chapter 5. Table 8.1 describes the statistical parameters of the subjects.

Grouping of cries by type

Participants in this study were divided into two main groups – spontaneous cries and pain-induced cries – as those two general groups are often used in infant cry analysis. For the first group (*SP* group), recordings were started when infants began crying spontaneously. None of these cries were pain-related or triggered by any known cause. To exclude causes like sleepiness, hunger or discomfort, it was assured the infant was awake, properly fed (but not right after feeding) and wore dry diapers. For the second group (*PI* group), cries were recorded during the phenylketonuria screening (PKU) as part of the routine newborn screening. During the PKU, a blood sample was drawn by heel prick. If the heel was not warmed by socks before, the heel was warmed with warm water to achieve a good circulation of the blood. A Microtainer lancet was used for the heel prick. Recordings were started before the prick to ensure that the first cry after the pain stimulus was not missed.

In addition to those two general groups, one especially homogeneous subgroup was extracted from each of the two main groups to explore if reliability is higher when focusing on special cries within each main group. Figure 8.1 visualizes the grouping of the two general groups and the two subgroups.

Within the spontaneous group, a special kind of spontaneous cry was identified by acoustic analysis: the *non-distressed cry*. The characteristic of this type is a harmonic structure of the signal with a continuous contour of F0 as well as of the intensity. Both contours are without shifts and breaks (Truby & Lind, 1965). Furthermore, this type has a clear rising in intensity at the beginning and a clear falling at the end. Additionally, Lester (1976) described a reduced intensity compared to

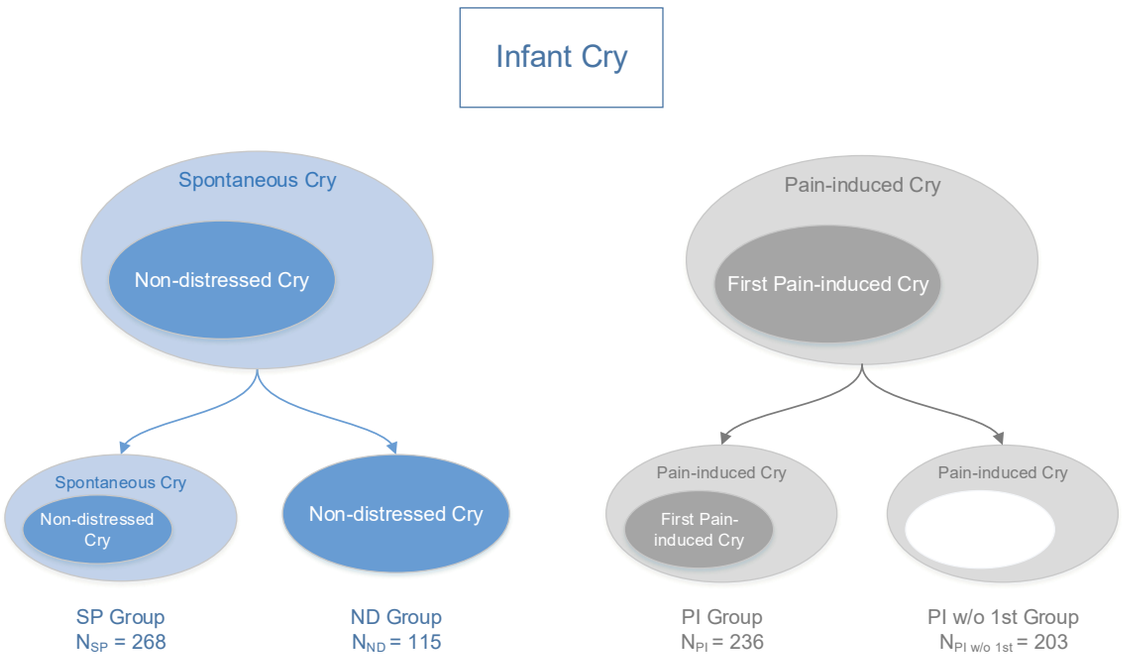


Figure 8.1.: Grouping of cries by type of crying

other cry types. In contrast to other subtypes of spontaneous cries, this cry type can be identified by spectral analysis and therefore is clearly recognizable. Non-distressed cries were assigned to a third group (*ND* group). This group contained non-distressed cries only, whereas the SP group contained non-distressed cries as well as other spontaneous cries.

For the pain-induced cries a subgroup was created by removing the first cry after the pain stimulus. Runefors et al. (2000) described the first cry after a painful stimulus as being different compared to the following cries. Pain-induced cries without the first cry were assigned to a fourth group (*PI w/o 1st* group).

Summarizing, four groups were defined after this grouping process: (1) the *SP* group containing 268 spontaneous cries, (2) the *ND* subgroup containing only spontaneous, non-distressed cries ($N = 115$), (3) the *PI* group containing 236 pain-induced cries, (4) and the *PI w/o 1st* subgroup containing 203 pain-cries without the first cry after the pain stimulus.

Acoustic analysis

Infant cries were analyzed on their acoustic parameters as described in chapter 6. For each cry utterance 19 acoustic parameters were computed with Praat software forming the data basis for

the statistical analyses.

Statistical analysis

Infant cry reliability: Krippendorff's Alpha The homogeneity of infant cries can be seen as a form of reliability. For each infant, an acoustic parameter, e.g., the fundamental frequency median, is computed for each of the infant's cries. Comparing those acoustic parameter values to each other allows an estimation about how reproducible the results for an infant are. Because the computation algorithms for acoustic parameters are perfectly reproducible (for identical signals, the algorithms compute always the same results), reliability estimation analyzes the reliability of the cry production itself.

Because of these considerations an algorithm for computing inter-rater reliabilities was chosen to quantify the extent of agreement (i.e., the reliability) among the single cries. Krippendorff's Alpha (Hayes & Krippendorff, 2007; Krippendorff, 2003) is a coefficient used in content analysis to compute inter-rater reliabilities (IRR). The inter-rater reliability measures for given events (called units), how exactly multiple observers (called raters) rate given units. If all observers give similar ratings, the IRR is high and it can be assumed that the rating results are reliable. If the observers give completely different ratings, the IRR is low and it can be assumed that the ratings are given more randomly and therefore are unreliable.

For analyzing the similarity of infant cries, Krippendorff's Alpha was adapted. For each acoustic parameter, one Krippendorff's Alpha value is computed. Here, the "units" are defined by the individual infants. Each cry is a "rater" for the real value of the infants' acoustic parameter. By this, Krippendorff's Alpha computes the consistency of one acoustic parameter over all cries of an infant, averaged over all infants.

To allow a better understanding of the adaption and to provide scientific transparency to the validity of this approach, algorithmic details of Krippendorff's Alpha computation are presented. For that purpose, the Krippendorff's Alpha algorithm as used for infant cry research is explained by an example. To compute the alpha coefficient, four steps are to be performed (Krippendorff, 2003):

a) Construct the reliability matrix
b) Tabulate coincidence within units
c) Compute difference between values
d) Compute the α-coefficient

In the following, the steps are described in detail.

Construct the reliability data matrix First, the reliability data matrix is computed (Matrix 8.1).

Infant:	$I1_{F_0}$	$I2_{F_0}$	$I3_{F_0}$	
Cry				
1	334	479	497	
2	373	360	492	
3	345	378	.	
m_i	3	3	2	$\sum_i m_i = 8$

(8.1)

The acoustic parameter of each cry given by one infant is noted; In the example, this is the median of the fundamental frequency F_0. The matrix has as many lines as cries were recorded per infant at most. In the example, 3 cries was the maximum number of recorded cries per infant. If no 3 cries were recorded for an infant, the F0 median values for the unrecorded cries are marked as missing values (marked as "."). In the bottom line the number m_i of valid (not missing) cries for infant i is noted. At the end of this line, the overall number of cries is summed up.

Tabulate coincidences within infants For each *value* occurring in the reliability data matrix, the observed coincidence is computed. The observed coincidence is the probability of a value to appear together with the other values as observed in the dataset (for numeric values, the distance between two values will be computed, see section 8.1.1). This information will later be used to determine if the fundamental frequency values of an infant may be by chance or not.

To construct the coincidence for all possible pairs of values, a coincidence matrix is calculated (Matrix 8.2).

F_0 values	1	$\cdots$	w	$\cdots$	
1	O_{11}	$\cdots$	O_{1w}	$\cdots$	N_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
v	O_{v1}	$\cdots$	O_{vw}	$\cdots$	$N_v = \sum_w O_{vw}$
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
	N_1	$\cdots$	N_w	$\cdots$	$N = \sum_{v,w} N_{vw}$

(8.2)

The rows (v) and columns (w) in this matrix represent all F0 median values occurring in the dataset. Each entry O in the matrix at point (v,w) is computed as

$$O_{vw} = \sum_i \frac{\text{Number of } (v,w) \text{ pairs in infant } i}{m_i - 1} \tag{8.3}$$

with v and w as F0 median values and m_i as number of cries for infant i.

For the given example, the coincidence matrix shown in Matrix 8.4 is computed.

F_0	334	345	360	373	378	479	492	497	
334	.	0.5	.	0.5	.	.	.	.	1
345	0.5	.	.	0.5	.	.	.	.	1
360	.	.	.	.	0.5	0.5	.	.	1
373	0.5	0.5	.	.	.	.	.	.	1
378	.	.	0.5	.	.	0.5	.	.	1
479	.	.	0.5	.	0.5	.	.	.	1
492	.	.	.	.	.	.	.	1	1
497	.	.	.	.	.	.	1	.	1
	1	1	1	1	1	1	1	1	8

(8.4)

Compute distance matrix To determine *how different* two F0 median values are, a distance function is used. Krippendorff's Alpha uses diverse distance functions according to the level of measurement of the data. It supports nominal, ordinal, interval and ratio scale for distance computation. For a comprehensive description of all distance functions see Krippendorff's book on content analysis (Krippendorff, 2003).

In this example, data are of ratio scale. The appropriate distance function between the ratio variables v and w is defined as:

$$_{\text{ratio}}\delta^2_{vw} = \left(\frac{v-w}{v+w}\right)^2 \tag{8.5}$$

Distances between all (v,w) pairs are calculated in a distance matrix. In the example the distance

matrix is:

δ^2_{vw}	334	345	360	373	378	479	492	497
334	.000	.000	.001	.003	.004	.032	.037	.038
345	.000	.000	.000	.002	.002	.026	.031	.033
360	.001	.000	.000	.000	.001	.020	.024	.026
373	.003	.002	.000	.000	.000	.015	.019	.020
378	.004	.002	.001	.000	.000	.014	.017	.018
479	.032	.026	.020	.015	.014	.000	.000	.000
492	.037	.031	.024	.019	.017	.000	.000	.000
497	.038	.033	.026	.020	.018	.000	.000	.000

(8.6)

Compute α coefficient Finally, Krippendorff's Alpha coefficient is computed as the ratio between the observed disagreement D_o among infant cry parameters and the disagreement D_e one would expect when the parameters are attributable to chance instead to the properties of the cry:

$$\alpha = 1 - \frac{D_o}{D_e} \tag{8.7}$$

For ratio values, Krippendorff's Alpha coefficient is defined as:

$$_{\text{ratio}}\alpha = 1 - (N-1)\frac{\sum_v \sum_{w>v} O_{vw}\ {}_{\text{ratio}}\delta^2_{vw}}{\sum_v \sum_{w>v} N_v N_w\ {}_{\text{ratio}}\delta^2_{vw}} \tag{8.8}$$

where N is the number of cries, O_{vw} is the coincidence for the pair (v,w) of the coincidence matrix (Matrix 8.4), $_{\text{ratio}}\delta^2_{vw}$ is the distance between both items of the pair as noted in the distance matrix (Matrix 8.6), N_v and N_w are the number of times the items v and w occur. For other levels of measurement, only the distance computation changes.

Inserting the corresponding values computes the alpha coefficient (the formula is abbreviated for readability reasons):

$$\begin{aligned} _{\text{ratio}}\alpha &= 1 - (8-1)\frac{0.5\cdot 0 + 0.5\cdot 0.003 + \ldots + 1\cdot 0}{1\cdot 1\cdot 0 + 1\cdot 1\cdot 0.001 + \ldots + 1\cdot 1\cdot 0} \\ &= 0.634 \end{aligned}$$

Krippendorff's Alpha was chosen as IRR coefficient for infant cry reliability for the following reasons. For infant cry reliability, algorithms that are able to cope with multiple units (acoustic

Table 8.2.: Interpretation of alpha coefficients according to Landis and Koch (1977)

Alpha	Interpretation
$\alpha \leq 0.0$	Poor
$0.0 > \alpha \geq 0.2$	Slight
$0.2 > \alpha \geq 0.4$	Fair
$0.4 > \alpha \geq 0.6$	Moderate
$0.6 > \alpha \geq 0.8$	Substantial
$0.8 > \alpha \geq 1.0$	Perfect

parameter) and multiple raters (the measured value for the acoustic parameter. Each rater is one cry in an episode of crying) are required. In addition, the algorithm must be able to handle missing values (not all infants had the same number of cries in an episode of crying, so some "ratings" were missing). Finally, acoustic parameters are interval-scaled data which required the algorithm to support this level of measurement. Given those criteria, many inter-rater reliability coefficients were excluded (see Hayes' comparison of IRR coefficients for a discussion of IRR coefficient properties: (Hayes & Krippendorff, 2007)). Krippendorff's Alpha and the Intra-Class coefficient (ICC: (Shrout & Fleiss, 1979)) were considered as appropriate algorithms. Krippendorff's Alpha was used as it allows to compute inter-rater agreement for all levels of measurement while the ICC is fixed to metric data. In this study, only interval-scaled acoustic parameters were used. However, infant cry research already explored nominal properties of cries (e.g., "is bi-phonetic or is not bi-phonetic"). So using Krippendorff's Alpha would allow to extend reliability analysis and compare results on nominal data with the results in this study.

Interpretation of Krippendorff's Alpha coefficient For interpreting inter-rater reliability coefficients like Krippendorff's Alpha, the best known conventions are those proposed by Landis and Koch (1977). They categorize reliability coefficients into six ranges as shown in table 8.2. Values less than 0.0 have a poor agreement and can be interpreted as having a great statistical spreading and as being very unequal to each other. Values less than 0.2 can be interpreted as a slight agreement between the cries. Values between 0.2 and 0.4 are said to have fair agreement. A moderate agreement can be assumed at values up to 0.6. Values between 0.6 and 0.8 can be interpreted as having a substantial agreement. Values between 0.8 and 1.0 can be interpreted as a perfect agreement between the single cries.

For the research field of content analysis (where Krippendorff's Alpha originates from), Krippendorff declared alpha values above 0.8 as good reliability and values above 0.667 as acceptable ones (Krippendorff, 2003). For interval-scaled data, especially in medical and language studies, alpha

values higher than 0.4 are considered adequate (Artstein & Poesio, 2008; Rietveld & van Hout, 1993).

In infant cry research, some degree of dispersion within the cries of an infant was assumed as normal. For this reason, a relaxed interpretation of alpha values is proposed as follows: Alpha values above 0.4 are declared as acceptable reliability and alpha values above 0.667 as good reliability.

Differences in reliability between cry types To identify if the overall distribution of Krippendorff's Alpha values were significantly different between groups, a Kruskal-Wallis test was performed. In a new dataset, the group and Krippendorff's Alpha value were defined as variables. For each group, 19 Krippendorff's Alpha values from the acoustic parameters were added as items. This distribution of alpha values was then compared between groups. A non-parametric test was chosen, because a Shapiro-Wilk test revealed that the alpha values for the acoustic parameters were not normally distributed.

Validation of the Krippendorff's Alpha approach To explore the validity of Krippendorff's Alpha results, the reliability of infant cries was computed with a second algorithm for inter-rater reliability; the intraclass correlation coefficient (ICC). As the ICC cannot deal with missing values, they were replaced by the group mean. For all acoustic parameters and the four groups, an intraclass correlation coefficient type $ICC(3,1)$ was computed according to Shrout and Fleiss (1979). Spearman's correlation coefficient was calculated to analyze the similarity of Krippendorff's Alpha and ICC.

8.1.2. Results

Acoustic parameters

Table 8.3 provides an overview over the results of the acoustic analysis.

Reliability of acoustic parameters

Krippendorff's Alpha was computed for all 19 acoustic parameters. Table 8.4 summarizes the results of Krippendorff's Alpha computation.

Table 8.3.: Mean (averaged over all cries within a group) and standard deviation of acoustic parameters over groups

	Group			
	SP *N=268*	ND *N=115*	PI *N=236*	PI w/o 1st *N=203*
Parameter	**Mean ± SD**	**Mean ± SD**	**Mean ± SD**	**Mean ± SD**
Cry duration [ms]	1056.88 ± 525.70	855.15 ± 364.46	1103.78 ± 562.73	1021.87 ± 460.48
F0 P10 [Hz]	371.95 ± 107.55	427.93 ± 77.43	344.95 ± 111.85	346.45 ± 113.37
F0 median [Hz]	468.57 ± 87.28	473.25 ± 79.43	456.87 ± 112.35	451.38 ± 108.82
F0 IQR [Hz]	79.06 ± 79.94	39.95 ± 24.87	106.56 ± 100.42	104.89 ± 95.86
F0 P90 [Hz]	527.19 ± 93.05	501.74 ± 81.46	538.77 ± 125.93	529.88 ± 117.20
F1 median [Hz]	1288.36 ± 231.50	1213.33 ± 225.75	1221.22 ± 265.86	1210.97 ± 245.41
F2 median [Hz]	2434.31 ± 440.93	2305.13 ± 430.86	2317.51 ± 343.73	2304.43 ± 341.74
F3 median [Hz]	3635.89 ± 393.02	3525.46 ± 387.83	3563.16 ± 415.65	3549.79 ± 399.69
F4 median [Hz]	4913.86 ± 235.78	4810.99 ± 234.86	4778.42 ± 354.76	4772.86 ± 377.11
F5 median [Hz]	6220.62 ± 279.48	6213.82 ± 272.80	5964.41 ± 428.29	5966.38 ± 415.30
F6 median [Hz]	7237.30 ± 247.48	7240.01 ± 249.42	7173.16 ± 226.14	7174.18 ± 210.38
Intensity P10 [dB]	66.50 ± 5.33	66.81 ± 5.01	55.85 ± 6.89	56.34 ± 6.31
Intensity median [dB]	71.99 ± 6.09	71.72 ± 4.98	60.40 ± 6.72	60.69 ± 5.13
Intensity IQR [dB]	4.67 ± 2.49	4.27 ± 2.18	4.63 ± 2.71	4.46 ± 2.70
Intensity P90 [dB]	75.21 ± 5.32	74.57 ± 5.18	64.30 ± 6.53	64.43 ± 6.00
Jitter (local) [%]	0.68 ± 0.36	0.49 ± 1.01	1.05 ± 0.51	1.04 ± 1.01
Shimmer (local) [%]	4.75 ± 2.27	3.87 ± 2.18	9.32 ± 3.82	9.39 ± 3.88
HNR mean [dB]	14.70 ± 5.19	18.35 ± 5.11	11.80 ± 5.14	11.70 ± 3.49
HNR mean SD [dB]	5.83 ± 1.55	4.66 ± 1.37	4.94 ± 1.68	4.83 ± 1.46

Table 8.4.: Results of Krippendorff's Alpha for the acoustic parameters grouped by the type of cry

	Kripp. Alpha for Group			
Parameter	**SP**	**ND**	**PI**	**PI w/o 1st**
Cry duration	0.368	0.385	0.337	0.379
F0 P10	0.350	0.558	0.269	0.266
F0 median	0.544	0.727	0.312	0.349
F0 IQR	0.229	0.257	0.223	0.234
F0 P90	0.631	0.708	0.406	0.489
F1 median	0.492	0.550	0.370	0.418
F2 median	0.530	0.548	0.490	0.507
F3 median	0.578	0.634	0.440	0.494
F4 median	0.392	0.488	0.339	0.336
F5 median	0.475	0.512	0.433	0.470
F6 median	0.184	0.053	0.456	0.463
Intensity P10	0.580	0.624	0.554	0.575
Intensity median	0.702	0.773	0.698	0.736
Intensity IQR	0.201	0.219	0.142	0.149
Intensity P90	0.728	0.779	0.663	0.718
Jitter (local)	0.439	0.582	0.529	0.544
Shimmer (local)	0.454	0.592	0.655	0.652
HNR mean	0.416	0.518	0.681	0.687
HNR mean SD	0.339	0.395	0.279	0.319

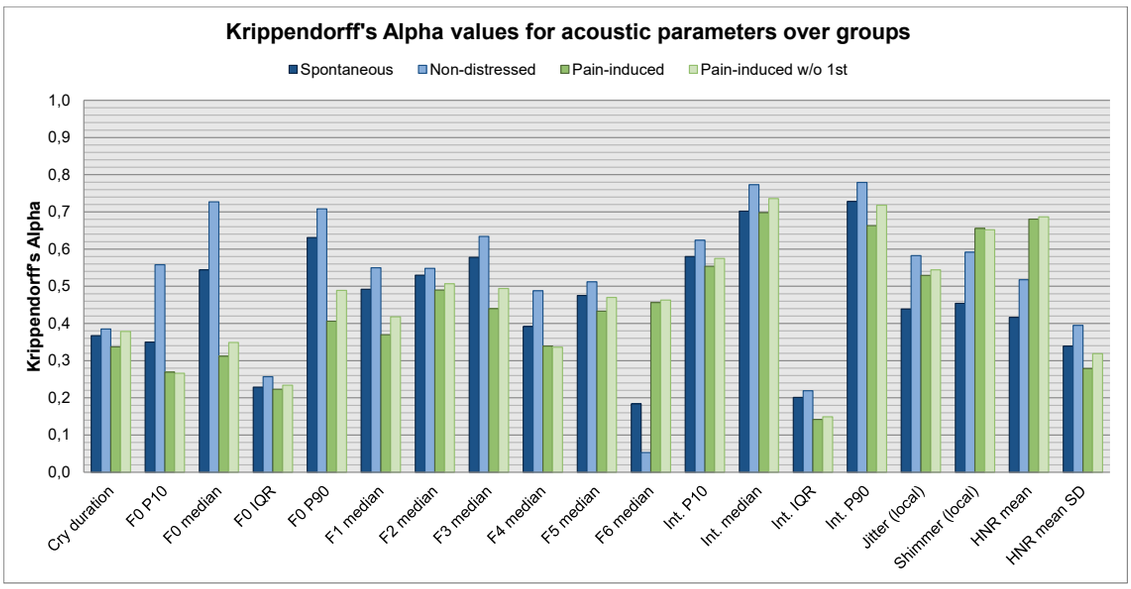


Figure 8.2.: Comparison of Krippendorff's Alpha values for the different cry types

For the spontaneous cries (*SP*) 2 out of 19 acoustic parameters had good alpha values ($\alpha > 0.667$): the intensity median and the 90th percentile (P90) of intensity. In the non-distressed group (*ND*), the F0 median, F0 P90, as well as the intensity median and intensity P90 reached good reliability values. For the pain-induced cries (*PI*), good reliability was achieved for intensity median and HNR mean. The pain-induced cries without the first cry (*PI w/o 1st*) reached good reliability values in HNR mean and intensity median as well as in intensity P90.

Differences in reliability between cry types

To test if one of the cry types shows significantly better reliability values in all acoustic parameters, a non-parametric Kruskal-Wallis test was computed. No significant differences ($p = 0.92$) were found between the four groups when including the Krippendorff's Alpha values for all 19 acoustic parameters.

However, Krippendorff's Alpha values were visualized in a box diagram to identify trends between the groups (8.2). By trend, the non-distressed cry (*ND* group) has the most reliable alpha values in 16 of 19 acoustic parameters. The acoustic parameters cry duration, all the parameters of F0 and intensity, formants 1 to 5, as well as the jitter and the HNR mean SD have their highest alpha values in the non-distressed group. The remaining three parameters F6, shimmer and HNR mean in the non-distressed group had alpha values below those in the other groups. For F6 and the HNR

mean, the alpha values in the pain-induced cries without the first cry (*PI w/o 1st*) were higher. In the case of the shimmer, the pain-induced group (*PI*) reached the highest values.

When specifically exploring the three acoustic parameters for which high differences in Krippendorff's Alpha between groups occurred (F0 P10, F0 median and F0 P90), significant differences can be verified between the ND group and the PI group ($p = 0.005$) as well as the ND group and the PI w/o 1st group ($p = 0.010$).

Validation of the Krippendorff's Alpha approach

Figure 8.3 compares the Krippendorff's Alpha values with the intraclass correlation coefficient for all 19 parameters and all 4 groups. Spearman's correlation coefficient revealed significant, moderate correlation ($R = 0.598, p = 0.00$) between Krippendorff's Alpha and ICC.

8.1.3. Interpretation

As described in section 8.1.2, Krippendorff's Alpha values over all groups were not very high for many of the acoustic parameters. As for all alpha coefficients, the threshold for acceptable similarity must be defined with respect to the research context and the research goals. When using acoustic parameters with low consistency values ($\alpha < 0.4$) for determining differences between groups, the impact of the low alpha values should be regarded. Low Krippendorff's Alpha values correlate with a greater variance within a group (figure 8.4), making it more difficult to identify differences between groups. Especially small differences between groups may get lost in high variances within groups.

For developing diagnostic instruments based on the infant cry, it is still to be defined which alpha values are satisfying and which are not. In this study a threshold of 0.4 for acceptable alpha values and a threshold of 0.667 for good alpha values was used. It is still to be evaluated if those threshold values prove to be appropriate when developing diagnostic instruments based on the infant cry.

By exploring which cry type has the most similar cries, statistically significant differences could not be found even though all parameters have been included. However, some conclusions may be drawn by trend (figure 8.2).

First of all, comparing spontaneous and pain-induced cries, the spontaneous cries are more reliable in 15 out of 19 parameters. This seems to refute the expectation that pain-induced cries might be more reliable as the trigger of cries is more standardized than for spontaneous cries.

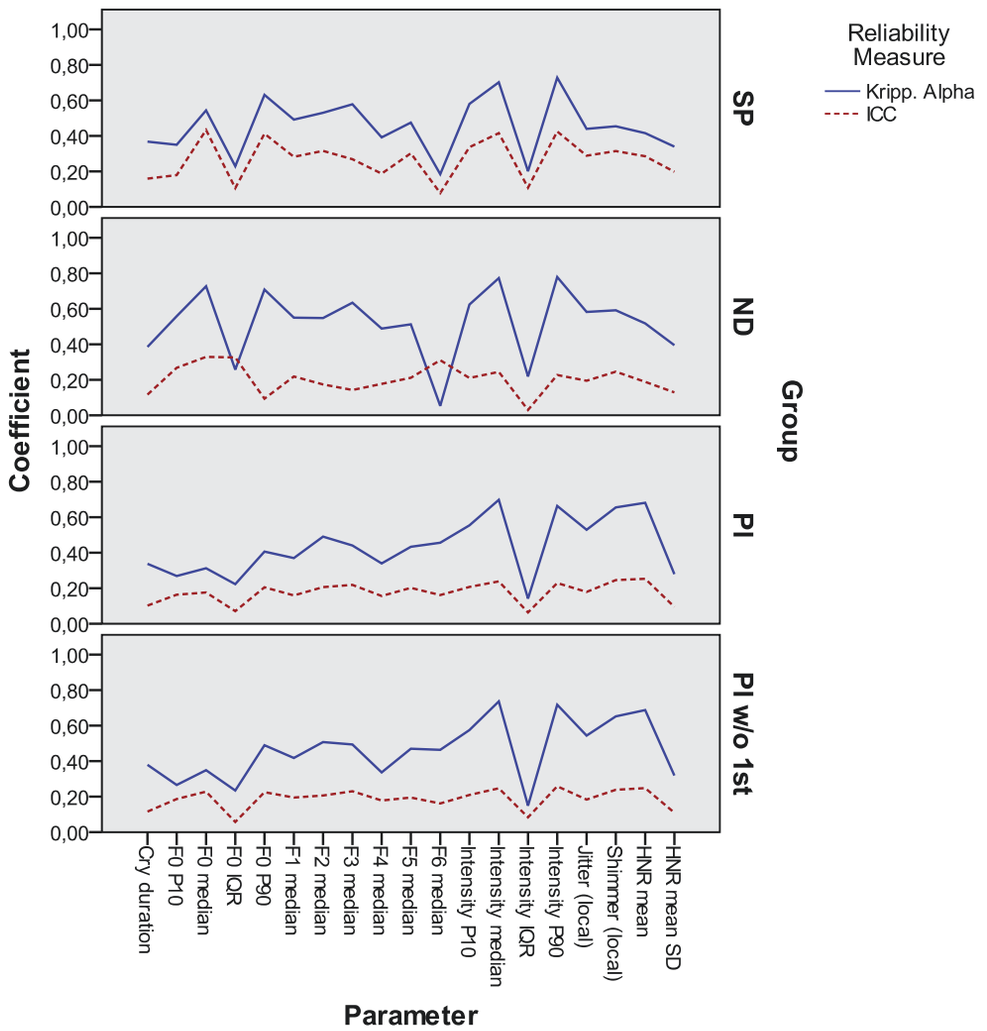
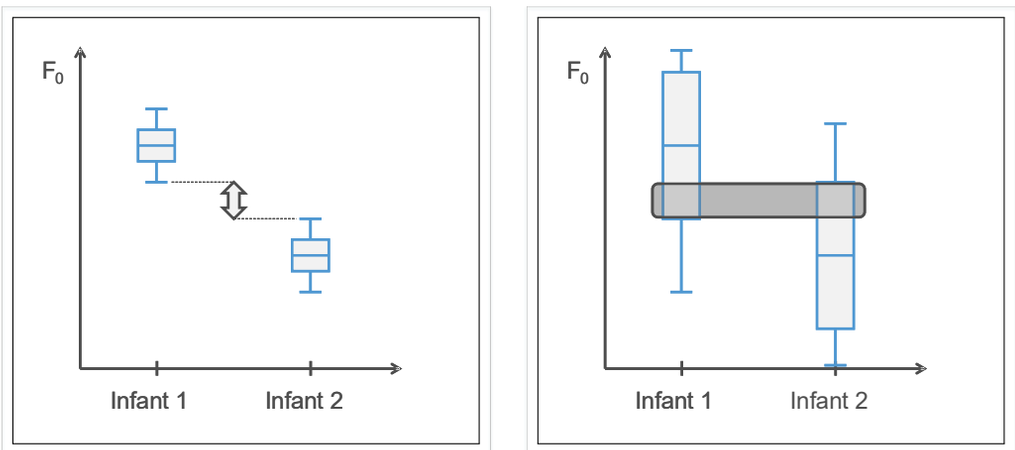


Figure 8.3.: Comparison of Krippendorff's Alpha and Intraclass Correlation Coefficient for all 19 acoustic parameters over the four groups of cry types.



(a) High Krippendorff's Alpha value **(b)** Low Krippendorff's Alpha value

Figure 8.4.: Krippendorff's Alpha influencing the variability of acoustic parameters

Scrutinizing the pain-induced cries (*PI*) and the subgroup of the pain-induced cries without the first cry (*PI w/o 1st*), the *PI w/o 1st* subgroup reached better values in 16 out of 19 parameters by trend. However, the differences between both groups are only marginal in most of the cases. By these marginal differences, it could not be confirmed that excluding the first cry after the pain stimulus clearly improves reliability of pain-induced cries as was expected due to the findings of Runefors et al. (2000) about the differences between the first cry and the remaining ones.

Comparing the spontaneous cries (*SP*) to their subgroup, the non-distressed cries (*ND*), the non-distressed cries reached better alpha values in 18 out of 19 parameters by trend. This finding might reflect the clean acoustic structure of non-distressed cries in contrast to spontaneous cries in general, which for example often include non-harmonic parts. Compared to all other groups in this study, the non-distressed cry provided the highest alpha values by trend for 16 of 19 acoustic parameters stating the non-distressed cry as the preferable type of crying in infant cry analysis.

For non-distressed cries, acoustic parameters describing the fundamental frequency (F0 P10, F0 median, F0 P90) and the intensity (Intensity P10, Intensity median, Intensity P90) provide the best reliability by trend. Most of the formants (F1 to F5), the jitter, shimmer and the HNR mean still show acceptable reliability. Acoustic parameters like the cry duration, the interquartile ranges of F0 and intensity, the sixth formant and the HNR mean SD have poor reliability values and should therefore be used with caution in infant cry analysis.

If only the acoustic parameters with the highest differences between groups (F0 P10, F0 median, F0 P90) are included in the analysis, the *ND* group compared to the *PI* and *PI w/o 1st* group have

significantly higher Krippendorff's Alpha values. Therefore, fundamental frequencies of infants seem to be more stable for non-distressed cries than for pain cries, supporting a preference for the non-distressed group.

Considering Krippendorff's Alpha and ICC, the results of the comparison can be interpreted as following: in figure 8.3, the ICC line follows the shape of the Krippendorff's Alpha line quite fairly. In addition, Spearman's coefficient revealed significant, moderate correlation between Krippendorff's Alpha and ICC. Therefore, results of the intraclass correlation are comparable to those computed by Krippendorff's Alpha. This supports the validity of the present findings. However, ICC values are consistently lower than Krippendorff's Alpha. Krippendorff's Alpha can deal with missing values naturally, whereas ICC needs to replace them by group means. Therefore, Krippendorff's Alpha might be the more valid coefficient here.

For the results of this study there are two main threats to validity:

a) only healthy infants were included and,
b) group sizes were not equal.

Regarding point a): In order to explore differences between healthy infants and infants with any kind of disorder, it is important to know which type of cry promises the most consistent acoustic parameters and therefore is better suited for infant cry analysis. Based on analyzing the healthy infant cry, the study provided answers regarding which type to prefer; the spontaneous cry, especially the non-distressed cry should be used. For analyzing differences between healthy infants and infants with disorders, spontaneous cries in general are a good choice, too; at least the healthy group is known to be as consistent as possible, then. When looking for differences between groups, the similarity of cries in the other groups should be explored, too, as this provides a better understanding of how difficult it will be to find differences between the groups. Furthermore, it must be verified if the non-distressed cry occurs in cries of infants with different pathologies in the same way as in healthy cries. Only little research exists about the non-distressed cry in infants with developmental disorders. For infants suffering from cleft lip and palate, it is known that they occur as often as in healthy cries (Wermke et al., 2011). For other groups like infants suffering from asphyxia or brain damage, it must be examined if non-distressed cries occur in these groups as infants from these groups are said to have problems with holding F0-stability (Accardo, 2013; Sahak, Mansor, Lee, Yassin, & Zabidi, 2010; Verduzco-Mendoza et al., 2012).

Regarding point b): Group sizes of the cry types were considered important as Krippendorff's Alpha might be influenced by this factor. For smaller groups it might be easier to achieve better

alpha values as less data items have to be similar. For this reason, this study verified the results by drawing a randomized sample from each group in the size of the smallest group. By this procedure it was possible to get four groups equal in size. Krippendorff's Alpha computation and comparison of groups were repeated for those randomized groups. The results were not very different from the results of the whole groups and did not lead to any other conclusions.

Summarizing, the reliability of infant cries is the best for spontaneous cries, especially for non-distressed cries. The results of this work indicate that using the non-distressed cry may be the best choice when homogeneity of cries is required. However, non-distressed cries are still ranging from very low similarity values to acceptable ones.

8.2. Reliability of pathological infant cries

In order to verify the findings of section 8.1 for infants with various pathologies, the approach described in section 8.1 was applied to the cries of pathological infants, too.

Since the approach for analyzing the reliability of pathological cries is very similar to the one described in section 8.1.1, the method section only describes points where the method differs. The remaining parts of the method are equivalent to the ones described before.

8.2.1. Method

Subjects

Cries of infants suffering from unilateral cleft lip and palate (UCLP), hearing impairment (HI), laryngomalacia (LA), asphyxia (AS) and brain damage (BD) were analyzed with Krippendorff's Alpha. 19 infants were hearing impaired with a threshold above 60 dB HL. 10 infants had a unilateral cleft lip and palate. 3 infants suffered from asphyxia, 2 infants with brain damage and 4 infants with laryngomalacia were part of the dataset. Inclusion and exclusion criteria for all infant groups are described in section 3.2.

Grouping of cries by type

For infants with pathologies, only spontaneous cries could be included in this thesis for the following reason: in their early days, when infants have medical vaccinations, pathological conditions haven't often been completely confirmed yet. For example, the degree and type of hearing impairment is often determined later when most vaccinations have already been applied. Eliciting additional pain stimuli without medical needs are unjustifiable for ethical reasons. For this reason and because spontaneous cries of healthy infants were the most reliability ones, reliability for pathological cries was tested on spontaneous cries only.

In section 8.1, the sub-group of non-distressed cries was identified within the group of spontaneous cries. Non-distressed cries by trend achieved higher reliability values than the average spontaneous cry. However, acoustical analyses of pathological cries revealed that for some of the clinical pictures infants did not produce non-distressed cries at all.

Figure 8.5 shows how high the percentage of non-distressed cries is in comparison to the spontaneous cries ratio for the different infant groups.

For healthy infants, from 35 healthy infants included in the study described in section 8.1.1, 268 spontaneous cries could be recorded and 115 of these cries were identified as non-distressed cries. Hence, 43 % of the spontaneous cries of healthy infants fall into the category of non-distressed cries.

For infants with pathological developments, the ratio of non-distressed cries is as follows: For infants suffering from cleft lip and palate, an average of 39 % of non-distressed cries (11 out of 28 spontaneous cries) was identified. Infants suffering from hearing impairment had 34 % of non-distressed cries (14 out of 41 spontaneous cries), infants suffering from laryngomalacia had 30 % (21 out of 70 spontaneous cries). An average of 10 % was reached by infants suffering from asphyxia (2 out of 19 spontaneous cries) and infants with brain damage showed no non-distressed cries out of 24 cries.

Because infants with pathological conditions show no or at least less non-distressed cries than healthy infants, the group of spontaneous cries was chosen to verify reliability for pathological cries. Spontaneous cries reached similar reliability values as non-distressed cries and are produced by infants with pathological conditions in a sufficient amount. A grouping and sub-grouping as conducted in the previous study was omitted for testing reliability for pathological cries.

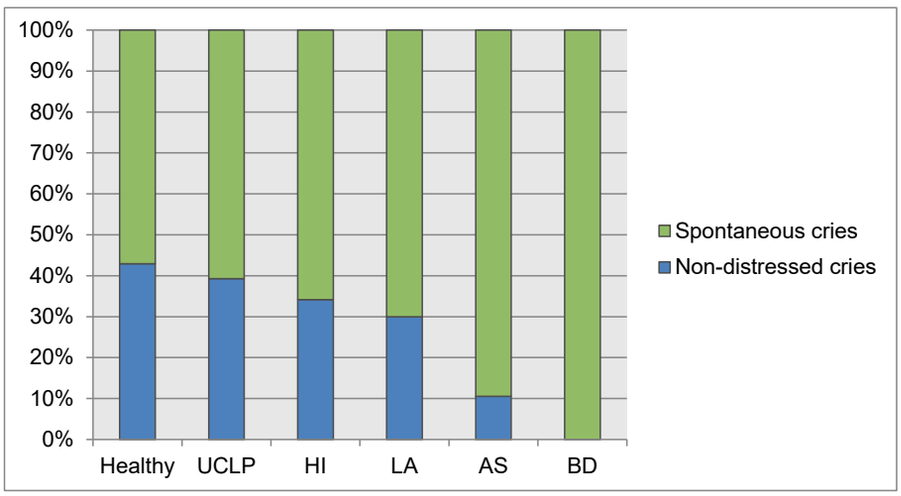


Figure 8.5.: Ratio of non-distressed cries compared to spontaneous cries for the pathological infant groups

8.2.2. Results

Reliability of acoustic parameters

Krippendorff's Alpha was computed for the same 19 acoustic parameters as described in section 8.1.2. Table 8.5 summarizes the results of Krippendorff's Alpha computation for the pathological cries.

The Krippendorff's Alpha values for the healthy group are from the study of spontaneous cries described in section 8.1.2. Krippendorff's Alpha values are interpreted in the same way as suggested in section 8.1.1.

For the cries of infants suffering from unilateral cleft lip and palate, non of the 19 acoustic parameters reached good alpha values ($\alpha > 0.667$). Acceptable values ($\alpha > 0.4$) were reached for 4 acoustic parameters: F1 median, F6 median, intensity median and intensity P90. The other 14 acoustic parameters reached values between a slight and a fair agreement (α between $0.0 - 0.4$). Only F0 P90 reached a poor agreement ($\alpha < 0.0$).

Infants suffering from hearing impairment reached good Krippendorff's Alpha values with $\alpha \geq 0.667$ in 3 out of 19 acoustic parameters. These values were reached for the intensity P10 and intensity P90 as well as for the intensity median. The formants F4 and F5 and the intensity IQR reached acceptable values. The other 13 acoustic parameters achieved values between a slight and fair agreement.

Table 8.5.: Krippendorff's Alpha values for spontaneous cries of infants from different groups

	Krippendorff's Alpha for Group					
Parameter	**Healthy**	**CLP**	**HI**	**LA**	**AS**	**BD**
Cry duration	0.368	0.191	0.115	0.054	0.157	0.876
F0 P10	0.35	0.237	0.111	0.167	0.163	0.69
F0 median	0.544	0.073	0.289	0.231	0.275	0.591
F0 IQR	0.229	0.267	0.089	0.192	0.071	0.323
F0 P90	0.631	-0.146	0.272	0.166	-0.014	-0.146
F1 median	0.492	0.532	0.327	0.248	0.225	0.691
F2 median	0.53	0.179	0.223	0.13	0.032	0.534
F3 median	0.578	0.126	0.281	0.145	0.179	0.507
F4 median	0.392	0.227	0.425	0.383	-0.01	0.095
F5 median	0.475	0.308	0.402	0.077	0.04	0.041
F6 median	0.184	0.556	0.273	0.387	-0.018	0.077
Intensity P10	0.58	0.281	0.813	0.615	0.124	0.404
Intensitiy median	0.702	0.404	0.786	0.615	0.156	0.294
Intensitiy IQR	0.201	0.076	0.44	0.298	0.145	0.699
Intensitiy P90	0.728	0.543	0.805	0.493	0.122	-0.014
Jitter (local)	0.439	0.204	0.308	0.287	0.161	0.22
Shimmer (local)	0.454	0.25	0.315	0.206	0.029	0.031
HNR mean	0.416	0.21	0.105	0.414	0.004	0.239
HNR mean SD	0.339	0.308	0.12	0.053	0.554	0.026

For the group of infants with laryngomalacia, no good agreements were reached. Intensity P10, intensity P90, intensity median as well as the mean of the HNR got acceptable alpha values. The 15 remaining acoustic parameters are between a slight and a fair agreement.

The asphyxia cry group achieved no alpha values above 0.667 and therefore had no good agreement of the acoustic parameters. Acceptable agreement was reached for the HNR mean SD. 15 acoustic parameters lay between 0.0 – 0.4. The F0 P10, the F4 median and the F6 median got values less than ($\alpha > 0.0$) and therefore reached a poor agreement.

The group containing cries of infants with brain damage achieved good agreement in 4 out of the 19 acoustic parameters: the cry duration, the F0 P10, the F1 median and the intensity IQR. Values above 0.4 were reached for the F0 median, the median of F2 as well as for the median of F3 and the intensity P10. 9 acoustic parameters had values between a slight and a fair agreement. Poor agreement was reached by 2 acoustic parameters: the F0 P90 and the intensity P90.

Differences in reliability between health states

Comparing the Krippendorff's Alpha values of infant cries in case of developmental diseases and cries of healthy infants, the pathological cries show more values of Krippendorff's Alpha between a poor ($\alpha < 0.0$), slight ($0.0 \leq \alpha \leq 0.2$) and fair ($0.2 \leq \alpha \leq 0.4$) agreement than the healthy infant cries. Hence, acceptable or good agreement of the Krippendorff's Alpha values of the acoustic parameters were achieved more rarely compared to the healthy infant cry.

Figure 8.6 visualizes the ratio between the Krippendorff's Alpha values of the healthy cry group and the pathological cry groups. The figure shows that healthy infant cries show values < 0.4 in 7 out of the 19 calculated acoustic parameters. The asphyxia cry group had 17 parameters under 0.4. Infants with unilateral cleft lip and palate as well as infants suffering from laryngomalacia reached values less than 0.4 in 15 out of the 19 acoustic parameters. For the hearing impaired group 13 acoustic parameters and for the brain damage group 11 acoustic parameters reached values under 0.4.

8.2.3. Interpretation

In summary, cries of infants with diverse pathologies are less similar to each other and therefore less reliable then healthy infant cries. For the healthy infant cries, most of the computed acoustic parameters show acceptable or good values and only a few parameters have poor reliability values.

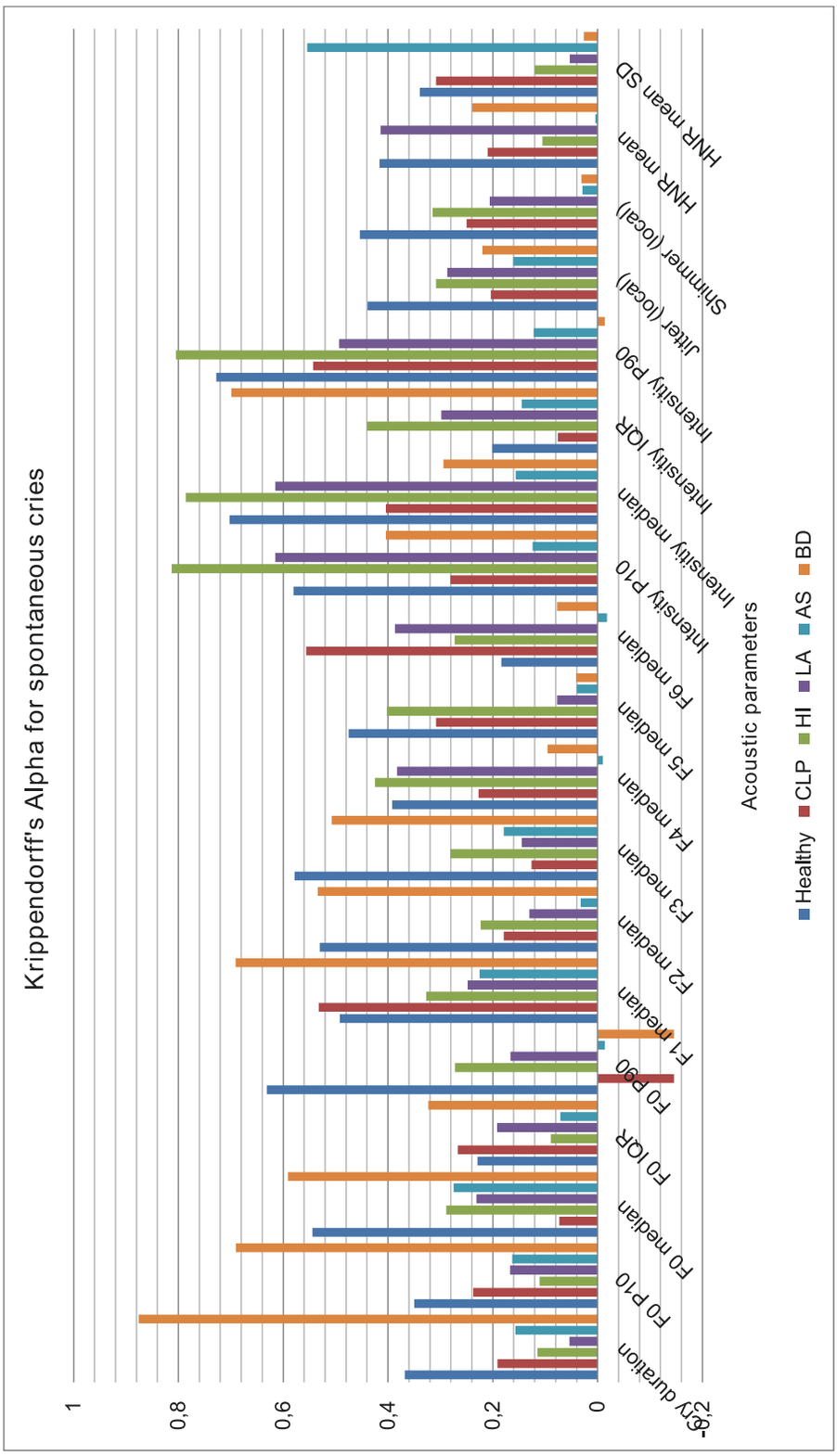


Figure 8.6.: Krippendorff's Alpha values for spontaneous cries of infants from different groups

For the pathological cries, most of the calculated acoustic parameters show low Krippendorff's Alpha values and therefore poor reliability. Pathological cries have a greater spread than healthy infant cries. Cries having a greater spread and hence a bigger variability will need more robust methods for classifying cries according to their health state.

8.3. Summary of the findings regarding the reliability of infant cries

Research question 1 — which type of crying is best suited for infant cry analysis — was answered in this chapter. To estimate the degree of consistency among single cries in an episode of crying, healthy infants were allocated to the four groups spontaneous cries, spontaneous non-distressed cries, pain-induced cries and pain-induced cries without the first cry after pain stimulus. 19 acoustic parameters were computed and statistically analyzed on their reliability with Krippendorff's Alpha.

Krippendorff's Alpha values between 0.184 and 0.779 were reached in all groups. No significant differences between the cry groups were found. However, the non-distressed cries reached the highest alpha values in 16 out of 19 acoustic parameters by trend. The results show that the single cries within an infant's episode of crying are not very reliable in general. For the cry types, the non-distressed cry is the one with the best reliability making it the preferred option for infant cry analysis analyzing healthy cries.

For the cries of infants suffering from various pathologies, the non-distressed cry exists very rarely; e.g., cries of infants suffering from brain damage contained no non-distressed cries at all.

For this reason, the spontaneous cry as the second reliable one was chosen in this thesis for analyzing cries of healthy infants and infants suffering from pathologies. Krippendorff's Alpha values between 0.004 and 0.876 indicate that spontaneous cries are only slightly less reliable than non-distressed cries.

Chapter 9.

Validity of infant cry analysis

Many previous studies defined various characteristic acoustic properties for certain pathological developments (e.g., LaGasse et al., 2005). For example, infants suffering from unilateral cleft lip and palate show malformations of the oral and pharyngeal systems and were said to show differences in the formant frequencies as they are influenced by the shape of the vocal tract (Wermke, Mende, et al., 2002). Infants suffering from hearing impairment were said to have longer cry durations and higher energy levels of their crying, because they do not realize their crying through the auditory system, but through their tactile sense (Möller & Schönweiler, 1999).

In this chapter, the second research objective — to assess, if valid methods exist for classifying the state of health of infants by acoustic parameters — was explored.

In infant cry research, mainly two approaches to rate the health state of infants by acoustic parameter have been used: the *acoustic analysis* of cry signals and the *auditory discrimination* of cries by listeners. This chapter explores and compares the validity of both approaches.

For automated infant cry classification based on statistically processing the acoustic parameters of infant cries, two approaches have been applied by researchers:

1. Identifying significant differences in acoustic parameters across groups of infants with varying health states using analysis of variances (ANOVA) and trying to find cut-off values separating healthy infants form unhealthy ones.
2. Training supervised-learning models on acoustic parameters of infant cries for which the health state of the infants is known to automatically extract rules from the data in order to classify infant cries with unknown health state.

In this thesis, both approaches are analyzed concerning their validity for infant cry classification. Section 9.1 uses a One-Way-ANOVA to find significant differences between cries of healthy infants and infants suffering from hearing impairment and unilateral cleft lip and palate. Although significant differences were found, they were not specific enough to be suited for infant cry classification.

In section 9.2, supervised-learning algorithms that have been used in infant cry classification and additional algorithms that have not been used yet are identified. The algorithms are reviewed in a systematic way to rate which algorithms are suited for infant cry classification.

Section 9.3 explores the ability of humans to identify the health states of infants by listening to their crying. A listening experiment was conducted and the performance of the human listeners to distinguish healthy cries from pathological ones as well as to separate between various pathologies is analyzed. The results are compared to automatic classification approaches.

Section 9.4 summarizes and interprets the findings and provides the answers for the research questions 2 and 3 that were postulated for the second research objective if valid methods exist for infant cry classification.

9.1. Validity of infant cry classification using analysis of variances[1]

This explorative study aimed to compare acoustic parameters of cries of infants with two well-defined infant disorders to cries of typically-developing infants. The research objective was to identify acoustic parameters in infant crying that are specific for certain developmental diseases. The cries of healthy infants were compared to the cries of infants with hearing impairment (HI) and with unilateral cleft lip and palate (UCLP). Infants with hearing impairment suffer from dysfunctions of (parts of) the hearing system and have delayed onset of babbling and cooing (Eilers & Oller, 1994; Koopmans-van Beinum, Clement, & van den Dikkenberg-Pot, 2001; Scheiner, Hammerschmidt, Jürgens, & Zwirner, 2004). In addition, hearing impairment may lead to disturbances in speech and language acquisition (M. P. Silva, Comerlatto, Bevilacqua, & Lopes-Herrera, 2011). UCLP originates from malformations in the oral-pharyngeal system and often leads to articulatory disorders during language acquisition (Prandini, Pegoraro-Krook, de Cassia Rillo Dutka, & de Castro Marino, 2011). UCLP and hearing impairment were selected as two exemplary pathological disorders already examined regarding their impact on acoustic parameters of the infant cry.

[1] The results of the study described in this section were published in Etz, T., Reetz, H., & Wegener, C. (2012). A classification model for infant cries with hearing impairment and unilateral cleft lip and palate. *Folia Phoniatrica et Logopaedica*, *64*(5), 254–261. doi:10.1159/000343994

As both disorders have different origins, it was assumed they might differ in their influence on acoustic parameters; not only compared to healthy infants but also compared to each other. The influence of only one factor in comparison to the healthy infant cry has already been examined by previous studies (Lester & Boukydis, 1985).

In this study, significant differences between each of the groups were identified by analysis of variance (ANOVA). In addition (and in contrast to prior research), it was tested if these differences in the cries were specific enough to classify them into the three groups Healthy, HI or UCLP.

9.1.1. Method

Subjects

A total of 29 infants were included in the study. 10 infants were found to be healthy by paediatricians at the routine postpartum examination. The infants had no adumbration for cleft or further anomalies. The hearing function of these infants was assessed by otoacoustic emissions of both ears. No hearing impairment was found. Healthy infants were between 1 and 6 months of age (mean: 3.2 months). Profound hearing impairment was found in another 9 infants by otoacoustic emissions and evoked brainstem response. Hearing impairment was verified by confirmation diagnosis (threshold of -60 dB HL). The infants were between 1 and 4 months of age (mean: 2.0 months). 10 infants had a unilateral cleft lip and palate and no adumbration for a hearing impairment. These infants were between 4 and 7 months of age (mean: 5.5 months). Further exclusion criteria for all groups were (1) low birth weight ($\leq$ 1500 g), (2) gestational age < 37 weeks at birth and (3) any diagnosis influencing normal development. After questioning the parents, there was no indication that the participants had an existing cold or middle ear disease at the time of recording. All parents of the infants were native speakers of German and gave written informed consent to participate in this study. The study was confirmed by the Ethic Review Committee of the Fresenius University of Applied Science (appendix B).

Data acquisition

Infant cries were recorded as described in chapter 4 and single cry utterances were extracted from each period of crying as described in chapter 5.

Acoustic analysis

The infant cries were analyzed acoustically with Praat Software 5.2.22. A cry utterance was defined to be a cry during the expiratory phase lasting at least 0.4 seconds. Altogether 128 cries were extracted. From both the Healthy group and the UCLP group, 43 cries were extracted. For the HI group, 42 cries were gathered.

For each cry utterance, the following acoustic parameters were computed with Praat: fundamental frequency (F0) and intensity (for both, the minimum and maximum value was computed, as well as the median for the whole cry), cry duration, the first six formants as references to frequency ranges with high spectral intensities (F1–F6, the upper frequency for this analysis was set to 8000 Hz), the micro-variability of vocal fold vibration (shimmer and jitter) and the harmonics-to-noise ratio (HNR) mean and standard deviation.

Settings and reasons for the acoustic parameters have been described in chapter 6.

Statistical analysis

The acoustic parameters of the infant cries were analyzed using the software SPSS Statistics 19 (IBM, 2011). An explorative data analysis revealed that 10 of 16 parameters were not normally distributed. According to Bortz (2010), parametric tests are still applicable, if group sizes are similar, as in the present study. As the age of the infants was significantly different between groups, Spearman's Rho correlation coefficient was computed to test the effect of age on the acoustic parameters. Analyses of variance (One-way ANOVAs) were used to test if group membership (one factor with three groups: Healthy, UCLP and HI) had significant impact on the dependent variables (the acoustic cry parameters). Sheffé post-hoc tests were used to identify the source of significant effects.

9.1.2. Results

Correlation analysis

First, the influence of the age on the acoustic parameters was tested. Spearman's Rho found low correlations (r < 0.3) between the different ages and 7 acoustic parameters: Cry duration ($r = 0.214,\ p = 0.015$), F0 median ($r = -0.175,\ p = 0.048$), F2 ($r = -0.189,\ p = 0.033$), F4

Table 9.1.: Results of the analysis of variance (One-way ANOVA)

	Mean Square between groups	F	Sig.
Cry duration [s]	**3.924**	**9.445**	**.000**
F0 min [Hz]	9 472.103	.863	.424
F0 median [Hz]	**54 647.603**	**7.906**	**.001**
F0 max [Hz]	**334 136.844**	**18.071**	**.000**
F1 [Hz]	75 176.388	.605	.548
F2 [Hz]	**1 222 349.403**	**9.091**	**.000**
F3 [Hz]	84 551.031	.302	.740
F4 [Hz]	**2 122 496.890**	**10.072**	**.000**
F5 [Hz]	5 791.980	.022	.978
F6 [Hz]	274 182.173	2.923	.057
Intensity min [dB]	**260.681**	**5.336**	**.006**
Intensity median [dB]	76.894	1.598	.206
Intensity max [dB]	73.185	1.518	.223
Jitter [%]	**4.089**	**4.527**	**.013**
Shimmer [%]	**166.104**	**12.614**	**.000**
HNR mean [dB]	**299.580**	**8.963**	**.000**
HNR standard deviation [dB]	.832	.394	.675

($r = 0.218,\ p = 0.014$), jitter ($r = -0.209,\ p = 0.018$), shimmer ($r = -0.185,\ p = 0.037$) and HNR mean ($r = 0.268,\ p = 0.002$).

Analysis of variances

Table 9.1 shows the results of the analyses of variance. Statistically significant differences are highlighted in bold. The following cry parameters are different between the cry groups: F0 median ($F = 9.445,\ p = 0.001$), F0 max ($F = 18.071,\ p \leq 0.0001$), F2 ($F = 9.091,\ p \leq 0.0001$), F4 ($F = 10.072,\ p \leq 0.0001$), Intensity min ($F = 5.336,\ p = 0.006$), jitter ($F = 4.527,\ p = 0.013$), shimmer ($F = 12.614,\ p \leq 0.0001$) and HNR mean ($F = 8.963,\ p \leq 0.0001$).

Table 9.2 summarizes the results of the Scheffé post-hoc comparisons for this test. Significant differences are highlighted in bold.

Compared to healthy infants, those with UCLP had a higher F0 median ($p = 0.001$) and max ($p \leq 0.0001$), a higher shimmer ($p \leq 0.0001$) and jitter ($p = 0.017$), and a lower HNR mean ($p = 0.003$).

Table 9.2.: Summary of significant pairwise differences between the three groups

	Pairwise comparison[a]								
	Healthy / UCLP			***Healthy / HI***			***UCLP / HI***		
Parameter	**Healthy mean (SD)**	**UCLP mean (SD)**	**Sig.**	**Healthy mean (SD)**	**HI mean (SD)**	**Sig.**	**UCLP mean (SD)**	**HI mean (SD)**	**Sig.**
Cry duration [s]	1.13 (0.08)	1.05 (0.07)	0.855	**1.13 (0.08)**	**1.62 (0.14)**	**0.003**	**1.05 (0.07)**	**1.62 (0.14)**	**0.001**
F0 median [Hz]	**376.31 (9.85)**	**447.46 (13.73)**	**0.001**	376.31 (9.85)	407.88 (14.22)	0.220	447.46 (13.73)	407.88 (14.22)	0.094
F0 max [Hz]	**437.85 (9.37)**	**614.15 (27.98)**	**0.000**	**437.85 (9.37)**	**525.82 (20.71)**	**0.014**	**614.15 (27.98)**	**525.82 (2.71)**	**0.013**
F2 [Hz]	2292.74 (49.77)	2295.16 (60.17)	1.000	**2292.74 (49.77)**	**1999.62 (58.01)**	**0.002**	**2295.16 (60.17)**	**1999.62 (58.01)**	**0.001**
F4 [Hz]	4631.89 (65.33)	4737.21 (77.57)	0.569	**4631.89 (65.33)**	**5061.35 (67.15)**	**0.000**	**4737.21 (77.57)**	**5061.35 (67.15)**	**0.006**
Intensity min [dB]	56.66 (1.13)	53.92 (1.20)	0.195	**56.66 (1.13)**	**51.72 (0.83)**	**0.006**	53.92 (1.20)	51.72 (0.83)	0.353
Jitter [%]	**0.78 (0.08)**	**1.38 (0.18)**	**0.017**	0.78 (0.08)	0.94 (0.16)	0.762	1.38 (0.18)	0.94 (0.16)	0.105
Shimmer [%]	**4.30 (0.42)**	**8.23 (0.57)**	**0.000**	**4.30 (0.42)**	**6.37 (0.65)**	**0.035**	8.23 (0.57)	6.37 (0.65)	0.065
HNR mean [dB]	**16.02 (0.88)**	**11.61 (0.74)**	**0.003**	16.02 (0.88)	16.35 (1.02)	0.966	**11.61 (0.74)**	**16.35 (1.02)**	**0.001**

[a] Summary of the results of the Sheffé post-hoc comparisons, and group means and standard deviations (SD). Cry parameters not being significantly different between any of the groups are not listed in the table

Compared to healthy infants, those with HI had a longer cry duration ($p = 0.003$), a higher F0 max ($p = 0.014$), a higher F2 ($p = 0.002$) and F4 ($p \leq 0.0001$) and a higher shimmer ($p = 0.035$).

Compared to infants with UCLP, those with HI had a longer cry duration ($p = 0.001$), a higher F0 max ($p = 0.013$), a higher F2 ($p = 0.001$) and F4 ($p = 0.006$), and a higher HNR median ($p = 0.001$).

9.1.3. Interpretation

Compared to the results of previous studies on infants with hearing impairment as compared to typically-developing infants, the study found similar results for most of the acoustic differences. For example, longer cry durations and a lower F2 for infants with HI compared to healthy infants were found confirming the findings of Möller and Schönweiler (1999). Additionally, the study identified higher values for F4 (the fourth formant was not examined in the study of Möller & Schönweiler). According to Jones (1971), significant differences related to F0 were discovered: for infants with hearing impairment, higher values in F0 maximum were observed, whereas Jones (1971) reported higher values in F0 median. Additionally higher shimmer values were identified for the HI group.

Infants in the UCLP group showed a higher F0 median and F0 maximum compared to healthy infants.Wermke, Hauser, et al. (2002) found increased F0 values in one of five examined patients at the age between one and three months. In addition to the higher F0 median and maximum, differences in the microvariability of vocal fold vibration (shimmer and jitter) as well as in the harmonics-to-noise ratio were found.

Only one acoustic parameter – the F0 maximum – was significantly different between each of the three groups. In addition, looking at the range between the 75- and 25-percentile of F0 maximum (table 9.3) revealed that the range of the Healthy group and the HI group overlap. In this range the central 50 % of the cries are located. That is, even if the median between both groups is significantly different, the cries are still too similar to be clearly separated by F0 maximum only. It was concluded that only looking for significant differences between groups is not enough to separate cries by their acoustic parameters. In order to separate the groups, multivariate learning algorithms should be used allowing to look at multiple acoustic parameters when trying to separate groups.

Table 9.3.: 75-, 50- and 25-Percentile of F0 maximum

			Percentiles		
		Group	**25**	**50**	**75**
Tukey's Hinges	F0 max [Hz]	Healthy	398.88	421.95	**488.05**
		UCLP	497.53	549.55	654.13
		HI	**443.40**	484.02	538.02

9.2. Validity of infant cry classification using supervised-learning models[2]

To overcome the problems that were identified in the last section when trying to classify infant cries using analysis of variances, more effective statistical approaches were selected for classification. Supervised-learning algorithms were identified as an appropriate approach for the classification of infant cries. The supervised-learning approach was introduced and justified in chapter 7. A short repetition: supervised-learning algorithms are trained on a training dataset containing predictor variables (e.g., acoustic features of an infant cry) for which the classification result is known (e.g., the state of health of an infant). After the training phase has been finished, the classification model can be used to classify data for which the result is not known. By this, an infant's state of health can be predicted by applying a classification model on the acoustic features measured for the infant's cry.

Results of previous studies applying supervised-learning models for infant cry classification are promising. However, there is no consent which classification models are suited best for infant cry classification. Available reports in literature are not comparable as models were trained on different data. This section aims at providing a systematic comparison of various models used for infant cry classification.

A *systematic classification model comparison* was conducted with a literature review aiming at identifying all possible classification models and at comparing them systematically. For this purpose, the review consisted of two major parts, a literature search and the systematic model rating. The literature itself was not reviewed (as would be done in systematic literature reviews), the *classification models* used were identified. Next, the classification models were applied to a common reference dataset and rated systematically.

[2]The results of the study described in this section were published in Fuhr, T., Reetz, H., & Wegener, C. (2015). Comparison of supervised-learning models for infant cry classification. *International Journal of Health Professions*, *2*(1), 4–15. doi:10.1515/ijhp-2015-0005

The remaining section is structured as follows: Section 9.2.1 introduces the structure of the systematic literature search and the classification model review. Section 9.2.2 presents the major results of this review. Models and model ratings are provided. Finally, section 9.2.3 interprets the ratings and provides recommendations for future use of classification models in infant cry research.

9.2.1. Method

The model review was divided into two phases. First, a systematic literature search was conducted to identify which classification models have already been used in infant cry research. Second, models were compared and rated on their performance in infant cry classification in a model review.

Systematic literature search

For the literature search, the research question as well as inclusion and exclusion criteria were specified.

For the literature search and the model review, the following major research question was stated: Which supervised-learning classification model is suited best to classify infant cries according to their state of health? This main research question was split into two subquestions: (a) Which supervised-learning classification models have already been used in infant cry research? and (b) Which one of these models is suited best to classify infant cries according to their state of health?

Inclusion and exclusion criteria were defined. Both sets cover criteria for the publication as well as for the model approach itself. The included publications had to describe the application of classification models in infant cry research and they had to be written in English or German. The described modeling approach had to be applicable to metric explanatory variables and to nominal predictor variables and had to be robust against non-normally distributed data.

Publications were excluded if they did not provide a sufficient description of the modeling approach in order to understand how the approach worked.

For the literature search, 7 online libraries and indexes covering the research field were selected: ACM digital library (ACM, 2014), DBLP (DBLP, 2014), IEEE XPlore (IEEE, 2014), SSG (GBV, 2014), DIMDI (DIMDI, 2014), Medpilot (Deutsche Zentralbibliothek für Medizin, 2014) and

Web of Knowledge (Reuters, 2014). All data sources were searched for articles describing the application of classification models in infant cry research.

For the search strategy the search term was composed to cover major keywords relevant for the research question, synonyms and related terms as well as broader terms and narrower terms in order to find all the relevant literature. Based on these terms, the following search string was composed:

```
(classif* OR predict* OR forecast*
    OR "machine*learning" OR "supervised*learning")
AND (model* OR algorithm* OR approach*)
AND (infant* OR baby OR babies OR newborn* OR neonate*)
AND (cry* OR cries)
```

Applying this search string to the libraries sometimes required adaption to the specific search engine (e.g., some libraries used the symbols '`&&`' instead of the keyword '`AND`'; others did not support wildcards like '`*`' which required to write down all possible morphological variants of a word, e.g., "predictive", "prediction", "predictions", "predicting" instead of "predict*").

After defining the setting, the literature search was conducted. The result sets of articles from each data source were collected in one central bibliography. The articles that were gathered by the initial search were scanned (title, abstract) and obviously irrelevant articles were excluded. In addition, duplicates were identified and removed. The references of the previously selected articles were explored to identify additional literature.

Concluding the literature search, all articles that remained in the bibliography after the filtering were read in detail, focusing on understanding what models were applied to infant cry classification and how these models work. The relevance of the articles was confirmed and supervised-learning classification models were extracted.

Classification model review

For conducting the model review, a common framework for comparing and rating the classification models was defined. Criteria for classification model quality were identified and weightings were provided. Table 9.4 summarizes the rating scheme that was defined. All criteria were rated by an ordinal value between 0 (lowest rating) and 4 (highest rating). Values that were of metric nature were categorized into these five ratings. Rating criteria are described in the following.

Accuracy is the most important aspect for rating the classification models. It is defined as the *precision* of the model on the *test dataset* (a sample not used for training the classification models, but only for validating their accuracy; see section 9.2.1).

$$\text{Accuracy} = \frac{N_{11}}{N}$$

where N_{11} is the number of infant cries that were classified correctly and N the overall number of cries. Higher accuracy values on the test dataset lead to better ratings. To be comparable with the other criteria, the accuracy value was categorized into an ordinal accuracy category *Acc* from 0 to 4. 0 is the worst accuracy category (accuracy below 50 %) and 4 is the best (accuracy above 96 %).

The *degree of overfitting* describes how much the model is generalizable to classify unknown cries correctly. It is computed as the difference between the accuracy of the model on the test dataset and the training dataset (section 9.2.1):

$$\text{Degree of overfitting} = \text{Accuracy}_{Test} - \text{Accuracy}_{Training}$$

Small values indicate better generalizability. The degree of overfitting was categorized, too. The $OFit$ categories are from 0 (the worst category, degree of overfitting higher than 30 %) to 4 (the best category, degree of overfitting smaller than 5 %; negative values are allowed and fall into the best category, too).

Conformability (Conf) depicts how well a classification model can be conformed by experts. Ratings are given in categories from 0 (no conformability) to 4 (the highest conformability category). Basic conformability (category 1) requires that experts can understand how the prediction of the model came about (i.e., the way a data item was categorized is transparent). For higher conformability values, the model must provide information about the importance of the exploratory variables (category 2) or about cut-off values describing what value ranges are typical for a predicted group (category 3; here, the knowledge about value ranges was rated higher than the knowledge about which variables were most influencing for model predictions). If feature importance as well as cut-off values are provided, the highest rating (category 4) is given. Higher conformability ratings are better for evaluating the correctness of the classification model. As the rating of conformability might be subjective, this factor was rated by two independent reviewers. In cases where ratings differed between reviewers, the ratings were discussed and one rating acceptable to both reviewers was picked.

Table 9.4.: Rating scheme for the systematic classification model review

		Fullfillment	
Aspect	**Importance factor**	**Category**	**Points**
Accuracy (*Acc*)	2	96 % - 100 %	4
		91 % - 95 %	3
		81 % - 90 %	2
		51 % - 80 %	1
		< 50 %	0
Degree of overfitting (*OFit*)	1	< 5 %	4
		5 % - 10 %	3
		10 % - 20 %	2
		20 % - 30 %	1
		> 30 %	0
Conformability (*Conf*)	1.5	Cut-off values and feature importance provided	4
		Only cut-off values provided	3
		Only feature importance provided	2
		Basic conformability	1
		No conformability	0

The overall rating (R) was computed as the weighted average of the three categorized criteria:

$$R = \frac{2 \cdot Acc + 1 \cdot OFit + 1.5 \cdot Conf}{4.5}$$

Therefore, the overall rating takes a value between 0 (the worst rating) and 4 (the best rating). The selection of weights is discussed in section 9.2.3.

After having defined the rating scheme, four software systems were evaluated on their ability to compute the classification models that had been identified during the literature search; two proprietary software systems (IBM SPSS Statistics 20 (IBM, 2013b) and IBM SPSS Modeler 15 (IBM, 2013a)) and two open-source systems (R 3.0.2 (The R. Foundation for Statistical Computing, 2014) and RapidMiner 6 (Rapid-I, 2014)). These four systems implemented most of the classification algorithms. Models for which no implementing software was found, had to be excluded from the review.

Classification models were then applied to a reference dataset. The reference dataset contained 468 cry samples of healthy infants as well as infants with various disorders. The reference dataset is described in section 9.2.1 in detail.

The performance of classification models is heavily influenced by defining their parameters correctly. Wrong parameter settings can lead to bad classification performance although the classification approach in general may be well suited for classifying infant cries. To find the best settings for each model, a pre-selection of different settings for each parameter was chosen. The pre-selection for each parameter contained different settings that are suited for the given dataset. For example for decision trees, the parameter "minimum number of items within child tree nodes" influences the complexity of the generated decision trees by pruning sub-trees when they get to fine-grained and lead to a high overfitting. Here, the settings 1 %, 2 % and 5 % were included in the pre-selection. Next, all possible combinations of settings in the pre-selections for all parameters were computed automatically and models with these settings were trained. By this, all relevant combinations of parameter settings were evaluated to find the best parameter setting for each model. The best combination of parameter settings was chosen by automatically comparing the performance of the models.

Classification models were rated according to the rating scheme and results were documented. Model ratings were assessed and suggestions about which classification model approach to use for infant cry classification were provided. The interpretation of these findings is provided in section 9.2.3.

Data selection

Altogether, cry signals of 69 infants were recorded. 31 of these infants were healthy, full-term babies without any pregnancy complications and indication of physical or neurological disorders. 19 infants were hearing impaired with a threshold above 60 dB HL. 10 infants had a unilateral cleft lip and palate (UCLP). 3 infants suffering from asphyxia, 2 infants with brain damage and 4 infants with laryngomalacia were contributed to the dataset. The infants were between 1 and 7 months of age. For all the infants in the pathological groups, it was ensured that they did not suffer from any other disease than the one representing the group. All cries were recorded on a Zoom H2n recorder with a sampling rate of 48 kHz, placed about 30 cm from the infants' mouth. The study was approved by the Ethic Review Committee of the Fresenius University of Applied Science.

Depending on the overall duration of an infant's cry episode, 5 to 11 single cry utterances were extracted from the episode resulting in 468 cry utterances in total. To allow the rating of the model accuracy, the dataset was split into a training dataset used for training a classification model and a separate test dataset used only for rating the model accuracy. 30 % of the cry samples from

Table 9.5.: Number of cry samples per group and dataset

	Healthy	Hearing impaired	UCLP	Asphyxia	Brain damage	Laryngomalacia	∑
Training dataset	200	28	21	11	18	43	321
Test dataset	86	13	7	8	6	27	147
∑	286	41	28	19	24	70	468

each group were allocated to the test dataset by chance; the remaining 70 % of samples formed the training dataset. Table 9.5 provides an overview about group sizes for the training and test dataset.

Acoustic analysis

The cries were acoustically analyzed with Praat Version 5.3.39 (Boersma & Weenink, 2013b). Overall, 19 acoustic parameters were computed for each cry. Settings and reasons for the acoustic parameters have been described in chapter 6.

Cry parameters were aggregated for each cry, i.e., they were stationary and no temporal development was captured (stationary analysis is a common approach in infant cry research; c.f. Branco et al. (2007), Goberman and Robb (2005), K. Michelsson et al. (2002), Rautava et al. (2007), Robb et al. (2007)).

9.2.2. Results

Systematic literature search

Overall, 579 articles were found by the initial search. Table 9.6 summarizes how many articles were found in each database. 64 out of 579 articles remained in the bibliography after filtering (31 articles of medical bibliography databases and 33 articles of computer science bibliographies). No additional literature was identified by exploring the bibliography of the 64 relevant articles.

By reading the articles, nine different types of classification models that had been used for infant cry classification were identified. Table 9.7 summarizes the models and lists studies that used them.

Table 9.6.: Search result statistics for the different databases

Database	No. of articles in search result set
Web of Knowledge	94
DIMDI	45
Medpilot	138
DBLP	0
IEEE	35
SSG	17
ACM	250
Σ	579

Classification model review

The four identified software systems (IBM SPSS Statistics 20, IBM SPSS Modeler 15, R 3.0.2, RapidMiner 6) implemented most of the classification algorithms. However, three algorithms could not be evaluated in the review.

Classification algorithms based on *fuzzy logic* were neither implemented in one of the given systems, nor was it possible to find any ready-to-use software system providing fuzzy classification. Therefore, fuzzy classification could not be applied to the dataset. The same problem occurred for *weighted rough sets*; no ready-to-use software was found, either.

Hidden Markov models were available as plug-in for one of the given systems. However, for Markov models it became apparent that they were not suited for the given dataset: Markov models are focused on temporal data, whereas the reference dataset is based on stationary parameters. To keep classification models comparable, Markov models were excluded as they would require different data from the other classification models.

Apart from the classification models identified in the infant cry classification literature, the software systems provided four additional classification models which were included in the review: three additional classification tree approaches (C&R decision tree, CHAID decision tree and Quest decision tree) and logistic regression. Logistic regression models provide linear formulas for classifying items into categorical groups.

Figure 9.1 summarizes all classification models that were applied to the reference dataset in this article. Algorithms in filled boxes have already been used in the literature and algorithms with blank boxes have not yet been used in the literature and were included in the review.

Table 9.7.: List of classification model types and the studies in which they were used

Classification models and Studies
Artificial Neural Network Amaro-Camargo and Reyes-García (2007), Barajas-Montiel and Reyes-García (2005), Cano, Suaste, Escobedo, Reyes-García, and Ekkel (2006), Galaviz and García (2005), J. O. García and Reyes-García (2003a, 2003b), Hariharan, Fook, Sindhu, Ilias, and Yaacob (2012), Hariharan, Saraswathy, Sindhu, Khairunizam, and Yaacob (2012), Hariharan, Sindhu, and Yaacob (2012), Hariharan, Yaacob, and Awang (2011), Kheddache and Tadj (2012), Mohd Ali, Mansor, Lee, and Zabidi (2012), Möller and Schönweiler (1999), Orozco-Garcia and Reyes-García (2003), Orozco and Reyes-García (2003), Ortiz, Escobedo Beceiro, and Ekkel (2004), Petroni, Malowany, Johnston, and Stevens (1995a, 1995b), Poel and Ekkel (2006), Reyes-Galaviz, Arch-Tirado, and Reyes-García (2004), Reyes-Galaviz, Cano-Ortiz, and Reyes-García (2008), Reyes-Galaviz and Reyes-García (2004, 2005), Reyes-Galaviz, Verduzco-Mendoza, Arch-Tirado, and Reyes-García (2005), Reyes-García et al. (2010), Rosales-Pérez, Reyes-García, and Gómez-Gil (2011), Saraswathy, Hariharan, Vijean, Yaacob, and Khairunizam (2012), Schönweiler, Kaese, Möller, Rinscheid, and Ptok (1996a, 1996b, 1996c), Suaste-Rivas, Diaz-Mendez, Reyes-García, and Reyes-Galaviz (2006), Suaste-Rivas, Reyes-Galaviz, Diaz-Mendez, and Reyes-García (2004a, 2004b), Zabidi, Khuan, Mansor, Yassin, and Sahak (2010, 2011), Zabidi, Mansor, Khuan, Yassin, and Sahak (2010, 2011), Zabidi, Mansor, Lee, Yassin, and Sahak (2011)
Bayes Classifier Amaro-Camargo and Reyes-García (2007)
Hidden Markov Model Abdulaziz and Ahmad (2010), Aucouturier, Nonaka, Katahira, and Okanoya (2011), Honda, Kitahara, Matsunaga, Yamashita, and Shinohara (2012), Lederman et al. (2002), Lederman, Zmora, Hauschildt, Stellzig–Eisenhauer, and Wermke (2008), Singh, Mukhopadhyay, and Rao (2013)
Linear Discriminant Analysis (LDA) Fuller (1991)
Support Vector Machine (SVM) Amaro-Camargo and Reyes-García (2007), Brahnam, Chuang, Shih, and Slack (2006), Guanming, Xiaonan, and Haibo (2008), Sahak, Lee, Mansor, Yassin, and Zabidi (2010), Sahak, Lee, Mansor, Zabidi, and Yassin (2011), Sahak, Mansor, Khuan, Zabidi, and Yassin (2012), Sahak, Mansor, Lee, Yassin, and Zabidi (2010)
Fuzzy Logic Barajas-Montiel and Reyes (2005), Cano-Ortiz, Reyes-García, Reyes-Galaviz, Escobedo Beceiro, and Cano-Otero (2013), Kia, Kia, Davoudi, and Biniazan (2012), Reyes-Galaviz, Arch-Tirado, and Reyes-García (2004), Santiago-Sanchez, Reyes-García, and Gómez-Gil (2009)
Decision Tree Amaro-Camargo and Reyes-García (2007), Etz, Reetz, and Wegener (2012)
K-nearest Neighbor (KNN) Cohen and Lavner (2012)
Weighted Rough Set Framework Own and Abraham (2012)

Table 9.8.: Rating results for the classification models

Classification model	**No. of computed models**	**Accuracy training dataset**	**Accuracy test dataset**		**Degree of overfitting**		**Comformability**		**Rating**
		% correct	% correct	Rating Acc	Value	Rating OFit	Category	Rating Conf	
Decision trees									
C&R decision tree	144	98.10 %	73.30 %	1	24.80 %	1	Cut-off values and feature importance	4	2.00
Quest decision tree	48	89.13 %	69.32 %	1	19.81 %	2	Cut-off values and feature importance	4	2.22
Chaid decision tree	384	99.46 %	77.84 %	1	21.62 %	1	Cut-off values and feature importance	4	2.00
C5 decision tree	96	97.20 %	98.64 %	4	−1.44%	4	Cut-off values and feature importance	4	4.00
J48 decision tree	64	97.63 %	82.93 %	2	14.70 %	2	Cut-off values and feature importance	4	2.67
K-nearest neighbour (KNN)	128	95.92 %	91.48 %	3	4.44 %	4	Basic	1	2.56
Bayes classifier	128	91.03 %	80.68 %	2	10.35 %	2	Cut-off values	3	2.33
Linear discriminant analysis	160	90.49 %	79.55 %	1	10.94 %	2	Feature importance	2	1.56
Logistic regression	576	95.11 %	69.89 %	1	25.22 %	1	Feature importance	2	1.33
Support Vector Machine (SVM)	384	52.17 %	53.98 %	1	−1.81%	4	Feature importance	2	2.00
Artificial neural networks									
Neural network: multilayer perceptron	192	96.30 %	79.50 %	1	16.80 %	2	Feature importance	2	1.56
Neural network: radial basis function	48	78.70 %	73.90 %	1	4.80 %	4	Feature importance	2	2.00

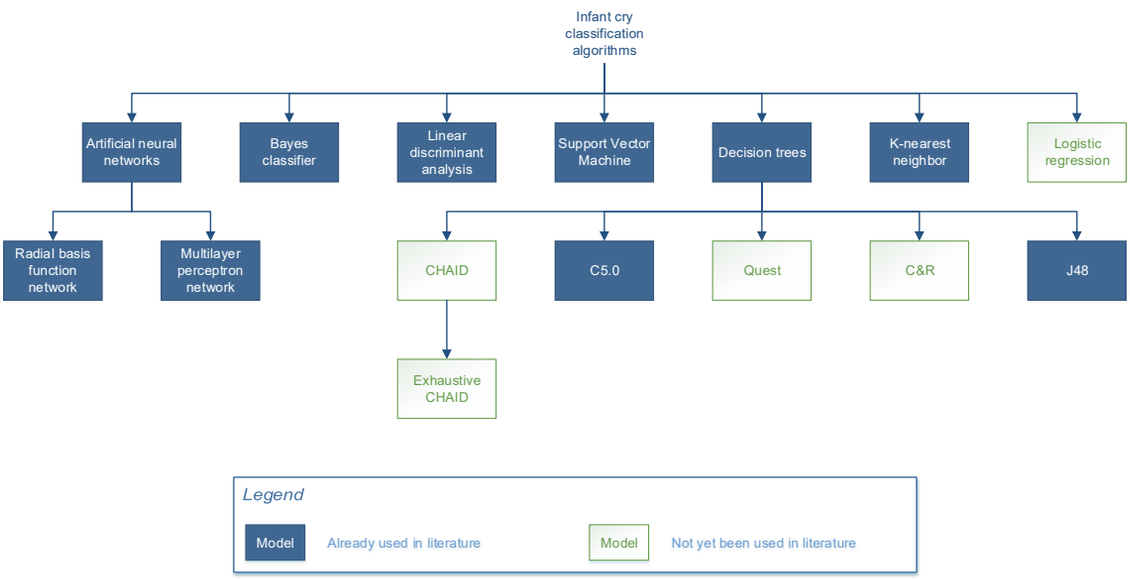


Figure 9.1.: Overview of classification model algorithms that have been included in the model review.

The classification models were applied to the reference dataset and rated according to the rating scheme.

Table 9.8 provides the results of the classification model ratings. The ratings are briefly explained in the following.

Decision Tree Ratings Five different algorithms for training decision trees were applied to the reference dataset. Accuracy on the test dataset varied between 69.32 % and 98.64 % resulting in accuracy ratings Acc from 1 to 4. The highest degree of overfitting (difference between the accuracy on training and test datasets) was 24.80 %, the lowest was -1.44 % (i.e., the accuracy on the test dataset was higher than on the training dataset). Ratings for overfitting $OFit$ were between 1 and 4. Conformability of all decision trees was rated with 4 ("Cut-off values and feature importance provided") because decision rules within the trees allow identifying characteristic properties of each cry group. The overall rating ($R = \frac{2 \cdot Acc + OFit + 1.5 \cdot Conf}{\sum \text{Importance factors}}$) was between 2.00 and 4.0 for the decision trees.

K-nearest Neighbor The best k-nearest neighbor model reached an accuracy value of 91.48 % on the test dataset resulting in a rating of 3 for Acc. A degree of overfitting at 4.44 % led to an $OFit$ rating of 4. Conformability of k-nearest neighbor models were rated with the basic degree ($Conf = 1$). No characteristic cut-off values could be extracted from the model representation.

In addition, feature importance was limited to the top three most important features. Further information about the importance of the remaining acoustic properties of the cry signal is missing. Overall, the k-nearest neighbor model was rated with 2.56.

Bayes Classifier The Bayes classifier had an accuracy of 80.68 % on the test dataset ($Acc = 2$) and an overfitting of 10.35 % ($OFit = 2$). Cut-off values were recognizable by exploring the probability statistics of the Bayes classifier results ($Conf = 3$). In summary, the Bayes classifier got a rating of 2.33.

Linear Discriminant Analysis Linear discriminant analysis achieved 79.55 % accuracy on the test dataset resulting in a rating Acc of 1. With a degree of overfitting of 10.94 %, the model reached a rating $OFit$ of 2. Cut-off values for each acoustic feature were not recognizable. However, information about the feature importance was provided, resulting in a rating $Conf$ of 2. The overall rating for linear discriminant analysis was 1.56.

Logistic Regression The best logistic regression model reached a rating Acc of 1 with 69.89 % accuracy on the test dataset; the overfitting of 25.22 % was rated with $OFit = 1$. Similar to linear discriminant analysis models, only the feature importance was recognizable, and therefore $Conf$ was rated with 2. Overall, logistic regression got a rating of 1.33.

Support Vector Machine The best support vector machine model only reached an accuracy of 53.98 % on the test dataset resulting in a rating Acc of 1. However, accuracy on the test dataset was even higher than on the training dataset, providing a high rating $OFit = 4$ for the degree of overfitting. As for discriminant analysis, no cut-off values were provided; this resulted in a rating $Conf$ of 2 for providing only feature importance information. Aggregating the single rating values resulted in an overall rating of 2.00 for the support vector machine.

Artificial Neural Networks Two types of artificial neural networks were trained, multilayer perceptron based network and radial basis function based network. Accuracy values for these two types of models were 79.50 % and 73.90 % respectively on the test dataset; both values resulted in a rating of 1 for Acc. A degree of overfitting at 16.80 % resulted in a rating of 2 for $OFit$ for the multilayer perceptron network; $OFit$ for the radial basis network was rated with 4 (4.80 %). Both types of artificial neural networks only provide information about the feature importance

($Conf = 2$). No information about what characteristics of the cry signal lead to a prediction was given. Overall, multilayer perceptron networks were rated with 1.56; radial basis function networks were rated with 2.00 in summary.

9.2.3. Interpretation

The interpretation of the review results is presented first and threats to validity are discussed thereafter.

Interpretation of the review results

The systematic literature search identified many studies that explored supervised-learning models for infant cry classification. Many different approaches have been used so far but no systematic comparison of models has been conducted.

On the training dataset, most classification models achieved high accuracy values confirming the promising results reported in the literature. However, when validating the models on an independent test dataset, accuracy often dropped significantly. On the test dataset, only two models had accuracy values beyond 90 % (C5.0 decision trees with 98.64 % and K-nearest neighbor with 91.48 %). The remaining models seem to have serious problems with correctly predicting an infant's state of health by its cry. The rating performance of the best-performing C5 decision tree is presented in tables 9.9 to 9.11. On the test dataset, only two cries were misclassified: two cries of infants with hearing impaired were interpreted as cries of healthy infants.

Additionally, the gap between high accuracy on the training dataset and much lower accuracy on an independent test dataset indicates another problem of classification models: model overfitting. Overfitting occurs when models are too specialized on the training dataset. It leads to a lack of abstraction and therefore prevents correct classification of data that is slightly different from the training dataset.

Analysis of model validity should not depend on statistical values only. Therefore, understanding how models predict data and verifying the correctness of this prediction by experts, is very important. Here, several models suffered from too complex algorithms and too little insight into the classification process. Here, decision tree approaches provided the best conformability as they allow to verify the decision rules of the trees visually. However, trees should not grow too large as they can get very complicated to understand, too.

Table 9.9.: Rating accuracy for the C5 decision tree on the training and test partitions

Partition	Training		Test	
Correct	312	97.20 %	145	98.64 %
Wrong	9	2.80 %	2	1.36 %
Total	321		147	

Table 9.10.: Correlation matrix showing the performance of the C5 decision tree on the trainings dataset

Training	HE	LA	HI	UCLP	AS	BD
HE	197	1	0	2	0	0
LA	1	42	0	0	0	0
HI	0	0	27	0	0	1
UCLP	2	0	0	19	0	0
AS	0	1	0	0	10	0
BD	1	0	0	0	0	17

Table 9.11.: Correlation matrix showing the performance of the C5 decision tree on the test dataset

Test	HE	LA	HI	UCLP	AS	BD
HE	86	0	0	0	0	0
LA	0	27	0	0	0	0
HI	2	0	11	0	0	0
UCLP	0	0	0	7	0	0
AS	0	0	0	0	8	0
BD	0	0	0	0	0	6

Overall, suitability of classification models for infant cry analysis varied between high suitability and low suitability. Interestingly, classification models that have not been explored in infant cry research very well (e.g., several decision tree approaches or k-nearest neighbor) performed better than modeling approaches that have often been used in infant cry classification (e.g., neural networks). Based on this observation, the major recommendation derived from the model review would be to give those "exotic" classification models a try and explore their suitability for infant cry classification in more detail in future research.

Threats to validity

The validity of systematic review processes may be influenced by various factors. In this systematic classification review, the following potential factors were identified and addressed to improve validity as much as possible.

First, the rating framework might be biased to favor certain models. For this reason, a short literature search was conducted prior to the review in order to identify rating factors used in other classification model reviews (classifying various data). Parameters were selected and adapted to fit the classification of infant cries. In addition, the weighting of the single rating factors is arbitrary. Here, the selection followed the author's view on what is important for infant cry classification. As the most important factor, the accuracy of the model on an independent test dataset was chosen. With a slightly lower priority — but still very important — conformability was considered. It is not sufficient to rely on statistical accuracy values but it is necessary to validate model ratings by experts; this requires high transparency of the rating process. As the least important aspect, the degree of overfitting was chosen as it is not quite as important as the first two factors.

The reference dataset was constructed to cover healthy cries as well as multiple pathologies in order to identify the models' abilities to discriminate among many groups. For some pathologies, it is difficult to recruit many subjects resulting in different group sizes in the dataset. As a threat to validity, this might introduce bias when training classification models. However, for infant cry classification used as an early indicator of an infant's state of health, it is more important to be able to discriminate between healthy and non-healthy cries. Here, the reference dataset contains about 60 % healthy cry samples and 40 % non-healthy ones. Classification models were rated on their overall accuracy and not on their ability to predict certain groups for this reason.

The quality (especially the accuracy) of models might vary depending on the parameter selection for the classification algorithms. For this reason, multiple combinations of reasonable parameter values were selected and models were trained with each parameter setting automatically. By this,

the parameters providing the model with the highest quality were selected for each classification approach to ensure that the best result was achieved.

9.3. Validity of infant cry classification by human listeners

As showed in the previous section, automatic classification of infant cries based on their medical condition is a powerful opportunity to rate the health status of an infant by means of their acoustic features. However, for many identification or classification tasks, humans have shown better results than computational algorithms in previous studies. As an example, humans perform much better in speech and language processing than computational models (Luxton, 2016; Norvig, 2012). For this reason, the ability of humans to classify infants according to their health status is explored in this section.

Studies investigating the link between specific cry characteristics and the *cry perception* of listeners showed that multiple acoustic parameters are perceived as negative or abnormal to listeners. An increased or more variable fundamental frequency, as well as more dysphonated or hyperphonated parts of a cry are often perceived as distressing, sick, arousing, aversive and urgent (LaGasse et al., 2005; Möller & Schönweiler, 1999; Schuetze et al., 2003). Cries of infants suffering from autism were also perceived as more distressed than cries of infants with normal development (Esposito, Nakazawa, Venuti, & Bornstein, 2012). Venuti et al. (2012) found out that cries of infants with autism activate brain regions associated with emotional processing more intensively than cries of healthy infants, and therefore, listeners perceived these cries as more distressed and aversive.

In addition to examining the perception of cries, studies explored the human ability to distinguish between different types of crying, too. Listeners were found to be able to differentiate between cries of infants with perinatal complications and cries of healthy infants (Zeskind & Lester, 1978). Hearing cries of infants with asphyxia, down syndrome, cri-du-chat syndrome or autism, listeners showed differences in their behavior and their reaction on hearing those cries (Esposito et al., 2012; Frodi & Senchak, 1990; Venuti et al., 2012). These studies indicate that acoustic differences between cries of healthy and non-healthy infants can be perceived by human listeners and therefore might allow listeners to “hear the health status” of infants.

Based on this assumption, Möller and Schönweiler (1999) compared the ability of nurses, parents and otolaryngologists to distinguish cries of healthy infants from cries of infants with hearing impairment. Here, nurses reached significantly better results indicating that experience with hearing infant cries might influence the ability to discriminate between infant cries auditorily. Morsbach

and Murphy (1979) also described that nurses reached better results in classifying healthy infants and infants with hearing impairment than naive listeners or parents because of their daily contact to various healthy and non-healthy infants. A listening experiment comparing the hearing capacity of mothers as well as the hearing capacity of naive listeners was able to show that naive listeners can reach better results in discriminating healthy neonate cries than mothers (Nolten, 1984). In these studies, the participants were not trained in discriminating infant cries before conducting the listening experiment.

Gladding (1979) tested if listeners can be trained to discriminate cries correctly. Subjects with training showed significantly better results in distinguishing various types of crying than subjects without listening training.

Summarizing, a significant amount of research has been conducted to explore the acoustic properties of infant cries and the potential to identify differences in those properties between healthy and non-healthy cries by computational models and algorithms as well as by human listeners. However, previous research did not examine sufficiently if human listeners are able to differentiate not only between healthy and non-healthy cries but also between different types of pathologies. In addition, a comparison of the classification skills of computational models in contrast to the skills of human listeners is still missing.

The present section describes a listening experiment not conducted only to test the human ability to distinguish between healthy an non-healthy cries, but also to discriminate auditorily between different types of pathologies. For the listening experiment, naive listeners (students and parents) and expert listeners (nurses/midwives and therapists) were trained to auditorily discriminate cries of healthy infants as well as infants with various pathologies (hearing impairment, cleft-lip-and palate (CLP), asphyxia, laryngomalacia, brain damage). After training, listeners rated cries of infants with different health states and their rating skills were compared to the classification skills of computation models.

To answer the main research question 4, if human listeners are able to discriminate auditorily between healthy infant cries and non-healthy infant cries and if they are able to differentiate between different pathologies, the question was extended by the following sub-questions (SRQ):

SRQ 1: Are there differences in the discrimination skills between the listener groups?

SRQ 2: Are there differences in the listeners' rating performance between the types of crying (e.g., healthy, hearing impaired, ...)?

SRQ 3: Do listeners rate infant cries that were used during training more accurately than unknown cries?

SRQ 4: Do sociodemographic parameters like age influence the rating skills of human listeners?

SRQ 5: Do human listeners perform more or less accurately in discriminating between infant cries than computational models?

The remainder of this section is structured as follows. Section 9.3.1 describes the method of the study. Here the infant cry samples, the setting of the listening experiment, the training of the supervised-learning models and the statistics used to answer the research questions are described. Section 9.3.2 describes the results of the listening experiment, the computational infant cry classification and the statistical evaluation. Section 9.3.3 interprets the results and discusses decisions and threats to validity.

9.3.1. Method

For exploring the human listener's ability to classify infant cries and for comparing their performance to the rating performance of computational models, both, humans and computational models were applied to the same process of training and prediction. Figure 9.2 visualizes the training phase and the rating phase for human listeners and computational models.

The ability of *human listeners* to hear the difference between healthy and pathological cries and between different pathologies was trained in a listening training using 18 training cries. After training, human listeners predicted the health state of infants on 18 unknown cries. The training and prediction for human listeners is described in more detail in the following.

To train computational models, various *supervised-learning algorithms* were trained on the same training cries as the human listeners. Similar to humans, supervised-learning algorithms learn patterns by analyzing training data for which the result is known (here, acoustic parameters of infant cries, for which the health state of the infant is known, are analyzed). After training, the algorithms create supervised-learning models, that represent the knowledge learned during the training phase. These models are then applied to the same 18 unknown cries that were rated by the human listeners and the models predict the health state of infants based on the acoustic parameters of the cries.

In this setting (Setting A), the supervised-learning algorithms use the same training set of cries as the human listeners. This provides the same setting for both, human listeners and computational algorithms, and allows comparing both. However, supervised-learning models originally were designed to be trained on large datasets to avoid fitting the models too exactly to the training data, loosing the ability to predict unknown data correctly (overfitting).

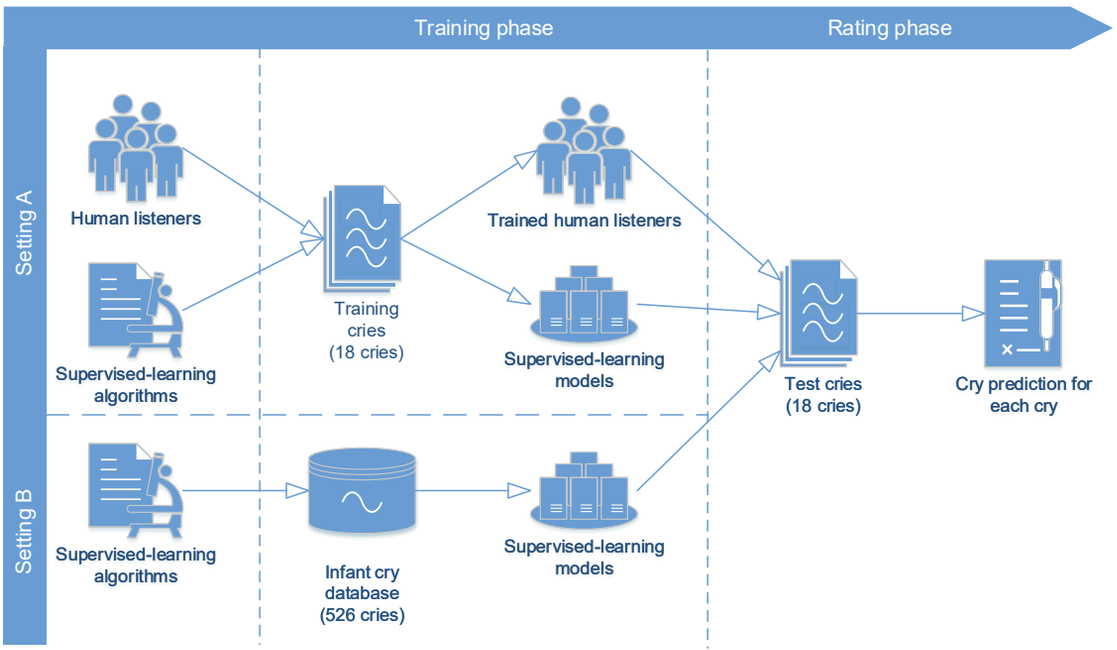


Figure 9.2.: Overview of the training phase and rating phase for the human listeners and for the computational models. From the infant cry database in Setting B, the 18 cries used in the rating phase were excluded.

For this reason, a second setting was included in the study. In Setting B, the supervised-learning algorithms were trained on a larger infant cry dataset to get more general models and to avoid overfitting to the training data. The resulting supervised-learning models were then applied to the same test cries as before.

Infant cry dataset

All infant cry samples used in this study were taken from the dataset of infant cries described in part II.

Listening experiment

The listening experiment was divided into a training phase and a rating phase. Participants were first trained in hearing cries of healthy infants and infants with various pathologies. In the rating phase, listeners had to allocate unknown cries to the different groups of health states.

Participants A total of 120 participants were included in the study and divided into the 4 groups *naive listeners* (group 1), *parents* (group 2), *nurses/midwives* (group 3) and *therapists* (group 4). All participants were female and German and were without hearing impairments. Because almost all participants in the groups nurses/midwives and therapists were female, the small number of male participants was excluded from the study to avoid any statistical error that might occur in an unbalanced study design. The impact of this decision on generalizing the results is discussed in section 9.3.3.

Table 9.12 provides sociodemographic parameters for each listener group. Group 1 contained 30 female naive listeners without children and without being in close contact to infants. Group 2 consisted of 30 mothers caring for infants being younger than 2 years old. Participants of the first two groups had jobs not related to the health system. Group 3 included 30 midwives and female pediatric nurses. Group 4 contained 30 female therapists. In this group, physical therapists (11 persons), occupational therapists (5 persons) and speech and language pathologists (14 persons) were included. For all midwives, nurses and all therapists, a professional experience of at least 4 years and a frequent contact to infants and young children with developmental diseases were defined as inclusion criteria.

A non-parametric Kruskal-Wallis test revealed significant differences in the distribution of age between naive listeners and the remaining groups, as well as the distribution of the number of children between therapists and parents. No significant difference was found in the professional experience across groups[3]. Balancing the groups for the parameters age and number of children was not possible with the given pool of participants. Because of the definition of the groups, the participants of the naive group had to be significantly younger than in the other groups as higher age did highly correlate with a higher number of children. Therefore, most participants with children were older, whereas most younger participants had no children. To cope with this variation of the sociodemographic parameters across groups, a correlation analysis has been used to analyze the possible effects of the parameters on the test results.

Training phase of human listeners In the training phase of the listening experiment, the participants had to listen to acoustic cry samples of healthy infants and infants with 5 different pathologies. According to Tsukamoto and Tohkura (1990), 2 to 5 cries build a perceptual unit for infant

[3]In the case of the number of children, the naive listeners were not included in the Kruskal-Wallis test and for the professional experience, the naive listeners and parents were excluded, as the differences to these groups result from the characteristics of the groups.

Table 9.12.: Sociodemographic parameters of the listener groups

Listener group: *(N = 120)*	**Naive listeners** *(N = 30)*		**Parents** *(N = 30)*		**Nurses/midwives** *(N = 30)*		**Therapists** *(N = 30)*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age [years]	23.3	2.6	31.4	5.4	32.4	6.4	35.5	7.1
No. children	-	-	1.6	0.7	1.2	1.2	0.8	1.0
Prof. experience [years]	-	-	-	-	9.5	5.8	10.3	5.8

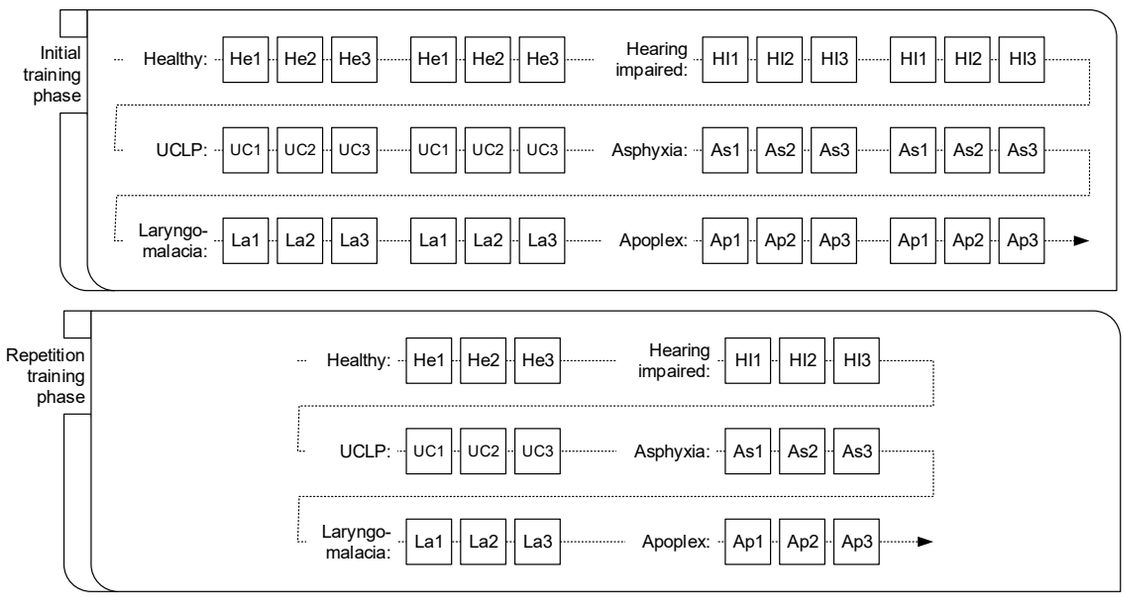


Figure 9.3.: Schema of the listening experiment

cry categorization. Therefore, three cries of each cry group were randomly selected from the infant cry dataset, summing up to a total of 18 cry samples for the training phase. All participants were trained on the same set of cry samples.

The training was held with all participants in a quiet room. The participants were told which cry type they would hear next and then the 3 cries of the cry group were played via speakers. The same 3 cries were then repeated and the participants took notes what they thought would be characteristic for the cry group. This procedure was repeated for each cry group. After this session, all cry groups were played again, but without repetitions. Figure 9.3 visualizes the schema of the training.

Altogether, the listeners heard the 3 cries from each of the 6 cry groups 3 times. This approach was used to ensure that the listeners were able to memorize the cry impressions. For the training

phase, listeners were asked to take personal notes about their hearing perception of each cry group in order to support the training effect.

The listening training was mandatory, as the present listening experiment examined not only the listeners' abilities to discriminate healthy and pathological cries, but also their abilities to discriminate various pathologies. Here, it could not be assumed that the listeners would know how cries of infants with various pathologies sound. In addition, Gladding (1979) showed that listeners with training reached significantly better results in distinguishing various types of crying than listeners without training.

Rating phase of human listeners In the rating phase, the participants had to listen to 18 cry samples and had to allocate each sample to one of the six cry groups. The questionnaire can be found in appendix D. From each cry group, 3 cry samples were presented, but the listeners were not told how many samples from each group were in the set.

One of the three cries was a cry sample that had already been used in the listening training. This approach was chosen to determine if cries known from the training phase can be allocated better to the six groups than unknown cries. The remaining two samples were randomly selected from the infant cry dataset. These cries had not been used in the training phase. All participants listened to the same set of cries.

Infant cry classification using supervised-learning models

Infant cry classification aims at finding a computational model that is able to automatically classify infant cries according to their acoustic properties into given categories of cries. Computational models work similar to human listeners rating infant cries: first, the acoustic properties of a cry must be extracted in an acoustic analysis. Second, a computational model must be trained on a training dataset for which the cry categories are known ("supervised-learning") in order to learn how to categorize the cries. Finally, the computational model can be applied to unknown cries to categorize them.

Following the comparison of supervised-learning models described in section 9.2, the following supervised-learning algorithms were selected for the study: Artificial neural networks, Bayes classifiers, Linear discriminant analyses, Support Vector Machines, Logistic regressions and Decision trees.

Acoustic analysis The following acoustic parameters were computed in this study: the median as well as lower and upper bounds — represented through the 10th and 90th percentile — of the fundamental frequency and intensity, the first six formants, jitter and shimmer values as well as the relation of phonated and non-phonated parts, number and degree of voice breaks and the cry duration were measured. Details for computing these parameters are described in chapter 6.

Training phase of supervised-learning models For training the supervised learning models, two different sets of training data were chosen: In *Setting A*, all supervised-learning models were trained on the same set of 18 infant cries that was used for training the human listeners. In *Setting B*, 526 cry samples from the cry database were used for training; the 18 infant cries used in the rating phase were excluded.

The supervised-learning algorithms described above were applied to the training datasets. Each algorithm of the supervised-learning model follows its own strategy for identifying rules to categorize the cries. All algorithms implement techniques to avoid overfitting of the models to the training data and thus, to be able to categorize unknown cries as correctly as possible.

For training the models, IBM's SPSS Modeler 18.0 was used. During the training phase, the software automatically varies different parameters of the algorithms to find the best settings of the algorithms.

After training, each algorithm creates a supervised-learning model that represents the classification rules for categorizing infant cries.

Rating phase of supervised-learning models After training, each model was applied to the test set of infant cries that was also presented to the human listeners. Here, the models of both training Settings A as well as B were applied to the same set of test cries.

Based on the rules learned during the training phase, all models categorized the cry samples to predict the health state of the infants.

Statistical analysis

To answer the research questions described in this section, various statistical methods were used. They are described in the following.

Covariate analysis First of all, a possible influence on the rating performance of the sociodemographic parameters age, number of children and professional experience was analyzed using a correlation analysis between these parameters and the rating correctness (SRQ 4). Non-parametric Spearman's rho correlation was computed as the sociodemographic parameters were not normally distributed. As no significant correlations between the parameters and the rating performance were found, the sociodemographic parameters were not included in any further statistical analyses.

Descriptive statistics To analyze the rating performance of the human listeners in the listening experiment (Research question 4), a confusion matrix was computed to compare the listeners' ratings and the actual cry types.

The following quality coefficients were computed on the confusion matrix to quantify the listeners' performances in discriminating between healthy and pathological cries as well as between the various pathologies:

- *Cohen's kappa coefficient* (κ) was computed to quantify the overall agreement of listener ratings with the actual cry types. In contrast to simple percentage agreement, κ takes into account any agreement occurring by chance.
- *Sensitivity* of the healthy group was computed to rate the listener's ability to identify healthy infant cries correctly.
- *Specificity* of the healthy group was computed to rate the listeners' ability to identify cries with one of the pathologies as not healthy (excluding the ability to differentiate between the various pathologies).

Analysis of variances for human listeners To analyze the influence of various factors on the classification performance of human listeners and to identify effects between these factors, a Generalized Linear Mixed Model (GLMM) was computed. GLMMs allow to analyze the influence of multiple fixed and random effects as well as effects of their interaction on one target variable that may have any scale or distribution.

In this analysis, the correctness of the cry ratings was chosen as a binomial-scaled (0=wrong rating, 1=correct rating) target variable. The GLMM was parameterized to use a binomial probability distribution and a Logit link function.

The following nominal variables were included as fixed factors:

The listener group was included to test if listeners of any group perform better in infant cry classification than the listeners of other groups (SRQ 1).

The cry type was included to test if any type of crying (e.g., cries of healthy infants or cries of infants with hearing impairment) was identified more precisely than the other types (SRQ 2).

The knowledge about cries was included to test if cries that were presented during the training phase are rated more precisely than unknown cries (SRQ 3).

To cope with possible differences in the rating performance between single listeners or between single cry samples, these two variables were added as random factors.

For exploring significant differences in more detail, pairwise comparisons with Bonferroni correction were conducted.

Analysis of variances between human listeners and computer models For comparing the rating performance of human listeners and supervised-learning models (SRQ 5), all ratings of human listeners and computer models were included in an additional analysis of variances. The groups "Human listeners", "Models, Setting A" and "Models, Setting B" were analyzed with the same statistics that were used for exploring variances between the human listener groups.

9.3.2. Results

Listening experiment

With all 120 participations, the listening experiment was conducted without dropouts.

Infant cry classification

All models were trained on the datasets for Setting A and Setting B. The trained models were then applied to the same set of 18 cry samples that was rated by the human listeners.

Statistical analysis

The statistical methods described in section 9.3.1 were computed using IBM's SPSS Statistics 23.0 (IBM, 2016). The results are described in the following.

Table 9.13.: Correlation analysis to analyze the influence of the sociodemographic covariates on the rating correctness.

		Age	No. children	Prof. experience	Correctness
Age	Correlation coefficient[a]	1.000	.633**	.599**	-0.024
	Sig. (2-tailed)		0.000	0.000	0.256
No. children	Correlation Coefficient[a]	.633**	1.000	.205**	0.010
	Sig. (2-tailed)	0.000		0.000	0.632
Prof. experience	Correlation Coefficient[a]	.599**	.205**	1.000	0.019
	Sig. (2-tailed)	0.000	0.000		0.386
Correctness	Correlation coefficient[a]	-0.024	0.010	0.019	1.000
	Sig. (2-tailed)	0.256	0.632	0.386	

[a] Spearman's rho
** Correlation is significant at the 0.01 level (2-tailed)

Covariate analysis Table 9.13 provides the results of the correlation analysis which explores the impact of the covariates *age*, *number of children* and *professional experience* on the rating performance of the human listeners. None of these covariates significantly correlate with the rating correctness.

Descriptive statistics for the human listeners To describe the rating performance of the listeners, the confusion matrix presented in table 9.14 was computed.

The overall ability of the listeners to correctly rate cries of healthy infants and infants with various pathologies was computed on the confusion matrix and is represented by the Kappa values shown in table 9.15.

The ability to differentiate between healthy and non-healthy cries was quantified by computing the sensitivity and specificity for rating healthy infant cries. Table 9.16 shows the sensitivity and specificity values for the listener groups. The sensitivity value represents the listeners' ability to identify healthy infants as healthy correctly. The specificity value represents their ability to identify infants with one of the pathologies as non-healthy.

Descriptive statistics for the classification models The classification performance of the models on the test dataset is shown in table 9.17. The models in Setting A were trained on the same 18 cries that were used for training the human listeners. The models in Setting B were trained on the complete dataset described in section 9.3.1.

Table 9.14.: Confusion matrix of the ratings of the participants in the listening experiment

		Listener rating						
		He	**CLP**	**HI**	**La**	**As**	**BD**	∑
Actual cry type	**He**	230 63.9 %	48 13.3 %	55 15.3 %	2 0.6 %	10 2.8 %	15 4.2 %	360
	CLP	116 32.2 %	108 30.0 %	74 20.6 %	10 2.8 %	23 6.4 %	29 8.1 %	360
	HI	49 13.6 %	73 20.3 %	172 47.8 %	12 3.3 %	9 2.5 %	45 12.5 %	360
	La	4 1.1 %	12 3.3 %	22 6.1 %	308 85.6 %	3 0.8 %	11 3.1 %	360
	As	4 1.1 %	29 8.1 %	14 3.9 %	6 1.7 %	259 71.9 %	48 13.3 %	360
	BD	31 8.6 %	58 16.1 %	30 8.3 %	13 3.6 %	61 16.9 %	167 46.4 %	360
	∑	434	328	367	351	365	315	2160

Table 9.15.: Kappa statistics for the listener groups and for all listeners

Listener Group	**Kappa Value**	**Asymptotic Standard Error[a]**	**Approximate T[b]**	**Approximate Significance**
Nurses	0.520	0.025	270.092	0.000
Naive listeners	0.498	0.025	250.884	0.000
Parents	0.471	0.026	240.514	0.000
Therapists	0.476	0.026	240.728	0.000
Total	0.491	0.013	510.095	0.000

[a] Not assuming the null hypothesis.
[b] Using the asymptotic standard error assuming the null hypothesis.

Table 9.16.: Sensitivity and specificity values of the human listeners for identifying healthy infants

Listener group	**Sensitivity**	**Specificity**
Nurses	0.70	0.88
Naive listeners	0.59	0.89
Parents	0.69	0.89
Therapists	0.58	0.88
Total	0.64	0.89

Table 9.17.: Confusion matrix presenting the classifications of the supervised-learning models for the training Settings A and B compared to the actual cry types

			Rating						
			He	**CLP**	**HI**	**La**	**As**	**BD**	∑
Models, Setting A	*Actual cry type*	**He**	23 85.2 %	0 0.0 %	2 7.4 %	2 7.4 %	0 0.0 %	0 0.0 %	27 100.0 %
		CLP	0 0.0 %	26 96.3 %	0 0.0 %	0 0.0 %	0 0.0 %	1 3.7 %	27 100.0 %
		HI	2 7.4 %	1 3.7 %	23 85.2 %	0 0.0 %	0 0.0 %	1 3.7 %	27 100.0 %
		La	1 3.7 %	0 0.0 %	0 0.0 %	26 96.3 %	0 0.0 %	0 0.0 %	27 100.0 %
		As	3 11.1 %	1 3.7 %	0 0.0 %	1 3.7 %	19 70.4 %	3 11.1 %	27 100.0 %
		BD	2 7.4 %	1 3.7 %	1 3.7 %	0 0.0 %	0 0.0 %	23 85.2 %	27 100.0 %
		∑	31	29	26	29	19	28	162
Models, Setting B	*Actual cry Type*	**He**	18 66.7 %	5 18.5 %	4 14.8 %	0 0.0 %	0 0.0 %	0 0.0 %	27 100.0 %
		CLP	4 14.8 %	21 77.8 %	2 7.4 %	0 0.0 %	0 0.0 %	0 0.0 %	27 100.0 %
		HI	3 11.1 %	4 14.8 %	20 74.1 %	0 0.0 %	0 0.0 %	0 0.0 %	27 100.0 %
		La	7 25.9 %	0 0.0 %	6 22.2 %	14 51.9 %	0 0.0 %	0 0.0 %	27 100.0 %
		As	1 3.7 %	1 3.7 %	1 3.7 %	1 3.7 %	22 81.5 %	1 3.7 %	27 100.0 %
		BD	0 0.0 %	1 3.7 %	0 0.0 %	0 0.0 %	0 0.0 %	26 96.3 %	27 100.0 %
		∑	33	32	33	15	22	27	162

Table 9.18.: Kappa statistics for the models of Settings A and B

Model Group	Kappa Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Models, Setting A	0.837	0.032	23.872	0.000
Models, Setting B	0.696	0.041	19.939	0.000
Total	0.527	0.012	58.796	0.000

[a] Not assuming the null hypothesis.
[b] Using the asymptotic standard error assuming the null hypothesis.

Table 9.19.: Sensitivity and specificity values of the classification models for identifying healthy infants

Rater Group	Sensitivity	Specificity
Models. Setting A	0.85	0.94
Models. Setting B	0.67	0.90
Total	0.76	0.92

Table 9.18 presents the Kappa values for the computer models of Setting A and Setting B, representing their overall ability to classify infant cries correctly.

Table 9.19 shows the sensitivity and specificity values representing the models' ability to identify healthy infants as healthy and infants with any of the pathologies as non-healthy.

Analysis of variances for human listeners Computing the Generalized Linear Mixed Model (GLMM) in SPSS resulted in a model with an accuracy value of 69.5 %, i.e., almost 70 % of the variance in the data can be explained by this model.

Table 9.20 shows the impact of the fixed factors on the rating correctness. The rating correctness does not vary significantly at the $p = 0.05$ level across the listener groups. However, the real cry type (e.g., healthy or CLP cries) and the knowledge about the cry (i.e., if the cry was known because it had already been used in the training phase, or if it was not known) had a significant impact on the rating performance.

To explore the significant fixed factors RealCryType and TestOrUnknownCry in more detail, pairwise comparisons were computed.

Table 9.21 shows the pairwise comparisons for the RealCryType factor[4].

[4]Symmetric contrasts were removed from the table

Table 9.20.: Fixed effects impact on the rating correctness

Source	F	df1	df2	Sig.
Corrected Model	30.056	9	2150	0.000
ListenerGroup	0.637	3	2150	0.591
RealCryType	53.099	5	2150	0.000
TestCries	7.078	1	2150	0.008

Probability distribution: Binomial
Link function: Logit

Table 9.21.: Pairwise contrasts of the real cry type groups

Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95 % Confidence Interval	
						Lower	Upper
He - CLP	0.344	0.035	9.774	2150	0.000	0.241	0.447
He - HI	0.162	0.037	4.442	2150	0.000	0.075	0.250
He - BD	0.176	0.037	4.829	2150	0.000	0.080	0.273
HI - CLP	0.182	0.036	5.010	2150	0.000	0.084	0.279
HI - BD	0.014	0.038	0.376	2150	0.707	-0.060	0.088
La - He	0.212	0.031	6.860	2150	0.000	0.126	0.298
La - CLP	0.556	0.031	18.231	2150	0.000	0.467	0.645
La - HI	0.375	0.032	11.621	2150	0.000	0.282	0.467
La - As	0.132	0.029	4.516	2150	0.000	0.059	0.206
La - BD	0.389	0.032	12.063	2150	0.000	0.297	0.480
As - He	0.080	0.034	2.337	2150	0.039	0.003	0.157
As - CLP	0.424	0.034	12.516	2150	0.000	0.326	0.522
As - HI	0.242	0.035	6.859	2150	0.000	0.144	0.340
As - BD	0.256	0.035	7.261	2150	0.000	0.157	0.355
BD - CLP	0.168	0.036	4.622	2150	0.000	0.074	0.261

The sequential Bonferroni adjusted significance level is 0.05.
Confidence interval bounds are approximate.

Table 9.22.: Simple contrast of the known and unknown cries

Simple Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95 % Confidence Interval	
						Lower	Upper
UKN cries - KN cries	-0.066	0.024	-2.692	2150	0.007	-0.113	-0.018

The sequential Bonferroni adjusted significance level is 0.05.
Confidence interval bounds are approximate.

Table 9.23.: Random effect covariances

Random Effect Covariance	Estimate	Std. Error	Z	Sig.	95 % Confidence Interval	
					Lower	Upper
ListenerGroup * ProbandID	.101	.064	1.566	.117	.029	.352
ListenerGroup * ProbandID * RealCryType * TestCries * CryNumber	.030	.076	.390	.696	.000	4.487

Covariance Structure: Variance components

Table 9.22 shows the contrast between the known cries and the unknown cries. Known cries were rated significantly better than unknown cries, but the effect size is not very large with -0.063, i.e., the rating performance was not that much better for known cries.

Random effect covariances were evaluated to estimate the influence of between-listeners variance and between-cry-samples variance.

Table 9.23 shows the random effect covariances. The between-listeners variance as well as the between-cry-samples variance are not significant. Therefore, both have no significant impact on the variance in the data.

Analysis of variances between human listeners and computer models The GLMM model for analyzing the impact of the group factors on the classification correctness reached an overall accuracy of 71.0 %, i.e., 71 percent of the variability in the data can be explained by the model.

The effects of the fixed factors RaterGroup, RealCryType and TestOrUnknownCry are presented in table 9.24. All three effects are significant at the $p = 0.05$ level.

Table 9.24.: Fixed effects impact on the rating correctness of computer models and human listeners

Source	F	df1	df2	Sig.
Corrected Model	34.340	8	2475	0.000
RaterGroup	26.660	2	2475	0.000
RealCryType	45.894	5	2475	0.000
TestOrUnknownCry	11.497	1	2475	0.001

Probability distribution: Binomial
Link function: Logit

Table 9.25.: Pairwise contrasts of the RaterGroup factor

Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95 % Confidence Interval	
						Lower	Upper
Humans - Models, Setting A	-0.290	0.028	-10.179	2475	0.000	-0.358	-0.222
Humans - Models, Setting B	-0.181	0.039	-4.613	2475	0.000	-0.269	-0.093
Models, Setting A - Models, Setting B	0.109	0.045	2.434	2475	0.015	0.021	0.196

The sequential Bonferroni adjusted significance level is 0.05.
Confidence interval bounds are approximate.

Table 9.25 shows the pairwise contrasts of the RaterGroup factor. All pairwise contrasts are significant at the $p = 0.05$ level. Human listeners are 29 % less precise in rating infant cries than models trained in Setting A, and they are 18 % less precise than models trained in Setting B. Comparing the models trained in the Settings A and B, models from Setting A are 11 % more precise than models from Setting B.

Table 9.26 shows the pairwise contrasts of the RealCryType factor across all rater groups[5].

Table 9.27 shows the simple contrasts between the rating of unknown cries (UKN cries) and known cries (KN cries). Known cries are rated slightly but significantly better than unknown cries.

Table 9.28 shows the random effect on the rating performance. The between-rater variance is significant at the $p = 0.05$ level.

[5]Symmetric contrasts were removed from the table

Table 9.26.: Pairwise contrasts of the RealCryType factor

Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95 % Confidence Interval	
						Lower	Upper
He - CLP	0.253	0.032	7.993	2475	0.000	0.163	0.343
He - HI	0.108	0.027	3.947	2475	0.000	0.038	0.179
He - BD	0.106	0.027	3.880	2475	0.000	0.038	0.174
HI - CLP	0.145	0.034	4.314	2475	0.000	0.056	0.234
HI - BD	-0.002	0.030	-0.071	2475	0.943	-0.061	0.057
La - He	0.116	0.021	5.510	2475	0.000	0.059	0.174
La - CLP	0.370	0.032	11.420	2475	0.000	0.275	0.465
La - HI	0.225	0.027	8.202	2475	0.000	0.145	0.304
La - As	0.069	0.018	3.832	2475	0.000	0.025	0.112
La - BD	0.223	0.027	8.155	2475	0.000	0.144	0.301
As - He	0.048	0.022	2.185	2475	0.058	-0.001	0.097
As - CLP	0.301	0.031	9.571	2475	0.000	0.209	0.393
As - HI	0.156	0.027	5.810	2475	0.000	0.081	0.231
As - BD	0.154	0.027	5.750	2475	0.000	0.080	0.228
BD - CLP	0.147	0.034	4.384	2475	0.000	0.057	0.237

The sequential Bonferroni adjusted significance level is 0.05.
Confidence interval bounds are approximate.

Table 9.27.: Simple contrast for the TestOrUnknownCry factor across all groups.

Simple Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95 % Confidence Interval	
						Lower	Upper
UKN cries - KN cries	-0.055	0.016	-3.433	2475	0.001	-0.087	-0.024

The sequential Bonferroni adjusted significance level is 0.05.
Confidence interval bounds are approximate.

Table 9.28.: Random effect covariances

Random Effect Covariance	Estimate	Std. Error	Z	Sig.	95 % Confidence Interval	
					Lower	Upper
RaterGroup * RaterID	0.098	0.045	2.197	0.028	0.040	0.240
RaterGroup * RaterID * RealCryType * TestCries	1.619E-19[a]					

Covariance Structure: Unknown
[a] This parameter is redundant

9.3.3. Interpretation

Correlation of the results with the research questions

In this section, the results that have been described in section 9.3.2 are interpreted and are put into correlation with the research questions of the study.

Are human listeners able to discriminate auditorily between healthy infant cries and non-healthy infant cries and are they able to differentiate between the different pathologies? The confusion matrix for the human listeners' ratings (table 9.14) provides an overview of the overall rating performance of human listeners. Laringomalacia cries are identified quite reliably (85.6 %). Asphixia cries and healthy cries also show a good rating accuracy with 71.9 % and 63.9 %. Although the remaining cry types are rated with lower accuracy, all cry types are identified more accurate than by chance (accuracy by chance is 16.67 %, assuming equal chance across all cry types). Hence, training human listeners to hear the health state of an infant seems to be possible. In addition, the performance of identifying healthy infants and distinguishing between various pathologies is better than by chance.

Cohen's Kappa values as an overall value (table 9.15) for the rating performance of human listeners are similar in all listener groups with an overall average value of 0.491. Following Landis and Koch (1977), this Kappa value can be interpreted as medium accuracy, backing the interpretation of the confusion matrix that human listeners have an average performance in identifying healthy infants and infants with various pathologies.

The sensitivity and specificity for identifying healthy infants as indicators for the listeners' performance to distinguish between healthy and non-healthy is similar between the listeners groups, too. The sensitivity value of 0.64 indicates a medium performance in identifying healthy infants as healthy. The specificity value of 0.89 shows that non-healthy infants are identified with high confidence. This observation backs the studies of Bisping (1986), who suspected that humans have the genetic ability to identify pathological states of health.

Summarizing, humans are well able to identify non-healthy infants by their cry. When distinguishing between various pathologies, the performance of humans is only average, but higher than by chance.

Are there differences in the discrimination skills between the listener groups? Analyzing the variances between the listener groups using GLMM showed no significant variances between listener groups. Therefore, the amount of contact of humans to infants with pathologies does not seem to influence the listeners' rating performance. These results contrast the study of Möller and Schönweiler (1999), who found a significant difference in the rating performance of parents and nurses when rating healthy infants and infants with hearing impairment. Although identifying significant differences, this study had only a small effect size in the variances.

Are there differences in the rating performance between the types of crying e.g., healthy, hearing impaired, ...) There are significant differences in the classification correctness across the different cry types. Evaluating the significant contrasts in table 9.21, the following statements about the cry types can be made.

1. Cleft lip and palate cries are rated less accurately than the other cry groups.
2. Healthy cries are rated more accurately than CLP, HI and BD cries.
3. Hearing impaired cries are rated more accurately than CLP cries.
4. Laryngomalacia cries are rated more accurately than He, CLP, HI, As and BD cries.
5. Asphyxiated cries are rated more accurately than He, CLP, HI and BD cries.
6. Brain damage cries are rated more accurately than CLP cries.

Cleft lip and palate disorders seem to have fewer auditory cry characteristics that are recognizable by humans than the other cry groups. Deformations in the orofacial tract do not seem to affect the cry signal very much, so an auditory identification is complicated.

Cries of infants suffering from laryngomalacia are rated most accurately. These cries are mostly high pitched, showing a lot of variation in the fundamental frequency and showing high intensity. These characteristics and the direct pathological impact of laryngomalacia on the vocal folds and the larynx, seem to result in auditory characteristics of the cries that are well recognizable by humans.

Do listeners rate infant cries that were used during training more accurately than unknown cries? Cry samples that were known to human listeners from the training phase were rated significantly better during the rating phase (table 9.22). Although, the effect size of 0.066 is not very high (i.e., known cries are rated by 6.7 % more accurately than unknown cries), so human

listeners seem to mainly learn the characteristics of the cry groups during training, instead of only recognizing certain cry samples they have already heard.

Do sociodemographic parameters like age influence the rating skills of human listeners? The correlation analysis (table 9.13) did not show any significant correlations between the rating correctness and the sociodemographic parameters age, number of children and professional experience. Therefore, these parameters do not seem to influence the rating skills of the listeners.

However, the age of the listeners strongly correlates with the number of children and the professional experience, which is somehow expected, as with higher age, it is likely to have one or more children and it is more likely to have a higher professional experience.

Do human listeners perform more or less accurately in discriminating between infant cries than computational models? The computational models trained in Setting A as well as those trained in Setting B perform significantly better than the human listeners at the $p = 0.05$ level (table 9.25).

The confusion matrix in table 9.17 for the computational models presents correctness values for the various cry types between 70 % and 100 % for models trained in Setting A, and between 51 % and 100 % for models trained in Setting B.

Kappa values of 0.696 for models trained in Setting B and 0.837 for models trained in Setting A stand for a substantial agreement between the classification and the actual health state of the infants.

The sensitivity and specificity values are above those of the human listeners, too. However, the specificity values of the classification models are only 0.03 points higher than those of the human listeners. Therefore, human listeners can identify pathological infant cries with a confidence similar to the models.

As for the human listeners, there are significant differences in the classification performance for the different cry types (table 9.26). The interpretation of these contrasts is similar to the interpretation for the human listeners. Hence, characteristic acoustic properties, that are relevant for the human listeners when classifying infant cries, seem to be relevant for the computational models, too.

Summarizing, computational models rate healthy infant cries and cries with various pathologies significantly better than human listeners. However, the rating performance in identifying pathological cries in general is very similar between humans and computational models.

Comparison of the approach to previous studies

Previous studies described that persons with frequent contact to healthy and ill infants perform better in identifying infant cries than persons without daily contact to infants (naive listeners) or persons having only close contact to one or two infants (parents) (Möller & Schönweiler, 1999; Morsbach & Murphy, 1979). In contrast, this study showed no differences between the listener groups. Here, the listening training seems to be an effective approach to train listeners in classifying infant cries. After a hearing training, experience in listening to infant cries has no impact on the rating accuracy.

In contrast to other studies (Möller & Schönweiler, 1999; Morsbach & Murphy, 1979), which examined how listeners perform in distinguishing between cries of healthy infants and infants with one pathology, this study examined if it is possible to distinguish between various pathologies. Here, a listening training is essential to ensure that listeners can learn to recognize acoustic properties specific to the various pathologies and thus, enable the listeners to distinguish pathologies auditorily.

The study could show that computational classification of infant cries reached better results and is more suitable for identifying pathologies by the cries than auditory discrimination by human listeners. Although listeners perform well in identifying cries as pathological, distinguishing between various pathologies seems to be very difficult and leads to bad classification results.

9.4. Summary of the findings regarding the validity of infant cries

To answer research questions 2 to 4, the validity of infant cries was analyzed for three different approaches to classify infant cries: (a) using analysis of variances for classification, (b) using supervised-learning models for classification and (c) using auditory discrimination by human listeners.

In this thesis, the approach to use analysis of variances for classifying infant cries was found to be inadequate for valid classification results. Although significant differences in acoustic parameters between healthy cries and cries of infants with hearing impairment and unilateral cleft lip and palate existed, single acoustic parameters were not specific enough to discriminate between all groups. For this reason, the validity of the ANOVA approach was found to be not sufficient for screening purposes.

Different supervised-learning approaches have successfully been applied in previous studies. To rate the validity of supervised-learning models, a systematic review and rating was conducted to identify valid models and to compare their performance for infant cry classification.

In the review, 9 classification models were identified that have been used for infant cry classification in the past. These classification models, as well as 3 new approaches were applied to a reference dataset containing cries of healthy infants and cries of infants suffering from laryngomalacia, cleft lip and palate, hearing impairment, asphyxia and brain damage. The classification models were evaluated according to a rating schema considering the aspects accuracy, degree of overfitting and conformability.

The results of this thesis indicate that many models have issues with accuracy and conformability. However, one model — the C5.0 decision tree — provides the most promising results in infant cry classification for diagnostic purpose. C5.0 decision trees reached the best ratings for all rating parameters and are therefore recommended for infant cry classification in this thesis.

To find out if the human brain performs better in infant cry classification compared to computational classification approaches, an auditory listening experiment was performed to examine if various listener groups (naive listeners, parents, nurses/midwives and therapists) are able to distinguish auditorily between healthy and pathological cries as well as to differentiate various pathologies from each other.

Listeners were trained in hearing cries of healthy infants and cries of infants suffering from cleft-lip-and-palate, hearing impairment, laryngomalacia, asphyxia and brain damage. After the training, the classification skills of the listeners were tested by allocating 18 unknown infant cries to the cry groups.

With a Kappa value of 0.491, listeners allocated the cries to the healthy and the five pathological groups with moderate performance. With a sensitivity of 0.64 and a specificity of 0.89, listeners were able to identify that a cry is a pathological one with higher confidence than separating between the single pathologies. GLMM found no significant differences between the classification accuracy of the listener groups. Significant differences between the pathological cry types were found.

Compared to supervised-learning algorithms that were trained and tested on the same dataset than the human listeners, the supervised-learning classification models performed significantly better than the human listeners in classifying infant cries. The models reached an overall Kappa value of up to 0.837.

Concluding, the best validity in classifying infant cries was achieved by using supervised-learning models, especially C5.0 decision trees. Analysis of variances as well as the ratings of human listeners could not provide results that are valid enough for screening purposes.

Chapter 10.

Objectivity of infant cry analysis

Objectivity in a screening process that is based on infant cry classification is closely related to the approach that is used for classification.

Automated classification approaches — e.g., using supervised-learning models — aim at enhancing objectivity by reducing the amount of decisions influenced by the researchers that conduct the screening. Despite the high degree of automation, various steps in the screening processes are still prone to subjective decisions and must therefore be analyzed on their objectivity.

When relying on human listeners for infant cry classification (as was introduced in section 9.3 for the listening experiment), the subjectivity of the human classifications is a concern. Here, techniques to enhance objectivity are required and must be tested on their applicability.

This chapter explores the objectivity of computer-based and human-based classification approaches and proposes techniques to enhance objectivity for both approaches.

10.1. Method

In the following, the objectivity of each step of the screening process used in this thesis for infant cry analysis is examined. The single steps are visualized in figure 10.1 and figure 10.2. The descriptions in this section of approaches to achieve objectivity are aligned with the steps shown in these figures.

Because ratings of human listeners are naturally prone to subjectivity, the objectivity of human-based infant cry classification (research question 5, how objective are the ratings of infant cries according to their health status for human listeners?) is explored in detail, too.

Objectivity in screening processes

In scientific research, the term "objectivity" defines the ideal goal of researchers to provide scientific findings without personal bias of researchers and without bias of the research recipients.

For screening instruments, objectivity must be ensured for all steps of the screening process. Therefore, objectivity must be explored in the following steps of a screening process in order to develop objective screening instruments based on infant cry classification. For developing screening instruments as well as for applying them, the following points should be considered:

Selecting subjects for the training dataset must be objective (implementation objectivity) in order to build an unbiased dataset that properly represents the population on which the screening instrument shall be applied.

Processing infant crying must be objective (implementation objectivity), so the method of receiving infant cries and making them accessible to the screening instrument must be independent of the person developing or applying the screening instrument.

Rating infant cries must be objective (application objectivity) and thus, objective methods for rating infant cries according to the health state of infants are required.

Interpreting the ratings must be independent of the persons that interpret the ratings to decide if infants are healthy or not (interpretation objectivity).

Section 10.2 describes how to obtain objectivity for the above-mentioned screening steps for computer-based screening approaches and for human-based ones.

10.2. Results

The strategy how to reach objectivity for the single steps of the screening process as well as the steps that are influenced by human decisions are presented in the following.

Objectivity for computer-based screening approaches

Computer-based screening approaches aim at improving objectivity by automatizing the screening process so subjective influences of the testers are reduced as much as possible. However, many steps in the process of developing and applying computer-based screening instruments (Figure 10.1) still require human interaction and are therefore prone to subjectivity.

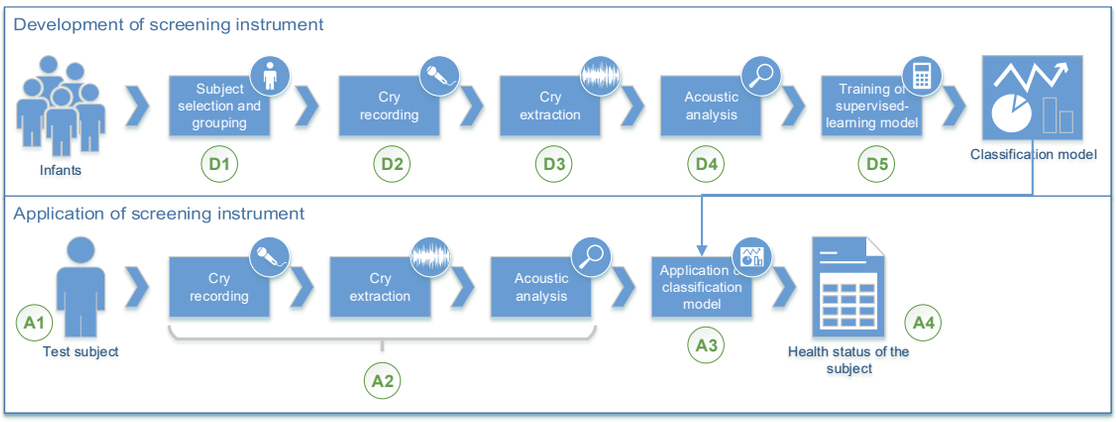


Figure 10.1.: Process for developing (training phase of supervised learning models) and applying (test phase of supervised learning models) computer-based screening instruments using supervised-learning classification models

Objectivity in developing computer-based screening instruments

When *developing* computer-based screening instruments for rating the health state of infants by their crying, the following steps of the development process include human test developers and are therefore possible threats to objectivity.

Subject selection and grouping (D1). Objectivity must be assured for allocating subjects to the different test groups. For infant cry classification, the groups are naturally defined by the health states of the infants (e.g., healthy, with hearing impairment, ...). A randomized allocation of subjects to groups as in randomized trials is therefore not applicable. Instead, the subjects included in the dataset must be selected by the screening instrument developers in a systematic way, so subjects are as homogeneous as possible across all groups regarding properties that are not related to their health state (e.g., age or gender). Only the properties that can clearly be related to the health states of the subjects may vary between the groups.

To achieve objectivity in the selection and grouping process, detailed inclusion and exclusion criteria must be defined. They are used to provide a precise description of which subjects shall be included in the training dataset and which properties are required for including subjects in each group. Applying these criteria in the selection process reduces subjectivity by standardizing the process for any screening instrument developer.

For this thesis, inclusion and exclusion criteria were defined for all subject groups as described in section 3.2. Subjects were selected following these criteria to build a balanced and homogeneous

dataset.

Cry recording (D2). Recordings of acoustic signals may be influenced by the conditions of the recording environment and by the recording process. For example for infant cry recordings, the distance between the microphone and the crying infant, the position of the infant itself and the general recording environment (e.g., background noise or echo) have a possible impact on the acoustic features of the recording.

To avoid bias introduced by varying researchers that record infant cries, a standardized approach for cry recordings is necessary. For this thesis, a detailed guide was provided describing how to achieve standardized recordings of infant cries (chapter 4). All recordings used in this thesis fitted these guidelines.

Cry extraction (D3). Extracting single cry utterances from the recordings can be done in two ways: (a) manually by humans that listen to and visualize the recordings to identify beginning and end of cry utterances, or, (b) automatically by applying computational algorithms that identify cry utterances based on statistics or rule-based systems.

The computational approach is the more objective one. Assuming that the computational algorithms are fully automatic and they were developed objectively, applying the algorithms to recordings will always result in the same cry extraction results, completely independent of the person applying the algorithm.

Extracting cries by humans requires standardization of the extraction process to achieve objectivity. Guidelines must be provided how to identify the beginning and ending of cry utterances and which quality criteria must be contained in the cry utterances.

For this thesis, cry extraction by humans was chosen and guidelines for the cry extraction process were provided as described in chapter 5. As explained in detail in chapter 5, current computational approaches for infant cry extraction have serious issues with validity. For this reason, computational approaches were not used although they may provide higher objectivity than human-based approaches.

Acoustic analysis (D4). Acoustic parameters for infant cries are computed by computational algorithms. In this step of the screening process, subjective influences may be introduced by the selection of algorithms and by setting parameters for them.

In this thesis, objectivity for the acoustic analysis is achieved by transparently describing and providing rationals for the algorithms and parameters that were chosen (chapter 6). Alternatives were discussed within the researcher group and choices were made in consensus to achieve the highest possible objectivity.

Training of supervised-learning models and interpretation of model quality (D5). Similar to the acoustic analysis, the classification of infant cries is based on computational algorithms that may be parameterized. For the class of algorithms used in this thesis – the supervised-learning algorithms – objective mathematical measures exist that rate the validity of the classification models on a test dataset. One can take advantage of this fact to objectify the selection and parametrization of an algorithm for valid infant cry classification: in this thesis, a set of many different supervised-learning algorithms was selected. For each algorithm, multiple reasonable values for all of its parameters were provided. Each algorithm was then trained with all possible combinations of the parameter values on the same training dataset. Following, all algorithms with all parameter combinations were evaluated on the same test dataset using objective mathematical measures to rate their accuracy. In a systematic review, subjective quality criteria of each algorithm were rated in a standardized review (section 9.2). Based on this transparent approach, the final supervised-learning algorithm was selected. Because of this automatic evaluation and systematic review, the choice of the algorithm and the selection of parameter values fulfill the requirements of an objective decision.

Objectivity in applying computer-based screening instruments

For *applying* a screening instrument based on computational infant cry classification, the steps shown in figure 10.1 are conducted and involve human interaction that may raise issues concerning the objectivity of the screening instrument as described in the following.

Test subject selection (A1). As first step in the screening process, the person responsible for conducting the screening has to decide if the screening instrument is applicable to an infant. To avoid subjective decisions, clear criteria must be provided to decide if the screening instrument is applicable to an infant or not.

Cry recording, cry extraction and acoustic analysis (A2). The approach of cry recording, cry extraction and acoustic analysis is equal to the approaches which were described for the development of computer-based screening instruments. The techniques to achieve objectivity are the same as described in section 10.2.

Application of the classification model (A3). Applying supervised-learning classification models to classify cries of infants with unknown health state does not involve any decision by humans. The classification process can be performed automatically by feeding the acoustic parameters of the cries into the classification model that was trained during the development phase. Taking measures against subjectivity is not required for this step of the screening.

Interpretation of the classification result (A4). Interpreting the classification result of supervised-learning models is straight-forward. All classification models return one single result, specifying the health state of the infant which is assumed by the classification model. An interpretation of the result is not necessary and thus, this step of applying the screening instrument is objective.

Objectivity for human-based screening approaches

When using the knowledge of human experts for classifying the health state of infants based on their crying, a development and application phase similar to the computer-based approach exist. Figure 10.2 shows the steps for developing and applying a screening approach by human listeners.

Objectivity in developing human-based screening approaches

Developing a screening instrument based on the rating of human listeners requires a training similar to the one for supervised-learning models as described in the following.

Subject selection, cry recording and cry extraction (D1). For training listeners in a systematic and comparable way, all listeners must be trained on the same dataset of infant cries. For this reason, infants are selected and their cries are recorded and extracted exactly like the training of computer-based approaches. A training dataset similar to the one for computer-based approaches is constructed and presented to all human listeners for training purposes. The guidelines to obtain objectivity in these steps is similar to the ones presented above.

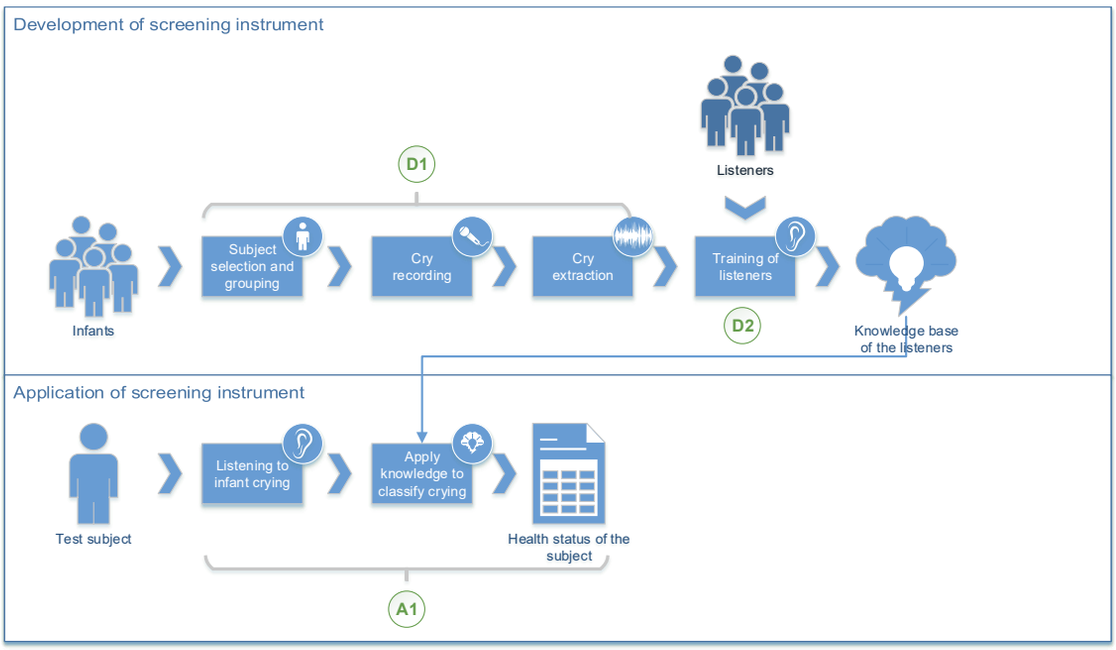


Figure 10.2.: Process of developing and applying a screening approach using human listeners for infant cry classification

Training of listeners (D2). The training of listeners is similar to the computer-based approach, too. Human listeners listen to cries of infants with various health states. For each cry, the health state is known and this information is given to the listeners. Each listener builds his own knowledge base by linking the audio impressions of the crying to the knowledge about the health states of the corresponding infants.

Two different aspects of this training are related to objectivity: the method of the training and the training-skills of the listeners.

The method of the training can be objectified by standardizing the training process. In this thesis, the cry samples of the training dataset were combined together with the information about the health states of the infants into one audio file as described in the method of the listening experiment in section 9.3.1. All listeners were then presented the same audio file in a standardized way for training purposes.

In contrast, the training-skills of the listeners cannot be objectified; every human has his own techniques for building knowledge, skills in learning and building knowledge also vary from human to human. This aspect of the training phase will always be subjective because of the trainee.

Table 10.1.: Krippendorff's Alpha for measruring the objectivty of human listeners in infant cry classification

Listener group	Krippendorff's Alpha
Naive listeners	0.3491
Parents	0.3031
Nurses	0.4061
Therapists	0.3309
Overall	0.3381

Objectivity in applying human-based ratings (A1)

Applying the knowledge of human listeners to classify infant cries is a purely subjective step. The way human listeners neurologically process the infant crying and how they access and interpret the knowledge obtained during training cannot be objectified in a standardized way. However, the effectiveness of the listening training and the homogeneity of the classifications made by the listeners can be measured.

During the listening experiment, each of the 120 participants rated the same set of 18 cries of healthy infants and infants with various pathologies. Measuring how homogeneously each cry was rated by the listeners provides insights about the objectivity of the classification.

To measure the similarity between the listener ratings, Krippendorff's Alpha coefficient was computed. For measuring the objectivity of the classifications, the 18 cries of the test dataset were defined as the observed units, the listeners were defined as the observers and the listener ratings for each cry were defined as the observations.

Table 10.1 shows the overall Krippendorff's Alpha value for all listeners and the values for the listeners within each listener group.

10.3. Interpretation

Objectivity is obligatory for developing and applying reliable and valid screening instruments.

Computer-based approaches for developing and applying screening instruments try to obtain objectivity by automating as much information of the screening process as possible. Still, many steps in the development phase of such a screening tool involve human decisions that may bias

the screening instrument. As described in section 10.2, objectivity in developing computer-based screening instruments can be achieved by standardizing all steps that require human interaction. This thesis successfully applied these techniques. Applying computer-based instruments in a screening process then is very objective leaving not a lot of room for subjective interpretations.

In contrast, objectivity is harder to achieve when relying on human classifications of infant cries for screening purposes. Efforts made to standardize the training process of human listeners did not lead to objective classifications as was shown by Krippendorff's Alpha coefficient that was measured to explore the homogeneity of infant cry classifications for human listeners. Alpha values between 0.3 and 0.4 in all listener groups certify a high heterogeneity in the classifications of human listeners. Hence, to answer research question 5, the ratings of the human listeners can not be seen as very objective.

10.4. Summary of the findings regarding the objectivity of infant cries

Summarizing, the objectivity in developing and applying computer-based screening processes based on infant cry classification can be achieved by a mixture of techniques to provide objective decisions: (a) a strict standardization of steps in the screening process that require activities of humans like recording the infant crying, (b) use of randomization where possible to prevent biased decisions, e.g., for selecting model-parameters for the classification models and (c) automation of as many steps as possible in the screening process to reduce the influence of subjective decisions and interpretations.

Objectivity for classification approaches relying on human listeners cannot be guaranteed. For this reason, the use of computer-based classification approaches is preferred in this thesis.

Part IV.

Finale

Chapter 11.

Discussion

This thesis explored the analysis and classification of infant cries from a screening perspective. The potential of infant cry classification for screening purposes was examined following the three main quality criteria of screening instruments: reliability, validity and objectivity. Different approaches that had previously been used in infant cry classification and approaches new for infant cry classification were investigated regarding their ability to achieve these three quality criteria.

The main findings of this thesis regarding the reliability, validity and objectivity of infant cry classification are discussed in the following. In addition, the method for obtaining the dataset of infant cries used in this thesis is discussed. Detailed rationals, e.g., for selecting certain acoustic parameters or methods, were given and discussed in the relevant chapters and will not be discussed again in this section.

11.1. Considering the research method for obtaining the infant cry dataset

In general, the premise for reliable research are data that are representative, balanced and complete regarding the properties of the subjects that are matter of the research questions. In this thesis, decision were made that contradict a balanced study design; they are discussed in the following.

11.1.1. The number of infants per cry group may vary

In a balanced study design, each cry group would contain the same number of infants. In infant cry research, this leads to problems because many pathological developments have a low incidence and

cries of such infants are therefore hard to record. To be able to include rare pathologies like brain damage in the thesis, the study design allowed different numbers of infants per cry groups. As supervised-learning algorithms are robust against varying group sizes (Russell & Norvig, 2010), this does not lead to any statistical problems during infant cry classification.

11.1.2. The number of cries per infant may vary

The number of cries that can be recorded from an infant varies; some infants cry more, others cry less. In a balanced study design, the number of cries per infant would be standardized after the recordings. Two approaches for standardization would be the following ones:

1. One average value over all cry utterances of an infant is computed for each acoustic parameter. By this, exactly one value per acoustic parameter per infant is computed for all infants. However, variations during an episode of crying that may be characteristic for pathological developments may get lost by applying this standardization technique and therefore, this technique is not applicable to infant cry classification.
2. A fixed number N_C of cry samples is randomly drawn from an infant's episodes of crying. If less than N_C cry samples exist for an infant, this infant must be excluded from the study. By using this technique, all infants provide the same number of cry samples to the dataset. However, drawing cry samples by chance might lead to excluding characteristic cry samples from the dataset. In addition, a difference in the number of cry samples might be a characteristic by itself (i.e., if infants cry more than other infants and therefore produce more cry samples in a given period of time.).

Because of the drawbacks of standardizing the number of cries per infant, it was decided to extract all cry samples from an infant's episodes of crying, independent of the number of cry samples that can be extracted. This approach ensures that all relevant and characteristic cries are included in the dataset and follows the state-of-the-art in infant cry research[1]. All cry samples are grouped according to the health states of the infants. The association of cries to single infants is mostly disregarded because the focus of the study was on classifying cries by health states instead of associating cries to individual infants.

[1] Cf. actual studies in infant cry classification like Lina Abou-Abbas, Tadj, Gargour, and Montazeri (2017), Chittora and Patil (2017), Farsaie Alaie, Abou-Abbas, and Tadj (2016).

11.2. Considering the reliability of infant cries

This thesis revealed the importance of statistically evaluating the reliability of infant cries. In most previous infant cry studies, the pain cry was generally accepted as the gold standard for cry analysis. For pain cries, the cause of crying is known (by a painful stimulus) and therefore, this cry type was assumed to be the most reliable cry type. However, this assumption was never statistically validated.

This thesis analyzed the reliability of different cry types statistically. The results show that no significant differences exist between pain cries and spontaneous cries. By trend, spontaneous cries are even more similar to each other than pain cries. One possible explanation why pain cries are not more reliable than spontaneous cries despite their known cause of crying, is the following: pain induced cries are produced with a higher level of energy and often contain more noisy parts than spontaneous cries. Algorithms for detecting acoustic features like the fundamental frequency or the formant frequencies seem to have trouble to deal with these very intensive and noisy cries. Hence, important information in the cries may be masked by these effects.

The lack of previous validity of the pain cry as the gold standard for cry analysis and the lack of significant differences to spontaneous cries found in this thesis raise doubts, if the pain cry should be further considered as gold standard for infant cry analysis in the future. Following the statistical trends found in the reliability analysis, this thesis proposes to use spontaneous infant cries instead of pain cries for developing screening instruments based on infant cry classification.

11.3. Considering the validity of infant cries

In this thesis, the validity is seen as the most important quality criterion, because the sensitivity and specificity of screening tools are important factors for the quality of the screening. This thesis showed that the validity of infant cry classification strongly depends on the classification techniques that are used.

Although analysis of variances was often used in previous studies for classifying infants according to their health state, the results of this thesis show that the validity of the ANOVA approach is not sufficient for screening purposes. Despite of the existence of significant differences in single acoustic parameters, the differences are not specific enough to separate between multiple pathologies. Therefore, the use of analysis of variances in screening instruments is not recommended in this thesis.

Auditory classification of infant cries by human listeners shows issues with the validity, too. While listeners were able to identify healthy infant cries with a specificity of 0.89, they had problems identifying pathological cries (sensitivity of 0.64). Additionally, listeners were not able to separate between different pathologies very well. The screening of infants by "hearing" is therefore regarded as an approach with low validity in this thesis.

This thesis recommends the use of supervised-learning models for infant cry classification as they provide the most valid classification results. However, the validity varies between the models. The systematic classification model review conducted for this thesis indicates that classification algorithms that had frequently been used in infant cry research (e.g., neural networks) have insufficiencies when trying to understand how the classification results are computed and if they are technically correct. Instead, the class of decision tree algorithms was identified as the most accurate and comprehensible approach for infant cry classification and is therefore recommended as most valid technique for screening purposes in this thesis.

11.4. Considering the objectivity of infant cries

Objectivity in infant cry classification was examined for computational classification approaches as well as for classification by human listeners through hearing.

For human listeners, the objectivity of their ratings was measured statistically based on their rating in the listening experiment conducted for this thesis. The results show that an objective rating of multiple listeners cannot be achieved, even when training the listeners in a standardized way. The human perception of infant cries seems to be too divergent as to be suited for screening purposes.

For computational infant cry classification based on supervised-learning models, objectivity is achieved by standardizing and automatizing the screening process in this thesis. All steps of the screening process where humans are involved have been standardized by providing detailed instructions, e.g., how to record the infant cries and how to extract single cry utterances from the recordings. The remaining steps of the screening process have been automatized by providing computational algorithms and models that process and classify the infant cries. The classification output unambiguously presents the health state of the infant and is therefore interpretable by humans in an objective way.

From an objectivity perspective, the use of computational classification techniques for infant cry classification is recommended in this thesis.

For all steps where humans are involved, the objectivity was obtained through standardization. Mostly in infant cry research, the cry recording process is described by two main pieces of information: the sampling rate and the recorder that was used. Although factors like the suspend of environment noise or the position of the infant or the distance between the mouth and the reorder are rarely described. Hence, this considerably influences the objectivity of the recording. Hence, including and excluding criteria of data are necessary to avoid bias and avoid to reach false effects. Because the group size is often small in infant cry analysis, side effects during the recording will mostly not be detected. Whenever humans are involved, the objectivity must be proven. For the auditory listening experiment, low interpretation objectivity of human listeners were calculated. In speech recognition, humans perform much better than computational algorithms. In the thesis, computational algorithms can be seen as more objective and also reached better results in their classification correctness.

Chapter 12.

Conclusion and future work

This section focuses on summarizing the main results as well as their placement in the context of the thesis.

As many previous studies indicated, the infant cry is not only a first step towards speech and language acquisition, but it is also a complex process involving many organs. Analyzing the acoustic features of infant cries promises to identify dysfunctions in these organs at a time when an infant is not yet able to tell us about potential problems or illnesses.

This thesis explored the potential of the acoustic characteristics of healthy infants' cries and infants with 5 different pathologies on their suitability for screening instruments based on the main quality criteria reliability, validity and objectivity.

From a reliability perspective, this thesis showed that the infant cry is very heterogeneous. Low values of the Krippendorff's Alpha coefficient indicated, that multiple cries of the same infant vary across a period of crying, making it difficult for screening instruments to differentiate between physiological and pathological variations of the infant cry.

Chapter 8 answered the first research question, namely which cry type is the most homogeneous and reliable one, and as a consequence, best suited for infant cry analysis. Pain cries were not found to be the most reliable cries — as was assumed by many previous studies because of their known cause of crying — but spontaneous cries, especially non-distressed cries reached higher Krippendorff's Alpha values than pain cries.

Backed by these results, spontaneous cries are proposed to be the more reliable ones compared to pain cries for use in screening instruments.

From a validity perspective, this thesis showed that analysis of variances is not suited for infant cry classification as acoustic parameters can overlap across groups and are therefore not specific to

certain pathologies, especially when including many different pathologies in the analysis process. Hence, the second research question, if differences in single acoustic parameters found by ANOVA are suited for infant cry classification, must be denied (section 9.1).

In addition to the ANOVA classification approach, various supervised learning models were examined on their validity. The third research question, which classification technique is suited best for discriminating infant cries was answered in section 9.2. Here, the different models were compared based on one referent dataset with six cry groups altogether and by the applying rating scheme which considered not only the rating accuracy, but also the degree of overfitting as well as the conformability. The C5 decision tree algorithm reached the most valid results in this rating.

Backed by these results, supervised-learning classification models, especially those with high accuracy and conformability like C5 decision trees, can be seen as a valid instrument for identifying the health state of infants based on acoustic parameters of their crying.

To answer the fourth research question, if the auditory discrimination of infant cries is also a valid method for identifying developmental pathologies, an auditory listening experiment was conducted. Section 9.3 showed that the auditory discrimination of humans is not sensitive to identifying various pathologies correctly, but human listeners were able to differentiate more adequately between healthy and non-healthy cries.

From an objectivity perspective, objectivity was achieved by standardization of all steps of the research process which can be influenced by human listeners to achieve application, implementation and interpretation objectivity (chapter 10). For human listeners, the interpretation objectivity was calculated with the Krippendorff's Alpha coefficient to answer the last research question of this thesis, namely how objective the ratings of infant cries according to their health status for human listeners are. Human listeners reached low Krippendorff's Alpha values. Hence, the interpretation objectivity for human listeners is not very high and therefore their ratings should be handled with care when used for screening purposes.

Summing up, the thesis examined the research field of infant cry analysis with regard to the infant cry's potential to fulfill the three quality criteria of screening instruments: objectivity, reliability and validity. Following the results of the thesis, it can be stated that infant cry classification can be conducted with objective, reliable and valid techniques. The infant cry is therefore recommended as to be suited for developing screening instruments based on infant cry analysis.

12.1. Future work

The two main lessons learned during this thesis are: (a) evaluating the quality of infant cry classification is very important when trying to use the infant cry for screening purposes as the current approaches vary widely regarding their reliability, validity and objectivity, and (b) screening approaches using a standardized process and supervised-learning models for classification are able to identify the health state of infants in a reliable, valid and objective way.

Hence, future research should analyze reliability, validity and objectivity more thoroughly when classifying infant cries. This is essential for bringing infant cry research one step further towards using the infant cry in screening instruments.

Experiences from this thesis indicate that future studies should concentrate on analyzing spontaneous infant cries instead of pain cries. Additionally, studies should select acoustic parameters for characterizing infant cries according to some physio-acoustic model instead of using simplified general models originating from acoustic science that are not related to acoustic properties of the infantile vocal tract. For the selection of suitable classification models, not only the accuracy but also the comprehensibility should be taken into account during the selection process.

Following the recommendations of this thesis can enable researchers to develop screening instruments based on infant cry classification. To develop an instrument that is ready to be used in clinical practice, the following challenges must be mastered in future research:

1. A larger and more complex dataset is required for training the classification models. As the supervised-learning models can only identify what they know from training, this dataset must include as many health states as possible, at least the ones that should be screened. A statistically representative sample of infants must be included in the study.
2. The extraction of cry utterances from recordings should be automatized. Approaches that currently exist should be improved to achieve higher validity when extracting cries. The automation of the cry extraction would transform one manual step in the screening process into an automatic one, improving the objectivity of the process.
3. The classification model that finally shall be used for screening must be validated on a large test dataset that contains a random sample of all possible infant conditions that may occur in the screening process. The validity of the model must be validated statistically as well as from a professional point of view bv experts.
4. Detailed instructions must be developed for clinicians that shall apply the screening instrument in order to provide objective results. Starting from the criteria that infants must fulfill

to be suited for the screening to the process of the screening itself, thorough training of the clinicians is required.

5. Finally, the complete screening instrument must be tested on its objectivity, reliability and validity in field experiments to prove the suitability of the instrument in clinical practice.

This thesis created the base for tackling these challenges in the future.

Bibliography

Abdulaziz, Y., & Ahmad, S. M. S. (2010). An accurate infant cry classification system based on continuos hidden markov model. In *Proceedings of the international symposium in information technology* (Vol. 3, pp. 1648–1652).

Abou-Abbas, L. [L.], Alaie, H. F., & Tadj, C. (2015). Automatic detection of the expiratory and inspiratory phases in newborn cry signals. *Biomedical Signal Processing and Control*, *19*, 35–43. doi:10.1016/j.bspc.2015.03.007

Abou-Abbas, L. [Lina], Tadj, C., Gargour, C., & Montazeri, L. (2017). Expiratory and inspiratory cries detection using different signals' decomposition techniques. *Journal of voice : official journal of the Voice Foundation*, *31*(2), 259.e13–259.e28. doi:10.1016/j.jvoice.2016.05.015

Accardo, P. J. (2013). 50 years ago in the journal of pediatrics. the cry latencies of normal infants and those with brain damage. *The Journal of pediatrics*, *162*(5), 902–917. doi:10.1016/j.jpeds.2012.11.044

ACM. (2014). Acm digital library. Retrieved January 22, 2014, from http://dl.acm.org/dl.cfm

Allemand, A., Stanca, M., Sposato, M., Santoro, F., Danti, F. R., Dosi, C., & Allemand, F. (2013). Neonatal asphyxia: Neurologic outcome. *Minerva pediatrica*, *65*(4), 399–410.

Amaro-Camargo, E., & Reyes-García, C. A. (2007). Applying statistical vectors of acoustic characteristics for the automatic classification of infant cry. *Advances in Pattern Recognition*, *4681*, 1078–1085.

American Speech-Language-Hearing-Association. (2015). Hearing screening. Retrieved January 9, 2015, from http://www.asha.org/public/hearing/Hearing-Screening/

Andersen, N. (1974). On the calculation of filter coefficients for maximum entropy spectral analysis. *Geophysics*, *39*(1), 69–72.

Apgar, V. (1953). A proposal for a new method of evaluation of the newborn infant. *Current researches in anesthesia & analgesia*, *32*(4), 260–267.

Arch-Tirado, E., Mandujano, M., Garcia-Torices, L., Martinez-Cruz, C. F., Reyes-García, C. A., & Taboada-Picazo, V. (2004). Cry analysis of hypoacoustic children and normal hearing children. *Cirugia y cirujanos*, *72*(4), 271–276.

Arruda, G. V., Dell' Aringa, A. R., Dell' Aringa, A. H., Esteves, M. C., & Nardi, J. C. (2009). Brainstem evoked response audiometry in normal hearing subjects. *Brazilian journal of otorhinolaryngology*, *75*(3), 420–425.

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, *34*(4), 555–596.

Aucouturier, J. J., Nonaka, Y., Katahira, K., & Okanoya, K. (2011). Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden markov models. *The Journal of the Acoustical Society of America*, *130*(5), 2969–2977. doi:10.1121/1.3641377

Ayari, S., Aubertin, G., Girschig, H., Van Den Abbeele, T., & Mondain, M. (2012). Pathophysiology and diagnostic approach to laryngomalacia in infants. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, *129*(5), 257–263. doi:10.1016/j.anorl.2012.03.005

Baeck, H. E., & de Souza, M. N. (2007). Longitudinal study of the fundamental frequency of hunger cries along the first 6 months of healthy babies. *Journal of voice*, *21*(5), 551–559.

Barajas-Montiel, S. E., & Reyes-García, C. A. (2005). Identifying pain and hunger in infant cry with classifiers ensembles. In *Proceedings of the international conference on computational intelligence for modelling, control and automation, and international conference on intelligent agents, web technologies and internet commerce* (Vol. 2, pp. 770–775).

Barajas-Montiel, S. E., & Reyes, C. A. (2005). Your fuzzy relational neural network parameters optimization with a genetic algorithm. In *Proceedings of the 14th ieee international conference on fuzzy systems* (pp. 684–689). IEEE.

Barr, R. G., Fairbrother, N., Pauwels, J., Green, J. A., Chen, M., & Brant, R. (2014). Maternal frustration, emotional and behavioural responses to prolonged infant crying. *Infant Behavior and Development*, *37*(4), 652–664. doi:10.1016/j.infbeh.2014.08.012

Barr, R. G., Hopkins, B., & Green, J. A. (Eds.). (2000). *Crying as a sign, a symptom, & a signal. clinical, emotional, and developmental aspects of infant and toddler crying* (1st ed.). Clinics in developmental medicine. London: Mac Keith Press and Cambridge University Press.

Bauer, H. (1968). Phoniatorischer Beitrag zum Cri-du-Chat Syndrom. *HNO: Wegweiser für die Fachärztliche Praxis*, *6*, 185–187.

Bellieni, C. V., Sisto, R., Cordelli, D. M., & Buonocore, G. (2004). Cry features reflect pain intensity in term newborns: An alarm threshold. *Pediatric research*, *55*(1), 142–146. doi:10.1203/01.PDR.0000099793.99608.CB

Berkowitz, S. (2013). *Cleft lip and palate. diagnosis and management* (3rd ed. 2013). Berlin, Heidelberg: Springer. Retrieved from http://dx.doi.org/10.1007/978-3-642-30770-6

Bielecki, I., Horbulewicz, A., & Wolan, T. (2012). Prevalence and risk factors for auditory neuropathy spectrum disorder in a screened newborn population at risk for hearing loss. *In-*

ternational Journal of Pediatric Otorhinolaryngology, *76*(11), 1668–1670. doi:10.1016/j.ijporl.2012.08.001

Bisping, R. (1986). *Der Schrei des Neugeborenen: Struktur und Wirkung*. Lehr- und Forschungstexte Psychologie. Berlin, Heidelberg: Springer-Verlag.

Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., ... Tooher, R. (2001). Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. *Journal of Speech Language and Hearing Research*, *44*(2), 264–285.

Blinick, G., Tavolga, W. N., & Antopol, W. (1971). Variations in birth cries of newborn infants from narcotic-addicted and normal mothers. *American journal of obstetrics and gynecology*, *110*(7), 948–958.

Boero, D. L., Weber, G., Vigone, M. C., & Lenti, C. (2000). Crying abnormalities in congenital hypothyroidism. preliminary spectrographic study. *Journal of child neurology*, *15*(9), 603–608.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the institute of phonetic sciences, amsterdam* (pp. 97–110).

Boersma, P. (2009). Should jitter be measured by peak picking or by waveform matching? *Folia Phoniatrica et Logopaedica*, *61*(5), 305–308.

Boersma, P., & Weenink, D. (2013a). Praat: Doing phonetics by computer. manual. Retrieved April 11, 2013, from http://www.fon.hum.uva.nl/praat/manual/Intro.html

Boersma, P., & Weenink, D. (2013b). Praat: Doing phonetics by computer. version 5.3.39. Retrieved April 11, 2013, from http://www.praat.org/

Borden, G. J., & Harris, K. S. (1984). *Speech science primer. physiology, acoustics, and perception of speech* (2nd ed.). Baltimore: Williams & Wilkins.

Bornstein, M. H., & Esposito, G. (2014). Beyond cry and laugh: Toward a multilevel model of language production. *Behavioral and brain sciences*, *37*(06), 548–549. doi:10.1017/S0140525X13003968

Bortz, J. (2010). *Statistik für Human- und Sozialwissenschaftler* (7th ed.). Berlin: Springer.

Boudewyns, A., Claes, J., & van de Heyning, P. (2010). Clinical practice: An approach to stridor in infants and children. *European journal of pediatrics*, *169*(2), 135–141. doi:10.1007/s00431-009-1044-7

Brahnam, S., Chuang, C. F., Shih, F. Y., & Slack. (2006). Machine recognition and representation of neonatal facial displays of acute pain. *Artificial Intelligence in Medicine*, *36*(3), 211–222.

Branco, A., Fekete, S. M. W., Rugolo, L. M. S. S., & Rehder, M. I. (2007). The newborn pain cry. descriptive acoustic spectrographic analysis. *International Journal of Pediatric Otorhinolaryngology*, *71*(4), 539–546.

Cacace, A. T., Robb, M. P., Saxman, J. H., Risemberg, H., & Koltai, P. (1995). Acoustic features of normal-hearing pre-term infant cry. *International Journal of Pediatric Otorhinolaryngology*, *33*(3), 213–224.

Cano-Ortiz, S. D., Reyes-García, C. A., Reyes-Galaviz, O. F., Escobedo Beceiro, D. L., & Cano-Otero, J. D. (2013). Emergence of a new alternative on cry analysis. the fuzzy approach. In J. Folgueras Méndez, T. Y. Aznielle Rodríguez, C. F. Calderón Marín, S. B. Llanusa Ruiz, J. Castro Medina, H. Vega Vázquez, ... R. Rodríguez Rojas (Eds.), *Proceedings of the 5th latin american congress on biomedical engineering. sustainable technologies for the health of all* (pp. 846–849). IFMBE Proceedings. Berlin: Springer.

Cano, S., Suaste, I., Escobedo, D., Reyes-García, C. A., & Ekkel, T. (2006). A combined classifier of cry units with new acoustic attributes. *Advances in Pattern Recognition*, *4225*, 416–425.

Cecchini, M., Lai, C., & Langher, V. (2010). Dysphonic newborn cries allow prediction of their perceived meaning. *Infant Behavior and Development*, *33*(3), 314–320.

Childers, D. G. (Ed.). (1978). *Modern spectrum analysis*. New York: IEEE Press.

Chittora, A., & Patil, H. A. (2017). Data collection of infant cries for research and analysis. *Journal of voice : official journal of the Voice Foundation*, *31*(2), 252.e15–252.e26. doi:10.1016/j.jvoice.2016.07.007

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *ASHA*, (23), 493–500.

Cleophas, T. J., & Zwinderman, A. H. (2013). *Machine learning in medicine*. doi:10.1007/978-94-007-6886-4

Cohen, R., & Lavner, Y. (2012). Infant cry analysis and detection. In *Proceedings of the ieee 27th convention of electrical electronics engineers in israel* (pp. 1–5). IEEE.

Colton, R. H., Steinschneider, A., Black, L., & Gleason, J. (1985). The newborn infant cry: Its potential implications for development and sids. In B. M. Lester & C. F. Z. Boukydis (Eds.), *Infant crying. theoretical and research perspectives* (pp. 119–138). New York: Plenum Press.

Committee on Fetus and Newborn American Academy of Pediatrics. (1996). Use and abuse of the apgar score. *Pediatrics*, *98*(1), 141–142.

Corwin, M. J., Lester, B. M., Sepkoski, C., McLaughlin, S., Kayne, H., & Golub, H. L. (1992). Effects of in utero cocaine exposure on newborn acoustical cry characteristics. *Pediatrics*, *89*(6 Pt 2), 1199–1203.

Corwin, M. J., Lester, B. M., Sepkoski, C., Peucker, M., Kayne, H., & Golub, H. L. (1995). Newborn acoustic cry characteristics of infants subsequently dying of sudden infant death syndrome. *Pediatrics*, *96*(1), 73–77.

Crowe, H. P., & Zeskind, P. S. (1992). Psychophysiological and perceptual responses to infant cries varying in pitch. comparison of adults with low and high scores on the child abuse potential inventory. *Child abuse & neglect*, *16*(1), 19–29.

DBLP. (2014). Digital bibliography & library project. Retrieved January 22, 2014, from http://www.dblp.org/

de la Peña, J. A. (2007). A web-based application and acoustic signal analysis of the infant cry. Budapest. Retrieved from http://e-archivo.uc3m.es/bitstream/handle/10016/11115/PFC_Jorge_Arranz_de_la_Pena.pdf?sequence=2

de Pisapia, N., Bornstein, M. H., Rigo, P., Esposito, G., de Falco, S., & Venuti, P. (2013). Sex differences in directional brain responses to infant hunger cries. *NeuroReport*, *24*(3), 142–146. doi:10.1097/WNR.0b013e32835df4fa

Denner, M. B. (2007). *Untersuchung spektraler und melodischer Eigenschaften vorsprachlicher Laute von Säuglingen mit einer familiären Disposition für eine spezifische Sprachentwicklungsstörung* (Dissertation, Bayrische Julius-Maximilians-Universität, Würzburg).

Deutsche Zentralbibliothek für Medizin. (2014). Medpilot. Retrieved January 22, 2014, from http://www.medpilot.de/

DIMDI. (2014). DIMDI Medizinwissen. Retrieved January 22, 2014, from http://www.dimdi.de/static/de/index.html

Dobbie, A. M., & White, D. R. (2013). Laryngomalacia. *Pediatric Clinics of North America*, *60*(4), 893–902. doi:10.1016/j.pcl.2013.04.013

Eckel, H. E., Sprinzl, G. M., Sittel, C., Koebke, J., Damm, M., & Stennert, E. (2000). Anatomy of the glottis and subglottis in the pediatric larynx. *HNO*, *48*(7), 501–507.

Editore, S. (2014). *Idiopathic sudden sensorineural hearing loss*. EBM Ebooks. SICS. Retrieved from https://books.google.de/books?id=2C%5C_tBQAAQBAJ

Eilers, R. E., & Oller, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *The Journal of pediatrics*, *124*(2), 199–203.

Esposito, G., Nakazawa, J., Venuti, P., & Bornstein, M. H. (2012). Perceptions of distress in young children with autism compared to typically developing children: A cultural comparison between japan and italy. *Research in developmental disabilities*, *33*(4), 1059–1067. doi:10.1016/j.ridd.2012.01.014

Esposito, G., Nakazawa, J., Venuti, P., & Bornstein, M. H. (2013). Componential deconstruction of infant distress vocalizations via tree-based models. a study of cry in autism spectrum disorder and typical development. *Research in developmental disabilities*, *34*(9), 2717–2724. doi:10.1016/j.ridd.2013.05.036

Etz, T., & Pörschmann, K. (2011). *Der Säuglingsschrei - Ein potentieller Frühindikator für sich entwickelnde Pathologien?* (Masterarbeit, Hochschule Fresenius, Idstein).

Etz, T., Reetz, H., & Wegener, C. (2012). A classification model for infant cries with hearing impairment and unilateral cleft lip and palate. *Folia Phoniatrica et Logopaedica*, *64*(5), 254–261. doi:10.1159/000343994

Etz, T., Reetz, H., Wegener, C., & Bahlmann, F. (2014). Infant cry reliability: Acoustic homogeneity of spontaneous cries and pain-induced cries. *Speech Communication*, *58*, 91–100. doi:10.1016/j.specom.2013.11.006

Fairbrother, N., Barr, R. G., Pauwels, J., Brant, R., & Green, J. A. (2014). Maternal thoughts of harm in response to infant crying: An experimental analysis. *Archives of women's mental health*. doi:10.1007/s00737-014-0471-2

Fant, G. (1971). *Acoustic theory of speech production. with calculations based on x-ray studies of russian articulations* (Reprint 2012). Description and Analysis of Contemporary Standard Russian. Berlin/Boston: de Gruyter and De Gruyter Mouton.

Farhadi, M., Mahmoudian, S., Mohammad, K., & Daneshi, A. (2006). The pilot study of a nationwide neonatal hearing screening in iran: Akbarabadi and mirzakouchak khan hospitals in tehran. *Hakim Res J*, (3), 65–75.

Farsaie Alaie, H., Abou-Abbas, L., & Tadj, C. (2016). Cry-based infant pathology classification using gmms. *Speech communication*, *77*, 28–52. doi:10.1016/j.specom.2015.12.001

Ferriero, D. M. (2004). Neonatal brain injury. *The New England journal of medicine*, *351*(19), 1985–1995. doi:10.1056/NEJMra041996

Fischer, A. (2009). *Prosodik organization in the babbling of german-learning infants betrween the age of six and twelve months* (Doctoral dissertation, Humboldt-Universität, Berlin).

Fisichelli, V. R., Coxe, M., Rosenfeld, L., Haber, A., Davis, J., & Karelitz, S. (1966). The phonetic content of the cries of normal infants and those with brain damage. *The Journal of psychology*, *64*(1), 119–126.

Fox, N. A., Kimmerly, N. L., & Schafer, W. D. (1991). Attachment to mother/attachment to father. a meta-analysis. *Child development*, *62*(1), 210–225.

Frodi, A., & Senchak, M. (1990). Verbal and behavioral responsiveness to the cries of atypical infants. *Child development*, *61*(1), 76–84.

Fuhr, T., Reetz, H., & Wegener, C. (2015). Comparison of supervised-learning models for infant cry classification. *International Journal of Health Professions*, *2*(1), 4–15. doi:10.1515/ijhp-2015-0005

Fuller, B. F. (1991). Acoustic discrimination of three types of infant cries. *Nursing research*, *40*(3), 156–160.

Furlow, F. B. (1997). Human neonatal cry quality as an honest signal of fitness. *Evolution and Human Behavior*, *18*(3), 175–193.

Galaviz, O. F. R., & García, C. A. R. (2005). Infant cry classification to identify hypo acoustics and asphyxia comparing an evolutionary-neural system with a neural network system. In A. Gelbukh, Á. de Albornoz, & H. Terashima-Marín (Eds.), *Micai 2005: Advances in artificial intelligence* (Vol. 3789, pp. 949–958). Lecture Notes in Computer Science. doi:10.1007/11579427_97

García, J. O., & Reyes-García, C. A. (2003a). Medical applications - acoustic features analysis for recognition of normal and hipoacusic infant cry based on neural networks. *Lecture notes in computer science*, *2687*, 615–622.

García, J. O., & Reyes-García, C. A. (2003b). Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neural networks. In *Proceedings of the international joint conference on neural networks* (Vol. 4, 3140–3145 vol.4). Piscataway, NJ: IEEE.

GBV. (2014). Online Contents - SSG Mathematik und Informatik. Retrieved January 22, 2014, from http://gso.gbv.de/DB=2.77/LNG=DU/

Gilbert, H. R., & Robb, M. P. (1996). Vocal fundamental frequency characteristics of infant hunger cries. birth to 12 months. *International Journal of Pediatric Otorhinolaryngology*, *34*(3), 237–243. doi:10.1016/0165-5876(95)01273-7

Gladding, S. T. (1979). Effects of training versus non-training in identification of infant cry-signals: A longitudinal study. *Perceptual and motor skills*, *48*(3 Pt 1), 752–754.

Goberman, A. M., & Robb, M. P. (2005). Acoustic characteristics of crying in infantile laryngomalacia. *Logopedics, phoniatrics, vocology*, *30*(2), 79–84. doi:10.1080/14015430510006703

Golden, C. J., Espe-Pfeifer, P., & Wachsler-Felder, J. (2002). *Neuropsychological interpretation of objective psychological tests*. Critical Issues in Neuropsychology. Boston, MA: Kluwer Academic Publishers. Retrieved from http://dx.doi.org/10.1007/b107998

Golub, H. L., & Corwin, M. J. (1982). Infant cry. a clue to diagnosis. *Pediatrics*, *69*(2), 197–201.

Golub, H. L., & Corwin, M. J. (2000). A physioacoustic model of the infant cry. In R. G. Barr, B. Hopkins, & J. A. Green (Eds.), *Crying as a sign, a symptom, & a signal. clinical, emotional, and developmental aspects of infant and toddler crying* (pp. 59–82). Clinics in developmental medicine. doi:10.1007/978-1-4613-2381-5_3

Grau, S. M., Robb, M. P., & Cacace, A. T. (1995). Acoustic correlates of inspiratory phonation during infant cry. *Journal of speech and hearing research*, *38*(2), 373–381.

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*(2), 271–279.

Grow, J., & Barks, J. D. E. (2002). Pathogenesis of hypoxic-ischemic cerebral injury in the term infant: Current concepts. *Clinics in perinatology*, *29*(4), 585–602, v.

Guanming, L., Xiaonan, L., & Haibo, L. (2008). Research on recognition for facial expression of pain in neonates. *Acta Optica Sinica*, *28*(11), 2109–2114.

Gustafson, G. E., Wood, R. M., & Green, J. A. (2000). Can we hear the causes of infants' crying? In R. G. Barr, B. Hopkins, & J. A. Green (Eds.), *Crying as a sign, a symptom, & a signal. clinical, emotional, and developmental aspects of infant and toddler crying* (pp. 772–780). Clinics in developmental medicine. London: Mac Keith Press and Cambridge University Press.

Haghshenas, M., Fard, H. A., Delavari, K., Gorji, H., Zadeh, P. Y., Javadian, Y., & Panjaki, H. S. (2014). Auditory screening in infants for early detection of permanent hearing loss in northern iran. *Annals of Medical and Health Sciences Research*, *4*(3), 340. doi:10.4103/2141-9248.133456

Hariharan, M., Chee, L. S., & Yaacob, S. (2010). Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network. *Journal of medical systems*, *36*(3), 1309–1315. doi:10.1007/s10916-010-9591-z

Hariharan, M., Fook, C. Y., Sindhu, R., Ilias, B., & Yaacob, S. (2012). A comparative study of wavelet families for classification of wrist motions. *Computers & Electrical Engineering*, *38*(6), 1798–1807.

Hariharan, M., Saraswathy, J., Sindhu, R., Khairunizam, W., & Yaacob, S. (2012). Infant cry classification to identify asphyxia using time-frequency analysis and radial basis neural networks. *Expert Systems with Applications*, *39*(10), 9515–9523. doi:10.1016/j.eswa.2012.02.102

Hariharan, M., Sindhu, R., & Yaacob, S. (2012). Normal and hypoacoustic infant cry signal classification using time-frequency analysis and general regression neural network. *Computer methods and programs in biomedicine*, *108*(2), 559–569. doi:10.1016/j.cmpb.2011.07.010

Hariharan, M., Yaacob, S., & Awang, S. A. (2011). Pathological infant cry analysis using wavelet packet transform and probabilistic neural network. *Expert Systems with Applications*, *38*(12), 15377–15382. doi:10.1016/j.eswa.2011.06.025

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89. doi:10.1080/19312450709336664

Honda, K., Kitahara, K., Matsunaga, S., Yamashita, M., & Shinohara, K. (2012). Emotion classification of infant cries with consideration for local and global features. In *Proceedings of the signal information processing association annual summit and conference* (pp. 1–4).

Howard, C. R., Lanphear, N., Lanphear, B. P., Eberly, S., & Lawrence, R. A. (2006). Parental responses to infant crying and colic. the effect on breastfeeding duration. *Breastfeeding medicine : the official journal of the Academy of Breastfeeding Medicine*, *1*(3), 146–155. doi:10.1089/bfm.2006.1.146

IBM. (2011). Spss statistics software. version 19.0.0.1.

IBM. (2013a). Spss modeler. version 15.

IBM. (2013b). Spss statistics. version 20.

IBM. (2016). Spss statistics 23.0.

IEEE. (2014). Ieee xplore digital library. Retrieved January 22, 2014, from http://ieeexplore.ieee.org/

Illingworth, R. S. (1955). Crying in infants and children. *British Medical Journal*, *1*, 75–78.

Jadcherla, S. R., Hogan, W. J., & Shaker, R. (2010). Physiology and pathophysiology of glottic reflexes and pulmonary aspiration: From neonates to adults. *Seminars in respiratory and critical care medicine*, *31*(5), 554–560. doi:10.1055/s-0030-1265896

Jambu, M. (1991). *Exploratory and multivariate data analysis*. Statistical Modeling and Decision Science. San Diego: Elsevier Science.

Jones, M. C. (1971). Diagnostic implications of acoustic cry features. *Journal of Communication Disorders*, *4*(4), 310–316. doi:10.1016/0021-9924(71)90010-4

Kent, R. D. [Ray D.]. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of speech and hearing research*, *19*, 421–447.

Kent, R. D. [Raymond D.]. (1997). *The speech sciences*. San Diego: Singular Publ. Group.

Kheddache, Y., & Tadj, C. (2012). Newborn's pathological cry identification system. In *Proceedings of the 11th international conference on information science, signal processing and their applications* (pp. 1024–1029).

Kia, M., Kia, S., Davoudi, N., & Biniazan, R. (2012). A detection system of infant cry using fuzzy classification including dialing alarm calls function. In *Proceedings of the second international conference on innovative computing technology* (pp. 224–229).

Kiese-Himmel, C. (Ed.). (2016). *Körperinstrument Stimme*. doi:10.1007/978-3-662-49648-0

Koitschev, A., & Sittel, C. (2012). Laryngomalazie. *HNO*, *60*(7), 573–580. doi:10.1007/s00106-011-2379-8

Koivisto, M., Wasz-Höckert, O., Vuorenkoski, V., Partanen, T. J., & Lind, J. (1970). Cry studies in neonatal hyperbilirubinemia. *Acta paediatrica Scandinavica. Supplement*, *206*(Suppl. s206), 26+.

Koopmans-van Beinum, F. J., Clement, C. J., & van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception. a matter of coordination? *Developmental Science*, *4*(1), 61–70. doi:10.1111/1467-7687.00149

Krippendorff, K. (2003). *Content analysis. an introduction to its methodology* (2nd ed.). Thousand Oaks: Sage Publications.

LaGasse, L. L., Neal, A. R., & Lester, B. M. (2005). Assessment of infant cry. acoustic cry analysis and parental perception. *Mental retardation and developmental disabilities research reviews*, *11*(1), 83–93. doi:10.1002/mrdd.20050

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Laufer, M. Z., & Horii, Y. (1977). Fundamental frequency characteristics of infant non-distress vocalisation during the first 24 weeks. *Journal of Child Language*, (4), 171–184.

Laurent, H. K., Stevens, A., & Ablow, J. C. (2011). Neural correlates of hypothalamic-pituitary-adrenal regulation of mothers with their infants. *Biological psychiatry*, *70*(9).

Lederman, D. (2010). Estimation of infants' cry fundamental frequency using a modified sift algorithm. *Computing Research Repository*.

Lederman, D., Cohen, A., Zmora, E., Wermke, K., Hauschildt, S., & Stellzig–Eisenhauer, A. (2002). On the use of hidden markov models in infants' cry classification. In *Proceedings of the 22nd convention of electrical and electronics engineers in israel* (pp. 350–352). Piscataway, NJ: IEEE.

Lederman, D., Zmora, E., Hauschildt, S., Stellzig–Eisenhauer, A., & Wermke, K. (2008). Classification of cries of infants with cleft-palate using parallel hidden markov models. *Medical & Biological Engineering & Computing*, *46*(10), 965–975. doi:10.1007/s11517-008-0334-y

Leerkes, E. M., Parade, S. H., & Burney, R. V. (2010). Origins of mothers' and fathers' beliefs about infant crying. *Journal of Applied Developmental Psychology*, *31*(6), 467–474. doi:10.1016/j.appdev.2010.09.003

Lester, B. M. (1976). Spectrum analysis of the cry sounds of well-nourished and malnourished infants. *Child development*, *47*(1), 237–241.

Lester, B. M., & Boukydis, C. F. Z. (Eds.). (1985). *Infant crying. theoretical and research perspectives*. New York: Plenum Press.

Lester, B. M., & Boukydis, C. F. Z. (1990). No language but a cry. In H. Papousek, J. Jurgens, & M. Papousek (Eds.), *Nonverbal vocal communication: Comparative and delopmental approaches.* (pp. 41–69). New York: Cambridge University Press.

Lester, B. M., Tronick, E. Z., LaGasse, L. L., Seifer, R., Bauer, C. R., Shankaran, S., ... Maza, P. L. (2002). The maternal lifestyle study. effects of substance exposure during pregnancy on neurodevelopmental outcome in 1-month-old infants. *Pediatrics*, *110*(6), 1182–1192.

Lind, J. (1965). Newborn infant cry. *Acta paediatrica Scandinavica*, (163), 3–128.

Lind, J., Wasz-Höckert, O., Rosberg, G., Theorell, K., Valanne, E. H., Partanen, T. J., & Vuorenkoski, V. (1967). Sound spectrography in pediatric diagnosis. *Acta paediatrica Scandinavica*, *177*(Suppl. s177), 113–119.

Lind, K., & Wermke, K. (2002). Development of the vocal fundamental frequency of spontaneous cries during the first 3 months. *International Journal of Pediatric Otorhinolaryngology*, *64*(2), 97–104.

Lüdge, W., & Gips, P. (1989). Microcomputer-aided studies of cry jitter uttered by newborn children based upon high-resolution analysis of fundamental frequencies. *Computer methods and programs in biomedicine*, *28*(3), 151–156.

Luxton, D. D. (2016). *Artificial intelligence in behavioral and mental health care* (1. Aufl.). s.l.: Elsevier Reference Monographs. Retrieved from http://gbv.eblib.com/patron/FullRecord.aspx?p=4003762

Manfredi, C., Bocchi, L., Orlandi, S., Calisti, M., Spaccaterra, L., & Donzelli, G. P. (2008). Non-invasive distress evaluation in preterm newborn infants. *Conference proceedings: Engineering in Medicine and Biology Society*, 2908–2911.

Manfredi, C., Bocchi, L., Orlandi, S., Spaccaterra, L., & Donzelli, G. P. (2009). High-resolution cry analysis in preterm newborn infants. *Medical engineering & physics*, *31*(5), 528–532. doi:10.1016/j.medengphy.2008.10.003

Michelsson, K. (1971). Cry analyses of symptomless low birth weight neonates and of asphyxiated newborn infants. *Acta paediatrica Scandinavica. Supplement*, *216*, 1–45.

Michelsson, K., Eklund, K., Leppänen, P., & Lyytinen, H. (2002). Cry characteristics of 172 healthy 1- to 7-day-old infants. *Folia Phoniatrica et Logopaedica*, *54*(4), 190–200. doi:10.1159/000063190

Michelsson, K., & Michelsson, O. (1999). Phonation in the newborn, infant cry. *International Journal of Pediatric Otorhinolaryngology*, *49*(Suppl. s1), 297–301.

Michelsson, K., Raes, J., Wasz-Höckert, O., & Thoden, C. J. (1981). Sound spectrographic cry analysis in neonatal diagnostics. an evaluative study. *Journal of Phonetics*, *10*, 79–88.

Michelsson, K., Sirviö, P., Koivisto, M., Sovijarvi, A., & Wasz-Höckert, O. (1975). Spectrographic analysis of pain cry in neonates with cleft palate. *Biology of the neonate*, *26*(5-6), 353–358.

Michelsson, K., Sirviö, P., & Wasz-Höckert, O. (1977). Pain cry in full-term asphyxiated newborn infants correlated with late findings. *Acta paediatrica Scandinavica*, *66*(5), 611–616.

Michelsson, K., Tuppurainen, N., & Aula, P. (1980). Cry analysis of infants with karyotype abnormality. *Neuropediatrics*, *11*(4), 365–376. doi:10.1055/s-2008-1071403

Mijovic, B., Silva, M. P., van den Bergh, B. R., Allegaert, K., Aerts, J.-M., Berckmans, D., & van Huffel, S. (2010). Assessment of pain expression in infant cry signals using empirical mode decomposition. *Methods of information in medicine*, *49*(5), 448–452.

Mohd Ali, M., Mansor, W., Lee, Y. K., & Zabidi, A. (2012). Asphyxiated infant cry classification using simulink model. In *Proceedings of the ieee 8th international colloquium on signal processing and its applications* (pp. 491–494). IEEE.

Möller, S., & Schönweiler, R. (1999). Analysis of infant cries for the early detection of hearing impairment. *Speech Communication*, *28*(3), 175–193. doi:10.1016/S0167-6393(99)00016-3

Morsbach, G., & Murphy, M. C. (1979). Recognition of individual neonates' cries by experienced and inexperienced adults. *Journal of Child Language*, *6*(01), 175–179. doi:10.1017/S030500090000773X

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, *14*(3), 477–492.

Mühler, R., & Hoth, S. (2014). Objektive audiologische Diagnostik im Kindesalter. *HNO*, *62*(10), 702–717. doi:10.1007/s00106-014-2920-7

Newman, J. D. (2007). Neural circuits underlying crying and cry responding in mammals. *Behavioural brain research*, *182*(2), 155–165. doi:10.1016/j.bbr.2007.02.011

Nolten, G. (1984). *Discrimination of neonate cries by mothers, non-mothers and computer analysis*. Retrieved from http://books.google.de/books?id=abo7HQAACAAJ

Norvig, P. (2012). Artificial intelligence. everyday ai. *New Scientist*, *216*(2889), iv–v. doi:10.1016/S0262-4079(12)62784-5

Orlandi, S., Reyes-García, C. A., Bandini, A., Donzelli, G., & Manfredi, C. (2015). Application of pattern recognition techniques to the classification of full-term and preterm infant cry. *Journal of voice*. doi:10.1016/j.jvoice.2015.08.007

Orozco-Garcia, J., & Reyes-García, C. A. (2003). A study on the recognition of patterns of infant cry for the identification of deafness in just born babies with neural networks. *Lecture notes in computer science*, *2905*, 342–349.

Orozco, J., & Reyes-García, C. A. (2003). Detecting pathologies from infant cry applying scaled conjugate gradient neural networks. In M. Verleysen (Ed.), *Proceedings of the 11th european symposium on artificial neural networks* (pp. 349–345). Evere: d-side.

Ortiz, S. D. C., Escobedo Beceiro, D. L., & Ekkel, T. (2004). A radial basis function network oriented for infant cry classification. *Lecture notes in computer science*, *3287*, 374–380.

Own, H. S., & Abraham, A. (2012). A new weighted rough set framework based classification for egyptian neonatal jaundice. *Applied Soft Computing*, *12*(3), 999–1005.

Papile, L. A., Burstein, J., Burstein, R., & Koffler, H. (1978). Incidence and evolution of subependymal and intraventricular hemorrhage: A study of infants with birth weights less than 1,500 gm. *The Journal of pediatrics*, *92*(4), 529–534.

Parsons, C. E., Young, K. S., Jegindo, E.-M. E., Vuust, P., Stein, A., & Kringelbach, M. L. (2014). Music training and empathy positively impact adults' sensitivity to infant distress. *Frontiers in psychology*, *5*, 1440. doi:10.3389/fpsyg.2014.01440

Partanen, T. J., Wasz-Höckert, O., Vuorenkoski, K., Valanne, E. H., & Lind, J. (1967). Auditory identification of pain cry signals of young infants in pathological conditions and its sound spectographic basis. *Annales paediatriae Fenniae*, *13*(2), 56–63.

Perlman, J. M. (2004). Brain injury in the term infant. *Seminars in perinatology*, *28*(6), 415–424.

Petroni, M., Malowany, M., Johnston, C. C., & Stevens, B. J. (1995a). A comparison of neural network architectures for the classification of three types of infant cry vocalizations. In *Proceedings of the 17th annual conference engineering in medicine and biology society* (Vol. 1, 821–822 vol.1). IEEE.

Petroni, M., Malowany, M., Johnston, C. C., & Stevens, B. J. (1995b). On the use of artificial neural networks (anns) for the classification of three types of infant cries. In *Proceedings of the ieee pacific rim conference on communications, computers, and signal processing* (pp. 501–504). IEEE.

Pinyerd, B. J. (1994). Infant cries: Physiology and assessment. *Neonatal network : NN*, *13*(4), 15–20.

Poel, M., & Ekkel, T. (2006). Analyzing infant cries using a committee of neural networks in order to detect hypoxia related disorder. *International journal on artificial intelligence tools*, *15*(3), 397–410.

Porter, F. L., Miller, R. H., & Marshall, R. E. (1986). Neonatal pain cries. effect of circumcision on acoustic features and perceived urgency. *Child development*, *57*(3), 790–802.

Prakash, M., & Johnny, J. C. (2015). Whats special in a child's larynx? *Journal of pharmacy & bioallied sciences*, *7*(Suppl 1), 55–58. doi:10.4103/0975-7406.155797

Prandini, E. L., Pegoraro-Krook, M. I., de Cassia Rillo Dutka, J., & de Castro Marino, V. C. (2011). Occurrence of consonant production errors in liquid phonemes in children with operated cleft lip and palate. *Journal of applied oral science : revista FOB*, *19*(6), 579–585.

Prescott, R. (1975). Infant cry sound; developmental features. *The Journal of the Acoustical Society of America*, *57*(5), 1186–1191.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (2002). *Numerical recipes in c. the art of scientific computing* (2nd ed.). Cambridge: Cambridge University Press.

Proceedings of the ieee 8th international colloquium on signal processing and its applications. (2012), IEEE.

Proceedings of the ieee international conference on computer applications and industrial electronics. (2011), IEEE.

Proceedings of the international conference on computer applications and industrial electronics. (2010).

Prochnow, A. (2013). *Der Erwerb melodisch-rhythmischer Grundbausteine im Rahmen der vorsprachlichen Entwicklung – eine vergleichende Analyse der Schreie von schwedis-*

chen und deutschen Neugeborenen (Inaugural-Dissertation, Julius-Maximilians-Universität, Würzburg).

Protopapas, A., & Eimas, P. D. (1997). Perceptual differences in infant cries revealed by modifications of acoustic features. *The Journal of the Acoustical Society of America*, *102*(6), 3723–3734.

Quick, Z. L., Robb, M. P., & Woodward, L. J. (2009). Acoustic cry characteristics of infants exposed to methadone during pregnancy. *Acta paediatrica*, *98*(1), 74–79.

Quinlan, J. R. (2003). *C4.5: Programs for machine learning* (5th ed.). The Morgan Kaufmann series in machine learning. San Mateo, CA: Morgan Kaufmann.

Radulova, P., & Slancheva, B. (2014). Neonatal hypoxic-ischemic brain injury: Pathogenesis and neuropathology. *Akusherstvo i ginekologiia*, *53*(3), 41–47.

Raes, J., Michelsson, K., Dehaen, F., & Despontin, M. (1982). Cry analysis in infants with infections and congenital disorders og the larynx. *Journal of Pediatric Otorhinolaryngology*, *4*, 156–169.

Raes, J., Michelsson, K., & Despontin, M. (1980). Spectrographic analysis of the crying of infants with laryngeal disorders. *Acta oto-rhino-laryngologica Belgica*, *34*(3), 224–237.

Raiha, H., Lehtonen, L., Huhtala, V., Saleva, K., & Korvenranta, H. (2002). Excessively crying infant in the family: Mother-infant, father-infant and mother-father interaction. *Child: care, health and development*, *28*(5), 419–429.

Rapid-I. (2014). Rapidminer studio 6. Retrieved January 22, 2014, from http://rapidminer.com/products-2/rapidminer-studio/

Rautava, L., Lempinen, A., Ojala, S., Parkkola, R., Rikalainen, H., Lapinleimu, H., ... Lehtonen, L. (2007). Acoustic quality of cry in very-low-birth-weight infants at the age of 1 1/2 years. *Early human development*, *83*(1), 5–12. doi:10.1016/j.earlhumdev.2006.03.004

Reetz, H. (2003). *Artikulatorische und akustische Phonetik* (2nd ed.). Trier: WVT Wiss. Verl. Trier.

Reggiannini, B., Sheinkopf, S. J., Silverman, H. F., Li, X., & Lester, B. M. (2013). A flexible analysis tool for the quantitative acoustic assessment of infant cry. *Journal of Speech Language and Hearing Research*, *56*(5), 1416. doi:10.1044/1092-4388(2013/11-0298)

Reinhard, A., & Sandu, K. (2014). Laryngomalacia: Principal cause of stridor in infants and small children. *Revue medicale suisse*, *10*(444), 1816–1819.

Reuters, T. (2014). Web of science. Retrieved January 22, 2014, from http://www.isiknowledge.com/WOS

Reyes-Galaviz, O. F., Arch-Tirado, E., & Reyes-García, C. A. (2004). Elderly and disabled people. therapy and care - classification of infant crying to identify pathologies in recently born babies with anfis. *Lecture notes in computer science*, *3118*, 408–415.

Reyes-Galaviz, O. F., Cano-Ortiz, S. D., & Reyes-García, C. A. (2008). Validation of the cry unit as primary element for cry analysis using an evolutionary-neural approach. In *Proceedings of the mexican international conference on computer science* (pp. 261–267).

Reyes-Galaviz, O. F., & Reyes-García, C. A. (2004). A system for the processing of infant cry to recognize pathologies in recently born babies with neural networks. In *Proceedings of the ninth international conference speech and computer*, Saint-Petersburg: Publ. House Anatolya.

Reyes-Galaviz, O. F., & Reyes-García, C. A. (2005). Bioinformatics and medical applications - infant cry classification to identify hypo acoustics and asphyxia comparing an evolutionary-neural system with a neural network system. *Lecture notes in computer science*, *3789*, 949–958.

Reyes-Galaviz, O. F., Verduzco-Mendoza, A., Arch-Tirado, E., & Reyes-García, C. A. (2005). Analysis of an infant cry recognizer for the early identification of pathologies. In G. Chollet, A. Esposito, M. Faundez-Zanuy, & M. Marinaro (Eds.), *Nonlinear speech modeling and applications* (pp. 404–409). Springer-Verlag.

Reyes-García, C. A., Reyes-Galaviz, O. F., Cano-Ortiz, S. D., Escobedo Beceiro, D. L., Zatarain, R., & Barrón-Estrada, L. (2010). Soft computing approaches to the problem of infant cry classification with diagnostic purposes. In P. Melin, J. Kacprzyk, & W. Pedrycz (Eds.), *Soft computing for recognition based on biometrics* (Vol. 312, pp. 3–18). Studies in Computational Intelligence. Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-642-15111-8_1

Rietveld, T., & van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*. doi:10.1515/9783110871609

Robb, M. P., & Cacace, A. T. (1995). Estimation of formant frequencies in infant cry. *International Journal of Pediatric Otorhinolaryngology*, *32*(1), 57–67.

Robb, M. P., Crowell, D. H., & Dunn-Rankin, P. (2007). Cry analysis in infants resuscitated for apnea of infancy. *International Journal of Pediatric Otorhinolaryngology*, *71*(7), 1117–1123.

Robb, M. P., Crowell, D. H., & Dunn-Rankin, P. (2013). Sudden infant death syndrome: Cry characteristics. *International Journal of Pediatric Otorhinolaryngology*, *77*(8), 1263–1267. doi:10.1016/j.ijporl.2013.05.005

Robb, M. P., Goberman, A. M., & Cacace, A. T. (1997). An acoustic template of newborn infant crying. *Folia Phoniatrica et Logopaedica*, *49*(1), 35–41.

Rosales-Pérez, A., Reyes-García, C. A., & Gómez-Gil, P. (2011). Genetic fuzzy relational neural network for infant cry classification. In *Proceedings of the third mexican conference on pattern recognition* (pp. 288–296). MCPR'11. Berlin: Springer-Verlag.

Rothgänger, H., Lüdge, W., & Grauel, E. L. (1990). Jitter index of the fundamental frequency of infant's cry as a possible diagnostic tool to predict future developmental problems. part 1: Physiological consideration. *Early Child Development and Care*, (65), 145–152.

Runefors, P., & Arnbjörnsson, E. (2005). A sound spectrogram analysis of children's crying after painful stimuli during the first year of life. *Folia Phoniatrica et Logopaedica*, *57*(2), 90–95. doi:10.1159/000083570

Runefors, P., Arnbjörnsson, E., Elander, G., & Michelsson, K. (2000). Newborn infants' cry after heel-prick. analysis with sound spectrogram. *Acta paediatrica*, *89*(1), 68–72.

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence. a modern approach* (3. ed.). Prentice-Hall series in artificial intelligence. Upper Saddle River, NJ: Prentice-Hall.

Sahak, R., Lee, Y. K. [Y. K.], Mansor, W., Yassin, I. M., & Zabidi, A. (2010). Optimized support vector machine for classifying infant cries with asphyxia using orthogonal least square. In *Proceedings of the international conference on computer applications and industrial electronics* (pp. 692–696).

Sahak, R., Lee, Y. K., Mansor, W., Zabidi, A., & Yassin, I. M. (2011). Detection of asphyxia from infant cry by linear kernel support vector machine enhanced with features from orthogonal least square. In *Proceedings of the ieee international conference on computer applications and industrial electronics* (pp. 341–345). IEEE.

Sahak, R., Mansor, W., Khuan, L. Y., Zabidi, A., & Yassin, I. M. (2012). Detection of asphyxia from infant cry using support vector machine and multilayer perceptron integrated with orthogonal least square. In *Proceedings of the ieee-embs international conference on biomedical and health informatics* (pp. 906–909).

Sahak, R., Mansor, W., Lee, Y. K. [Y. K.], Yassin, I. M., & Zabidi, A. (2010). Performance of combined support vector machine and principal component analysis in recognizing infant cry with asphyxia. In *Proceedings of the 32nd annual international conference of the ieee engineering in medicine and biology society* (pp. 6292–6295). doi:10.1109/IEMBS.2010.5628084

Santiago-Sanchez, K., Reyes-García, C. A., & Gómez-Gil, P. (2009). Type-2 fuzzy sets applied to pattern matching for the classification of cries of infants under neurological risk. *Advances in Pattern Recognition*, *5754*, 201–210.

Sapienza, C. M., Ruddy, B. H., & Baker, S. (2004). Laryngeal structure and function in the pediatric larynx: Clinical applications. *Language, speech, and hearing services in schools*, *35*(4), 299–307.

Saraswathy, J., Hariharan, M., Nadarajaw, T., Khairunizam, W., & Yaacob, S. (2014). Optimal selection of mother wavelet for accurate infant cry classification. *Australasian physical & engineering sciences in medicine / supported by the Australasian College of Physical Scien-*

tists in Medicine and the Australasian Association of Physical Sciences in Medicine, *37*(2), 439–456. doi:10.1007/s13246-014-0264-y

Saraswathy, J., Hariharan, M., Vijean, V., Yaacob, S., & Khairunizam, W. (2012). Performance comparison of daubechies wavelet family in infant cry classification. In *Proceedings of the ieee 8th international colloquium on signal processing and its applications* (pp. 451–455). IEEE.

Sato, K., Hirano, M., & Nagashima, T. (2001). Fine structure of the human newborn and infant vocal fold mucosae. *Annals of otology, rhinology, and laryngology*, *110*(5 Pt 1), 417–424.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of voice*, *16*(4), 509–529.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2004). The influence of hearing impairment on preverbal emotional vocalizations of infants. *Folia Phoniatrica et Logopaedica*, *56*(1), 27–40. doi:10.1159/000075326

Schönweiler, R., Kaese, S., Möller, S., Rinscheid, A., & Ptok, M. (1996a). Anwendung selbstorganisierender neuronaler Netzwerke (Kohonen-Maps) zur Klassifizierung von Stimmschallsignalen am Beispiel der Säuglingsstimme mit und ohne zeitverzögerter auditiver Rückkopplung. *HNO*, *44*(4).

Schönweiler, R., Kaese, S., Möller, S., Rinscheid, A., & Ptok, M. (1996b). Classification of spectrographic voice patterns using self-organizing neuronal networks (kohonen maps) in the evaluation of the infant cry with and without time-delayed feedback. *HNO*, *44*(4), 201–206.

Schönweiler, R., Kaese, S., Möller, S., Rinscheid, A., & Ptok, M. (1996c). Neuronal networks and self-organizing maps. new computer techniques in the acoustic evaluation of the infant cry. *International Journal of Pediatric Otorhinolaryngology*, *38*(1), 1–11.

Schuetze, P., & Zeskind, P. S. (2001). Relations between women's depressive symptoms and perceptions of infant distress signals varying in pitch. *Infancy*, *2*(4), 483–499. doi:10.1207/S15327078IN0204_06

Schuetze, P., Zeskind, P. S., & Eiden, R. D. (2003). The perceptions of infant distress signals varying in pitch by cocaine-using mothers. *Infancy*, *4*(1), 65–83. doi:10.1207/S15327078IN0401_4

Sheinkopf, S. J., Iverson, J. M., Rinaldi, M. L., & Lester, B. M. (2012). Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. *Autism research*, *5*(5), 331–339. doi:10.1002/aur.1244

Sheppard, W. C., & Lane, H. L. (1968). Development of the prosodic features of infant vocalizing. *Journal of speech and hearing research*, *11*(1), 94–108.

Shinya, Y., Kawai, M., Niwa, F., & Myowa-Yamakoshi, M. (2014). Preterm birth is associated with an increased fundamental frequency of spontaneous crying in human infants at term-equivalent age. *Biology Letters*, *10*(8), 20140350. doi:10.1098/rsbl.2014.0350

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations. uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420–428.

Silva, D. G., Oliveira, L. C., & Andrea, M. (2009). Jitter estimation algorithms for detection of pathological voices. *EURASIP Journal on Advances in Signal Processing*, *2009*(1), 567875. doi:10.1155/2009/567875

Silva, M. P., Comerlatto, A. A., Junior, Bevilacqua, M. C., & Lopes-Herrera, S. A. (2011). Instruments to assess the oral language of children fitted with a cochlear implant. a systematic review. *Journal of applied oral science : revista FOB*, *19*(6), 549–553.

Silva, M., Mijovic, B., van den Bergh, B. R. H., Allegaert, K., Aerts, J.-M., van Huffel, S., & Berckmans, D. (2010). Decoupling between fundamental frequency and energy envelope of neonate cries. *Early human development*, *86*(1), 35–40. doi:10.1016/j.earlhumdev.2009.12.006

Singh, A. K., Mukhopadhyay, J., & Rao, K. S. (2013). Classification of infant cries using epoch and spectral features. In *Proceedings of the national conference on communications* (pp. 1–5).

Sirviö, P., & Michelsson, K. (1976). Sound-spectrographic cry analysis of normal and abnormal newborn infants. a review and a recommendation for standardization of the cry characteristics. *Folia phoniatrica*, *28*(3), 161–173.

Sohner, L., & Mitchell, P. (1991). Phnonatory and phonetic characteristics of prelinguistic vocal development in cri du chat syndrome. *Journal of Communication Disorders*, *24*(1), 13–20.

Soltis, J. (2004). The signal functions of early infant crying. *Behavioral and brain sciences*, *27*(4), 443–490.

Stenton, J. (2010). *Long term effects of middle ear problems on learning and behaviour. an investigation into the long term effects of the fluctuating, conductive hearing loss caused by otitis media with effusion for adolescent students*. Saarbrücken: VDM Verlag Dr. Müller.

Stevens, B., McGrath, P., Gibbins, S., Beyene, J., Breau, L., Camfield, C., ... Yamada, J. (2007). Determining behavioural and physiological responses to pain in infants at risk for neurological impairment. *Pain*, *127*(1-2), 94–102. doi:10.1016/j.pain.2006.08.012

Suaste-Rivas, I., Diaz-Mendez, A., Reyes-García, C. A., & Reyes-Galaviz, O. F. (2006). Iv dialogue - hybrid neural network design and implementation on fpga for infant cry recognition. *Lecture notes in computer science*, *4188*, 703–710.

Suaste-Rivas, I., Reyes-Galaviz, O. F., Diaz-Mendez, A., & Reyes-García, C. A. (2004a). A fuzzy relational neural network for pattern classification. *Advances in Pattern Recognition*, *3287*, 358–365.

Suaste-Rivas, I., Reyes-Galaviz, O. F., Diaz-Mendez, A., & Reyes-García, C. A. (2004b). Implementation of a linguistic fuzzy relational neural network for detecting pathologies by infant cry recognition. *Advances in Pattern Recognition*, *3315*, 953–962.

Suthaharan, S. (Ed.). (2016). *Machine learning models and algorithms for big data classification*. Integrated Series in Information Systems. doi:10.1007/978-1-4899-7641-3

Teixeira, J. P., & Gonçalves, A. (2014). Accuracy of jitter and shimmer measurements. *Procedia Technology*, *16*, 1190–1199. doi:10.1016/j.protcy.2014.10.134

The R. Foundation for Statistical Computing. (2014). The r project for statistical computing. Retrieved January 22, 2014, from http://www.r-project.org/

Thoden, C. J., & Koivisto, M. (1980). Acoustic analysis of the normal pain cry. In T. Murry & J. Murry (Eds.), *Infant communication. cry and early speech* (pp. 124–151). Houston: College-Hill Press.

Thoden, C. J., & Michelsson, K. (1979). Sound spectrographic cry analysis in krabbe's disease. *Developmental medicine and child neurology*, *21*(3), 400–401.

Thomas, D. R., & Hodges, I. (2010). *Designing and managing your research project. core skills for social and health research*. Los Angeles: SAGE. Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10546189

Truby, H. M., & Lind, J. (1965). Cry sounds of the newborn infant. *Acta paediatrica*, *54*(Suppl. s163), 8–59. doi:10.1111/j.1651-2227.1965.tb09308.x

Tsukamoto, T., & Tohkura, Y. (1990). Perceptual units of the infant cry. *Early Child Development and Care*, *65*(1), 167–178. doi:10.1080/0300443900650119

Tucker, H. M. (1993). *The larynx* (2. ed.). New York: Thieme Med. Publ.

Vallotton, C. D. (2009). Do infants influence their quality of care? infants' communicative gestures predict caregivers' responsiveness. *Infant Behavior and Development*, *32*(4).

Várallyay, G. (2007). The melody of crying. *International Journal of Pediatric Otorhinolaryngology*, *71*(11), 1699–1708.

Várallyay, G., Benyó, Z., Illényi, A., Farkas, Z., & Kovacs, L. (2004). Acoustic analysis of the infant cry: Classical and new methods. In D. Hudson (Ed.), *Proceedings of the 26th annual international conference of the ieee engineering in medicine and biology society* (pp. 313–316). doi:10.1109/IEMBS.2004.1403155

Vassella, F., Joss, E., Luchsinger, R., Dubois, C., Gloor, R., & Wiesmann, U. (1967). Cri-du-chat syndrome. a case report with spectral analysis of the "meowing". *Helvetica paediatrica acta*, *22*(1), 13–21.

Venuti, P., Caria, A., Esposito, G., de Pisapia, N., Bornstein, M. H., & de Falco, S. (2012). Differential brain responses to cries of infants with autistic disorder and typical development: An fmri study. *Research in developmental disabilities*, *33*(6), 2255–2264. doi:10.1016/j.ridd.2012.06.011

Verduzco-Mendoza, A., Arch-Tirado, E., Reyes-García, C. A., Leybon-Ibarra, J., & Licona-Bonilla, J. (2012). Spectrographic cry analysis in newborns with profound hearing loss and perinatal high-risk newborns. *Cirugia y cirujanos*, *80*(1), 3–10.

Vidakovic, B. (2011). Sensitivity, specificity, and relatives. In B. Vidakovic (Ed.), *Statistics for bioengineering sciences* (pp. 109–130). Springer Texts in Statistics. doi:10.1007/978-1-4614-0394-4_4

Vohr, B. R., Lester, B. M., Rapisardi, G., O'Dea, C., Brown, L., Peucker, M., ... Oh, W. (1989). Abnormal brain-stem function (brain-stem auditory evoked response) correlates with acoustic cry features in term infants with hyperbilirubinemia. *The Journal of pediatrics*, *115*(2), 303–308.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., ... Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America*, *125*(3), 1666–1678. doi:10.1121/1.3075589

Wang, X., Pedrycz, W., Chan, P., & He, Q. (2014). *Machine learning and cybernetics*. doi:10.1007/978-3-662-45652-1

Wasz-Höckert, O., Koivisto, M., Vuorenkoski, V., Partanen, T. J., & Lind, J. (1971). Spectrographic analysis of pain cry in hyperbilirubinemia. *Biology of the neonate*, *17*(3), 260–271.

Wasz-Höckert, O., Lind, J., Vuorenkoski, V., Partanen, T. J., & Valanne, E. H. (1968). *The infant cry: A spectrographic and auditory analysis* (1st ed.). Clinics in developmental medicine. Cambridge University Press.

Weiner, I. B., Freedheim, D. K., Graham, J. R., Schinka, J. A., & Velicer, W. F. (2003). *Handbook of psychology, assessment psychology*. Wiley. Retrieved from http://books.google.de/books?id=L6rjmrz6J1sC

Weiss, S. M., & Kulikowski, C. A. (1994). *Computer systems that learn. classification and prediction methods from statistics, neural nets, machine learning, and expert systems*. San Mateo: Kaufmann.

Wermke, K. (2002). *Untersuchung der Melodieentwicklung im Säuglingsschrei von monozygoten Zwillingen in den ersten 5 Lebensmonaten* (Dissertation, Humboldt-Universität, Berlin).

Wermke, K. (2008). Melodie und Rhythmus in Babylauten und ihr potenzieller Wert zur Frühindikation von Sprachentwicklungsstörungen. *Interdisziplinär*, *16*(3), 190–195.

Wermke, K., Birr, M., Voelter, C., Shehata–Dieler, W., Jurkutat, A., Wermke, P., & Stellzig–Eisenhauer, A. (2011). Cry melody in 2–month–old infants with and without clefts. *The Cleft Palate-Craniofacial Journal*, *48*(3), 321–330. doi:10.1597/09-055

Wermke, K., Hain, J., Oehler, K., Wermke, P., & Hesse, V. (2014). Sex hormone influence on human infants' sound characteristics: Melody in spontaneous crying. *Biology Letters*, *10*(5), 20140095. doi:10.1098/rsbl.2014.0095

Wermke, K., Hauser, C., Komposch, G., & Stellzig–Eisenhauer, A. (2002). Spectral analysis of prespeech sounds (spontaneous cries) in infants with unilateral cleft lip and palate (uclp). a pilot study. *The Cleft Palate-Craniofacial Journal*, *39*(3), 285–294. doi:10.1597/1545-1569(2002)039<0285:SAOPSS>2.0.CO;2

Wermke, K., Leising, D., & Stellzig–Eisenhauer, A. (2007). Relation of melody complexity in infants' cries to language outcome in the second year of life: A longitudinal study. *Clinical linguistics & phonetics*, *21*(11-12), 961–973. doi:10.1080/02699200701659243

Wermke, K., & Mende, W. (1992). Sprache beginnt mit dem ersten Schrei. *Spektrum der Wissenschaft*, *14*(7), 115–118.

Wermke, K., Mende, W., Manfredi, C., & Bruscaglioni, P. (2002). Developmental aspects of infant's cry melody and formants. *Medical engineering & physics*, *24*(7-8), 501–514.

Wermke, K., & Robb, M. P. (2010). Fundamental frequency of neonatal crying. does body size matter? *Journal of voice*, *24*(4), 388–394.

Wermke, K., Ruan, Y., Feng, Y., Dobnig, D., Stephan, S., Wermke, P., ... Shu, H. (2016). Fundamental frequency variation in crying of mandarin and german neonates. *Journal of voice*. doi:10.1016/j.jvoice.2016.06.009

White, K. R., Vohr, B. R., Meyer, S., Widen, J. E., Johnson, J. L., Gravel, J. S., ... Weirather, Y. (2005). A multisite study to examine the efficacy of the otoacoustic emission/automated auditory brainstem response newborn hearing screening protocol: Research design and results of the study. *American journal of audiology*, *14*(2), S186–199. doi:10.1044/1059-0889(2005/021)

Wolff, P. H. (1969). The natural history of crying and other vocalizations in early infancy. In B. M. Foss (Ed.), *Determinants of infant behavior* (pp. 81–109). Methuen.

Wood, R. M., & Gustafson, G. E. (2001). Infant crying and adults' anticipated caregiving responses. acoustic and contextual influences. *Child development*, *72*(5), 1287–1300. doi:10.1111/1467-8624.00348

Wyszynski, D. F. (Ed.). (2002). *Cleft lip and palate. from origin to treatment*. Oxford: Oxford University Press.

Zabidi, A., Khuan, L. Y., & Mansor, W. (2012). Asphyxia screening kit. In *Proceedings of the annual international conference of the ieee engineering in medicine and biology society* (pp. 1298–1301). Piscataway, NJ: IEEE.

Zabidi, A., Khuan, L. Y., Mansor, W., Yassin, I. M., & Sahak, R. (2010). Classification of infant cries with asphyxia using multilayer perceptron neural network. In *Proceedings of the second international conference on computer engineering and applications* (pp. 204–208). IEEE.

Zabidi, A., Khuan, L. Y., Mansor, W., Yassin, I. M., & Sahak, R. (2011). Binary particle swarm optimization for feature selection in detection of infants with hypothyroidism. In *Proceedings of the annual international conference of the ieee engineering in medicine and biology society* (pp. 2772–2775).

Zabidi, A., Mansor, W., Khuan, L. Y., Yassin, I. M., & Sahak, R. (2010). Discrete mutative particle swarm optimisation of mfcc computation for classifying hypothyroidal infant cry. In *Proceedings of the international conference on computer applications and industrial electronics* (pp. 588–592).

Zabidi, A., Mansor, W., Khuan, L. Y., Yassin, I. M., & Sahak, R. (2011). Three-dimensional particle swam optimisation of mel frequency cepstrum coefficient computation and multilayer perceptron neural network for classifying asphyxiated infant cry. In *Proceedings of the ieee international conference on computer applications and industrial electronics* (pp. 290–293). IEEE.

Zabidi, A., Mansor, W., Lee, Y. K. [Yoot Khuan], Yassin, I. M., & Sahak, R. (2011). Binary particle swarm optimization for selection of features in the recognition of infants cries with asphyxia. In *Proceedings of the ieee 7th international colloquium on signal processing and its applications* (pp. 272–276).

Zeskind, P. S., & Lester, B. M. (1978). Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications. *Child development*, *49*(3), 580–589.

Zeskind, P. S., & Lester, B. M. (2001). Analysis of infant crying. In L. T. Singer & P. S. Zeskind (Eds.), *Biobehavioral assessment of the infant* (pp. 149–166). New York: Guilford Press.

Zeskind, P. S., McMurray, M. S., Cox-Lippard, E. T., Grewen, K. M., Garber, K. A., Johns, J. M., & Cooper, B. G. (2014). Translational analysis of effects of prenatal cocaine exposure on human infant cries and rat pup ultrasonic vocalizations. *PLoS ONE*, *9*(10), e110349. doi:10.1371/journal.pone.0110349

Zeskind, P. S., McMurray, M. S., Garber, K. A., Neuspiel, J. M., Cox-Lippard, E. T., Grewen, K. M., … Johns, J. M. (2011). Development of translational methods in spectral analysis of human infant crying and rat pup ultrasonic vocalizations for early neurobehavioral assessment. *Frontiers in psychiatry*, *2*, 56. doi:10.3389/fpsyt.2011.00056

Appendix

A. Praat script

```
form Schallanalytische Untersuchung von Saeuglingsschreien
  comment Auswahl der Gruppenzugehörigkeit
    optionmenu Gruppe: 1
      option Healthy
      option Healthy, spontaneous
      option Healthy, non-distressed
      option Healthy, pain-induced
      option Healthy, pain-induced without 1st
      option Hearing-impaired
      option UCLP
      option Asphyxia
      option Apoplex
      option Laryngomalacia

  comment Soll eine bereits vorhandene Ausgabedatei
   überschrieben werden (Vorhandene Daten werden gelöscht)? Wenn
   nicht, wird die Datei um die neuen Daten erweitert!
    boolean Ausgabedatei_ueberschreiben 0

  comment Soll die Ergebnis-Datei nachbearbeitet werden, so dass
    sie mit SPSS kompatibel ist?
    boolean Ausgabedatei_nachbearbeiten 1

  comment Grundfrequenzanalyse
    # Bereich ([min_F0_(Hz), max_F0_(Hz)]) in dem die
     Grundfrequenz identifiziert werden soll
    natural Minimale_F0_(Hz) 100
    natural Maximale_F0_(Hz) 1000
```

```
25
26   comment Formantenanalyse
27     # Anzahl Formanten und maximale Frequenz für die
       Identifizierung von Formanten
28     natural Anzahl_Formanten 6
29     natural Maximale_Formanten_Frequenz_(Hz) 8000
30 endform
31
32 # "Datei Speichern"-Dialog zur Auswahl, wohin die Ergebnisse
   geschrieben werden sollen
33 outfile$ = chooseWriteFile$ ("Waehlen Sie eine Datei zum
   Speichern der Ergebnisse...", "results.csv")
34
35 if outfile$ = ""
36   exit Keine Ausgabedatei gewählt! Breche Analyse ab...
37 endif
38
39 # Lösche ggf. eine existierende Ausgabe-Datei und leere das
   Praat Ausgabefenster
40 if ausgabedatei_ueberschreiben = 1
41   filedelete 'outfile$'
42 endif
43 clearinfo
44
45 # Verwendet um Bearbeitungszeit zu messen
46 stopwatch
47
48 # Frage Anzahl der zu analysierenden Sounds ab
49 n = numberOfSelected ("Sound")
50
51
52 # Frage alle ausgewählten Sound-Dateien ab
53 for i to n
54   sound'i' = selected ("Sound", i)
55 endfor
56
```

```
# Schreibe die Spaltenbezeichner in die CSV-Ausgabedatei, falls diese neu angelegt wurde
if fileReadable (outfile$) = 0
fileappend "'outfile$'" Filename;Group;Infant ID;Cry utterance; Sampling frequency [Hz];Cry duration [s];
fileappend "'outfile$'" F0 min [Hz];F0 P10 [Hz];F0 median [Hz]; F0 IQR [Hz];F0 mean [Hz];F0 mean SD [Hz];F0 P90 [Hz];F0 max [Hz];
for formant to anzahl_Formanten
  # Spalten für jeden Formanten hinzufügen
  fileappend "'outfile$'" F'formant' min [Hz];F'formant' P10 [Hz];F'formant' median [Hz];F'formant' IQR [Hz];F'formant' at max itensity [Hz];F'formant' mean [Hz];F'formant' mean SD [Hz];F'formant' P90 [Hz];F'formant' max [Hz];
endfor
fileappend "'outfile$'" Intensity min [dB];Intensity P10 [dB]; Intensity median [dB];Intensity IQR [dB];Intensity mean [dB]; Intensity mean SD [dB];Intensity P90 [dB];Intensity max [dB];
fileappend "'outfile$'" Jitter (local) [%];Shimmer (local) [%];
fileappend "'outfile$'" Fraction of locally unvoiced pitch frames [%];Number of voice breaks;Degree of voice breaks [%];
fileappend "'outfile$'" HNR mean [dB];HNR mean SD [dB]
fileappend "'outfile$'" 'newline$'
endif

# Analysiere jede ausgewählte Sound-Datei
for i to n

  select sound'i'
  name$ = selected$ ("Sound")

  # Identifizieren der Säuglings-ID und der Schreinummer
  # Muss an das jeweilige Dateiformat angepasst werden!
  group$ = left$(name$, index(name$,"-")-1)
  infantId$ = group$ + mid$(name$, index(name$,"-")+1, 3)
  cryUtterance$ = mid$(name$, index(name$,"-")+5, 2)
```

```
83
84  # Frage den Namen der i-ten Sound-Datei ab
85  printline >>> Analysiere Sound-Datei 'name$'...
86
87  # Frage allgemeine Werte ab
88  sampling_frequenz = Get sampling frequency
89  schreidauer = Get total duration
90
91  # Berechne die Grundfrequenz (F0)
92  printline 'tab$'Berechne Grundfrequenz...
93  noprogress To Pitch (ac)... 0 minimale_F0 15 yes 0.03 0.45
     0.01 0.35 0.14 maximale_F0
94  f0_min = Get minimum... 0 0 Hertz Parabolic
95  f0_max = Get maximum... 0 0 Hertz Parabolic
96  f0_median = Get quantile... 0 0 0.50 Hertz
97  f0_p10 = Get quantile... 0 0 0.10 Hertz
98  f0_p90 = Get quantile... 0 0 0.90 Hertz
99  f0_p25 = Get quantile... 0 0 0.25 Hertz
100 f0_p75 = Get quantile... 0 0 0.75 Hertz
101 f0_iqr = f0_p75 - f0_p25
102 f0_mean = Get mean... 0 0 Hertz
103 f0_mean_stdev = Get standard deviation... 0 0 Hertz
104 Remove
105 select sound'i'
106
107 # Berechne die Intensität
108 printline 'tab$'Berechne Intensitaet...
109 noprogress To Intensity... minimale_F0 0 yes
110 int_min = Get minimum... 0 0 Parabolic
111 int_max = Get maximum... 0 0 Parabolic
112 int_max_at_time = Get time of maximum... 0 0 Parabolic
113 int_median = Get quantile... 0 0 0.5
114 int_p10 = Get quantile... 0 0 0.1
115 int_p90 = Get quantile... 0 0 0.9
116 int_p25 = Get quantile... 0 0 0.25
117 int_p75 = Get quantile... 0 0 0.75
```

```
int_iqr = int_p75 - int_p25
int_mean = Get mean... 0 0 energy
int_mean_stdev = Get standard deviation... 0.0 0.0
Remove
select sound'i'

# Berechne die Formanten
printline 'tab$'Berechne Formanten...
noprogress To Formant (burg)... 0 anzahl_Formanten maximale_Formanten_Frequenz 0.025 50
for formant to anzahl_Formanten
  # Berechne Formant-Werte
  f'formant'_min = Get minimum... 'formant' 0 0 Hertz Parabolic
  f'formant'_p10 = Get quantile... 'formant' 0 0 Hertz 0.1
  f'formant'_p25 = Get quantile... 'formant' 0 0 Hertz 0.25
  f'formant'_median = Get quantile... 'formant' 0 0 Hertz 0.5
  f'formant'_mean = Get mean... 'formant' 0 0 Hertz
  f'formant'_mean_stdev = Get standard deviation... 'formant' 0 0 Hertz
  f'formant'_p75 = Get quantile... 'formant' 0 0 Hertz 0.75
  f'formant'_p90 = Get quantile... 'formant' 0 0 Hertz 0.9
  f'formant'_max = Get maximum... 'formant' 0 0 Hertz Parabolic
  f'formant'_iqr = f'formant'_p75 - f'formant'_p25
endfor
Remove
select sound'i'

# Berechne die "Harmonics-to-Noise-Ratio" (HNR) Werte
printline 'tab$'Berechne HNR-Werte...
noprogress To Harmonicity (cc)... 0.01 minimale_F0 0.1 1
hnr_mean = Get mean... 0 0
hnr_mean_stdev = Get standard deviation... 0 0
Remove
select sound'i'
```

```
150
151  # Berechne Jitter, Shimmer und Voice Breaks
152  printline 'tab$'Berechne Jitter, Shimmer und Voice-Breaks...
153  noprogress To Pitch (cc)... 0 minimale_F0 15 yes 0.03 0.45
      0.01 0.35 0.14 maximale_F0
154  pitch = selected ("Pitch")
155  select sound'i'
156  plus pitch
157  noprogress To PointProcess (cc)
158  pointprocess = selected ("PointProcess")
159
160  # Berechne Jitter
161  jitter = Get jitter (local)... 0 0 0.0001 0.02 1.3
162  jitter = jitter * 100
163
164  # Berechne Shimmer
165  select sound'i'
166  plus pointprocess
167  shimmer = Get shimmer (local)... 0 0 0.0001 0.02 1.3 1.6
168  shimmer = shimmer * 100
169
170  # Berechne Voice Breaks über den Voice Report
171  select sound'i'
172  plus pitch
173  plus pointprocess
174  voiceReport$ = Voice report... 0 0 100 1000 1.3 1.6 0.03 0.45
175  fractionOfLocallyUnvoicedFrames = extractNumber (voiceReport$,
       "Fraction of locally unvoiced frames: ") * 100
176  noOfVoiceBreaks = extractNumber (voiceReport$, "Number of
      voice breaks: ")
177  degreeOfVoiceBreaks = extractNumber (voiceReport$, "Degree of
      voice breaks: ") * 100
178  #noFrames = Get number of frames
179  #noVoicedFrames = Count voiced frames
180  #noUnVoicedFrames = noFrames - noVoicedFrames
```

```
181 #fractionOfLocallyUnvoicedFrames = (noUnVoicedFrames /
     noFrames) * 100
182
183 # Entferne temporäre Objekte
184 select pointprocess
185 Remove
186 select pitch
187 Remove
188 select sound'i'
189
190 # Schreibe die Schallparameter des aktuellen Sounds in die CSV
     -Ausgabedatei
191 printline 'tab$'Schreibe Ergebnisse in CSV-Ausgabedatei...
192 fileappend "'outfile$'" 'name$';'Gruppe';'infantId$';'
     cryUtterance$:2';'sampling_frequenz';'schreidauer:4';
193 fileappend "'outfile$'" 'f0_min:3';'f0_p10:3';'f0_median:3';'
     f0_iqr:3';'f0_mean:3';'f0_mean_stdev:3';'f0_p90:3';'f0_max:3';
194 for formant to anzahl_Formanten
195   f_min = f'formant'_min
196   f_p10 = f'formant'_p10
197   f_med = f'formant'_median
198   f_at_max_int = f'formant'_at_max_int
199   f_iqr = f'formant'_iqr
200   f_mean = f'formant'_mean
201   f_mean_stdev = f'formant'_mean_stdev
202   f_p90 = f'formant'_p90
203   f_max = f'formant'_max
204   # Gebe alle formanten aus
205   fileappend "'outfile$'" 'f_min:3';'f_p10:3';'f_med:3';'f_iqr
       :3';'f_at_max_int:3';'f_mean:3';'f_mean_stdev:3';'f_p90:3';'
       f_max:3';
206 endfor
207 fileappend "'outfile$'" 'int_min:3';'int_p10:3';'int_median
     :3';'int_iqr:3';'int_mean:3';'int_mean_stdev:3';'int_p90:3';'
     int_max:3';
208 fileappend "'outfile$'" 'jitter:3';'shimmer:3';
```

```
209   fileappend "'outfile$'" 'fractionOfLocallyUnvoicedFrames:3';'
       noOfVoiceBreaks';'degreeOfVoiceBreaks:3';
210   fileappend "'outfile$'" 'hnr_mean:3';'hnr_mean_stdev:3'
211   fileappend "'outfile$'" 'newline$'
212
213   printline <<< Analyse 'i' von 'n' (Sound-Datei 'name$')
       abgeschlossen!
214
215 endfor
216
217 # Bearbeite Ausgabedatei nach
218 if ausgabedatei_nachbearbeiten = 1
219   if fileReadable("'outfile$'")
220     printline 'tab$'Stelle Kompatibilität der Ausgabedatei mit
       SPSS her...
221     result$ < 'outfile$'
222
223     # Ersetze alle '.' in Zahlen durch das deutsche
       Dezimaltrennzeichen ','
224     newResult$ = replace$(result$, ".", ",", 0)
225     newResult$ = replace$(newResult$, "--undefined--", "", 0)
226     filedelete 'outfile$'
227     fileappend "'outfile$'" 'newResult$'
228   endif
229 endif
230
231 time = stopwatch
232 printline
233 printline ANALYSE VON 'n' SOUND-DATEIEN IN 'time:2' SEKUNDEN
     ABGESCHLOSSEN.
234 printline ERGEBNISSE WURDEN IN FOLGENDE CSV-DATEI GESPEICHERT: '
     outfile$'!
235
236 # Stelle die Selektion im Praat Objects Fenster wieder her
237 if n >= 1
238   select sound1
```

```
239   for i from 2 to n
240     plus sound'i'
241   endfor
242 endif
```

B. Ethical clearing I

HOCHSCHULE FRESENIUS
UNIVERSITY OF APPLIED SCIENCES

Hochschule Fresenius gemeinnützige GmbH· Limburger Straße 2 · D-65510 Idstein

Frau
Tanja Etz
Müllerwies 14

65232 Taunusstein

HOCHSCHULE FRESENIUS
STANDORT IDSTEIN

Limburger Straße 2
D - 65510 Idstein

idstein@hs-fresenius.de
www.hs-fresenius.de

Fachbereich Gesundheit
Studiengang Naturheilkunde
und komplementäre Medizin

Fon +49 (0)61 26. 93 52 - 901
Fax +49 (0)61 26. 93 52 - 177
guendling@hs-fresenius.de

Idstein, 10.10.2012

Antrag an die Ethikkommission der Hochschule Fresenius
Der Säuglingsschrei – Ein Frühindikator für Störungsbilder mit einer potentiell pathologischen Sprachentwicklung?

Sehr geehrte Frau Etz,

Ihr o.g. Antrag wurde eingehend in der Ethikkommission der Hochschule Fresenius beraten.

Gegen die Studie „**Der Säuglingsschrei – Ein Frühindikator für Störungsbilder mit einer potentiell pathologischen Sprachentwicklung?**" bestehen keine berufsethischen Bedenken.

Die Ethikkommission bittet um zeitnahe Unterrichtung über alle schwerwiegenden oder unerwarteten unerwünschten Ereignisse, die während der Studie auftreten und die Sicherheit der Studienteilnehmer oder die Durchführung der Studie beeinträchtigen könnten. Dies gilt auch, wenn die Studie aus unvorhergesehenen Gründen abgebrochen wird.

Es wird weiterhin darauf hingewiesen, dass Änderungen oder Erweiterungen des Versuchsplanes oder dessen Anlagen der Ethikkommission umgehend schriftlich anzuzeigen sind und ggf. eine erneute Beratung erforderlich wird. Wir bitten, die einzureichenden Änderungen und/oder Erweiterungen der Studienunterlagen deutlich zu kennzeichnen.

Die fachliche und juristische Verantwortung des Studienleiters und der an der Studie teilnehmenden Therapeuten und Prüfer bleibt entsprechend der Beratungsfunktion der Ethikkommission durch unsere Stellungnahme unberührt.

Wir wünschen Ihnen viel Erfolg bei der Durchführung Ihrer Studie und bitten um eine Kopie Ihrer Abschlussarbeit nach Abschluss der Untersuchung.

Mit herzlichen Grüßen

Prof. Dr. med. Peter W. Gündling MSc
Vorsitzender der Ethikkommission der Hochschule Fresenius

Schulträger: Hochschule Fresenius gemeinnützige GmbH · Limburger Straße 2 · D-65510 Idstein
Geschäftsführer: Dipl.-Kfm. Hermann Kögler · Amtsgericht Wiesbaden · HRB 19044
Nassauische Sparkasse KTO-Nr. 104 000 363 · BLZ 510 500 15
Frankfurter Sparkasse KTO-Nr. 200 386 654 · BLZ 500 502 01
vr Bank Untertaunus eG KTO-Nr. 12 587 708 · BLZ 510 917 00
Finanzamt-Nr. 2222 · Steuer-Nr. 22/870/01919 ; USt.-Identifikations-Nr. DE196697659

IDSTEIN KÖLN HAMBURG MÜNCHEN

C. Ethical clearing II

HOCHSCHULE FRESENIUS
UNIVERSITY OF APPLIED SCIENCES

Hochschule Fresenius gemeinnützige GmbH· Limburger Straße 2 · D-65510 Idstein

Frau
Tanja Etz
Müllerwies 14

65232 Taunusstein

HOCHSCHULE FRESENIUS
STANDORT IDSTEIN

Limburger Straße 2
D - 65510 Idstein

idstein@hs-fresenius.de
www.hs-fresenius.de

Fachbereich Gesundheit
Studiengang Naturheilkunde
und komplementäre Medizin

Fon +49 (0)61 26. 93 52 – 901
Fax +49 (0)61 26. 93 52 – 177
guendling@hs-fresenius.de

Idstein, 03.04.2013

Antrag an die Ethikkommission der Hochschule Fresenius
Analyse der Reliabilität von schmerzinduzierten Schreien im Vergleich zum spontanen Säuglingsschrei

Sehr geehrte Frau Etz,

Ihr o.g. Antrag wurde eingehend in der Ethikkommission der Hochschule Fresenius beraten.

Gegen die Studie „**Analyse der Reliabilität von schmerzinduzierten Schreien im Vergleich zum spontanen Säuglingsschrei**" bestehen keine berufsethischen Bedenken.

Die Ethikkommission bittet um zeitnahe Unterrichtung über alle schwerwiegenden oder unerwarteten unerwünschten Ereignisse, die während der Studie auftreten und die Sicherheit der Studienteilnehmer oder die Durchführung der Studie beeinträchtigen könnten. Dies gilt auch, wenn die Studie aus unvorhergesehenen Gründen abgebrochen wird.

Es wird weiterhin darauf hingewiesen, dass Änderungen oder Erweiterungen des Versuchsplanes oder dessen Anlagen der Ethikkommission umgehend schriftlich anzuzeigen sind und ggf. eine erneute Beratung erforderlich wird. Wir bitten, die einzureichenden Änderungen und/oder Erweiterungen der Studienunterlagen deutlich zu kennzeichnen.

Die fachliche und juristische Verantwortung des Studienleiters und der an der Studie teilnehmenden Therapeuten und Prüfer bleibt entsprechend der Beratungsfunktion der Ethikkommission durch unsere Stellungnahme unberührt.

Wir wünschen Ihnen viel Erfolg bei der Durchführung Ihrer Studie und bitten um eine Kopie Ihrer Abschlussarbeit nach Abschluss der Untersuchung.

Mit herzlichen Grüßen

Prof. Dr. med. Peter W. Gündling MSc MME
Vorsitzender der Ethikkommission der Hochschule Fresenius

Schulträger: Hochschule Fresenius gemeinnützige GmbH · Limburger Straße 2 · D-65510 Idstein
Geschäftsführer: Dipl.-Kfm. Hermann Kögler · Amtsgericht Wiesbaden · HRB 19044

Nassauische Sparkasse KTO-Nr. 104 000 363 · BLZ 510 500 15
Frankfurter Sparkasse KTO-Nr. 200 386 654 · BLZ 500 502 01
vr Bank Untertaunus eG KTO-Nr. 12 587 708 · BLZ 510 917 00

Finanzamt-Nr. 2222 · Steuer-Nr. 22/870/01919 · USt.-Identifikations-Nr. DE196697659

IDSTEIN KÖLN HAMBURG MÜNCHEN

D. Questionnaire for the listening experiment

Fragebogen PI Hörversuch

Bitte machen Sie folgende allgemeine Angaben!

Geschlecht:

männlich ..... ☐
weiblich ..... ☐

Alter:

____ ____ Jahre

Berufserfahrung seit:

____ ____ Jahren

Haben Sie Kinder?

Nein ..... ☐
Ja (Anzahl der Kinder?) ☐

_____

Sind bei Ihnen Hörschwierigkeiten bekannt?

Nein ..... ☐
Ja (welche?) ☐

_____

_____

Anleitung

Im Folgenden werden Ihnen die akustischen Eigenschaften des gesunden Säuglingsschreis wie auch Schreie von Säuglingen mit verschiedenen Störungsbildern beschrieben und vorgespielt. Im Anschluss bekommen Sie insgesamt 18 Hörbeispiele dargeboten. Bitte ordnen Sie die 18 Hörbeispiele den jeweiligen Gruppen

A) gesunder Säugling
B) Säugling mit Hörstörung
C) Säugling mit Lippen-Kiefer-Gaumenspalte (LKGS)
D) Säugling mit Sauerstoffmangel
E) Säugling mit Kehlkopferweichung
F) Säugling mit Schlaganfall

zu. Bitte kreuzen Sie pro Hörbeispiel nur 1 Gruppe an.

Bitte ordnen Sie die 18 Hörbeispiele jeweils einer Gruppe zu!

☞ ***Kreuzen Sie je Hörbeispiel nur eine Gruppe an!***

	Gesund	**Hörstörung**	**LKGS**	**Sauerstoffmangel**	**Kehlkopferweichung**	**Schlaganfall**
Hörbeispiel 1	☐	☐	☐	☐	☐	☐
Hörbeispiel 2	☐	☐	☐	☐	☐	☐
Hörbeispiel 3	☐	☐	☐	☐	☐	☐
Hörbeispiel 4	☐	☐	☐	☐	☐	☐
Hörbeispiel 5	☐	☐	☐	☐	☐	☐
Hörbeispiel 6	☐	☐	☐	☐	☐	☐
Hörbeispiel 7	☐	☐	☐	☐	☐	☐
Hörbeispiel 8	☐	☐	☐	☐	☐	☐
Hörbeispiel 9	☐	☐	☐	☐	☐	☐
Hörbeispiel 10	☐	☐	☐	☐	☐	☐
Hörbeispiel 11	☐	☐	☐	☐	☐	☐
Hörbeispiel 12	☐	☐	☐	☐	☐	☐
Hörbeispiel 13	☐	☐	☐	☐	☐	☐
Hörbeispiel 14	☐	☐	☐	☐	☐	☐
Hörbeispiel 15	☐	☐	☐	☐	☐	☐
Hörbeispiel 16	☐	☐	☐	☐	☐	☐
Hörbeispiel 17	☐	☐	☐	☐	☐	☐
Hörbeispiel 18	☐	☐	☐	☐	☐	☐

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit mit dem Titel “Exploring the ability of acoustic infant cry analysis for discriminating developmental pathologies ” selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel in Anspruch genommen habe, sowie die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, durch Angabe der Quellen kenntlich gemacht habe.

Taunusstein, 24.07.2018 Tanja Fuhr