



FAKULTÄT
Informatik, Mathematik und Naturwissenschaften

VISUALISIERUNG KONTINUIERLICHER, MULTIMODALER SCHMERZ SCORES AM BEISPIEL AKUSTISCHER SIGNALE

Abschlussarbeit zur Erlangung des akademischen Grades
MASTER OF SCIENCE

Studiengang MEDIENINFORMATIK

Autor:
Franz ANDERS

Erstbetreuer HTWK:
Prof. Dr. Mario HLAWITSCHKA

Zweitbetreuer LaBP:
Mirco FUCHS

17. März 2017

Eidesstattliche Versicherung

Ich erkläre hiermit, dass ich diese Masterarbeit selbstständig ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel verfasst habe. Alle den benutzten Quellen wörtlich oder sinngemäß entnommenen Stellen sind als solche einzeln kenntlich gemacht. Diese Arbeit ist bislang keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht worden. Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Leipzig, 17. März 2017

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen der Schmerzbewertung mit Hilfe akustischer Signale	3
2.1	Schmerz und Weinen bei Neugeborenen aus medizinischer Sicht	3
2.1.1	Pain Scales	3
2.1.2	Weinen bei Neugeborenen	6
2.2	Signalverarbeitung	7
2.2.1	Grundlegende Definitionen	8
2.2.2	Statistische Merkmale	8
2.2.3	Fehlersignale	9
2.2.4	Kurzzeit-Fourier-Transformation	10
2.2.5	Akustische Modellierung der menschlichen Stimme	12
2.3	Schreiforschung	17
2.3.1	Physio-Akustische Modellierung des Weinens	18
2.3.2	Diskussion	20
2.4	Klassifizierung und Regression	21
2.4.1	ID3 und C4.5	23
2.4.2	Gütemaße binärer Klassifikatoren	26
3	Ableitung der Schmerz Score	28
3.1	Segmentierung	28
3.2	Extrahierung von Eigenschaften und Ableitung der Schmerzscores	31
3.2.1	Extrahierung von Eigenschaften	33
3.2.2	Ableitung der Pain-Score	36
4	Visualisierung	41
5	Zusammenfassung	45
	Appendices	49
A	Auswertungsdetails VAD	50

Abbildungsverzeichnis

2.1	Statistische Merkmale eines Beispielsignals über dem Intervall [50,200] . . .	9
2.2	Ein 1.8-Sekunden langes Signal. Oben: Der Zeitbereich mit drei klar erkennbaren Events. Unten: Das Frequenz-Spectrum des gesamten Signals mit logarithmisierten Achsen.	11
2.3	Windowing: Die Zerlegung eines Signals in kürzere Fenster.	11
2.4	Das Hamming-Window	12
2.5	STFT des Beispiel-Signals aus Abbildung 2.3	12
2.6	Schematische Übersicht über die Organe der Spracherzeugung. Lung = Lunge, Vocal Chords = Stimmbänder, Pharynx = Rachen, Velum = Halszäpfchen, Mouth Cavity = Mundraum, Nasal Cavity = Nasenraum [24]	13
2.7	Schematische Übersicht über das Source-Filter-Model [13, nach Source estimation, S. 17]	14
2.8	Zeitbereiche der periodischen und der turbulenten Quelle [25, Source]	14
2.9	Betrachtung der Frequenzbereiche des Source-Filter-Model (nach: [13, Source Estimation, S. 3])	15
2.10	Grundfrequenz und harmonische Obertöne eines periodischen Sprachsignals.	16
2.11	Formanten im Sprach-Signal (nach: [2])	16
2.12	Spektrogramm einer Audioaufnahme eines Babys. Rot $\hat{=}$ hohen Amplituden, Blau $\hat{=}$ niedrigen Amplituden. Oben: Zeitbereich. Mitte: Spektrogramm mit einer Fensterlänge von 185 ms(8192-Sample DFT). Unten: Spektrogramm mit einer Fensterlänge von 5 ms(265-Sample DFT)	17
2.13	Veranschaulichung des Grundvokabulars	18
2.14	(1) Pitch of Shift (2) Maximale Grundfrequenz (3) Minimum der Grundfrequenz (4) Biphonation (5) Double Harmonic Break (6) Vibrato (7) Glide (8) Furcation [36, S. 142]	20
2.15	Entscheidungsbaum, der durch den ID3-Algorithmus für den Datensatz aus Beispiel 2.3 erzeugt wurde.	23
2.16	Aufspaltung einer kontinuierlichen Variable im Entscheidungsbaum	26
2.17	Confusion-Matrix (nach: [20, S. 214])	26
3.1	Beispiel für eine Segmentierung mit einem Grenzwert von $t_s = 6$ s	30
3.2	Beispiel für die Ableitung von Pain Scores für ein Signal nach einer fiktiven Pain Scale ohne Beobachtungszeitraum oder Aktualisierungsintervall	39
3.3	Beispiel für die Ableitung von Pain Scores für ein Signal nach einer fiktiven Pain Scale mit $t_{act} = 7.5$ s und $t_{obs} = 15$ s	39
4.1	Oben: Ein Beispielsignal, sowie zwei Segmente, die mit $t_s = 10$ s gefunden wurden. Darunter: Pain Scores, die für die beiden Segmente berechnet wurden. Darunter: Schematische Darstellung des Signals als Regionen. Unten: Scores der Regionen.	42

4.2	Erstellte Farbpaletten mit 2, 3, 4 und 5 Farben inklusive der zugehörigen RGB-Koordinaten im Hex-Code.	44
A.1	Boxplot-Auswertung über Sensitivität, Spezifität und Genauigkeit der beiden besten Modelle	51

Tabellenverzeichnis

2.1	Neonatal Infant Pain Scale (NIPS) [14]	4
2.2	Übersicht über Pain-Scales. [29, S. 98] [37] [33] [9] [3] [18] [15] [5] [28] [9] . .	6
2.3	Beispieldatensatz D für die Klassifikation mit ID3	23
3.1	Fiktive Pain Scale	38
4.1	Fiktive Pain Scales zur Erläuterung der Visualisierung	41
4.2	Fixpunkte als Basis der Farbpaletten	44
A.1	Klassifikationsgenauigkeiten der durch den REPTree entworfenen Modelle . .	50

1 Einleitung

Die Schmerzdiagnostik Neugeborener ist eine Aufgabe, die in vielen medizinischen Kontexten notwendig ist: Angefangen von der Überwachung prozeduraler Eingriffe wie Impfungen, über die Therapie von Entzugserscheinungen infolge des Drogenkonsums der Mutter bis hin zu Komplikationen bei der Geburt, welche die Beobachtung der Neugeborenen über mehrere Wochen bis Monate erfordert. Es gibt starke Hinweise darauf, dass insbesondere lang anhaltender Schmerz, welcher nicht korrekt erkannt und dessen Ursache somit nicht beseitigt wird, das Schmerzempfinden des Betroffenen bis in das Erwachsenenalter nachhaltig beeinflussen kann.[17, S. 402]

Babys äußern das Vorhandensein von Schmerz anhand einer Reihe von Merkmalen. Die für den Menschen am leichtesten sichtbaren Zeichen sind beispielsweise das Verkrampfen des Gesichtsausdruckes oder das Weinen. Aber auch physiologische Veränderungen wie der Erhöhung der Herzfrequenz oder des Blutdruckes lassen Rückschlüsse auf den Schmerzgrad zu.[32, S. 440] Auf Basis dieser Indikatoren werden im klinischen Alltag zur Schmerzdiagnostik sogenannte *multimodale Pain Scales* eingesetzt. Eine Pain Scale listet eine Reihe von Merkmalen, wie zum Beispiel den Hautfarbe vor das Weinen, für die eine Korrelation mit Schmerz festgestellt wurde. Für jedes Merkmal werden, je nach Grad der Ausprägung, Punkte vergeben, um schlussendlich den Schmerzgrad als *Pain Score* zu quantifizieren.[17, S. 406]

Forschungsbestrebungen der letzten Jahre haben sich zum Ziel gesetzt, die Schmerzbewertung zu automatisieren. Das Ziel dieser Arbeit ist die Entwicklung eines Konzeptes zur kontinuierlichen Schmerzbewertung bei Neugeborenen in einem multimodalen Verbund. Dabei soll der Schmerzgrad nicht nur abgeleitet, sondern auch visualisiert werden. Der Fokus liegt dabei auf der Analyse des Schmerzindikators *Weinen*, das heisst der Analyse akustischer Signale. Ausgangspunkt aller Entwürfe und Überlegungen sind die über viele Jahre erprobten Pain Scales.

Die Lösung dieser Aufgabenstellung erfordert die Kombination verschiedener Wissenschaftsdisziplin, wie zum Beispiel der medizinischen Schreiforschung oder der akustischen Modellierung der menschlichen Stimme. Kapitel 2 bietet eine Einführung in die wichtigsten Grundlagen. In Kapitel ?? wird ein Überblick über Veröffentlichungen mit ähnlichen Zielstellungen gegeben sowie dass in dieser Arbeit entwickelte Konzept grundlegend erläutert. Die folgenden Kapitel beleuchten die einzelnen Bausteine dieses Konzeptes genauer. In Kapitel ?? werden Methoden zur Feststellung des Vorhandenseins von Weingeräuschen in einem akustischen Signal vorgestellt. Dazu werden klassische Methoden der Voice Activity Detection bezüglich ihrer Leistungsfähigkeit im speziellen Anwendungsfall kindlicher Lautäußerungen eruiert und evaluiert. In Kapitel 2.3.1 werden Methoden entworfen, um aus den festgestellten Weingeräuschen den Schmerzgrad abzuleiten. Möglichkeiten der Visualisierung der Pain Score werden abschließend in Kapitel 4 vorgestellt.

Diese Arbeit leistet vor allem den Entwurf eines modular aufgebauten und erweiterbaren Konzeptes, welches die kontinuierliche Feststellung und Visualisierung von Schmerz Scores

ermöglicht. Das Modul, welches am ausgiebigsten evaluiert werden konnte, ist die Voice Activity Detection. Da alle darauf folgenden Module ausgiebigere Datenerhebungen in Zusammenarbeit mit medizinischen Fachkräften benötigen, wurden hier begründete Vorschläge zur Aufgabenbewältigung entworfen, welche jedoch nicht evaluiert werden konnten. Das Ziel dieser Arbeit war insbesondere nicht die Erlangung neuer Erkenntnisse über die tatsächliche Schmerzwahrnehmung Neugeborener.

2 Grundlagen der Schmerzbewertung mit Hilfe akustischer Signale

Das Ziel dieses Kapitels ist es, wichtig Grundlagen zum Verständnis der Schmerzbewertung bei Neugeborenen auf Basis akustischer Signale zu legen. Dazu wird in Kapitel 2.1 zunächst erläutert, wie die Schmerzbewertung aus Sicht medizinischer Fachkräfte im klinischen Alltag durchgeführt wird. Der Fokus liegt dabei insbesondere auf der Ableitung des Schmerzgrades aus dem Weinen des Babys. Um die Stimme des Babys automatisiert analysieren zu können, werden Methoden der Signalverarbeitung verwendet. Daher werden in Kapitel 2.2 Grundlagen zur akustischen Modellierung der menschlichen Stimme erläutert. In Kapitel 2.3 wird eine Einführung in die „klassische Schreiforschung“ gegeben. Dabei handelt es sich um Wissenschaftsgebiet, bei dem Methoden der Signalverarbeitung verwendet werden, um tieferegehende Analysen des Weinens Neugeborener durchzuführen. Da sich das in dieser Arbeit vorgestellte Konzept als Erweiterung der klassischen Methoden versteht, ist ein Verständnis dieses Wissenschaftsgebietes unerlässlich. In Kapitel 2.4 werden Grundlagen des überwachten maschinellen Lernens erläutert, da diese bei der Spracherkennung ausgiebigen Einsatz erfahren.

2.1 Schmerz und Weinen bei Neugeborenen aus medizinischer Sicht

Schmerz wird definiert als eine „ein unangenehmes Sinnes- oder Gefühlserlebnis, das mit tatsächlicher oder potenzieller Gewebeschädigung einhergeht“.[32, S. 438] Abseits von dieser theoretischen Definition hat der Mensch ein intuitives Verständnis für Schmerz, da jeder ihn in seine Leben mindestens Einmal erfahren musste. In der ersten Hälfte des 20ten Jahrhunderts war die vorherrschende Meinung, dass Neugeborene keinen Schmerz empfinden können. Beispielsweise bekamen sie nach Operationen keine Schmerzmittel verabreicht. Der aktuelle Stand der Forschung besagt, dass Neugeborene im selben Maße wie Erwachsene Schmerz empfinden können. Die freien Nervenenden, die in der Lage sind, physische Schäden am Körper festzustellen, sind bei Neugeborenen ebenso wie bei Erwachsenen über den Körper verteilt. Die hormonelle Reaktion ist ebenfalls vergleichbar. [17, S. 402] [32, S. 438]

2.1.1 Pain Scales

Es gibt diverse Gründe, die bei Neugeborenen Schmerz verursachen können. Sie reichen über physische Schäden, aufgrund von Komplikationen bei der Geburt oder Gewalteinwirkungen, über Erkrankungen, wie Kopfschmerzen oder Infektionen, bis hin zu therapeutischen Prozeduren, wie Injektionen oder Desinfektionen von Wunden. Das Vorhandensein von Schmerz ist anhand diverser physiologischer, biochemischer, verhaltensbezogener und psychologischer

Veränderungen messbar. Die für diese Arbeit wichtigste Verhaltensänderung ist das Weinen, welches zu den verhaltensbezogenen Veränderungen gezählt wird.[32, S. 441]

Schlussendlich ist Schmerz ein subjektives Empfinden. Das heißt, dass ein und der selbe Stimulus bei zwei verschiedenen Personen zu einem unterschiedlichem Schmerzempfinden führen kann. Daher wird der Schmerzgrad bei Erwachsenen typischerweise durch eine Selbsteinschätzung des Patienten unter der Leitung gezielter Fragen des Arztes festgestellt. Bei Kindern unter 3 Jahren ist diese Selbsteinschätzung nicht möglich. Diese Einschätzung muss daher von anderen Personen vorgenommen werden. Im klinischen Kontext sind dies medizinische Fachkräfte, wie beispielsweise Ärzte, Krankenpfleger oder Geburtshelfer. Die von außen am leichtesten feststellbaren Indikatoren von Schmerz sind die verhaltensbasierten Merkmale, wie zum Beispiel ein Verkrampfen des Gesichtsausdrucks, erhöhte Körperbewegungen oder lang anhaltendes Weinen.[32, S. 438] Die Schmerzdiagnostik durch eine andere Person ist, genau wie das Schmerzempfinden, etwas inherent subjektives und abhängig von Faktoren wie dem Alter, Geschlecht, kulturellen Hintergrund, persönlichen Erfahrungen mit Schmerz usw.[12, S. 3] Um die Schmerzdiagnostik objektiver zu gestalten, wurden daher sogenannte *Pain Scales* entwickelt, welche mit Hilfe eines Punktesystems den Schmerzgrad des Babys quantifizieren.[32, S. 438 - 439] Es existieren *monomodale / unidimensionale Pain Scales*, bei denen der Schmerzgrad aus der Beobachtung *eines* Merkmals geschlossen wird, wie beispielsweise des Gesichtsausdrucks. Ein Merkmal wird in diesem Zusammenhang als *Schmerzindikator* bezeichnet. *Multimodale / Multidimensionale Pain Scales* beziehen mehrere Schmerzindikatoren in das Scoring mit ein.[19, S. 69 - 71].

Tabelle 2.1 zeigt das Scoring-System „Neonatal Infant Pain Scale“ (NIPS) als Beispiel für eine multimodale Pain Scale. Diese Pain Scale ist für Babys von 0 bis 1 Jahr geeignet. Sie ist vor allem für die Diagnose schon Schmerz geeignet, der während einer Prozedur entsteht. Sie wurde auf der Basis der Erfahrungen von Krankenschwestern erarbeitet. Das Baby soll bei der Anwendung dieser Pain Scale für ungefähr eine Minute beobachtet werden. Wird die Scale verwendet, um den Schmerzgrad nach der Prozedur festzustellen, wird empfohlen, die Diagnose alle 30 Minuten durchzuführen. Für jede der aufgeführten Kategorie (Schmerzindikatoren) werden ein, zwei oder drei Punkte vergeben und anschließend aufsummiert. Ein insgesamt Wert von > 3 zeigt moderaten Schmerz an, ein Wert von > 4 großen Schmerz.[14] [29, S. 98]

Tabelle 2.1: Neonatal Infant Pain Scale (NIPS) [14]

NIPS	0 points	1 point	2 points
Facial Expr.	Relaxed	Contracted	-
Cry	Absent	Mumbling	Vigorous
Breathing	Relaxed	Different than basal	-
Arms	Relaxed	flexed/stretched	-
Legs	Relaxed	flexed/stretched	-
Alertness	Sleeping	uncomfortable	-

Nach dem Muster der NIPS existieren viele weitere Pain Scales. Sie unterscheiden sich hinsichtlich der Schmerzindikatoren, die betrachtet werden, dem Punktesystem, der Art des Schmerzes, die festzustellen ist, dem Beobachtungszeitraum usw. Einige Pain Scales sind beispielsweise auf die Schmerzdiagnostik während eines Eingriffes spezialisiert, andere auf den darauf folgenden Heilungsprozess. In den meisten multimodalen Pain Scales wird das Weinen oder Schreien der Babys als Schmerzindikator mit einbezogen. In der englischen Fachliteratur ist von „Cry“ die Rede.[29, S. 97 - 98] In dieser Arbeit wird „Cry“ mit „Weinen“

oder mit dem neutraleren Begriff „kindliche Lautäußerungen“ übersetzt. Tabelle 2.2 zeigt eine Übersicht über einige multimodale Pain Scales. Die Übersicht zeigt vor allem, nach welchen Kriterien das Weinen in den jeweiligen Scales bewertet wird. Außerdem wird für jede Pain Scale angegeben, für welches Alter sie geeignet ist, welcher Schmerz-Typ diagnostiziert wird, sowie der zur Diagnose vorgesehene Beobachtungszeitraum und -Intervall. Angaben, die mit einem ? verzeichnet wurden, konnten nicht in Erfahrung gebracht werden. Es handelt sich hierbei nur eine Übersicht über die wichtigsten Fakten der Pain Scales. Die Anleitungen der jeweiligen Pain Scales geben weitere Anweisungen zur Benutzung.

System	P.	Description	other Ind.	Comments
FLACC	0	No cry (awake or asleep)	Face,	Age: 2 months - 7 years
	1	Moans or whimpers; occasional complaint	Legs, Activity,	Observe for: 1 - 5 minutes
	2	Crying steadily, screams or sobs, frequent complaints	Consolability	Observe every: ? Pain-Type: Ongoing
N-PASS	-2	No cry with painful stimuli	Behaviour, Facial Expr., Extremities, Vital Signs	Age: 0 - 100 days
	-1	Moans or cries minimally with painful stimuli		Observe for: ?
	0	Appropriate Crying		Observe every: 2 - 4 hours
	1	Irritable or Crying at Intervals. Consolable		Pain-Type: Ongoing
	2	High-pitched or silent-continuous crying. Not consolable		
BPSN	0	No Crying	Alertness,	Age: ?
	1	Crying less than 2 minutes	Skin Color, Eyebrows,	Observe for: ?
	2	Crying more than 2 minutes	...	Observe every: ?
	3	Shrill Crying more than 2 minutes		Pain Type: ?
CRIES	0	If no cry or cry which is not high pitched	O2,	Age: 0 - 6 Months
	1	If cry high pitched but baby is easily consoled	Vital Signs, Expression, Sleeplessness	Observe for: ?
	2	If cry is high pitched and baby is inconsolable		Observe every: 1 hour Pain-Type: Post Operative
COVERS	0	No Cry	O2,	Age: ?
	1	High-Pitched or visibly crying	Vital Signs, Expression,	Observe for: ?
	2	Inconsolable or difficult to soothe	...	Observe every: ? Pain Type: Procedural
PAT	0	No	Posture, Sleep Pattern,	Age: 0 - 3 months
	1	When disturbed, doesn't settle after handling, loud, whimper, whining	Expression, ...	Observe for: 15 - 30 sec Observe every: 30 min Pain Type: Post Operative
DAN	0	Moans Briefly	Facial Exp., Limb Mov.	Age: 0 - 2 years
	1	Intermittent Crying		Observe for: ?
	2	Long-Lasting Crying, Continuous howl		Observe every: ? Pain Type: Procedural
COMFORT	0	No crying	Alertness, Calmness, Respiration, ...	Age: 0 - 3 years
	1	Sobbing or gasping		Observe for: ?
	2	Moaning		Observe every: ?
	3	Crying		Pain: Post Operative
	4	Screaming		
MBPS	0	Laughing or giggling	Facial Exp., Movement	Age: ?

1	Not Crying	Observe for: ? Observe every: ?
2	Moaning quiet vocalizing gentle or whimpering cry	Pain Type: Procedural
3	Full lunged cry or sobbing	
4	Full lunged cry more than baseline cry	

Tabelle 2.2: Übersicht über Pain-Scales. [29, S. 98 | [37] [33] [9] [3] [18] [15] [5] [28] [9]

Da die Begriffe *Pain Scale* und *Pain Score* in einigen Veröffentlichungen inkonsistent verwendet werden, wird in dieser Arbeit die Konvention getroffen, dass mit *Pain Scale* das System zur Schmerzdiagnostik gemeint ist und mit *Pain Score* die auf Basis der Pain Scale vergebene Punktzahl. *NIPS* ist also beispielsweise eine Pain Scale, und 3 eine Pain Score.

Die folgende Schlussfolgerungen werden bezüglich der Pain Scales aus Tabelle 2.2 gezogen:

1. Die Kriterien zur Bewertung des Weinens werden zum größten Teil mit *subjektiv behafteten Begriffen* beschrieben. Beispielsweise wird bei dem *N-PASS*-System ein Score von drei für „High-pitched or silent-continuous crying“ vergeben. Die Begriffe „high-pitched“ und „silent-continuous“ werden nicht näher definiert. Auch die Anwendungsvorschriften der Pain Scales geben keine festen Definitionen. Dies erleichtert den praktischen Einsatz der Pain Scales, führt jedoch zu einem Interpretationsspielraum und somit zu einem von der diagnostizierenden Person abhängigen Scoring. Die *BPSN*-Scale nutzt als einzige der vorgestellten Scales objektiv messbare Eigenschaften.
2. Die Pain Scales fokussieren unterschiedliche Eigenschaften zur Bestimmung der Pain Score bezüglich des Weinen-Indikators. Bei *CRIES* ist die Tonhöhe, bei *BPSN* die Länge und bei *COMFORT* die Art des Weinens ausschlaggebend für ein höheres Scoring.
3. Die Beschreibungen sind kurz und prägnant gehalten, die diagnostizierende Person hat bei keiner Pain Scale auf mehr als drei Eigenschaften des Weinens zu achten.

2.1.2 Weinen bei Neugeborenen

An dieser Stelle stellt sich der Leser eventuell die Frage, woher die unterschiedlichen Bewertungskriterien für das Weinen in den Pain Scales stammen. Gibt es eine „beste“ Pain Scale? Dieser Frage unterliegen zwei grundlegendere Fragen:

1. Ist es möglich, aus den akustischen Eigenschaften den motivierenden Grund für die Lautäußerung abzuleiten? Klingt ein durch Hunger bedingtes Weinen anders als ein durch Schmerz bedingtes?
2. Ist es möglich, anhand der akustischen Eigenschaften den Schweregrad dieses motivierenden Grundes abzuleiten?

Die Annahme, dass es möglich sei, aus den Eigenschaften des Weinens den Grund abzulesen zu können, wird als „Cry-Types Hypothesis“ bezeichnet. Die berühmtesten Befürworter dieser Hypothese ist eine skandinavische Forschungsgruppe, auch bezeichnet als „Scandinavian Cry-Group“, die die Idee in dem Buch „Infant Crying: Theoretical and Research Perspectives“ [21] publik machte. Die Hypothese besagt, dass die Empfindungen *Hunger*, *Freude*, *Schmerz*, *Geburt* sowie Sonstiges klare Unterschiede hinsichtlich der akustischen Merkmale des

Weinens aufweisen würden. Diese Unterschiede seien im Spektrogramm sichtbar (Siehe Kapitel 2.2.4). Wenige Jahre Später zeigten Müller et al. [8], dass bei leichter Veränderung der Experimentbedingungen die Unterscheidung nicht mehr möglich sei. Die Gegenhypothese ist, dass Weinen „nichts als undifferenziertes Rauschen“ sei. Bis heute liegt kein anerkannter Beweis für die eine oder andere Hypothese vor. Es gibt lediglich starke Hinweise dafür, dass sich die Plötzlichkeit des Eintretens des Grundes in den akustischen Eigenschaften bemerkbar macht. Ein plötzliches Ereignis, wie ein Nadelstich oder ein lautes Geräusch, führen auch zu einem plötzlich beginnenden Weinen. Ein langsam eintretendes Ereignis, wie ein langsam zunehmender Schmerz oder Hunger führen auch zu einem langsam eintretenden Weinen. Da nach Kenntnis des Autors bis heute keine wissenschaftlich belastbarer Beweis vorgelegt wurde, wird empfohlen, den Grund aus dem Kontext abzuleiten.[36, S. 9 - 13, 17 - 19]

Die Zweite Frage nach der Ableitung der Stärke des Unwohlseins aus den akustischen Eigenschaften des Weinens wird in der Fachliteratur unter dem Begriff *Cry as a graded Signal* subsumiert. Je „stärker“ das Weinen, desto höher sei das Unwohlsein (*Level of Distress (LoD)*) des Säuglings. Tatsächlich bemessen wird dabei der von dem Beobachter vermutete Grad des Unwohlseins des Babys, und nicht der tatsächliche Grad, da dieser ohne die Möglichkeit der direkten Befragung des Babys nie mit absoluter Sicherheit bestimmt werden kann. Ein hohes Unwohlsein hat vor allem eine schnelle Reaktion der Aufsichtspersonen zur Beruhigung des Babys zur Folge, womit dem Weinen eine Art Alarmfunktion zukommt. Es gibt starke Hinweise darauf, dass das Level of Distress anhand objektiv messbarer Eigenschaften des Audiosignals bestimmt werden kann. So herrscht beispielsweise weitestgehend Einigung darüber, dass ein „lang“ anhaltendes Wein auf einen hohen Level of Distress hinweist. Insofern aus dem Kontext des Weinens Schmerz als die wahrscheinlichste Ursache eingegrenzt werden kann, kann aus einem hohen Level of Distress ein hoher Schmerz abgeleitet werden. [36, S. 13 - 17] [35] Es herrscht wiederum keine Einigung darüber, welche akustischen Eigenschaften im Detail ein hohes Level of Distress anzeigen. Carlo V Bellieni et al. [5] haben festgestellt, dass bei sehr hohem Schmerz in Bezug auf die DAN-Scale (siehe Tabelle 2.2) die Tonhöhe steigt. Qiaobing Xie et al. [35] haben festgestellt, dass häufiges und dysphoniertes Schreien auf einen hohen Level of Distress hinweist.

2.2 Signalverarbeitung

In Kapitel 2.1 wurde erläutert, wie Weinen von Neugeborenen mit Hilfe subjektiv behafteter Begriffe eingeschätzt werden kann. Möchte man das Weinen objektiv beschreiben und messbar machen, so verwendet man die Methoden der digitalen Signalverarbeitung. An dieser Stelle wird eine Einführung in die wichtigsten Themen, die im Zusammenhang mit der Audiosignalverarbeitung größere Bedeutung haben. Es wird ein grundlegendes Verständnis der Signalverarbeitung vorausgesetzt, da aus Platzgründen keine für Neulinge geeignete Einführung in das Themengebiet gewährleisten werden kann. Falls dieses Wissen nicht vorhanden ist, wird zur Einarbeitung das Buch „The Scientist and Engineer’s Guide to Digital Signal Processing“ von Steven W. Smith empfohlen.[39], welches kostenlos als E-Book bereitgestellt wird.

2.2.1 Grundlegende Definitionen

In dieser Arbeit sind nur *digitale Signale* von Bedeutung. Ein digitales Signal $x[\]$ ist nach Formel 2.1 eine beliebige Zahlenfolge mit diskretem Definitionsbereich. Dem Definitionsbereich kommt die Bedeutung *Zeit* zu.[39, S. 11-12] In dieser Arbeit gilt die Konvention, dass mit $x[\]$ das gesamte Signal gemeint ist und mit $x[n]$ ein Wert des Signals zum Zeitpunkt/Index n . Ein Wert $x[n]$ wird auch als *Sample* bezeichnet. Die Samplingfrequenz des digitalen Signals wird mit f_s bezeichnet.

$$x[\] := \forall n \in \mathbb{Z} : x[n] = s \quad (2.1)$$

Der Definitionsbereich eines Signals erstreckt sich implizit immer von negativer bis positiver Unendlichkeit. Das heißt nicht, dass alle Samples des Signals auch Informationen enthalten müssen. Der *Support* ist das kleinst mögliche Zeitintervall, das alle Samples enthält, die nicht den Wert 0 haben, wie Formel 2.2 definiert. Wird also auf ein Sample zugegriffen, das außerhalb des Supportes liegt, hat dieses Sample den Wert 0 (bezeichnet als „0-Sample“)[30, S. 24]

$$\begin{aligned} \text{Sup}(x[\]) &= [sup_s, sup_e] \quad , sup_s, sup_e \in \mathbb{Z} \\ , x[sup_s] &\neq 0 \wedge x[sup_e] \neq 0 \wedge \forall n \notin [sup_s, sup_e] : x[n] = 0 \end{aligned} \quad (2.2)$$

Die *Dauer* eines Signals ist die Länge des Supportes nach Formel 2.3. In dieser Arbeit herrscht die Konvention, dass die Länge des Signals kurz mit der Variable N abgekürzt wird. Wenn nicht anders definiert, erstreckt sich der Support eines Signals von $0, \dots, N - 1$. [30, S. 24]

$$\text{Length}(x[\]) = sup_e - sup_s + 1 = N \quad (2.3)$$

2.2.2 Statistische Merkmale

Im folgenden wird ein Überblick über häufig berechnete Merkmale von Signalen gegeben. Abbildung 2.1 visualisiert die Erläuterungen.

1. Der **Maximalwert** / **Minimalwert** beschreibt den höchsten / niedrigsten in $x[\]$ enthaltenen Wert nach den Formel 2.4.

$$\begin{aligned} \max(x[\]) &= \max_{n \in \text{Sup}(x[\])} \{ x[n] \} \\ \min(x[\]) &= \min_{n \in \text{Sup}(x[\])} \{ x[n] \} \end{aligned} \quad (2.4)$$

2. Der **Durchschnittswert** / **Average Value** beschreibt den durchschnittlichen Wert aller Samples von $x[\]$ nach Formel 2.5. Dieser Durchschnittswert wird über ein beliebiges Intervall $[n_1, n_2]$ berechnet.

$$\text{AVG}(x[\]) = \frac{1}{n_2 - n_1 + 1} \sum_{n=n_1}^{n_2} x[n] \quad (2.5)$$

3. Der **Mean Squared Value** (*MSV*) beschreibt den quadrierten Durchschnittswert über eine bestimmtes Intervall nach Formel 2.6. Er wird auch als *durchschnittliche Energie* oder *average Power* bezeichnet.

$$\text{MSV}(x[\]) = \frac{1}{n_2 - n_1 + 1} \sum_{n=n_1}^{n_2} x[n]^2 \quad (2.6)$$

4. Das **Root Mean Square** (*RMS*) wird definiert als die Wurzel des Mean Squared Value nach Formel 2.7. Der RMS kann im Vergleich zum MSV besser ins Verhältnis zu den Werten des Signals gesetzt werden kann. Er wird im Deutschen auch als **Effektivwert** oder **Durchschnittsleistung** bezeichnet. Da die deutschen Begriffe in einigen Quellen jedoch auch für den MSV verwendet werden, wird in dieser Arbeit nur mit den englischen Begriffen gearbeitet.

$$\text{RMS}(x[\]) = \sqrt{\frac{1}{n_2 - n_1 + 1} \sum_{n=n_1}^{n_2} x[n]^2} \quad (2.7)$$

5. Die **Energie** / **Energy** eines Signals wird nach Formel 2.8 definiert. Sie entspricht dem MSV-Wert multipliziert mit der Länge des Intervalls. [30, S. 27-28]

$$E(x[\]) = \sum_{n=n_1}^{n_2} x[n]^2 \quad (2.8)$$

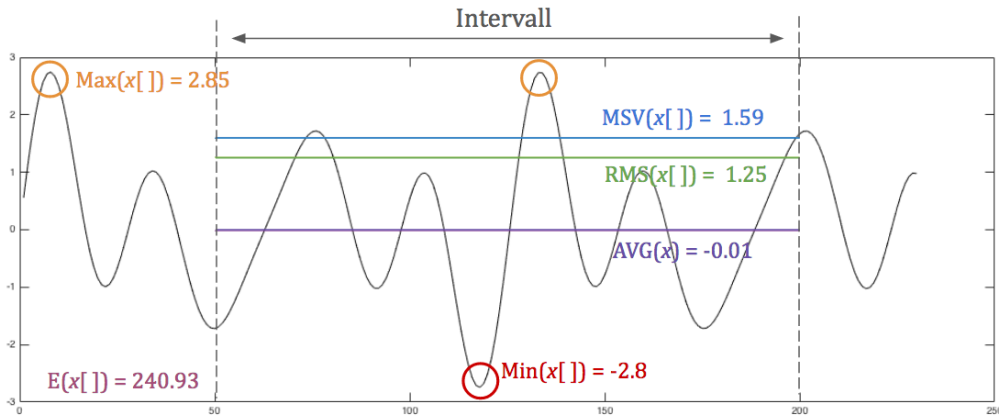


Abbildung 2.1: Statistische Merkmale eines Beispielsignals über dem Intervall [50,200]

2.2.3 Fehlersignale

Angenommen, ein Signal $x[\]$ wird übertragen, auf dem Übertragungsweg jedoch durch ein anderes Störsignal wie z.B. Rauschen $e[\]$ überlagert. $e[\]$ wird in diesem Zusammenhang als das *Fehlersignal* bezeichnet. Das resultierende *Nutzsignal* $x'[\]$ wird nach Formel 2.9 berechnet.

$$x'[\] := \bigvee_{n=n_1}^{n_2} : x'[n] = x[n] + e[n] \quad (2.9)$$

Eine Möglichkeit der Quantifizierung der Stärke des Rauschens im Vergleich zum Signal ist, den MSV des Eingangssignal ins Verhältnis zum MSV des Fehlersignals zu setzen. Formel 2.10 gibt die Definition.

$$\text{SNR}_{rel}(x[\], e[\]) = \frac{MSV(x[\])}{MSV(e[\])} \quad (2.10)$$

In der Praxis ist der MSV des Eingangssignals meist sehr viel höher als der des Fehlersignals. Um den Zahlenraum zu begrenzen, wird die Pseudoeinheit dB verwendet. Formel 2.11 definiert den *Signal/Rausch-Abstand* (SNR, englisch Signal-to-Noise-Ratio). Ein *niedriger* SNR weist auf ein *starkes* Rauschen hin, und ein *hoher* SNR auf ein *schwaches* Rauschen. Im Zusammenhang mit der Spracherkennung ist der Signal/Rausch-Abstand von Bedeutung, da ein höheres Rauschen die Verarbeitung des Nutzsignals, der Sprache, erschwert.

$$\text{SNR}(x[\], e[\]) = 10 \cdot \lg \left(\frac{MSV(x[\])}{MSV(e[\])} \right) \text{ dB} \quad (2.11)$$

2.2.4 Kurzzeit-Fourier-Transformation

Das Signal $x[\]$ befindet sich im *Zeitbereich*, da die unabhängige Variable die Zeit definiert. Gleichung 2.12 definiert die *komplexe diskrete Fouriertransformation*, kurz *DFT*, die das diskrete Signal $x[\]$ aus dem Zeitbereich in den Frequenzbereich $X[\]$ transformiert. Das Signal des Frequenzbereiches ist, ebenso wie das Signal des Zeitbereiches, N punkte Lang und hat den Support $0, \dots, N-1$. Jedes Sample des Frequenzbereiches ist eine komplexe Zahl, deren Realteil $\Re(x[k])$ die Amplitude der entsprechenden Sinuswelle mit der Frequenz $f = k \frac{f_s}{N}$ bezeichnet und deren Imaginärteil $\Im(x[k])$ die Amplitude der entsprechenden Kosinuswelle bezeichnet.[39, S. 149, S. 567 - 571] [1, S. 60]

$$\text{DFT}\{x[\]\} = X[\] := \bigvee_{k=0}^{N-1} : X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j2\pi k \frac{n}{N}} \quad (2.12)$$

Das *Frequenz-Spektrum*, kurz auch nur als *Spektrum* bezeichnet, wird in dieser Arbeit nach Gleichung 2.13 definiert als der Absolutwert des Frequenzbereiches im Bereich $0, \dots, N/2$.

$$\text{Spektrum} := |X[0]|, \dots, |X[N/2]| \quad (2.13)$$

Abbildung 2.2 visualisiert die Transformation in den Frequenzbereich: In der Abbildung ist oben der Zeitbereich eines 1.8 Sekunden langen Signals zu sehen. Es können klar drei nacheinander gespielte Töne erkannt werden. Der Zeitbereich lässt erkennen, zu welchen Zeitpunkten die Töne beginnen und Enden, aber nicht, welche Frequenzkomponenten in den Tönen enthalten sind. Das heißt, es kann beispielsweise nicht erkannt werden, ob es sich um hohe oder tiefe Töne handelt. Unten ist das Spektrum abgebildet. Die x-Achse bezeichnet die Frequenz von 0 bis 22050 Hz und die y-Achse die Amplitude der entsprechenden Frequenz. Beide Achsen werden logarithmiert dargestellt. Das Frequenzspektrum zeigt, welche Frequenzkomponenten im dem Signal enthalten sind. So kann beispielsweise erkannt werden, dass keine Frequenzen unterhalb von 1000 Hz in dem Beispielsignal enthalten sind. Das Spektrum macht jedoch nicht erkennbar, zu welchen Zeitpunkten die Töne beginnen oder enden.

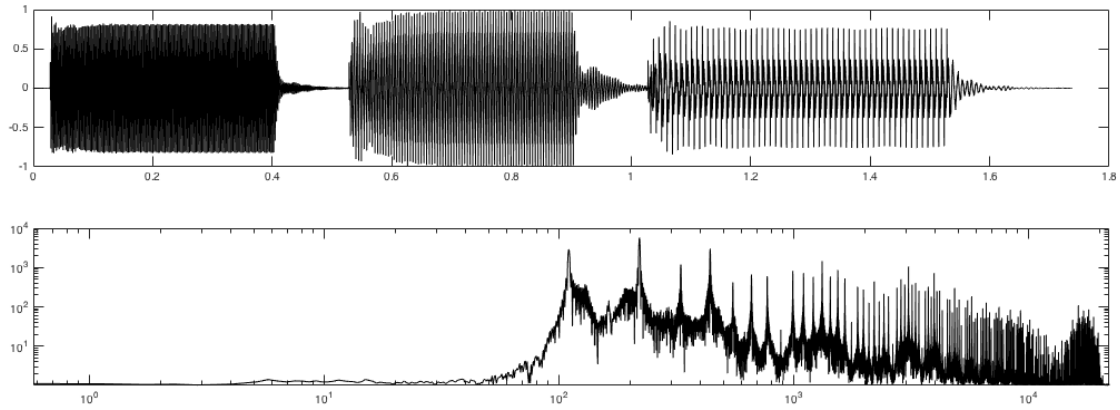


Abbildung 2.2: Ein 1.8-Sekunden langes Signal. Oben: Der Zeitbereich mit drei klar erkennbaren Events. Unten: Das Frequenz-Spectrum des gesamten Signals mit logarithmisierten Achsen.

Es ist wünschenswert, einen Kompromiss aus den Vorteilen beider Bereiche zu finden, in dem man das Spektrum kürzerer Zeitabschnitte des Signals bildet. Dazu wird der Zeitbereich $x[\]$ in Fenster der Länge M zerlegt. Die zeitliche Differenz zwischen zwei Fenstern wird als *Hoptime* R bezeichnet. Gleichung definiert die Bildung des *Signalfensters* $x_i[\]$. Die komplette Zerlegung eines Signals in Signalfenster wird als *Windowing* bezeichnet.[38]

$$x_i[\] := \begin{matrix} M-1 \\ \forall \\ n=0 \end{matrix} : x_m[n] = x[n + i \cdot R] \quad (2.14)$$

Abbildung 2.3 gibt ein Beispiel für die Zerlegung eines Signals $x[\]$ in die Signalfenster $x_0[\], \dots, x_4[\]$. Die Samplingrate des Signals ist $f_s = 44100$, die Fensterlänge beträgt $M = 22050/f_s = 0.5$ s und die Hoptime $R = M/2 = 0.25$ s.

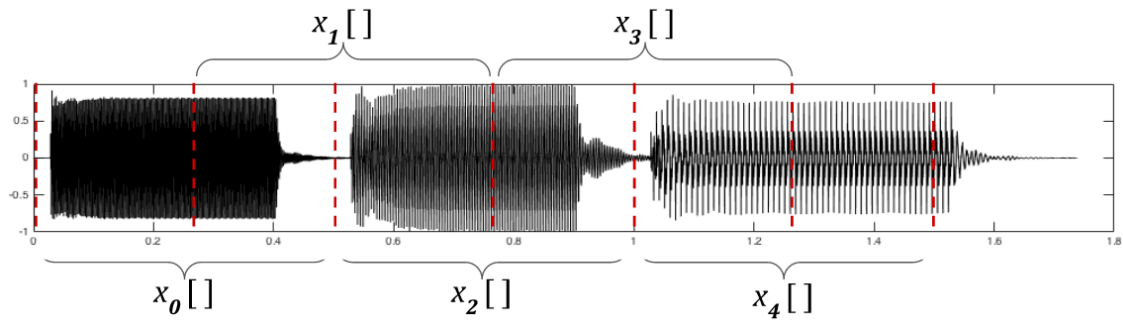


Abbildung 2.3: Windowing: Die Zerlegung eines Signals in kürzere Fenster.

Als Vorbereitungsschritt für die Transformation der Signalfenster in den Frequenzbereich wird nun jedes Fenster mit einer *Fensterfunktion* (engl. *window*) $w[\]$ multipliziert.[1, S. 69] Gleichung 2.15 definiert eine der am weitesten verbreiteten Fensterfunktionen, das *Hamming-Window*. Der Parameter M gibt die Länge des Fensters an. Abbildung 2.4 visualisiert das Hamming-Window. [39, S. 286]

$$w[\] := \begin{matrix} M-1 \\ \forall \\ n=0 \end{matrix} : w[n] = 0.54 - 0.46 \cos\left(\frac{2\pi n}{M}\right) \quad (2.15)$$

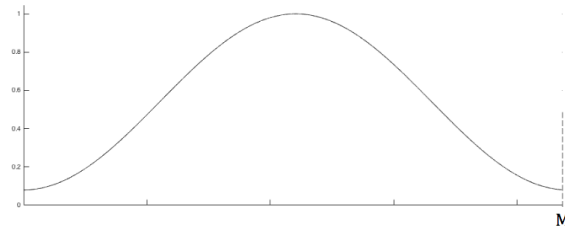


Abbildung 2.4: Das Hamming-Window

Die Gleichung 2.16 definiert die *Kurzzeit-Fourier-Transformation* (engl. *Short Time Fourier Transformation*, kurz *STFT*), implementiert mit Hilfe der DFT. Dabei wird das Signalfenster $x_i[\] = x[n + i \cdot R]$ mit der Fensterfunktion $w[\]$ multipliziert und in das *Frequenz-Fenster* $X_i[\]$ transformiert.[1, S. 69] [4] Abbildung 2.5 visualisiert die STFT des Beispiels aus Abbildung 2.3.

$$\text{STFT}_i\{x[\]\} = X_i[\] := \forall_{k=0}^{M-1} : X_i[k] = \sum_{n=0}^{M-1} x[n + i \cdot R] \cdot w[n] \cdot e^{-j2\pi k \frac{n}{N}} \quad (2.16)$$

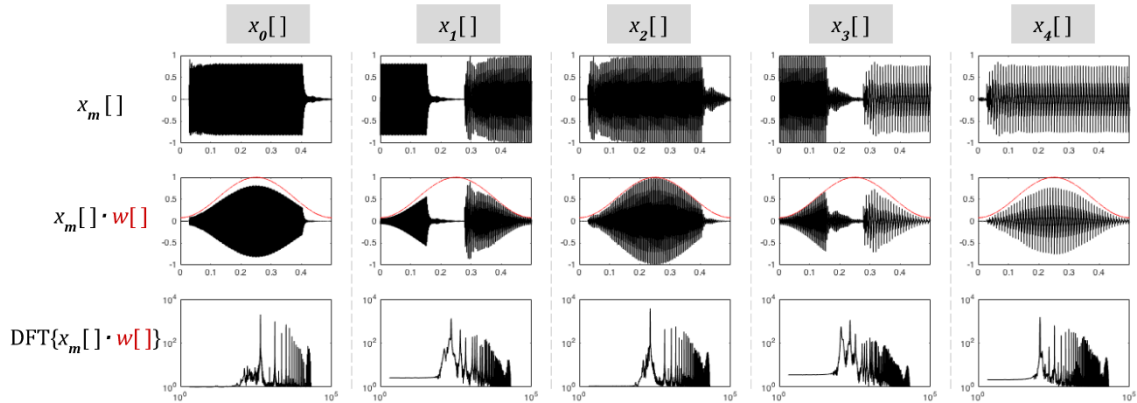


Abbildung 2.5: STFT des Beispiel-Signals aus Abbildung 2.3

2.2.5 Akustische Modellierung der menschlichen Stimme

Der menschliche Sprechapparat wird in die folgenden Komponenten unterteilt:

Schallproduktion: Die Lunge stößt Luft aus, welche die Stimmbänder passieren. Sind die Stimmbänder leicht gespannt, so wird der Luftstrom periodisch unterbrochen. Die Schwingfrequenz beträgt bei erwachsenen Männern etwa 120 Hz und bei Frauen 220 Hz. Die Frequenz kann während des Sprechens um bis zu einer Oktave variieren. Es wird so ein periodisches, akustisches Signal produziert, bezeichnet als „periodische Quelle“ (engl. „periodic Source“). Sind die Stimmbänder stark gespannt, so entstehen Turbulenzen, die sich akustisch als ein zischendes Geräusch ohne identifizierbare Tonhöhe äußern. Dieses stimmlose Signal wird bezeichnet als „Turbulenzquelle“ (engl. „turbulence Source“).

Klangformung: Das Signal der Stimmlippen passiert den Rachen, Mund- und Nasenraum, welche gemeinsam als „Vokaltrakt“ bezeichnet werden. Das Halszäpfchen bestimmt, ob der Luftstrom in den Mund- oder Nasenraum geleitet wird. Die Stellung der Artikulatoren, bestehend aus dem Kiefer, der Zunge usw. bestimmen die Beeinflussung des Klanges, der durch die Stimmbänder erzeugt wurde. Diese Beeinflussung wird als Filter angenähert. [16, S. 62] [1, S. 13] Abbildung 2.6 visualisiert diese Komponenten.

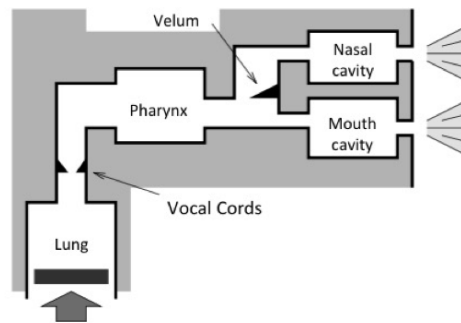


Abbildung 2.6: Schematische Übersicht über die Organe der Spracherzeugung. Lung = Lunge, Vocal Chords = Stimmbänder, Pharynx = Rachen, Velum = Halszäpfchen, Mouth Cavity = Mundraum, Nasal Cavity = Nasenraum [24]

Aus Sicht der Signalverarbeitung wird die menschliche Lautproduktion durch das sogenannte *Source-Filter-Modell* modelliert. Der durch die Stimmbänder erzeugte periodische Ton wird angenähert durch einen Impuls-Zug, welcher durch den Schlund als linearen Filter moduliert wird. Der stimmlose, nicht-periodische Ton wird durch weißes Rauschen angenähert. Der so erzeugte periodische oder nicht-periodische Ton wird als das Eingangssignal $u[]$ bezeichnet. Dieses Signal wird daraufhin an den Vokaltrakt weitergeben, welcher als lineares, zeitinvariantes Filter mit der Impulsantwort $v[]$ modelliert wird. Diese Impulsantwort ist abhängig von der Konfiguration der Organe des Vokaltraktes. Die Lippen werden als zweites lineares, zeitinvariantes Filter mit der Impulsantwort $r[]$ modelliert. $r[]$ wird auch als „radiant Model“ bezeichnet. Das tatsächliche Sprachsignal $y[]$ entsteht somit durch die Faltung des Signals $u[]$ mit den beiden linearen, zeitinvarianten Filtern nach Gleichung 2.17. Gleichung 2.18 definiert den Frequenzbereich des Ausgangssignals $Y[]$ durch die Multiplikation der Frequenzbereiche der drei Komponenten. Abbildung 2.7 visualisiert diesen Prozess schematisch. [16, S. 62 - 63] [24]

$$u[] * v[] * r[] = y[] \quad (2.17)$$

$$U[] \cdot V[] \cdot R[] = Y[] \quad (2.18)$$

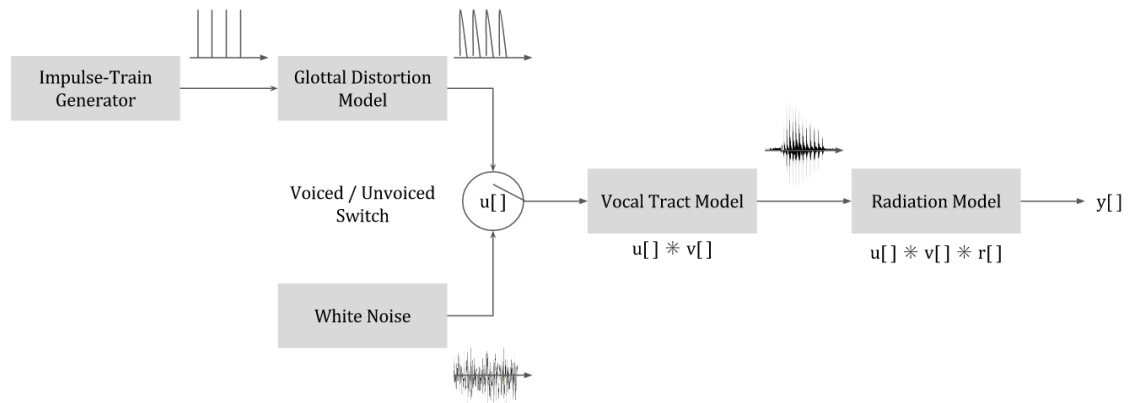


Abbildung 2.7: Schematische Übersicht über das Source-Filter-Model [13, nach Source estimation, S. 17]

Abbildung 2.8 zeigt die Zeitbereiche der stimmhaften und turbulenten Quelle im Vergleich. Wie zu sehen ist, bestimmt der zeitliche Abstand zwischen den Impulsen die Grundfrequenz der Stimme. Dieses Signal $p[]$ wird durch den Schlund als Filter $G\{\}$ gefiltert, wodurch der Zeitbereich der periodischen Quelle entsteht $G\{p[]\} = u_p[]$. Darunter ist der Zeitbereich des weißen Rauschen zu sehen. [25, Source]

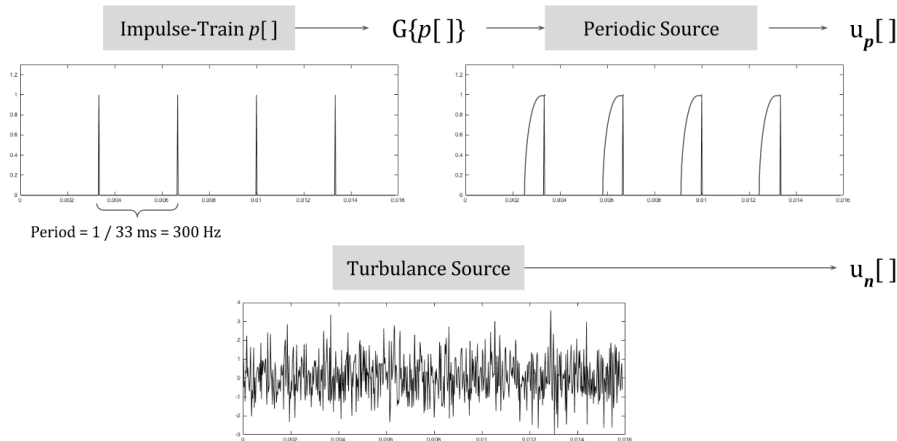


Abbildung 2.8: Zeitbereiche der periodischen und der turbulenten Quelle [25, Source]

Abbildung 2.9 zeigt die Frequenzbereiche der Komponenten des Source-Filter-Modells. Die periodische Quelle ($U[]$ links) zeichnet sich im Frequenzbereich durch gleichmäßig verteilte Spitzen aus, die mit steigender Frequenz an Amplitude verlieren. Rechts daneben ist der Frequenzbereich des weißen Rauschen zu sehen, welcher einer Zufallsverteilung entspricht. Die Frequenzantwort des Vokaltraktes $V[]$ zeichnet sich durch Resonanzfrequenzen aus, von denen in diesem Beispiel vier erkennbar sind. Die Übertragungsfunktion der Lippen $R[]$ wird als Hochpassfilter angenähert. Das Ausgangssignal $Y[] = U[] \cdot V[] \cdot R[]$ zeigt den Einfluss der Filter auf das jeweilige Eingangssignal.[13, Source estimation], [25, Vocal Tract Resonance]

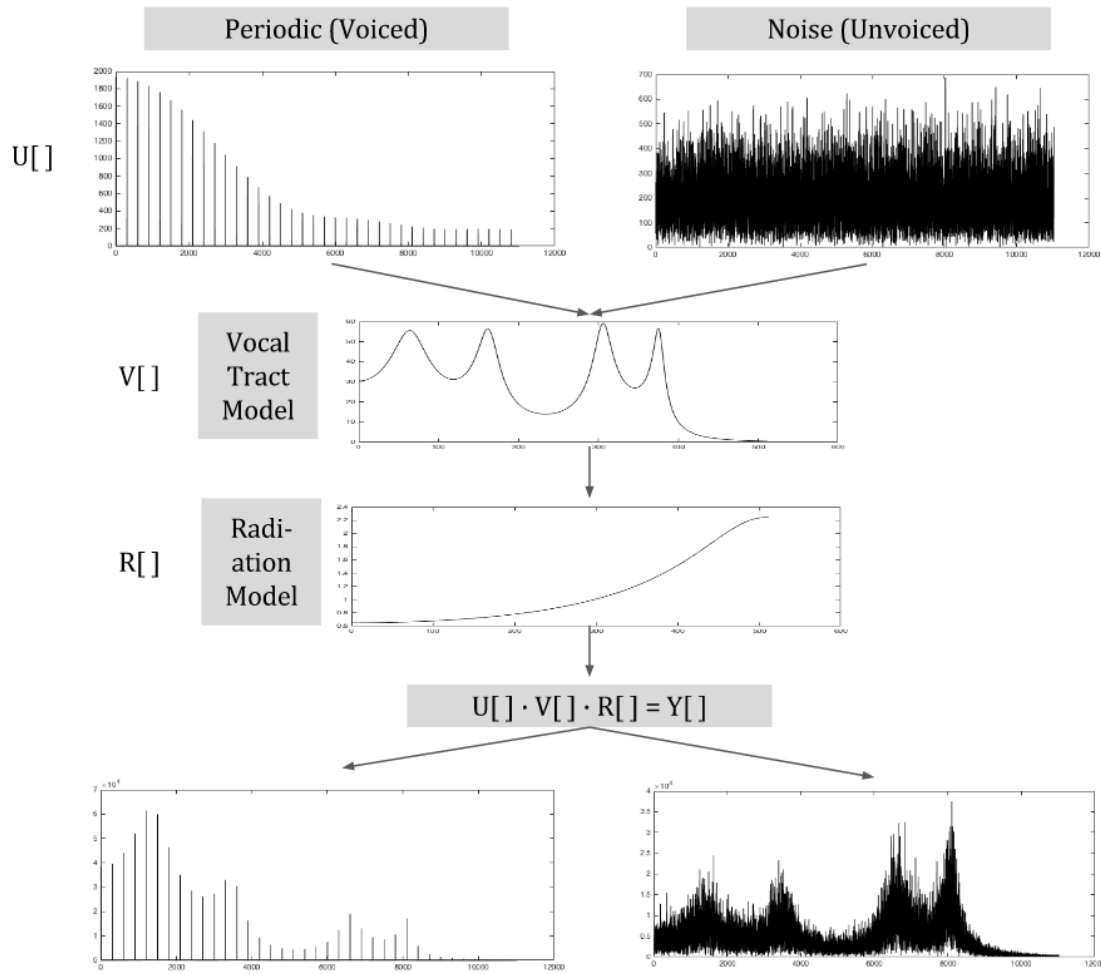


Abbildung 2.9: Betrachtung der Frequenzbereiche des Source-Filter-Model (nach: [13, Source Estimation, S. 3])

Abbildung 2.10 zeigt schematisch das Spektrum eines stimmhaften Sprachsignals. Sowohl die Grundfrequenz als auch die harmonischen Obertonwellen sind rein visuell als „viele, kurze Signalspitzen“ im Spektrum erkennbar. Der kleinste gemeinsame Teiler der Frequenzen dieser Signalspitzen entspricht der Grundfrequenz f_0 des Stimmsignals, in diesem Beispiel 250.7 Hz. Die Grundfrequenz ist ebenfalls an der Signalspitze mit der tiefsten Frequenz ablesbar. Die harmonischen Obertöne entsprechen der doppelten, dreifachen, ... Frequenz dieser Grundfrequenz, das heißt $2 \cdot f_0, 3 \cdot f_0, \dots$ und werden bezeichnet mit H_1, H_2, \dots . Die Grundfrequenz ist *nicht zwingend* die Spitze der höchsten Amplitude. Durch den Einfluss des Vokaltraktes als Filter können harmonische Oberwellen eine höhere Amplitude als die Grundfrequenz erhalten. Auf Basis des Spektrums lässt sich somit rein visuell ein stimmhaftes Signal von einem nicht stimmhaften (Rausch-)Signal unterscheiden, in dem das Spektrum nach dem Vorhandensein dieser regelmäßigen Signalspitzen überprüft wird (vergleiche mit Abbildung 2.9).[1, S. 52 - 53]

Abbildung 2.11 verdeutlicht, wie der als lineares, zeitinvariantes Filter modellierte Vokaltrakt durch Formanten bestimmt wird. Diese Formanten spielen vor allem bei der Beschreibung von Vokalen eine Rolle. Formanten sind lokale Maxima im Spektrum der Transferfunktion, die dadurch erzeugt werden, dass der Vokaltrakt Resonanzen erzeugt. Die

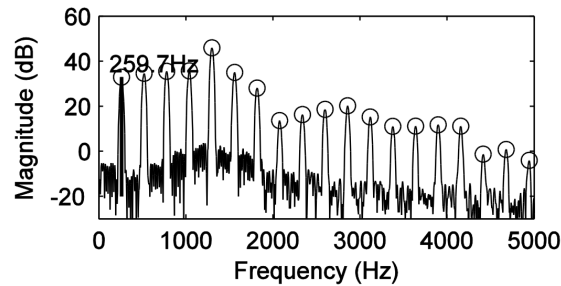


Abbildung 2.10: Grundfrequenz und harmonische Obertöne eines periodischen Sprachsignals.

Formanten werden von links nach rechts durchnummeriert, von F_1, \dots, F_n . Jeder Formant wird durch seine Mittenfrequenz, seine Bandbreite und seine Amplitude beschrieben. Das wichtigste Merkmal ist jedoch die Mittenfrequenz, da sie vom menschlichen Gehör am stärksten zur Identifikation und Unterscheidung der Vokale genutzt wird. Mit steigender Frequenz nimmt die Amplitude der Formanten ab, der dominanteste Formant ist somit immer der erste. Daher werden meist nur die ersten 2 oder 3 Formanten zur Beschreibung eines Vokals angegeben, auch, wenn theoretisch weitaus mehr vom Vokaltrakt erzeugt werden. Für verschiedene Sprachen sind allerlei Tabellen zu finden, welche die Formantenfrequenzen der Vokale auflisten.[1, S. 19]

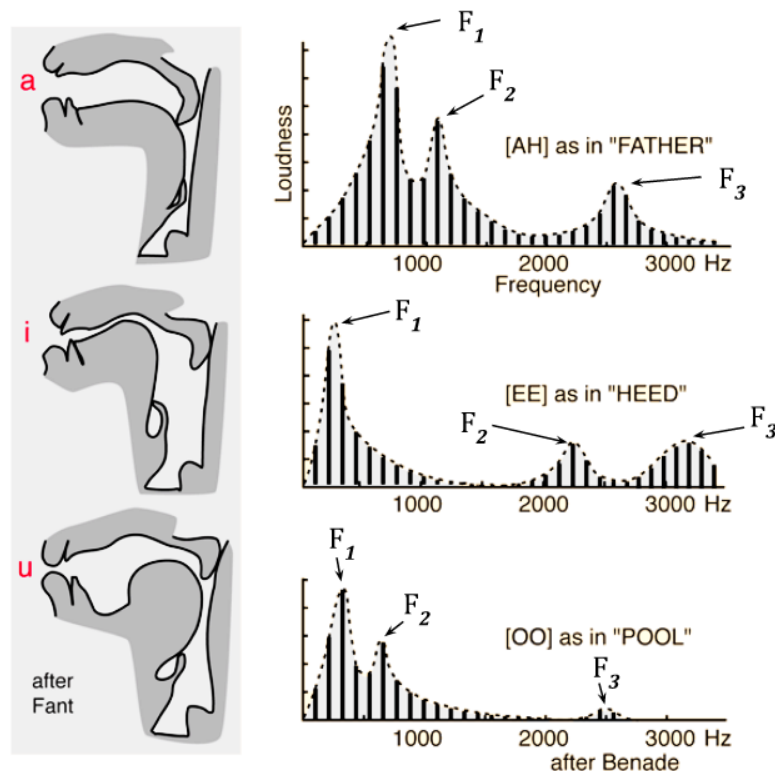


Abbildung 2.11: Formanten im Sprach-Signal (nach: [2])

Beim Sprechen befinden sich sowohl das Signal der Stimmbänder als auch das Filter des Vokaltraktes und der Lippen in ständiger Veränderung. Ein stimmhaftes Sprachsignal

gilt nur über kurze Zeitbereiche weniger Millisekunden als periodisch, und selbst in diesen kurzen Zeitbereichen ist die Stimme nicht perfekt, sondern nur annähernd periodisch. Da die Informationen der Sprache vor allem im Frequenzbereich codiert sind, wird die in Kapitel 2.2.4 vorgestellte Kurzzeit-Fourier-Transformation zur Analyse von Sprache eingesetzt. Die Visualisierung der STFT wird als *Spektrogramm* bezeichnet. Dabei werden auf der x-Achse die Zeitpunkte der Fenster und auf der y-Achse die Frequenz dargestellt. Die Frequenzfenster werden „auf die Seite gelegt“, damit ihr zeitlicher Verlauf übersichtlich betrachtet werden kann. Die Amplitude der entsprechenden Frequenzen wird farblich oder durch Helligkeiten codiert, abhängig von der konkreten Implementierung des Spektrogramms. Je länger das Zeitfenster der STFT, desto höher ist die Auflösung bezüglich des Frequenzbereiches und desto niedriger die Auflösung bezüglich der Zeitbereiche. Je kürzer die Zeitfenster der STFT, desto höher ist die Auflösung bezüglich des Zeitbereiches, und desto niedriger die Auflösung des Frequenzbereiches.[1, S. 45 - 50] [25, Acoustic Representations of Speech].

Abbildung 2.12 zeigt ein Beispiel für zwei Spektrogramme mit unterschiedlichen Fensterlängen der STFT, angewandt auf einer 9 Sekunden langen Aufnahme eines weinenden Babys. Es ist zu erkennen, wie bei der geringeren Fensterlänge der zeitliche Verlauf besser erkennbar, jedoch die einzelnen harmonischen Obertöne weniger gut voneinander unterscheidbar sind. Bei der längeren Fensterlänge sind die Formanten leichter zu unterscheiden, der Beginn und das Ende der Lautäußerungen jedoch schwerer zu lokalisieren.

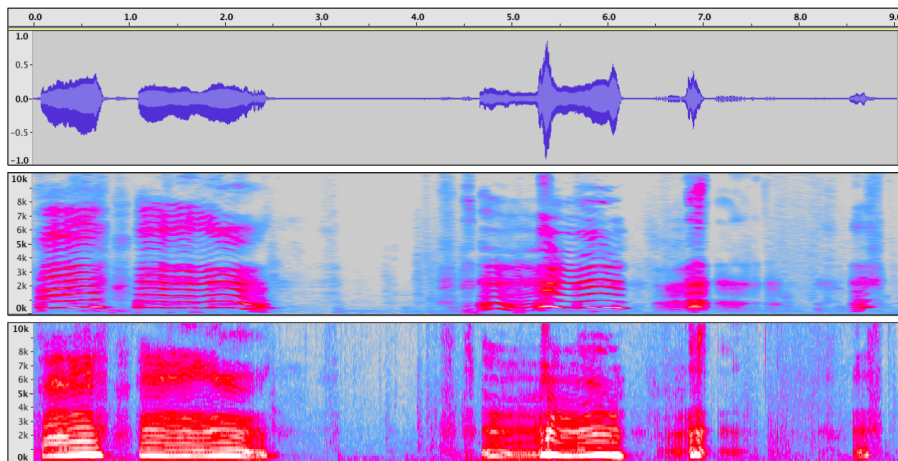


Abbildung 2.12: Spektrogramm einer Audioaufnahme eines Babys. Rot $\hat{=}$ hohen Amplituden, Blau $\hat{=}$ niedrigen Amplituden. Oben: Zeitbereich. Mitte: Spektrogramm mit einer Fensterlänge von 185 ms(8192-Sample DFT). Unten: Spektrogramm mit einer Fensterlänge von 5 ms(265-Sample DFT)

2.3 Schreiforschung

Das Wissenschaftsgebiet, welches sich mit der Analyse und Interpretation von Lautäußerungen Neugeborener auseinandersetzt, wird als „Schreiforschung“ bezeichnet. Die bis heute wohl prominenteste Forschungsgruppe dieses Wissenschaftsgebietes ist die in Kapitel 2.1.2 erwähnte „Scandinavian Cry-Group“[21], welche zwischen 1960 und 1990 die Laute von Babys systematisch erforscht haben. Das wichtigste Werkzeug zur Analyse der Lautäußerungen war das Spektrogramm, welches damals auf analogen Technologien basierte. Das Ziel der frühen Schreiforschung war es, mit Hilfe des Spektrogramms Muster zur Unterscheidung eines

abnormalem Weinen von einem normalen Weinen zu finden, um beispielsweise Krankheiten erkennen zu können.[36, S. 142]

Teil der Scandinavian Cry-Group waren H. Golub und M. Corwin, die in der Veröffentlichung „A Physioacoustic Model of the Infant Cry“[16] ein Vokabular zur Beschreibung typischer, im Spektrogramm erkennbarer Muster festgelegt haben. Da das Vokabular bis heute Einsatz findet, wird an dieser Stelle eine Übersicht über die wichtigsten Begriffe gegeben. Weiterhin werden Begriffe eingeführt, die von Zeskind et al. in „Rhythmic organization of the Sound of Infant Cry “ veröffentlicht wurden.[34]

2.3.1 Physio-Akustische Modellierung des Weinens

Das Weinen von Babys lässt sich im allgemeinen als das „rhythmische Wiederholen eines beim Ausatmen erzeugen Geräusches, einer kurzen Pause, einem Einatmungsgeräusch, einer zweiten Pause, und dem erneuten Beginn des Ausatemungsgeräusches“beschreiben. [41].

Die folgenden Begriffe werden in Abbildung 2.13 veranschaulicht.

- **Expiration (Ausatmung):** Der Klang, der bei einem einzelnen, ununterbrochenen Ausatmen mit Aktivierung der Stimmbänder durch das Baby erzeugt wird. [34]. Der von Golub et al. [16, S. 61] verwendete Begriff **Cry-Unit** wird in dieser Arbeit synonym verwendet. Umgangssprachlich ist handelt es sich um einen einzelnen, ununterbrochenen *Schrei*.
- **Inspiration (Einatmung):** Der Klang, der beim Einatmen durch das Baby erzeugt wird.
- **Burst:** Die Einheit einer Ausatmung und der darauf folgenden Einatmung. Das heisst, dass die zeitliche Dauer eines Bursts sowohl die Ausatmung, die Einatmung als auch die beiden Pausen zwischen diesen Geräuschen umfasst.[34]¹
- **Cry:** Die gesamte klangliche Antwort zu einem spezifischen Stimulus. Eine Gruppe mehrerer Cry-Units.[16, S. 61] In dieser Arbeit wird ein *Cry* auch als **Cry-Segment** bezeichnet, um Verwechslungen zu vermeiden.

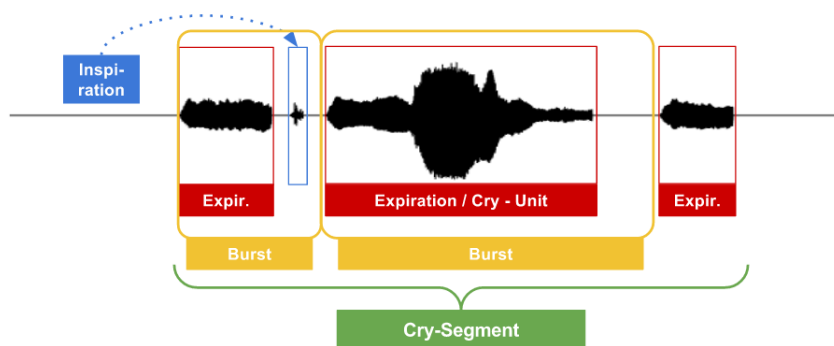


Abbildung 2.13: Veranschaulichung des Grundvokabulars

¹Praktisch ergibt sich das Problem, dass vor allem bei stärkerem Hintergrundrauschen die Einatmung häufig weder hörbar noch auf dem Spektrogramm erkennbar ist. Daher wird die Zeitdauer eines Bursts von Beginn einer Ausatmung bis zum Beginn der darauf folgenden Ausatmung definiert und somit allein von den Ausatemungsgeräuschen auf die Bursts geschlossen. Implizit wird somit eine Einatmung zwischen zwei Ausatmungen angenommen.

Cry-Units werden von H. Golub und M. Corwin in eine der drei folgenden Kategorien eingeordnet, bezeichnet als *Cry-Types*: [16, S. 61 - 62]

- **Phonation** beschreibt eine Cry-Unit mit einer „vollen Vibration der Stimmbänder“ und einer Grundfrequenz zwischen 250 und 700 Hz. Entspricht umgangssprachlich einem Weinen mit einem „klaren, hörbaren Ton“.
- **Hyper-Phonation** beschreibt eine Cry-Unit mit einer „falsetto-artigen Vibration der Stimmbänder“ mit einer Grundfrequenz zwischen 1000 und 2000 Hz. Entspricht umgangssprachlich einem Weinen mit einem „sehr hohen, aber klar hörbaren Ton“.
- **Dysphonation** beschreibt eine Cry-Unit ohne klar feststellbare Tonhöhe, produziert durch Turbulenzen an den Stimmbändern. Entspricht umgangssprachlich dem „Brüllen oder Krächzen“.

Die folgenden weiteren Eigenschaften können für einzelne Cry-Units extrahiert werden:

- **Duration:** Die zeitliche Dauer der Cry-Unit.
- **Duration of Inspiration:** Die zeitliche Dauer der Pause zwischen zwei Cry-Units.
- **Grundfrequenz:** Für eine Cry-Unit kann die durchschnittliche, die höchste und die niedrigste Grundfrequenz sowie die Varianz festgestellt werden.
- **Frequenz der Formanten:** Wie bei der Grundfrequenz kann der Durchschnitt, das Maximum, Minimum usw. für eine Cry-Unit berechnet werden.
- **Ratio2:** Verhältnis zwischen den Energien der Frequenzen unterhalb von 2000 Hz zu den Frequenzen oberhalb von 2000 Hz
- **Cry-Mode Changes:** Häufigkeit des Wechsels des Cry-Modes innerhalb einer Cry-Unit.
- **Amplitude:** Die Lautstärke der Cry-Unit, gemessen in Dezibel. [22, S. 85] [10, S. 156]

H. Golub und M. Corwin haben weiterhin eine Reihe von Eigenschaften vorgestellt, die das zeitliche Verhalten der Grundfrequenz und der harmonischen Obertöne innerhalb einer Cry-Unit beschreiben. [16, S. 73] Einige dieser Eigenschaften werden in Abbildung 2.14 in einem schematischen Spektrogramm dargestellt. Die schwarzen Linien zeigen den Verlauf der Grundfrequenz und der Formanten.

- **Pitch of Shift:** Grundfrequenz nach einem schnellen Anstieg zu Beginn der Cry-Unit
- **Glide:** Kurzes, starkes ansteigen der Grundfrequenz
- **Glottal Roll:** Dysphonation, die häufig am Ende einer Cry-Unit nach einem Abfall der Grundfrequenz beobachtet wird.
- **Vibrato:** Mehr als vier starke Schwankungen der Grundfrequenz innerhalb einer Cry-Unit.
- **Melody-Type:** einer Cry-Unit. Meist: fallend, steigend/fallend, steigend, fallend/-steigend, flach.
- **Continuity:** Verhältnis zwischen stimmhaften und nicht-stimmhaften Bereichen der Cry-Unit

- **Double Harmonic Break:** Das Aufkommen einer zweiten Serie von harmonischen Obertönen zwischen den eigentlichen harmonischen Obertönen der Cry-Unit.
- **Biphonation:** Das Aufkommen einer zweiten Grundfrequenz mit eigenen harmonischen Obertönen zusätzlich zu der eigentlichen Grundfrequenz.
- **Noise Concentration:** Starke Energiespitzen zwischen 2000 und 2300 Hz.
- **Furcation:** Plötzliches Aufteilen der Grundfrequenz und harmonischen Obertöne in mehrere, schwächere Obertöne.

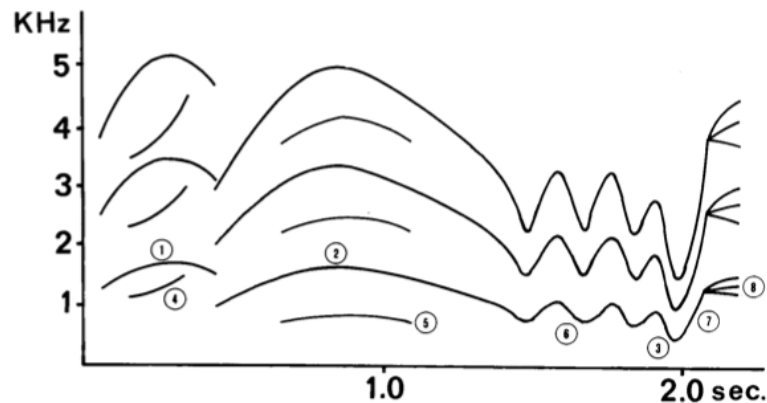


Abbildung 2.14: (1) Pitch of Shift (2) Maximale Grundfrequenz (3) Minimum der Grundfrequenz (4) Biphonation (5) Double Harmonic Break (6) Vibrato (7) Glide (8) Furcation [36, S. 142]

Die folgenden Eigenschaften werden in Bezug auf das gesamte Cry-Segment, oder zumindest auf eine Menge aufeinander folgender Cry-Units berechnet:

- **Cry Latence:** Zeit zwischen Stimulus, wie zum Beispiel einem Nadelstich, und der ersten Cry-Unit.
- **Utterances:** Anzahl der Cry-Units im Segment.
- **Short Utterances:** Anzahl stimmloser Cry-Units im Segment.
- und statistische Auswertungen bezüglich aller oben genannten Features, die sich auf eine Cry-Unit beziehen, wie beispielsweise der Durchschnitt aller Tonhöhen, Anzahl des Vorkommens bestimmter Melodiekonturen, Varianz der Länge der Cry-Units etc.[22, S. 85]

Einige Krankheiten wurden in Zusammenhang mit dem vermehrten Vorkommen bestimmter Eigenschaften bei kindlichen Lautäußerungen gebracht. So wurde eine Korrelation zwischen dem Anstieg der durchschnittlichen Grundfrequenz, häufiger Biphonation und geringer Duration in Zusammenhang mit Gehirnschäden beobachtet. Tendenziell niedrige Grundfrequenzen zeigen eine Korrelation mit Trisomie 13, 18 und 21.[22, S. 85]

2.3.2 Diskussion

Bis heute bleibt die Analyse von kindlichen Lautäußerungen weitestgehend unstandardisiert: [36, S. 142]

- Es gibt keine komplette Liste, welche einen Überblick über alle berechenbaren Eigenschaften für Cry-Units oder Segmente gibt. Viele Veröffentlichungen beziehen sich auf die Eigenschaften, die von H. Golub und M. Corwin vorgestellt wurden, und erweitern diese Liste mit eigenen Vorschlägen.
- Es gibt keine Einigung darüber, welche der Eigenschaften die wichtigsten sind. Beispielsweise konzentrierten sich H. Golub und M. Corwin [16] vermehrt auf die Erkennung von Mustern im Melodieverlauf, Zeskind et al. auf zeitliche Eigenschaften.[34]. Die Eigenschaft, die am häufigsten mit Schmerz, Krankheiten und sonstigen Abnormalitäten in Verbindung gebracht wird, ist eine abnormal hohe oder niedrige Tonhöhe. Bei einigen Features, die von H. Golub und M. Corwin verwendet wurden, ist nicht einmal gesichert, ob es sich nicht doch um technische Artefakte der damals verwendeten Analogtechnik handelt. [22, S. 84 - 85]
- Selbst, wenn in verschiedenen Studien die selbe Eigenschaft verwendet wird, wie zum Beispiel die durchschnittliche Tonhöhe, ist nicht standardisiert, wie dieses zu berechnen ist. Mit „durchschnittliche Tonhöhe des Segmentes“ kann gemeint sein: 1.) die durchschnittliche Tonhöhe, errechnet aus den durchschnittlichen Tonhöhen der der Cry-Units. 2.) Die durchschnittliche Tonhöhe aller festgestellten Tonhöhen. 3.) Die durchschnittliche Tonhöhe nur von Ausatemungslauten usw.
- Zusammenhänge, die zwischen bestimmten Eigenschaften des Weinens und bestimmten Krankheitsbildern festgestellt wurden, haben häufig eine hohe Spezifität, aber niedrige Sensitivität. So wurde zum Beispiel festgestellt, dass Kinder, die am plötzlichen Kindstod sterben, fast immer eine Erhöhung der Frequenz des ersten Formanten in Verbindung mit häufigen Cry-Mode-Changes zeigen. Viele Babys, die nicht am plötzlichen Kindstod sterben, zeigen jedoch die selben Merkmale.[22, S. 85]
- H. Golub und M. Corwin behaupten, bereits in den achtziger Jahren ein System zur computergestützten und voll automatisierten Analyse von Cry-Segmenten implementiert zu haben. Das System nimmt 1.) eine Audioaufnahme, gespeichert auf einer Kasette an, 2.) berechnet Formanten, Grundfrequenz und Amplitude gegen die Zeit, 3.) samplt die Grundfrequenz-Kontur, 4.) berechnet insgesamt 88 akkumulierte Features für das gesamte Segment und 5.) zieht Schlussfolgerungen aus den 88 Features, wie zum Beispiel die Diagnose einer bestimmten Krankheit.[16, S. 75 - 76] Abseits der kurzen Erwähnung der Existenz dieser „Mutter aller automatisierten Analysesysteme für das Weinen von Babys“ konnte der Autor dieser Arbeit keine Implementierungsdetails oder sonstige genaueren Ausführungen über das System finden, welche für diese Arbeit von höchstem Interesse gewesen wären.

2.4 Klassifizierung und Regression

Klassifizierung und Regression sind Teilgebiete des Wissenschaftsgebietes des *Überwachten Lernens*, einem Teilgebiet des Wissenschaftsgebietes des *maschinellen Lernens*. Das Ziel beim Überwachten Lernen ist es, einen *Prädiktor*, auch bezeichnet als *Modell*, zu entwerfen, der aus den Eigenschaften einer Instanz dessen Kategorie oder Wert ableiten kann. Im Zusammenhang mit der Schreiforschung könnte eine Instanz eine Baby sein, dessen Eigenschaften 1.) die durchschnittliche Tonhöhe beim Weinen und 2.) die Augenfarbe ist. Der Prädiktor hat nun die Aufgabe, aus diesen beiden Eigenschaften eine Klasse abzuleiten, wie zum Beispiel das Geschlecht des Babys, oder einen Wert, wie beispielsweise das Alter. Das

Lernen basiert dabei auf dem Generalisieren einer Liste von Beispielen, die der Algorithmus zur Verfügung gestellt bekommt. In diesem Zusammenhang wäre dies eine Liste an Babys, bei der für jede Instanz das Geschlecht oder das Alter bereits bekannt ist. Der Algorithmus versucht nun, diese Beispiele soweit zu Verallgemeinern, dass er für neue, bisher unbekannte Babys die Klasse oder den Wert korrekt voraussagen kann.[26, S. 6 - 7]

Eine Instanz x ist ein Vektor $x = (f_1 \in F_1, \dots, f_n \in F_n)$. F_i wird in diesem Zusammenhang als *Eigenschaft*, *Feature* oder *Attribut* bezeichnet. In Bezug auf das eben genannte Beispiel wäre das erste Feature $F_1 = \text{durchschnittliche Tonhöhe}$ und das zweite Feature $F_2 = \text{Augenfarbe}$. Eine Instanz wäre in diesem Fall ein Tupel mit zwei beliebigen Werten dieser Attribute, wie zum Beispiel $x = (300 \text{ Hz}, \text{blau})$. Features, die einen kontinuierlichen Wertebereich mit einem quantitativen Charakter haben, wie zum Beispiel das Gewicht, werden als *kontinuierliche* Features bezeichnet. Features, die einen diskreten Wertebereich mit einem qualitativen Charakter haben, wie zum Beispiel die Augenfarbe, werden als *diskrete* Features bezeichnet. Die Menge aller möglichen Kombination der Features $F_1 \times \dots \times F_n$ wird als *Feature-Raum* bezeichnet. Der Trainings-Datensatz D_{training} besteht aus einer Liste an Instanzen, wobei für jede Instanz die Kategorie oder der Wert, gemeinsam bezeichnet als *Output* oder *Target* $y \in Y$, bekannt ist. Ein Tupel aus einer Instanz zu einem Output wird als *Example* $e = (x, y)$ bezeichnet. Y bezeichnet die Menge aller möglichen Outputs des Problems. Das heißt, $D_{\text{training}} = ((x_1, y_1), \dots, (x_N, y_N))$. Der Prädiktor P ist nun eine Funktion, die von einer Instanz auf den Output abbildet, also $P : X \mapsto Y$. Die Fehlerfunktion E berechnet, wie häufig sich der Prädiktor bei der Bestimmung der Targets eines Test-Datensatzes D_{test} irrt. Der Test- und der Trainings-Datensatz können die selben Instanzen, teilweise die selben oder gar keine gemeinsamen Instanzen beinhalten.[26, S. 6 - 7, 18 - 19] [7, S. 8 - 9]

Bei der **Klassifizierung** wird eine Target als *Klasse* bezeichnet. Die Menge aller möglichen Klassen eines bestimmten Problems $Y = \{y_1, \dots, y_n\}$ ist dabei diskret und hat einen *qualitativen* Charakter. Das heißt, dass keine Klasse „besser“ oder „höher“ ist als eine andere. Ein Beispiel für ein Klassifizierungsproblem wäre die also die Ableitung des Geschlechtes für eine Instanz, also $Y = \{m, w\}$. Der Prädiktor wird in diesem Fall als Klassifikator C bezeichnet.² [11, S. 28, 127]

Bei der *Regression* ist die Menge der möglichen Targets eines bestimmten Problems *kontinuierlich* und hat einen „quantitativen Charakter. Das heißt, es kann eine interne Ordnung in der Menge der Outputs festgelegt werden. Ein Beispiel für ein Regressions-Problem wäre die also die Ableitung des Alters des Babys, also $Y = \{0, \dots, 3\}$. Der Prädiktor wird in diesem Fall auch als *Regressor* R bezeichnet.[7, S. 24] [26, S. 8] [11, S. 28]

Es gibt eine Vielzahl an Algorithmen zum Finden des Klassifikators oder Regressors. Welcher Algorithmus der „beste“ ist, das heißt für einen Test-Datensatz eine möglichst hohe *Genauigkeit* oder einen möglichst geringen *Klassifikationsfehler* erzeugt, ist abhängig von der konkreten Problemstellung. Auf die Bestimmung der Genauigkeit wird weiter in Kapitel 2.4.2 eingegangen. Ein Algorithmus, der in dieser Arbeit zur Klassifizierung eingesetzt wird, ist der *ID3*- und der *C4.5*-Algorithmus, welcher genauer in Kapitel 2.4.1 beschrieben wird.

²In einigen Quellen werden die Begriffe *Klassifizierung* und *Klassifikation* inkonsistent verwendet. *Klassifizierung* ist ein Prozess, dessen Ergebnis die *Klassifikation* ist. Daher wird von einem *Klassifizierungs-Algorithmus* gesprochen, da sich der Algorithmus auf den Prozess des Klassifizierens konzentriert, aber vom *Klassifikationsfehler*, da der Fehler des Ergebnisses der Klassifizierung bestimmt wird. In einigen Fällen ist es unerheblich, ob das Ergebnis oder der Prozess im Fokus des Diskussionsgegenstandes steht. Dann können *Klassifikation* und *Klassifizierung* synonym gebraucht werden.

2.4.1 ID3 und C4.5

Der ID3-Algorithmus zählt zu den sogenannten Entscheidungsbäumen, da der durch den Algorithmus entworfene Klassifikator die Form eines Entscheidungsbaumes annimmt. Die Voraussetzung ist, dass alle Features diskret und nicht kontinuierlich sind. Tabelle 2.3 zeigt einen Beispieldatensatz zur Erläuterung des Algorithmus. Die Instanzen sind Babys, die Features die „Häufigkeit des Weinens“ und die „Lautstärke des Weinens“, und die beiden Klassen { Ja, Nein } geben an, ob das Kind an chronischem Schmerz leidet

Tabelle 2.3: Beispieldatensatz D für die Klassifikation mit ID3

x_i	$f_1 \in \text{Häufigkeit}$	$f_2 \in \text{Lautstärke}$	$y_i = \text{chron. Schmerz}$
x_1	oft	laut	Ja
x_2	selten	laut	Ja
x_3	normal	leise	Nein
x_4	selten	leise	Nein
x_5	normal	laut	Ja
x_6	oft	leise	Ja

Abbildung 2.15 zeigt einen Klassifikator, den der ID-3 Algorithmus für diesen Datensatz erzeugt. Es handelt sich um einen Entscheidungsbaum. In jedem Knoten steht ein Feature, welches einen Ast für jeden möglichen Wert dieses Features bildet. In den Blättern stehen die Klassen.[26, S. 134]

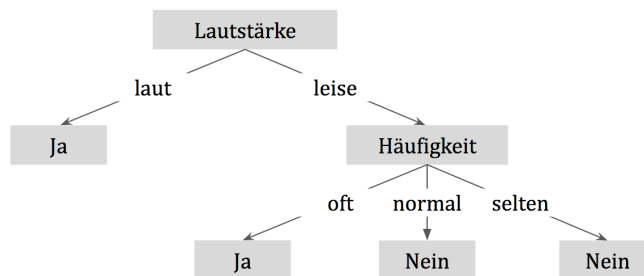


Abbildung 2.15: Entscheidungsbaum, der durch den ID3-Algorithmus für den Datensatz aus Beispiel 2.3 erzeugt wurde.

Der Entscheidungsbaum lässt sich in eine Reihe von `if ... then ...`-Regeln transformieren. Jeder Weg von der Wurzel bis zu einem Blatt ergibt eine Entscheidungsregel, bei der die Feature-Werte der entsprechenden Kanten konjunktiv Verknüpft werden und die Klasse implizieren. Die Entscheidungsregeln für den Baum aus Abbildung 2.15 sind: [26, S. 134]

- `if Tageszeit = Tag then Spaß = Ja`
- `if Tageszeit = Nacht and Temperatur = warm then Spaß = Ja`
- ...

Der Entscheidungsbaum wird beim ID3 Algorithmus nach folgenden Muster erstellt: Die Konstruktion wird Top-Down vollzogen, dass heisst beginnend bei der Wurzel bis zu den Blättern. In jedem Knoten wird ein Feature in alle seine möglichen Attribute aufgespalten.

Um an der Wurzel zu entscheiden, welches Feature zuerst aufgespalten werden soll, wird jedes Feature einem statistischen Test unterzogen, um festzustellen, wie „gut“ dieses Feature zur Klassifikation der Trainings-Daten beiträgt. Das „beste“ Attribut wird ausgewählt und als Wurzel festgelegt. Nun wird ein Kind für jeden möglichen Wert des Features gebildet. Der Datensatz des Elternknotens wird in disjunkte Teilmengen aufteilt, wobei jedes Kind die Untermenge mit denjenigen Instanzen erhält, die den jeweiligen Feature-Wert besitzen. Daraufhin beginnt für jedes Kind der Prozess des Auswählens des „besten“ Attributes von vorn. Ein Kind wird dann zu einem Blatt, wenn seine Teilmenge an Daten nur noch aus Instanzen einer Klasse besteht und somit kein weiteres Aufteilen notwendig ist.[27, S. 55]

Zur Quantifizierung der Information wird die Entropie nach Formel 2.19 als Hilfsmittel definiert. p_i ist die Wahrscheinlichkeit, dass in einem Datensatz D ein Example mit der Klasse $i \in Y$ angetroffen wird. Die Entropie quantifiziert die *Unreinheit des Datensatzes*. Ein Datensatz, dessen Instanzen alle der selben Klasse angehören, hat die Entropie 0. Ist die *Unreinheit des Datensatzes* hingegen maximal, das heißt, dass der Datensatz exakt gleich viele Instanzen jeder Klasse beinhaltet, ist die Entropie 1. [26, S. 135]

$$H(p) = - \sum_{i \in Y} p_i \cdot \log_2 p_i \quad (2.19)$$

Es ist das Attribut in einem Knoten zu wählen, welches den höchsten *Informationsgewinn* gewährleistet, das heißt, zu einer bestmöglichen *Reinheit* in den Kindsknoten der alleinigen Unterteilung des Datensatzes auf Basis dieses Attributs führt. Der Informationsgewinn eines Features F für den Datensatz D wird nach Formel 2.20 definiert. f sind alle möglichen Werte dieses Features. $|D|$ beschreibt die Anzahl an Instanzen des Datensatzes. D_f ist die Untermenge an Instanzen, die für das Feature F den Wert f besitzen.[26, S. 136 - 137]

$$\text{Gain}(D, F) = H(D) - \sum_{f \in F} \frac{|D_f|}{|D|} H(D_f) \quad (2.20)$$

Die Erstellung eines Entscheidungsbaumes mit Hilfe des ID3-Algorithmus wird folgendermaßen als Pseudocode definiert. Der Input des Algorithmus ist der Datensatz D und der Feature-Raum F_{all} , der Output ist der Entscheidungsbaum.[26, S. 139]

ID3(D, F_{all})

- **Wenn** alle Examples $e \in D$ das selbe Label haben:
 - **return** eine Blatt mit diesem Label
- **Sonst: Wenn**
 - **return** ein Blatt mit dem häufigsten Label in dem Datensatz
- **Sonst:**
 - Wähle ein Feature \hat{F} als den nächsten Knoten, dass den Informationsgewinn für den Datensatz D nach Formel 2.20 maximiert.
 - Füge einen Ast für jeden möglichen Wert $f \in \hat{F}$ von dem Knoten hinzu.
 - Für jeden Ast:

- * Berechne D_f , in dem \hat{F} von der Liste der Features entfernt wird.
 - * Rufe **ID3**($D_f, F_{all}/\hat{F}$) rekursiv auf.
-

Der ID3-Algorithmus hat folgende **Nachteile**

- Der Algorithmus akzeptiert keine kontinuierlichen Features.[27, S. 72]
- Der Algorithmus neigt zu *Overfitting*. Overfitting bedeutet, dass der Klassifikator C zwar einen möglichst geringen Klassifikationsfehler in Bezug auf den Trainings-Datensatz erzeugt, es jedoch einen anderen Klassifikator C' gibt, welcher für den Trainings-Datensatz einen höheren Fehler erzeugt, jedoch einen geringeren Fehler als C in Bezug auf *alle möglichen Instanzen dieses Typs* erzeugt. Anders formuliert bedeutet Overfitting, dass der Klassifikator den Trainings-Datensatz „auswendig gelernt hat“ und nicht genügend generalisiert, um auf im Training nicht enthaltene Instanzen angewandt werden zu können. Overfitting im Zusammenhang mit dem ID-3 Algorithmus wird durch *Rauschen im Trainings-Datensatz* bedingt.
- Der Algorithmus bevorzugt greedy Attribute, die zum Zeitpunkt der Berechnung den höchsten Informationsgewinn gewährleisten. Dabei besteht die Gefahr, dass der Algorithmus in ein lokales Maximum läuft.[27, S. 66 - 70]

Der $C4.5$ -Algorithmus erweitert den ID3, um dessen Nachteile auszumerzen, das heißt die Möglichkeit der Verwendung kontinuierlicher Attribute sowie Lösungsansätze für das Overfitting.

Bei einem kontinuierlichen Attribut wird beim $C4.5$ -Algorithmus im Gegensatz zu einem diskreten Attribut *nicht* ein Ast für jeden möglichen Wert gebildet, sondern genau zwei Äste. Es wird also ein Grenzwert für das Feature festgelegt, bei dessen Unterschreitung der linke, und bei dessen Überschreitung der rechte Ast gewählt wird (oder, je nach Implementiert, umgedreht). Das Vorgehen zum finden eines solchen Grenzwertes ist wie folgt:

1. Ordne alle Examples nach ihrem jeweiligen Wertes des kontinuierlichen Feature F_c , für das der Grenzwert gesucht wird.
2. Identifiziere benachbarter Examples mit unterschiedlichen Klassen. Die Feature-Werte dieser Examples sind mögliche Kandidaten für einen Grenzwert.
3. Berechne den Informationsgewinn bei Setzung des Grenzwertes auf jeden gefundenen Kandidaten.
4. Wähle denjenigen Grenzwert, der den höchsten Informationsgewinn bringt. [27, S. 73]

Abbildung 2.16 visualisiert einen Knoten mit einer kontinuierlichen Variable F_c , der nach einem Grenzwert in zwei Äste aufgespalten wird.

Das als Overfitting beschriebene Problem lässt sich vermeiden, in dem die Tiefe des Entscheidungsbaumes reduziert wird. Diese Begrenzung wird als *Beschneiden* oder *Pruning* bezeichnet. Es gibt grundlegend zwei verschiedene Ansätze:

Pre-Pruning: Ab der Überschreitung einer bestimmten Tiefe wird der Algorithmus frühzeitig gestoppt und ein Knoten, welcher die maximale Tiefe überschreitet, zwangsweise zu einem Blatt umgewandelt.

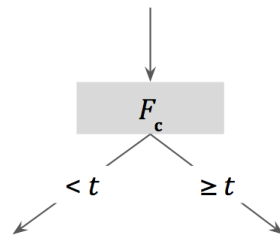


Abbildung 2.16: Aufspaltung einer kontinuierlichen Variable im Entscheidungsbaum

Post-Pruning: Zuerst wird der komplette Entscheidungsbaum aufgebaut und Overfitting zugelassen. Im Nachhinein wird der Entscheidungsbaum in seiner Tiefe reduziert. Eines der am weitesten verbreiteten Post-Pruning-Algorithmen ist das sogenannte *Reduced Error Pruning*. Dabei wird ein Knoten des Entscheidungsbaumes zu einem Blatt umgewandelt und diesem Blatt das Label zugewiesen, welches in seinem Sub-Baum am häufigsten vorkommt. Daraufhin wird der originale Entscheidungsbaum und sowie der beschnittene Entscheidungsbaum verwendet, um den Test-Datensatz zu klassifizieren. Ist der Klassifizierungsfehler des beschnittenen Baumes nicht schlechter als der des originalen Baumes, wird das Pruning übernommen. Dieses Vorgehen wird für jeden Knoten des Entscheidungsbaumes angewandt. [27, S. 68 - 70]

2.4.2 Gütemaße binärer Klassifikatoren

Ein binärer Klassifikation ist eine, bei dem es nur zwei Klassen gibt, das heißt $|Y| = 2$. Applikationsabhängig werden die beiden Klassen beispielsweise als *Positive* und *Negative*, 1 und 0 oder *True* und *False* bezeichnet. Wird bei einer Klassifizierung ein tatsächliches Positive korrekt als Positive vorhergesagt wird, spricht man von einem *True Positive* [TP]. Wird hingegen ein tatsächliches Positive fälschlicherweise als Negative vorhergesagt, spricht man von einem *False Negative* [FN]. Bei der Klassifizierung von Negatives spricht man dementsprechend von *True Negatives* [TN] und *False Positives* [FP]. Die *Confusion Matrix* in Abbildung 2.17 gibt eine Übersicht über die vier möglichen Klassifikations-Ergebnisse. [20, S. 213 - 214]

		Predicted Class	
		Positive	Negative
Real Class	Positive	True-Positive	False-Negative
	Negative	False-Positive	True-Negative

Abbildung 2.17: Confusion-Matrix (nach: [20, S. 214])

Die insgesamt Güte einer Klassifikation wird durch die *Genauigkeit* (engl. *Accuracy*) nach Formel 2.21 bestimmt. Eine Genauigkeit von 100% bedeutet, dass *alle* Instanzen richtig klassifiziert wurden, eine Genauigkeit von 50% bedeutet, dass die Hälfte aller Instanzen richtig klassifiziert wurden. Je höher die Genauigkeit, desto geringer der Klassifikationsfehler. [20, S. 214]

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (2.21)$$

Mit Hilfe der Genauigkeit lässt sich die insgesamt Performance des Klassifikators messen. Der Wert allein gibt jedoch keinen Aufschluss darüber, ob der Klassifikator eher eine Tendenz zur falschen Klassifizierung von Positives oder Negatives hat. Bei einer Datenbank mit der selben Anzahl an Positives und Negatives kann eine Genauigkeit von 50% beispielsweise dadurch entstehen, dass *alle* Instanzen als Positives markiert werden. Das heißt, dass alle Positives richtigerweise als Positives, aber alle Negatives fälschlicherweise ebenfalls als Positives klassifiziert werden. Im Umgedrehten Fall ergibt die Klassifizierung aller Instanzen als Negatives ebenfalls eine Genauigkeit von 50%. In einem dritten Fall irrt sich die Klassifikator gleich oft bei der Einordnung der Negatives und Positives.

Die Maße *Sensitivität* (engl. *Sensitivity*) und *Spezifität* (engl. *Specificity*) geben Aufschluss über die Performance des Klassifikators bei der Prädiktion der Positives und Negatives. Die Sensitivität, auch bezeichnet als *True-Positive-Rate*, bemisst den Anteil tatsächlicher Positives, die auch als solche erkannt wurden, nach Formel 2.22. Eine Sensitivität von 100% bedeutet, dass alle in der Datenbasis enthaltenen tatsächlichen Positives auch als solche erkannt wurden. Die Erkennungsrate der Negatives hat keinen Einfluss auf die Sensitivität. Eine hohe Sensitivität lässt sich somit „einfach“ erzielen, in dem man *alle* Instanzen immer als Positives klassifiziert.[20, S. 222]

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2.22)$$

Die Spezifität nach Formel 2.23 bestimmt analog zur Sensitivität den Anteil der Negatives, die als solche klassifiziert wurden.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2.23)$$

Ein Klassifikator, der alle Instanzen als Positives markiert, hat zwar eine Sensitivität von 100%, aber eine Spezifität von 0%. Ergeben zwei verschiedene Klassifikationsmodelle sehr ähnliche Genauigkeiten, hilft die Bestimmung der Sensitivität und der Spezifität bei der Auswahl des für den Anwendungsfall adäquateren Klassifikators. So ist beispielsweise bei der Bestimmung von schweren Krankheiten eventuell ein Klassifikator mit höherer Sensitivität wünschenswert, um die Wahrscheinlichkeit zu minimieren, dass die entsprechende Krankheit nicht erkannt wird. [23] [20, S. 222]

3 Ableitung der Schmerz Score

In Kapitel ?? wurden Methoden vorgestellt, durch die einem Audiosignal die Cry-Units erkannt und markiert werden. Diese Cry-Units bilden den Ausgangspunkt, um eine Schmerz Score ableiten zu können.

Wie in Kapitel 2.1.1 erläutert wurde, wird eine Schmerz Score typischerweise nicht aus den Informationen einer einzelnen Cry-Unit, sondern aus Gruppen mehrerer Cry-Units innerhalb eines längeren Beobachtungszeitraumes geschlossen. Dafür ist es zunächst notwendig, diejenigen Cry-Units zu identifizieren, die gemeinsam einem schmerzversuchenden Stimulus zugeordnet werden können. Diese Gruppierung von Cry-Units wird in dieser Arbeit als *Segmentierung* bezeichnet und in Kapitel 3.1 vorgestellt. Die so entstandenen Segmente bilden die Basis zur Ableitung von Schmerz Scores, welche in Kapitel ?? diskutiert wird.

3.1 Segmentierung

Wie in Kapitel 2.1.1 erläutert wurde, wird die Schmerzdiagnose für Post-Operativen Schmerz mit Pain Scales typischerweise umgesetzt, indem das Baby in bestimmten Intervallen besucht und für einen festgelegten Zeitraum beobachtet wird, worauf den Schmerzgrad für einen bestimmten Zeitpunkt diagnostiziert wird. So empfiehlt die Pain Scale PAT, das Baby alle 30 Minuten zu besuchen und für 15 bis 30 Sekunden zu beobachten. Für viele Pain Scales konnten diese Beobachtungsintervalle und Zeiträume nicht in Erfahrung gebracht werden.

Dieses Vorgehen der Schmerzdiagnose nur zu festgelegten Zeitintervallen macht keinen Sinn für ein kontinuierliches System, da es den Vorteil der Kontinuität eliminieren würde. Die Frage ist also, wann eine medizinische Fachkraft den Anfangszeitpunkt einer Schmerzdiagnose festlegen würde, würde es das entsprechende Baby rund um die Uhr beobachten, und nicht nur eine Visite zu festgelegten Zeitpunkten durchführen. Die logisch nächstliegende Antwort ist, die Schmerzdiagnose genau dann zu starten, wenn das Baby anfängt, zu schreien.

Die nächste Frage ist, wann man die Beobachtung wieder beendet und eine Schmerz Score für den beobachteten Zeitraum festlegt. Da einige Pain Scales aus Tabelle 2.2 Beobachtungszeiträume angeben, kann argumentiert werden, dass die Beobachtung direkt nach dem für die jeweilige Scale angegebenen Zeitraum beendet werden und eine Pain Score abgeleitet werden kann, auch, wenn das Baby weiterhin schreit. Dabei stellt sich die Frage, welche Zeiträume man für die Pain Scales festlegt, die keinen selber keinen vorschreiben. Die zweite Frage ist, ob eine Vorzeitige Beendigung der Beobachtung überhaupt sinnvoll ist, wenn ein automatisierte System zur Verfügung steht und der Beobachter somit „unendlich viel Zeit hat“. Daher wird an dieser Stelle zuerst die grundlegendere Frage gestellt: Was ist der längste sinnvolle Zeitraum, für den eine Pain Score abgeleitet werden kann? Die nächstliegende Antwort ist: Dann, wenn das Kind aufhört, zu schreien.

Auf Basis dieser Argumentation wurde das folgende Vorgehen zur kontinuierlichen Seg-

mentierung entwickelt: Wenn das Baby keine Äußerungen von sich gibt, weil es beispielsweise schläft, wird keine Cry-Unit festgestellt, und somit existiert auch momentan kein offenes Segment. Fängt das Baby an, einen Laut von sich zu geben, also eine Cry-Unit zu produzieren, wird ein neues Segment eröffnet und die Cry-Unit diesem Segment hinzugefügt. Weitere Cry-Units werden so lange diesem Segment hinzugefügt, wie die Dauer der Stille nach einer Cry-Unit einen festgelegten Grenzwert t_s nicht überschreitet. Ein Cry-Segment wird dann geschlossen, wenn das Baby für einen festgelegten Zeitraum keine Laute mehr von sich gibt, also „aufhört, zu weinen“. Das Endzeitpunkt des Segmentes wird als der Endzeitpunkt der letzten Cry-Unit des Segmentes festgelegt.

Formel 3.1 definiert ein *Cry-Segment* $[CS]$ als Datentyp. Ein Cry-Segment ist eine Liste von Cry-Units. Alle Cry-Units erfüllen die Nebenbedingung 3.2, das heißt, dass die Distanzen aller benachbarter Cry-Units eines Cry-Segments unterhalb des Grenzwertes t_s liegen.

$$CS = [cu_0, \dots, cu_n] \quad (3.1)$$

$$\forall cs \in CS : \forall i = 0 \dots \text{length}(cs) - 2 : d(cs[i], cs[i + 1]) < t_s \quad (3.2)$$

Der Start-Zeitpunkt eines Cry-Segments wird nach Formel 3.3 als der Startzeitpunkt der ersten Cry-Unit des Segments definiert. Das Ende eines Segmentes wird definiert als der Endzeitpunkt der letzten Cry-Unit nach Gleichung 3.4.

$$\text{start}(cs) = cs[0].\text{start} \quad (3.3)$$

$$\text{end}(cs) = cs[n].\text{end} \quad (3.4)$$

Algorithmus 1 zeigt einen Pseudocode, wie die Segmentierung nach dem beschriebenen Prinzipien offline durchgeführt wird. Input des Algorithmus ist die Liste aller Cry-Units $CU_{all} = [cu_0, \dots, cu_m]$, die durch das Decision-Smoothing nach Algorithmus ?? entstanden ist. Das Ergebnis des Algorithmus ist die Liste CS_{all} , die alle gefundene Cry-Segmente $[cs_0, \dots, cs_n]$ enthält. Der Algorithmus eignet sich nicht für eine Online-Segmentierung, da das Ende eines Segmentes erst bei Beginn eines neuen Segmentes festgestellt wird, wobei beliebig viel Zeit zwischen den Beiden Segmenten liegen kann. Wurde beispielsweise $t_s = 1$ min festgelegt, und die Pause zwischen zwei Segmenten beträgt eine Stunden, so wäre das Ende des ersten Segmentes 59 Minuten zu spät festgestellt worden. Bei einer online durchgeführten Segmentierung empfiehlt es sich, die Dauer der Stille nach jeder neu erkannten Cry-Unit kontinuierlich zu messen und ein Segment sofort zu beenden, wenn ein Stillezeitraum den Grenzwert t_s überschreitet.

Abbildung 3.1 zeigt die nach dieser Methode durchgeführte Segmentierung anhand eines Beispiels.

Das hier vorgestellte Vorgehen wurde absichtlich möglichst einfach gehalten, damit der Sinn des Parameters t_s leicht ersichtlich ist und somit von der medizinischen Fachkraft selbständig festgelegt werden kann. Schlussendlich ist eines der Hauptziele dieser Segmentierung, unnötige Berechnungen von Schmerz-Scores in den nachfolgenden Schritten zu vermeiden, so lange keine Cry-Units vorliegen. Das Ende eines Segmentes ist außerdem ein günstiger Zeitpunkt, um die Parameter des Kompressors im Pre-Processing auf Basis des

Algorithm 1 Gruppierung von Cry-Units zu Cry-Segments

```

1: function SEGMENTCRYUNITS( $CU_{all}, t_s$ )
2:    $CS_{all} \leftarrow []$ 
3:    $cs \leftarrow [CU_{all}[0]]$ 
4:   for  $i = 1, \dots, \text{length}(CU_{all}) - 1$  do
5:      $cu_i \leftarrow CU_{all}[i]$ 
6:      $cu_{i-1} \leftarrow CU_{all}[i - 1]$ 
7:     if  $d(cu_{i-1}, cu_i) < t_{seg-max}$  then
8:        $cs \leftarrow [cs_i, cu_i]$ 
9:     else
10:       $CS_{all} \leftarrow [CS_{all}, cs]$ 
11:       $cs \leftarrow [cu_i]$ 
12:    end if
13:  end for return  $CS_{all}$ 
14: end function

```

RMS-Wertes des Segmentes zu aktualisieren (siehe Kapitel ??). Trotz der Trivialität dieser laufenden Segmentierung liegt hier ein wichtiger Unterschied im Gegensatz zu vergleichbaren Systemen, wie zum Beispiel das von Cohen et al. [6], bei dem die Entscheidung über Cry/not-Cry für Segmente mit einer festen Fenstergröße von 10 Sekunden vorgenommen wird.

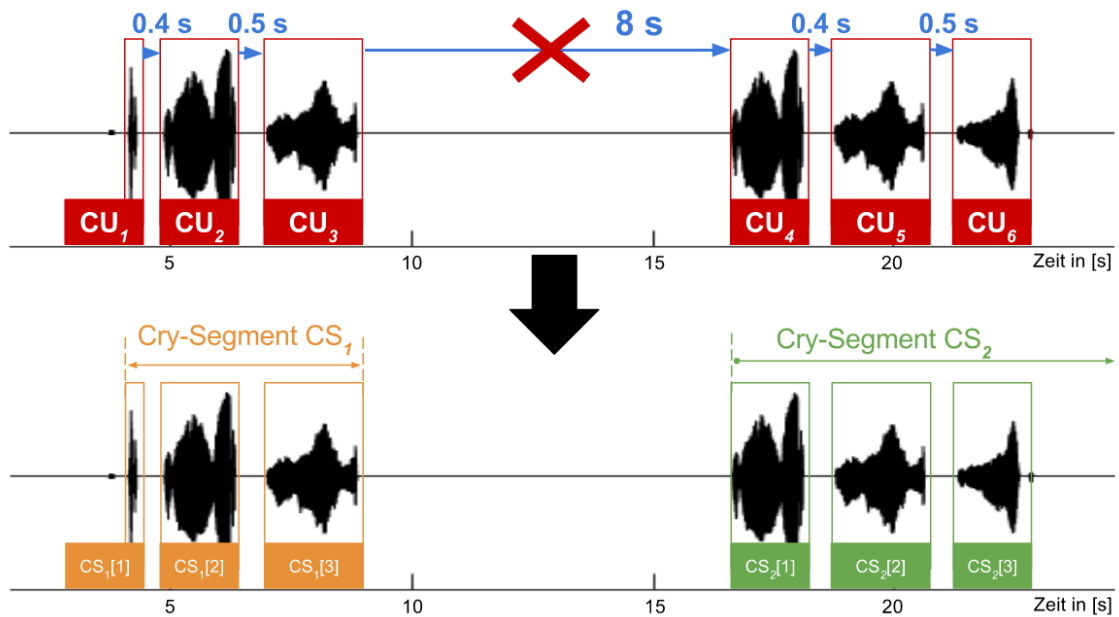


Abbildung 3.1: Beispiel für eine Segmentierung mit einem Grenzwert von $t_s = 6$ s

3.2 Extrahierung von Eigenschaften und Ableitung der Schmerzscore

Das Ergebnis der Segmentierung ist eine Liste an Cry-Segmenten cs_0, \dots, c_n . Diese Cry-Segmente bilden nun die Basis für die Ableitung der Pain Score. Die medizinische Fachkraft, die das System verwendet, muss dabei zuerst die Wahl treffen, welche Pain Scale verwendet werden soll. Das einfachste denkbare Vorgehen ist die Ableitung genau einer Pain Score aus den globalen Eigenschaften eines Segmentes, wobei diese Ableitung erst vollzogen werden kann, sobald ein Segment abgeschlossen wurde und alle Informationen für dieses Segment vorliegen. Es wird also jedem Segment genau eine Pain Score zugewiesen. Das Vorgehen wird am Beispiel der NIPS aus Tabelle 2.1 verdeutlicht: Dabei steht die Abwesenheit von Weinen für null Punkte, „mumbling“ (murmeln) für einen Punkt und „vigorous“ (energisch) für zwei Punkte. Bei Abwesenheit von Lautäußerungen, also der Zeitraum zwischen den Segmenten, werden also keine Punkte = null Punkte vergeben. Ein Segment, dessen Qualität insgesamt als „murmelnd“ bewertet wird, erhält einen Punkt, und ein Segment, welches als insgesamt als „energisch“ bewertet wird, zwei Punkte. Das Problem ist offensichtlich: „murmelnd“ und „energisch“ sind subjektiv behaftete Begriffe und lassen sich nicht ohne weiteres aus den Eigenschaften eines Segmentes feststellen.

Es werden zwei verschiedene Lösungs-Strategien für dieses Problem vorgestellt.

Strategie 1

... löst das Problem mit Hilfe von *Regression* (Siehe Kapitel ??):

1. Man erstellt eine Datenbank mit Aufnahmen von kindlichen Lautäußerungen, die man segmentiert.
2. Man errechnet „so viele *objektiv* messabare Eigenschaften wie möglich“ für jedes Segment, wie zum Beispiel die insgesamte Länge, die durchschnittliche Länge der enthaltenen Cry-Units, durchschnittliche Tonhöhe usw.
3. Man bittet medizinische Fachkräfte, für jedes Segment der Datenbank eine Score bezüglich einer Pain Scale zu vergeben. Dadurch erhält man eine gelabelte Test-Datenbank.
4. Man verwendet einen *Regressionsalgorithmus*, um den Zusammenhang zwischen den in Schritt 2 objektiv gemessenen Eigenschaften der Segmente und den in Schritt 3 vergebenen *Scores* herzustellen. An dieser Stelle kann zum Beispiel die in Kapitel ?? beschriebene multiple lineare Regression verwendet werden. Man erhält somit einen Regressor für jede Pain Scale.
5. Möchte man für neue, unbekannte Segmente die Pain Score prognostizieren, nutzt man den entsprechenden Regressor.

Das Vorteil dieses Vorgehens ist, dass das Problem der Übersetzung der objektiv messbaren Parameter in die subjektiv behafteten Begriffe überbrückt wird, indem die Regression direkt von den objektiv messbaren Parametern auf die Pain Score durchgeführt wird. Der Nachteil ist, dass eine Testdatenbank für jede Pain Scale aufgebaut werden muss. Wird ein neue Pain Scale eingeführt, muss der Regressor für diese Scale durch erneutes Labeln festgestellt werden. Ein weiterer Effekt der Abbildung des Problems als Regression ist, dass ein Regressor

in einen kontinuierlichen Zahlenraum abbildet. Es sind also Regressionsergebnisse wie zum Beispiel 2.8 denkbar. Diese „bessere Auflösung“ kann als Vorteil betrachtet werden. Ist jedoch eine direkte Übersetzung der Pain Scale inklusive der ganzzahligen Punktzahlen gewünscht, so stellt sich die Frage, ob eine 2.8 auf- oder abzurunden ist.

Strategie 2

... löst das Problem mit Hilfe von Klassifizierung (Siehe Kapitel ??):

1. und 2. entsprechen Strategie 1
3. Man sammelt alle subjektiven Begriffe, die in Pain Scales verwendet werden, wie zum Beispiel „murmelnd“, „energisch“, usw.
4. Man bittet medizinische Fachkräfte, jedes Segment der Datenbank mit denjenigen Begriffen zu labeln, die die jeweilige Person für zutreffend hält.
5. Man Verwendet einen *Klassifizierungsalgorithmus*, um einen Zusammenhang zwischen den in Schritt 2 festgestellten objektiv messbaren Eigenschaften der Segmente und den *subjektiv behafteten Begriffen* zu finden. Man erhält somit einen Klassifikator für jeden Begriff, der binär in *positive = zutreffend* und *negative = nicht zutreffend* klassifiziert.
6. Möchte man für neue, unbekannte Segmente die Pain Score prognostizieren, so wird für jede mögliche Score der Pain Scale überprüft, ob für alle subjektiv beschreibenden Begriffe der entsprechende Klassifikator ein positive prognostiziert. Die Ableitung der Score ist somit ein weiteres Klassifizierungsproblem, wobei eine Score einer Klasse entspricht und genau dann abgeleitet werden kann, wenn alle Voraussetzungen für die Klasse erfüllt sind.

Der Vorteil dieser Methode ist, dass auch zum Zeitpunkt der Erstellung der Testdatenbank unbekannte Pain Scales zu einem späteren Zeitpunkt eingebunden werden können, insofern alle in dieser neuen Pain Scale verwendeten subjektiv behafteten Begriffe bereits gelabelt vorliegen, weil sie auch in anderen Pain Scales verwendet wurden. Das Vorgehen erlaubt somit eine gewissen Flexibilität bezüglich zukünftig entwickelter Pain Scales. Der Nachteil dieser Methode ist, dass durch die Umwandlung der eigentlich quantitativ geordneten Score einer Pain Scale in qualitative Klassen aus einem implizit als Regression zu betrachtenden Problem ein Klassifizierungsproblem macht. Dies wirft neue Fragen auf, wie zum Beispiel: Angenommen, bei einer fiktiven Pain Scale wird jede Score mit jeweils drei subjektiv behafteten Begriffen beschrieben, und bei der Klassifizierung eines Segmentes wird festgestellt, dass für jede Punktzahl genau zwei der drei Begriffe erfüllt werden. Welche Score wird dann prognostiziert? Ein anderes Beispiel wird am Beispiel der NIPS-Score aus Tabelle 2.1 verdeutlicht: Angenommen, ein Cry-Segment enthält hörbar „starkes“ Schreien, es kann jedoch weder „mumbling (murmelnd)“ noch „vigorous (energisch)“ abgeleitet werden. Demzufolge müsste dieses Segment eine Score von 0 Punkten erhalten, wobei ein Mensch in dieser Situation eventuell „stark“ zu „heftig“ uminterpretieren und 2 Punkte vergeben hätte. Strategie 1 ist weniger anfällig für dieses Problem.

In jedem Fall werden medizinische Fachkräfte benötigt, um das Labeling der Cry-Segmente durchzuführen, was aus Zeitgründen im Rahmen dieser Arbeit nicht möglich ist. Die Aquisierung von Audioaufnahmen von Babys sowie das Labeling der Aufnahmen erfordern nicht nur Zeit, sondern das Fachwissen über das Führen und die Auswertung von Interviews.

3.2.1 Extrahierung von Eigenschaften

Im vergangenen Kapitel wurde erläutert, dass die Basis für die Ableitung einer Pain Score für ein Segment die Extraktion von „so vielen Features wie möglich“ ist. In diesem Kapitel wird präzisiert, welche Features gemeint sind. Varallyay [40, S. 16 - 17] schlug vor, drei Kategorien an Features zu betrachten: 1.) Features des Zeitbereichs, 2.) Features der Frequenzbereichs, und 3.) Melodie-bezogene Attribute. Diese Kategorisierung wurde für diese Arbeit übernommen.

In Kapitel 2.3.1 wurde beschrieben, welche Features in der medizinischen Schreiforschung typischerweise extrahiert wurden. In Kapitel 2.3.2 wurde diskutiert, dass 1.) nicht bewiesen ist, welche Features die „wichtigsten“ sind und 2.) keine Einigung darüber herrscht, wie genau bestimmte Features zu berechnen sind. Basierend auf den in diesem Kapitel vorgestellten Features werden in diesem Kapitel konkrete Berechnungsvorschriften definiert. Welche von diesen Features tatsächlich im Zusammenhang mit Schmerz stehen, lässt sich erst in der anschließenden Nutzung der Features zur Regression oder Klassifizierung der Pain Scales feststellen, welche jedoch im Rahmen dieser Arbeit nicht durchgeführt werden kann.

Features des Zeitbereiches

Mit Features des Zeitbereiches sind solche gemeint, die sich allein aus Kenntnis der der Start- und Endzeitpunkte der im Segment enthaltenen Cry Units sowie deren Zeitbereiche gewinnen lassen, wie beispielsweise die durchschnittliche Länge der Cry-Units oder die durchschnittliche Energie der Cry-Units. In diesem Kapitel gilt die Konvention, dass eine Cry-Segment cs insgesamt N Cry-Units enthält, die Indexierung wird mit $0 \dots N - 1$ definiert.

Segment-Length: Zeitliche Länge des Segmentes:

$$\text{Segment-Length}(cs) = cs[N - 1].end - cs[0].start \quad (3.5)$$

Density: Relativer Anteil der Cry-Units an der Länge des Segmentes („Dichte“)

$$\text{Density}(cs) = \frac{\sum_{i=0}^{N-1} \lambda(cs[i])}{\text{Segment-Length}(cs)} \quad (3.6)$$

Tempo: Das Verhältnis zwischen der Dauer des Segmentes und der Anzahl der Cry-Units. Dieses Feature ähnelt dem von LaGasse et al. [22, S. 85] als *Utterances* bezeichneten Feature.

$$\text{Tempo}(cs) = \frac{N}{\text{Segment-Length}(cs)} \quad (3.7)$$

Statistics of Cry-Units: Statistische Auswertungen bezüglich der *Länge der Cry-Units* $\text{stats}_{cu}(cs)$: Durchschnitt, Median, Minimum, Maximum und Standardabweichung

der Cry-Units. Das $\text{mean}_{cu}(cs)$ -Feature wird von LaGasse et al. [22, S. 85] und vielen weiteren Schreiforschern als *Mean Duration* bezeichnet.

$$\text{stats}_{cu}(cs) = \begin{cases} \text{mean}_{cu}(cs) = \text{mean}_{i=0 \dots N-1} \{\lambda(cs[i])\} \\ \text{median}_{cu}(cs) = \text{median}_{i=0 \dots N-1} \{\lambda(cs[i])\} \\ \text{min}_{cu}(cs) = \min_{i=0 \dots N-1} \{\lambda(cs[i])\} \\ \text{max}_{cu}(cs) = \max_{i=0 \dots N-1} \{\lambda(cs[i])\} \\ \sigma_{cu}(cs) = \sigma_{i=0 \dots N-1} \{\lambda(cs[i])\} \end{cases} \quad (3.8)$$

Statistics of Bursts: ¹ Die in Gleichung 3.8 definierten Features können ebenso in Bezug auf die *Längen der Bursts* errechnet werden, in dem in jeder Gleichung $\lambda(cs[i])$ ersetzt wird durch $cs[i].\text{start} - cs[i-1].\text{start}$. Die Indexierung muss auf $i = 1, \dots, N-1$ begrenzt werden.

$$\text{stats}_{burst}(cs) = \begin{cases} \text{mean}_{burst}(cs) = \text{mean}_{i=1 \dots N-1} \{cs[i].\text{start} - cs[i-1].\text{start}\} \\ \text{median}_{burst}(cs) = \text{median}_{i=1 \dots N-1} \{cs[i].\text{start} - cs[i-1].\text{start}\} \\ \dots \end{cases} \quad (3.9)$$

Statistics of Pauses: Nach dem selben Muster werden die statistischen Auswertungen bezüglich der *Längen der Pausen* ermittelt. Eine Pause entspricht in diesem Zusammenhang der Distanz zwischen zwei aufeinander folgenden Cry-Units nach Gleichung ??.

$$\text{stats}_{pause}(cs) = \begin{cases} \text{mean}_{pause}(cs) = \text{mean}_{i=1 \dots N-1} \{d(cs[i-1], cs[i])\} \\ \text{median}_{pause}(cs) = \dots \end{cases} \quad (3.10)$$

Statistics of Energies: Zunächst wird die Liste aller in den Cry-Units enthaltenen Signalfenster definiert nach Gleichung 3.11. Eine Cry-Unit hat die Signalfenster $cu.\text{windows} = x_0[], \dots, x_m[]$

$$x_{seg}[] = cs[0].\text{windows}[0], \dots, cs[N-1].\text{windows}[m] \quad (3.11)$$

Die Liste $x_{seg}[]$ hat R Elemente, die Indexierung wird definiert mit $0, \dots, R-1$. Gleichung 3.12 definiert die Features bezüglich der MSV-Werte („Lautstärken“) des Segmentes. Der MSV-Wert als Maß des durchschnittlichen Energiegehaltes wurde in Gleichung 2.6 definiert.

$$\text{stats}_{msv}(cs) = \begin{cases} \text{mean}_{msv}(cs) = \text{mean}_{i=0 \dots R-1} \{MSV(x_{seg}[i])\} \\ \text{median}_{msv}(cs) = \dots \end{cases} \quad (3.12)$$

¹Erläuterung zum Begriff *Burst* in Kapitel 2.3.1

Diese statistischen Auswertungen bezüglich der Länge der Cry-Units und Bursts wurden beispielsweise von Zeskind et al. [34] vorgenommen, wenn auch nicht Computer-gestützt. Es ist zu bemerken, dass in der klassischen Schreiforschung zeitliche Features im geringeren Maße in Betracht gezogen wurden als Features des Frequenz-Bereiches. Die einzigen zeitliche Features, die zum Beispiel von Wasz-Hockert et al. [31], Fuller [10] und LaGasse et al. [22] berechneten, sind *die durchschnittliche Länge der Cry-Units* (hier $\text{mean}_{cu}(cs)$) und die *Latenz zwischen Reiz und erster Cry-Unit*, welche nur auf Basis des Audiosignals nicht feststellbar ist. Es spricht jedoch nichts dagegen, die hier vorgestellten Features trotzdem zu erproben. Die anschließende Nutzung der Features zur Regression/Klassifizierung wird Auskunft darüber geben, welchen Beitrag diese Features zur Schmerzdiagnose leisten können.

Features des Frequenzbereiches und der Melodie

Mit Features des Frequenz-Bereiches sind diejenigen Features gemeint, die sich aus der Short Time Fourier Transformation der Cry-Units gewinnen lassen. Um die Features durch mathematische Formeln definieren zu können, wird zuerst das *Spectrum des Segmentes* $X_{seg}[\]$ nach Formel 3.13 als die Liste aller Frequenz-Bereiche der Signalfenster der Cry-Units des Segmentes definiert. Die Indexierung von $X_{seg}[\]$ läuft, wie bei $x_{seg}[\]$ von $0, \dots, R-1$. Nach dem selben Muster wird das *Cepstrum des Segmentes* $c_{seg}[\]$ definiert.

$$X_{seg}[\] := \bigvee_{x_i[\] \in x_{seg}} : |DFT\{x_i[\] \cdot w[\]\}| \quad (3.13)$$

Die folgenden Features des Frequenzbereiches lassen sich mit den in dieser Arbeit vorgestellten Methoden berechnen:

Tensness: Das Feature, welches in Kapitel 2.3.1 als „Ratio2“ beschrieben wurde. Es wurde von Fuller [10] eingeführt und beschreibt die Spannung des Vokaltraktes als Verhältnis der Energien oberhalb von 2000 Hz zu unter 2000 Hz. Wie bei den statistischen Auswertungen der Features des Zeitbereiches kann für das gesamte Segment der Durchschnitt, Median, Maximum, Minimum und Standardabweichung berechnet werden.

$$\text{stats}(Tensness) = \begin{cases} \text{mean}_{Tens}(cs) = \text{mean}_{i=0 \dots R-1} \left\{ \frac{\sum_{k=0}^{2000 \text{ Hz}} X_{sec}[i][k]}{\sum_{j=2000 \text{ Hz}}^{f_s} X_{sec}[i][j]} \right\} \\ \text{median}_{Tens}(cs) = \dots \end{cases} \quad (3.14)$$

Clarity: Wie in Kapitel ?? erläutert wurde, lässt eine stark ausgebildete Spitze im oberen Cepstrum-Bereich auf ein stimmhaftes Signal schließen. Ein hoher Anteil stärkerer Cepstrum-Peaks lässt also auf vermehrt phonierte Laute schließen, geringere Cepstrum-Peaks auf dysphoniertere Laute (Siehe Kapitel 2.3.1). Dieses durchschnittliche Wert dieses Features trifft eine Aussagen über den Anteil dysphonierter Laute, die Standardabweichung ähnelt dem in Kapitel 2.3.1 vorgestellten *Cry-Mode Changes*-Feature.

$$\text{stats}_{clarity}(cs) = \begin{cases} \text{mean}_{Clarity}(cs) = \text{mean}_{i=0 \dots R-1} \left\{ Ceps_{mag}(c_{seg}[i]) \right\} \\ \text{median}_{Clarity}(cs) = \dots \end{cases} \quad (3.15)$$

Alle weiteren Features, die in Kapitel 2.3.1 vorgestellt wurden und sich auf den Frequenzbereich beziehen, lassen sich nicht mehr mit den in dieser Arbeit vorgestellten Methoden extrahieren. Entweder beziehen sie sich auf die Lage der Formanten, oder basieren auf der Feststellung der Grundtonhöhe. In dieser Arbeit konnten aus Platzgründen jedoch keine Methoden zur Extraktion dieser Informationen vorgestellt werden. Gleiches gilt für die Feststellung des Melodieverlaufs, welche ebenfalls auf der Feststellung der Grundtonhöhe basiert. Das Muster, nach dem diese Features berechnet werden können, sollte aus den bisher vorgestellten Features ersichtlich sein. So lassen sich beispielsweise die Features bezüglich der Grundtonhöhe nach Formel 3.16 ableiten. Dabei sei $f_0(X_i[])$ eine idealisierte Funktion, welche die Grundtonhöhe f_0 für das Frequenzfenster $X_i[]$ berechnet. Da für die Definition der weiteren Features idealisierte ebenfalls Funktionen angenommen werden müssten, wird die Festlegung weiterer Features an dieser Stelle nicht fortgeführt.

$$\text{stats}_{pitch}(cs) = \begin{cases} \text{mean}_{Pitch}(cs) = \text{mean}_{i=0 \dots R-1} \left\{ f_0(X_{seg}[i]) \right\} \\ \text{median}_{Pitch}(cs) = \dots \end{cases} \quad (3.16)$$

Diskussion

Bei allen vorgestellten Features handelt es sich, nach dem Vorbild der in Kapitel 2.3.1 vorgestellten Features der klassischen Schreiforschung, um solche, bei denen die Reihenfolge der Cry-Units nicht mit in Betracht gezogen wird. Angenommen, ein Segment besteht aus n Cry-Units, wobei genau eine Hälfte der Cry-Units kurz und die andere Hälfte der Cry-Units lang ist. Das $\text{stats}_{cu}(cs)$ -Feature wird bezüglich des Durchschnittes, Minimum, Maximum etc. die selben Werte berechnen, unabhängig davon, ob sich die kurzen Cry-Units allesamt am Beginn des Segmentes, am Ende des Segmentes oder mit den langen Cry-Units durchmischt befinden. Bei der anschließenden Nutzung der Features zu Regression/Klassifizierung wird sich zeigen, wie sehr sich diese Features zur Ableitung von Pain-Scores eignen. Stellt sich heraus, dass sich die Features nicht eignen, ist es eventuell notwendig, neue Features zu definieren, die die Position der Cry-Units mit in Betracht ziehen.

3.2.2 Ableitung der Pain-Score

Mathematisch wird die Ableitung einer Pain Score als Funktion $PS_{Scale} : cs \mapsto \mathbb{N}$ definiert, welche eine Cry-Segment cs unter Verwendung der Pain-Scale auf einen Pain-Score \mathbb{N} abbildet. Der konkrete Inhalt der Funktion ist für die entsprechende Pain Scale mit dem in diesem Kapitel beschriebenen Vorgehen zu ermitteln. Die Ableitung genau einer Pain Score für ein Cry-Segment stellt den einfachsten Fall dar. Dies ist für bestimmte Anwendungsfälle eventuell nicht ausreichend:

1. Die Score kann erst nach der Beendigung eines Segmentes abgeleitet werden, was für einigen Kontexte möglicherweise zu spät ist. Besonders die Schmerzdiagnostik während Schmerzverursachenden Prozeduren kann das häufigere „Aktualisieren“ der Schmerzscores notwendig machen.
2. Falls der Schmerz innerhalb eines Segmentes stark ab- oder zunimmt, ist dieser Verlauf nicht erkennbar. Es würde lediglich der „durchschnittliche Schmerz“ des Segmentes abgeleitet werden.

Das vorgestellte Prinzip wird daher erweitert, indem ein Aktualisierungsintervall t_{act} und Beobachtungszeiträume t_{obs} eingeführt wird.

Aktualisierungsintervall

Die Grundlegende Idee des Aktualisierungsintervalls ist, bei einem momentan offenen Segment in regelmäßigen Abständen die Features abzufragen und direkt die Pain Score abzuleiten, um Zwischenergebnisse zu erhalten. Der am häufigsten umsetzbare Fall ist, ein Aktualisierung nach jeder neu dem Segment hinzugefügten Cry-Unit vorzunehmen. Der am wenigsten häufige Fall ist der bereits genannte, die Aktualisierung erst bei Beendigung eines Segmentes durchzuführen. Eine offensichtliche Variante zur Festlegung von t_{act} ist die Bestimmung eines zeitlichen Wertes. Ein t_{act} von beispielsweise 10s würden bedeuten, dass alle 10 Sekunden ein neuer Pain Score für ein Segment berechnet wird. Die Beendigung eines Segmentes würde in jedem Fall eine Ableitung der Pain Score auslösen und einen „erzwungenen Aktualisierungszeitpunkt“ darstellen. Die folgenden Möglichkeiten zur Festlegung von t_{act} sind denkbar:

- t_{act} als globaler, nicht veränderbarer Wert. Da in der Literatur keine Vorschlag diesbezüglich gefunden werden konnte, müsste ein sinnvoller Wert in Absprache mit medizinischen Fachkräften eruiert werden.
- Man überlässt der medizinischen Fachkraft, die das System überwacht, die Festlegung des Aktualisierungsintervalls. So kann die Person selber bestimmen, wie häufig sie eine Aktualisierung der Pain Score wünscht.
- Die feste Bindung des Aktualisierungsintervalls an die verwendete Pain Scale. Die CRIES-Scale ist beispielsweise für das post-operative Monitoring gedacht und benötigt somit möglicherweise weniger häufige Aktualisierungen als der DAN, welcher zur Schmerzdiagnostik während einer Operation eingesetzt werden kann (siehe Tabelle 2.2). Da die Pain Scales nicht für die kontinuierliche Schmerzdiagnostik ausgelegt sind, lässt sich aus den für einige Pain Scales eventuell vorgeschriebenen Beobachtungsintervallen kein Aktualisierungsintervall für ein kontinuierliches System argumentieren.

Wenn t_{act} als zeitlicher Wert definiert wird, kann es passieren, dass eine Aktualisierung in einem offenen Segment durchgeführt wird, während gerade eine neue Cry-Unit markiert wird und noch nicht abgeschlossen wurde. Da die Funktion $PS_{Scale} : cs \mapsto \mathbb{N}$ nur für Cry-Segmente definiert wurde, die vollständige Cry-Units enthalten, wird diese „halbe Cry-Unit“ nicht mit zur Ableitung der Pain Score verwendet. Der Hintergrund für diese Entscheidung ist, dass bestimmte Funktionen zum Ableiten der Features für das Segment ansonsten fehlerhafte Ergebnisse liefern können. Die Funktion PS wird folglich auf das *Subsegment* cs_{sub} angewandt, welches alle Cry-Units des möglicherweise noch offenen Segmentes cs beinhaltet, die zum Aktualisierungszeitpunkt t vollständig begonnen und beendet wurden. Der Endzeitpunkt des Subsegmentes wird, wie bei herkömmlichen Segmenten, auf den Endzeitpunkt der letzten vollständigen Cry-Unit im Subsegment gelegt.

Beobachtungszeitraum

Es gibt Eigenschaften, die sich implizit auf den gesamten Zeitraum *Beginn des Segmentes* $start(cs)$ bis *Aktualisierungszeitpunkt* t beziehen, wie beispielsweise die *Zeitliche Länge des*

Segmentes nach Formel 3.5. Dieser Zeitraum ist gleichzeitig der längst mögliche Zeitraum innerhalb eines Segmentes, der für die Ableitung der Pain Score mit einbezogen werden kann. Es ist jedoch auch möglich, einen kürzere Beobachtungszeitraum t_{obs} zu wählen. Dies hat zur Folge, dass bei der Ableitung die ersten Cry-Units des Segmentes ausgelassen werden, die außerhalb des Beobachtungszeitraums liegen. So können zeitliche Veränderungen der Pain-Score innerhalb eines Segmentes detaillierter dargestellt werden. Es sind wiederum verschiedene Varianten zur Festlegung von t_{obs} denkbar:

- Festlegung eines globalen oder eines von der medizinischen Fachkraft frei wählbaren Wertes, so wie bei dem Aktualisierungsintervall t_{act} .
- Die feste Bindung des Aktualisierungsintervalls an die verwendete Pain Scale. Einige Pain Scale empfehlen bestimmte Beobachtungszeiträume. So wird beispielsweise bei der NIPS-Scale ein Beobachtungszeitraum von einer Minute empfohlen (Siehe Kapitel 2.1.2). Es müsste wiederum in Zusammenarbeit mit medizinischen Fachkräften eruiert werden, ob diese, für die manuelle Schmerzdiagnostik vorgesehenen Werte auch für ein automatisiertes System Sinn machen.
- Eine weitere Variante ist, t_{obs} an den Wert von t_{act} zu binden. Fall t_{obs} frei festlegbar sein soll, muss das Personal nicht zwei Werte festlegen. Ein Verhältnis von $t_{obs} = k \cdot t_{act}$ würde mit $k = 1$ nicht-überlappende Beobachtungszeiträume und mit $k = 2$ überlappende Beobachtungszeiträume erzeugen.

Der Beobachtungszeitraum t_{obs} schränkt somit die Länge des Subsegmentes cs_{sub} weiter ein, und zwar in diesem Fall bezüglich der Startzeitpunktes. Der Zeitraum innerhalb des Segmentes cs , der zur Bildung von cs_{sub} genutzt wird, ist der Zeitraum *Aktualisierungszeitpunkt* $t - t_{obs}$ bis t . Es werden nur solche Cry-Units von cs in cs_{sub} übernommen, die innerhalb dieses Zeitraumes vollständig begonnen und beendet werden konnten. Der Anfangszeitpunkt des Subsegments ist somit der Anfangszeitpunkt der ersten vollständigen Cry-Unit innerhalb Beobachtungszeitraumes, der Endzeitpunkt des Subsegmentes entspricht dem Ende der zum Aktualisierungszeitpunkt zuletzt vollständig beendeten Cry-Unit.

Beispiel

Die in diesem Kapitel vorgestellten Methoden zur Ableitung von Pain Scores werden anhand eines Beispiels verdeutlicht. Tabelle 3.1 definiert eine fiktive Pain Scale. Es werden die für Pain Scales typischen, subjektiv behafteten Worte zum Scoring verwendet. Es werden sowohl die subjektiv beschreibenden Kriterien angegeben, als auch die objektiv messbaren Kriterien, die mit einer gefunden wurden. In diesem Fall hängt die Pain Score von der Durchschnittslänge der Cry-Units des Segmentes $mean_{cu}(cs)$ ab (siehe Gleichung 3.8).

Tabelle 3.1: Fiktive Pain Scale

	0 Punkte	1 Punkt	2 Punkte
Fiction Scale	kein oder sehr wenig Weinen	normales Weinen	starkes Weinen

Mit Hilfe der in Kapitel 3 beschriebenen Strategien wurde die Funktion $PS_{Fiction}$ ermittelt,

definiert in Gleichung 3.17, welche die Ableitung des Schmerz Score für eine Cry-Segment mit Hilfe objektiv messbarer Features erlaubt.

$$PS_{Fiction}(cs) = \begin{cases} 0 & , \text{wenn } mean_{cu}(cs) < 0.3 \text{ s} \\ 1 & , \text{wenn } mean_{cu}(cs) < 1 \text{ s} \\ 2 & , \text{wenn } mean_{cu}(cs) \geq 1 \text{ s} \end{cases} \quad (3.17)$$

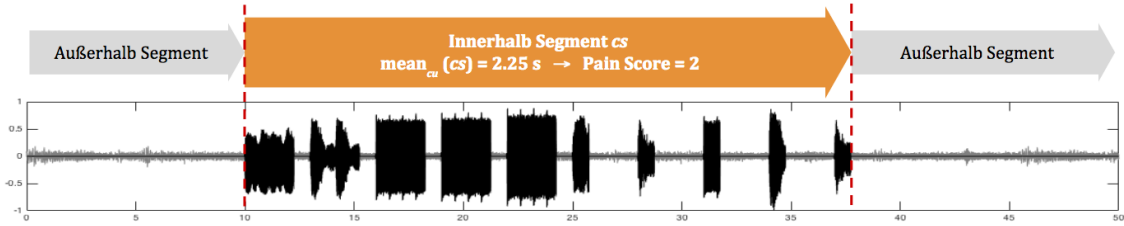


Abbildung 3.2: Beispiel für die Ableitung von Pain Scores für ein Signal nach einer fiktiven Pain Scale ohne Beobachtungszeitraum oder Aktualisierungsintervall

Abbildung ?? zeigt ein Beispielsignal, für das Pain Scores nach dieser Pain Scale abgeleitet werden. In dem Signal werden die stimmhaften Signalbereiche schwarz und das Hintergrundrauschen grau dargestellt. Es sind insgesamt 10 Cry-Units zu erkennen. Die ersten fünf Cry-Units haben jeweils eine Länge von 2.25 s, die letzten fünf Cry-Units eine jeweilige Länge von 0.75 s. Das Signal wurde nach der in Kapitel 3.1 beschriebenen Methode segmentiert mit $t_s = 5 \text{ s}$ und so alle 10 Cry-Units zu einem Segment zusammengefasst. Das Segment erstreckt sich von Sekunde 10 bis Sekunde 37.5. Für das Segment wurde eine durchschnittliche Länge der Cry-Units von $mean_{cu}(cs) = 1.5 \text{ s}$ gemessen und dem zufolge eine Pain Score von 2 abgeleitet. In diesem Fall wurde ohne Beobachtungs- und Aktualisierungsintervall gearbeitet. Wäre die Analyse also kontinuierlich vorgenommen worden, so wäre nach Feststellung der ersten Cry-Unit das Segment eröffnet, nach Überschreitung der maximal zulässigen Stille von $t_s = 5 \text{ s}$ nach der 10. Cry-Unit das Segment geschlossen, und daraufhin $PS_{Fiction}(cs) = 2$ berechnet worden.

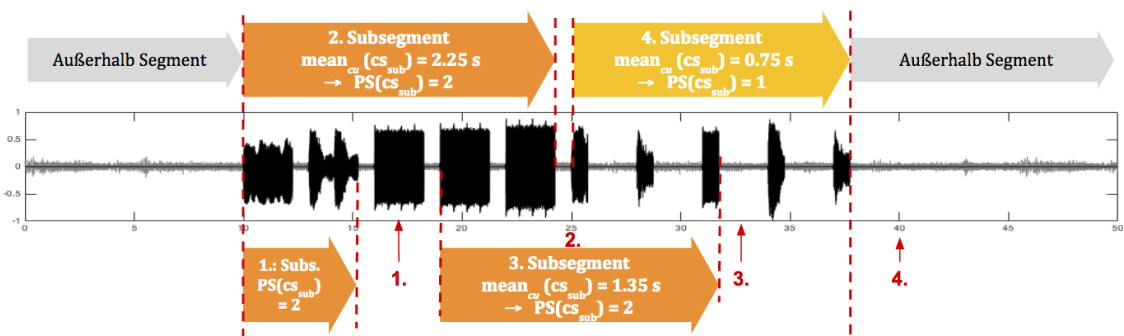


Abbildung 3.3: Beispiel für die Ableitung von Pain Scores für ein Signal nach einer fiktiven Pain Scale mit $t_{act} = 7.5 \text{ s}$ und $t_{obs} = 15 \text{ s}$

Abbildung ?? zeigt die Ableitung der Schmerz Scores, wenn zusätzlich ein Aktualisierungsintervall von 7.5 s und ein Beobachtungszeitraum von 15 s gewählt wird. Nach dem das Segment durch die Cry-Unit an Sekunde 10 eröffnet wurde, werden Aktualisierungen zu den Zeitpunkten $t = 17.5 \text{ s}, 25 \text{ s}, 32.5 \text{ s}$ und 40 s durchgeführt, verdeutlicht durch die

kleinen, roten Pfeile in der Abbildung. Wie zu sehen ist, wird bei jeder Aktualisierung innerhalb des Beobachtungszeitraumes ein Subsegment gebildet, für das Subsegment die Features errechnet und die Pain Score abgeleitet. Der Anfangszeitpunkt jedes Subsegmentes ist der Anfang der erste Cry-Unit innerhalb des jeweiligen Beobachtungszeitraums, und das Ende des Subsegmentes das Ende letzten Cry-Unit im jeweiligen Beobachtungszeitraum. Beispielsweise erstreckt sich das bei der 3. Aktualisierung der Beobachtungszeitraum von $17.5 - 32.5$ s, das Subsegment jedoch von $19 - 32$ s aufgrund der Lage der Cry-Units. Durch die Verwendung des Beobachtungs- und Aktualisierungsintervalls wird erkennbar, dass in diesem Beispiel der Schmerzgrad innerhalb des Segmentes nach hinten hin abnimmt.

4 Visualisierung

Ziel dieses Kapitels ist es, das erarbeitete Konzept zur Visualisierung der Pain Scores vorzustellen. Das Konzept wird anhand eines Beispielsignals mit einer Aufnahme des Weinenes eines Babys enthält. Das Signal wird in Abbildung 4.1 oben gezeigt. Der gezeigte Ausschnitt ist insgesamt 220 Sekunden lang. Die Signalabschnitte, die Stimme des Babys enthalten, werden Schwarz dargestellt, das Hintergrundrauschen grau. Das Signal wurde segmentiert mit $t_s = 10$ s und so zwei Segmente gefunden. Die beiden Segmente sind 30.5 Sekunden voneinander entfernt. Zur Schmerzdiagnostik auf Basis zwei verschiedener fiktiver Pain Scales durchgeführt, welche in Tabelle 4.1 aufgeführt werden. Die „Length-Scale“ bewertet den Schmerzgrad nach der Länge des Weinens und vergibt einen maximalen Score von 2, die „Min-Scale“ bewertet die Qualität des Weinens und vergibt einen maximalen Score von 3.

Tabelle 4.1: Fiktive Pain Scales zur Erläuterung der Visualisierung

	„Length-Scale“	„Min-Scale“
0 Punkte	kein Weinen	Lachen
1 Punkt	kurzes Weinen	leichtes Weinen
2 Punkte	langes Weinen	mittleres Weinen
3 Punkte	-	starkes weinen

Mit Hilfe einer der in Kapitel 3 vorgestellten Strategien wurden Formeln zur Ableitung der Schmerz Scores nach objektiv messbaren Eigenschaften gefunden. Gleichung 4.2 definiert die Funktion PS_{Length} , welche die Pain Score für Segmente nach der „Length-Scale“ ableitet. Aus der Gleichung geht hervor, dass ein Pain Score von 0 für ein Segment nicht abgeleitet werden kann, da die Anwesenheit eines Segmentes der Abwesenheit von Weinen implizit widerspricht. Gleichung 4.2 definiert die Funktion PS_{Min} zur Schmerzableitung nach der „Min-Scale“.

$$PS_{Length}(cs) = \begin{cases} 1 & , \text{ wenn } S\text{-Length}(cs) \leq 1 \text{ min} \\ 2 & , \text{ wenn } S\text{-Length}(cs) > 1 \text{ min} \end{cases} \quad (4.1)$$

$$PS_{Min}(cs) = \begin{cases} 0 & , \text{ wenn } min_{cu}(cs) < 0.3 \text{ s} \\ 1 & , \text{ wenn } 0.3 \text{ s} \leq min_{cu}(cs) \leq 1 \text{ s} \\ 2 & , \text{ wenn } 1 \text{ s} < min_{cu}(cs) \leq 2 \text{ s} \\ 3 & \text{ sonst} \end{cases} \quad (4.2)$$

Mit Hilfe der Gleichungen werden die Pain Scores für die beiden Segmente des Beispielsignals in Abbildung 4.1 abgeleitet. Es wurden dabei keine Aktualisierungsintervalle oder Beobachtungszeiträume genutzt. Wie zu sehen ist, hat das erste Segment nach der „Length-Scale“ einen Score von 1 und nach der „Min-Scale“ einen Score von 2, das zweite

Segment hat nach der „Length-Scale“ einen Score von 2 und nach der „Min-Scale“ einen Score von 1.

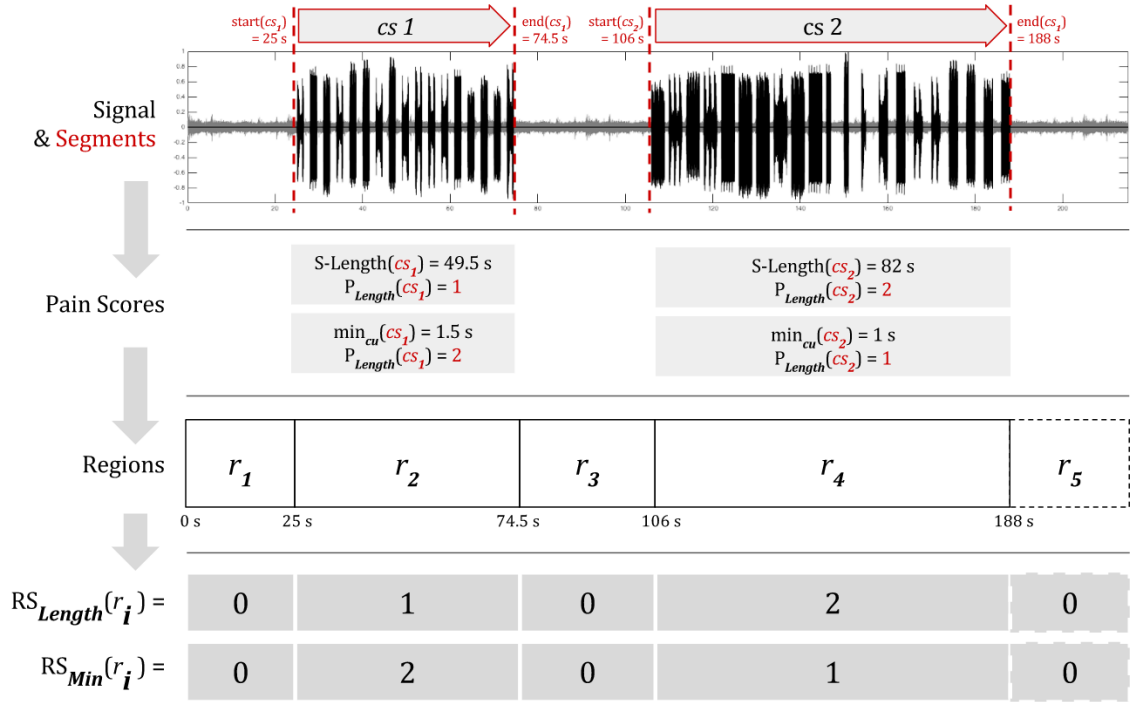


Abbildung 4.1: Oben: Ein Beispielsignal, sowie zwei Segmente, die mit $t_s = 10$ s gefunden wurden. Darunter: Pain Scores, die für die beiden Segmente berechnet wurden. Darunter: Schematische Darstellung des Signals als Regionen. Unten: Scores der Regionen.

Die Grundlegende Idee der Visualisierung ist nun, den zeitlichen Verlauf des Signals schematisch als einen *Balken* darzustellen. Dieser Balken wird in *Regionen* eingeteilt. Eine Region r beinhaltet entweder ein Cry-Segment oder den Raum zwischen zwei Cry-Segmenten. Die Länge einer Region entspricht der zeitlichen Länge des jeweiligen Stille- oder Cry-Segmentes. In Abbildung 4.1 werden fünf Regionen r_1, \dots, r_5 markiert, wobei die Region r_2 und r_4 Cry-Segmente enthalten. Die letzte Region wurde noch nicht abgeschlossen, da in der Abbildung nur ein Ausschnitt des sich eigentlich noch weiter fortsetzenden Signals gezeigt wird.

Jeder Region wird ein *Region Score* RS nach Formel 4.3 zugewiesen. Beinhaltet eine Region ein Cry-Segment, so wird der Region die entsprechende Pain Score zugewiesen. Beinhaltet die Region kein Cry-Segment, so erhält die Region automatisch einen Score von 0. In Abbildung 4.1 unten werden die Region Scores für das Beispielsignal nach den beiden fiktiven Pain Scales gezeigt.

$$RS_{Scale}(r) = \begin{cases} 0 & , \text{ wenn } r \text{ kein Cry-Segment beinhaltet} \\ PS_{Scale}(cs) & , \text{ wenn } r \text{ ein Cry-Segment } cs \text{ beinhaltet} \end{cases} \quad (4.3)$$

Das Ziel ist es nun, jede Region mit einer Farbe einzufärben, die die Region Score der entsprechenden Region anzeigt. Dazu wird eine Funktion $F_{Scale} : RS \mapsto P_{Scale}$ benötigt, welche in Abhängigkeit von der verwendeten Pain Scale eines Region Score auf eine Farbe

abbildet. Die Abbildungsfunktion F_{Scale} wird in diesem Zusammenhang als *Farbschema* bezeichnet und der Funktionsbereich P_{Scale} als *Farbpalette*. Das Farbschema und die Farbpalette soll die folgenden Kriterien erfüllen.

Diskrete Farbpalette Die Scores aller in Kapitel 2.1.1 vorgestellten Pain Scales haben diskrete Wertebereich. Ebenso soll P_{Scale} eine diskrete Menge an Farben enthalten, so dass $|RS_{Scale}| \leq |P_{Scale}|$.

Kleinst mögliche Veränderung des Farbschemas bei Wechsel der Pain Scale Wird zur Schmerzdiagnostik die Pain Scale verwendet, soll das Farbschema von der alten Pain Scale verwendete Farbschema möglichst einfach auf die neue Pain Scale übertragbar sein. Verwenden zwei Pain Scales den selben Wertebereich bezüglich der Scores, soll das selbe Farbschema verwendet werden. Wird angenommen, dass der geringste Pain Score jeder Pain Scale 0 beträgt¹, so soll P_{Scale} allein abhängig von dem maximalen Score der jeweiligen Scale sein.

Intuitive Farbsemantik durch Ampelschema Das Farbschema soll eine intuitive Zuordnung zwischen der Höhe des Score und der jeweiligen Farbe ermöglichen. In dieser Arbeit wurde sich für ein Ampelschema entschieden. Das heißt, dass der niedrigste Score „grün“, der höchste Score „rot“ und ein „mittlerer“ Score „gelb“ dargestellt werden soll. Definiert eine Pain Scale einen maximalen Score von 2, so wie beispielsweise das FLACC-System, so ergibt sich die Abbildung $0 \mapsto \text{grün}$, $1 \mapsto \text{gelb}$, $2 \mapsto \text{rot}$. Hat die Pain Scale mehr Scores, wie beispielsweise das MBPS mit insgesamt fünf möglichen Scores, so müssen geeignete Zwischenfarben definiert werden.

Visuelle Gleichabständigkeit der Farben Pain Scales definieren die Scores zwar in einer Reihenfolge, gewährleisten aber keine Vergleichbarkeit. Das heißt, dass bei einer Pain Scale ein Score von 4 in jedem Fall schlimmer ist als ein Score von 2, jedoch, das jedoch nicht heißt, dass ein Score von 4 automatisch doppelt so schlimm ist wie ein Score von 2. Die Farben eines Farbschemas sollen eine visuelle Gleichabständigkeit gewährleisten. Das heißt, dass die Farben, auf die zwei aufeinander folgende Scores einer Scale abgebildet werden, visuelle den gleichen Abstand zueinander haben. So wird verhindert, dass ein Farbschema eine Nähe oder einen Abstand zweier Scores suggeriert, der durch die Pain Scale nicht vorhanden ist.

Visuelle Gleichwichtigkeit der Farben Es kann nicht davon ausgegangen werden, dass ein bestimmter Score für die medizinische Fachkraft interessanter ist als andere Scores. Daher soll keine Farbe einer Farbpalette suggerieren, wichtiger als die anderen Farben der Palette zu sein.

Das Kriterium der Gleichabständigkeit legt die Verwendung des *CIELAB*-Raum zur Farbdefinition für die Farbpaletten nahe. Der Farbraum ist in Bezug auf die Menschliche Farbwahrnehmung gleichförmig. Das heißt, dass die euklidische Distanz zwischen zwei Farben ihrer wahrgenommenen Unterschiedlichkeit entsprechen. Da die Farbdefinition im *CIELAB*-Raum jedoch auf für den Menschen unintuitiven Parametern beruht, wird weiterhin der *LCH*-Raum verwendet, der zylindrischen Transformation des *CIELAB*-Raumes. Dieser erlaubt die Farbdefinition auf Basis der für den Menschen intuitiveren Parameter **L**uminance (Luminanz), **C**hroma (Buntheit) und **H**ue (Farbton). Dies Erleichtert die Erfüllung der

¹Die Pain Scale N-PASS bildet hier eine Ausnahme, da sie die Pain Score -2 und -1 definiert. Da diese jedoch die den Grad der Beruhigung diagnostizieren, können sie nur durch die manuelle Zuführung eines Schmerzstimulus festgestellt werden.

Farbsemantik, da sich die Farben Grün, Gelb und Rot in der H-Dimension in direkter Nachbarschaft befinden.

zur Zusammenstellung der konkreten Farbpaletten wird das Unterstützungswertkeuzug von David Johnstone verwendet.² Das Tool erlaubt die Wahl von n Farben LCH-Raum, bezeichnet als „Fixpunkte“. Auf Basis dieser Fixpunkte generiert das Tool eine Farbpalette mit m Farben. Der erste und der letzte Fixpunkt definieren die erste und die letzte Farbe der Farbpalette. Ist $n = m$, so entspricht die Farbpalette den definierten Fixpunkten. Ist $m > n$, so interpoliert das Tool zwischen den Fixpunkten, um die Zwischenfarben zu finden. Findet diese Interpolation im LCH-Raum statt, so ist die visuelle Gleichabständigkeit der interpolierten Farben zu den benachbarten Fixpunkten gewährleistet.

Ein „reines Rot“, codiert im RGB-Raum mit $[255, 0, 0]$, hat im LCH-Raum die Koordinaten $[53, 105, 40]$. „Reines grün“ hat die Koordinaten $\text{RGB} = [0, 255, 0]$, was $\text{LCH} = [88, 120, 136]$ entspricht. Das „Leicht rötliche Gelb“ definiert mit $\text{RGB} = [255, 240, 0] \hat{=} \text{LCH} = [93, 93, 99]$. Diese Farben haben im LCH Raum eine unterschiedliche Buntheit, weshalb sie in einer Farbpalette eine unterschiedliche Wichtigkeit suggerieren würden. Da wurde der Chroma-Wert aller drei Fixpunkte auf den niedrigsten der drei Werte, 93, gesetzt. Die tatsächlichen Parameter der drei Fixpunkte *Rot*, *Gelb* und *Grün* werden in Tabelle 4.2 für den RGB- und den LCH-Farbraum definiert.

Tabelle 4.2: Fixpunkte als Basis der Farbpaletten

	RGB			LCH		
	R	G	B	L	C	H
Rot	243	45	22	53	93	40
Gelb	255	240	0	93	93	99
Grün	110	250	86	88	93	136

Abbildung 4.2 zeigt die Farbpaletten, die mit Hilfe des Tools für zwei bis sieben Farben erstellt wurde. Bei der Farbpalette mit ungerader Farbanzahl befinden sich die Fixpunkte an erster, letzter und mittlerer Position, während die zusätzlichen Farben durch Interpolation der benachbarten Farben erzeugt wurden. Bei den Farbpaletten mit gerader Anzahl an Farben ergibt sich das Problem, dass der gelbe Fixpunkt nicht mehr mit enthalten ist.

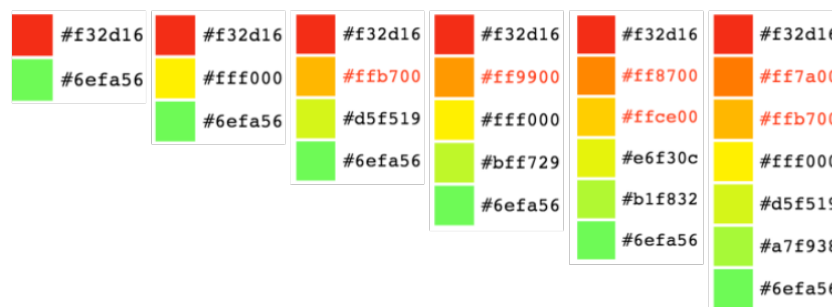


Abbildung 4.2: Erstellte Farbpaletten mit 2, 3, 4 und 5 Farben inklusive der zugehörigen RGB-Koordinaten im Hex-Code.

Auf Basis die Farbpaletten wurden die Farbschemen P_{Scale} definiert, welche die Region-Scores

²<http://davidjohnstone.net/pages/lch-lab-colour-gradient-picker>

5 Zusammenfassung

Literaturverzeichnis

- [1] Tobias Kaufmann Beat Pfister. *Sprachverarbeitung*. Springer, Berlin, 2008.
- [2] Arthur H Benade. *Fundamentals of Musical Acoustics*. 1976.
- [3] Judy Bildner. *CRIES Instrument Assessment Tool of Pain in Neonates*. City of Hope Pain, 1997. Online unter <http://prc.coh.org/pdf/CRIES.pdf>.
- [4] Richard Brown. The short time fourier transform, 2014. Online erhältlich unter: http://spinlab.wpi.edu/courses/ece503_2014/12-6stft.pdf.
- [5] R Sisto & Giuseppe Buonocore Carlo Bellieni, Franco Bagnoli. Cry features reflect pain intensity in term newborns: An alarm threshold. *Pediatric Research*, 5:142–146, 1. Online unter https://www.researchgate.net/publication/297827342_Cry_features_reflect_pain_intensity_in_term_newborns_An_alarm_threshold.
- [6] Rami Cohen and Yizhar Lavner. Infant Cry Analysis and Detection. In *27th Convention of Electrical and Electronics Engineers in Israel*. IEEE, 2012. Online unter https://www.researchgate.net/publication/261116332_Infant_cry_analysis_and_detection.
- [7] Alin Dobra. Introduction to classification and regression, 2005. Online erhältlich unter: <https://www.cise.ufl.edu/~adobra/datamining/classif-intro.pdf>.
- [8] H. Hollien & T Murry E Müller. Perceptual responses to infant crying: identification of cry types. *Journal of Child Language*, 1(1):89–95, 1974. Online unter <https://www.cambridge.org/core/journals/journal-of-child-language/article/perceptual-responses-to-infant-crying-identification-of-cry-types/4F0F8088116FCE381851D8D560697A5F>.
- [9] Jan Hamers Eva Cignac, Romano Mueller and Peter Gessler. Pain assessment in the neonate using the Bernese Pain Scale for Neonates. *Early Human Development*, 78(2):125–131, 2004. Online unter https://www.researchgate.net/publication/8485535_Pain_assessment_in_the_Neonate_using_the_Bernese_Pain_Scale_for_Neonates.
- [10] Barbara Fuller. Acoustic Discrimination of three Cry Types. *Nursing Research*, 40(3), 1991. Online erhältlich unter: https://www.researchgate.net/publication/21125005_Acoustic_Discrimination_of_Three_Types_of_Infant_Cries.
- [11] Trevor Hastie Gareth James, Daniela Witten and Robert Tibshirani. *An Introduction to Statistical Learning*. Springer, 2013.
- [12] Dmitry Goldgof Rangachar Kasturi Yu Sun Ghada Zamzmi, Chih-Yun Pai and Terri Ashmeade. Machine-based Multimodal Pain Assessment Tool for Infants: A Review, 2016. Online unter <https://arxiv.org/ftp/arxiv/papers/1607/1607.00331.pdf>.
- [13] Ricardo Gutierrez-Osuna. Introduction to Speech Processing. Online unter http://courses.cs.tamu.edu/rgutier/csce689_s11/.

- [14] Health Facts For You. *Using Pediatric Pain Scales Neonatal Infant Pain Scale (NIPS)*, 2014. Online unter <https://www.uwhealth.org/healthfacts/parenting/7711.pdf> und unter <https://com-jax-emergency-pami.sites.medinfo.ufl.edu/files/2015/02/Neonatal-Infant-Pain-Scale-NIPS-pain-scale.pdf>.
- [15] Hodgkinson. Neonatal Pain Assessment Tool, 2012. Online unter http://www.rch.org.au/rchcpg/hospital_clinical_guideline_index/Neonatal_Pain_Assessment/#The%20Pain%20Assessment%20Tool.
- [16] Michael J Corwin Howard L Golub. A Physioacoustic Model of the Infant Cry. In *Infant Crying - Theoretical and Research Perspectives*, chapter 3, pages 59 – 82. Plenum, 1985.
- [17] Bonnie Stevens Huda Huijer Abu-Saad, Gerrie Bours and Jan Hamers. Assessment of pain in Neonates. *Seminars in Perinatology*, 2(5):402–416, 1998. Online unter <https://www.ncbi.nlm.nih.gov/pubmed/9820565>.
- [18] Donna Geiss Laura Wozniak & Charles Hall Ivan Hand, Lawrence Noble. COVERS Neonatal Pain Scale: Development and Validation. *International Journal of Pediatrics*, 2010, 2010. Online unter <https://www.hindawi.com/journals/ijpedi/2010/496719/>.
- [19] Bonnie J. Stevens K. J. S. Anand and Patrick J. McGrath. *Pain in Neonates and Infants*. Elsevier, 2007.
- [20] Miroslav Kubat. *An Introduction to Machine Learning*. Springer, 2015.
- [21] Barry Lester and Zachariah Boukydis. *Infant Crying: Theoretical and Research Perspectives*. Springer, 1985.
- [22] A. Rebecca Neal Linda L. LaGasse and Barry M. Lester. Assessment of infant cry: Acoustic cry analysis and parental perception. *Mental retardation and developmental disabilities*, 11(1):83–93, 2005. Online unter <https://www.ncbi.nlm.nih.gov/pubmed/15856439>.
- [23] Tze-Wey Loong. Understanding sensitivity and specificity with the right side of the brain. *BMJ*, 327(7417), 2003. Online unter <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC200804/>.
- [24] Michael Lutter. Speech production, 2015. Online erhältlich unter: <http://recognize-speech.com/speech/speech-production>.
- [25] Robert Mannell. Acoustic theory of speech production, 2015. Online erhältlich unter: http://clas.mq.edu.au/speech/acoustics/frequency/acoustic_theory.html.
- [26] Stephen Marsland. *Machine Learning - An Algorithmic Perspective*. Chapman & Hall / CRC, 2009.
- [27] Tom M Mitchell. *Machine Learning*. WCB McGraw-Hill, 1997.
- [28] Hans M Koot Dick Tibboel Jan Passchier & Hugo Duivenvoorden Monique van Dijk, Josien de Boer. The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain*, 84(2):367—377, 2000. Online unter <http://www.sciencedirect.com/science/article/pii/S0304395999002390> und unter.
- [29] Sinno Simons Monique van Dijk and Dick Tibboel. Pain assessment in neonates. *Paediatric and Perinatal Drug Therapy*, 6(2):97–103, 2004. Online unter <http://www>.

- sciencedirect.com/science/article/pii/S0304395999002390.
- [30] D L Neuhoff. *Signal and Systems I - EECS 206 Laboratory*. The University of Michigan, 2002. Online erhältlich unter: <http://www.eecs.umich.edu/courses/eecs206/archive/spring02/> abgerufen am 11. Januar 2016.
 - [31] Katarina Michelsson Ole Wasz-Hockert and John Lind. Twenty-Five Years of Scandinavian Cry Research. In *Infant Crying - Theoretical and Research Perspectives*, chapter 3, pages 59 – 82. Plenung, 1985.
 - [32] J L Mathew P J Mathew. Assessment and management of pain in infants. *Postgrad Med J*, 79:438–443, 2003. Online unter <http://pmj.bmj.com/content/79/934/438.full>.
 - [33] Steven Creech Patricia Hummel, Mary Puchalski and Marc Weiss. N-PASS: Neonatal Pain, Agitation and Sedation Scale – Reliability and Validity. *Pediatrics/Neonatology*, 2(6), 2004. Online unter <http://www.anesthesiarianimazione.com/2004/06c.asp>.
 - [34] Susan Parker-Price & Ronald Barr Philip Zeskind. Rythmic organization of the Sound of Infant Cry. *Dev Psychobiol*, 26(6):321–333, 1993. Online unter <https://www.ncbi.nlm.nih.gov/pubmed/8119482>.
 - [35] R Ward & C Laszlo Qiaobing Xie. Automatic Assessment of Infants Levels-of-Distress from the Cry Signals. *IEEE Transanctions on Speech and Audio Processing*, 4(4):253–265, 1996. Online unter <http://ieeexplore.ieee.org/document/506929/>.
 - [36] Brian Hopkins & James Green Ronald Barr. *Crying as a Sign, a Symptom, and a Signal*. Mac Keith Press, 2000.
 - [37] J R Shayevitz & Shobha Malviya Sandra Merkel, Terri Voepel-Lewis. The FLACC: A Behavioral Scale for Scoring Postoperative Pain in Young Children. *Pediatric Nursing*, 23(3):293–7, 1996. Online unter https://www.researchgate.net/publication/13998379_The_FLACC_A_Behavioral_Scale_for_Scoring_Postoperative_Pain_in_Young_Children.
 - [38] Julius Smith. *Spectral Audio Signal Processing*. Center for Computer Research in Music and Acoustics (CCRMA), 1993. Online unter https://www.dsprelated.com/freebooks/sasp/Short_Time_Fourier_Transform.html.
 - [39] Steven W. Smith. *The Scientist and Engineer’s Guide to Digital Signal Processing*. California Technical Publishing, 1999. Online erhältlich unter: <http://www.dspguide.com/pdfbook.htm>.
 - [40] Gyorgy Ivan Varallyay. *Analysis of the Infant Cry with Objective Methods*. PhD thesis, Budapest University of Technology and Economics, 2009. Online erhältlich unter: <https://pdfs.semanticscholar.org/5c38/b368dc71d67cbfab3077a50536b086d8eec.pdf>.
 - [41] P H Wolff. The role fo biological rhythms in early psychological development. *Bulletin of the Menninger Clinic*, 31(1):197–218, 1967.

Appendices

A Auswertungsdetails VAD

Tabelle A.1: Klassifikationsgenauigkeiten der durch den REPTree entworfenen Modelle

$SNR_{Training}$	3 dB				50 dB				50+3 dB			
SNR_{Test}	3 dB	50 dB	7 dB*	Mean	3 dB	50 dB	7 dB*	Mean	3 dB	50 dB	7 dB*	Mean
Zeit	77.81%	79.02%	86.04%	80,96%	49.33%	94.70%	48.66%	64,23%	77.54%	92.47%	84.38%	84,80%
Spektrum	82.05%	89.28%	82.71%	84,68%	70.52%	94.37%	55.06%	73,31%	81.75%	91.22%	74.90%	82,62%
Cepstrum	88.98%	94.72%	92.96%	92,22%	86.83%	94.68%	92.83%	91,45%	88.98%	94.72%	92.96%	92,22%
Autokorr.	80.45%	73.47%	84.89%	79,60%	73.07%	87.14%	77.98%	79,39%	77.90%	84.88%	82.84%	81,87%
Zeit+Spektrum	82.05%	89.28%	82.71%	84,68%	70.52%	94.37%	55.06%	73,31%	81.75%	91.22%	74.90%	82,62%
Zeit+Cepstrum	88.98%	94.72%	92.96%	92,22%	86.83%	94.68%	92.83%	91,45%	88.98%	94.72%	92.96%	92,22%
Zeit+Autokorr	80.45%	73.47%	84.89%	79,60%	49.33%	94.70%	48.66%	64,23%	80.32%	92.35%	88.22%	86,96%
Spektrum+Ceps	88.98%	94.72%	92.96%	92,22%	70.65%	94.75%	55.06%	73,49%	88.98%	94.72%	92.96%	92,22%
Spektrum+Autokorr.	82.05%	89.28%	82.71%	84,68%	70.52%	95.60%	95.60%	87,24%	81.75%	94.42%	74.90%	83,69%

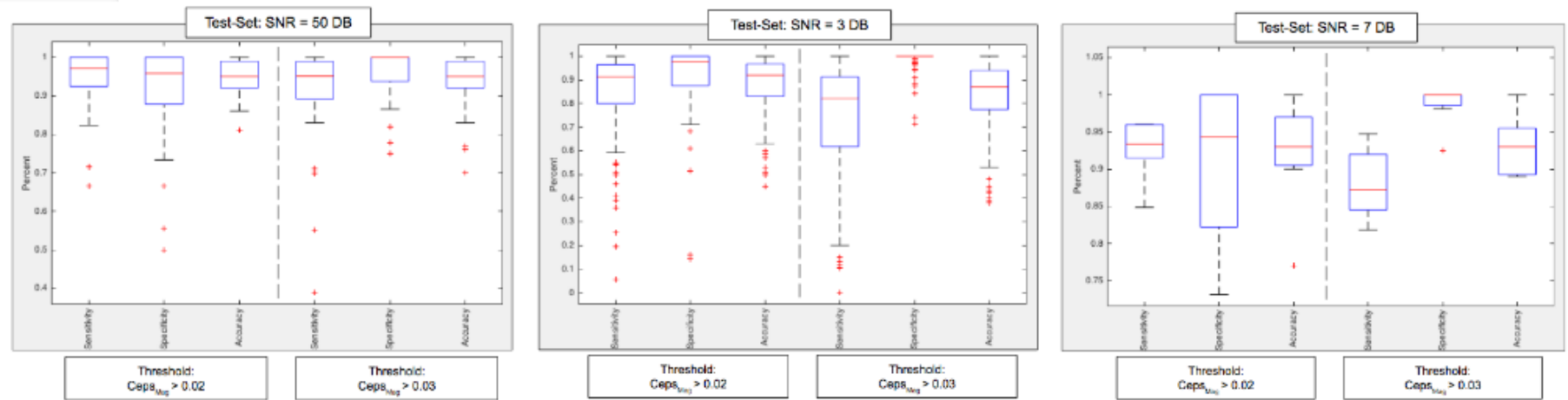


Abbildung A.1: Boxplot-Auswertung über Sensitivität, Spezifität und Genauigkeit der beiden besten Modelle