

Wissensbasierte Modellierung von Audio-Signal-Klassifikatoren

Zur Bioakustik von *Tursiops truncatus*

Dissertation zur Erlangung des Doktorgrades der Philosophie

von

Sebastian Hübner

Eingereicht bei der humanwissenschaftlichen Fakultät
der Universität Potsdam

Potsdam, im August 2006

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Universitätsverlag Potsdam 2008

<http://info.ub.uni-potsdam.de/verlag.htm>

Universitätsverlag Potsdam, Am Neuen Palais 10, 14469 Potsdam

Tel.: +49 (0)331 977 4623 / Fax: 4625

e-mail: ubpub@uni-potsdam.de

Dieses Werk ist urheberrechtlich geschützt.

2., überarbeitete Auflage

Elektronisch veröffentlicht auf dem

Publikationsserver der Universität Potsdam

URL <http://pub.ub.uni-potsdam.de/volltexte/2008/1663/>

URN [urn:nbn:de:kobv:517-opus-16631](http://nbn-resolving.org/urn:nbn:de:kobv:517-opus-16631)

[<http://nbn-resolving.org/urn:nbn:de:kobv:517-opus-16631>]

Danksagung

Der Autor möchte sich an dieser Stelle bei allen bedanken, die geholfen haben, die vorliegende Arbeit Wirklichkeit werden zu lassen.

Ohne die geduldige Unterstützung des gesamten Projektes durch meinen Doktorvater Prof. Dr. James Douglas Saddy (University of Reading UK) wäre ein Zustandekommen dieser Arbeit nicht möglich gewesen. Prof. Dr. Dietmar Todt und Dr. Frank Veit (beide Freie Universität Berlin) verdanke ich die einzigartige Möglichkeit, im Reeflab in Eilat (Israel) Aufnahmen der akustischen Signale des Großen Tümmlers herzustellen. Auch gilt mein Dank den Gutachtern Frau Prof. Dr. Marianne Vater und Herrn Prof. Dr. Joachim Selbig (beide Universität Potsdam) sowie dem Vorsitzenden der Prüfungskommission Herrn Prof. Dr. Ralf Engbert (Universität Potsdam).

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der wissensbasierten Modellierung von Audio-Signal-Klassifikatoren (ASK) für die Bioakustik. Sie behandelt ein interdisziplinäres Problem, das viele Facetten umfasst. Zu diesen gehören artspezifische bioakustische Fragen, mathematisch-algorithmische Details und Probleme der Repräsentation von Expertenwissen.

Es wird eine universelle praktisch anwendbare Methode zur wissensbasierten Modellierung bioakustischer ASK dargestellt und evaluiert. Das Problem der Modellierung von ASK wird dabei durchgängig aus KDD-Perspektive (Knowledge Discovery in Databases) betrachtet. Der grundlegende Ansatz besteht darin, mit Hilfe von modifizierten KDD-Methoden und Data-Mining-Verfahren die Modellierung von ASK wesentlich zu erleichtern.

Das etablierte KDD-Paradigma wird mit Hilfe eines detaillierten formalen Modells auf den Bereich der Modellierung von ASK übertragen. Neunzehn elementare KDD-Verfahren bilden die Grundlage eines umfassenden Systems zur wissensbasierten Modellierung von ASK. Methode und Algorithmen werden evaluiert, indem eine sehr umfangreiche Sammlung akustischer Signale des Großen Tümmlers mit ihrer Hilfe untersucht wird. Die Sammlung wurde speziell für diese Arbeit in Eilat (Israel) angefertigt. Insgesamt werden auf Grundlage dieses Audiomaterials vier empirische Einzelstudien durchgeführt:

1. Auf der Basis von oszillographischen und spektrographischen Darstellungen wird ein phänomenologisches Klassifikationssystem für die vielfältigen Laute des Großen Tümmlers dargestellt.
2. Mit Hilfe eines Korpus halbsynthetischer Audiodaten werden verschiedene grundlegende Verfahren zur Modellierung und Anwendung von ASK in Hinblick auf ihre Genauigkeit und Robustheit untersucht.
3. Mit einem speziell entwickelten Clustering-Verfahren werden mehrere Tausend natürliche Pfifflaute des Großen Tümmlers untersucht. Die Ergebnisse werden visualisiert und diskutiert.
4. Durch maschinelles mustererkennungsbasiertes akustisches Monitoring wird die Emissionsdynamik verschiedener Lauttypen im Verlaufe von vier Wochen untersucht. Etwa 2.5 Millionen Klicklaute werden im Anschluss auf ihre spektralen Charakteristika hin untersucht.

Die beschriebene Methode und die dargestellten Algorithmen sind in vielfältiger Hinsicht erweiterbar, ohne dass an ihrer grundlegenden Architektur etwas geändert werden muss. Sie lassen sich leicht in dem gesamten Gebiet der Bioakustik einsetzen. Hiermit besitzen sie auch für angrenzende Disziplinen ein hohes Potential, denn exaktes Wissen über die akustischen Kommunikations- und Sonarsysteme der Tiere wird in der theoretischen Biologie, in den Kognitionswissenschaften, aber auch im praktischen Naturschutz, in Zukunft eine wichtige Rolle spielen.

Abkürzungen

KDD	Knowledge Discovery in Databases
KDSE	Knowledge Discovery Support Environment
ASK	Audio-Signal-Klassifikator(en)
FLC	First-Level-Klassifikator(en)
SLC	Second-Level-Klassifikator(en)
SSDM	Symbolic-Similarity-Driven-Merge
FLCD	First-Level-Classifer-Discovery
SLCD	Second-Level-Classifer-Discovery
P-Laut	Pfiff laut
BP-Laut	Burst-Pulsed-Laut
NFP-Laut	Niederfrequenter pulsierter Laut
NFNP-Laut	Niederfrequenter nichtpulsierter Laut
NF-Klicklaut	Niederfrequenter Klicklaut
MF-Klicklaut	Klicklaut im mittleren Frequenzbereich
FM	Frequenzmodulation
AM	Amplitudenmodulation
PM	Phasenmodulation
PFM	Pulsfrequenz-Modulation
PAM	Pulsamplituden-Modulation
PPM	Pulsphasen-Modulation
PDM	Pulsdauer-Modulation
PSM	Pulsspektrum-Modulation

Inhaltsverzeichnis

Einleitung	1
I Wissensbasierte Modellierung von Audio-Signal-Klassifikatoren	9
1 Interdisziplinärer Kontext	11
1.1 KDD und Data-Mining	11
1.1.1 Hintergründe und Ziele des KDD	11
1.1.2 Konzepte des KDD	12
1.1.3 Der KDD-Prozess	14
1.1.4 Data-Mining-Aufgaben	17
1.1.5 Data-Mining-Verfahren	18
1.1.6 Aktuelle Forschungsthemen	19
1.2 KDD in der Bioakustik	21
1.2.1 Bioakustik	21
1.2.2 Perspektiven für KDD-Methoden	32
2 KDD und Audio-Signal-Klassifikatoren	35
2.1 Manuelle versus maschinelle Annotation	35
2.2 Unterschiede zur Spracherkennung	36
2.3 Das Annotationsproblem der Bioakustik	37
2.4 Maschinelle Annotation von Audiodaten	40
2.4.1 Grundlegende Konzepte	41
2.4.2 First-Level-Klassifikatoren	41
2.4.3 Second-Level-Klassifikatoren	43
2.4.4 Zusammenfassung	45
2.5 Der FLCD-Prozess	46
2.6 Der SLCD-Prozess	49
2.7 Modellierungs-Aufgaben	51
3 Modellierung von Audio-Signal-Klassifikatoren	53
3.1 Vorbemerkung	53
3.2 Spektrographische Visualisierung	54

3.3	Modellierung spektrographischer FLC	56
3.3.1	Spektrographische Signaturen	57
3.3.2	KKF für spektrographische Signaturen	59
3.3.3	Algorithmische Annotation	61
3.3.4	Manuelle Programmierung von FLC	62
3.3.5	Notation und Visualisierung	63
3.3.6	Ausdünnen von Merkmalen	63
3.3.7	Klonen von Signaturen	64
3.3.8	Automatische Extraktion spektrographischer Signaturen . . .	65
3.3.9	Verschmelzen von spektrographischen Signaturen	67
3.3.10	Ähnlichkeitsbasiertes Verschmelzen spektrographischer Signa- turen	69
3.3.11	Induktive Programmierung	72
3.3.12	Diskussion	76
3.4	Modellierung von SLC	76
3.4.1	Modellierung symbolischer Signaturen durch Interpretation . .	77
3.4.2	Algorithmische Annotation	80
3.4.3	Notation und Visualisierung	82
3.4.4	Automatische Extraktion von symbolischen Signaturen	84
3.4.5	Verschmelzen symbolischer Signaturen	86
3.4.6	Ähnlichkeitsbasiertes Verschmelzen symbolischer Signaturen .	89
3.4.7	Induktive Programmierung	92
3.4.8	Diskussion	94

II Zur Bioakustik von *Tursiops truncatus* 95

4	Zur Bioakustik von <i>Tursiops truncatus</i>	97
4.1	Der Große Tümmler	97
4.1.1	Taxonomie und Abstammung	97
4.1.2	Lebensweise	98
4.1.3	Sinne und Gehirn	98
4.2	Das Eilat-Korpus	99
4.2.1	Aufnahmetechnik	99
4.2.2	Randbedingungen	101
4.3	Akustische Phänomene im Eilat-Korpus	103
4.3.1	Ziele	103
4.3.2	Material und Methode	103
4.4	Schwingungstypen und Grundtypen	104
4.4.1	Schwingungstypen	104
4.4.2	Grundtypen	105
4.4.3	Zusammenfassung	113
4.5	Phonetische Modulation	118

4.5.1	Allgemeine Prinzipien	118
4.5.2	Modulation und Kodierung	122
4.5.3	Spezielle Prinzipien	124
4.5.4	Phonetische Modulation bei <i>Tursiops truncatus</i>	125
4.5.5	Zusammenfassung	135
5	Klassifikation synthetischer Pfiff-laute	139
5.1	Die Signaturpiff-Hypothese	139
5.2	Fragestellung	140
5.3	Methodische Vorbemerkungen	141
5.4	Audiomaterialien	141
5.5	Robustheit von FLC und SNR	145
5.6	Überlagerungen von Pfiff-laute	145
5.7	Partiell instanziierte Pfiff-laute	147
5.8	Verschiebungen der Grundfrequenz	148
5.9	Induktion von FLC für Pfiff-laute	149
5.10	Diskussion	151
6	Pfiff-typen im Eilat-Korpus	153
6.1	Fragestellung	153
6.2	Materialien	154
6.3	Methode	154
6.4	Ergebnisse	158
6.4.1	Die Menge Φ_{10}^{AEK}	160
6.4.2	Die Menge Φ_{35}^{AEK}	163
6.4.3	Die Menge Φ_{10}^{EK}	166
6.4.4	Die Menge Φ_{35}^{EK}	169
6.4.5	Automatische Extraktion	172
6.4.6	Verhalten des SSDM-Algorithmus	173
6.4.7	Eigenschaften verschmolzener Signaturen	173
6.5	Diskussion	174
7	Zur Natur von MF-Klick-laute	177
7.1	Fragestellung	177
7.2	Materialien	178
7.3	Methode	182
7.4	Ergebnisse	184
7.5	Diskussion	184
8	Schluss	195
A	KDSE-Instrumente in DSPROLOG	209

Abbildungsverzeichnis

1.1	Der KDD-Prozess	15
2.1	Annotation von Audiodaten	38
2.2	Annotation mit FLC und SLC	46
2.3	Der FLCD-Prozess	47
2.4	Der SLCD-Prozess	49
3.1	Graustufenpalette und kontrastoptimierte Farbpalette I	56
3.2	Graustufenpalette und kontrastoptimierte Farbpalette II	57
3.3	Generieren einer spektrographischen Signatur	60
3.4	Arbeitsweise des Annotationsalgorithmus	62
3.5	Visualisierung einer Signatur	63
3.6	Visuelles Feedback	64
3.7	Ausdünnen von Merkmalen	65
3.8	Klonen von Signaturen	66
3.9	Annotationsentscheidungen geklonter FLC	66
3.10	FLC mit Suchraum	68
3.11	Annotationen	68
3.12	Verschmelzen von Signaturen I	70
3.13	Verschmelzen von Signaturen II	70
3.14	Ähnlichkeitsbasiertes Verschmelzen von Signaturen	73
3.15	Induktive Programmierung von FLC	75
3.16	Modellierung durch Interpretation	79
3.17	Textuelle Notation eines SLC.	83
3.18	Visualisierung von Konstituenten, Hierarchie und Toleranz in einem SLC	83
3.19	Visualisierung von SLC-Annotationen	84
3.20	Automatische Extraktion von Φ_{β}^S	87
3.21	Verschmelzen von symbolischen Signaturen	89
3.22	Vergleich zweier Signaturen	91
3.23	Visualisierung und Bereinigung symbolischer Signaturen	93
4.1	Aufnahmezeiten im Eilat-Korpus	100
4.2	Oszillogramm typischer Unterwassergeräusche in Eilat	102

4.3	Spektrogramm typischer Unterwassergeräusche in Eilat	102
4.4	Beispielaufnahme	103
4.5	Sinusartiger Schwingungstyp oszillographisch	106
4.6	Pulsierter Schwingungstyp oszillographisch	107
4.7	Klick-Schwingungstyp oszillographisch	108
4.8	Sequenz von Pfifflauten	109
4.9	Spektrogramm eines Pfifflauts mit max. 32,8 kHz	110
4.10	Niederfrequente nichtpulierte Laute	111
4.11	Sequenz von niederfrequenten pulsierten Lauten	112
4.12	Dominante Frequenzen von NFP-Lauten	113
4.13	Sequenz von Burst-Pulsed-Lauten	114
4.14	Sequenz von niederfrequenten Klicklauten	115
4.15	Sequenz von mittelfrequenten Klicklauten	116
4.16	Klassifikation akustischer Phänomene	117
4.17	Modulation in der Nachrichtentechnik	118
4.18	Phonetische Modulation in der Bioakustik	120
4.19	Spektrogramm eines idealen diskreten pulsierten Signals	121
4.20	Allgemeine Prinzipien der phonetischen Modulation	123
4.21	Glissando und Harmonische in einem Pfifflaut	126
4.22	Pfifflaut ohne Harmonische	127
4.23	Vibrato in einem Pfifflaut (Detail)	127
4.24	Seitenband in einem Pfifflaut	128
4.25	Zischen in einem Pfifflaut	128
4.26	Frequenzsprünge in einem Pfifflaut	129
4.27	NFNP-Laut mit Formanten	129
4.28	NFNP-Laut ohne Formanten	130
4.29	Sequenz von NFP-Lauten (Übersicht)	131
4.30	Sequenz von NFP-Lauten (Detail)	131
4.31	Pulsfrequenz-Glissando bei NFP-Laut	132
4.32	Sequenz von Burst-Pulsed-Lauten	133
4.33	NF-Klicklaute 1	134
4.34	NF-Klicklaute 2	134
4.35	MF-Klicklaute 1	135
4.36	MF-Klicklaute 2	136
4.37	MF-Klicklaute 3	136
4.38	MF-Klicklaute 4	137
5.1	Künstliche Pfifflaute	142
5.2	Überlagerte künstliche Pfifflaute	144
5.3	Weglassungen in künstlichen Pfifflaute	144
5.4	Performanz von FLC gegenüber Änderungen der SNR	146
5.5	Performanz von FLC bei Überlagerungen	147
5.6	Performanz bei partiell gelöschten Pfifflauten	148

5.7	Performanz bei Frequenzabweichungen	150
5.8	Die Menge K	151
6.1	Annotation unter schwierigen Bedingungen	155
6.2	Clustering synthetischer Pfiffklaute I	160
6.3	Streudiagramme 1.1	160
6.4	Streudiagramme 1.2	161
6.5	Streudiagramme 1.3	161
6.6	Streudiagramme 1.4	162
6.7	Signaturen synthetischer Pfiffklaute	162
6.8	Clustering synthetischer Pfiffklaute II	163
6.9	Streudiagramme 2.1	163
6.10	Streudiagramme 2.2	164
6.11	Streudiagramme 2.3	164
6.12	Streudiagramme 2.4	165
6.13	Signaturen synthetischer Pfiffklaute	165
6.14	Clustering natürlicher Pfiffklaute I	166
6.15	Streudiagramme 3.1	166
6.16	Streudiagramme 3.2	167
6.17	Streudiagramme 3.3	167
6.18	Streudiagramme 3.4	168
6.19	Fluktuationen	168
6.20	Clustering natürlicher Pfiffklaute II	169
6.21	Streudiagramme 4.1	169
6.22	Streudiagramme 4.2	170
6.23	Streudiagramme 4.3	170
6.24	Streudiagramme 4.4	171
6.25	Signaturen natürlicher Pfiffklaute	171
6.26	Symbolische Signaturen verschiedener Signaturpfeife	172
7.1	FLC-Modellierung I	180
7.2	FLC-Modellierung II	181
7.3	FLC-Modellierung III	182
7.4	MF-Klicklaute im Eilat-Korpus I	185
7.5	MF-Klicklaute im Eilat-Korpus II	186
7.6	MFK-Trains im Eilat-Korpus	187
7.7	Burst-Pulsed-Laute im Eilat-Korpus	188
7.8	Elemente in F^{EK}	189
7.9	Pfiffklaute im Eilat-Korpus	190
7.10	Okkurrenzen von zwei verschiedenen Pfiffklauten	191
7.11	Dominante Frequenzen in MF-Klicklauten I	193
7.12	Dominante Frequenzen in MF-Klicklauten II	194

Tabellenverzeichnis

4.1	Vokalisations-Grundtypen im Bereich von 0-48 kHz.	117
4.2	Phonetische Modulation im Bereich von 0-48 kHz.	137
5.1	Klassifikationsvariablen	141
6.1	Streuverhalten I	173
6.2	Streuverhalten II	173
7.1	Performanz von nicht-optimiertem und optimiertem FLC	181
A.1	KDSE-Instrumente in DSPROLOG	209

Einleitung

Motivation und Ziele der Arbeit

Akustische Signalsysteme im Tierreich können außerordentlich vielschichtig und komplex sein. Viele Tierarten, darunter Insekten, Amphibien, Singvögel, Meeres-säuger und Primaten, kommunizieren mit Hilfe von mannigfaltigen, kompliziert aufgebauten Lauten und Lautsequenzen. Nicht wenige Arten, darunter Fledermäuse und Zahnwale, haben hochspezifische biologische Echolokationssysteme.

Empirische bioakustische Studien besitzen in der theoretischen Biologie, in den Kognitionswissenschaften und auch im praktischen Naturschutz ein hohes Potential. Um in diesen Bereichen statistisch relevante Aussagen zu erzielen, müssen oftmals umfangreiche Mengen von Audiodaten ausgewertet werden. Dabei treten in der Regel beträchtliche Schwierigkeiten auf, da das Audiomaterial klassifiziert und bewertet werden muss. Sind sehr große Datenvolumina im Spiel, ist eine manuelle Auswertung in akzeptabler Zeit nur schwer möglich. Erst eine automatische oder semiautomatische Klassifikation und Annotation der Audiodaten kann hier Abhilfe schaffen.

Es verwundert nicht, dass in der Bioakustik und in verwandten Disziplinen ein großes Interesse an Verfahren zur Datenanalyse mit Hilfe von **Audio-Signal-Klassifikatoren** (im Folgenden ASK) existiert. Die Schwierigkeiten bei der Herstellung bioakustischer ASK sind allerdings so groß, dass bisher nur sehr wenige Systeme zur automatischen Klassifikation von Tierlauten entwickelt wurden. Die allermeisten davon befinden sich gegenwärtig noch im Laborstadium.

Die Entwicklung bioakustischer ASK ist ein nicht-trivialer Prozess, da die physikalischen und biologischen Realitäten akustischer Kommunikation und Echolokation in der Tierwelt außerordentlich komplex sind. In letzter Konsequenz steht die Wissenschaft hier vor einer immensen Aufgabe: Möchte man bioakustische Prozesse in natürlichen Biotopen automatisiert erfassen und auswerten, so können nicht nur einzelne Arten berücksichtigt werden. Es bedarf vielmehr effizienter ASK für die akustischen Signalsysteme tausender Spezies in weitverzweigten Ökosystemen.

Das Problem der Entwicklung von effizienten bioakustischen ASK umfasst viele Facetten. Zu diesen gehören:

1. Biologische und ökologische Fragen,
2. artspezifische bioakustische Details,

3. mathematisch-algorithmische Probleme und
4. Fragen der Repräsentation von Expertenwissen.

In der Regel handelt es sich hierbei um eine hochgradig wissensintensive Aufgabe. Bei der Modellierung bioakustischer ASK spielen nicht nur mathematisch-algorithmische Aspekte eine Rolle. Fast immer muss biologisches und bioakustisches *Expertenwissen* in den Modellierungsprozess miteinbezogen werden.

Eine wichtige Voraussetzung für zukünftige Fortschritte in diesem Bereich liegt daher darin, eine universelle praktisch anwendbare Methode zur *wissensbasierten* Modellierung bioakustischer ASK zu schaffen. In dieser müssen alle Facetten des oben genannten Problems explizit berücksichtigt werden. Gleichzeitig bedarf es algorithmischer Werkzeuge, welche gestatten alle Einzelschritte dieser Methode auf dem Rechner auszuführen.

Im Fehlen einer solchen Methode und im Fehlen geeigneter Werkzeuge zur wissensbasierten Modellierung von ASK liegt ein wesentlicher Grund für den bisherigen Mangel an effizienten Systemen zur maschinellen Klassifikation bioakustischer Audiodaten. In diesem Kontext verfolgt die vorliegende Arbeit folgende Hauptziele:

1. Die Schaffung methodischer Grundlagen für die wissensbasierte Modellierung von bioakustischen ASK;
2. Die Beschreibung grundlegender Algorithmen, mit deren Hilfe die wissensbasierte Modellierung effizienter bioakustischer ASK im praktischen Anwendungsfalle erfolgen kann;
3. Eine Evaluation der erarbeiteten methodischen und algorithmischen Werkzeuge anhand von synthetischen und natürlichen bioakustischen Daten.
4. Eine empirische Untersuchung von ausgewählten Phänomenen, Häufigkeiten, Verteilungen und Zeit-Frequenz Charakteristika von Sonar- und Kommunikationslauten der Spezies *Tursiops truncatus* mit Hilfe der erarbeiteten methodischen und algorithmischen Werkzeuge;

Ansatz und Aufbau der Arbeit

In den vergangenen zwei Jahrzehnten wurde eine Anzahl neuer Methoden zur *Wissensentdeckung* in großen Datenmengen entwickelt. Diese werden unter dem Begriff **Knowledge Discovery in Databases** (im Folgenden KDD) zusammengefasst (siehe Kapitel 1). In dieser Arbeit wird die Frage der Modellierung bioakustischer ASK aus KDD-Perspektive betrachtet. Der grundlegende Ansatz besteht darin, KDD-Methoden und Data-Mining-Verfahren so zu modifizieren, dass der Prozess der Modellierung bioakustischer ASK computergestützt und wissensbasiert erfolgen kann. Die vorliegende Arbeit ist somit nicht als ethologische Studie aufzufassen, obwohl im empirischen Teil auf die Bioakustik des Großen Tümmlers (*Tursiops truncatus*) näher eingegangen wird. Sie ist wie folgt gegliedert:

Teil I: Wissensbasierte Modellierung von Audio-Signal-Klassifikatoren

In diesem Teil werden zunächst theoretische und methodische Fragen diskutiert. Im Anschluss hieran erfolgt die detaillierte Beschreibung einer Reihe von Verfahren zur Modellierung von ASK. Teil I ist in folgende Kapitel untergliedert:

- **Kapitel 1. Interdisziplinärer Kontext:** Dieses Kapitel dient der Einordnung der vorliegenden Arbeit in den interdisziplinären Kontext und der Konkretisierung allgemeiner Grundkonzepte. Es ist in zwei Abschnitte untergliedert: Der Abschnitt „KDD und Data-Mining“ beinhaltet eine kurze Einführung in den Themenbereich des KDD und Data-Mining. Der Abschnitt „KDD in der Bioakustik“ enthält eine allgemein gehaltene Einführung in das Gebiet der Bioakustik und in damit zusammenhängende Fragestellungen. Im Anschluss wird dargestellt, welches die wichtigsten potentiellen Einsatzgebiete für KDD-Methoden in der Bioakustik sind.
- **Kapitel 2. KDD und Audio-Signal-Klassifikatoren:** In diesem Kapitel wird untersucht, wie sich der Prozess der wissensbasierten Modellierung von bioakustischen ASK im Rahmen des gängigen KDD-Paradigmas aus formaler und methodischer Perspektive darstellt.
- **Kapitel 3. Modellierung von Audio-Signal-Klassifikatoren:** In diesem Kapitel wird beschrieben, wie sich der Prozess der wissensbasierten Modellierung von ASK im Rahmen des KDD-Paradigmas aus algorithmischer Perspektive darstellt. Es wird eine Anzahl von Verfahren dargestellt, die von grundlegendem Interesse für die KDD-basierte Modellierung von ASK sind.

Teil II: Zur Bioakustik von *Tursiops truncatus* Dieser Teil dient der Evaluation der zuvor dargestellten Methoden und Verfahren anhand von Aufzeichnungen der Vokalisationen des Großen Tümmlers. Hierzu erfolgt zunächst eine Darstellung grundlegender Phänomene in den akustischen Signalen dieser Tierart. Im Anschluss daran werden ausgewählte Fragestellungen experimentell untersucht. Teil II ist in folgende Kapitel untergliedert:

- **Kapitel 4. Zur Bioakustik von *Tursiops truncatus*:** Dieses Kapitel enthält eine Beschreibung und Klassifikation grundlegender Phänomene in den akustischen Signalen des Großen Tümmlers. Die Darstellung beruht auf einem ca. 87 Gigabyte umfassenden Korpus von Audiodateien, welches eigens für diese Arbeit in Eilat (Israel) aufgezeichnet wurde (im Folgenden kurz 'Eilat-Korpus').
- **Kapitel 5. Klassifikation synthetischer Pfifflaute:** Dieses Kapitel enthält eine Evaluation der in Teil I beschriebenen Methoden zur Modellierung von ASK. Die Evaluation erfolgt anhand von synthetischen Pfifflauten, welche denen von *Tursiops truncatus* nachempfunden sind. Es wird der Frage nachgegangen, ob sich mit Hilfe der in Teil I dargestellten Verfahren ASK modellieren

lassen, die unter kontrollierten Bedingungen zu sinnvollen Ergebnissen führen. Hierbei stehen Fragen der Performanz und Stabilität im Vordergrund.

- **Kapitel 6. Pfifftypen im Eilat-Korpus:** In diesem Kapitel wird ein in Teil I beschriebenes Verfahren zur Induktion symbolischer Signaturen (Kennlinien) akustischer Ereignisse zum Clustering von natürlichen Pfifflauten eingesetzt. Ziel hierbei ist, allgemeine spektrographische Charakteristika und die Häufigkeiten von Pfifftypen im Eilat-Korpus festzustellen.
- **Kapitel 7. Zur Natur von MF-Klicklauten:** In diesem Kapitel wird eine bestimmte Klasse von Lauten von *Tursiops truncatus* in Hinblick auf ihre fundamentale biologische Funktion (Sonar oder Kommunikation) untersucht. Es handelt sich hierbei um Klicklaute im mittleren Frequenzbereich (im Folgenden auch MF-Klicklaute). Die Untersuchung erfolgt auf Grundlage der in Teil I beschriebenen KDD-Verfahren. In ihr werden mehrere Millionen Einzellaute maschinell analysiert.

Im Schlussteil wird auf einige der wichtigsten weiterführenden Forschungsfragen im Bereich der KDD-basierten Modellierung von ASK hingewiesen.

Beiträge zum wissenschaftlichen Diskurs

Im Rahmen der vorliegenden Arbeit sind auf verschiedenen Gebieten wesentliche Ergebnisse erzielt worden. Im Einzelnen sind dies:

- Das etablierte KDD-Paradigma wurde auf den Bereich der Modellierung bioakustischer ASK übertragen (siehe Abschnitte 2.5 und 2.6). Grundlage hierfür bildet ein mathematisches Modell, welches ASK auf zwei Ebenen beschreibt: First-Level-ASK und Second-Level-ASK (siehe Abschnitt 2.4). Das Modell gestattet eine Unterscheidung von mathematisch-algorithmischen Aspekten und Expertenaufgaben bei der Modellierung von ASK.
- Es werden insgesamt 19 elementare KDD-Verfahren beschrieben, welche die Grundlage eines umfassenden Systems zur wissensbasierten Modellierung von ASK bilden. Es handelt sich um:
 1. ein Verfahren zur Erweiterung des sichtbaren Wertebereichs in Spektrogrammen mit Hilfe von kontrastoptimierten Farbpaletten (siehe Abschnitt 3.2),
 2. ein Verfahren zur Extraktion spektrographischer Signaturen akustischer Ereignisse aus Spektrogrammen (siehe Abschnitt 3.3.1),
 3. ein Verfahren zur Berechnung einer Klassen-Korrelationsfunktion für spektrographische Signaturen (siehe Abschnitt 3.3.2),

4. ein Verfahren zur algorithmischen Annotation von Audiodaten mit Hilfe von spektrographischen Signaturen (siehe Abschnitt 3.3.3),
 5. ein Verfahren zur Visualisierung von First-Level-Annotationsfakten in Audiodaten (siehe Abschnitt 3.3.5),
 6. ein Verfahren zum Ausdünnen von Merkmalsvektoren in spektrographischen Signaturen (siehe Abschnitt 3.3.6),
 7. ein Verfahren zum Klonen von spektrographischen Signaturen (siehe Abschnitt 3.3.7),
 8. ein Verfahren zur automatischen Extraktion von spektrographischen Signaturen aus Audiodaten (siehe Abschnitt 3.3.8),
 9. ein Verfahren zum Verschmelzen von spektrographischen Signaturen (siehe Abschnitt 3.3.9),
 10. ein Verfahren zum ähnlichkeitsbasierten Verschmelzen von spektrographischen Signaturen (siehe Abschnitt 3.3.10),
 11. ein Verfahren zur induktiven Programmierung von First-Level-ASK (siehe Abschnitt 3.3.11),
 12. ein Verfahren zur Modellierung symbolischer Signaturen akustischer Ereignisse durch Interpretation von Audiodaten (siehe Abschnitt 3.4.1),
 13. ein Verfahren zur algorithmischen Annotation von Audiodaten mit Hilfe von symbolischen Signaturen akustischer Ereignisse (siehe Abschnitt 3.4.2),
 14. ein Verfahren zur Visualisierung von Second-Level-Annotationsfakten in Audiodaten (siehe Abschnitt 3.4.3),
 15. ein Verfahren zur automatischen Extraktion von symbolischen Signaturen aus Mengen von Annotationsfakten (siehe Abschnitt 3.4.4),
 16. ein Verfahren zum Verschmelzen von symbolischen Signaturen (siehe Abschnitt 3.4.5),
 17. ein Verfahren zum ähnlichkeitsbasierten Verschmelzen von symbolischen Signaturen (siehe Abschnitt 3.4.6),
 18. ein Verfahren zur Visualisierung von verschmolzenen symbolischen Signaturen (siehe Abschnitt 3.4.6),
 19. ein Verfahren zur induktiven Programmierung von Second-Level-ASK (siehe Abschnitt 3.4.7).
- Es wurde ein Softwarepaket zur Modellierung von ASK für die Bioakustik entwickelt, welches die KDD-basierte Programmierung von First-Level- und Second-Level-ASK gestattet. Die Software stellt eine Verbindung von logischer Programmierung und digitaler Signalverarbeitung her und wurde daher

DSPROLOG¹ genannt. Alle oben aufgelisteten Verfahren sind in DSPROLOG implementiert².

- Im 'Dolphin-Reef' in Eilat wurde für diese Arbeit ein umfangreiches Korpus von Audio-Aufzeichnungen der akustischen Signale des Großen Tümmlers hergestellt. Das Korpus umfasst ca. 87 Gigabyte (270 Stunden) von Aufnahmen in sehr hoher Qualität (96 kHz, 16 Bit). In den Aufnahmen wird ein Frequenzbereich von 0-48 kHz abgedeckt. Es handelt sich um eine der derzeit umfangreichsten digitalen Sammlungen akustischer Signale dieser Spezies³.
- Eine Untersuchung des Eilat-Korpus mit KDD-Methoden erbrachte Resultate, welche einen wichtigen Beitrag zur Erforschung des akustischen Signalsystems des Großen Tümmlers leisten:
 1. Eine phänomenologische Analyse des Audiokorpus zeigte eine Anzahl akustischer Phänomene, die so noch nicht in der Literatur beschrieben wurden. Hierzu gehören: (1) Pfiff-laute mit einer Grundfrequenz oberhalb von 32 kHz (siehe Abschnitt 4.4.2), (2) zahlreiche sequentiell angeordnete pulsierte Laute mit starker Formantenstruktur im unteren Frequenzbereich (siehe Abschnitt 4.5.4), (3) das Phänomen des Pulsspektrum-Drifts bei Klicklauten (siehe Abschnitt 4.5.4).
 2. Mit Hilfe eines symbolischen Lernverfahrens wurden unter verschiedenen Bedingungen Cluster für Pfiff-laute im Eilat-Korpus generiert. Die Cluster spiegeln die Gruppenähnlichkeit von Pfiff-typen in der Delphingruppe wieder. Ein Vergleich mit synthetischen Pfiff-lauten zeigt, dass die natürlichen Laute sehr starke Varianzen aufweisen. Zudem kann eine starke Prävalenz von nur *zwei* Lauttypen in den Daten nachgewiesen werden (siehe Kapitel 6).
 3. Durch maschinelle Zählung von einzelnen Klicklauten (ca. 2.5 Millionen), Klick-Trains (ca. 38000), Burst-Pulsed-Lauten (ca. 11000) und Pfiff-lauten (ca. 12000) mit Hilfe von speziell hierfür modellierten ASK wurden statistische Verteilungsprofile erstellt, welche die Dynamik der Emission dieser Lauttypen im Verlaufe von drei Wochen widerspiegeln (siehe Kapitel 7).
 4. Ein Vergleich der Verteilungsprofile der verschiedenen Lauttypen und eine Spektralanalyse von ca. 2.5 Millionen Klicklauten legt nahe, dass es sich bei Klicklauten im mittleren Frequenzbereich (bis 48 kHz) um Sonar-laute und nicht um kommunikative Laute handelt. Sie weisen Eigenschaften auf, die bisher unbekannt waren. Hierzu gehört unter anderem, dass sie für Sonar-laute sehr niedrige dominante Frequenzen besitzen. Möglicherweise handelt es sich bei dem Phänomen des Pulsspektrum-Drifts um ein

¹Nach DSP (**D**igital **S**ignal **P**rocessing) und der logischen Programmiersprache PROLOG.

²Siehe: <http://tocsy.agnld.uni-potsdam.de/dsprolog.php>

³Die Aufnahmen befinden sich im Tierstimmenarchiv der Humboldt-Universität zu Berlin.

Modulationsprinzip, welches von *Tursiops truncatus* genutzt wird, um Sonarlaute gezielt zu modulieren (siehe Kapitel 7).

Die in der vorliegenden Dissertation erarbeiteten Methoden, Verfahren und Algorithmen sind in vielfältiger Hinsicht erweiterbar. Eine auf allen Ebenen konsequent durchgeführte Modularisierung gestattet, in Zukunft Modifikationen vorzunehmen, ohne dass an der grundlegenden Architektur von Methode und Algorithmik etwas geändert werden muss. Für die dargestellten Verfahren ergeben sich mannigfaltige Anwendungsperspektiven: Sie eignen sich zur wissensbasierten Modellierung von ASK für nahezu alle Spezies – seien es Insekten, Singvögel, Meeressäuger oder Primaten.

Teil I

Wissensbasierte Modellierung von Audio-Signal-Klassifikatoren

Kapitel 1

Interdisziplinärer Kontext

1.1 KDD und Data-Mining

Um ein besseres Verständnis der in den Kapiteln 2 und 3 beschriebenen Modellierungsprozesse zu ermöglichen, werden in diesem Abschnitt zunächst die grundlegenden Konzepte und Themen des KDD und Data-Mining dargestellt. Hierbei fasst der Autor KDD als umfassenden methodischen Prozess auf, in dem der Mensch eine Schlüsselrolle spielt.

1.1.1 Hintergründe und Ziele des KDD

Schon seit langem ist es dem unbewaffneten menschlichen Verstand nicht mehr möglich, die immensen Datenmengen in modernen Datenbanksystemen sinnvoll zu interpretieren. Daher wurde bereits in den 1980er Jahren damit begonnen, spezialisierte Werkzeuge zur Analyse und Auswertung von großen Datenmengen zu schaffen. Es entstand eine neue Disziplin mit Namen **K**nowledge **D**iscovery in **D**atabases (KDD). Seitdem sind mehr als zwanzig Jahre vergangen, und KDD ist zu einem wichtigen Gebiet avanciert, das die Grenzen der Informatik weit überschreitet.

Beispiele für sehr große Datenbanken sind heute in fast allen Wissenschaftsbereichen zu finden: GIS-Datenbanken, Gendatenbanken oder digitale Bibliotheken können leicht die Terabyte Grenze überschreiten. Der Bedarf an maschinellen Hilfsmitteln zum Umgang mit derartigen Datenvolumina ist evident. Von besonderer Wichtigkeit hierbei sind intelligente Methoden zum Klassifizieren, Interpretieren und Visualisieren von Datenmengen sowie Verfahren zum Auffinden von neuartigen und interessanten Mustern in ihnen.

Heute werden von den meisten Autoren die Begriffe KDD und Data-Mining wie folgt verwendet [36, 37, 12, 104, 38, 39, 11]:

- **KDD:** KDD ist der Oberbegriff für alle Verfahren zur intelligenten Datenauswertung, die auf dem methodischen Schritt des Data-Mining beruhen oder diesen beinhalten. Der KDD-Prozess umfasst auch allgemeine Vorbereitungen und die Evaluation von Ergebnissen.

- **Data-Mining:** Der Begriff Data-Mining bezieht sich auf den Schritt der induktiven Musterbildung (Pattern-Formation) mit Hilfe formaler algorithmischer Methoden. Data-Mining ist der zentrale, aber nicht der einzige Schritt im KDD-Prozess.

KDD als wissenschaftliche Disziplin basiert auf den Erkenntnissen verschiedener Teilgebiete der Informatik. Beiträge kommen aus der Mustererkennung, dem maschinellen Lernen, der KI, der Computerlinguistik, der Statistik, dem Software-Engineering und der Computervisualistik.

1.1.2 Konzepte des KDD

In einem der grundlegenden Werke zum Thema KDD findet sich folgende oft zitierte Definition: „*Knowledge Discovery in Databases is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data* [38].“

Beim KDD-Prozess handelt es sich um einen mehrstufigen nicht-trivialen methodisch strukturierten Prozess, der Datenvorbereitung, Data-Mining und Wissensevaluation umfasst. Der Prozess ist durch mehrere Einzelschritte und rekursive Schleifen gekennzeichnet. Er wird im nächsten Abschnitt ausführlich dargestellt. Zuvor bedarf es jedoch der Konkretisierung einiger grundlegender Begriffe¹:

1. **Daten:** Daten können als eine Menge F von Fakten in symbolischer Repräsentationsform aufgefasst werden. In einer Datenbank ist F typischerweise in Form von Tabellen oder Listen repräsentiert.
2. **Muster:** Ein Muster ist ein Ausdruck E in einer formalen Sprache L , der eine Untermenge F_E von F beschreibt und dabei in gewissem Sinne einfacher ist als die direkte Aufzählung von F_E .
3. **Gültigkeit:** Im KDD-Prozess entdeckte Muster sollten für F , aber auch für neue Datensätze aus einer gleichartigen oder ähnlichen Problemdomäne mit spezifischer Gewissheit gültig sein. Als Maß für die Gewissheit (*Certainty*) wird oft eine Funktion C verwendet, die Ausdrücke aus L auf einen reellwertigen Messraum M_C abbildet. Einem Ausdruck $E \in L$ kann so eine Maßzahl $c = C(E, F)$ für die Gültigkeit zugeordnet werden. C kann auch eine Boolesche Funktion sein.
4. **Neuartigkeit:** Im KDD-Prozess entdeckte Muster sollten neuartig sein, d.h. wesentliche Unterschiede zu bereits vorhandenem Wissen aufweisen. Als Maß für die Neuartigkeit wird oft eine Funktion N verwendet, die Ausdrücke aus L auf einen reellwertigen Messraum M_N abbildet. Einem Ausdruck $E \in L$ kann so eine Maßzahl $n = N(E, F)$ für die Neuartigkeit zugeordnet werden. N kann auch eine Boolesche Funktion sein.

¹Alle KDD Begriffe in diesem Abschnitt nach [11]

5. **Nützlichkeit:** Im KDD-Prozess entdeckte Muster sollten nützlich sein bzw. zu nützlichen Aktionen führen können. Im wissenschaftlichen Kontext ist ein Muster z.B. dann nützlich, wenn es zur Bildung einer bestimmten wissenschaftlichen Theorie beiträgt oder das Potential hierfür besitzt. Als Maß für den Nutzen (*Usability*) wird oft eine Funktion U verwendet, die Ausdrücke aus L auf einen reellwertigen Messraum M_U abbildet. Einem Ausdruck $E \in L$ kann so eine Maßzahl $u = U(E, F)$ für die Nützlichkeit zugeordnet werden. U kann auch eine Boolesche Funktion sein.
6. **Verständlichkeit:** Im KDD-Prozess entdeckte Muster sollten für den Menschen verständlich sein oder mit geeigneten Verfahren (z.B. Visualisierungsverfahren) verständlich gemacht werden können. Maße für den Grad der Verständlichkeit sind in der Regel nicht leicht zu finden. Unspezifische Begriffsbestimmungen, wie „leicht nachvollziehbar“ bis hin zu restriktiven quantitativen Maßen wie „Länge in Bit“ sind verbreitet. Analog zu den obigen Maßen, kann für die Verständlichkeit eine Funktion S verwendet werden, die Ausdrücke aus L auf einen reellwertigen Messraum M_S abbildet. Einem Ausdruck $E \in L$ kann so eine Maßzahl $s = S(E, F)$ für die Verständlichkeit zugeordnet werden. S kann auch eine Boolesche Funktion sein.
7. **Interessantheit:** Ein weiteres wichtiges Maß für die Relevanz von entdeckten Mustern ist deren Interessantheit. Einige KDD-Systeme gestatten die explizite Verwendung einer Interessantheitsfunktion

$$I : L \rightarrow M_I, M_I \subseteq \mathbb{R}, \quad (1.1)$$

die Ausdrücke aus L auf einen reellwertigen Messraum M_I abbildet (ein konkretes Beispiel findet sich in Kapitel 6). Einem Ausdruck $E \in L$ kann so eine Maßzahl für die Interessantheit zugeordnet werden. I kann aber auch eine Boolesche Funktion sein. In KDD-Systemen, in denen keine Funktion für die Interessantheit existiert, kann eine Ordnung über gefundenen Mustern aufgestellt werden, die Rückschlüsse auf deren Interessantheit zulässt.

Wissen ist ebenfalls ein zentrales Konzept des KDD, entzieht sich jedoch mehr als die anderen einer formalen Spezifikation. Im wissenschaftlichen KDD kann Wissen als Hypothese aufgefasst werden, die einerseits aus einer Theorie ableitbar ist, andererseits empirisch verifiziert wurde. Im Sinne der Wissenschaftstheorie nach Bunge (1967) [12] ist eine Theorie dabei ein konsistentes System von allgemeingültigen Gesetzen und universalen Hypothesen. Sie dient dazu, Erklärungen für Phänomene eines Gegenstandsbereiches zu finden und zu formulieren. Muster, die aus einem KDD-Prozess resultieren und das Kriterium der Gültigkeit für empirische Daten erfüllen, sind typischerweise empirisch verifizierte Hypothesen, da sie sich aus empirischen Daten ableiten und dabei statistisch signifikant sind. Diese Muster sind idealerweise auch theoriekonsistent, da bei einem sorgfältig durchgeführten KDD-Prozess, die Sprache L nicht im Widerspruch zu wissenschaftlichen Theorien über den Forschungsgegenstand steht.

1.1.3 Der KDD-Prozess

Zu Beginn dieses Kapitels wurde auf die enge konzeptuelle Verbindung von KDD und Data-Mining hingewiesen. Data-Mining lässt sich als essentieller Schritt im KDD-Prozess verstehen, bei dem mit Hilfe von Data-Mining-Algorithmen Muster gefunden werden, die im günstigen Falle den obigen Kriterien Gültigkeit, Neuartigkeit, Nützlichkeit, Verständlichkeit und Interessantheit genügen. Der KDD-Prozess kann als Prozess der systematischen, zielgerichteten Anwendung von konkreten Data-Mining-Methoden aufgefasst werden, der mit der Motivation ausgeführt wird, aus einer Datenbank Wissen zu gewinnen, das spezifischen Bewertungskriterien genügt. Der KDD-Prozess umfasst die Gesamtheit aller hierfür notwendigen Schritte einschließlich aller Vorbereitungen und Nacharbeiten.

Die bisherigen Definitionen erwecken den Anschein, dass KDD ein autonomer Prozess ist oder zumindest sein könnte. Dies ist jedoch nicht der Fall, da die vielen heterogenen Schritte des KDD derzeit nur vom Menschen sinnvoll und zielgerichtet ausgeführt werden können. Der Mensch ist Kontrollinstanz des Gesamtprozesses und jedes einzelnen Schrittes. KDD soll den Menschen dabei unterstützen, die Prozesse zu verstehen, welche den Daten in großen Datenbanken zugrunde liegen.

Im folgenden wird dargestellt, wie ein KDD-Prozess typischerweise abläuft. Jeder Schritt kann ggf. iteriert werden. Auch beinhaltet jeder Schritt möglicherweise zahlreiche Entscheidungen, die vom Anwender auf der Grundlage seines Hintergrundwissens getroffen werden müssen. Abbildung 1.1 veranschaulicht die einzelnen Schritte.

1. **Vorwissen:** Es versteht sich fast von selbst, dass die erste Voraussetzung für sinnvolles KDD immer ein adäquates Verständnis des Anwendungsgebietes sein muss. In den meisten Fällen muss das Vorwissen explizit herausgearbeitet werden, damit die Ziele des KDD in einer konkreten Anwendung besser sichtbar werden. Man spricht in diesem Zusammenhang auch von Ziel-Findung (Task-Discovery), einem mitunter langwierigen und schwierigen Prozess. Auch wenn es oftmals scheint, dass bestimmte Fragestellungen eigentlich schon klar und verständlich sind, müssen sie für den KDD-Prozess meist nochmals präzisiert und formalisiert werden. Überspringt man diesen Schritt, kann es allzu leicht passieren, dass man entweder keine Antworten auf gestellte Fragen findet oder falsche. Verschiedene KDD-Systeme gestatten es, Hintergrundwissen in Form von Metadaten (Metatags) einzubeziehen und algorithmisch zu nutzen.
2. **Datenbasis:** Die Datenbasis repräsentiert die Menge aller Fakten. Das Vorwissen bedingt auch, welche Daten überhaupt in den Scopus der Untersuchung miteinbezogen werden sollen. In den Fällen, in denen die vorab zu Verfügung stehende Datenmenge nicht ausreichend oder gar nicht vorhanden ist, muss die Datenbasis unter Einbezug des Vorwissens zunächst erzeugt bzw. erweitert werden.
3. **Selektion:** In den meisten Fällen ist es zunächst notwendig, aus einer Flut von

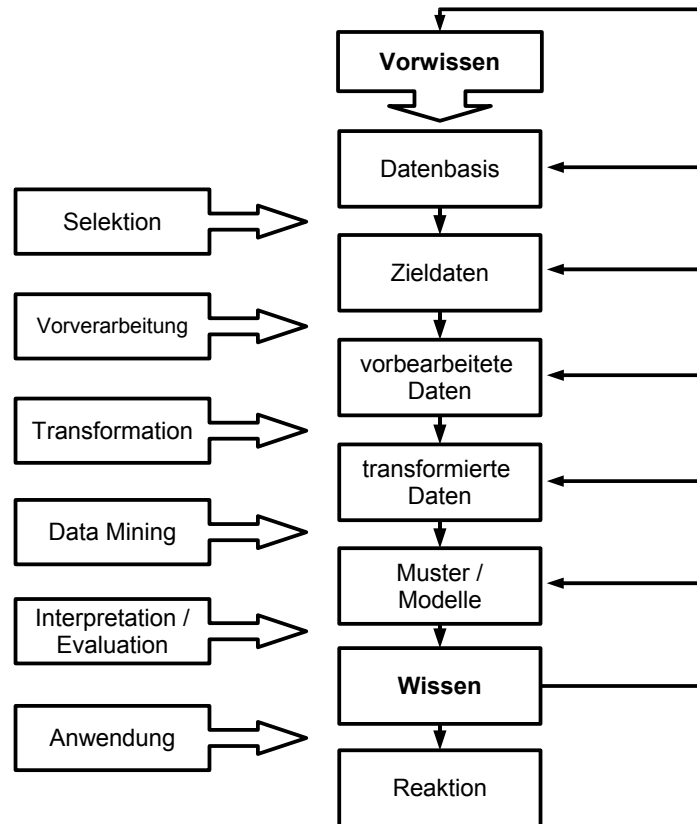


Abbildung 1.1: Der KDD-Prozess.

Rohdaten eine Teilmenge auszuwählen, die für die Untersuchung am geeignetsten erscheint. In die Entscheidung darüber, welche Daten tatsächlich in den nachfolgenden KDD-Schritten weiterverwendet werden, fließen Betrachtungen über Inhalt, Qualität, Quantität und Struktur der Daten ein. Die verbleibenden Daten nennt man Zieldaten.

4. **Datenbereinigung:** Die Menge der Zieldaten muss in der Regel noch bereinigt werden. Idealerweise wird die Menge fehlerhafter Fakten schon bei der Datenerfassung minimiert, so dass dieser Schritt gelegentlich übersprungen werden kann.
5. **Transformation:** Die Menge der bereinigten Zieldaten wird nun durch eine geeignete Projektion weiter reduziert. In diesem Schritt wird angestrebt, durch den Einsatz von geeigneten dimensionsvermindernden Transformationsmethoden die Anzahl der auszuwertenden Variablen zu minimieren. Wie schon bei der Selektion fließen hierbei Betrachtungen über Inhalt, Qualität, Quantität und Struktur der Daten ein. Dieser Schritt schließt die Vorbereitung der Daten

ab.

6. **Wahl des Data-Mining-Ziels:** Bevor man einen konkreten Data-Mining-Algorithmus wählt, muss entschieden werden, was das eigentliche Ziel des KDD-Prozesses ist. Mögliche Ziele sind beispielsweise Klassifizierung, Regression, Clustering oder die Entdeckung von Veränderungen und Abweichungen.
7. **Wahl des Data-Mining-Algorithmus:** Die Wahl des Data-Mining-Ziels bedingt die Klasse von möglichen Data-Mining-Algorithmen. Aus dieser muss ein konkreter Algorithmus ausgewählt werden. Diese Entscheidung beinhaltet weitere detaillierte Betrachtungen über Parameter und Modelle sowie das Gesamtumfeld des KDD-Prozesses. Oftmals stellt sich an dieser Stelle heraus, dass bestehende Algorithmen nicht ausreichen, um das Problem adäquat zu behandeln. In diesen Fällen müssen existierende Algorithmen modifiziert oder neue Algorithmen erarbeitet werden.
8. **Anwendung des Data-Mining-Algorithmus:** Die Anwendung des Data-Mining-Algorithmus zur Suche nach Mustern (z.B. Klassifizierungsregeln oder Clustern), erfolgt erst, nachdem alle vorausgehenden Schritte sorgfältig ausgeführt worden sind. Dieser Schritt kann autonom (*unsupervised*) oder überwacht (*supervised*) ablaufen.
9. **Interpretation:** Eine erfolgreiche Anwendung des gewählten Data-Mining-Algorithmus auf die bereinigten und transformierten Zieldaten resultiert in einer Menge von Mustern. Diese müssen nun mit Hilfe der oben beschriebenen Kriterien interpretiert und bewertet werden. Meist kehrt man danach zu einem der vorausgegangenen Schritte zurück und führt vor einem erneuten Durchlauf des Data-Mining-Schrittes Modifikationen an Daten bzw. Algorithmen durch. Sind alle Schritte des KDD-Prozesses mit der notwendigen Sorgfalt ausgeführt worden, können die gewonnenen Muster zur Bildung oder Modifikation von Wissen über die Problemdomäne führen, so etwa zur Neubildung von wissenschaftlichen Hypothesen oder zur Änderung von bestehenden Theorien.
10. **Nutzung des Wissens:** Das gewonnene Wissen kann nun dokumentiert und genutzt werden. Es kann sich bewähren und festigen oder aber verworfen werden. Potentielle Konflikte mit anderen Erkenntnissen müssen dabei erkannt und aufgelöst werden.

Durch Rückkoppelungen und Iterationen im KDD-Prozess ergeben sich oft komplizierte und schwer zu kontrollierende Abläufe. Die Speicherung von Zwischenergebnissen ist daher meist erforderlich. In den vielen Fällen ist eine integrierte KDD-Arbeitsumgebung mit ergonomischer Benutzerführung, intelligenter Daten-Verwaltung und Feedback über Zwischenergebnisse notwendig. Man nennt eine solche Umgebung auch **Knowledge Discovery Support Environment (KDSE)**. Das Fehlen einer solchen Umgebung kann den KDD-Prozess erheblich erschweren.

1.1.4 Data-Mining-Aufgaben

Die iterative Anwendung von Data-Mining-Methoden spielt eine zentrale Rolle im KDD-Prozess. Data-Mining kann als der Prozess definiert werden, in dem unter Verwendung einer Datenmenge Muster angepaßt oder neu generiert werden. Ob die auf diesem Wege gefundenen oder erzeugten Muster den obigen Evaluationskriterien genügen, muss im korrespondierenden KDD-Gesamtprozess herausgefunden werden. Hierzu können formale Bewertungsmodelle, aber auch subjektive Beurteilungen hinzugezogen werden.

Die übergeordnete Aufgabe des Data-Mining liegt darin, Muster zu gewinnen, die Vorhersagen gestatten, Klassifikationen ermöglichen oder Informationen komprimieren. Gefundene Muster können auch genutzt werden, um unbekannte Werte von Variablen vorherzusagen. Sie sollten so gestaltet sein, dass sie kompakte für den Menschen verständliche Beschreibungen von Datenmengen liefern. Leider sind diese Ziele mitunter gegensätzlich, so geht z.B. die Vorhersagegenauigkeit oftmals auf Kosten der Verständlichkeit. Häufige Ziele des Data-Mining sind:

- **Klassifikation:** Ein Klassifikator kann als eine Funktion aufgefasst werden, die eine Menge von Daten einem Element aus einer Menge von Klassen zuordnet. Klassifikationsverfahren basieren meist auf dem Erlernen von Klassifikator-Funktionen anhand von Trainingsdaten.
- **Regression:** Unter dem Begriff Regression werden solche numerischen und statistischen Verfahren zusammengefasst, die dem Erlernen von linearen und nichtlinearen Funktionen dienen, die Datensätze auf reellwertige Vorhersagevariablen abbilden.
- **Gruppenbildung:** Unter dem Begriff Gruppenbildung (*Clustering*) werden Methoden zusammengefasst, die dem ähnlichkeitsbasierten Erlernen von endlichen Mengen von Gruppen (*Clustern*) dienen. Cluster können gegenseitig ausschließend, überlappend und hierarchisch gegliedert sein. Gruppenbildung und Klassifikation sind miteinander eng verwandt.
- **Abhängigkeitsentdeckung:** Verfahren zur Abhängigkeitsentdeckung (*Dependency Modeling*) dienen dem Auffinden von Mustern, die statistisch signifikante Abhängigkeiten zwischen Variablen beschreiben. Derartige Relationen können qualitativ (strukturell) oder quantitativ sein. *Dependency Modeling* dient nicht notwendigerweise Klassifikationszwecken.
- **Abweichungsentdeckung:** Abweichungsentdeckung (*Change and Derivation Detection*) dient dazu, in vorgegebenen Objektmengen statistisch auffällige Veränderungen und Abweichungen von Durchschnittswerten zu entdecken.

Jedes dieser Methodengebiete stellt eine große Anzahl von Algorithmen zur Verfügung. Die Wahl, welcher der zahlreichen Algorithmen im konkreten Anwendungsfall tatsächlich eingesetzt wird, hängt im Wesentlichen von den Zielen des KDD ab, aber auch von technischen und praktischen Randbedingungen.

1.1.5 Data-Mining-Verfahren

Viele Data-Mining-Verfahren besitzen einen charakteristischen Grundaufbau. Sie sind meist (aber nicht immer) wie folgt gekennzeichnet:

- **Modelldarstellung:** Um die Menge der entdeckbaren Muster einzuschränken, muss eine geeignete formale Sprache L gefunden werden. Bei der Definition von L ist darauf zu achten, dass sie weder zu beschränkt noch zu sehr übergeneralisierend ist. Das letztere führt zu einer schlechten Vorhersagegenauigkeit durch Überangepasstheit (*Overfitting*), das erstere *a priori* zum Ausschluss von möglicherweise interessanten Mustern.
- **Suchmethode:** Suchmethoden dienen dem Auffinden von interessanten Mustern. Man kann dabei zwei Arten von Suche unterscheiden: die Parametersuche (*Model Fitting, Parameter Search*) und die Modellsuche (*Model Search*). Mit der Parametersuche wird versucht, die besten Parameter zu einem festen Modell zu finden. Die Modellsuche ist der Parametersuche übergeordnet. In ihr werden einzelne Modelle in Bezug zu Modellklassen gesetzt. Auch fließt hier ggf. Hintergrundwissen ein, um die Menge möglicher Modelle einzuschränken.
- **Modellevaluation:** Die Modellevaluation dient der Bewertung von Zwischenergebnissen und Resultaten der Suche. Hierzu wird in Abhängigkeit des jeweiligen KDD-Kontextes auf verschiedenartige quantitative und qualitative Maße und Bewertungskriterien zurückgegriffen (siehe Abschnitt 1.1.2). Bei der Modellevaluation wird in der Regel zwischen Trainings- und Testdaten unterschieden.

Im Folgenden werden einige Verfahren aufgezählt, die in ihrem Grundaufbau dem obigen Schema entsprechen. Sie finden Verbreitung in vielen wissenschaftlichen KDD-Systemen.

1. **Entscheidungsbaumverfahren:** Diese Klasse von Methoden stammt aus der Statistik. Sie zeichnet sich durch eine einfach zu verstehende Darstellungsform aus. Es werden Bäume produziert, mit deren Hilfe kategoriale oder numerische Variablen vorhergesagt werden. Den Knoten eines Baumes entsprechen Attributabfragen, den Teilbäumen vorgefundene Attributwerte und den Blättern vorhergesagte Klassen.
2. **Nichtlineare Regression:** Hierbei handelt es sich um eine Familie von Klassifikationsmethoden und Voraussagetechniken. Sie zeichnet sich dadurch aus, dass Kombinationen von linearen und nichtlinearen Elementarfunktionen an Vektoren von Inputvariablen angepasst werden. Künstliche neuronale Netze (KNN) und adaptive Regressions-Splines zählen zu den nichtlinearen Regressionsverfahren [79, 41]. KNN werden im KDD-Kontext aufgrund mangelnder Interpretierbarkeit häufig kritisiert.

3. **Regellernverfahren:** Aus Datensätzen werden relationale oder propositionale Regeln generiert [24]. Diese Regeln können auch probabilistisch oder unscharf sein [122]. Regeln können zur Vorhersage und Klassifikation genutzt werden. Induktiv logische Programmierung ist eine Klasse von Regellernverfahren, die sich der Prädikatenlogik bedient. ILP-Verfahren können gut in relationalen Datenbanken verwendet werden.
4. **Bayes-Verfahren:** Diese Verfahren dienen in der Regel der Klassifikation. Unter Ausnutzung des Bayes-Theorems werden Datensätze auf Klassen abgebildet, die in einem gegebenen Kontext die größte *a posteriori* Wahrscheinlichkeit besitzen.
5. **Support-Vektor-Maschinen:** Es handelt sich um überwachte Methoden zur Klassifikation. Support-Vektor-Maschinen beruhen im Prinzip darauf, dass versucht wird, eine optimal trennende Hyperebene in einem hochdimensionalen Eingaberaum zu finden. Der Eingaberaum ist dabei ein spezieller Merkmalsraum mit höherer Dimension als der ursprüngliche. Die Idee ist, dass mittels einer geeigneten nichtlinearen Abbildung in eine höhere Dimension Daten aus unterschiedlichen Kategorien mit einer Hyperebene immer optimal getrennt werden können. Ein Vorteil von SVM liegt darin, dass dem Problem des Overfitting bei gleichzeitig sehr guter Generalisierung begegnet wird.

1.1.6 Aktuelle Forschungsthemen

Im Bereich der KDD-Forschung lassen sich zwei Trends ausmachen: (1) Forschung an prinzipiellen theoretischen Problemen und (2) die Entwicklung von neuen aufwändigen Methoden in komplizierten Anwendungsgebieten. Forschungsfragen der ersten Kategorie umfassen folgende Aspekte:

- Vorhersagegenauigkeit, Verlässlichkeit und Signifikanz: Wie muss man KDD-Systeme modifizieren, um die Vorhersagegenauigkeit in praktischen Anwendungen zu steigern? Wie kann man Overfitting verhindern? Wie kann man vermeiden, dass zufällige Fluktuationen in Datenbeständen als signifikante Muster erkannt werden?
- Skalierbarkeit: Wie müssen KDD- und Data-Mining-Verfahren gestaltet werden, dass man sie auch auf Datenbestände im Gigabyte- und Terabytebereich anwenden kann? Wie geht man mit hochdimensionalen Datenräumen um?
- Evaluation von Ergebnissen: Wie kann man die Kriterien zur Beurteilung von Mustern, Modellen und anderen KDD-Zwischenergebnissen verbessern? Wie können die Funktionen für Neuartigkeit, Interessantheit usw. im konkreten Fall spezifiziert werden?

- Benutzerinteraktion und Verständlichkeit von Ergebnissen: Wie veranschaulicht man Muster, Modelle und andere KDD-Zwischenergebnisse? Wie kann man KDD-Systeme besser an Benutzerbedürfnisse anpassen?
- Integration von Expertenwissen in KDD-Systeme und Informations-Infrastrukturen: Wie lässt sich die Expertise von Anwendern sinnvoll in den KDD-Prozess miteinbeziehen? Wie kann man KDD-Systeme so gestalten, dass sie sich besser in existierende Informations-Infrastrukturen einfügen?
- Temporäre, veränderliche und fehlerhafte Daten: Wie geht man mit fehlerhaften Daten und leeren Feldern um? Wie vermeidet man Fehler bei der Datenakquise? Wie geht man mit Veränderungen von Daten im Zeitverlauf um?

Aktuelle Fragen der zweiten Kategorie umfassen u.a. folgende Aspekte, die auch heterogene und nichtstrukturierte Datenbestände einschließen:

- Texte und Web-Daten: Wie können Textsammlungen zur Extraktion von Wissen genutzt werden? Wie kann man Inhalte von Texten sinnvoll klassifizieren oder indizieren? Wie können die Ressourcen des Internets zur Extraktion von Wissen genutzt werden? Wie kann man Inhalte von Webseiten sinnvoll klassifizieren oder indizieren?
- Multimediale Daten, räumliche Daten und Zeitreihen: Wie können Audiodaten, Videos und Bilder für KDD-Analysen zugänglich gemacht werden? Wie können Daten mit Raumbezug (z.B. geographische Daten) KDD-Analysen zugänglich gemacht werden? Wie können Daten mit Zeitbezug (z.B. Messreihen) verwendet werden?

Wissenschaftliches KDD in bioakustischen Audiodaten nimmt in diesem Kontext eine Sonderstellung ein. Es handelt sich um eine Spezialform des KDD, in der Audiodaten im Mittelpunkt stehen. Im Gegensatz zu vielen anderen Einsatzgebieten des KDD muss bei dieser Form besonders hohen wissenschaftlichen Qualitätsansprüchen Rechnung getragen werden.

1.2 KDD in der Bioakustik

*Nothing in biology makes sense
except in the light of evolution.*
Theodosius Dobzhansky 1937 [30]

Im vorausgehenden Abschnitt wurden grundlegende Konzepte des KDD dargestellt. Das folgende Kapitel enthält eine kurze Einführung in das Gebiet der Bioakustik und damit zusammenhängende interdisziplinäre Forschungsfragen. Im Weiteren wird dargestellt, welches die wichtigsten potentiellen Einsatzgebiete für KDD-Methoden in der Bioakustik sind.

1.2.1 Bioakustik

Forschungsgegenstand

Die Bioakustik ist die Wissenschaft von den „*akustischen Erscheinungen bei Tieren und untersucht sie strukturell, funktionell, informationell sowie unter den Gesichtspunkten der Veränderungen in der Zeit* [109].“ In ihr wird zwischen akustischen Kommunikationssystemen und Biosonarsystemen (Echolotsystemen) unterschieden.

Die überwiegende Mehrheit akustischer Emissionen im Tierreich dient der Kommunikation. Nicht wenige Spezies, darunter Zahnwale (*Odontoceti*), Fledertiere (*Chiroptera*) und einige Vogelarten, wie z.B. der höhlenbewohnende Fettschwalm (*Steatornis caripensis*), besitzen aber auch biologische Echolotsysteme. Sie dienen der Orientierung bei schlechten Sichtverhältnissen und/oder zum Orten von Beute.

Oft finden kommunikative Prozesse statt, indem gleichzeitig oder in schneller Abfolge mehrere Sinnesmodalitäten angesprochen werden. Eine Reihe von Spezies hat aber auch Kommunikationssysteme ausgebildet, die ganz oder in wesentlichen Teilen auf dem Austausch akustischer Signale basieren. Hierzu gehören viele Insekten- und Vogelarten, Zahnwale und etliche landlebende Säugetiere.

Akustischer Signaltransfer hat entscheidende Vorteile gegenüber anderen Arten der Kommunikation: Er funktioniert auch dann, wenn kein Körper- oder Sichtkontakt besteht und der Austausch von Botenstoffen nicht möglich oder nicht ausreichend ist. Schallsignale können bei geringem Energieaufwand gut moduliert und in praktisch unbegrenzter Vielgestaltigkeit produziert werden. Sie breiten sich sehr schnell im Raum aus, überwinden Sichthindernisse und können eine große Reichweite besitzen.

Die Aufgaben der Bioakustik gliedern sich in theoretische und praktische. Die *theoretische* Bioakustik entwickelt wissenschaftliche Theorien über die akustischen Kommunikations- und Echolotsysteme der Tiere. Sie befasst sich vorrangig mit folgenden Themen:

- Funktionelle Morphologie der Stimm- und Sonarorgane;
- Struktur und Organisation bioakustischer Signal- bzw. Zeichensysteme;

- Funktionsweise bioakustischer Sonarsysteme;
- Evolution und Kommunikation;
- Sprachevolution;
- Verarbeitung akustischer Information im Tiergehirn;

Exakte Theorien bioakustischer Kommunikationssysteme besitzen das Potential, eine Schlüsselrolle bei der Beantwortung einer Reihe von Leitfragen der Kognitions- und Sprachwissenschaften, der Ethologie und der Evolutionsbiologie zu spielen. Hierzu gehören Fragen nach Ursprung und Evolution der menschlichen Sprache, nach kognitiven und kommunikativen Fähigkeiten der Tiere sowie nach den allgemeinen Zusammenhängen von Kommunikation und Evolution.

Die Aufgaben der *praktischen* Bioakustik bestehen in der Entwicklung und Anwendung von Methoden zum Umgang mit akustischen Prozessen im Tierreich. Sie befasst sich im Wesentlichen mit folgenden Themen:

- Aufzeichnung, Archivierung und Verwaltung von Audiodaten;
- Analoge und digitale Verarbeitung bioakustischer Signale;
- Analyse bioakustischer Signale;
- Sonifikation und wissenschaftliche Visualisierung von Audiodaten;
- Klassifikation und Retrieval von Audiodaten;

Bioakustik und Ethologie

In ihrer heutigen Form wird die Bioakustik oft als Teildisziplin der Verhaltensbiologie bzw. Ethologie² angesehen. Dies geht auf die Ansicht zurück, dass auch Vokalisationen von Tieren im Grunde genommen nichts anderes sind als eine Form von Verhalten. Wie der Begriff Verhalten zu verstehen ist, hat sich im Laufe der Geschichte der Ethologie allerdings mehrfach geändert³.

²Die Begriffe Ethologie (*Ethology*) und Verhaltensbiologie (*Behavioural Biology*) werden meist synonym gebraucht – so auch hier. Es handelt sich um ein Teilgebiet der Biologie, welches auf den Arbeiten von Wissenschaftlern wie Charles Darwin (1809-1882), Oskar Heinroth (1871-1945), Konrad Lorenz (1903-1989), William Morton Wheeler (1865-1937), Nikolaas Tinbergen (1907-1988) und vielen anderen beruht. Der Begriff Ethologie ist abgeleitet von griech. Ethos (Charakter, Geist, Moral). Die Ethologie als Wissenschaft bezieht sich auf die Dinge, welche die mentalen Eigenschaften der Tiere betreffen.

³Der heutigen Auffassung des Verhaltensbegriffes ist eine Kontroverse von Vitalisten und Mechanisten vorausgegangen. Schon Darwin hatte auf die Analogie von stereotypen komplexen Verhaltensweisen und anatomischen Merkmalen hingewiesen. Er versuchte, die Gründe der Entstehung von beidem mit Hilfe seiner evolutionären Konzepte von Variabilität und Selektion zu erklären. Zu seinem Antagonisten wurde Alfred Russel Wallace (1823-1913). Wallace propagierte eine „vitalistische Kraft“ in der Natur als Grund für die evolutionären Anpassungen einschließlich der Instinkte.

Die moderne Konzeption des Verhaltensbegriffes basiert im Wesentlichen auf Arbeiten von Oskar Heinroth und Konrad Lorenz. Heinroth beobachtete in seinen Studien an Gänsen und Enten „ärteigene Triebhandlungen“, so z.B. immer wiederkehrende Gesten und Körperhaltungen bei der Balz. Er entdeckte auch, dass ähnliche Arten diese Verhaltensweisen in gleichen Kontexten in nur leichter Abwandlung zeigen. Für diese Verhaltensweisen fanden Heinroth und Lorenz eine evolutionäre Deutung – das ethologische Instinktkonzept: Die Instinktbewegungen sind in den Genen verankert und werden von Schlüsselreizen ausgelöst. Die Evolution hat unter dem Druck der Notwendigkeit der Arterhaltung das Ineinandergreifen von äußerem Auslöser, Handlungsbereitschaft und Verhaltensweise hervorgebracht.

Nicht zuletzt auf Lorenz geht auch die Forderung zurück, dass ethologische Instinktforschung auf genauester Beobachtung des Verhaltens von Tieren unter natürlichen Umweltbedingungen in der freien Natur erfolgen müsse. Charakteristisch für die Ethologie sind seitdem Ethogramme, d.h. detaillierte Kataloge über typische Verhaltensweisen von Tieren unter Berücksichtigung der zeitlichen Abfolge der Beobachtungen.

Mit dem Begriff Instinktbewegung wurden bis zum Ende der 1960er Jahre Aktivitäten bezeichnet, die rein angeboren sind. Heute lässt sich diese restriktive Interpretation instinktiven Verhaltens jedoch nicht mehr aufrecht erhalten, da sich herausgestellt hat, dass starre Reaktionen auf Umweltreize eher die Ausnahme als die Regel sind. Selbst einfachste Reflexe ändern sich im Laufe der Ontogenese eines Organismus in Abhängigkeit von inneren und äußeren Faktoren. Instinktives Verhalten beruht fast immer auf einem diffizilen Zusammenspiel von genetischer Prädisposition, Reizexposition und Umwelteinflüssen. Dies gilt auch für akustisches Verhalten.

Bioakustik und Sprachevolution

Im Rahmen der Bioakustik nimmt die menschliche Sprache eine Sonderrolle ein. Hierbei ist die Frage nach der Sprachevolution von besonderem Interesse.

Kommunikationssysteme unterliegen nicht nur der Evolution, sondern sie beeinflussen auch die Evolution selbst. Sie sind unlösbar mit dem evolutionären Gesamtprozess verwoben. Es bestehen hier auf mehreren Ebenen nichtlineare Zusammenhänge:

1. In komplexen Tiergesellschaften können gut ausgebildete kommunikative Fähigkeiten die Überlebens- und Fortpflanzungschancen einzelner Individuen signifikant erhöhen.

In der ersten Hälfte des 20. Jahrhunderts setzten sich zunehmend mechanistische Anschauungen durch. Die Mechanisten fassten Verhalten als Reagieren auf Außenreize (Stimuli) auf, die eine Kette von Reflexen auslösen. Zu den Vertretern dieser Richtung gehören der Nobelpreisträger (1904) Iwan Pawlow (1848-1936) und der Psychologe Burrhus Frederic Skinner (1904-1990), der auch als Begründer des Behaviorismus gilt.

2. Eine Spezies kann eine andere durch überlegene Kommunikationsfähigkeiten im Extremfall ganz verdrängen.
3. Interspezifische Kommunikation kann Voraussetzung für die Koevolution von Arten sein. Es entstehen hier vielfältige Wechselbeziehungen, die Gegenstand der aktuellen Forschung sind [123].
4. Kommunikation spielt bei der sexuellen Auslese eine zentrale Rolle. Beispielsweise sind Gesänge dazu geeignet, als biologische Indikatoren für Gesundheitszustand und Kraft von Individuen zu fungieren.
5. Kommunikationssysteme verändern sich mit der Zeit. Diese Veränderungen können räumlich begrenzt sein und dazu führen, dass Individuen einer Spezies sich nicht mehr fortpflanzen, da sie nicht mehr miteinander kommunizieren können. Hierin liegt möglicherweise ein Schlüsselmoment bei der Entstehung neuer Arten.

Kommunikation spielt bei der Evolution der Lebewesen eine tragende Rolle. Biologische Kommunikationssysteme sind integraler Bestandteil aller Spezies, der in die Lebensumwelt ihrer Träger tief eingebettet ist. Sie sind Produkt eines Jahrmillionen währenden evolutionären Prozesses und ihr Motor in einem. Der Mensch und die natürliche Sprache bilden hier keine Ausnahme.

Mit der Darwinschen Evolutionslehre hat sich der Gedanke durchgesetzt, dass auch die natürliche Sprache das Ergebnis eines evolutionären Prozesses ist, der möglicherweise viele Millionen Jahre umfasste. Die evolutionäre Sichtweise auf die Sprachentstehung (Sprachevolution) setzt voraus, dass sich unsere heutige Sprache aus einer Tiersprache entwickelt hat. Tatsächlich gibt es viele Evidenzen dafür, dass der heutige Mensch und der Schimpanse vor fünf bis sechs Millionen Jahren einen gemeinsamen affenähnlichen Vorfahren hatten. Es ist plausibel, dass dieser gemeinsame Vorfahr – ein Tier – ein relativ einfaches Kommunikationssystem besaß, das als Vorläufer unserer heutigen Sprache anzusehen ist. Die natürliche Sprache hat sich demzufolge erst nach der Teilung der evolutionären Linien von Mensch und Schimpanse allein auf der menschlichen Linie entwickelt. Ob sich die Evolution von urzeitlicher Tiersprache zur heutigen natürlichen Sprache kontinuierlich oder in Sprüngen vollzogen hat, ist dabei offen. In jedem Falle aber war der Ausgangspunkt eine Tiersprache. Die gegenteilige Hypothese, dass der gemeinsame Vorfahr bereits ein sprachähnliches komplexes Kommunikationssystem besaß, und nur der Schimpanse dieses zurückgebildet habe, gilt als unwahrscheinlich. Hochentwickelte kommunikative Fähigkeiten sind in der von den Vorfahren der Schimpansen und unseren Vorfahren über lange Zeiträume hinweg gemeinsam bewohnten Umwelt als evolutionärer Vorteil anzusehen.

Die evolutionstheoretische Sichtweise auf die Entstehung der natürlichen Sprache geht somit davon aus, dass sich die menschliche Sprache innerhalb der vergangenen fünf bis sechs Millionen Jahre aus einem tierischen Kommunikationssystem entwickelt hat. Wie dieser evolutionäre Prozess im Detail ausgesehen haben mag, ist

außerordentlich schwierig zu beantworten, da sich weder im Tierreich intermediäre Stufen der Sprachevolution konserviert haben, noch es natürliche Sprachen auf der Welt gibt, die im Vergleich zu anderen auf einer evolutionär vorgeordneten Stufe stehen. Möchte man etwas über die evolutionäre Metamorphose von Tiersprache zu menschlicher Sprache erfahren, so bedarf es einer Theorie der Sprachevolution, die Ausgangspunkte, Endpunkte und Schlüsselmomente der Sprachevolution miteinander verknüpfen muss.

Ein denkbarer hypothetischer Ausgangspunkt für die Evolution der natürlichen Sprache ist das Kommunikationssystem, welches der gemeinsame Vorfahr von Mensch und Schimpanse vor fünf bis sechs Millionen Jahren gehabt haben muss. Eine auf diesem Ausgangspunkt fußende Theorie der Sprachevolution braucht nur einen (in erdgeschichtlicher Hinsicht) relativ kurzen Zeitraum zu umfassen. Der gemeinsame Vorfahr ähnelte in Gestalt und Lebensweise den heutigen Schimpansen. Die Vermutung liegt nahe, dass sein Kommunikationssystem ähnlichen Anforderungen genügen musste wie das Kommunikationssystem der heute lebenden Schimpansen. Um eine nachvollziehbare Grundlage für eine exakte Beschreibung des Ausgangspunktes der Sprachevolution zu schaffen, ist es daher möglich, das Kommunikationssystem der heutigen Schimpansen als Referenz hinzuzuziehen.

Es ist allerdings auch denkbar, dass der eigentliche Ausgangspunkt der Sprachevolution sehr viel weiter in die Vergangenheit zurückreicht. Möglicherweise findet Informationsverarbeitung in den Gehirnen höherer Tiere schon seit viel früheren Zeiten mit Hilfe zeichenverarbeitender regelbasierter hierarchischer Systeme statt. Dies würde bedeuten, dass die Gehirne vieler höherer Tiere *universale* Prinzipien der diskreten regelbasierten Informationsverarbeitung nutzen. Diese haben zwei entscheidende Vorteile:

1. Informations-Kompression: Informationen (z.B. Sinneseindrücke) können mit Hilfe von Zeichencodes und Objekthierarchien nahezu verlustfrei komprimiert und im Gedächtnis organisiert werden;
2. Sequentielle intracerebrale Kommunikation: Informationen (z.B. Gedächtnisinhalte) können in Form von Zeichencodes über sequentiell arbeitende Reizleitungsnetzwerke innerhalb des Gehirns *verschickt* werden.

Die natürliche Sprache könnte demzufolge im Prinzip nichts anderes sein, als ein besonderes externalisiertes System universaler Prinzipien der zeichenbasierten Informationsverarbeitung. Durch sprachliche Zeichen können nunmehr mentale Zustände in komprimierter Form nicht nur zwischen verschiedenen Regionen eines Gehirns, sondern auch von Gehirn zu Gehirn kommuniziert werden. Cerebrale Reizleitungsnetzwerke werden durch eine neuartige Kommunikationsschnittstelle ergänzt, die auf Stimm- und Sinnesorganen basiert.

Für eine Theorie der Evolutionsgeschichte der Sprache kann es hilfreich sein, die Unterschiede von menschlicher Sprache und bioakustischen Kommunikationssystemen genau zu bestimmen. Durch ein Verständnis von Gemeinsamkeiten und

Unterschieden der verschiedenen Typen von Kommunikationssystemen können die Endpunkte der Sprachevolution, und somit das, was die natürliche Sprache eigentlich ausmacht, genau charakterisiert werden.

Ob der Mensch ein qualitativ distinktes Unicum der Evolution ist, oder ob er sich nur durch ein graduelles Eigenschaftsprofil von Tieren unterscheidet, ist keine Frage von körperlichen Eigenschaften, sondern eine von mentalen. Die herausragendste mentale Eigenschaft des Menschen ist seine Sprachfähigkeit⁴. Komparative Studien zwischen den Kommunikationssystemen der Tiere und der menschlichen Sprache können daher auch zu einer schärferen Abgrenzung dessen beitragen, was den Menschen an sich ausmacht. An dieser Stelle übernimmt die Wissenschaft eine gesellschaftliche Funktion, die zuvor der Religion vorbehalten war⁵.

Jedes soziale Tier besitzt verschiedene, teilweise recht komplizierte Arten zu kommunizieren. Diese können jedoch nur in den seltensten Fällen als Sprache bezeichnet werden [48]. Von den meisten Wissenschaftlern wird heute davon ausgegangen, dass Zeichensysteme im Tierreich weit weniger ausdrucksvoll und komplex sind, als die menschliche Sprache. Man schreibt der menschlichen Sprache eine Reihe von Eigenschaften zu, die in ihrer Gesamtheit und durch den Grad ihrer Ausprägung ihre Einzigartigkeit ausmachen:

- Willkürlichkeit: In der menschliche Sprache können willkürliche Beziehungen zwischen akustischer und/oder visueller Form von sprachlichen Zeichen und ihrer Bedeutung aufgestellt werden. Es existiert nicht notwendigerweise eine direkte Beziehung zwischen akustischer oder visueller Darstellung eines Zeichens und seiner Bedeutung.
- Kulturelle Weitergabe: Menschliche Sprache wird im Kontext einer kulturell geprägten Umwelt von den Sprachträgern einer Generation auf die der nächsten bewusst und unbewusst weitergegeben.
- Diskretheit: Menschliche Sprache kann als System von diskreten Elementen verstanden werden, die anhand von syntaktischen Regeln zu komplexen bedeutungstragenden Einheiten zusammensetzbar sind.

⁴Die meisten der evolutionären Anpassungen des Menschen sind nur durch den Grad ihrer Ausprägung einzigartig. Zum Beispiel nutzen viele Tiere Werkzeuge, darunter alle großen Affen, [108, 78, 114], Maulwürfe [105] und Spinnen. Manche Tierarten sind in der Lage Werkzeuge selbst herzustellen, so etwa Schimpansen und Bonobos [9, 75]. Auch die Weitergabe von Wissen von Generation zu Generation ist in der Tierwelt nicht unbekannt. Unterschiedliche kulturell tradierte Verhaltensmuster sind z.B. bei Schimpansen [28, 5, 119], Orang-Utans [114] und Walen [94] beobachtet worden.

⁵Im wissenschaftshistorischen und soziologischen Kontext ist die Frage nach Ähnlichkeiten und Unterschieden von menschlicher Sprache und bioakustischen Kommunikationssystemen von großer Tragweite. In unserem Kulturraum, der zwei Jahrtausende durch die biblische Theologie und ihre Sicht auf dieses Thema geprägt wurde, besitzt die Frage nach dem Ursprung der Sprache besondere Brisanz. Möglicherweise liegt hier ein wichtiger Grund für die große Aufmerksamkeit, welche dieses Thema heute in der Öffentlichkeit erzielt. Es verwundert nicht, dass moderne Wissenschaftsautoren wie Steven Pinker mit Monographien zu diesem Thema Millionenauflagen erzielen und Hörsäle zum Bersten füllen.

- Produktivität: Mit einer begrenzten Anzahl von diskreten Elementen kann eine praktisch unbegrenzte Anzahl von bedeutungstragenden Einheiten produziert werden.
- Räumliche und zeitliche Versetzung: Menschliche Sprache kann verwendet werden, um Ideen über Dinge mitzuteilen, die sich nicht in der unmittelbaren räumlichen bzw. zeitlichen Umgebung befinden.
- Zwei Ebenen: Menschliche Sprache besitzt eine syntaktische und eine semantische Ebene.
- Metalinguistik: In menschlicher Sprache kann über die menschliche Sprache selbst gesprochen werden.

Die oben angeführten Charakteristika natürlicher Sprache werden unterschiedlich gewichtet. Pinker [87] definiert Sprache im klassischen Sinne als ein System, welches zwei Komponenten – Worte und Grammatik – umfasst. Die Fähigkeit, Sprache zu entwickeln, ist dabei dem Menschen angeboren. Pinker spricht in diesem Zusammenhang auch von einem Sprach-Instinkt (*Language Instinct*). Für die moderne Linguistik spielt der grammatikalische Aspekt von Sprache eine herausragende Rolle. Chomsky [22, 21, 23] sieht in der Syntax den Kern und das wichtigste konstituierende Merkmal menschlicher Sprache.

Von besonderer Bedeutung für die menschliche Sprache ist auch die symbolische Natur von Worten und komplexeren sprachlichen Äußerungen [25]. Symbolische Zeichen sind etwas anderes als bloße Signale, die beliebige Arten von Information transportieren können.

Neuere Ergebnisse legen nahe, dass Kommunikationssysteme im Tierreich zumindest in qualitativer Hinsicht über die Eigenschaften Willkürlichkeit, kulturelle Weitergabe, räumliche bzw. zeitliche Versetzung und symbolverarbeitende Mechanismen verfügen. Ob die Eigenschaften Diskretheit und Produktivität bei Tieren vorliegen, ist umstritten. Die Syntax natürlicher Sprachen mit der von bioakustischen Zeichensystemen zu vergleichen, erscheint gänzlich unzweckmäßig [22]. Tatsache ist zwar, dass viele akustische Prozesse im Tierreich (z.B. Gesänge von Vögeln) gut mit *formalen Sprachen* modelliert werden können, die Komplexität solcher Modelle aber ist sehr gering im Vergleich zu der von Grammatiken natürlicher Sprachen. Gänzlich nicht nachgewiesen werden konnte die Eigenschaft Metalinguistik.

Sprache dient der Kommunikation, weist aber auch eine Reihe von zusätzlichen Eigenschaften auf. Somit ist Sprache und Kommunikation nicht dasselbe [8]. Sprachliche Kommunikation ist aber die wahrscheinlich effizienteste Lösung, wenn Informationen in hochgradig komprimierter Form übertragen bzw. verarbeitet werden müssen. Sie scheint aus der Notwendigkeit einer Effizienzsteigerung von Kommunikationsprozessen heraus entstanden zu sein [87]. Eine solche Effizienzsteigerung kann auf unterschiedlichem Wege erreicht werden, so (1) durch eine Steigerung der Geschwindigkeit aller beteiligten informationsverarbeitenden Systeme, (2) durch eine

Erhöhung von Gedächtniskapazitäten, (3) durch eine Parallelisierung informationsverarbeitender Prozesse, (4) durch eine Vergrößerung des Inventars kommunikativer Signale und (5) durch den Einsatz von regelbasierten Zeichensystemen.

In der Tierwelt, einschließlich der Spezies *Homo Sapiens*, lassen sich alle diese Prinzipien beobachten. Die ersten vier führen zu eingeschränkten Effizienzsteigerungen. Nur das letzte Prinzip – die Verwendung von regelbasierten Zeichensystemen – bedeutet eine sprunghafte Effizienzsteigerung um viele Größenordnungen: Durch Produktionssysteme, welche Zeichenentitäten und hierarchisch organisierte syntaktische Regeln umfassen, kann Information maximal effizient organisiert, repräsentiert und kommuniziert werden.

Viele Wissenschaftler gehen heute davon aus, dass es in der Evolutionsgeschichte der Sprache einen besonderen Moment gegeben haben muss, an dem erstmals ein kognitives Prinzip zum Tragen kam, welches unseren Vorfahren im Verlauf ihrer Ontogenese gestattete, kommunizierte Informationen mit Hilfe von regelbasierten Zeichensystemen im Gedächtnis zu organisieren. Dieser Moment kann als Wendepunkt in der Evolution der natürlichen Sprache verstanden werden. Hierdurch wurde das Kommunikationssystem der Vormenschen plötzlich um ein Vielfaches leistungsfähiger als die Systeme ihrer Kospezies.

Funktional referentielle Kodierung

Neben strukturellen Betrachtungen der Vokalisationen von Tieren ist die Funktion der verschiedenen Laute von Interesse [69]. Aus traditioneller ethologischer Perspektive schrieb man den Vokalisationen der Tiere rein affektiven Charakter zu. In jüngerer Vergangenheit wurde diese Sichtweise zusehends in Frage gestellt. Es zeigte sich, dass die Verwendung des Konzeptes einer symbolischen Referenzebene auch dann sinnvolle Erklärungen zulässt, wenn das alte Erklärungsmodell versagt.

Viele Tierarten benutzen Alarmrufe und Futterrufe, die wirken, als besäßen sie eine semantische Referenz. Seyfarth et al. (1980) zeigten, dass die grüne Meerkatze (*Cercopithecus Aethiops*) über drei verschiedene Alarmrufe verfügt, denen drei verschiedene Feinde – Schlangen, Leoparde und Adler – korrespondieren. Dies lässt sich aus den durch die Rufe ausgelösten spezifischen Verhaltensweisen der Tiere schließen: Der Schlangen-Alarmruf bewirkt, dass die Tiere aufstehen und sich umsehen. Der Leoparde-Alarmruf bewirkt, dass die Tiere in die Baumkronen flüchten, und der Adler-Alarmruf führt dazu, dass die Tiere sich im Buschwerk verstecken [103]. Man spricht in diesem Zusammenhang auch von „funktional referentieller Kodierung“ (*referential signalling*). Funktional referentielle Rufe sind Laute mit diskreter akustischer Struktur, kontextunabhängiger Perzeption und erlerntem stimuluspezifischem Einsatz (nach [33]).

Bemerkenswert in diesem Kontext ist eine Beobachtung von Marler (1998): Die verschiedenen Rufe der Meerkatzen scheinen angeboren zu sein, lediglich die Assoziation zu den drei Feind-Typen wird erlernt. Jungtiere verwenden in verschiedensten Situationen den Adler-Alarmruf, nur erwachsene Tiere verwenden ihn ausschließlich

in Situationen, in denen tatsächlich Adler oder ähnliche Vögel in der Nähe sind [69]. Seit der Studie von Seyfarth et al. (1980) wurden Rufe mit funktionaler Referenz bei einer Vielzahl anderer Spezies beobachtet [77, 34, 47, 70]. Hierzu gehören (1) das Haushuhn (*Gallus gallus domesticus*) und sein wilder Verwandter, das Bankivahuhn (*Gallus gallus*) [34, 69, 48], (2) einige Arten von Sperlingsvögeln (*Passeriformes*) [69], (3) Murmeltiere (*Marmota*) [69], (4) verschiedene Arten von Baumhörnchen (*Sciurini*) [45, 69] und (5) Primaten (*Primates*) [69, 48].

Häufiger noch als Alarm- und Futterrufe sind Vokalisationen in sozialem Kontext [32, 27]. Für komparative Studien sind diese Arten von Lauten besonders interessant, jedoch wirft die Aufdeckung von funktionalen Referenzen hier weitaus größere Probleme auf. Cheney et al. (1995, 1996) versuchten in aufwendigen Studien die sozialen Vokalisationen von Pavianen zu entschlüsseln. Sie beobachteten, wie die Mitglieder einer Gruppe Informationen aus den Rufen ihrer Artgenossen übernehmen, jedoch konnten sie keine funktionalen Zusammenhänge für das Sozialverhalten der Tiere nachweisen [19, 18]. Rendall et al. (1999) fanden, dass Paviane Kontaktrufe (*Contact Barks*) verwenden, um zerstreute Mitglieder einer Gruppe zusammenzuhalten [93].

Eine wichtige Eigenschaft der Rufe mit funktionaler Referenz besteht in ihrer nicht-ikonischen Lautstruktur [72]. Nicht-Ikonizität ist auch eines der Merkmale menschlicher Sprachlaute. Ob die Rufe intern tatsächlich eine symbolische Repräsentation ansprechen, ist Gegenstand der Forschung. Nach Martin (1998) könnten neurologische Studien mit bildgebenden Verfahren hier weiteren Aufschluss geben: Das mentale Abbild, welches eine Meerkatze von einem Adler besitzt, sollte nicht nur beim Sehen eines Adlers aktiviert werden, sondern auch beim Hören des entsprechenden Rufes [71]. Die Aktivierung eines mentalen Abbildes auf neurologischer Ebene wird hier mit der bloßen affektiven Auslösung einer Handlung kontrastiert. Das Aktivitätsmuster wäre das neurologische Korrelat der symbolischen Repräsentationsebene, und somit hätten die Rufe tatsächlich symbolischen Charakter.

Tierpsychologische Forschung

Die tierpsychologische Forschung hat in den letzten Jahren immer wieder zu kuriosen Ergebnissen geführt. Diese legen nahe, dass die Ontogenese bei der Ausbildung von kognitiven und kommunikativen Fähigkeiten für viele Spezies von größerer Bedeutung ist, als noch vor wenigen Jahrzehnten angenommen wurde.

In den 1970er Jahren gab es Versuche, Affen und großen Tümmlern menschliche Sprache beizubringen. Diese Versuche ergaben keine greifbaren Ergebnisse [26, 25]. Teilweise ist die schlechte Performanz der Tiere in diesen Experimenten auf die Anatomie der Vokalisationstrakte der Tiere zurückzuführen⁶. Besser schneiden Tiere ab, wenn das Experimentaldesign diesem Umstand Rechnung trägt und Zeichen-

⁶Die Fähigkeit vieler Vogelarten, menschliche Sprachlaute zu imitieren, ist bekannt. Landlebende Säugetiere scheinen dies generell nicht zu tun. Deacon (1977) weist aber auf einen Seehund hin, der menschliche Sprache imitierte [25].

modalitäten verwendet, die von den beteiligten Tierarten besser beherrscht werden können. Interessante Ergebnisse mit solchen gewissermaßen für Tiere ergonomisch optimierten artfremden künstlichen Zeichensystemen lieferten Studien an Schimpansen [96, 43], Bonobos [99], Gorillas [84, 83], Orang-Utans [76, 25], großen Tümmlern [51] und Papageien [86, 85].

Weitreichende Fähigkeiten zum Umgang mit artifiziellen Zeichensprachen ließen sich bei Schimpansen nachweisen. Der Schimpanse Washoe z.B. konnte eine große Anzahl von Handzeichen richtig interpretieren und einsetzen [43]. Die Schimpansin Lana erlernte den Umgang mit Symbolen auf Plastikchips und einer Computertastatur [96]. Savage-Rumbaugh et al. (1978, 1980) führten ein Experiment mit den Schimpansen Sherman und Austin durch, die in sozialem Kontext den Umgang mit einer Anzahl von Symbolen erlernen sollten. Die Faktoren Kommunikation, Verstehen und Produktion standen hier im Zentrum der Untersuchung. Das Experiment zeigte, dass Schimpansen fähig sind, sich in sozialem Kontext mit Hilfe von Symbolen zu verständigen [25]. Nach Savage-Rumbaugh und Lewin (1994) lag den beobachteten Kommunikationsprozessen eine kommunikative Absicht (*Communicative Intent*) zugrunde. Die gemeinsame symbolische Referenzebene wurde dabei von den Schimpansen Sherman und Austin selbst festgesetzt, indem einer der beiden abwechselnd auf einen Gegenstand (z.B. eine Frucht) und auf ein bestimmtes Symbol auf einer Tastatur zeigte. Danach diente das Symbol auf der Tastatur den Tieren als Bezeichner für den betreffenden Gegenstand [98]. Ältere Schimpansen erlernen den Umgang mit Zeichensystemen erheblich schwerer als Jungtiere. Für einen erfolgreichen Lernprozess bei Jungtieren kann eine Immersion in eine entsprechende Lernumwelt schon ausreichend sein. So versagte die Bonobo-Mutter Matata bei einem Lernexperiment, während ihr beiläufig anwesendes Junges Kanzi die Aufgabe schon bald mit Leichtigkeit meisterte [97]. Hierin liegt eine Parallele zum Spracherwerb beim Menschen, der seine Muttersprache ebenfalls am besten durch Immersion in eine Sprachumwelt während einer kritischen Periode der Kindheit erlernt [25, 46, 7]. Nach Savage-Rumbaugh et al. (1993) liegt die kritische Periode für das Erlernen von Symbolen bei Schimpansen im Alter von $0 - 2\frac{1}{2}$ Jahren. Danach verlieren die Tiere ihre Fähigkeit zum Erlernen von Symbolen gänzlich. Die Fähigkeit, während der kritischen Periode symbolische Zeichensysteme zu erlernen, reichte bei Kanzi so weit, dass er einfache Sätze wie „Give the lighter to Rose“ in gesprochenem Englisch verstand. Er erreichte dabei angeblich das Niveau eines etwa $2\frac{1}{2}$ jährigen Kindes.

Große Tümmler sind relativ einfach in Gefangenschaft zu halten und gelten als kooperativ und lernbereit. Herman et al. (1984) führten einige Studien durch [51], die sich von den oben beschriebenen dadurch unterscheiden, dass sie vorwiegend auf das Verständnis von symbolischen Zeichensystemen abzielen. Die Delphine Akeakamai und Phoenix wurden darin unterrichtet, Sätze von zwei verschiedenen Zeichensprachen zu verstehen und darauf zu reagieren. Bei der einen handelte es sich um eine Zeichensprache, die auf Gesten und Handzeichen eines Trainers basierte. Bei der anderen handelte es sich um eine Zeichensprache, die Piffe verwendete, welche von einem Computer generiert und über einen Unterwasserlautsprecher übertragen wur-

den. Beide Zeichensysteme besaßen eine einfache wohlgeformte Syntax, welche die Reihenfolge von Wortklassen festlegte. Die Sätze dieser Sprachen hatten einen imperativen Charakter, d.h. es handelte sich um Handlungsanweisungen, die von den Delphinen ausgeführt werden mussten. Grammatische Subjekte waren nicht in den Sätzen enthalten. Die für Phoenix verwendete Grammatik verwendete die Reihenfolge: Direktes Objekt – Aktion – Indirektes Objekt. Die für Akeakamai verwendete Grammatik verwendete die umgekehrte Reihenfolge: Indirektes Objekt – Aktion – Direktes Objekt. Die Zeichensprachen enthielten auch Modifikatoren für Objekt- und Aktionszeichen. Manche Handlungsanweisungen hatten nur ein direktes Objekt, andere beinhalteten beide Arten von Objekten. Gleiche Gegenstände konnten sowohl als direktes, als auch als indirektes Objekt verwendet werden. Beide Tiere lernten, auf Sätze in ihrer Zeichensprache korrekt zu reagieren, so dass Handlungsanweisungen wie „Hoop fetch pipe“ (Fetch the hoop to the pipe!) und „Pipe fetch hoop“ richtig ausgeführt wurden. Die Erfolgsquote für neue, nicht im Training verwendete Sätze, lag für beide Tiere bei 80% und darüber. Bei semantisch reversiblen Sätzen, in denen eine richtige Interpretation der Syntax wichtig war, erzielte Phoenix 59% und Akeakamai 77%. Das direkte Objekt wurde in über 90% der Fälle richtig identifiziert. Herman et al. (1984) schließen aus diesen Ergebnissen, dass große Tümmler in der Lage sind, willkürliche festgesetzte Regeln und Symbole zu erlernen und danach Semantik, Wortreihenfolge und Wortklassen richtig zu interpretieren. Sie weisen auch auf die Unabhängigkeit von sensorischen Modalitäten hin: Die artifizielle Pfiffsprache weist Ähnlichkeiten mit dem natürlichen Kommunikationssystem der Tiere auf, die Gestensprache tut dies nicht. Die Zeichen der Gestensprache hatten ikonischen Charakter, die der Pfiffsprache hatten dies nicht. Dennoch konnten beide Zeichensprachen in etwa gleich gut erlernt werden.

Graupapageien (*Psittacus erithacus*) leben in engen sozialen Verbänden in einer hochkomplexen Lebensumwelt. Dies ist sicherlich ein wichtiger Grund für ihre bemerkenswerten kognitiven und kommunikativen Fähigkeiten. Pepperberg et al. (1998, 1999, 2001) gelang es, mit Hilfe von gesprochenen Wörtern der englischen Sprache mit Graupapageien zu kommunizieren. Pepperberg et al. nutzten bei ihren Experimenten das *Model-Rival*-Lernparadigma⁷ nach Todt (1975) [110], welches sich als geeigneter herausstellte als klassische Konditionierungstechniken, die in vorausgegangenen weniger erfolgreichen Studien genutzt wurden. Zu den von Pepperberg, Todt und anderen gezeigten Fähigkeiten von Graupapageien gehören das korrekte Bezeichnen von Objekten (*Object Labeling*), der Gebrauch hierarchischer Konzepte und die richtige Interpretation einfacher Aussagesätze [86, 85].

Es ist bekannt, dass Vogelgesänge kulturell weitergegeben werden und regionale Dialekte keine Seltenheit sind [120, 6]. Das Erlernen des Gesangs scheint bei einigen Vogelarten angeborenen restriktiven Prozeduren zu folgen. Sehr viele Vogelarten müssen den Gesang ihrer ausgewachsenen Artgenossen während einer kurzen sensiblen Periode hören, um selbst entsprechende Fähigkeiten zu entwickeln [120]. Wha-

⁷Es handelt sich um ein soziales Lernparadigma. Eine Übersicht über soziale Lernparadigmata findet sich in [63]

ling et al. (1997) sprechen in diesem Zusammenhang auch von einem „Song acquisition device“ [117]. Die Parallele zur menschlichen Sprache, jedoch auch Unterschiede werden hier sichtbar: Zwar scheinen auch wir in unserer Kindheit Sprachen wesentlich schneller und besser zu erlernen, jedoch ist das Erlernen einer oder mehrerer Fremdsprachen auch im Erwachsenenalter noch gut möglich. Auch die dem Sprachrespektive Gesangserwerb zugrundeliegenden physiologischen und neuropsychologischen Strukturen unterscheiden sich erheblich. Es wurde daher argumentiert, dass komparative Studien zur Sprachevolution besser auf Säugetieren aufbauen sollten [120]. Interessanterweise existiert hier eine Reihe von Spezies, deren Vokalisationen mehr dem Gesang von Vögeln als der menschlichen Sprache ähneln. Hierzu gehören Walgesänge, die wahrscheinlich auch kulturell weitergegeben werden [118, 80], die Gesänge von Gibbons [49], mit denen wir relativ eng verwandt sind, und auch die Ultraschallgesänge von Mäusen[54].

1.2.2 Perspektiven für KDD-Methoden

In den letzten Jahren haben sich im Bereich der Hard- und Software revolutionäre Änderungen vollzogen, die erhebliche Auswirkungen auf viele Aspekte der Bioakustik haben. Es ist nun möglich, fast das gesamte Methodenarsenal dieser Disziplin von Analog- auf Digitaltechnik zu portieren. In Analogie zur Computer-Linguistik wird die traditionelle Bioakustik mehr und mehr zur Computer-Bioakustik.

Der Rechner erleichtert den Umgang mit bioakustischem Material in vielfältiger Art und Weise. Studien, die bisher sehr aufwändig waren, sind nun leicht zu realisieren. Dies wird u.a. durch folgende Neuerungen bewirkt:

- **Aufzeichnung von akustischen Signalen:** Durch Hard-Disk-Recording von Audiosignalen (siehe auch Teil II) entfällt heute die Notwendigkeit große teure Tonbandgeräte zu benutzen. Audiosignale können mit entsprechender Soundkarte in Echtzeit auf Festplatte geschrieben werden.
- **Archivierung und Verwaltung von Audiodaten:** Akustisches Material und damit in Verbindung stehende Daten lassen sich auf dem Rechner archivieren und in Datenbanken verwalten. Es wird wesentlich einfacher, große Mengen auch heterogener Daten zu verwalten und im Bedarfsfall darauf zuzugreifen.
- **Wissenschaftliche Visualisierung von Audiodaten:** Zeit-Frequenz-Darstellungen von Audiodaten lassen sich viele Dutzend mal schneller als noch vor zehn Jahren berechnen. Es ist möglich, ergonomische visuelle interaktive Schnittstellen zu großen digitalen Audiodatenbanken zu schaffen. Es vollzieht sich ein Wandel vom wissenschaftlichen Diagramm zum visuellen wissenschaftlichen Dateninterface.
- **Digitale Signalverarbeitung:** Hierzu gehören digitale Filter, Zeit-Frequenz-Analysen und vieles mehr. Auch hier liegt die entscheidende Neuerung darin,

dass Rechengeschwindigkeit und Speicherkapazitäten heute ausreichen, um mit bioakustischen Daten sinnvoll zu arbeiten. Eine Übersicht über DSP-Verfahren in der Bioakustik findet sich in [55].

KDD-Methoden bilden innerhalb der Computer-Bioakustik einen eigenen Methodenkomplex mit speziellen Anwendungsperspektiven. Der Einsatz von formalen Modellen, maschineller Wissensverarbeitung und KDD-Verfahren in der Bioakustik bedeutet eine qualitative Neuerung gegenüber traditionellen Methoden. Es liegt hier eine Methodenklasse mit vielen interessanten Anwendungsperspektiven vor:

1. **Wissensbasierte Modellierung von ASK:** Klassifikatoren für Tierlaute können verwendet werden, um bioakustische Audiodaten automatisch oder semiautomatisch zu klassifizieren. Mit Hilfe von KDD- und Data-Mining-Methoden können solche Klassifikatoren wissensbasiert modelliert werden (siehe folgende Kapitel).
2. **Erlernen von syntaktischen Beschreibungen:** Methoden des strukturellen *Dependency Modeling* und der grammatischen Inferenz (*Grammatical Inference*) können verwendet werden, um syntaktische Beziehungen zwischen einzelnen Lauten auf der Grundlage von Mengen von Lautsequenzen aufzufinden.
3. **Abweichungsentdeckung:** *Change and Derivation Detection* kann verwendet werden, um in Mengen von Lauten oder Lautsequenzen statistisch auffällige Veränderungen und Abweichungen der akustischen und/oder der temporalen Strukturen zu entdecken. Diese wiederum können mit Hilfe des *Dependency Modelling* Kontextinformationen zugeordnet werden.
4. **Aufdecken referentieller Kodierungen:** Durch eine Auswertung von Beobachtungsprotokollen und symbolischen Beschreibungen von Audioaufnahmen mit Methoden des *Dependency Modeling* können Korrelationen zwischen Lautkategorien und Verhaltensweisen aufgedeckt werden. Die hierzu erforderlichen symbolischen Beschreibungen von akustischen Prozessen können mit Hilfe entsprechender Audio-Signal-Klassifikatoren erhalten werden.

KDD-Methoden sind im Rahmen der Forschung an bioakustischen Signalsystemen einerseits bloßes Hilfsmittel, welches eine Beschleunigung der Datenauswertung und eine Automatisierung von Klassifikationsaufgaben gestattet. Andererseits besitzen sie das Potential, den wissenschaftlichen Theoriebildungsprozess wesentlich zu unterstützen. Data-Mining-Verfahren sind in der Lage, auch solche statistisch auffälligen Zusammenhänge in empirischen Daten aufzudecken, die man mit dem unbewaffneten Verstand nicht erkennen kann. Dies kann zum Finden neuartiger, für bioakustische Signalsysteme maßgeblicher Gesetzmäßigkeiten führen.

Kapitel 2

KDD und Audio-Signal-Klassifikatoren

Generell lassen sich KDD- und Data-Mining-Verfahren zur Auswertung von Beobachtungsdaten aller Art hinzuziehen. Symbolische Data-Mining-Algorithmen können allerdings mit Audiodaten nicht ohne weiteres umgehen (vgl. Kapitel 1). In empirischen KDD-Studien der Bioakustik müssen daher alle in den Audiodaten enthaltenen relevanten Informationen auf eine symbolische Repräsentationsebene abgebildet werden.

Die Abbildung der Audiodaten in die symbolische Repräsentationsebene kann auch als *Annotation* bezeichnet werden, da die Daten selbst hierbei in der Regel nicht verändert werden. Sie basiert auf **Audio-Signal-Klassifikatoren** (ASK), also Algorithmen, die in der Lage sind, akustische Ereignisse in Audiodaten zu lokalisieren und zu klassifizieren.

In vielen Bereichen der Bioakustik ist das Problem der Modellierung von effizienten ASK selbst ein wissensintensives Problem. Es bedarf hierzu nicht nur der Kenntnis der akustischen Phänomene, sondern auch wissenschaftlicher Theorien über die Prozesse, welche ihnen zugrunde liegen. Die Modellierung von bioakustischen ASK kann als eigenständiger KDD-Prozess verstanden werden. Im folgenden Kapitel wird untersucht, wie sich dieser aus Sicht des KDD-Paradigmas darstellt.

2.1 Manuelle versus maschinelle Annotation

Aus Mangel an geeigneten Verfahren wurden bisher in der Mehrzahl der Fälle keine maschinellen Systeme zur Annotation von bioakustischen Audiodaten eingesetzt. In den meisten empirischen Untersuchungen wird die Klassifikation bzw. Annotation der Audiodaten auch heute noch von Hand vorgenommen. Ein solches Vorgehen erscheint nicht effizient, denn die manuelle Klassifikation hat viele Nachteile:

1. Eingeschränkte Verfügbarkeit und hohe Kosten menschlicher Klassifikatoren;

2. Nur geringe Datenmengen können bei gleichzeitig hohem Zeitaufwand ausgewertet werden;
3. Menschen müssen für die Klassifikationsaufgaben geschult werden;
4. Klassifikationsentscheidungen von Menschen sind fehlerbehaftet und nicht immer transparent;
5. Klassifikationsentscheidungen von Menschen können formal nur schlecht evaluiert werden;
6. Nicht alle Klassifikationsaufgaben können von Menschen gelöst werden.

Ungeachtet dieser Nachteile ist der Einsatz von menschlichen Klassifikatoren in vielen Fällen die einzige Möglichkeit, bioakustische Daten zu annotieren, denn passende maschinelle Verfahren existieren bisher nur für sehr wenige Problemklassen. Insbesondere das maschinelle Erkennen komplexer akustischer Muster, die durch variable Zeit- oder Frequenzparameter charakterisiert sind, bereitet größere Probleme. In diesem Bereich scheinen menschliche Klassifikatoren derzeit noch überlegen zu sein. Gegenwärtig finden Verfahren zur maschinellen Klassifikation von Tierlauten allerdings zusehends Eingang in die Wissenschaft (siehe z.B. [42, 17, 20, 44, 50, 88]).

2.2 Unterschiede zur Spracherkennung

Im Falle des wahrscheinlich komplexesten bioakustischen Kommunikationssystems – der menschlichen Sprache – ist das Problem der maschinellen Audio-Signal-Klassifikation in weiten Teilen bereits gelöst. Moderne Spracherkennungstechnologie leistet eine Abbildung aus der Repräsentationsebene digitalisierter Audiodaten in die Ebene von Buchstaben, Worten und Sätzen. Warum also existieren entsprechende Technologien nicht auch für die Bioakustik?

Für den Mangel an geeigneten Algorithmen zur Klassifikation von bioakustischen Audiodaten existiert bei näherer Betrachtung eine Reihe von Ursachen, die in ihrer Summe dazu geführt haben, dass diese Problemklasse bis heute für die meisten Tierarten unbearbeitet ist:

1. Spracherkennung basiert auf der jahrzehntelangen Arbeit vieler Forschungsgruppen im universitären Bereich und soliden Investitionen von Großkonzernen. Die Bioakustik ist personell und materiell bei weitem nicht so gut ausgestattet.
2. Die Sprachwissenschaft ist eine traditionsreiche Wissenschaft, die im Verlauf der letzten Jahrhunderte die menschliche Sprache sehr detailliert erforscht hat. Die Bioakustik steht erst am Anfang ihrer Entwicklung [109].

3. Die akustischen Kommunikationssysteme im Tierreich sind zwar ein jedes für sich mit einiger Sicherheit weit weniger komplex als die menschliche Sprache und also im Prinzip auch einfacher zu modellieren. Es existieren aber sehr viele Tierarten, die auf akustischer Ebene kommunizieren – so leben weltweit ca. 10000 Vogelarten und rund 5500 Säugetierarten.
4. Viele Zeichensysteme im Tierreich sind komplexer, als man gemeinhin annimmt. Alleine die Männchen der Nachtigall beherrschen ca. 120-260 unterschiedliche Strophentypen [56, 111].
5. Die Laute von vielen Tieren weisen eine komplizierte Charakteristik von physikalischen Merkmalen auf. Diese lassen sich mathematisch nur schwer fassen. Die Programmierung von Algorithmen zum Erkennen solcher Charakteristika kann sehr aufwendig sein.
6. Die Form und Bedeutung von Lauten der menschlichen Sprache ist für uns intuitiv verständlich und durch Introspektion vertraut. Maschinelle Spracherkennung kann auf diesem Wissen aufbauen. Im Falle von Tierlauten ist dies generell nicht möglich.

Der biologische Sinn von bioakustischen Signalen kann nicht analog zu der semantischen Ebene menschlicher Sprachlaute verstanden werden. Bioakustische Signalsysteme müssen notwendigerweise andersartig modelliert werden. Die Klassifikation bioakustischer Signale muss einerseits auf den physikalischen Eigenschaften der akustischen Prozesse und andererseits auf einem theoretischen Netzwerk wissenschaftlicher Annahmen über deren Form und Funktionen aufbauen. Ein solches Netzwerk zu erstellen ist nicht nur ein algorithmisches, sondern vielmehr ein interdisziplinäres Problem, in dem Expertenwissen eine zentrale Rolle spielt.

2.3 Das Annotationsproblem der Bioakustik

Die Aufgaben maschineller Annotation von bioakustischen Audiodaten sind problemspezifisch. Sie variieren in Abhängigkeit von der jeweiligen wissenschaftlichen Fragestellung. Grundsätzlich müssen die Daten aus einem hochdimensionalen Raum in einen Raum mit geringerer Dimension transformiert werden, ohne dass es hierbei zu einem Verlust von relevanten Informationen kommt. Durch die Annotation muss die Komplexität weiterführender Analysen soweit reduziert werden, dass diese mit akzeptablem Aufwand durchgeführt werden können.

Maschinelle Annotation basiert auf der Klassifikation von Audiodaten mit Hilfe von ASK. Ein System von ASK kann als ein Klassifikationssystem für Audiodaten verstanden werden. Die Modellierung von bioakustischen ASK kann relativ einfach, aber auch sehr komplex sein. In vielen Fällen handelt es sich um ein nicht triviales Problem der Programmierung von Mustererkennungsalgorithmen. Folgende Aspekte sind hierbei zu unterscheiden:

1. die Aufgaben der Annotation im jeweiligen wissenschaftlichen Kontext,
2. der Prozess der Erstellung des Klassifikationssystems,
3. die Struktur des Klassifikationssystems,
4. die Eigenschaften der Klassifikatoren,
5. die Veranschaulichung von Klassifikatoren und Klassifikationsentscheidungen,
6. der Prozess der Klassifikation von Audiomaterial,
7. die Evaluation von Klassifikatoren und Klassifikationssystem.

Die grundlegende Aufgabe eines Systems zur maschinellen Annotation besteht darin, in einer Menge A von Audiodateien mit Hilfe einer Menge von algorithmischen ASK alle Instanzen der Elemente einer Menge R von Repräsentationen akustischer Phänomene zu finden. Zu jeder gefundenen Instanz wird ggf. eine Menge von Umgebungsvariablen gemessen. Die Instanzen und die damit verbundenen Messwerte werden zu Annotationsfakten (im Folgenden auch Annotationen) zusammengefasst und in einer Datenbank o.ä. gespeichert. Anschließend können sie visualisiert und ausgewertet werden. Abbildung 2.1 veranschaulicht diesen Prozess.

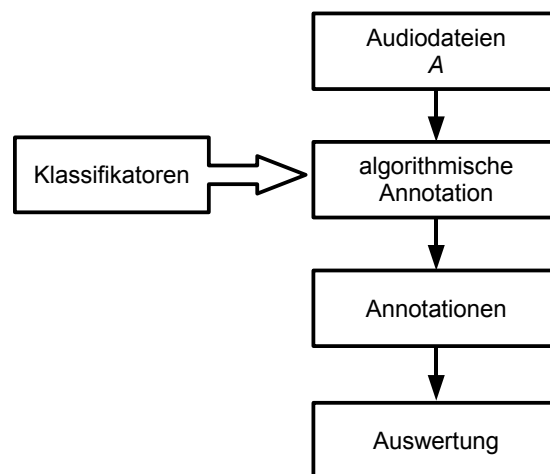


Abbildung 2.1: Annotation von Audiodaten

Beispiel: Gegeben sei eine Spezies, die über ein akustisches Zeichensystem mit einem Inventar von drei Zeichen $Z = \{z_1, z_2, z_3\}$ verfügt. Jedes Zeichen in Z sei durch eine Anzahl akustischer Phänomene gekennzeichnet, so dass $|R| \geq |Z|$. Die Zeichen werden von der Spezies sequentiell ohne direkt erkennbare Ordnung emittiert. Eine empirische Untersuchung soll die Frage klären, ob die akustischen Zeichen der Spezies regelhaft in Sequenzen oder in zufälliger Abfolge angeordnet sind. Für die

Untersuchung steht eine Sammlung A von Audiodateien mit akustischen Signalen der Spezies zu Verfügung, welche eine unbekannte Anzahl von Zeichen enthält. Idee ist, die Daten mit Hilfe eines geeigneten Data-Mining-Verfahrens zum Aufdecken struktureller Abhängigkeiten in Symbolsequenzen zu analysieren.

Das Verfahren benötigt als Input eine Menge F von symbolischen Fakten, die Informationen über alle Zeitpunkte des Auftretens der verschiedenen Zeichen in A enthält. Jedes Faktum $F_i \in F$ muss die Form $z(\tau)$ haben, wobei z Symbol für ein Zeichen aus Z ist und τ einen Zeitpunkt in A repräsentiert. Die Aufgabe ist also, die Dateien in A so zu annotieren, dass durch die Annotation die Menge F erhalten wird.

Ist $|R| = |Z|$, so beschränkt sich die Aufgabe darauf, für jedes Element in R genau einen ASK zu finden. Ist $|R| > |Z|$ muss bei der Annotation berücksichtigt werden, durch welche Konstellation welcher akustischen Phänomene die verschiedenen Zeichen in A zu erkennen sind. Es ergeben sich zwei verschiedene Annotationsprobleme für bioakustische Daten:

- **Phänomenorientierte Annotation:** Finde für eine Menge R von Repräsentationen akustischer Phänomene eine Menge von korrespondierenden ASK, so dass in einer Menge A von Audiodateien möglichst jede Instanz jedes Elementes aus R mit Hilfe eines ASK identifiziert werden kann. Bilde mit Hilfe der ASK und A die Menge der phänomenorientierten Annotationen.
- **Zeichenorientierte Annotation:** Finde für eine Menge Z von Zeichen eines bioakustischen Zeichensystems eine Menge von korrespondierenden ASK, so dass in einer Menge A von Audiodateien möglichst jede Instanz jedes Elementes aus Z mit Hilfe eines ASK identifiziert werden kann. Bilde mit Hilfe der ASK und A die Menge der zeichenorientierten Annotationen.

Ist unbekannt, durch welche Konstellation welcher akustischen Phänomene die verschiedenen Zeichen in Z zu erkennen sind oder wieviele Zeichen das Zeicheninventar einer Spezies umfasst, kann versucht werden, das Wissen hierüber im Rahmen eines KDD-Prozesses zum Entdecken von zeitlich-strukturellen Abhängigkeiten in Mengen von Klassifikationsentscheidungen zu gewinnen.

Ein algorithmischer ASK kann explizit oder im Rahmen eines KDD-Prozesses modelliert werden (siehe Abschnitte 2.5 und 2.6). Eine explizite Programmierung, d.h. eine Programmierung ohne zwischengeschalteten Data-Mining-Schritt, kann vorgenommen werden, wenn das Wissen, welches zur Programmierung des Klassifikators notwendig ist, vollständig bekannt und vollständig verwertbar ist. Ist dies nicht der Fall, kann die Modellierung durch einen geeigneten KDD-Prozess wesentlich unterstützt bzw. überhaupt erst ermöglicht werden.

Lassen sich folgende Fragen nicht eindeutig beantworten, so deutet dies darauf hin, dass das Wissen für die Modellierung eines bioakustischen ASK nicht hinreichend ist:

- Korrespondiert einem bestimmten akustischen Phänomen tatsächlich ein biologisch relevanter Prozess oder handelt es sich lediglich um Noise?

- Korrespondiert einem bestimmten akustischen Phänomen genau ein Zeichen eines bioakustischen Zeichensystems oder ist das Phänomen lediglich eines von mehreren Merkmalen eines Zeichens?
- Ist ein komplexes akustisches Phänomen in diskrete Einzelphänomene dekomponierbar oder nicht? Welche Regeln gelten für die Dekomposition?
- Ist ein bestimmtes, scheinbar alleinstehendes akustisches Phänomen Bestandteil eines Lautkomplexes? Wie kann der Lautkomplex beschrieben werden?
- Welche Merkmale eines bioakustischen Zeichens müssen in einem Klassifikator repräsentiert werden? Wie sind die Merkmale zu gewichten?
- Ist ein bestimmter Klassifikationsalgorithmus überhaupt für eine konkrete akustische Phänomenklasse geeignet? Wie müssen die zu einem ASK gehörenden Parameter gewichtet werden?

In der Bioakustik ist eine eindeutige Beantwortung der obigen Fragen nur für sehr einfache Problemstellungen möglich. Das Annotationsproblem der Bioakustik kann daher in den meisten Fällen als Problem der wissensbasierten Modellierung von algorithmischen ASK verstanden werden.

Im Bereich der wissensbasierten Modellierung ist eine Trennung von Expertenwissen und Klassifikator nur schwer vorzunehmen. Ein Klassifikator kann hier als computerimplementiertes Expertenwissen verstanden werden. Zur Modellierung von ASK muss der Experte die Möglichkeit haben, sein Wissen auf diese zu übertragen und ihr Klassifikationsverhalten gemäß seiner Expertise zu evaluieren. Es bedarf daher eines speziellen Systems zur Modellierung von ASK, das Bioakustikern gestattet, ihre Expertise direkt auf wissensbasierte algorithmische Klassifikationssysteme für Audiodaten zu übertragen. Die theoretischen Grundkonzepte eines solchen Systems werden in den nächsten Abschnitten beschrieben.

2.4 Maschinelle Annotation von Audiodaten

Es existierten viele Möglichkeiten, ASK zu modellieren. ASK können nach ihrer Aufgabe, ihrem Input, ihrem Output, ihrer Architektur, der Art und Weise der Programmierung und nach Formen ihres Trainings unterschieden werden. Zudem können ASK kontextfrei oder kontextsensitiv sein.

Ein System zur wissensbasierten Modellierung von bioakustischen ASK darf den Freiraum bei der Gestaltung der ASK nicht zu sehr beschränken, da Tierlaute ausserordentlich vielgestaltig sein können. Um dies zu gewährleisten, wird in diesem Kapitel zunächst der theoretische Grundriss eines solchen Systems beschrieben. Von den Details spezifischer ASK wird dabei abstrahiert. Konkrete Algorithmen und Verfahren finden sich im anschließenden Kapitel.

2.4.1 Grundlegende Konzepte

Eine *Audiodatei* \vec{s} ist eine zeitdiskrete Messreihe (Zeitreihe) von äquidistanten PCM-kodierten Messwerten. Eine Zeitreihe mit N Elementen kann als Vektor der Dimension N oder als Punkt in einem N -dimensionalen Raum verstanden werden:

$$\vec{s} = [s_0, \dots, s_{N-1}], N \in \mathbb{N}, s_n \in \mathbb{Z} \quad (2.1)$$

Ein *Audiokorpus* A ist eine Menge von Audiodateien:

$$A = \{\vec{s}_0, \dots, \vec{s}_{M-1}\}, M \in \mathbb{N} \quad (2.2)$$

Jeder Datei $\vec{s}_m \in A$ kann genau ein Label zugeordnet werden. Die Menge $\Psi = \{0, \dots, M-1\}$ sei die Menge aller Dateilabel des Audiokorpus A .

Jede Datei $\vec{s}_m \in A$ besitzt einen expliziten Zeitstempel, der den Startzeitpunkt t_0^m des Beginns der Aufnahme angibt. Die Elemente (Messwerte) in den \vec{s}_m besitzen keinen expliziten Zeitstempel. Die *Abtastfrequenz* δ der Zeitreihen in A ist konstant. Der Zeitpunkt t des i -ten Elementes in jedem $\vec{s}_m \in A$ ist implizit gegeben durch:

$$t(\vec{s}_m(i)) = t_0^m + \delta i, \delta \in \mathbb{R} \quad (2.3)$$

Eine *Sequenz* benachbarter Elemente in \vec{s} von einem Element mit dem Zeitpunkt t_1 bis zu einem Element mit dem Zeitpunkt t_2 , $t_1 < t_2$ wird durch $\vec{s}_{t_1:t_2}$ bezeichnet. Eine nicht näher spezifizierte Subsequenz einer Audiodatei in A wird im Folgenden kurz mit \vec{x} bezeichnet. Die *Länge* l einer Sequenz ist ihre Kardinalität $l(\vec{x}) = |\vec{x}|$. Die zeitliche Ausdehnung Δt einer Sequenz ist gegeben durch:

$$\Delta t(\vec{x}) = \delta l(\vec{x}), \delta \in \mathbb{R} \quad (2.4)$$

Ein *Phänomen* ist eine Erscheinung, die direkt oder indirekt mit den Sinnen wahrnehmbar ist. Direkt wahrnehmbar sind z.B. akustische Erscheinungen, die unmittelbar mit dem Gehörsinn wahrgenommen werden können. Indirekt wahrnehmbar sind z.B. akustische Erscheinungen, die einer wissenschaftlichen Visualisierung (z.B. eines Spektrogramms) bedürfen um wahrgenommen werden zu können. Ein akustisch wahrnehmbares Phänomen heisst im Folgenden kurz *akustisches Phänomen*. Ein visuell wahrnehmbares Phänomen heisst im Folgenden kurz *visuelles Phänomen*.

2.4.2 First-Level-Klassifikatoren

Die zweidimensionale *Visualisierung* V einer Sequenz \vec{x} der Länge l ist eine $x \times y$ -Matrix über \mathbb{R} mit der zeitlichen Ausdehnung $\Delta t^V = \Delta t(\vec{x})$, die durch eine Abbildung $v : \mathbb{Z}^l \rightarrow \{V(x \times y, \mathbb{R})\}$ entsteht, so dass

$$V = v(\vec{x}) \quad (2.5)$$

Die Repräsentation eines akustischen Phänomens ist eine Audiosequenz \vec{x}^ϕ der Länge l , in welcher durch Abspielen eindeutig ein bestimmtes akustisches Phänomen auditiv wahrgenommen werden kann. Die *akustische Signatur* ϕ^A eines akustischen Phänomens ist ein Vektor \tilde{x}^ϕ der Länge l' , welcher \vec{x}^ϕ repräsentiert, aber einfacher als \vec{x}^ϕ aufgebaut ist. Die akustische Signatur wird durch eine Abbildung $E^A : \mathbb{Z}^l \rightarrow \mathbb{R}^{l'}$ erhalten, wobei gilt

$$\phi^A = E^A(\vec{x}^\phi) \quad (2.6)$$

Die Repräsentation eines visuellen Phänomens ist eine Visualisierung V^ϕ , in welcher durch grafische Darstellung eindeutig ein bestimmtes akustisches Phänomen visuell wahrgenommen werden kann. Die *visuelle Signatur* ϕ^V eines visuellen Phänomens ist eine $x' \times y'$ -Matrix über \mathbb{R} , welche das in V^ϕ beobachtbare Phänomen repräsentiert, aber einfacher als V^ϕ aufgebaut ist.

Eine visuelle Signatur wird durch eine Abbildung $E^V : \{V(x \times y, \mathbb{R})\} \rightarrow E^V : \{V(x' \times y', \mathbb{R})\}$ erhalten, so dass

$$\phi^V = E^V(v(\vec{x}^\phi)) \quad (2.7)$$

Ein akustisches Phänomen in A heisst *elementar*, wenn es durch genau eine Signatur exakt charakterisiert und repräsentiert werden kann. Die Signatur eines elementaren Phänomens heisst im Folgenden auch *elementare Signatur*.

Korollar: Jedes wahrnehmbare akustische oder visuelle Phänomen in A ist auf eine Subsequenz $\vec{s}_{t_1:t_2}$ in einem $\vec{s}_m \in A, m < M$ begrenzt. Es existiert für jedes wahrnehmbare Phänomen in A eine akustische oder eine visuelle Signatur. Alle akustischen Phänomene, die in einer Untersuchung von Interesse sind, können durch eine Menge akustischer Signaturen $\Phi^A = \{\phi_0^A, \dots, \phi_{N-1}^A\}, N \in \mathbb{N}$ repräsentiert werden. Alle visuellen Phänomene, die in einer Untersuchung von Interesse sind, können durch eine Menge visueller Signaturen

$$\Phi^V = \{\phi_0^V, \dots, \phi_{N-1}^V\}, N \in \mathbb{N} \quad (2.8)$$

repräsentiert werden.

Die *elementare Problemdomäne* Φ einer Untersuchung ist durch alle akustischen und visuellen Signaturen definiert, die in einem gegebenen wissenschaftlichen Kontext von Interesse sind, so dass

$$\Phi = \{\phi_0, \dots, \phi_{N-1}\} = \Phi^A \cup \Phi^V, N \in \mathbb{N} \quad (2.9)$$

Jeder Signatur $\phi_n \in \Phi$ kann genau ein Klassenlabel zugeordnet werden. Die Menge $\Omega = \{0, \dots, N-1\}, N \in \mathbb{N}$ sei die Menge aller Klassenlabel der Problemdomäne Φ .

Jeder Signatur $\phi_n \in \Phi$ kann eine signaturspezifische Klassenkorrelationsfunktion (im Folgenden auch KKF) zugeordnet werden. Die Menge

$$C = \{c_0, \dots, c_{N-1}\}, N \in \mathbb{N} \quad (2.10)$$

sei die Menge aller KKF der Problemdomäne. Eine KKF $c_n \in C$, erhält als Input eine Sequenz \vec{x} der Länge l . Durch c_n wird der Hypothese, dass in \vec{x} das der Signatur ϕ_n entsprechende Phänomen enthalten ist, eine reellwertige Maßzahl zuordnet. Diese Maßzahl wird im Folgenden auch Klassenähnlichkeit s (*Similarity*) genannt. Eine KKF ist somit eine Funktion $c : \mathbb{R}^l \rightarrow \mathbb{R}$, so dass

$$s = c(\vec{x}), c \in C \quad (2.11)$$

Der Menge C entspricht eine Menge $\Gamma = \{\gamma_0, \dots, \gamma_{N-1}\}$, $N \in \mathbb{N}$ von First-Level-Annotationsalgorithmen (im Folgenden auch First-Level-AA). Ein First-Level-AA $\gamma_n \in \Gamma$ sucht in A mit Hilfe der KKF $c_n \in C$ Instanzen des Elements $\phi_n \in \Phi$. Zeigt die KKF in einem bestimmten Intervall einen Wert an, der einem definierten Kriterium entspricht (z.B. einen festgesetzten Ähnlichkeitsschwellwert überschreitet), wird von dem First-Level-AA ein Annotationsfaktum generiert.

Ein First-Level-ASK (im Folgenden auch FLC = **F**irst-**L**evel **C**lassifier) κ ist ein 3-Tupel der Form

$$\kappa = \langle \phi, c, \gamma \rangle, \phi \in \Phi, c \in C, \gamma \in \Gamma \quad (2.12)$$

Der Problemdomäne Φ korrespondiert die Menge $K = \{\kappa_0, \dots, \kappa_{N-1}\}$, $N \in \mathbb{N}$ von FLC. Ein FLC $\kappa \in K$ erhält als Input eine Audiodatei $\vec{s} \in A$ und generiert daraus eine abzählbare Menge $F^\kappa = \{F_0^\kappa, F_1^\kappa, \dots\}$ von FLC-Annotationen, d.h. $\kappa : A \rightarrow F^\kappa$.

Eine FLC-Annotation $F_i^\kappa \in F^\kappa$ im Zeit-Frequenz-Raum ist ein 5-Tupel der Form

$$F_i^\kappa = \langle \psi, \omega, \tau, f, \Lambda \rangle \quad (2.13)$$

Hierbei ist $\psi \in \Psi$ ein Dateilabel, $\omega \in \Omega$ das Klassenlabel der zu κ gehörenden Signatur, τ ein Zeitpunkt in A , f ein Frequenzwert und $\Lambda = \{\lambda_1, \lambda_2, \dots\}$ eine optional zu besetzende Liste von Variablen, die in einem spezifischen Kontext erforderlich sein können.

Das Ziel der wissensbasierten Modellierung von FLC besteht darin, in einem gegebenen wissenschaftlichen Kontext eine elementare Problemdomäne Φ zu definieren und dazu eine geeignete Menge K von FLC zu finden, mit deren Hilfe die zu Φ gehörenden Phänomene in A zuverlässig lokalisiert werden können. Mit Hilfe von K kann dann aus A maschinell die Menge F^K der FLC-Annotationen erstellt werden.

2.4.3 Second-Level-Klassifikatoren

Oft sind bioakustische Signale und Signalsequenzen zu komplex, um sie durch FLC alleine zu modellieren. Viele komplexe Phänomene können aber als *Konstellationen* von elementaren Phänomenen im Zeit-Frequenzraum verstanden werden.

Ein *symbolisches Element* ϕ^s charakterisiert die Position einer elementaren Signatur $\phi \in \Phi$ in einem Ausschnitt des Zeit-Frequenz-Raums. Es kann als 4-Tupel

der Form

$$\phi^s = \langle \omega, \tau, f, \Lambda \rangle \quad (2.14)$$

aufgefasst werden. Hierbei ist $\omega \in \Omega$ das Klassenlabel der elementaren Signatur, τ ein zu einem hypothetischen Nullpunkt relativer Zeitpunkt, f ein Frequenzwert und $\Lambda = \{\lambda_0, \lambda_1, \dots\}$ eine abzählbare Liste von optional zu besetzenden Variablen.

Die *symbolische Signatur* eines komplexen Phänomens beschreibt eine Konstellation von symbolischen Elementen in einem Ausschnitt des Zeit-Frequenz-Raums. Sie kann als Menge der Form

$$\phi^S = \{\phi_0^s, \dots, \phi_{N-1}^s\}, N \in \mathbb{N} \quad (2.15)$$

aufgefasst werden.

Ein komplexes akustisches Phänomen in A heisst *symbolisch beschreibbar*, wenn es durch eine Konstellation von symbolischen Elementen exakt charakterisiert werden kann. Jedes symbolisch beschreibbare Phänomen in A kann durch eine symbolische Signatur repräsentiert werden.

Die *symbolische Problem-domäne* Φ^S einer Untersuchung ist durch alle symbolischen Signaturen definiert, die in einem gegebenen wissenschaftlichen Kontext von Interesse sind, so dass

$$\Phi^S = \{\phi_0^S, \dots, \phi_{M-1}^S\}, M \in \mathbb{N} \quad (2.16)$$

Jeder symbolischen Signatur $\phi^S \in \Phi^S$ kann genau ein Klassenlabel zugeordnet werden. Die Menge $\Omega^S = \{0, \dots, M-1\}$ sei die Menge aller Klassenlabel der symbolischen Problem-domäne Φ^S .

Die Suche nach komplexen akustischen Phänomenen in A kann als Suche nach Instanzen von symbolischen Signaturen aus Φ^S in einer geeigneten Menge F^K von FLC-Annotationen verstanden werden. In F^K müssen dabei Zeit-Frequenz-Konstellationen von Annotationsfakten gefunden werden, welche den Konstellationen der symbolischen Elemente der $\phi^S \in \Phi^S$ entsprechen.

Der Menge Φ^S entspricht eine Menge $\Gamma^S = \{\gamma_0^S, \dots, \gamma_{M-1}^S\}$ von Second-Level-Annotationsalgorithmen (im Folgenden auch Second-Level-AA). Ein Second-Level-AA $\gamma_i^S \in \Gamma^S$ sucht in F^K mit Hilfe einer geeigneten Fusionsregel Instanzen des Elements $\phi_i^S \in \Phi^S$. Entspricht eine Konstellation von Fakten in F^K einem durch die Fusionsregel definierten Kriterium, so wird von dem Second-Level-AA ein Annotationsfaktum generiert.

Ein Second-Level-ASK π (im Folgenden auch SLC = **S**econd-**L**evel-**C**lassifier) kann als 2-Tupel der Form

$$\pi = \langle \phi^S, \gamma^S \rangle, \phi^S \in \Phi^S, \gamma^S \in \Gamma^S \quad (2.17)$$

aufgefasst werden. Der symbolischen Problemdomäne Φ^S korrespondiert eine Menge $P = \{\pi_0, \dots, \pi_{M-1}\}$ von SLC. Ein SLC $\pi \in P$ erhält als Input eine Menge F^K von FLC-Annotationen und generiert daraus eine abzählbare Menge $F^\pi = \{F_0^\pi, F_1^\pi, \dots\}$ von SLC-Annotationen, d.h. $\pi : F^K \rightarrow F^\pi$.

Eine *Second-Level-Annotation* ist ein 6-Tupel der Form

$$F_i^P = \langle \psi, \omega^S, \tau, f, \Lambda, \Theta \rangle \quad (2.18)$$

Hierbei ist $\psi \in \Psi$ ein Dateilabel, $\omega^S \in \Omega^S$ das Klassenlabel der zu suchenden symbolischen Signatur, τ ein Zeitpunkt in A , f ein Frequenzwert und $\Lambda = \{\lambda_0, \lambda_1, \dots\}$ eine Liste von optional zu besetzenden Variablen. Θ ist eine durch die Instanziierung modifizierte symbolischen Signatur. Die Modifikation erfolgt dabei durch den AA bzw. die Fusionsregel. Modifikationen betreffen typischerweise die Positionen von symbolischen Elementen im Zeit-Frequenz-Raum. Θ kann – je nach Anwendung – aber auch eine Menge von FLC-Annotationen sein.

Das Ziel der wissensbasierten Modellierung von SLC besteht darin, in einem gegebenen wissenschaftlichen Kontext eine symbolische Problemdomäne Φ^S zu definieren und dazu eine geeignete Menge P von SLC zu finden, mit deren Hilfe die zu Φ^S gehörenden Phänomene in F^K zuverlässig lokalisiert werden können. Mit Hilfe von P kann dann aus F^K maschinell die Menge F^P der SLC-Annotationen erstellt werden.

2.4.4 Zusammenfassung

In dem oben dargestellten Modell werden grundsätzlich zwei verschiedene Typen von ASK zur Annotation bioakustischer Daten unterschieden: First-Level und Second-Level ASK (siehe auch Abbildung 2.2).

1. First-Level-ASK (FLC) basieren auf elementaren Signaturen (Kennlinien) von Phänomenen und arbeiten direkt über Audiodaten. Mit Hilfe einer Menge K von FLC wird über einer Menge A von Audiodaten die Menge F^K von FLC-Annotationen generiert.
2. Second-Level-ASK (SLC) basieren auf symbolischen Signaturen und arbeiten über der Menge F^K . Auf Grundlage einer Menge P von SLC generiert ein Suchverfahren über F^K die Menge F^P von SLC-Annotationen.

Visualisierungen, Signaturen, KKF und Annotationsalgorithmen sind in hohem Maße anwendungsspezifisch. Es können verschiedenartige Zeit-Frequenz-Darstellungen, Energiediagramme, Oszillogramme etc. verwendet werden. Die Filter zum Erstellen von elementaren Signaturen hängen von der Natur der Phänomene ab, die charakterisiert werden sollen. KKF können als harte Klassifikatoren, über das Intervall $[0..1]$ normierte weiche Klassifikatoren oder probabilistische Klassifikatoren ausgeführt werden [29]. Der Aufbau der Annotationsalgorithmen muss dem Anwendungskontext und den Signaturen entsprechen.

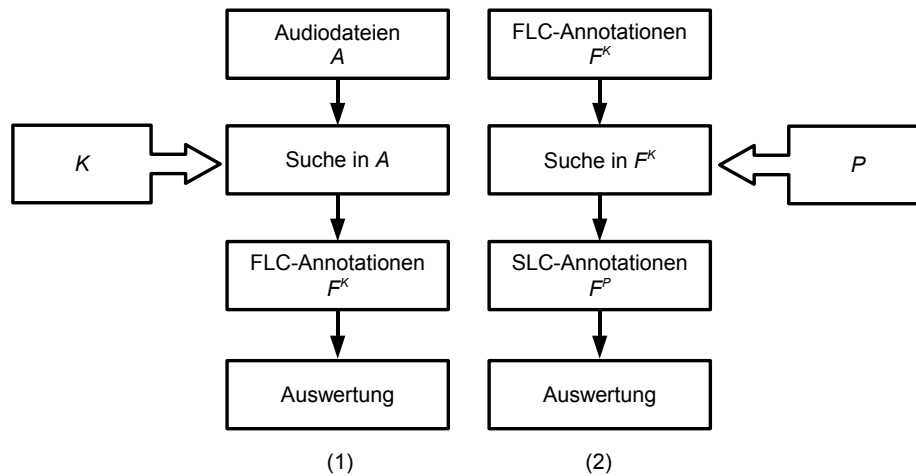


Abbildung 2.2: Annotation von Audiodaten mit FLC und SLC

Den Mengen K und P korrespondieren zwei Modellierungsprobleme. Im Folgenden wird gezeigt, wie die Modellierung beider Klassifikatortypen durch KDD-Methoden sinnvoll unterstützt werden kann.

2.5 Der FLCD-Prozess

FLC, die für eine bestimmte Problemdomäne optimiert sind, können als besondere Form von Expertenwissen verstanden werden. Die Modellierung von FLC ist – wenn man von den einfachsten Fällen absieht – ein wissensintensiver Prozess. FLC können im Rahmen eines modifizierten KDD-Prozesses (vgl. Abbildung 1.1) generiert, evaluiert und verfeinert werden. Es ergibt sich der in Abbildung 2.3 dargestellte FLC-Discovery-Prozess (FLCD-Prozess).

Das Ziel des FLCD-Prozesses besteht darin, in einem gegebenen wissenschaftlichen Kontext eine elementare Problemdomäne Φ zu definieren und dazu eine geeignete Menge K von FLC zu finden, mit deren Hilfe die zu Φ gehörenden Phänomene einer Menge A von Audiodateien zuverlässig lokalisiert und annotiert werden können.

In der folgenden Aufzählung werden die einzelnen Schritte des FLCD-Prozesses dargestellt. Jeder Schritt ist interaktiv und kann ggf. iteriert werden. Die Schritte können Entscheidungen beinhalten, die von dem jeweiligen Benutzer auf der Grundlage seines Hintergrundwissens getroffen werden müssen.

1. **Audiodaten-Basis:** Die Audiodaten müssen unter Berücksichtigung des Vorwissens und der jeweiligen wissenschaftlichen Fragestellung in geeigneter Art und Weise aufgezeichnet werden. Die Audio-Datenbasis A ist die Menge aller aufgezeichneten akustischen Prozesse.
2. **Audiodaten-Interface:** Um mit Audiodaten umgehen zu können, bedarf es

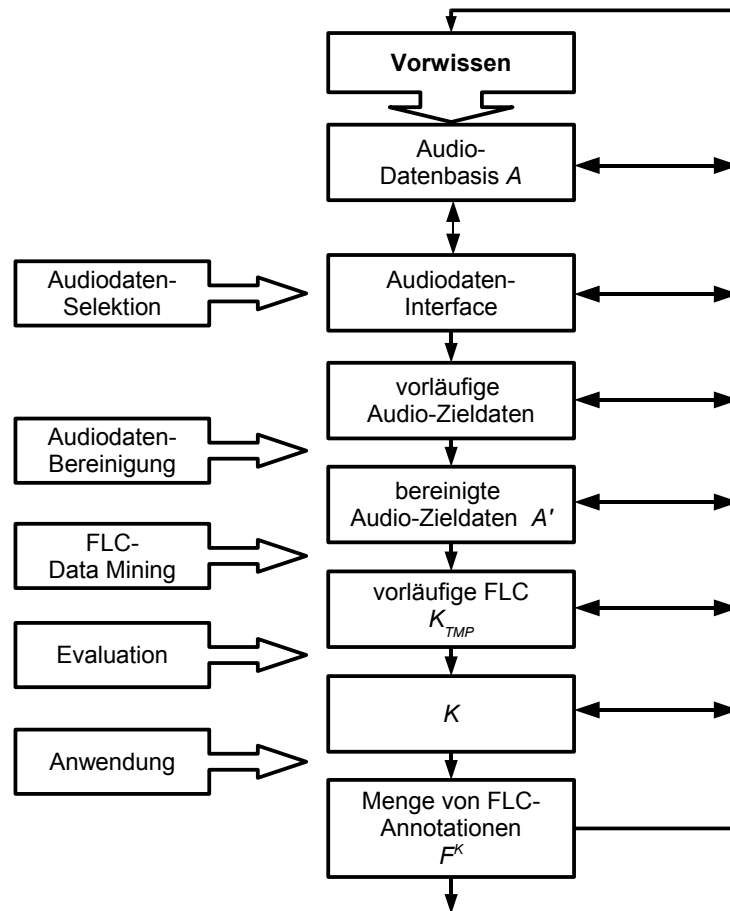


Abbildung 2.3: Der FLCD-Prozess

eines hierfür geeigneten Audiodaten-Interfaces, welches Werkzeuge zur wissenschaftlichen Visualisierung und zur Navigation in den Audiodaten zur Verfügung stellt. Es sollte auch Funktionen zur Markierung und Selektion sowie zum Kopieren, Einfügen und Löschen von Audiodaten bereitstellen.

3. **Audiodaten-Selektion:** In den meisten Fällen ist es notwendig, aus einer Menge von rohen Audiodaten eine Teilmenge auszuwählen, die für die Untersuchung am geeignetsten erscheint. In die Entscheidung darüber, welche Audiodaten in den nachfolgenden Schritten weiterverwendet werden, fließen Betrachtungen über Inhalt, Qualität, Quantität und Struktur der Daten ein. Es ergibt sich eine Menge von vorläufigen Audio-Zieldaten.
4. **Audiodaten-Bereinigung:** Die Menge der vorläufigen Audio-Zieldaten muss in der Regel noch bereinigt, d.h. von unbrauchbaren (gestörten, verzerrten, zu leisen) Aufnahmen befreit werden. Idealerweise wird die Menge fehlerhafter Aufnahmen schon während der Aufzeichnung minimiert, so dass dieser Schritt

gelegentlich übersprungen werden kann. Die verbleibenden Daten werden bereinigte Audio-Ziel Daten A' genannt.

5. **FLC-Data-Mining:** Das FLC-Data-Mining zielt darauf ab, neuartige und interessante elementare Signaturen zu finden. Diese können durch geeignete Data-Mining-Verfahren generiert oder verfeinert werden. Das FLC-Data-Mining umfasst folgende Schritte:
 - (a) **Wahl des Data-Mining-Algorithmus:** Es muss ein konkreter Algorithmus zum Finden von elementaren Signaturen für akustische Phänomene gewählt werden. Dieser Schritt beinhaltet detaillierte Betrachtungen über Parameter und mögliche Modelle sowie das Gesamtumfeld des FLC-Data-Mining-Prozesses.
 - (b) **Anwendung des Data-Mining-Algorithmus:** Die Anwendung des Data-Mining-Algorithmus zur Suche nach elementaren Signaturen kann *unsupervised* oder *supervised* ablaufen. Im Falle einer überwachten Anwendung des Data-Mining-Algorithmus müssen die Zwischenergebnisse evaluiert werden. Vom Data-Mining-Algorithmus entdeckte Signaturen können automatisch mit Klassifikationsfunktionen zu FLC verknüpft werden.
 - (c) **Evaluation:** Eine erfolgreiche Anwendung des Data-Mining-Algorithmus auf die bereinigten Ziel Daten A' resultiert in einer Menge K_{TMP} von vorläufigen, nicht-evaluierten FLC. Diese müssen bewertet werden. Im Bedarfsfalle kann der Data-Mining-Schritt wiederholt werden. Es entsteht die Menge K der evaluierten FLC.
6. **Die Menge K :** Das Produkt des FLC-Data-Mining-Prozesses ist eine Menge K von FLC. Diese Menge modelliert Wissen mit Referenz auf die Ebene elementarer akustischer Phänomene in den Audiodaten. Dieses Wissen kann zur Annotation der Ziel Daten in A' verwendet werden, es kann aber auch für sich interessante Informationen über die den Audiodaten zugrundeliegenden Prozesse enthalten.
7. **Anwendung:** Durch Anwendung der Klassifikatoren in K auf A' entsteht die Menge F^K der FLC-Annotationen. Diese Menge kann entweder direkt verwendet oder in weiterführenden KDD-Schritten auf höherer Ebene ausgewertet werden.

FLC müssen immer individuell für die im jeweiligen wissenschaftlichen Kontext zur Disposition stehenden Fragen gestaltet werden. Wenn es nicht möglich ist, eine Menge von FLC für einen bestimmten Forschungskontext zu definieren, kann die Menge F^K auch durch manuelle Annotation von A' erhalten werden. Hierdurch werden weiterführende KDD-Schritte möglich.

2.6 Der SLCD-Prozess

Auch SLC, die für eine bestimmte Problemdomäne optimiert sind, können als besondere Form von Expertenwissen verstanden werden. Die Modellierung von SLC ist – wenn man von den einfachsten Fällen absieht – ein wissensintensiver Prozess. SLC können im Rahmen eines modifizierten KDD-Prozesses generiert, evaluiert und verfeinert werden. Es ergibt sich der in Abbildung 2.4 dargestellte SLC-Discovery-Prozess (SLCD-Prozess).

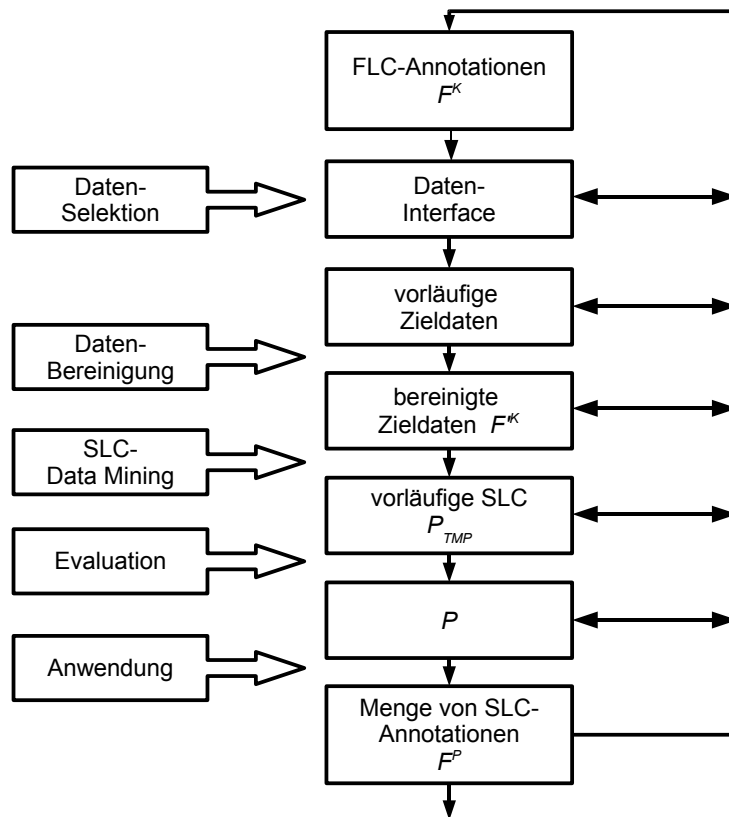


Abbildung 2.4: Der SLCD-Prozess

Das Ziel des SLCD-Prozesses besteht darin, in einem gegebenen wissenschaftlichen Kontext eine symbolische Problemdomäne Φ^S zu definieren und dazu eine geeignete Menge P von SLC zu finden, mit deren Hilfe die zu Φ^S gehörenden Phänomene in F^K zuverlässig lokalisiert und annotiert werden können.

In der folgenden Aufzählung werden die einzelnen Schritte des SLCD-Prozesses dargestellt. Jeder Schritt ist interaktiv und kann ggf. iteriert werden. Die Schritte können Entscheidungen beinhalten, die von dem jeweiligen Benutzer auf der Grundlage seines Hintergrundwissens getroffen werden müssen.

1. **FLC-Annotationen F^K :** Die Menge der FLC-Annotationen F^K entsteht

durch Anwendung der Klassifikatoren in K auf A' . F^K kann ggf. auch durch manuelle Annotationen von A' erhalten werden.

2. **Daten-Interface:** Um mit den Elementen in F^K operieren zu können, bedarf es eines speziell hierfür gestalteten Daten-Interfaces, welches Werkzeuge zur Visualisierung von Annotationsfakten und zur Navigation in Mengen von Annotationsfakten zur Verfügung stellt.
3. **Daten-Selektion:** Im SLCD-Prozess kann es notwendig sein, aus F^K eine Teilmenge auszuwählen, welche für die Untersuchung am geeignetsten erscheint. In die Entscheidung darüber, welche Daten in den nachfolgenden Schritten verwendet werden, fließen Betrachtungen über Inhalt, Qualität, Quantität und Struktur der Daten ein. Es entsteht die Menge der vorläufigen Zieldaten.
4. **Daten-Bereinigung:** Die Menge der vorläufigen Zieldaten muss ggf. bereinigt, d.h. von unbrauchbaren Elementen befreit werden. Idealerweise wird die Menge fehlerhafter Daten schon während der Annotation von A' minimiert. Die verbleibenden Daten werden bereinigte Zieldaten F'^K genannt.
5. **SLC-Data-Mining:** Das SLC-Data-Mining zielt darauf ab, neuartige und interessante symbolische Signaturen zu finden. Diese können durch geeignete Data-Mining-Verfahren generiert oder verfeinert werden. Das SLC-Data-Mining umfasst folgende Schritte:
 - (a) **Wahl des Data-Mining-Algorithmus:** Es muss ein konkreter Algorithmus zum Auffinden von neuartigen symbolischen Signaturen komplexer akustischer Phänomene gewählt werden. Dieser Schritt beinhaltet detaillierte Betrachtungen über Parameter und mögliche Modelle sowie das Gesamtumfeld des SLCD-Prozesses.
 - (b) **Anwendung des Data-Mining-Algorithmus:** Die Anwendung des Data-Mining-Algorithmus zur Suche nach symbolischen Signaturen komplexer akustischer Phänomene kann *unsupervised* oder *supervised* ablaufen. Im Falle einer überwachten Anwendung des Data-Mining-Algorithmus müssen die Zwischenergebnisse evaluiert werden. Vom Algorithmus entdeckte Signaturen werden automatisch mit Fusionsregeln zu SLC verknüpft.
 - (c) **Evaluation:** Eine erfolgreiche Anwendung des Data-Mining-Algorithmus auf die bereinigten Zieldaten F'^K resultiert in einer Menge P_{TMP} von vorläufigen, nicht-evaluierten Klassifikatoren für komplexe akustische Phänomenklassen. Diese müssen bewertet werden. Im Bedarfsfalle kann der Data-Mining-Schritt wiederholt werden, bis die Menge P der evaluierten SLC entsteht.

6. **Die Menge P :** Das Produkt des SLCD-Prozesses ist die Menge P . Diese Menge modelliert Wissen mit Referenz auf Konstellationen elementarer akustischer Phänomene im Zeit-Frequenz-Raum. Dieses Wissen kann zur Annotation der Zieldaten in F'^K verwendet werden, es kann aber auch für sich interessante Informationen über die den Audiodaten zugrundeliegenden Prozesse enthalten.
7. **Anwendung:** Durch Anwendung der Klassifikatoren in P auf F'^K entsteht die Menge F^P der SLC-Annotationen. Diese Menge kann entweder direkt verwendet oder in weiterführenden KDD-Schritten auf höherer Ebene ausgewertet werden.

2.7 Modellierungs-Aufgaben

Der oben beschriebene Grundriss gestattet, die bei der wissensbasierten Modellierung von ASK zu lösenden Aufgaben in zwei Gruppen einzuteilen:

- **KDSE-Aufgaben**, also Standard-Aufgaben, für die ein KDSE zur Modellierung von ASK vorgefertigte Lösungen bereitstellen sollte. Hierzu gehören: (1) Verfahren zur Generation von akustischen und visuellen Phänomenen; (2) Verfahren zur Extraktion von elementaren und symbolischen Signaturen; (3) Implementationen von KKF; (4) Implementationen von First-Level- und Second-Level-AA; (5) Data-Mining-Instrumente zur Suche nach neuartigen elementaren und symbolischen Signaturen; (6) Instrumente zur Visualisierung von Signaturen und Annotationen; (7) Tools zum Management aller KDD-Einzelschritte; (8) Datenbank-Tools zur Verwaltung der Mengen F^K und F^P .
- **Experten-Aufgaben**, also kontextspezifische Aufgaben, die von Experten im Rahmen des FLCD- oder SLCD-Prozesses gelöst werden müssen. Hierzu gehören: (1) Die Suche und Spezifikation von akustischen und visuellen Phänomenen; (2) Die Generation von elementaren und symbolischen Signaturen; (3) Die Auswahl von KKF und Annotationsalgorithmen; (4) Die Justierung von Parametern; (5) Die Auswahl und Anwendung von Data-Mining-Instrumenten zur Suche nach neuartigen elementaren und symbolischen Signaturen; (6) Die Ausführung und Überwachung des FLCD- und SLCD-Prozesses; (7) Die Evaluation und Anwendung von Ergebnissen des FLCD- und SLCD-Prozesses.

Im nächsten Kapitel wird für die wichtigsten der oben genannten KDSE-Aufgaben jeweils eine konkrete Lösung beschrieben. Es wird auch gezeigt, wie Experten ASK mit Hilfe dieser Instrumente modellieren können. In Teil II wird gezeigt, wie FLC und SLC für akustische Signale des Großen Tümmers mit Hilfe der KDSE-Instrumente wissensbasiert modelliert werden können.

Kapitel 3

Modellierung von Audio-Signal-Klassifikatoren

Bei der Modellierung von ASK lassen sich KDSE- und Experten-Aufgaben unterscheiden (vgl. Abschnitt 2.4.4). In diesem Kapitel werden Verfahren aus beiden Bereichen beschrieben. Die Darstellung beschränkt sich auf Algorithmen und Methoden, die von grundlegendem Interesse sind. Ziel der Ausführungen ist, für alle wichtigen Schritte des FLCD- und SLCD-Prozesses in exemplarischer Weise mindestens eine konkrete Lösung darzustellen. Die beschriebenen Verfahren können ohne großen Aufwand an konkrete Problemstellungen angepasst werden.

3.1 Vorbemerkung

Das Ziel des FLCD- und SLCD-Prozesses besteht in der Generierung von effizienten ASK. Diese sollen in empirischen Studien eingesetzt werden, die dem Aufbau oder der Modifikation von wissenschaftlichen Hypothesen dienen. Wissenschaftliches KDD zur Modellierung von ASK sollte einer Reihe von spezifischen Anforderungen genügen. Im Wesentlichen sind dies:

1. Expertenwissen, Metadaten und Beobachtungsdaten müssen in den Modellierungsprozess miteinbezogen werden können.
2. Aufzeichnungen bioakustischer Prozesse müssen ausgewertet und in den Modellierungsprozess miteinbezogen werden können.
3. Es muss möglich sein, auch sehr große Mengen von Audiodaten in akzeptabler Zeit auszuwerten.
4. ASK müssen im wissenschaftlichen Kontext sinnvoll evaluierbar sein.
5. Teile von ASK und Ergebnisse von rechenintensiven Klassifikationsprozessen sollten nach Möglichkeit wiederverwertbar sein.

Die Ausführung von KDD-Prozessen kann durch ein KDSE beträchtlich unterstützt werden (vgl. Kapitel 1). Ein Softwaresystem, das den FLCD-Prozess und den SLCD-Prozess über großen Audiodatenbanken ermöglichen soll, muss einer sehr großen Anzahl von Anforderungen genügen. Hierzu gehören vor allem funktionale Anforderungen und Qualitätsanforderungen [31]. Im Bereich der funktionalen Anforderungen ist die Entkoppelung von Programmierung und Wissensmodellierung von entscheidender Bedeutung. Im Bereich der Qualitätsanforderungen sind die Faktoren Geschwindigkeit, Stabilität und Genauigkeit essentiell.

Im Rahmen dieser Arbeit wurde ein KDSE entwickelt, das gestattet, den FLCD- und SLCD-Prozess über großen Audiodatenbanken zu realisieren. Dieses KDSE stellt eine Verbindung von logischer Programmierung und digitaler Signalverarbeitung her und wurde daher DSPROLOG genannt. Alle weiter unten dargestellten Verfahren wurden in DSPROLOG implementiert. In das System wurde auch eine Anzahl von Hilfsprogrammen integriert, die eine reibungslose Ausführung des FLCD- und SLCD-Prozesses gestatten. Eine kurze Übersicht über diese Werkzeuge gibt Tabelle A.1 im Anhang.

Im Folgenden werden nur solche Algorithmen und Verfahren dargestellt, die von grundlegendem Interesse für die Modellierung von ASK sind. Es handelt sich um Bausteine (*Building-Blocks*) für den FLCD- und SLCD-Prozess. Die Algorithmen und Verfahren können auch außerhalb des KDSE verwendet werden. Sie lassen sich ohne großen Aufwand modifizieren und erweitern.

3.2 Spektrographische Visualisierung

Spektrogramme von Audiodaten auf Grundlage der Fourier-Transformation gehören zu den wichtigsten Hilfsmitteln in der Bioakustik [55, 115]. Diese Form der Darstellung besitzt verschiedene Vor- und Nachteile: Spektrogramme sind ohne großen Aufwand zu berechnen und intuitiv verständlich. Das ursprüngliche Signal lässt sich aus einem Spektrogramm in der Regel jedoch nicht verlustfrei rekonstruieren¹.

Das *Spektrogramm* SP einer Sequenz \vec{x} ist eine $x \times y$ -Matrix über \mathbb{R} . Die Koeffizienten in den Spalten der Matrix werden erhalten, indem die gefensterter diskrete Fourier-Transformation (DFT) auf x Subsequenzen von \vec{x} angewendet wird. Die Subsequenzen können sich überlappen oder \vec{x} nur in Teilen abdecken².

Die DFT berechnet die Fourier-Koeffizienten für ein komplexes Signal \hat{x} der Länge W durch:

$$DFT(k) = \sum_{t=0}^{W-1} \hat{x}(t) \cdot \exp\left(-j2\pi \frac{kt}{W}\right) \quad (3.1)$$

¹Ein Vergleich zu anderen Zeit-Frequenz-Darstellungen (Wigner-Ville, Choi-Williams, Cone-Shaped, Gabor, u.a.) findet sich in [58].

²Eine ausführliche Darstellung findet sich in [90].

Da die Elemente in SP reellwertig sind, muss von jedem komplexen Element der Betrag berechnet werden:

$$SP_{ik} := \sqrt{DFT(k).R^2 + DFT(k).I^2} \quad (3.2)$$

Das Input-Signal \hat{x} für die DFT ergibt sich, indem die Elemente in den x Subsequenzen von \vec{x} zunächst mit einer Fensterfunktion multipliziert werden. Der imaginäre Anteil der Elemente in \hat{x} wird auf 0 gesetzt. Der reale Anteil entspricht den durch die Fensterfunktion gewichteten Elementen der Subsequenzen. Die hier immer verwendete Fensterfunktion *Blackman* (mit $\alpha = 15$, Diskussion in [57]) ist definiert durch:

$$Blackman(t) = \frac{\alpha + 1}{2} - 0.5 \cos\left(\frac{2\pi t}{N-1}\right) - \frac{\alpha}{2} \cos\left(\frac{4\pi t}{N-1}\right) \quad (3.3)$$

Eine *spektrographische Visualisierung* V sei im Folgenden die graphische Darstellung eines Spektrogramms. Auf der x-Achse ist die Zeit in ms abgetragen. Auf der y-Achse ist die Frequenz in Hz abgetragen. Der in einer Visualisierung sichtbare Bereich eines Spektrogramms kann ggf. eingeschränkt werden, beispielsweise indem ein oberes und unteres Frequenzband bestimmt wird. Die maximale Frequenz des oberen Frequenzbandes entspricht der Nyquist-Frequenz = *Abtastfrequenz* $\div 2$ der Audiodatei.

Ein besonderes Problem im Bereich der Bioakustik ist der hohe Dynamikumfang der Audiodaten. Um einen möglichst großen Wertebereich in einer spektrographischen Visualisierung sichtbar zu machen, können die Elemente der zugrundeliegenden Matrix logarithmisiert oder in spezieller Weise farbkodiert dargestellt werden. Die Farbkodierung wird erhalten, indem jedem Wert $v \in V$, ein Farbwert $color(v)$ zugewiesen wird. Die Funktion $color : \mathbb{R} \rightarrow \mathbb{N}$ wird im Folgenden auch *Farbpalette* der spektrographischen Visualisierung genannt. Für die hier verwendeten 24Bit-Bitmaps ist sie definiert durch:

$$color(v) = red(v) + 2^8 \cdot green(v) + 2^{16} \cdot blue(v)$$

Es sind insgesamt 2^{24} verschiedene Farbwerte aber nur 2^8 verschiedene Graustufen möglich. Der *Farbauszug* $red(v)$ der Palette ist definiert durch:

$$red(v) = c_{red} \cdot \begin{cases} 0 & \text{if } v \geq R_{pos} + R_{width} + 255 \\ R_{max} - v & \text{else if } v \geq R_{pos} + R_{width} \\ 255 & \text{else if } v \geq R_{pos} \\ v - R_{min} & \text{else if } v \geq R_{pos} - 255 \\ 0 & \text{sonst} \end{cases}$$

Die Funktion hängt von zwei internen Konstanten ab: R_{pos} und R_{width} . Der Wert R_{pos} ist der Wert, an dem die Farbe Rot ihren maximalen Wert 255 erreicht. Der

Wert $R_{pos} + R_{width}$ ist der Wert, ab dem die Farbe Rot in Einerschritten wieder abnimmt. Der Wert $R_{pos} - 255$ ist der erste Wert, ab dem die Farbe Rot angezeigt wird. Der Wert $R_{pos} + R_{width} + 255$ ist der letzte Wert, bis zu dem die Farbe Rot angezeigt wird. Analog werden die Farben Grün und Blau berechnet.

Durch geeignete Farbpaletten können die Werte in V mit Hilfe von Helligkeits- und Farbkontrasten in einem großen Dynamikumfang dargestellt werden. Dies hat gegenüber der traditionellen Graustufenendarstellung entscheidende Vorteile: Ein insgesamt größerer Wertebereich wird sichtbar und feine Abstufungen in benachbarten Werten treten durch Farbkontraste deutlich hervor. Die Abbildungen 3.1 und 3.2 veranschaulichen diesen Unterschied.

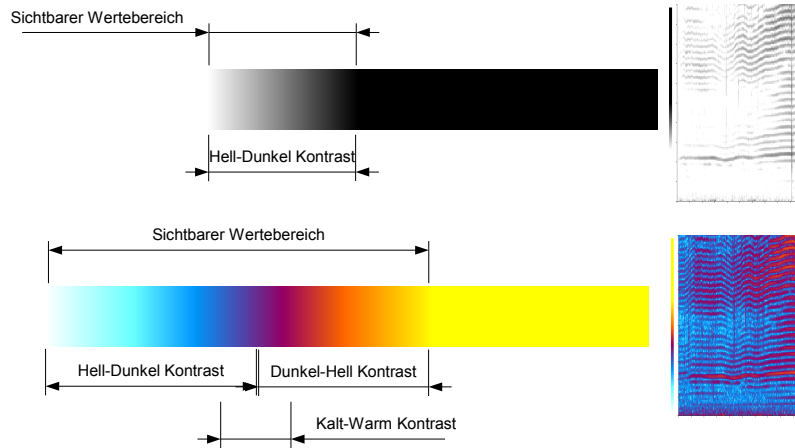


Abbildung 3.1: Graustufen- und kontrastoptimierte Farbpalette für logarithmische Darstellungen. Der sichtbare Wertebereich der Farbpalette ist ca. um den Faktor drei größer, als bei der Graustufenpalette. Der Kalt-Warm-Kontrast sorgt für eine Abgrenzung feiner Abstufungen im mittleren Wertebereich. Rechts im Bild ist ein Delphinlaut dargestellt. Nur bei der Farbkontrastdarstellung sind alle einzelnen Linien der Obertöne deutlich zu erkennen.

3.3 Modellierung spektrographischer FLC

In diesem Abschnitt wird dargestellt, wie FLC auf Grundlage von spektrographischen Visualisierungen erhalten werden können. Ein FLC $\kappa \in K$ kann als 3-Tupel verstanden werden (vgl. Abschnitt 2.4). Er besteht aus einer elementaren Signatur $\phi \in \Phi$, einer Klassen-Korrelationsfunktion (KKF) $c \in C$ und einem Annotationsalgorithmus (AA) $\gamma \in \Gamma$:

$$\kappa = \langle \phi, c, \gamma \rangle \quad (3.4)$$

Spektrographische Signaturen können entweder explizit oder induktiv programmiert werden. Explizite Programmierung bedeutet hier, dass ϕ auf der Grundlage

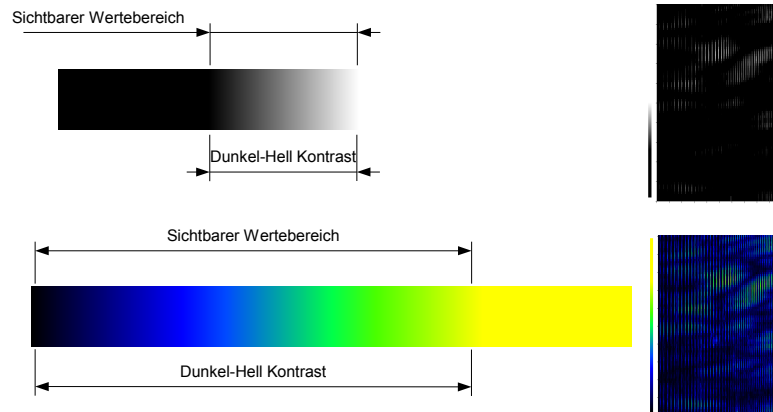


Abbildung 3.2: Graustufen- und kontrastoptimierte Farbpalette für lineare Darstellungen. Der sichtbare Wertebereich der Farbpalette ist ca. um den Faktor drei größer, als bei der Graustufenpalette. Rechts im Bild ist ein Delphinlaut dargestellt. Nur bei der Farbkontrastdarstellung sind Details des Musters zu erkennen.

explizit ausgewählter Beispiele generiert wird. Induktive Programmierung bedeutet, dass ϕ auf der Grundlage vieler Beispiele maschinell *supervised* oder *unsupervised* erlernt wird. Beide Arten der Programmierung erfolgen im Rahmen des FLC-
Prozesses.

3.3.1 Spektrographische Signaturen

Die Signatur, welche ein Phänomen charakterisiert, sollte möglichst kompakt sein und dabei alle relevanten Informationen beinhalten – andernfalls wird die Berechnung der KKF zu komplex und überflüssige Informationen könnten ihre Genauigkeit beeinträchtigen. Die Modellierung von Signaturen fällt in den Bereich der Experten-Aufgaben (vgl. Abschnitt 2.4).

Visuelle Signaturen werden über eine Funktion E^V zur Extraktion von Merkmalen aus Audiosignalen erhalten. Im Folgenden wird ein Verfahren zur Realisierung dieser Funktion auf der Basis von spektrographischen Visualisierungen (vgl. Abschnitt 3.2) beschrieben. Die Signaturen sind eine Untermenge der ursprünglichen spektrographischen Visualisierung. Das Verfahren umfasst folgende Schritte (siehe auch Abbildung 3.3):

Algorithmus: Extraktion spektrographischer Signaturen

Input: A

Output: ϕ

1. Generiere die Visualisierung V eines gewünschten Phänomens in A mit Hilfe der Methode aus Abschnitt 3.2. Hierbei ist zu beachten, dass alle Charakteris-

tika des Phänomens in V deutlich hervortreten müssen. V ist eine $x \times y$ -Matrix über \mathbb{R} .

2. Wähle in V einen Bereich V' aus, welcher der Position und Ausdehnung des Phänomens im Zeit-Frequenz-Raum entspricht. V' ist eine $x' \times y'$ -Matrix über \mathbb{R} , $0 \leq x' < x$, $0 \leq y' < y$, die eine Untermenge von V enthält.
3. Extrahiere aus V' die spektrographische Signatur ϕ des Phänomens. Gehe dabei wie folgt vor:

- (a) Wähle in V' alle Elemente aus, die größer sind als ein vorgegebener Schwellwert ς . Teile diese Menge in zwei Matrizen: (1) Die Matrix ϕ^+ , welche alle Elemente umfasst, die offensichtlich eine besondere Relevanz für Klassifikationsentscheidungen besitzen, und (2) die Matrix ϕ^0 , die alle restlichen Elemente umfasst, d.h. alle Elemente, von denen noch nicht bekannt ist, ob sie eine Relevanz für Klassifikationsentscheidungen besitzen.

In ϕ^+ sind nur lokale Maxima von Spalten und Zeilen in V' enthalten:

$$\phi^+ = \{v_{x,y} \in V' \mid v_{x,y} > \varsigma, f_M(v_{x,y}) = 1\} \quad (3.5)$$

Die Matrix ϕ^0 wird beim Verschmelzen von Signaturen benötigt. Sie enthält Datenpunkte, die durch Verschmelzungsoperationen möglicherweise zu lokalen Maxima von Spalten und Zeilen transformiert werden können (siehe Abschnitt 3.3.9):

$$\phi^0 = \{v_{x,y} \in V' \mid v_{x,y} > \varsigma, f_M(v_{x,y}) = 0\} \quad (3.6)$$

Die Funktion $f_M : V' \rightarrow \{1, 0\}$ gibt an, ob ein Element $v_{x,y}$ ein lokales Maximum in einer Spalte oder Zeile von V' ist. Dieses Vorgehen geht auf die Annahme zurück, dass die meisten *Real-World*-Zeitreihen durch ihre maximalen Fourier-Koeffizienten repräsentiert werden können [53, 1]. Die Funktion f_M ist definiert durch

$$f_M(v_{x,y}) = \begin{cases} 1 & \text{if } ((v_{x,y} > v_{x+1,y}) \wedge (v_{x,y} > v_{x-1,y})) \\ 1 & \text{if } ((v_{x,y} > v_{x,y+1}) \wedge (v_{x,y} > v_{x,y-1})) \\ 0 & \text{sonst} \end{cases} \quad (3.7)$$

- (b) Wähle anschließend in V' alle Elemente aus, die kleiner sind, als ς und sich zugleich in unmittelbarer Nachbarschaft zu Elementen aus ϕ^+ oder ϕ^0 befinden. Die Nachbarschaft ist durch je eine Abstandsvariable dx in horizontaler und dy vertikaler Richtung definiert. Diese Elemente bilden den *Umriss* ϕ^- der Visualisierung eines Phänomens, da sich hierin bei geeigneten Abstandsvariablen der Umriss des Phänomens widerspiegelt.

$$\phi^- = \{v_{x,y} \in V' \mid v < \varsigma, f_U(v_{x,y}) = 1\} \quad (3.8)$$

Hierbei ist $f_U : V' \rightarrow \{1, 0\}$ eine Funktion, die angibt, ob sich ein Datenpunkt $v_{x,y}$ in Nachbarschaft zu einem Punkt aus ϕ^+ oder ϕ^0 befindet. Die Funktion f_U ist definiert durch

$$f_U(v_{x,y}) = \begin{cases} 1 & \text{if } v_{x+dx,y} > \varsigma \\ 1 & \text{if } v_{x-dx,y} > \varsigma \\ 1 & \text{if } v_{x,y+dy} > \varsigma \\ 1 & \text{if } v_{x,y-dy} > \varsigma \\ 0 & \text{sonst} \end{cases} \quad (3.9)$$

- (c) Lösche aus ϕ^+, ϕ^0 , und ϕ^- manuell alle Elemente, die offensichtlich nicht zu dem Phänomen gehören, das modelliert werden soll. Es ergibt sich die spektrographische Signatur des Phänomens:

$$\phi = \langle \phi^+, \phi^0, \phi^- \rangle \quad (3.10)$$

4. Evaluiere die Eigenschaften der Signatur im gegebenen KDD-Kontext. Wiederhole ggf. die vorausgehenden Schritte.

Es ist zu sehen, dass dieses Verfahren – wie die meisten KDD-Schritte – ggf. iteriert werden muss. Eine schrittweise Verfeinerung gestattet es, Signaturen zu erhalten, die einer Fragestellung angemessen sind.

3.3.2 KKF für spektrographische Signaturen

Mit Hilfe der KKF vergleicht der Annotationsalgorithmus eine Signatur mit einer Audiosequenz. Durch die KKF wird der Hypothese, dass in der Sequenz das der Signatur entsprechende Phänomen enthalten ist, eine reellwertige Maßzahl s – die Klassenähnlichkeit – zugeordnet.

Da eine spektrographische Signatur als Untermenge eines Spektrogramms verstanden werden kann, eignet sich zur Berechnung der Klassenähnlichkeit ein Maß für den Abstand zwischen Signatur und Spektrogramm der Sequenz. Das im Folgenden beschriebene Verfahren zur Berechnung der KKF ist schwellwertbasiert. Es erfordert, dass die zeitliche Ausdehnung von Sequenz und Signatur einander entsprechen. Der Algorithmus zeichnet sich durch große Stabilität und einfaches Parameter-Fitting aus:

Algorithmus: KKF spektrographisch

Input: \vec{x}, ϕ

Output: s

1. Bilde aus \vec{x} ein Spektrogramm unter Verwendung *aller* Parameter, die zur Generierung des der Signatur zugrundeliegenden Spektrogramms verwendet wurden. Das Ergebnis ist ein Spektrogramm $SP^{\vec{x}}$, dessen Elemente sich direkt mit denen von ϕ vergleichen lassen.

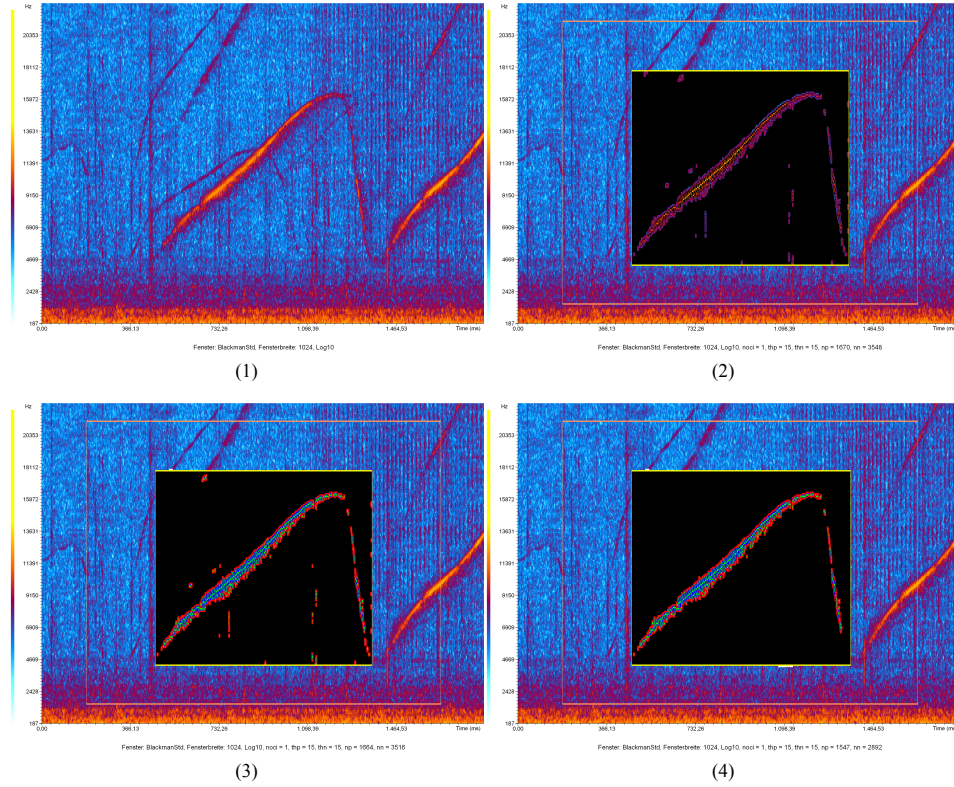


Abbildung 3.3: Generieren einer spektrographischen Signatur: (1) Visualisierung eines Pfifflautes von *Tursiops truncatus*; (2) Auswahlfenster V' (inneres Rechteck); (3) Signatur des Pfifflautes in Rohform. (4) Signatur des Pfifflautes nach manueller Bereinigung: Rot entspricht ϕ^- , blau ϕ^0 und grün ϕ^+ .

2. Extrahiere aus $SP^{\vec{x}}$ zwei Matrizen, deren Elemente den definierten Elementen in ϕ^+ und ϕ^- entsprechen:

$$X^+ = \{v_{x,y} \in SP^{\vec{x}} | Def(\phi_{x,y}^+) = True\} \quad (3.11)$$

$$X^- = \{v_{x,y} \in SP^{\vec{x}} | Def(\phi_{x,y}^-) = True\} \quad (3.12)$$

Es gilt nun $N = |\phi^+| = |X^+|$ und $M = |\phi^-| = |X^-|$.

3. Berechne die Ähnlichkeit zwischen ϕ und \vec{x} durch:

$$s(X^+, X^-, \phi) = dy_0 + c_0 \sqrt{d^+(X^+, \phi^+) \cdot d^-(X^-, \phi^-)} \quad (3.13)$$

Hierbei sind d^+ und d^- zwei unterschiedliche Funktionen zur Berechnung des Abstands:

$$d^+(X^+, \phi^+) = dy_1 + \frac{1}{N} \sum_{i=0}^{x-1} \sum_{j=0}^{y-1} c_1 \Delta^+(X_{ij}, \phi_{ij}^+) \quad (3.14)$$

$$d^-(X^-, \phi^-) = dy_2 + \frac{1}{M} \sum_{i=0}^{x-1} \sum_{j=0}^{y-1} c_2 \Delta^-(X_{ij}, \phi_{ij}^-) \quad (3.15)$$

Die Parameter $dy_0, c_0, dy_1, c_1, dy_2$ und c_2 sind Normierungsparameter. Die Abstandsfunktionen Δ^+ und Δ^- sind definiert durch:

$$\Delta^+(a, b) = \begin{cases} a - b & \text{if } a > b \\ 0 & \text{sonst} \end{cases} \quad (3.16)$$

$$\Delta^-(a, b) = \begin{cases} b - a & \text{if } a < b \\ 0 & \text{sonst} \end{cases} \quad (3.17)$$

Idee bei diesem Vorgehen ist, die Differenz zwischen einem Element in X^+ und einem Element in ϕ^+ nur dann in das Ergebnis mit einzubeziehen, wenn das erste Element kleiner ist, d.h. nicht der Vorgabe entspricht. Wenn es dagegen die Vorgabe übererfüllt, wird der resultierende negative Abstand nicht mit einbezogen. Dasselbe gilt für die Differenz zwischen einem Element in X^- und einem Element in ϕ^- : Nur wenn die Vorgabe nicht erfüllt ist, wird die Differenz zur Berechnung des Ergebnisses hinzugezogen. Mit dieser Art der Gewichtung wird dem hohen Dynamikumfang in bioakustischen Signalen besser Rechnung getragen als mit einfacheren Maßen, wie z.B. der Manhattan-Distanz.

3.3.3 Algorithmische Annotation

Das dritte Element eines FLC ist der Annotationsalgorithmus, der die KKF systematisch auf die Dateien des Audiokorpus anwendet und unter bestimmten Bedingungen FLC-Annotationen erzeugt. Im Folgenden wird ein Verfahren vorgeschlagen, das sich besonders gut für die oben beschriebene KKF eignet. Der AA berechnet anhand der KKF für eine Audiodatei \vec{s} zunächst einen geglätteten Korrelationsvektor \vec{D} . Aus diesem werden anschließend Intervalle isoliert, deren Elemente alle über einem bestimmten Schwellwert th liegen. Zu jedem Intervall wird dann zu dem Zeitpunkt eine FLC-Annotation generiert, an dem \vec{D} innerhalb des Intervalls sein Maximum hat. Die Annotation erfolgt *unsupervised* nach folgendem Verfahren:

Algorithmus: FLC-Annotation

Input: \vec{s}, th, ϕ

Output: F^K

1. Bilde aus einer Audiodatei \vec{s} mit Hilfe der signaturspezifischen spektrographischen KKF einen Korrelationsvektor \vec{D} , der für jeden möglichen Zeitpunkt eine Aussage über die Klassenähnlichkeit zwischen Signatur und Audiodatei beinhaltet. Glätte anschließend den Vektor.

2. Finde in dem geglätteten Korrelationsvektor \vec{D} die Intervalle I , deren Elemente alle größer sind als der vorgegebene Ähnlichkeits-Schwellwert th .
3. Generiere für jedes Intervall genau eine FLC-Annotation. Der Zeitpunkt der Annotation entspricht dabei dem Zeitpunkt, an dem der geglättete Korrelationsvektor innerhalb des Intervalls sein Maximum hat. Die Menge aller so generierten FLC-Annotationen ist F^K .

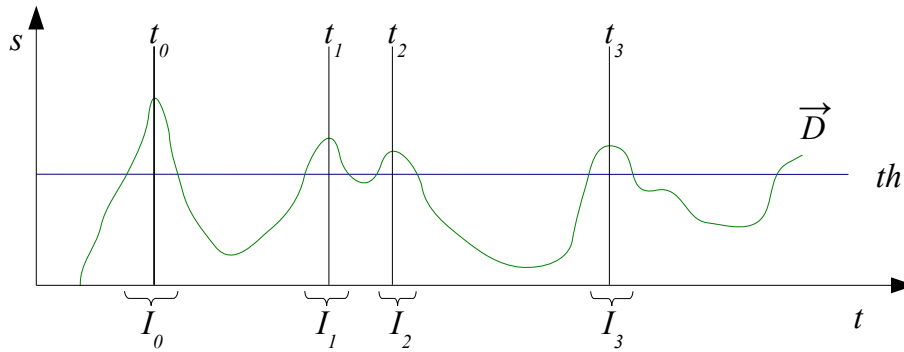


Abbildung 3.4: Arbeitsweise des Annotationsalgorithmus: \vec{D} ist der geglättete Korrelationsvektor. th ist der Ähnlichkeitsschwellwert. I_0, I_1, I_2, I_3 sind die Intervalle, an denen die Elemente von \vec{D} größer sind als th . t_0, t_1, t_2, t_3 sind die Zeitpunkte zu denen der Algorithmus eine Annotation generiert.

In Abbildung 3.4 ist die Arbeitsweise des Annotationsalgorithmus schematisch dargestellt. Idee bei diesem Verfahren ist, Annotationen immer nur dann zu generieren, wenn der Abstand zwischen Signatur und Audio-Datei oberhalb eines Schwellwertes liegt und gleichzeitig ein lokales Maximum ist. Die Glättung dient der Vermeidung von falschen Annotationsentscheidungen durch lokale Fluktuationen in \vec{D} .

3.3.4 Manuelle Programmierung von FLC

In vielen Fällen ist es ausreichend FLC (und auch SLC) manuell zu erstellen. Eine intuitiv einleuchtende Methode zur expliziten Programmierung von FLC besteht darin, Beispiele für Phänomen-Klassen mit Hilfe eines entsprechenden Audio-Daten-Interfaces auszuwählen und nach der oben beschriebenen Methode spektrographische Signaturen zu generieren. Anschließend können diese mit der KKF und dem Annotationsalgorithmus verknüpft werden. Alle notwendigen Parameter können von Hand justiert werden. Die so erhaltenen Klassifikatoren können schrittweise evaluiert und verfeinert werden. Es ist auch möglich, sie so zu gruppieren, dass eine Phänomenklasse durch mehrere Klassifikatoren abgedeckt wird. Der Prozess der expliziten Modellierung der Menge K entspricht dem FLCD-Prozess (vgl. Abschnitt 2.5), beinhaltet aber keinen Data-Mining-Schritt.

3.3.5 Notation und Visualisierung

Die Performanz eines FLC ist von seiner Signatur selbst, aber auch von einer Reihe algorithmischer Parameter abhängig. Unmittelbares visuelles Feedback über die folgenden Eigenschaften eines Klassifikators ist bei seiner Gestaltung und bei der Justierung der Parameter hilfreich (siehe auch Abbildung 3.5):

- Die Zeit-Frequenz Position der Datenpunkte in ϕ .
- Die Aufspaltung des Klassifikators in ϕ^+ , ϕ^0 und ϕ^- .

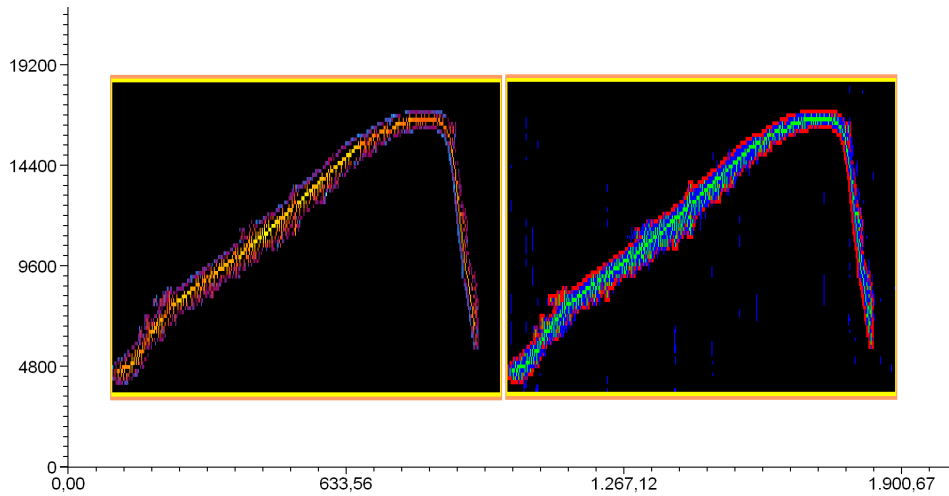


Abbildung 3.5: Visualisierung einer Signatur. Links: ϕ^+ und ϕ^- in Farben der Palette aus Abschnitt 3.2. Rechts: ϕ^+ (Grün), ϕ^0 (Blau) und ϕ^- (Rot) in Kontrastfarben.

In Prädikatenlogik kann eine FLC-Annotation als atomare Formel der Form $\omega(\psi, \tau, f, \Lambda)$ notiert werden. Diese Darstellungsform ist kompakt, lässt sich aber auf intuitivem Wege nur schwer mit dem annotierten Ereignis in ψ verbinden. Auch das Verhalten der KKF und ihrer Konstituenten kann im FLCD-Kontext von Interesse sein. Eine verständlichere Form der Darstellung von FLC-Annotationen und KKF zeigt Abbildung 3.6.

3.3.6 Ausdünnen von Merkmalen

Oft sind Signaturen trotz der in Abschnitt 3.3.1 beschriebenen Beschränkung auf die maximalen Fourier-Koeffizienten zu umfangreich, d.h. sie enthalten zu viele Datenpunkte. Dieser Umstand beeinflusst zwar nicht unbedingt die Genauigkeit der Klassifikatoren, wohl aber den Rechenaufwand und somit die Geschwindigkeit der maschinellen Auswertung von Audiodaten.

Eine Möglichkeit, dieses Problem zu lösen, besteht darin, die Datenmenge in einer Signatur durch einen weiteren Filter zu reduzieren. Eine einfache aber effiziente

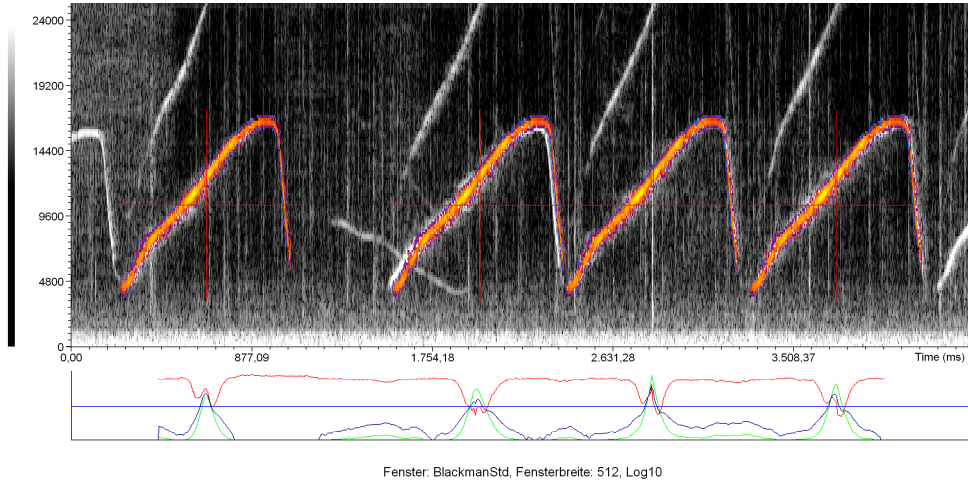


Abbildung 3.6: Visuelles Feedback für die Signatur aus Abb. 3.5. Oben: Annotationsentscheidungen werden durch Einblendung in Kontrastfarben und Fadenkreuz dargestellt. Unten: Graph der Korrelationsfunktion s (Blau) und ihrer Konstituenten d^+ (Grün) und d^- (Rot). Die waagerechte blaue Linie entspricht th .

Möglichkeit zur Ausdünnung von Merkmalen besteht darin, nur jedes n -te Element in horizontaler bzw. jedes m -te Element in vertikaler Richtung nach einer zweiten Filterstufe übrig zu lassen:

$$\phi_2^+ := \{v_{x,y} \in \phi^+ | (x \bmod n = 0), (y \bmod m = 0)\} \quad (3.18)$$

$$\phi_2^- := \{v_{x,y} \in \phi^- | (x \bmod n = 0), (y \bmod m = 0)\} \quad (3.19)$$

Dieses Vorgehen muss nicht sofort zu einer Herabsetzung der Genauigkeit eines FLC führen. Um das Verhalten einer auf diese Art manipulierten Signatur einschätzen zu können, empfiehlt sich eine Überwachung des Graphs der KKF. In Abbildung 3.7 ist ein Beispiel dargestellt.

3.3.7 Klonen von Signaturen

Oftmals ist es erforderlich, mehrere FLC so zu gruppieren, dass durch sie ein großer Frequenzbereich abgedeckt wird. Um diese Aufgabe zu erleichtern, ist es möglich, aus einer einzelnen Signatur eine Schar von Klonen zu erzeugen, die einen vorgegebenen Frequenzbereich abdeckt. Die Klone unterscheiden sich dabei von der ursprünglichen Signatur lediglich durch den Frequenzbereich, der ihnen zugeordnet ist.

Abbildung 3.8 veranschaulicht den Prozess des Klonens von FLC. In dem Beispiel wird für die Signatur eines Haar-Wavelets eine Schar von Klonen erzeugt. Zunächst wird aus dem Spektrogramm des Wavelets eine FLC-Signatur gebildet. Anschließend wird aus dieser eine Schar weiterer FLC erzeugt, die einen großen Frequenzbereich abdeckt.

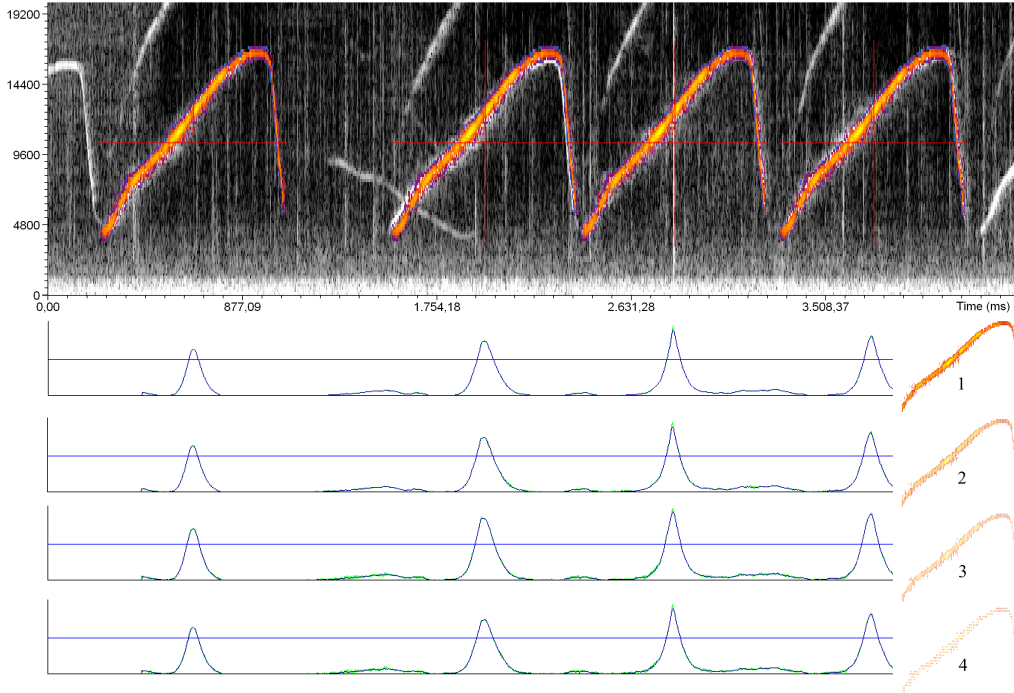


Abbildung 3.7: Ausdünnen von Merkmalen: Signatur 1 = 2600 Punkte; Signatur 2 = 1266 Punkte; Signatur 3 = 663 Punkte; Signatur 4 = 269 Punkte; Der Graph von s ist für alle vier Signaturen nahezu identisch.

Abbildung 3.9 zeigt, wie mittels einer solchen Klon-Schar eine Sequenz von Pfifflauten von *Tursiops truncatus* annotiert werden kann. Die einzelnen Klone liegen dabei so dicht beieinander, dass sich die Kontur der Pfiffe in ihren Annotationsentscheidungen widerspiegelt.

3.3.8 Automatische Extraktion spektrographischer Signaturen

Spektrographische Signaturen können aus einer Menge F^K von Annotationsfakten automatisch extrahiert werden. Voraussetzung hierfür ist, dass die Audiodateien, über denen die Menge F^K gebildet wurde, zur Verfügung stehen. Im Folgenden wird am Beispiel eines einzelnen Annotationsfaktums $F_i^\kappa \in F^K$ ein Verfahren zur automatischen Extraktion von spektrographischen Signaturen dargestellt:

Algorithmus: Extraktion spektrographischer Signaturen

Input: F_i^κ

Output: ϕ_i^κ

1. Gegeben sei die Annotationsentscheidung $F_i^\kappa \in F^K$ eines elementaren Klassi-

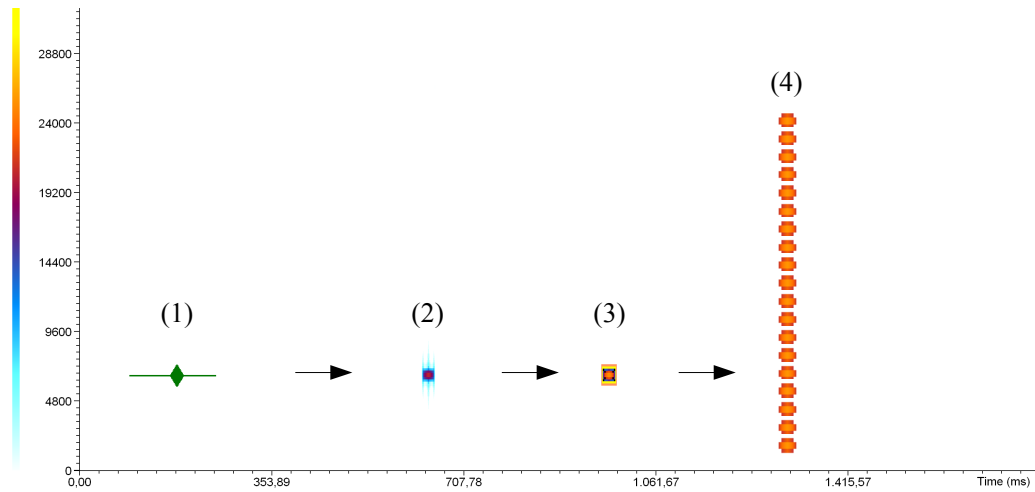


Abbildung 3.8: Klonen von Signaturen: (1) Oszillogramm eines Wavelets; (2) Spektrogramm des Wavelets; (3) Signatur des Wavelets; (4) Klon der Signatur

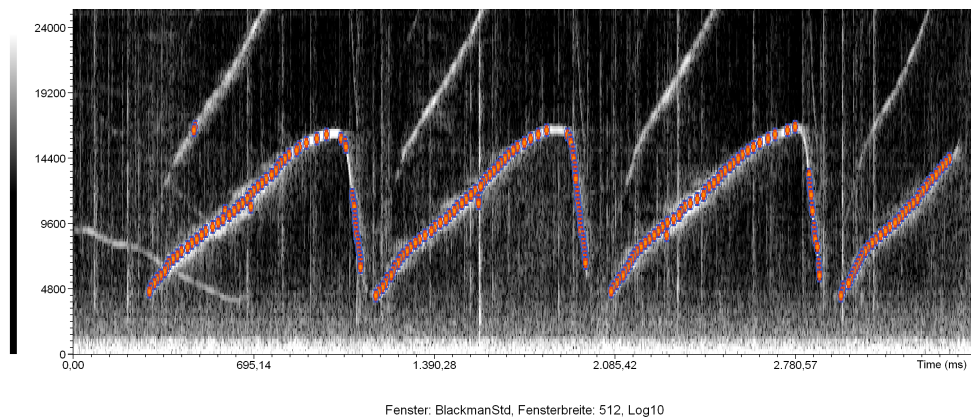


Abbildung 3.9: Annotationsentscheidungen der geklonten FLC aus Abbildung 3.8

fiktors $\kappa = \langle \phi, c, \gamma \rangle$. F_i^κ referiert auf ein Ereignis in einer Audiodatei $\vec{s} \in A$. Der Ausdehnung von ϕ im Zeit-Frequenz-Raum entspricht ein Rechteck $R = (t_1, f_2, t_2, f_1)$.

Hierbei sind t_1 und t_2 Zeitpunkte relativ zu einem nicht näher zu spezifizierenden Nullpunkt. f_1 und f_2 sind Werte auf der Frequenzskala in Hertz. Die zeitliche Ausdehnung von ϕ ist $\Delta t = t_2 - t_1$. Die Ausdehnung von ϕ im Frequenzbereich ist $\Delta f = f_2 - f_1$. t_1 und f_2 bilden die linke obere Ecke des Rechtecks, t_2 und f_1 bilden die rechte untere Ecke des Rechtecks³.

2. Zu der Annotationsentscheidung F_i^κ soll automatisch eine Signatur ϕ_i^κ extrahiert werden, wobei folgendes gelte:
 - (a) ϕ_i^κ soll auf einer spektrographischen Visualisierung beruhen, welche *nicht* notwendigerweise dieselben Parameter benutzt, die der Signatur ϕ von κ zugrunde liegen (vgl. Abschnitt 3.2).
 - (b) Die Extraktion der Signatur ϕ_i^κ soll nach dem selben Verfahren wie die Extraktion von ϕ vollzogen werden. Zur Extraktion der Signatur ϕ_i^κ sollen dieselben Parameter verwendet werden, die zur Extraktion von ϕ verwendet wurden (vgl. Abschnitt 3.3.1).
3. Definiere ausgehend von R einen Suchraum R_s für ϕ , so dass $R_s := (t_1^s, f_2^s, t_2^s, f_1^s)$. Hierbei kann, aber muss nicht $t_1^s \leq t_1, f_2^s \geq f_2, t_2^s \geq t_2, f_1^s \leq f_1$ gelten. Die zeitliche Ausdehnung von R_s ist $\Delta t^s = t_2^s - t_1^s$. Die Ausdehnung von R_s im Frequenzbereich ist $\Delta f^s = f_2^s - f_1^s$ (vgl. Abbildung 3.10).
4. Der Annotation F_i^κ entspricht ein Zeitpunkt t_i^κ in \vec{s} . Wähle ausgehend von diesem und den Zeitkoordinaten des Suchraums t_1^s, t_2^s aus \vec{s} eine Sequenz

$$\vec{s}_{t_1':t_2'} \text{ mit } t_1' = t_i^\kappa - (t_1 - t_1^s), t_2' = t_i^\kappa + \Delta t + (t_2^s - t_2) \quad (3.20)$$

5. Generiere nach dem Verfahren in 3.2 aus $\vec{s}_{t_1':t_2'}$ ein Spektrogramm SP . Setze hierfür ggf. andere Parameter als bei der Generierung von ϕ fest.
6. Generiere nach dem Verfahren in 3.3.1 aus SP eine spektrographische Signatur ϕ_i^κ . Verwende hierzu dieselben Parameter, wie zur Generierung von ϕ . Das Auswahlfenster, das den für die Extraktion der Signatur relevanten Bereich V' definiert, wird durch die Parameter f_1^s und f_2^s automatisch begrenzt.

3.3.9 Verschmelzen von spektrographischen Signaturen

Eine weitere elementare Technik für die Modellierung von FLC ist das Verschmelzen mehrerer Signaturen zu einer einzigen. Ziel des Verschmelzens ist, aus einer Anzahl

³Eine in der Computergrafik übliche Konvention.

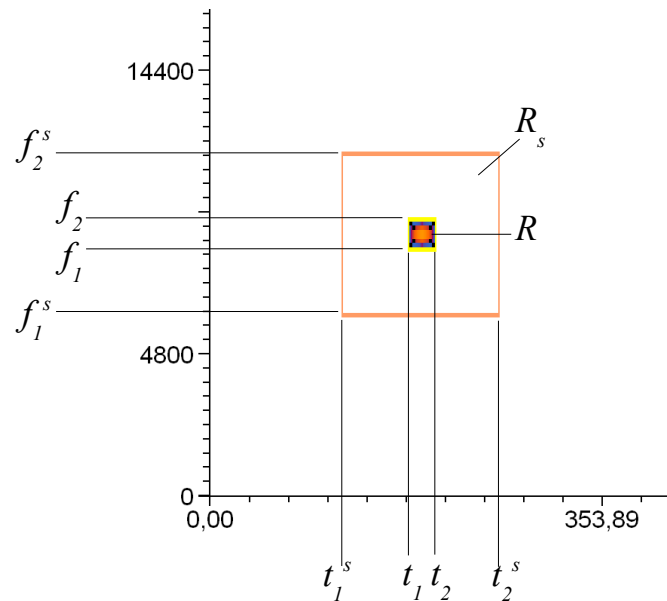


Abbildung 3.10: FLC mit Suchraum R_s

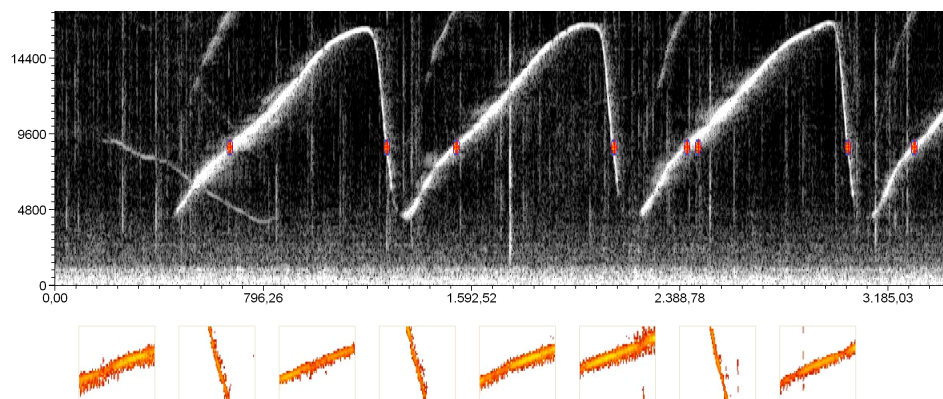


Abbildung 3.11: Oben: Annotationen des FLC aus Abb. 3.10. Unten: Aus den Annotationen automatisch extrahierte Signaturen

von Signaturen diejenigen Eigenschaften herauszufiltern, welche in der Mehrzahl der Signaturen vorhanden sind und gleichzeitig diejenigen Eigenschaften herauszumitteln, welche nur in einzelnen oder wenigen Beispielen vertreten sind.

Für das Verschmelzen einer Menge $\Phi = \{\phi_0, \dots, \phi_{N-1}\}$ von spektrographischen Signaturen zu einer einzigen verschmolzenen Signatur ϕ^M eignet sich das im Folgenden beschriebene Verfahren. Bedingung dabei ist, dass alle Signaturen auf Spektrogrammen mit identischen Parametern beruhen:

Algorithmus: Verschmelzen spektrographischer Signaturen

Input: Φ

Output: ϕ^M

1. Prüfe, ob die Signaturen in Φ auf Spektrogrammen mit identischen Parametern beruhen. Sortiere ggf. Signaturen aus, die nicht mit dem Rest verschmelzbar sind.
2. Generiere für jedes Element $\phi_i \in \Phi, 0 \leq i < N$ die Vereinigung

$$V_i := \phi_i^+ \cup \phi_i^0 \cup \phi_i^-, 0 \leq i < N, \quad (3.21)$$

so dass eine Menge $\mathbf{V} = \{V_0, \dots, V_{N-1}\}$ von $x \times y$ -Matrizen entsteht. Bilde nun eine $x \times y$ -Matrix M (Merged) mit den Mittelwerten der Koeffizienten der Matrizen in \mathbf{V} durch:

$$M := \frac{1}{N} \sum_{i=0}^{N-1} V_i \quad (3.22)$$

3. Extrahiere aus M nach dem Verfahren in Abschnitt 3.3.1 eine spektrographische Signatur der Form $\phi^M = \langle \phi^+, \phi^0, \phi^- \rangle$.

Abbildung 3.12 veranschaulicht das Verschmelzen von \mathbf{V} zu M . Das Verfahren eignet sich, um in einer Menge von Signaturen Noise und andere Störungen herauszumitteln. Es kann auch dazu verwendet werden, um herauszufinden, welche Merkmale häufig in einer Menge von Signaturen vertreten sind.

Durch das Verschmelzen kann aus einer Menge von verrauschten Signaturen eines akustischen Ereignistyps eine einzige Signatur generiert werden, die den Ereignistyp besser modelliert, als die einzelnen Elemente der Menge. Abbildung 3.13 zeigt dies am Beispiel von drei verschiedenen verschmolzenen Signaturen eines Pfifflauts von *Tursiops truncatus*.

3.3.10 Ähnlichkeitsbasiertes Verschmelzen spektrographischer Signaturen

Von besonderer Bedeutung für den Data-Mining-Schritt im FLCD-Prozess ist die Technik des ähnlichkeitsbasierten Verschmelzens von elementaren Signaturen.

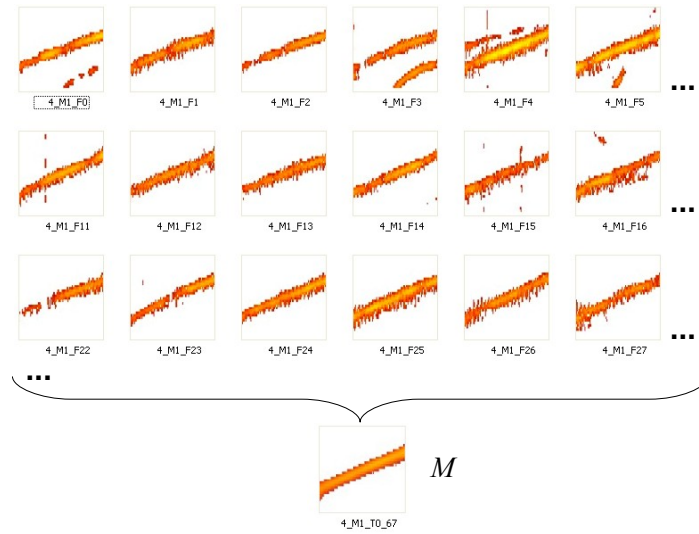


Abbildung 3.12: Verschmelzen von Signaturen: Aus 67 Signaturen (aus Platzgründen nicht alle abgebildet) wird eine einzige durch Verschmelzung generiert. Störungen werden dabei herausgemittelt.

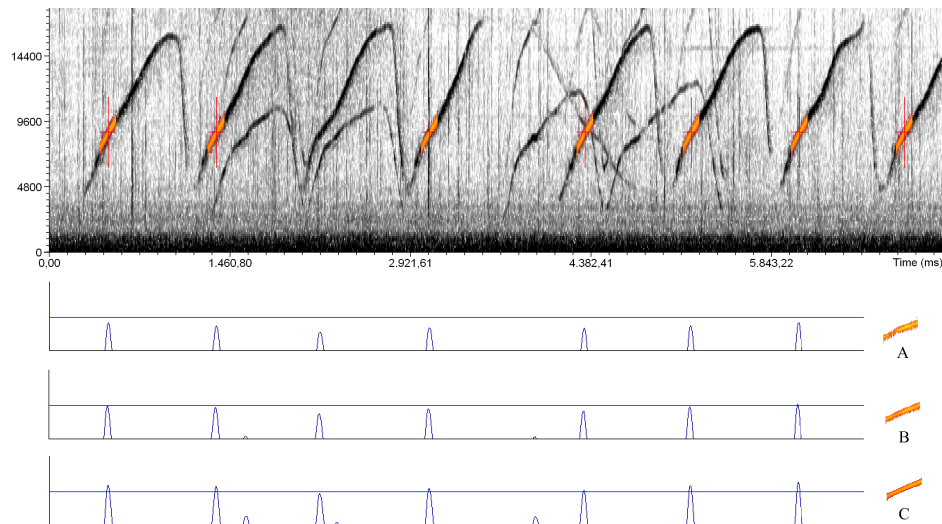


Abbildung 3.13: Graph von s für drei verschiedene Signaturen: A wurde aus einem Beispiel generiert. B wurde aus vier Signaturen durch Verschmelzung generiert. C wurde aus 67 Signaturen durch Verschmelzung generiert. Nur C trifft in dem Beispiel richtige Annotationsentscheidungen.

Für das ähnlichkeitsbasierte Verschmelzen einer Menge $\Phi = \{\phi_0, \dots, \phi_{N-1}\}$ von Signaturen zu einer Menge $\Phi^M = \{\phi_0^M, \dots, \phi_{N'-1}^M\}$ von verschmolzenen Signaturen eignet sich das im Folgenden beschriebene Verfahren. Der Verschmelzungsprozess ist dabei von einem Ähnlichkeitsparameter σ abhängig. Dieser bestimmt den für eine Verschmelzungsoperation minimal erforderlichen Abstand von zwei Signaturen.

Algorithmus: Similarity-Driven-Merge

Input: Φ, σ

Output: Φ^M

1. Prüfe, ob die Signaturen in Φ auf Spektrogrammen mit identischen Parametern beruhen. Sortiere ggf. Signaturen aus, die nicht mit dem Rest verschmelzbar sind.
2. Generiere für jedes Element $\phi_i \in \Phi, 0 \leq i < N$ die Vereinigung

$$V_i := \phi_i^+ \cup \phi_i^0 \cup \phi_i^-, 0 \leq i < N, \quad (3.23)$$

so dass eine Menge $\mathbf{V} = \{V_0, \dots, V_{N-1}\}$ von $x \times y$ -Matrizen entsteht. Verschmelze dann nach der Prozedur *SDMerge* \mathbf{V} zur einer Menge $\mathbf{M} = \{M_0, \dots, M_{N'-1}\}$ von verschmolzenen $x \times y$ -Matrizen.

Procedure: *SDMerge*

Input: \mathbf{V}, σ **Output:** \mathbf{M}

begin

$N := |\mathbf{V}|;$

$\mathbf{M} := \emptyset;$

Append(V_0, \mathbf{M});

for $i = 1$ to $N - 1$ do begin

$n := \text{Nearest}(V_i, \mathbf{M}, \sigma);$

if $n \neq -1$ then *Merge*(V_i, M_n) else *Append*(V_i, \mathbf{M});

end;

return \mathbf{M} ;

end.

Function: *Nearest*

Input: V_i, \mathbf{M}, σ **Output:** n

begin

$N := |\mathbf{M}|;$

$n := -1;$

$\sigma^{Best} := \sigma;$

for $j := 0$ to $N - 1$ do begin

$\sigma^{Current} := \text{Distance}(V_i, M_j);$

if $\sigma^{Current} \leq \sigma^{Best}$ then begin

```

     $\sigma^{Best} := \sigma^{Current};$ 
     $n := j;$ 
end;
end;
return  $n$ ;
end.

```

Die Funktion *Distance* entspricht dem Abstandsmaß in Abschnitt 3.3.2. Diese Funktion kann in Abhängigkeit von dem jeweiligen KDD-Kontext variiert werden. Einfachere Maße sind die Summe der absoluten Distanzen der Koeffizienten

$$Distance(V, M) = \sum_{i=0}^{x-1} \sum_{j=0}^{y-1} |V_{ij} - M_{ij}| \quad (3.24)$$

oder der positive Anteil der KKF:

$$Distance(V, M) = dy + \frac{1}{x \cdot y} \sum_{i=0}^{x-1} \sum_{j=0}^{y-1} c \cdot \Delta^+(V_{ij}, M_{ij}) \quad (3.25)$$

wobei $\Delta^+(V_{ij}, M_{ij})$ in Gleichung 3.16 beschrieben ist.

Die Funktion *Merge* entspricht dem Verfahren zur Verschmelzung von elementaren Signaturen in Abschnitt 3.3.9. Da hier aber die Anzahl der Verschmelzungen pro Signatur *a priori* nicht bekannt ist, muss für jedes Element in **M** die Anzahl *nom* (**number of merged**) der bisher durchgeführten Verschmelzungen protokolliert werden. Hierdurch wird es möglich, Koeffizienten in *V* und *M* vor der Mittelwertbildung adäquat zu gewichten:

$$M_{ij} := \frac{nom \cdot M_{ij} + V_{ij}}{nom + 1} \quad (3.26)$$

3. Extrahiere aus **M** nach dem Verfahren in Abschnitt 3.3.1 eine Menge $\Phi^{\mathbf{M}}$ von spektrographischen Signaturen der Form

$$\Phi^{\mathbf{M}} = \{\phi_0^{\mathbf{M}}, \dots, \phi_{N'-1}^{\mathbf{M}} | \phi_i^{\mathbf{M}} = \langle \phi_i^+, \phi_i^0, \phi_i^- \rangle\}, 0 \leq i < N'$$

Abbildung 3.14 veranschaulicht das ähnlichkeitsbasierte Verschmelzen von **V** zu **M**. Bei den Beispielen handelt es sich um die automatisch extrahierten Signaturen aus Abbildung 3.11.

3.3.11 Induktive Programmierung

In diesem Abschnitt wird ein Verfahren zum ähnlichkeitsbasierten Erlernen einer endlichen Menge von FLC beschrieben. Das Verfahren beruht auf den Methoden

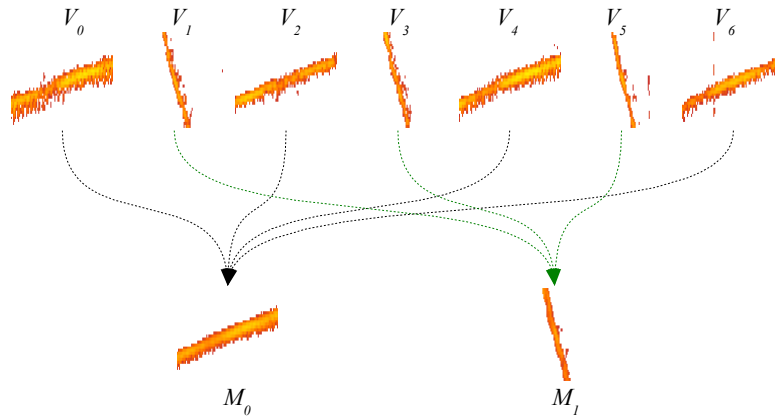


Abbildung 3.14: Ähnlichkeitsbasiertes Verschmelzen von Signaturen: Die Menge $\mathbf{V} = \{V_0, \dots, V_6\}$ wird zur Menge $\mathbf{M} = \{M_0, M_1\}$ verschmolzen.

zur automatischen Extraktion (vgl. Abschnitt 3.3.8) und ähnlichkeitsbasierten Verschmelzung (vgl. Abschnitt 3.3.10) von spektrographischen Signaturen. Es entspricht im FLC-CD-Prozess dem Data-Mining-Schritt (vgl. Abschnitt 2.5).

Die Idee ist, mit Hilfe eines übergeneralisierenden Klassenidentifikators in A eine Menge von spektrographischen Phänomenen zu identifizieren, automatisch Signaturen für diese Phänomene zu erzeugen, diese dann ähnlichkeitsbasiert zu verschmelzen und mit einer KKF und einem Annotationsalgorithmus zu FLC zu verknüpfen.

Mit dem Verfahren wird erreicht, dass spektrographische Phänomene in A (und neuen Daten) nach einem definierten Ähnlichkeitskriterium durch eine Menge von FLC zu Clustern gruppiert werden können. Für eine erfolgreichen Durchführung dieses Verfahrens müssen folgende Bedingungen erfüllt sein:

1. Die zu gruppierenden akustischen Phänomene müssen *alle* ein eindeutiges gemeinsames Merkmal besitzen, das mit Hilfe einer spektrographischen Signatur ϕ repräsentiert werden kann. Ist dies der Fall, so ist durch ϕ implizit die Klasse Ω^ϕ der ϕ -repräsentierbaren Phänomene in A definiert.
2. Für das Klassenmerkmal ϕ muss ein elementarer Klassifikator $\kappa = \langle \phi, c, \gamma \rangle$ existieren, mit dessen Hilfe Instanzen der Elemente der Klasse Ω^ϕ in A zuverlässig lokalisiert werden können. Ist dies der Fall, so ist durch κ implizit die Klasse Ω^κ der ϕ -identifizierbaren Phänomene in A definiert.

Sind diese Voraussetzungen erfüllt, kann durch die folgenden eine Menge K von FLC erlernt werden, welche die Menge Ω^ϕ ähnlichkeitsbasiert zu Clustern diversifiziert, wobei jedem Element in K genau ein Cluster entspricht. Mit Hilfe der FLC in K ist es möglich akustische Phänomene, die den jeweiligen Clustern entsprechen, in A maschinell zu lokalisieren:

Algorithmus: FLC-Induktion

Input: ϕ, A

Output: K

1. Prüfe, ob die oben beschriebenen Voraussetzungen für die zu gruppierenden akustischen Phänomene erfüllt sind.
2. Generiere für das gemeinsame Klassenmerkmal ϕ einen elementaren Klassifikator $\kappa = \langle \phi, c, \gamma \rangle$ mit dessen Hilfe Instanzen der Elemente der Klasse Ω^ϕ in A zuverlässig lokalisiert werden können.
3. Erzeuge mit Hilfe von κ eine Menge F^κ von elementaren Annotationsfakten über A .
4. Lege für κ einen Suchraum R_s fest. R_s muss so dimensioniert sein, dass die zu gruppierenden Phänomene aus Ω^ϕ durch R_s vollständig abgedeckt werden (vgl. Abschnitt 3.3.8).
5. Erzeuge mit Hilfe des in Abschnitt 3.3.8 beschriebenen Verfahrens aus F^κ und A eine Menge Φ^E von spektrographischen Signaturen der Phänomene in Ω^ϕ .
6. Bilde mit Hilfe des in Abschnitt 3.3.10 beschriebenen Verfahrens aus der Menge Φ^E die Menge $\Phi^{\mathbf{M}} = \{\phi_0^{\mathbf{M}}, \dots, \phi_{N-1}^{\mathbf{M}}\}$ der ähnlichkeitsbasiert verschmolzenen Signaturen.
7. Bilde durch Verknüpfung mit der KKF $c \in \kappa$ und dem Annotationsalgorithmus $\gamma \in \kappa$ aus $\Phi^{\mathbf{M}}$ eine Menge K_{TMP} von FLC der Form

$$K_{TMP} = \{\kappa_0, \dots, \kappa_{N-1} | \kappa_i = \langle \phi_i^{\mathbf{M}}, c, \gamma \rangle\}, 0 \leq i < N$$

8. Evaluiere die FLC in K_{TMP} und iteriere ggf. die vorausgehenden Schritte.

Alle FLC der Menge K_{TMP} müssen im Rahmen des FLCD-Prozesses (vgl. Abbildung 2.3) evaluiert und verfeinert werden. Die Verknüpfung von Signaturen mit KKF und Annotationsalgorithmen erfolgt in Schritt 7 automatisch. Parameter der KKF und des Annotationsalgorithmus müssen ggf. von Hand justiert werden.

Beispiel: Abbildung 3.15 veranschaulicht die einzelnen Schritte dieses Verfahrens. In dem Beispiel erscheint eine Sequenz von Pfifflauten. Die Laute bestehen aus einem Ton, der langsam ansteigt und dann relativ schnell abfällt. Nicht in allen Fällen existiert der zweite Teil des Lautes – so wird der letzte Pfiff laut nach dem Anstieg abrupt abgebrochen. Ziel ist hier, für solche Laute zwei verschiedene FLC zu generieren: Der erste soll dem ansteigenden Anteil der Pfiff laute entsprechen, der zweite dem abfallenden. Da Pfiff laute eine sinusartige Schwingung zugrunde liegt, können sie mit einem FLC κ für Kurzzeit-Sinusschwingungen (20ms) in A identifiziert werden. Annotationen eines solchen FLC sagen jedoch nichts darüber aus, ob

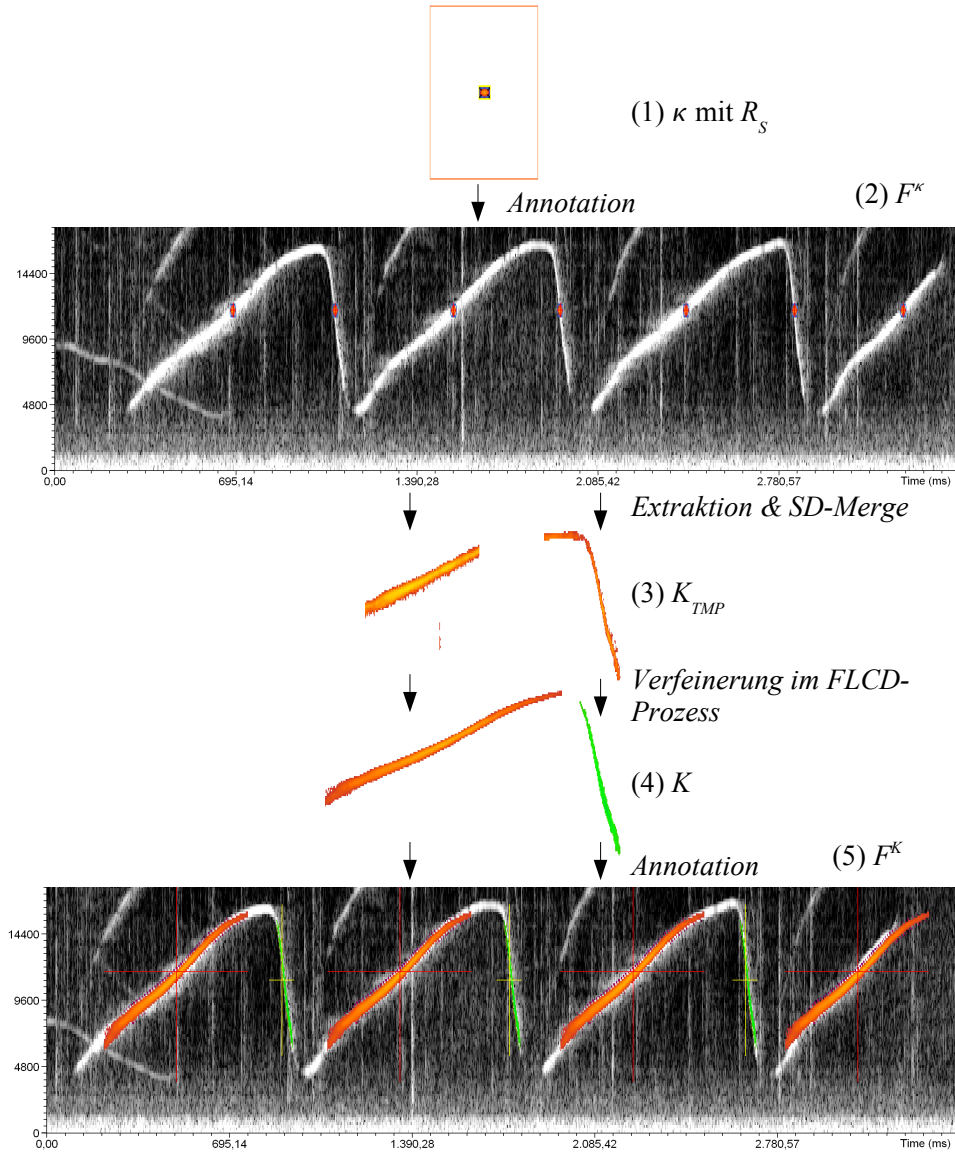


Abbildung 3.15: Induktive Programmierung von FLC: (1) κ mit Suchraum R_s . (2) Annotationen von κ in einer Sequenz von Piffflauten. (3) Verschmolzene Signaturen von K_{TMP} ; (4) Verfeinerte Signaturen von K . Die zweite Signatur ist zur besseren Unterscheidung mit einer anderen Farbpalette *grün* dargestellt; (5) Annotationsentscheidungen von K .

der identifizierte Laut dem ansteigenden oder abfallenden Lautanteil entspricht. Für κ wird daher ein Suchraum R_s festgelegt, der so dimensioniert ist, dass die Pfiff laut-Phänomene ungefähr von ihm abgedeckt werden. Aus der Menge der Annotationen von κ wird unter Beachtung von R_s eine Menge von Signaturen extrahiert und mit Hilfe des Verfahrens in Abschnitt 3.3.10 ähnlichkeitsbasiert verschmolzen. Durch Verknüpfung mit c und γ aus κ ergibt sich eine Menge K_{TMP} von nicht-evaluierten FLC. Diesen entsprechen auf- bzw. absteigende Lautanteile. Sie können im Rahmen des FLCD-Prozesses verbessert werden. Durch schrittweise Verfeinerung entsteht die Menge K der evaluierten FLC.

3.3.12 Diskussion

Bei den in diesem Abschnitt beschriebenen Verfahren handelt es sich um elementare Techniken zur wissensbasierten Modellierung von spektrographischen FLC. Diese können ohne größere Probleme so modifiziert werden, dass sie z.B. auch für die Auswertung von anderen Formen der wissenschaftlichen Visualisierung verwendbar sind.

Es existiert eine Reihe weiterer Techniken, die sich in diesem Bereich als nützlich erwiesen haben. Hierzu gehört die automatisierte Bereinigung der Menge F^K , das Spiegeln von Signaturen und eine Vielzahl von Operationen zur Manipulation und Visualisierung von Datenbanken mit Annotationsfakten. Da diese zwar von großer praktischer Bedeutung, aber für das theoretische Verständnis der folgenden Kapitel nicht erforderlich sind, wird an dieser Stelle auf eine Darstellung verzichtet.

Im Rahmen des FLCD-Prozesses können mit Hilfe der oben beschriebenen Methoden FLC modelliert werden, die auf spezifische Fragestellungen im empirischen Kontext abgestimmt sind. Die algorithmischen Anteile der Verfahren können von einem KDSE zu Verfügung gestellt werden.

3.4 Modellierung von SLC

In diesem Abschnitt wird dargestellt, wie SLC auf Grundlage von spektrographischen FLC erhalten werden können. Ein SLC P^S kann als 2-Tupel verstanden werden:

$$P^S = \langle \phi^S, \gamma^S \rangle \quad (3.27)$$

Hierbei ist ϕ^S eine symbolische Signatur und γ^S ein Second-Level-Annotationsalgorithmus (vgl. Abschnitt 2.4.3).

Die Aufgabe einer symbolischen Signatur besteht darin, ein komplexes akustisches Phänomen mit Hilfe einer Konstellation von symbolischen Elementen im Zeit-Frequenz-Raum zu beschreiben. Dies ist dann möglich, wenn das komplexe Phänomen in eine Menge von elementaren Phänomenen dekomponiert werden kann und für jedes der konstituierenden Einzelphänomene ein FLC existiert, der Instanzen des Phänomens in Audiodaten lokalisieren kann.

Symbolische Signaturen können durch manuelle Modellierung, durch automatische Interpretation einer Sequenz von Audiodaten oder durch Induktion über einer Menge von Annotationsfakten erhalten werden:

1. **Manuelle Modellierung:** Die Signatur ϕ^S wird durch manuelles Positionieren von symbolischen Elementen in einem virtuellen Ausschnitt des Zeit-Frequenz-Raumes generiert. Alle notwendigen Parameter werden manuell festgesetzt.
2. **Modellierung durch Interpretation:** Die Signatur ϕ^S wird erhalten, indem eine Signalsequenz \vec{x} zunächst mit Hilfe einer Menge K von FLC annotiert wird. Danach werden die entstandenen Annotationsfakten F^K automatisch zu einer symbolischen Signatur ϕ^S assembliert. Alle notwendigen Parameter werden maschinell festgesetzt.
3. **Induktive Programmierung:** Die Signatur ϕ^S wird *supervised* oder *unsupervised* durch ein maschinelles symbolisches Induktionsverfahren aus einer Menge von Annotationsfakten generiert. Alle notwendigen Parameter werden maschinell festgesetzt.

In den folgenden Abschnitten werden mehrere Ansätze und Verfahren zur interpretativen und induktiven Programmierung von SLC dargestellt. Manuelle Programmieretechniken werden an dieser Stelle nicht explizit beschrieben.

In den Ausführungen wird davon ausgegangen, dass die den Signaturen zugrundeliegenden symbolischen Elemente (vgl. Abschnitt 2.4.3) die Position von elementaren Annotationsfakten im Zeit-Frequenz-Raum widerspiegeln. Auf welchem Wege die Annotationsfakten erhalten wurden, ist dabei von untergeordneter Bedeutung. Es ist durchaus denkbar, dass sie rein manuell erstellt wurden. Auch ist denkbar, dass Frequenzparameter vollständig ignoriert werden (z.B. bei der Modellierung von rhythmischen Strukturen in bioakustischen Signalsequenzen).

3.4.1 Modellierung symbolischer Signaturen durch Interpretation

Die Modellierung einer symbolischen Signatur kann durch Interpretation einer Signalsequenz \vec{x} mit Hilfe einer Menge K von FLC erfolgen. Ziel ist, für ein komplexes Phänomen in \vec{x} eine symbolische Signatur ϕ^S zu finden. ϕ^S sollte nur die zur Charakterisierung des Phänomens relevanten Informationen beinhalten.

Es müssen dabei folgende Voraussetzungen erfüllt sein: Das komplexe Phänomen kann in eine Anzahl von Elementarereignissen dekomponiert werden. Jedem Elementarereignis kann genau ein Punkt im Zeit-Frequenz-Raum zugeordnet werden. Für jedes Elementarereignis existiert ein FLC mit dessen Hilfe das Ereignis in \vec{x} lokalisiert werden kann. Die Menge dieser FLC sei K .

Um aus der Sequenz von Audiodaten \vec{x} eine symbolische Signatur ϕ^S zu erhalten, die das komplexe Phänomen beschreibt, eignet sich folgendes Verfahren (siehe auch Abbildung 3.16):

Algorithmus: Modellierung durch Interpretation

Input: \vec{x}, K

Output: Φ^S

1. Generiere aus \vec{x} die Visualisierung V des zu beschreibenden Phänomens – z.B. mit Hilfe der Methode aus Abschnitt 3.2. Hierbei ist zu beachten, dass sich alle Charakteristika des Phänomens innerhalb des von V dargestellten Ausschnitts befinden müssen. Die Qualität der visuellen Darstellung spielt hier eine untergeordnete Rolle, da die Koeffizienten von V nicht direkt zur Bildung der Signatur verwendet werden.
2. Wähle entsprechend der Position und Ausdehnung des Phänomens in V einen Bereich V' aus. Setze einen Zeitpunkt τ_1 für den Beginn des komplexen Phänomens in \vec{x} fest.
3. Interpretiere das komplexe Phänomen in V' durch Annotation mit Hilfe der Elemente in K . Es entsteht eine Menge F^K von FLC-Annotationen der Form

$$F^K = \{F_0, \dots, F_{N-1} | F_i = \langle \psi, \omega_i, \tau_i, f_i, \Lambda_i \rangle\}, 0 \leq i < N, N = |F^K|. \quad (3.28)$$

4. Transformiere alle Elemente in F^K in symbolische Elemente, so dass eine symbolische Signatur ϕ'^S der Form

$$\phi'^S = \{\phi_0^s, \dots, \phi_{N-1}^s | \phi_i^s = \langle \omega_i, (\tau_i - \tau_1), f_i, \Lambda_i \rangle\}, 0 \leq i < N \quad (3.29)$$

entsteht. Die Signatur ϕ'^S ist nicht bereinigt, d.h. sie enthält u.U. noch symbolische Elemente, die auf Noise o.ä. in \vec{x} zurückzuführen sind.

5. Lösche aus ϕ'^S ggf. alle Elemente, die nicht zu dem Phänomen gehören, das modelliert werden soll. Es ergibt sich die *bereinigte* symbolische Signatur des komplexen Phänomens:

$$\phi^S = \{\phi_0^s, \dots, \phi_{M-1}^s\}, M \leq N \quad (3.30)$$

6. Evaluiere die Signatur im gegebenen KDD-Kontext. Wiederhole ggf. die vorhergehenden Schritte.

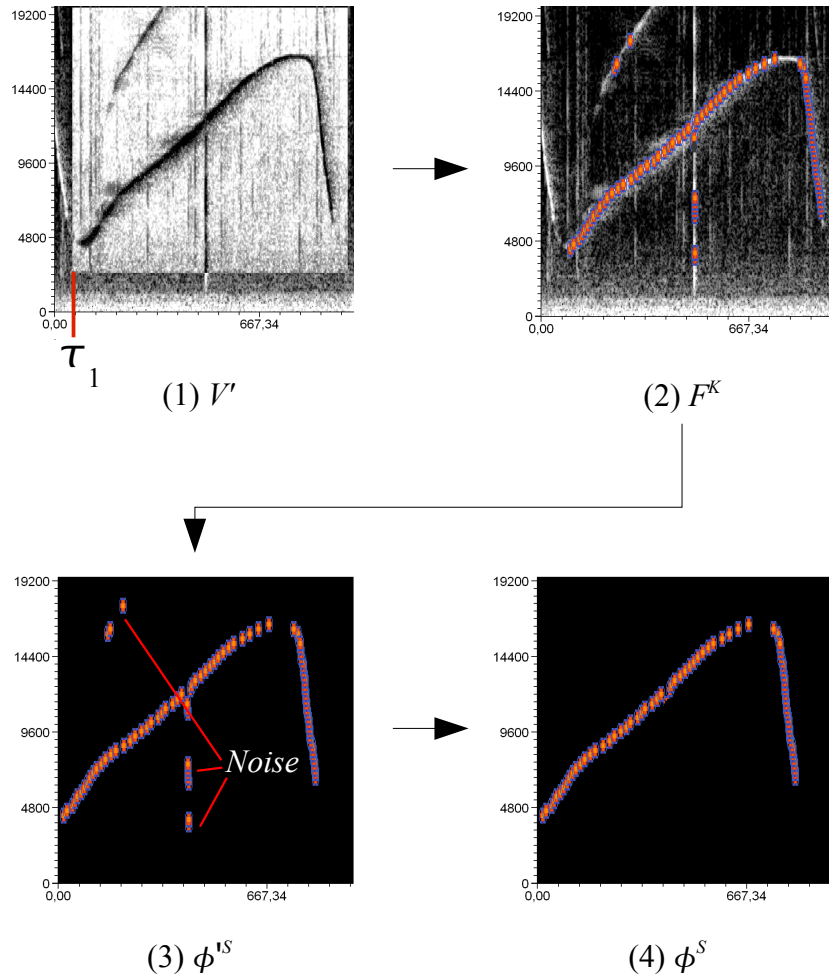


Abbildung 3.16: Modellierung durch Interpretation: (1) Der Ausschnitt V' (invertiert) in V mit einem komplexen Phänomen (hier ein Pfiff laut); (2) Visualisierung der Menge F^K generiert von einer Schar Haar-Wavelet basierter FLC; (3) Nicht bereinigte Signatur ϕ'^S – verschiedene symbolische Elemente gehen auf *Noise* im Bereich des Pfiff lautes zurück; (4) Bereinigte symbolische Signatur ϕ^S des Pfiff lautes

3.4.2 Algorithmische Annotation

Das zweite Element eines SLC ist der Annotationsalgorithmus γ^S , der in einer Menge F^K von elementaren Annotationsfakten nach Instanzen von ϕ^S sucht und unter bestimmten Bedingungen SLC-Annotationen generiert.

Das im Folgenden beschriebene Annotationsverfahren arbeitet hierarchiebasiert. Für jede Anfrage wird eine Hierarchie festgelegt, die den Suchvorgang steuert. Hierdurch kann unter günstigen Umständen eine wesentliche Reduktion des Rechenaufwandes erreicht werden. Das Verfahren generiert SLC-Annotationen *unsupervised*. Die Instanziierung von symbolischen Signaturen erfolgt in dreierlei Hinsicht fehler-tolerant:

1. **Zeitverschiebungen von Elementen:** Symbolische Elemente $\phi^s \in \phi^S$ werden unter Tolerierung einer Abweichung auf der Zeitachse instanziiert. Die Toleranzgrenze ist durch einen Parameter t^{tol} festgesetzt.
2. **Frequenzverschiebungen von Elementen:** Symbolische Elemente $\phi^s \in \phi^S$ werden unter Tolerierung einer Abweichung auf der Frequenzachse instanziiert. Die Toleranzgrenze ist durch einen Parameter f^{tol} festgesetzt.
3. **Nicht instanziiierbare Elemente:** ϕ^S wird auch dann instanziiert, wenn eine Anzahl von konstituierenden symbolischen Elementen $\phi^s \in \phi^S$ trotz der Toleranz im Zeit- und Frequenzbereich nicht instanziiert werden kann. Die Toleranzgrenze für nicht instanziiierbare Elemente ist durch einen Parameter i^{tol} festgelegt.

Aus Gründen der Übersichtlichkeit wird hier davon ausgegangen, dass F^K über genau einer Audiodatei mit dem Label ψ generiert wurde. Das Verfahren arbeitet wie folgt:

Algorithmus: SLC-Annotation

Input: $\phi^S, F^K, t^{tol}, f^{tol}, i^{tol}$

Hierbei ist

$$\phi^S = \{\phi_0^s, \dots, \phi_{N-1}^s | \phi_n^s = \langle \omega_n, \tau_n, f_n, \Lambda_n \rangle\}, 0 \leq n < N \quad (3.31)$$

und

$$F^K = \{F_0, \dots, F_{M-1} | F_m = \langle \psi, \omega_m, \tau_m, f_m, \Lambda_m \rangle\}, 0 \leq m < M \quad (3.32)$$

Output: F^P

Hierbei ist

$$F^P = \{F_0^P, \dots, F_{I-1}^P | F_i^P = \langle \psi, \omega_i^S, \tau_i, f_i, \Lambda_i, \Theta_i \rangle\}, 0 \leq i < I \quad (3.33)$$

wobei $\Theta_i \subseteq F^K$ eine Konstellation von FLC-Annotationen ist, die der Konstellation der symbolischen Elemente in ϕ^S entspricht.

Der Zeitpunkt τ_i sei der Zeitpunkt des zeitlich ersten Elementes in Θ_i . Für alle Θ_i aller $F_i^P \in F^P$ gelte

$$\Theta_i = \{F_0, \dots, F_{N-1} | F_n = \langle \psi, \omega_n^F, \tau_n^F, f_n^F, \Lambda_n^F \rangle\}, 0 \leq n < N \quad (3.34)$$

wobei die Bedingung

$$\begin{aligned} (\omega_n^F = \omega_n) \wedge (\tau_n^F \in [(\tau_i + \tau_n - t^{tol}) \dots (\tau_i + \tau_n + t^{tol})]) \wedge \\ (f_m^F \in [(f_n - f^{tol}) \dots (f_n + f^{tol})]) \end{aligned} \quad (3.35)$$

erfüllt ist.

Die Menge F^P ist also die Menge aller SLC-Annotationen, die alle zulässigen Konstellationen von FLC-Annotationen in der Menge F^K beschreibt. Zulässig sind dabei genau die Konstellationen, welche unter Beachtung der festgesetzten Toleranzgrenzen die Konstellation der symbolischen Elemente von ϕ^S im Zeit-Frequenz-Raum widerspiegeln. Der Algorithmus umfasst folgende Schritte:

1. Definiere für die Konstituenten der Signatur ϕ^S eine Hierarchie.

$$\Xi = \{1, \dots, H\}, H \in \mathbb{N} \quad (3.36)$$

sei die Menge der Label der hierarchischen Ebenen. Die oberste Hierarchieebene sei 1. Es gelte

$$\forall \phi^s \in \phi^S : \phi^s \rightarrow h, h \in \Xi \quad (3.37)$$

2. Gruppiere die Elemente in ϕ^S entsprechend ihrer Zugehörigkeit zu einer Hierarchieebene in H Teilsignaturen, so dass

$$\tilde{\phi}^S = \{\phi_1^S, \dots, \phi_H^S\} \quad (3.38)$$

3. Bilde über F^K aus der Teilsignatur ϕ_1^S entsprechend den Gleichungen 3.33 und 3.34 mit der Bedingung 3.35 die Menge der SLC-Annotationen F_C^P (C von *Current*). Generiere eine SLC-Annotation auch dann, wenn zu einem bestimmten Zeitpunkt für maximal i^{tol} Elemente aus ϕ_1^S keine Entsprechung in F^K gefunden wird. Speichere für jedes Element in F_C^P die Anzahl der nicht instanziierten Konstituenten mit Hilfe einer Variable y .

4. Setze $h := 2$ und $\phi_C^S := \phi_1^S$

5. Vereinige die Teilsignatur ϕ_C^S mit ϕ_h^S , d.h. $\phi_C^S := \phi_C^S \cup \phi_h^S$

6. Prüfe für jedes Element in F_C^P , ob es zu einer SLC-Annotation der Teilsignatur ϕ_C^S erweitert werden kann, welche die Gleichungen 3.33 und 3.34 mit der Bedingung 3.35 erfüllt. Erweitere eine SLC-Annotation auch dann, wenn für maximal $(i^{tol} - y)$ Elemente aus ϕ_C^S keine Entsprechung in F^K gefunden wird, wobei y die Anzahl der bereits nicht-instanzierten Konstituenten der Annotation ist. Speichere für jedes Element in F_C^P die aktuelle Gesamtanzahl der nicht instanziierten Elemente in y .
7. Lösche alle nicht erweiterten Elemente aus F_C^P
8. Erhöhe h um 1. Falls $h \leq H$ gehe zu Schritt 5
9. $F^P := F_C^P$;

Das beschriebene Verfahren zeichnet sich durch generelle Vor- und Nachteile hierarchiebasierter Suchverfahren aus. Einerseits kann die Suche in F^K wesentlich beschleunigt werden, andererseits können im Worst-Case gültige Konstellationen von Elementen in F^K übersehen werden. Dies ist z.B. dann der Fall, wenn alle Elemente einer Ebene zu einem bestimmten Zeitpunkt nicht instanziiert werden können, obwohl zu demselben Zeitpunkt eine äquivalente Signatur mit nur einer Ebene instanziiert wäre.

3.4.3 Notation und Visualisierung

In prädikatenlogischer Notation kann eine symbolische Signatur ϕ^S der Länge N als Liste der Form

$$\phi^S = [[\omega_0, \tau_0, f_0, [\lambda_0^0, \lambda_0^1, \dots]], \dots, [\omega_{N-1}, \tau_{N-1}, f_{N-1}, [\lambda_{N-1}^0, \lambda_{N-1}^1, \dots]]] \quad (3.39)$$

notiert werden. Hierbei sind ω_n, τ_n, f_n und $\lambda_n^0, \lambda_n^1, \dots$ Konstanten, welche die räumlichen Positionen der elementaren symbolischen Konstituenten der Signatur definieren. Einer Second-Level-Annotation entspricht eine atomare Formel der Form $\omega^S(\psi, \tau, f, \Lambda, \Theta)$.

Die Toleranzgrenzen t^{tol} und f^{tol} sowie der hierarchische Aufbau eines SLC sind von entscheidender Bedeutung für die Performanz des beschriebenen Annotationsverfahrens. In textueller Notation ist eine symbolische Signatur intuitiv nur schwer zu verstehen – siehe Abbildung 3.17. Zu den KDSE-Aufgaben gehört daher auch die Möglichkeit, symbolische Signaturen und SLC-Annotationen grafisch darzustellen.

Abbildung 3.18 zeigt eine Möglichkeit, SLC zu visualisieren: Hierarchische Ebenen werden farbkodiert dargestellt. Der Parameter t^{tol} beträgt hier 30% der zeitlichen Ausdehnung der Konstituenten von ϕ^S und wird durch senkrechte Linien angedeutet. f^{tol} ist hier Null, kann aber analog zu t^{tol} durch waagerechte Linien gekennzeichnet werden.


```

pfiff01(36,8,16711680,0,0,51939,16098,4615384615,10486,1538461538,1,5,
[Models\20msClones\b37.trm,0,6646,15384615385,[2221,0,0,1033,84615384615,b 3780,top,AND,4,0]],
[Models\20msClones\b39.trm,763,6941,53846153846,[2221,763,0,1033,84615384615,b 398763,top,AND,5,0]],
[Models\20msClones\b41.trm,1534,7236,92307692308,[2221,1534,0,1033,84615384615,b 4181534,top,AND,4,0]],
[Models\20msClones\b43.trm,2697,7532,30769230769,[2221,2697,0,1033,84615384615,b 4382697,top,AND,5,0]],
[Models\20msClones\b45.trm,3717,7827,69230769231,[2221,3717,0,1033,84615384615,b 4583717,top,AND,7,0]],
[Models\20msClones\b47.trm,4650,8123,07692307692,[2221,4650,0,1033,84615384615,b 4784650,top,AND,7,0]],
[Models\20msClones\b49.trm,5716,8418,46153846154,[2221,5716,0,1033,84615384615,b 4985716,top,AND,5,0]],
[Models\20msClones\b51.trm,7731,8713,84615384615,[2221,7731,0,1033,84615384615,b 5187731,top,AND,7,0]],
[Models\20msClones\b53.trm,8701,9009,23076923077,[2221,8701,0,1033,84615384615,b 5388701,top,AND,7,0]],
[Models\20msClones\b55.trm,10810,9304,61538461538,[2221,10810,0,1033,84615384615,b 55810810,top,AND,8,0]],
[Models\20msClones\b57.trm,12754,9600,61538461538,[2221,12754,0,1033,84615384615,b 57812754,top,AND,8,0]],
[Models\20msClones\b59.trm,14258,9895,38461538462,[2221,14258,0,1033,84615384615,b 59814258,top,AND,8,0]],
[Models\20msClones\b61.trm,16053,10190,7692307692,[2221,16053,0,1033,84615384615,b 61816053,top,AND,8,0]],
[Models\20msClones\b63.trm,18193,10486,1538461538,[2221,18193,0,1033,84615384615,b 63818193,top,AND,8,0]],
[Models\20msClones\b65.trm,19859,10781,5384615385,[2221,19859,0,1033,84615384615,b 65819859,top,AND,8,0]],
[Models\20msClones\b67.trm,21231,11076,9230769231,[2221,21231,0,1033,84615384615,b 67821231,top,AND,8,0]],
[Models\20msClones\b69.trm,22914,11372,3076923077,[2221,22914,0,1033,84615384615,b 69822914,top,AND,8,0]],
[Models\20msClones\b71.trm,24628,11667,6923076923,[2221,24628,0,1033,84615384615,b 71824628,top,AND,8,0]],
[Models\20msClones\b73.trm,26016,11963,0769230769,[2221,26016,0,1033,84615384615,b 73826016,top,AND,8,0]],
[Models\20msClones\b75.trm,28091,12258,4615384615,[2221,28091,0,1033,84615384615,b 75828091,top,AND,8,0]],
[Models\20msClones\b77.trm,29529,12553,8461538462,[2221,29529,0,1033,84615384615,b 77829529,top,AND,8,0]],
[Models\20msClones\b79.trm,31248,12849,2307692308,[2221,31248,0,1033,84615384615,b 79831248,top,AND,7,0]],
[Models\20msClones\b81.trm,32631,13144,6153846154,[2221,32631,0,1033,84615384615,b 81832631,top,AND,7,0]],
[Models\20msClones\b83.trm,33710,13440,6153846153,[2221,33710,0,1033,84615384615,b 83833710,top,AND,8,0]],
[Models\20msClones\b85.trm,35016,13735,3846153846,[2221,35016,0,1033,84615384615,b 85835016,top,AND,8,0]],
[Models\20msClones\b87.trm,36568,14030,7692307692,[2221,36568,0,1033,84615384615,b 87836568,top,AND,8,0]],
[Models\20msClones\b89.trm,38169,14326,1538461538,[2221,38169,0,1033,84615384615,b 89838169,top,AND,8,0]],
[Models\20msClones\b91.trm,39672,14621,5384615385,[2221,39672,0,1033,84615384615,b 91839672,top,AND,8,0]],
[Models\20msClones\b93.trm,41517,14916,9230769231,[2221,41517,0,1033,84615384615,b 93841517,top,AND,8,0]],
[Models\20msClones\b95.trm,43773,15212,3076923077,[2221,43773,0,1033,84615384615,b 95843773,top,AND,7,0]],
[Models\20msClones\b97.trm,45642,15507,6923076923,[2221,45642,0,1033,84615384615,b 97845642,top,AND,7,0]],
[Models\20msClones\b99.trm,46810,15803,0769230769,[2221,46810,0,1033,84615384615,b 99846810,top,AND,6,0]],
[Models\20msClones\b101.trm,47289,15803,0769230769,[2221,47289,0,1033,84615384615,b 99847289,top,AND,6,0]],
[Models\20msClones\b101.trm,48902,16098,4615384615,[2221,48902,0,1033,84615384615,b101848902,top,AND,4,0]],
[Models\20msClones\b101.trm,49424,16098,4615384615,[2221,49424,0,1033,84615384615,b101849424,top,AND,4,0]],
[Models\20msClones\b101.trm,49718,16098,4615384615,[2221,49718,0,1033,84615384615,b101849718,top,AND,4,0]]).

```

Abbildung 3.17: Textuelle Notation einer symbolischen Signatur. Die Eigenschaften des SLC sind intuitiv nur schwer zu erfassen.

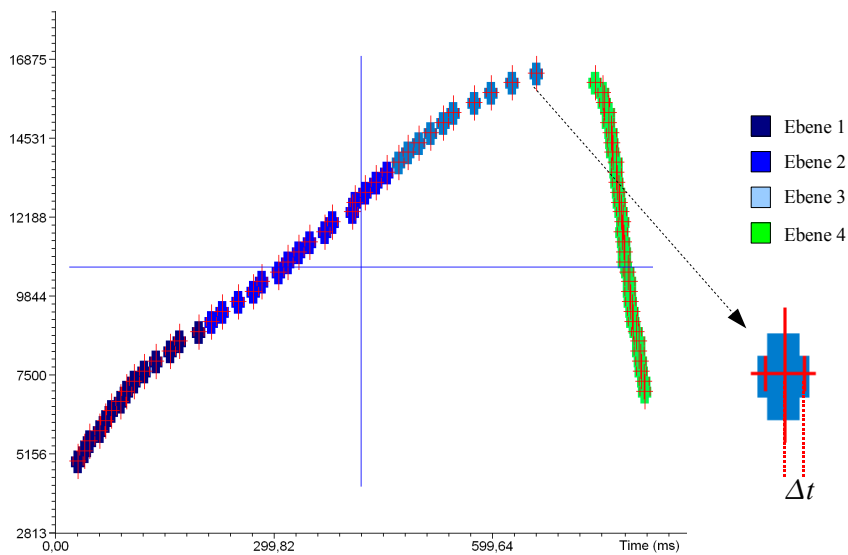


Abbildung 3.18: Visualisierung von Konstituenten, Hierarchie und Toleranz in einem SLC. Durch die grafische Darstellung sind die Eigenschaften des SLC intuitiv besser zu erfassen.

Im Rahmen der Visualisierung der Ergebnisse von KDD-Schritten kann von Interesse sein, welche Konstituenten einer symbolischen Signatur in welchen Situationen instanziiert werden können und welche nicht. Auch dies lässt sich mit einer farblichen Kodierung der entsprechenden Annotationsentscheidungen darstellen. Abbildung 3.19 zeigt ein Beispiel.

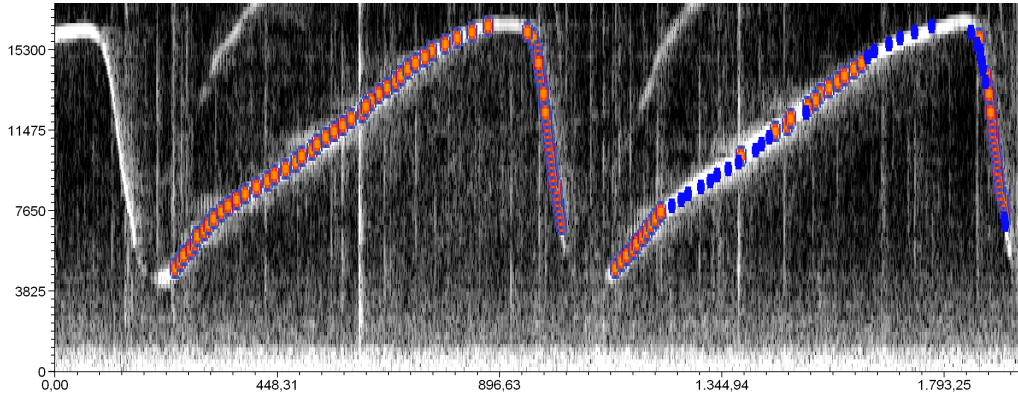


Abbildung 3.19: Visualisierung von SLC-Annotationen mit $i^{tol} = 25$: Links ein Pfiff laut, für den alle Konstituenten der symbolischen Signatur des SLC aus Abbildung 3.18 instanziiert werden konnten. Rechts ein ähnlicher Pfiff laut, für den nur ein Teil der Konstituenten instanziiert werden konnte. Alle nicht instanziierten Konstituenten sind blau dargestellt.

3.4.4 Automatische Extraktion von symbolischen Signaturen

Symbolische Signaturen können aus einer Menge F^K von Annotationsfakten automatisch extrahiert werden. Die Audiodateien, über denen die Menge F^K gebildet wurde, müssen dabei nicht zur Verfügung stehen. Im Folgenden wird ein allgemeines Verfahren zur automatischen Extraktion von *sequentiellen* symbolischen Signaturen dargestellt, das Nachbarschaftsbeziehungen von Elementen in F^K ausnutzt. Sequentiell bedeutet hier, dass die Konstituenten der Signaturen zeitlich geordnet sind. Aus Gründen der Übersichtlichkeit wird im Folgenden davon ausgegangen, dass F^K über genau einer Audiodatei generiert wurde.

Das Verfahren basiert auf folgenden Annahmen und Definitionen:

1. Es sei $\tau : F^K \rightarrow \mathbb{R}$ eine Funktion, die den Zeitpunkt eines Elementes in F^K angibt. Fernerhin sei $f : F^K \rightarrow \mathbb{R}$ eine Funktion, die den Frequenzmittelpunkt eines Elementes in F^K angibt.
2. Eine symbolische Signatur ϕ_X^S kann durch eine triviale Transformation einer Konstellation $X \subseteq F^K$ erhalten werden. Hierzu ist es lediglich notwendig die Annotationsfakten durch Weglassung des Dateilabels und Normalisierung der

Zeitinformation in symbolische Elemente zu transformieren. Sei $\zeta : \mathcal{P}(F^K) \rightarrow \Phi^S$ die Transformationsfunktion, so kann jede gesuchte symbolische Signatur mit Hilfe von ζ erhalten werden, d.h.

$$\forall X \subseteq F^K : X \xrightarrow{\zeta} \phi_X^S \quad (3.40)$$

3. Es kann ein Kriterium $\beta : \mathcal{P}(F^K) \rightarrow \{1, 0\}$ definiert werden, mit dessen Hilfe aus F^K die Menge \mathbf{X} aller zur ζ -Transformation *zulässigen* Untermengen ausgewählt werden kann, so dass

$$\mathbf{X} = \{X_0, \dots, X_{N-1} | X_n \subseteq F^K \wedge \beta(X_n) = 1\}, 0 \leq n < N \quad (3.41)$$

Somit ist die Menge Φ_β^S der gesuchten Signaturen gegeben durch

$$\Phi_\beta^S = \{\zeta(X_0), \dots, \zeta(X_{N-1}) | X_n \in \mathbf{X}\}, 0 \leq n < N \quad (3.42)$$

4. Sind die Elemente in F^K zeitlich geordnet, d.h. $\tau(F_i) \leq \tau(F_{i+1}), 0 \leq i < |F^K|$, so lässt sich β als Verkettung benachbarter Elemente definieren:

$$\beta(X) = \begin{cases} 1 & \text{if } \forall F_i \in X : \nu(F_i, F_{i+1}) = 1, 0 \leq i < |X| - 1 \\ 0 & \text{sonst} \end{cases} \quad (3.43)$$

wobei $\nu : F^K \times F^K \rightarrow \{1, 0\}$, eine Aussage darüber macht, ob zwei FLC-Annotationen F_a, F_b im Zeit-Frequenz-Raum benachbart sind.

5. Die Nachbarschaftsbeziehung ν kann wie folgt definiert werden: Δt_{max} und Δf_{max} seien Schwellwerte für maximal zulässige Entfernungen zweier Elemente auf der Zeit-bzw. Frequenzachse. Fernerhin seien Δt_{min} und Δf_{min} Schwellwerte für den minimalen Abstand zweier Elemente auf der Zeit-bzw. Frequenzachse. Es gelte

$$\nu(F_a, F_b) = \begin{cases} 1 & \text{if } (|\tau(F_b) - \tau(F_a)| < \Delta t_{max}) \wedge \\ & (|f(F_b) - f(F_a)| < \Delta f_{max}) \wedge \\ & (|\tau(F_b) - \tau(F_a)| > \Delta t_{min}) \wedge \\ & (|f(F_b) - f(F_a)| > \Delta f_{min}) \\ 0 & \text{sonst} \end{cases} \quad (3.44)$$

Ausgehend von diesen Definitionen kann Φ_β^S *supervised* oder *unsupervised* durch folgende Schritte aus F^K generiert werden.

Algorithmus: Extraktion symbolischer Signaturen

Input: $F^K, \Delta t_{min}, \Delta f_{min}, \Delta t_{max}, \Delta f_{max}$

Output: Φ_β^S

1. Ordne F^K nach der Zeit der Elemente.
2. Lösche alle Elemente aus F^K die entsprechend der Gleichung 3.44 keinen Nachbarn besitzen.
3. Gebe eine minimale Länge l_{min} und optional ein maximale Länge l_{max} für die gesuchten Signaturen an.
4. Generiere die Menge \mathbf{X} entsprechend der Gleichungen 3.41 und 3.43.
5. Transformiere \mathbf{X} entsprechend der Gleichung 3.42 in Φ_β^S .
6. Bereinige ggf. \mathbf{X} im Rahmen des gegebenen KDD-Kontexts.

Abbildung 3.20 veranschaulicht die automatische Extraktion von sequentiellen symbolischen Signaturen aus F^K . In *Real-World* Problemstellungen kann der Schritt der Bereinigung von \mathbf{X} weitere nicht-triviale Schritte erfordern.

3.4.5 Verschmelzen symbolischer Signaturen

Eine weitere grundlegende Technik für die Modellierung von SLC ist das Verschmelzen mehrerer Signaturen zu einer einzigen. In Analogie zu dem Verschmelzen von elementaren Signaturen ist auch hier das Ziel, aus einer Anzahl von Signaturen diejenigen Eigenschaften herauszufiltern, die in der Mehrzahl der Signaturen vorhanden sind und gleichzeitig diejenigen Eigenschaften herauszumitteln, welche nur in einzelnen oder wenigen Beispielen vertreten sind.

Eine Menge $\Phi^S = \{\phi_0^S, \dots, \phi_{N-1}^S\}$ von symbolischen Signaturen kann einfach durch die Vereinigung ihrer Konstituenten zusammengefasst werden, so dass sich eine einzige Signatur $\tilde{\phi}^S$ durch

$$\tilde{\phi}^S = \bigcup \phi_0^S, \dots, \phi_{N-1}^S \quad (3.45)$$

ergibt. Dieses Vorgehen ist in der Praxis allerdings unbrauchbar, da in der resultierenden Menge $\tilde{\phi}^S$ in der Regel mehr Elemente als nötig enthalten sind. Dies liegt daran, dass die Identität zweier symbolischer Elemente $\phi_1^s = \langle \omega_1, \tau_1, f_1, \Lambda_1 \rangle$ und $\phi_2^s = \langle \omega_2, \tau_2, f_2, \Lambda_2 \rangle$ in den seltensten Fällen gegeben ist, da τ_1, τ_2 und f_1, f_2 reellwertige Größen sind (Λ wird im Folgenden aus Gründen der Vereinfachung nicht beachtet).

Es bedarf somit einer Verfahrensweise, die geeignete Elemente in $\tilde{\phi}^S$ identifiziert und miteinander verschmilzt. Durch eine solche Verschmelzung kann die Kardinalität von $\tilde{\phi}^S$ erheblich reduziert werden. Gleichzeitig bietet sich hier eine Möglichkeit, typische Elemente eines Musters von untypischen zu unterscheiden.

Das Verfahren basiert auf folgenden Spezifikationen:

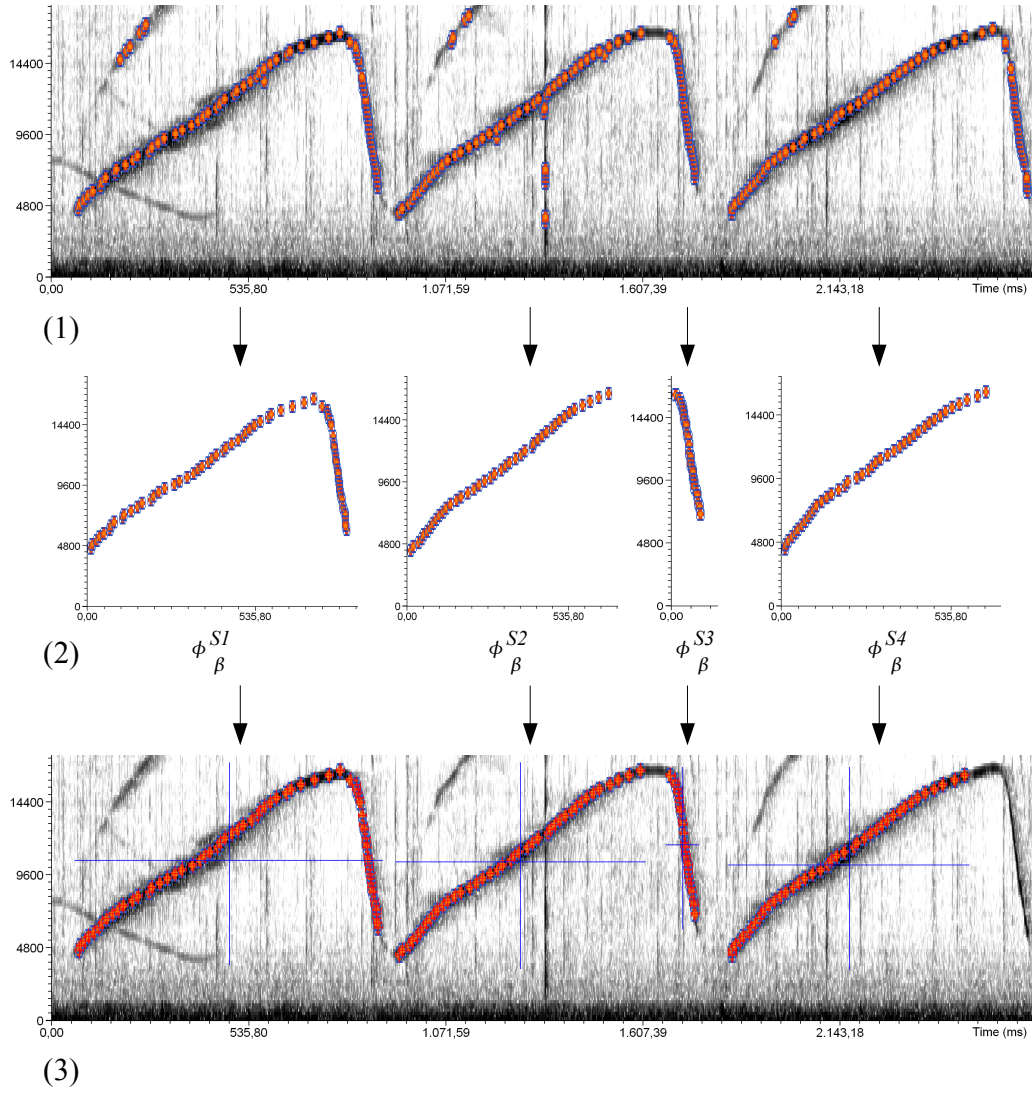


Abbildung 3.20: Automatische Extraktion von Φ_β^S : (1) Visualisierung von F^K ; (2) $\Phi_\beta^S = \{\phi_\beta^{S1}, \phi_\beta^{S2}, \phi_\beta^{S3}, \phi_\beta^{S4}\}$; (3) Annotationen F^P von Φ_β^S (blaue Fadenkreuze).

1. $\tau : \tilde{\phi}^S \rightarrow \mathbb{R}$ sei eine Funktion, die den Zeitpunkt eines symbolischen Elementes angibt. $f : \tilde{\phi}^S \rightarrow \mathbb{R}$ sei eine Funktion, die den Frequenzmittelpunkt eines symbolischen Elementes angibt. $\omega : \tilde{\phi}^S \rightarrow \Omega$ sei eine Funktion, die das Label eines symbolischen Elementes angibt. Es gelte fernerhin $M = |\tilde{\phi}^S|$.
2. $v : \tilde{\phi}^S \times \tilde{\phi}^S \rightarrow \{1, 0\}$ sei ein Kriterium, das angibt, ob zwei symbolische Elemente miteinander verschmelzbar sind oder nicht. Ausgehend von diesem Kriterium kann die Menge ϕ_V^S der verschmelzbaren symbolischen Elemente in $\tilde{\phi}^S$ definiert werden durch

$$\phi_V^S := \left\{ \phi_i^s \in \tilde{\phi}^S \mid \exists \phi_j^s : v(\phi_i^s, \phi_j^s) = 1 \right\}, i <> j, 0 \leq i < M, 0 \leq j < M \quad (3.46)$$

3. Δt_{max} und Δf_{max} seien Schwellwerte für maximal zulässige Entfernungen zweier symbolischer Elemente auf der Zeit-bzw. Frequenzachse. Das Verschmelzbarkeitskriterium v kann dann definiert werden durch

$$v(\phi_a^s, \phi_b^s) = \begin{cases} 1 & \text{if } (|\tau(\phi_b^s) - \tau(\phi_a^s)| < \Delta t_{max}) \wedge \\ & (|f(\phi_b^s) - f(\phi_a^s)| < \Delta f_{max}) \wedge \\ & (\omega(\phi_b^s) = \omega(\phi_a^s)) \\ 0 & \text{sonst} \end{cases} \quad (3.47)$$

4. Zwei miteinander verschmelzbare symbolische Elemente ϕ_a^s, ϕ_b^s in Φ_V^S werden zu einem Element ϕ_μ^s verschmolzen, indem für eines der beiden Elemente die Position im Zeit-Frequenz-Raum neu berechnet und das andere gelöscht wird.

Die neue Zeit τ_μ von ϕ_μ^s ist

$$\tau_\mu := \tau(\phi_a^s) + \frac{|\tau(\phi_a^s) - \tau(\phi_b^s)|}{2} \quad (3.48)$$

und der neue Frequenzmittelpunkt f_μ ist

$$f_\mu := \begin{cases} f(\phi_b^s) + \frac{|f(\phi_a^s) - f(\phi_b^s)|}{2} & \text{if } f(\phi_a^s) > f(\phi_b^s) \\ f(\phi_a^s) + \frac{|f(\phi_a^s) - f(\phi_b^s)|}{2} & \text{sonst} \end{cases} \quad (3.49)$$

Auf Grundlage dieser Konzepte lässt sich ein allgemeines Verfahren zum Verschmelzen einer Menge von symbolischen Signaturen wie folgt angeben:

Algorithmus: Verschmelzen symbolischer Signaturen

Input: $\Phi^S, \Delta t_{max}, \Delta f_{max}$

Output: ϕ_μ^S

1. Bilde die Vereinigung $\tilde{\phi}^S$ von Φ^S .

2. Bilde die Menge ϕ_V^S der verschmelzbaren symbolischen Elemente entsprechend Gleichung 3.46 und dem Verschmelzbarkeitskriterium in Gleichung 3.47.
3. Bilde die Menge $\phi_R^S := \tilde{\Phi}^S \setminus \Phi_V^S$ der nicht verschmelzbaren symbolischen Elemente.
4. Verschmelze alle miteinander verschmelzbaren Paare symbolischer Elemente in ϕ_V^S zu der Menge $\phi_\mu'^S$. Speichere für jedes Element in $\phi_\mu'^S$ die Anzahl der Verschmelzungs-Operationen in einer Variable *nom* (*number of merged*).
5. Bilde die verschmolzene Signatur durch $\phi_\mu^S := \phi_R^S \cup \phi_\mu'^S$

Das in diesem Abschnitt beschriebene Verfahren zum Verschmelzen von symbolischen Signaturen eignet sich insbesondere dazu, eine Menge von Signaturen zu verschmelzen, die aus einer Menge von Instanzen einer einzigen Signatur gewonnen wurden, da die erzwungene Verschmelzung durch die Vereinigungsoperation in Gleichung 3.45 nur dann sinnvoll ist, wenn alle Signaturen in Φ^S vom gleichen Grundtyp sind (vgl. Abbildung 3.21).

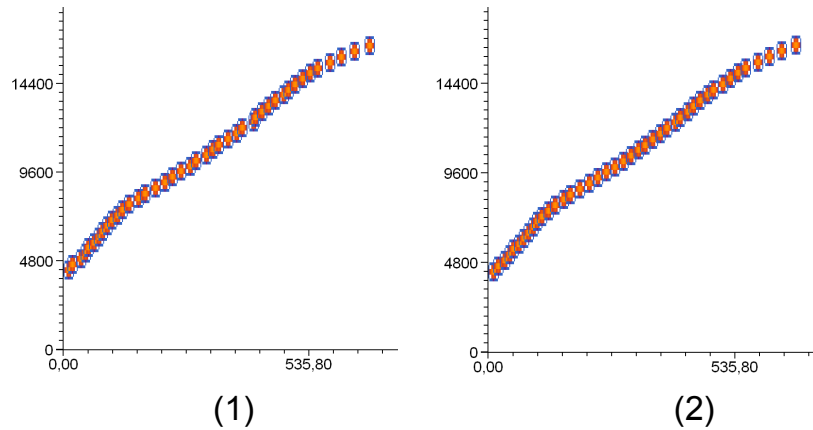


Abbildung 3.21: Verschmelzen von symbolischen Signaturen: (1) Eine automatisch extrahierte Signatur aus Abbildung 3.20; (2) Eine Signatur, die aus 19 Instanzen der ersten Signatur verschmolzen wurde. Zu erkennen ist eine gleichmäßigere Anordnung der Konstituenten

3.4.6 Ähnlichkeitsbasiertes Verschmelzen symbolischer Signaturen

In diesem Abschnitt wird eine Erweiterung des im vorausgehenden Abschnitt beschriebenen Verfahrens zum Verschmelzen symbolischer Signaturen dargestellt. Das Verfahren gestattet, Elemente innerhalb einer Menge von Signaturen unter Berücksichtigung eines Ähnlichkeitskriteriums automatisch zu verschmelzen. Diese Technik

kann dazu verwendet werden, um typische Muster oder Abweichungen von typischen Mustern in akustischen Daten zu entdecken.

Da unterschiedliche symbolische Signaturen unterschiedliche zeitliche Ausdehnungen besitzen können, muss beim ähnlichkeitsbasierten Verschmelzen von der absoluten Zeit der symbolischen Elemente der Signaturen abstrahiert werden. Anstatt dieser muss für den Ähnlichkeitsvergleich von zwei Signaturen eine optimale Zeitbasis $\Delta\tau^{opt}$ dynamisch berechnet werden. Das Verfahren basiert auf den Definitionen in Abschnitt 3.4.5 und einigen zusätzlichen Konzepten:

1. $\sigma^{\Delta\tau} : \Phi^S \times \Phi^S \rightarrow \mathbb{N}$ sei eine Funktion, die angibt, wieviele Elemente zweier symbolischer Signaturen $\phi_a^S, \phi_b^S \in \Phi^S$ entsprechend dem Verschmelzbarkeitskriterium in Gleichung 3.47 miteinander verschmelzbar sind, wenn alle symbolischen Elemente in ϕ_b^S um einen Wert $\Delta\tau$ auf der Zeitachse verschoben werden. $\Delta\tau$ kann positiv oder negativ sein.
2. Die für eine Verschmelzung optimale Zeitbasis $\Delta\tau^{opt}$ sei der *kleinst mögliche* Wert, um den die Elemente in ϕ_b^S auf der Zeitachse verschoben werden müssen, damit $\sigma^{\Delta\tau}$ den *maximal möglichen* Wert erreicht. $\sigma^{opt} : \Phi^S \times \Phi^S \rightarrow \mathbb{N}$ sei die maximal mögliche Anzahl miteinander verschmelzbarer symbolischer Elemente zweier symbolischer Signaturen.
3. Das Kriterium $\sigma^{min} \in \mathbb{N}$ gibt die für eine Verschmelzung minimal erforderliche Anzahl symbolischer Elemente an. σ^{min} kann konstant sein oder dynamisch für jeden Vergleich neu berechnet werden, z.B. in Prozent der Länge der zweiten Signatur durch $\sigma^{min} := c \cdot |\phi_b^S|/100$. Zwei Signaturen sind genau dann verschmelzbar, wenn $\sigma^{opt} \geq \sigma^{min}$.

Abbildung 3.22 veranschaulicht den Ähnlichkeitsvergleich zweier symbolischer Signaturen. Der Algorithmus *Symbolic-Similarity-Driven-Merge* (SSDM) generiert aus Φ^S eine Menge von verschmolzenen Signaturen Φ_μ^S . Die Funktion *Merge* entspricht dabei dem Verfahren in Abschnitt 3.4.5:

Algorithmus: Symbolic-Similarity-Driven-Merge

Input: Φ^S, σ^{min}

Output: Φ_μ^S

begin

$N := |\Phi^S|;$

$\Phi_\mu^S := \emptyset;$

Append(Φ_0^S, Φ_μ^S);

for $i = 1$ **to** $N - 1$ **do begin**

$n := \text{Nearest}(\Phi_i^S, \Phi_\mu^S, \sigma^{min});$

if $n < -1$ **then** *Merge*(Φ_i^S, Φ_μ^S) **else** *Append*(Φ_i^S, Φ_μ^S);

end;

return $\Phi_\mu^S;$

end.

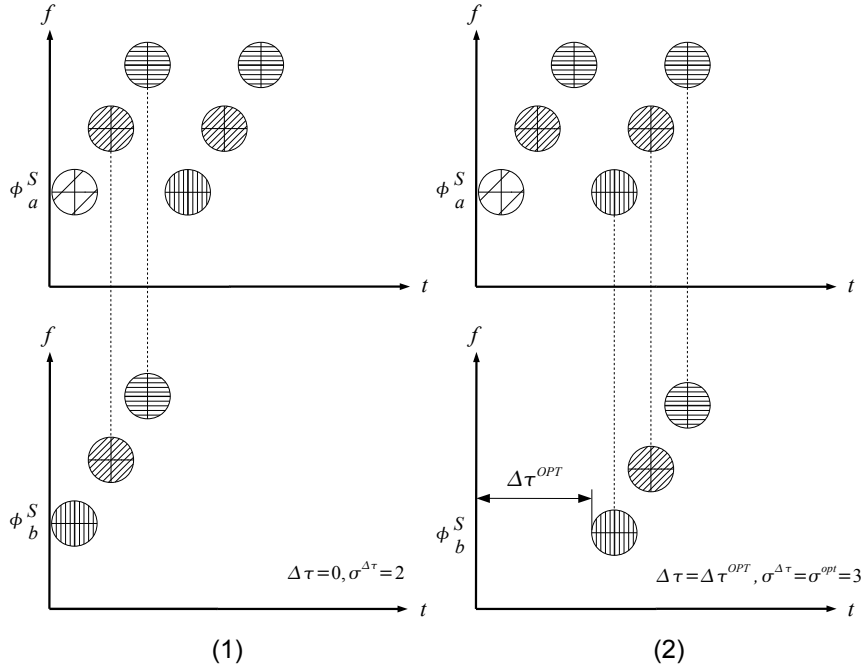


Abbildung 3.22: Vergleich von zwei symbolischen Signaturen: (1) Die Ähnlichkeit $\sigma^{\Delta\tau}$ von ϕ_a^S und ϕ_b^S ist 2 bei $\Delta\tau = 0$. (2) Die Ähnlichkeit von ϕ_a^S und ϕ_b^S ist 3 (also maximal) wenn die Zeitverschiebung optimal ist, d.h. $\Delta\tau = \Delta\tau^{opt}$. Die gestrichelten Linien symbolisieren eine Verschmelzbarkeit zwischen zwei Elementen. Die Kreise symbolisieren symbolische Elemente und ihre Position im Zeit-Frequenz-Raum. Die Schraffuren symbolisieren die verschiedenen Label der symbolischen Elemente.

Die Funktion *Nearest* findet das Element in Φ_μ^S mit der größten Ähnlichkeit zur Φ_i^S :

Function: Nearest

Input: ϕ^S, Φ_μ^S **Output:** n

```

begin
   $M := |\Phi_\mu^S|;$ 
   $n := -1;$ 
   $\sigma^{best} := \sigma^{min};$ 
  for  $j := 0$  to  $M - 1$  do begin
     $\sigma^{current} := \sigma^{opt}(\phi^S, \Phi_{\mu j}^S);$ 
    if  $\sigma^{current} \geq \sigma^{best}$  then begin
       $\sigma^{best} := \sigma^{current};$ 
       $n := j;$ 
    end;
  end;
  return  $n;$ 
end.
```

Mit geringfügigen Modifikationen kann das hier beschriebene Verfahren zum *Conceptual-Clustering* von Mustern im Zeit-Frequenz-Raum genutzt werden (siehe auch Kapitel 7). Das Verfahren ist nichtlinear, da sich die Signaturen bei jeder Verschmelzungsoperation verändern.

Im Anschluss an das einfache oder ähnlichkeitsbasierte Verschmelzen können symbolische Signaturen von nicht-häufigen Elementen bereinigt werden. Hierzu werden diejenigen symbolischen Elemente aus einer verschmolzenen Signatur gelöscht, denen ein niedriger *nom*-Wert entspricht (vgl. Abschnitt 3.4.5). Eine farbliche Kodierung des *nom*-Wertes ermöglicht einen direkten Überblick über die Ergebnisse von Verschmelzungsoperationen. Abbildung 3.23 zeigt eine mögliche Farbkodierung und veranschaulicht den Bereinigungsverfahren.

3.4.7 Induktive Programmierung

In diesem Abschnitt wird ein Verfahren zur Induktion einer endlichen Menge von symbolischen Signaturen beschrieben. Die mit Hilfe dieses Verfahrens generierten Signaturen fassen Mengen elementarer Annotationsfakten in F^K zu Gruppen zusammen. Das Verfahren beruht auf den Methoden zur automatischen Extraktion (vgl. Abschnitt 3.4.4) und ähnlichkeitsbasierten Verschmelzung (vgl. Abschnitt 3.4.6) von symbolischen Signaturen.

Die Idee besteht darin, in F^K Konstellationen von benachbarten Annotationsfakten zu suchen und mit Hilfe des Verfahrens der automatischen Extraktion (vgl. Abschnitt 3.4.4) in eine Menge von symbolischen Signaturen zu transformieren. Anschließend können die Elemente dieser Menge ähnlichkeitsbasiert verschmolzen werden. Durch Verknüpfung mit einem geeigneten Annotationsalgorithmus (vgl. Abschnitt 3.4.2) ergibt sich eine Menge P von SLC. Mit Hilfe der SLC in P ist es möglich, den Signaturen entsprechende akustische Phänomene in neuen Audiodaten maschinell zu lokalisieren:

Algorithmus: SLC-Induktion

Input: F^K

Output: P

1. Extrahiere aus F^K mit Hilfe des Verfahrens in Abschnitt 3.4.4 eine Menge Φ^S von symbolischen Signaturen.
2. Verschmelze Φ^S nach dem Verfahren in Abschnitt 3.4.6 ähnlichkeitsbasiert zu einer Menge Φ_μ^S verschmolzener symbolischer Signaturen.
3. Berechne die Signaturen Φ_μ^S von selten auftretenden symbolischen Elementen.
4. Evaluere die Signaturen in Φ_μ^S und iteriere ggf. die vorausgehenden Schritte.
5. Verknüpfe die evaluierten Signaturen in Φ_μ^S mit geeigneten Second-Level-AA zu einer Menge P von SLC.

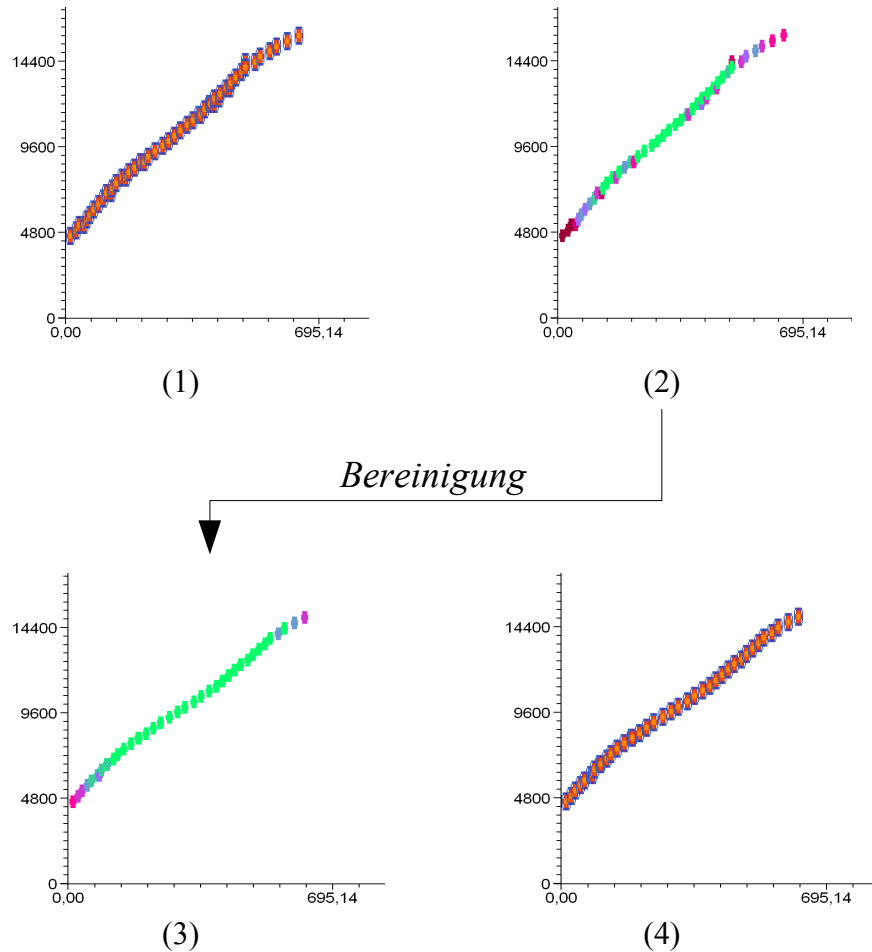


Abbildung 3.23: Visualisierung und Bereinigung symbolischer Signaturen: (1) Eine aus 41 symbolischen Signaturen verschmolzene Signatur; (2) Dieselbe Signatur mit farbkodierter Darstellung des Parameters *nom*. Die Farbe hellgrün entspricht hohen *nom*-Werten. Die Farbe dunkelrot entspricht sehr niedrigen *nom*-Werten; (3) Durch Löschen der Elemente mit *nom* < 2 bereinigte Signatur; (4) Die bereinigte Signatur in normaler nicht-farbkodierter Darstellung.

In Teil II wird gezeigt, wie dieses Verfahren zur Clusteranalyse von Delphinpfeifen eingesetzt werden kann. Ein wichtiges Evaluationskriterium für die Signaturen in Φ_μ^S ist die durch sie erzielte Kompressionsrate (siehe Kapitel 6).

3.4.8 Diskussion

Bei allen Verfahren in diesem Abschnitt handelt es sich um elementare Methoden zur Realisierung des SLCD-Prozesses. Daneben existiert eine Reihe weiterer nützlicher Hilfsmittel zur Modellierung von SLC. Hierzu gehört z.B. eine Vielzahl von Operationen zur Navigation in Datenbanken mit SLC-Annotationsfakten. Diese sind zwar von großer praktischer Bedeutung, aber für das theoretische Verständnis der folgenden Kapitel nicht erforderlich. Es wird daher auf eine Darstellung verzichtet.

Im Rahmen eines SLCD-Prozesses können mit Hilfe der oben beschriebenen Methoden SLC modelliert werden, die auf spezifische Fragestellungen im empirischen Kontext abgestimmt sind. Eine sorgfältige Durchführung des SLCD-Prozesses gestattet, Mengen von SLC zu erhalten, die für viele Problemstellungen im Bereich der maschinellen Annotation von Audiodaten geeignet sind.

Werden die algorithmischen Anteile der beschriebenen Verfahren von einem KDE zur Verfügung gestellt, ist es möglich, problemadäquate Mengen von FLC und SLC visuell-interaktiv auf der Grundlage von Expertenwissen alleine zu generieren. Hierin liegt ein wesentlicher Vorteil des Ansatzes, der in dieser Arbeit vorgestellt wird.

Teil II

Zur Bioakustik von *Tursiops truncatus*

Kapitel 4

Zur Bioakustik von *Tursiops truncatus*

KDD-Methoden besitzen ein großes Potential, um strukturelle und funktionale Eigenschaften bioakustischer Signalsysteme aufzudecken (vgl. Abschnitt 1.2). Insbesondere für datenintensive Untersuchungen an hochkomplexen Signalsystemen, wie denen der *Delphinidae*, erscheint der Einsatz solcher Methoden angezeigt. Eine wichtige Grundlage für derartige Studien sind effiziente ASK, die auf einem einheitlichen Klassifikationssystem beruhen, das alle akustischen Phänomene abdeckt, die für eine bestimmte Art relevant erscheinen. In diesem Kapitel werden die Grundzüge eines solchen Systems für die Spezies *Tursiops truncatus* dargestellt.

4.1 Der Große Tümmler

Der Große Tümmler (*Tursiops truncatus*) besitzt eine charakteristische kurze Schnauze – daher sein englischer Name *Bottlenose Dolphin*. Sein Körper ist grau, der Bauch ist weißlich. Er wird 1,9 bis 4,5m lang und wiegt 150 bis 650 kg. Seine durchschnittliche Größe beträgt ungefähr 2,40 m bei ca. 200 kg Körpergewicht. Der große Tümmler ist in allen Ozeanen beheimatet. In Europa gibt es ihn im Atlantik, im Mittelmeer, im schwarzen Meer, in der Nordsee und in der westlichen Ostsee. In tropischen Breiten ist er häufig in flachen Buchten und in Lagunen anzutreffen.

4.1.1 Taxonomie und Abstammung

Die Art *Tursiops truncatus* gehört der Familie der Delphine (*Delphinidae*) an, die ca. 33 Arten umfasst. Alle Delphine sind Zahnwale (*Odontoceti*). Neben den *Odontoceti* mit ca. 70 Arten umfasst die Ordnung der Wale (*Cetacea*) auch noch eine zweite Unterordnung – die Bartenwale (*Mysticeti*) mit ca. 15 Arten [95].

Wieviele Arten der Gattung *Tursiops* es gibt, ist unter Zoologen umstritten. Rice (1998) unterscheidet zwei Arten – den atlantischen Tümmler (*Tursiops truncatus*)

und den indopazifischen Tümmler (*Tursiops aduncus*) [95]. Andere Autoren sehen in dem pazifischen Tümmler (*Tursiops gillii*) eine dritte Art der Gattung *Tursiops*.

Die Urwale (*Archaeocetae*), die vor 30-50 Millionen Jahren lebten, gelten nicht als unmittelbare Vorfahren der heutigen *Cetacea*. Von Boisserie, Lihoreau und Brunet (2005) wurde aber anhand eines DNA-Vergleichs eine Verwandtschaft zwischen den *Cetacea* und den Flußpferden (*Hippopotamidae*) gezeigt [10]. Beide hatten anscheinend vor 60-70 Millionen Jahren einen gemeinsamen Vorfahren.

4.1.2 Lebensweise

Tursiops truncatus lebt in engen sozialen Verbänden. Die Tiere schwimmen täglich zwischen 60 und 100 km und können über 200 m tief tauchen. Eine Schule von Tümmlern besteht aus zwei bis fünfzehn Individuen. Der Schnitt liegt im Atlantik bei fünf Tieren, im Pazifik und im indischen Ozean liegt er höher. Im offenen Meer lebende Tümmler bilden tendenziell größere Schulen als küstennah lebende. Die Hierarchie in einer Schule wird meist von einem alten Männchen angeführt. Sonst gibt es im Verband nur Weibchen und Jungtiere. Jugendliche Männchen bilden eigene Schulen – sogenannte Junggesellenverbände. Die Nahrung der Tiere besteht aus kleinen Fischen, gelegentlich Tintenfischen und Krebstieren.

Die Tragzeit von *Tursiops truncatus* beträgt ein Jahr. Bei der Geburt sind die Jungen etwa 120 cm lang. Wie bei allen Walen wird meist nur ein Junges geboren. Es kommt mit der Schwanzflosse zuerst aus dem Mutterleib. Im Anschluss schubst die Mutter das Junge zu seinem ersten Atemzug an die Wasseroberfläche. Zwillinge sind selten und bleiben meist nicht lange am Leben, da die Muttermilch in der Regel für zwei Junge nicht ausreicht. Die Jungtiere bleiben ungefähr drei Jahre bei der Mutter, davon wird ein Jahr gesäugt. Die Tiere werden im Alter von 9-10 Jahren geschlechtsreif. Die Lebensdauer der Großen Tümmlers beträgt in freier Wildbahn ca. 25 Jahre, in Gefangenschaft bis zu 35 Jahre.

4.1.3 Sinne und Gehirn

Tursiops truncatus besitzt hochentwickelte Sinne. Hervorzuheben sind der Gesichtssinn, der Sonar- und Gehörsinn, der Geschmackssinn sowie der Tastsinn [67, 2, 4]. Auf diesen Sinnen basieren verschiedene Kommunikationsmodalitäten. Die akustische spielt dabei eine besonders wichtige Rolle. Zur Mimik ist die Gattung *Tursiops* nicht fähig, da die entsprechenden Gesichtsmuskeln fehlen.

Das Gehirn von *Tursiops truncatus* wiegt im Schnitt 1.7 kg. Das durchschnittliche Gehirngewicht liegt damit um mehrere hundert Gramm über dem des Menschen. Dieser Umstand gab in der Vergangenheit Anlass zu zahlreichen Spekulationen über den Zusammenhang von Gehirnmasse und kognitiver Performanz [121]. Da es nicht sinnvoll ist, das Gehirngewicht als einziges Maß für kognitive Performanz heranzuziehen, wurden noch weitere Aspekte in diesen Diskurs eingebracht. Hierzu gehören (1) das Verhältnis von Körper- zu Gehirnmasse, (2) das Verhältnis von Gehirnmasse

zum Durchmesser der Rückenmarkkanäle, (3) der Grad der Faltung des Neocortex, (4) die Anzahl neuronaler Layer im Neokortex (5 nicht 6-7, wie beim Menschen), (5) die mittlere Neuronendichte im Neokortex, (6) das Verhältnis von Größe des Kleinhirns zur Gesamtgröße des Gehirns und (7) die Gehirngröße zum Geburtszeitpunkt.

Der Diskurs zu diesem Thema beinhaltet eine Anzahl von Argumenten, die einen vorsichtigen Umgang mit Rückschlüssen auf eine direkte Korrelation zwischen cerebraler Masse und kognitiver Performanz nahelegen. Hierzu gehören prinzipielle Schwierigkeiten, kognitive Performanz zu messen und quantitativ zu beschreiben, aber auch der Umstand, dass in der Evolutionsgeschichte der Delphine ein erhöhter evolutionär Druck zur Minimierung der Gehirnmasse vermutlich gefehlt hat, da das fettreiche Gehirngewebe zusätzlichen Auftrieb in der knochenreichen Kopfgregion verleiht. Eine Besonderheit ist allerdings, dass *Tursiops truncatus* offenbar in der Lage ist, sich selbst im Spiegel zu erkennen [91]¹.

4.2 Das Eilat-Korpus

Ausgangsmaterial für die vorliegende Untersuchung ist ein Korpus von Aufzeichnungen der Vokalisationen des Großen Tümmlers. Ziel bei der Herstellung des Korpus war, so viele Signale wie möglich in maximal hoher Qualität aufzuzeichnen. Das Korpus wurde im Dolphin-Reef in Eilat/Israel (siehe <http://www.dolphinreef.co.il>) in der Zeit vom 24.5.2002 bis einschließlich 20.6.2002 an insgesamt 28 Aufnahmetagen hergestellt. Die Aufnahmen erfolgten meist in der Zeit zwischen 8:00 und 22:00. Am 1.06.2002 konnte aus technischen Gründen keine Aufnahme angefertigt werden.

Die Aufnahmen umfassen insgesamt 87,9 Gigabyte in 453 Dateien. Jede Datei entspricht dabei genau einer einzelnen nicht unterbrochenen ungekürzten Aufnahme. Die Größe der Dateien variiert zwischen 50 MB und 700 MB. Das Dateiformat ist unkomprimiertes MS-PCM, Mono, Enkodiertiefe 16 Bit, Abtastfrequenz 96kHz. Die Gesamtaufnahmedauer beträgt 273,27 Stunden. In Abbildung 4.1 sind Aufnahmezeiten, die Aufnahmedauer pro Tag und die Aufnahmedauer im zehn Minuten-Intervall nach der Tageszeit dargestellt.

4.2.1 Aufnahmetechnik

Für die Aufnahmen wurde das sphärische Breitband Hydrophon TC 4014 der Firma RESON verwendet. Das Hydrophon besitzt einen eingebauten Vorverstärker. Es hat einen relativ ausgeglichenen Frequenzverlauf im Bereich von 30 Hz bis weit über 200 kHz. Der Hersteller macht folgende Angaben zu dem Gerät: (1) Frequenzbereich: 15 Hz – 480 kHz; (2) Linearer Frequenzbereich: 30 Hz – 100 kHz \pm 2 dB, 25 Hz – 250 kHz \pm 3dB; (3) Horizontales Direktivitätsmuster: Omnidirektional \pm 2dB bei 100kHz; (4) Vertikales Direktivitätsmuster: 270° \pm 2dB bei 100 kHz (siehe <http://www.reson.com/TC4014.html>).

¹Filmmaterial hierzu unter: <http://www.pnas.org/cgi/content/full/101086398/DC1/1>

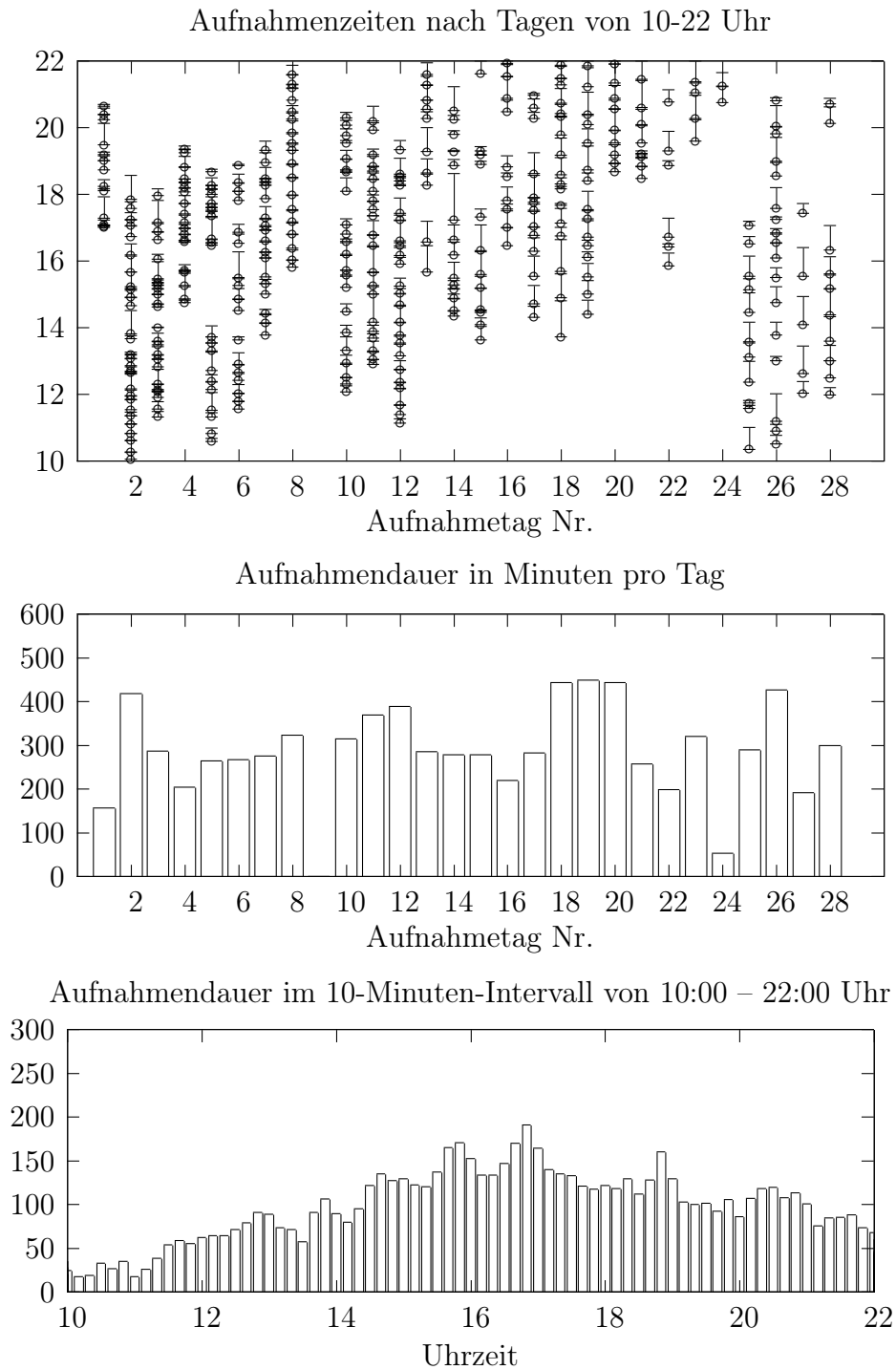


Abbildung 4.1: Aufnahmezeiten im Eilat-Korpus. Der erste Aufnahmetag war der 24.5.2002, der letzte der 20.6.2002.

Die Ausgangssignale des Hydrophons wurden über ein drei Meter langes abgeschirmtes Kabel mit der externen Soundkarte DSP24 der Firma Hoontech mit 96 kHz digitalisiert und in Echtzeit auf Festplatte geschrieben. Die DSP24 besitzt einen Vorverstärker, der so eingestellt wurde, dass Übersteuerungen in der Aufnahme minimiert wurden. Diese waren aufgrund der hohen Lautstärke der Klicklaute (s.u.) aber nicht immer zu vermeiden.

4.2.2 Randbedingungen

Das Dolphin-Reef in Eilat ist kein Delphinarium im herkömmlichen Sinn. Die Tiere leben in einem Unterwassergehege im Meer und können das Gehege jederzeit durch einen Ausgang verlassen. Das Gehege befindet sich an einem Sandstrand. Der Innenbereich des Geheges ist vom offenen Meer nur durch einen Unterwasserzaun aus Kunststoff getrennt. Das Hydrophon wurde für die Aufnahmen stationär ca. fünf Meter vom Ufer entfernt befestigt. Dies erfolgte außerhalb des für Besucher zugänglichen Areals. An insgesamt zehn Aufnahmetagen in der Zeit nach 17:00 wurde das Hydrophon unterhalb eines schwimmenden Steges im Zentrum des Unterwassergeheges angebracht.

Der Meeresboden im Dolphin-Reef besteht vorwiegend aus Sand. Akustische Störungen durch Echos waren daher gering. Geringfügige Verzerrungen der Signale durch Reflexionen an anderen im Wasser befindlichen Gegenständen (Unterwasserzaun, Boote, Schnorchler, Holz- und Plastikkonstruktionen, Schwimmsteg) waren unvermeidbar.

Die meisten Beeinträchtigungen der Aufzeichnungen wurden durch Klicklaute von Garnelen verursacht. Daneben gab es noch eine Reihe weiterer mehr oder weniger häufiger Störfaktoren: (1) eine Wasserpumpe, (2) Motorgeräusche von Schiffen, (3) akustisches Schiffs-Sonar, (4) pulsierte HF-Strahlung unbekannter Provenienz (evtl. verursacht durch Mobilfunktelefone), (5) Wellengeräusche und (6) temporäre Störungen im A/D-Wandler bei 29 und 35 kHz.

Aufgrund von sehr starken Störungen durch Motorgeräusche mussten einige Aufnahmen sofort gelöscht bzw. die Aufnahme unterbrochen werden. Die meisten Störungen konnten aber toleriert werden. Die Abbildungen 4.2 und 4.3 veranschaulichen dies. Sie zeigen nur Hintergrundgeräusche ohne Vokalisationen von *Tursiops truncatus*. Abbildung 4.4 zeigt eine typische Aufnahme von Pfiff- und Klicklauten aus dem Korpus. Pfffe und pulsierte Laute setzen sich sehr klar gegenüber den Hintergrundgeräuschen ab. Auch Obertöne und feine Modulationen sind gut zu erkennen. Zusammenfassend kann gesagt werden, dass mit der oben beschriebenen Anordnung sehr viele verwertbare Aufnahmen der Vokalisationen von *Tursiops truncatus* im Bereich von 50 Hz – 48 kHz erstellt werden konnten.

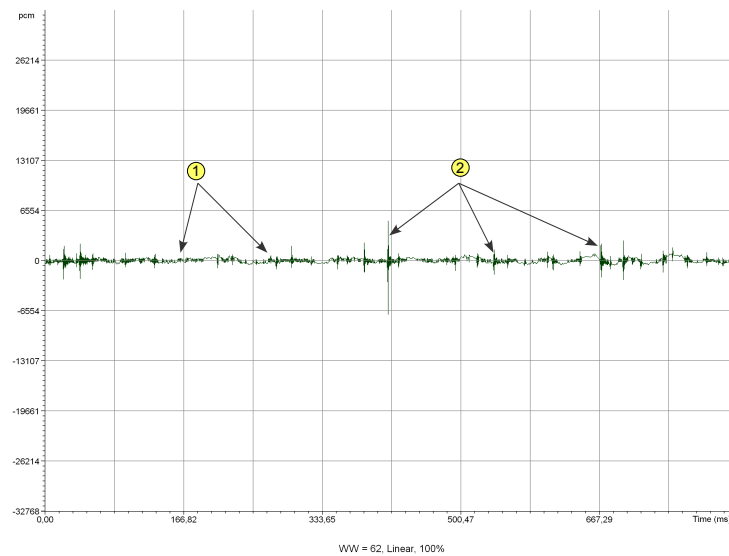


Abbildung 4.2: Oszillogramm typischer Unterwassergeräusche in Eilat. (1) Noise; (2) Klicklaute von Garnelen. Der Ausschnitt ist in Abbildung 4.3 in spektrographischer Darstellung abgebildet.

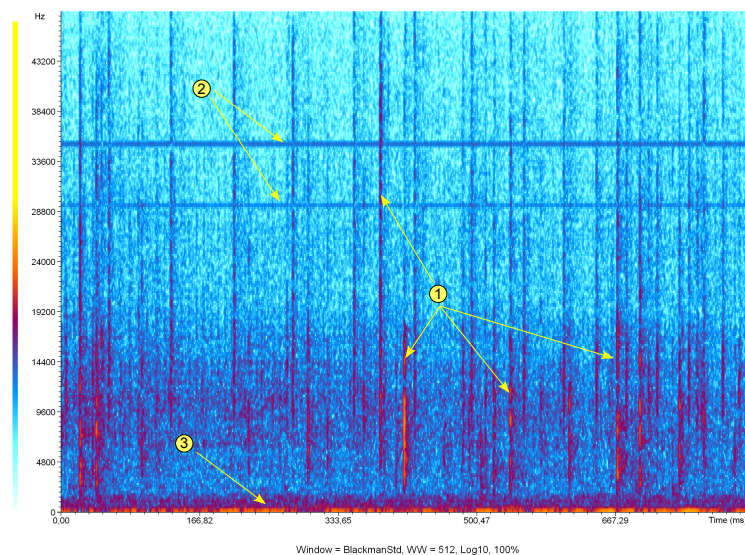


Abbildung 4.3: Spektrogramm typischer Unterwassergeräusche in Eilat. (1) Klicklaute von Garnelen; (2) Elektromagnetische Störungen; (3) Niederfrequenter Noise (Wellengeräusche etc.).

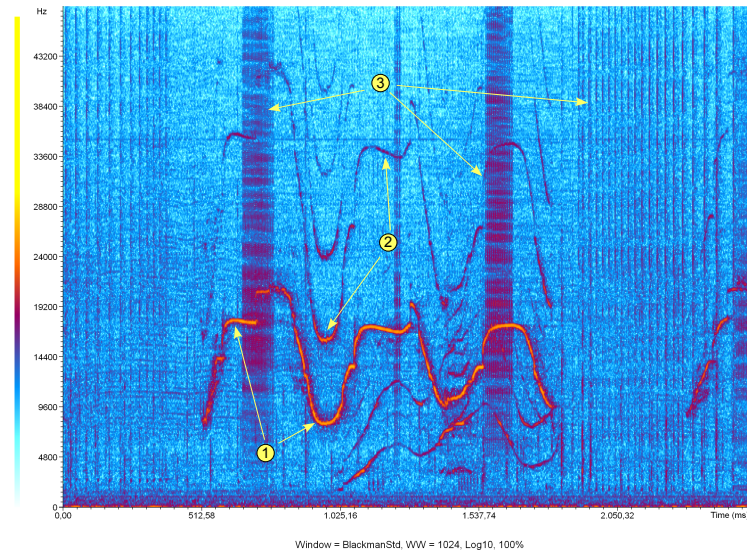


Abbildung 4.4: Beispiel einer verwertbaren Aufnahme von Vokalisationen des Großen Tümmlers. (1) Pfiffe; (2) Obertöne von Pfiffen; (3) Klicksequenzen (*Trains*);

4.3 Akustische Phänomene im Eilat-Korpus

In diesem Abschnitt werden die im Eilat-Korpus beobachtbaren akustischen Phänomene im Überblick dargestellt. Die Phänomene werden zunächst mit Hilfe eines zweistufigen Klassifikationssystems grundlegend charakterisiert. Im Anschluß hieran wird eine Vielzahl von phonetischen Modulationsphänomenen im Detail dargestellt.

Das in Schwingungs- und Grundtypen gegliederte Klassifikationssystem ist nicht zu verwechseln mit dem Lautinventar der Delphingruppe in Eilat. Auch geht nicht jedes dargestellte Phänomen notwendigerweise auf eine biologisch relevante Modulation zurück.

4.3.1 Ziele

Grundlage der Modellierung effizienter ASK für die Vokalisationen von *Tursiops truncatus* ist ein einheitliches, auf den physikalischen Eigenschaften der Signale basierendes Klassifikationssystem, das alle relevanten akustischen Phänomene abdeckt. Ziel der folgenden Untersuchung ist es, die Grundzüge eines solchen Systems darzustellen. Hierbei wird eine Vielzahl von Phänomenen systematisiert und visuell charakterisiert. Das Klassifikationssystem ist rein phänomenologisch, d.h. es wurde ohne weitere technische oder statistische Hilfsmittel erstellt.

4.3.2 Material und Methode

Das verwendete Material ist das in Abschnitt 4.2 beschriebene Eilat-Korpus. Jede der 453 Dateien wurde in die Untersuchung miteinbezogen.

Der Inhalt aller Dateien wurde mit Hilfe der in Abschnitt 3.2 beschriebenen Methode an einem hochauflösenden Bildschirm visualisiert. Zusätzlich wurde das Audiomaterial mit Hilfe von Oszillogrammen und durch Abhören in verschiedenen Geschwindigkeiten untersucht.

Interessante akustische Phänomene wurden mit Hilfe von Containerdateien archiviert. In jeder Containerdatei wurden insgesamt 31 verschiedene Parameter, darunter zeitliche Koordinaten, oszillographische, spektrographische und farbliche Einstellungen, gespeichert. Kriterien für die Klassifikation waren die elementaren Signaleigenschaften Schwingungsform (*Waveform*), Amplitude, Dauer und Verteilung der spektralen Energie sowie Arten der phonetischen Modulation (siehe Abschnitt 4.5.4). Die weiter unten abgebildeten Beispiele stellen einen Auszug aus dem so angelegten Phänomen-Archiv dar.

4.4 Schwingungstypen und Grundtypen

Die Vokalisationen von *Tursiops truncatus* lassen sich auf zwei grundlegenden Ebenen klassifizieren:

1. *Schwingungstypen* lassen sich benennen, wenn man das Signal auf der Ebene von Schwingungsform und Amplitude betrachtet.
2. *Grundtypen* lassen sich benennen, wenn man zusätzlich Vokalisationsdauer und spektrale Energieverteilung betrachtet.

Die allgemeinen Charakteristika von Schwingungs- und Grundtypen werden in den beiden folgenden Abschnitten dargestellt. Phonetische Modulationsprinzipien werden hierbei zunächst noch nicht behandelt.

4.4.1 Schwingungstypen

Traditionell unterscheidet man beim Großen Tümmler zwei Typen von Vokalisationen, die mit besonderen, im Kopf der Tiere befindlichen Organen produziert werden – *Pfiff-laute* und *pulsierte Laute* [35, 52, 89]. Pfiff-laute dienen der Kommunikation [66, 52, 82]. Pulsierte Laute dienen dem Sonar [2] und kommunikativen Zwecken [14, 66, 74, 82, 68]. Im Frequenzbereich von 0-48 kHz lassen sich Piffe und pulsierte Laute sehr oft beobachten. Sonar-laute können weit über 48 kHz hinausreichen (siehe aber Abschnitt 7). Hinzu kommen Laute, die mit anderen Körperteilen, so z.B. mit der Schwanzflosse, erzeugt werden.

Betrachtet man Schwingungsform und Amplitude der Vokalisationen gemeinsam, so lassen sich sinusartige Schwingungen, pulsierte Laute und Klick-laute voneinander unterscheiden. In dieser Arbeit wird daher zwischen folgenden Schwingungstypen unterschieden:

1. **Sinusartiger Schwingungstyp:** Die dominierende Schwingungsform ähnelt in der oszillographischen Darstellung einer Sinuskurve. Die Amplitude liegt im mittleren Bereich.
2. **Pulsierter Schwingungstyp:** Die dominierende Schwingungsform besteht aus Sequenzen von ineinander übergehenden Pulsen in schneller Abfolge. Die Amplitude liegt im unteren bis mittleren Bereich. Die einzelnen Pulse sind manchmal schwer als solche zu erkennen. Zwischen den einzelnen Pulsen lassen sich keine signalfreien Lücken erkennen.
3. **Klick-Schwingungstyp:** Die dominierende Schwingungsform besteht aus Sequenzen (*Trains*) von scharf voneinander abgesetzten, sehr energiereichen Klicklauten. Die Amplitude liegt im oberen Bereich. Zwischen den einzelnen Klicklauten sind signalfreie Lücken deutlich zu erkennen.

Die drei verschiedenen Schwingungstypen sind in den Abbildungen 4.5, 4.6 und 4.7 dargestellt.

Die obige Einteilung ist nicht immer verbindlich. Es gibt sinusartige Signale, die in pulsierte übergehen. Häufig gehen pulsierte Signale auch in Klick-Sequenzen über oder befinden sich in unmittelbarer Nachbarschaft zu diesen. Oft werden die verschiedenen Signaltypen zugleich, überlagert oder in schneller Abfolge hintereinander produziert.

4.4.2 Grundtypen

Für eine greifbare Klassifikation in Vokalisations-Grundtypen ist eine Unterteilung auf der alleinigen Grundlage von Schwingungsform und Amplitude nicht ausreichend. Bezieht man noch die Vokalisations-Dauer und Charakteristika der spektralen Energieverteilung in die Betrachtung mit ein, so ergeben sich im Bereich von 0-48 kHz folgende Grundtypen von Vokalisationen mit charakteristischen überlappenden Frequenzbereichen:

1. **Pfiff-laute:** Diese Laute basieren auf einer sinusartigen Schwingungsform, die in vielfältiger Art und Weise moduliert werden kann. Die Amplitude liegt meist im mittleren Bereich. Pfiff-laute können aber auch sehr laut werden. Sie sind sehr häufig im Bereich von ca. 2-22 kHz. Selten liegen sie darüber. Der Pfiff-laut mit der höchsten Grundfrequenz im Eilat-Korpus erreichte ca. 32,8 kHz (siehe Abbildung 4.9). Die Obertöne von Pfiff-lauten können Werte von über 48 kHz erreichen. Die grundlegenden Charakteristika dieses Vokalisationstypus sind in Abbildung 4.8 grafisch dargestellt.
2. **Niederfrequente nichtpulsierte Laute:** NFNP-Laute basieren auf einer sinusartigen Schwingungsform. Sie heben sich von pulsierten niederfrequenten Lauten durch das völlige Fehlen einer Pulsstruktur ab. Die Amplitude liegt im unteren Bereich. Die Grundfrequenz liegt typischerweise unterhalb von 700 Hz.

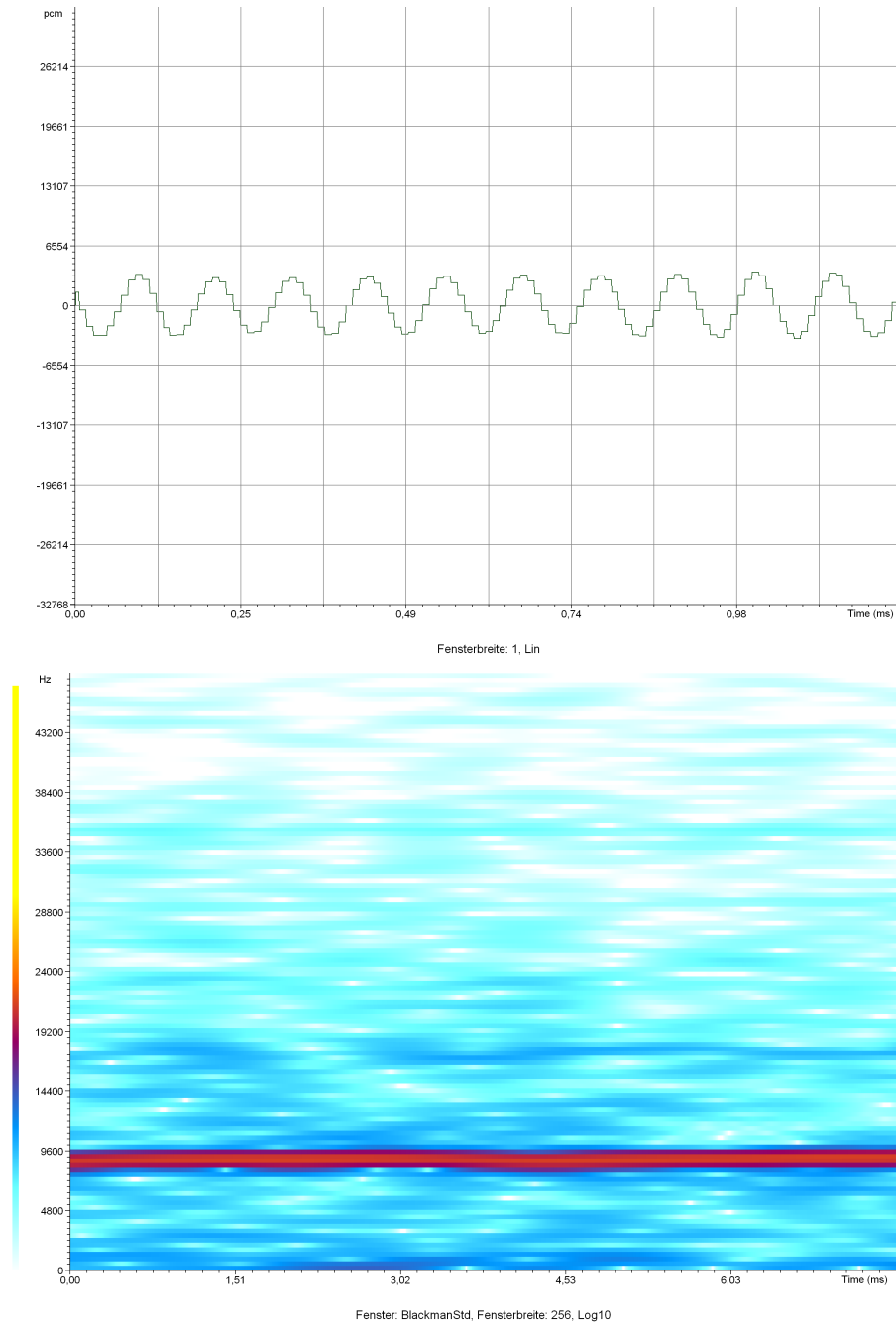


Abbildung 4.5: Oben: Sinusartiger Schwingungstyp in oszillographischer Darstellung. Unten: Sinusartiger Schwingungstyp in spektrographischer Darstellung.

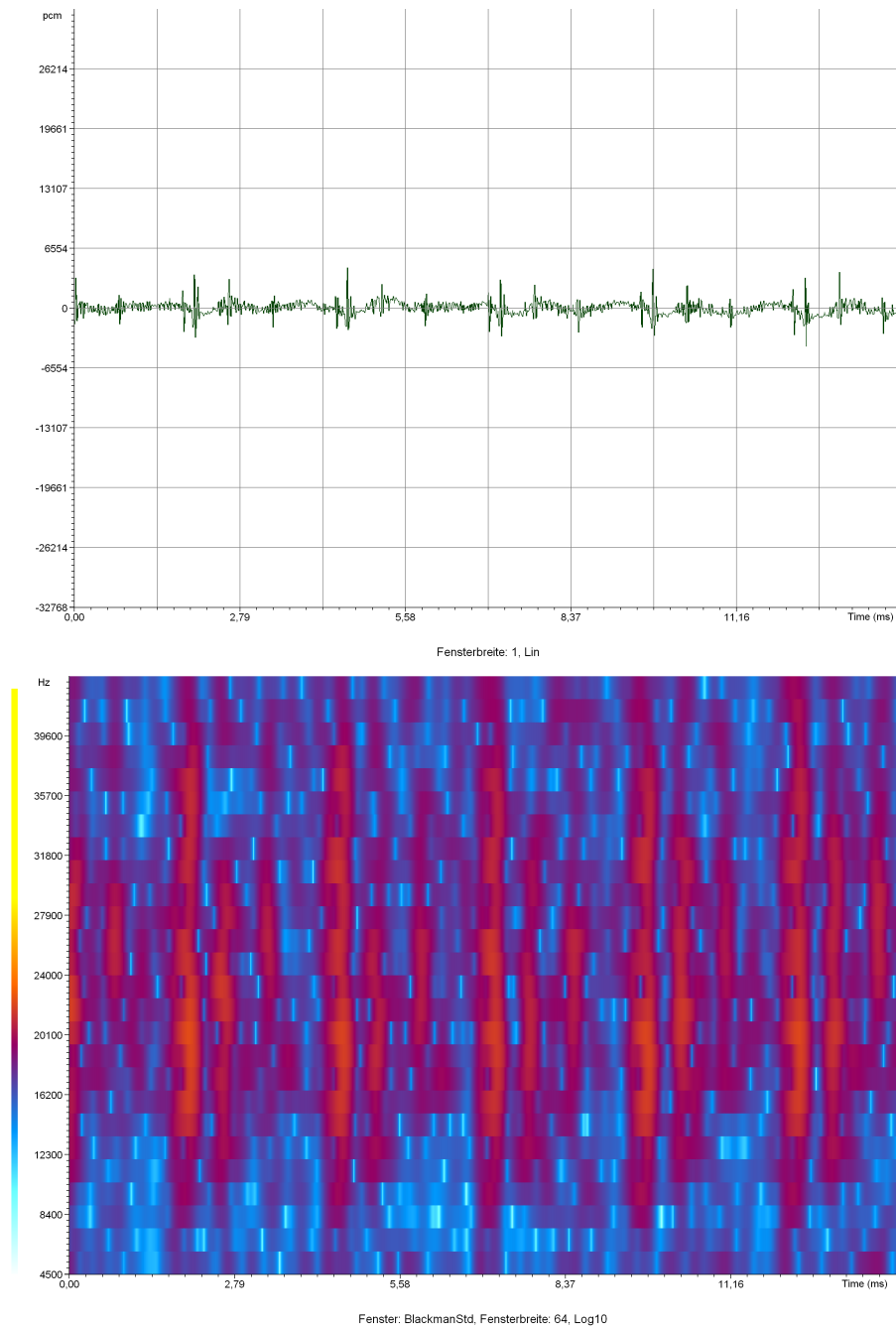


Abbildung 4.6: Oben: Pulsierter Schwingungstyp in oszillographischer Darstellung. Unten: Pulsierter Schwingungstyp in spektrographischer Darstellung.

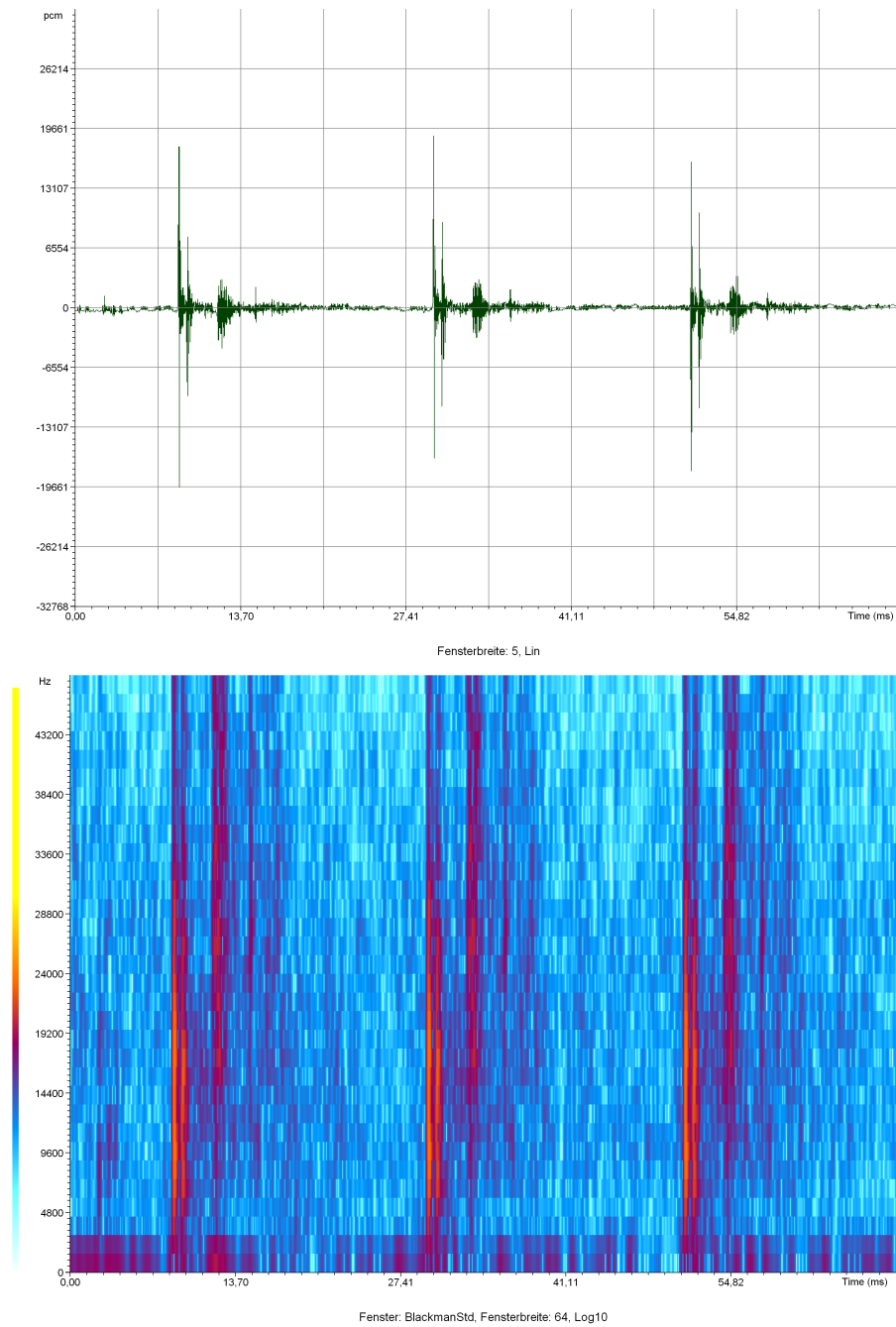


Abbildung 4.7: Oben: Klick-Schwingungstyp in oszillographischer Darstellung. Unten: Klick-Schwingungstyp in spektrographischer Darstellung.

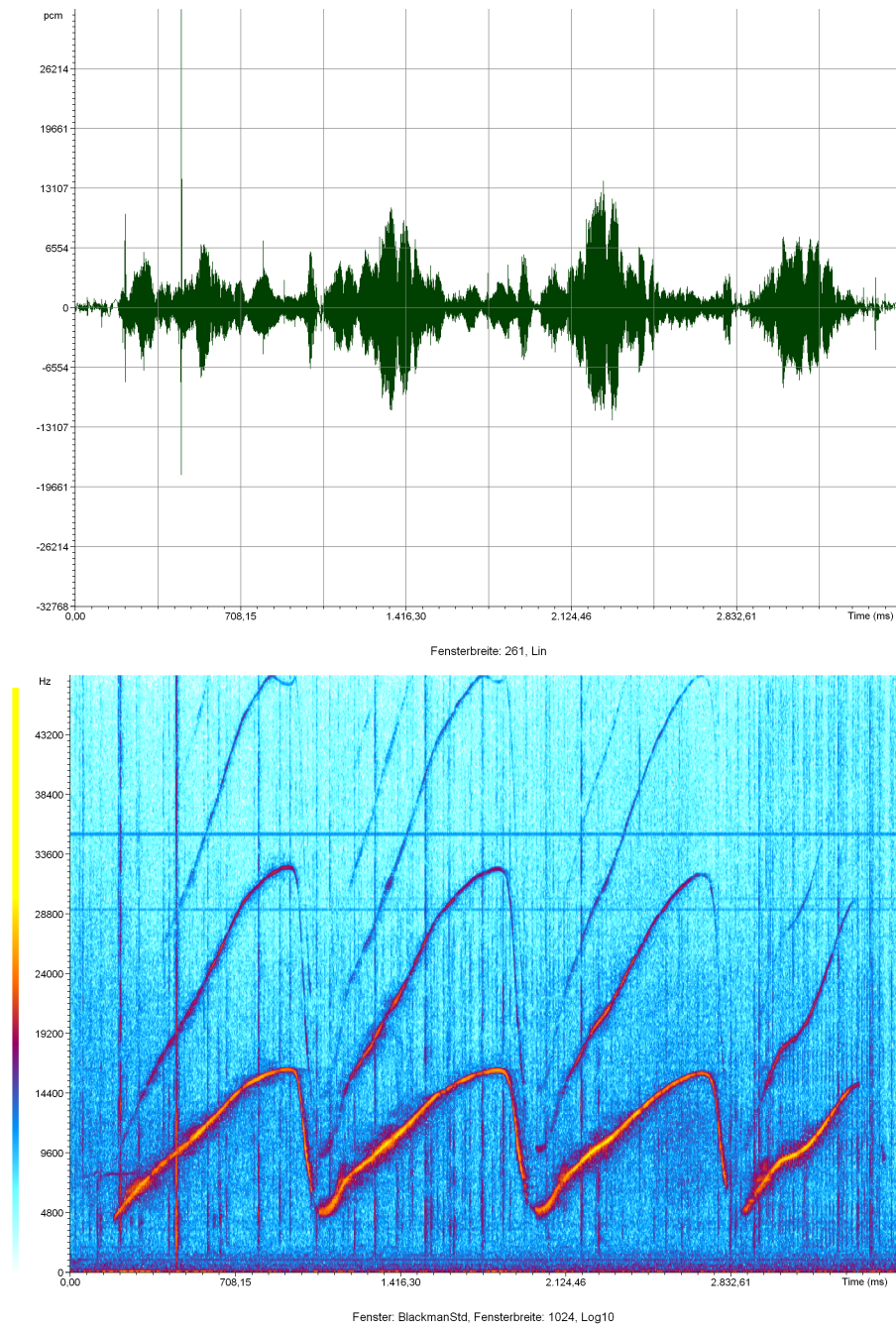


Abbildung 4.8: Oben: Oszillogramm einer Sequenz von Pffifflauten. Unten: Spektrogramm derselben Sequenz.

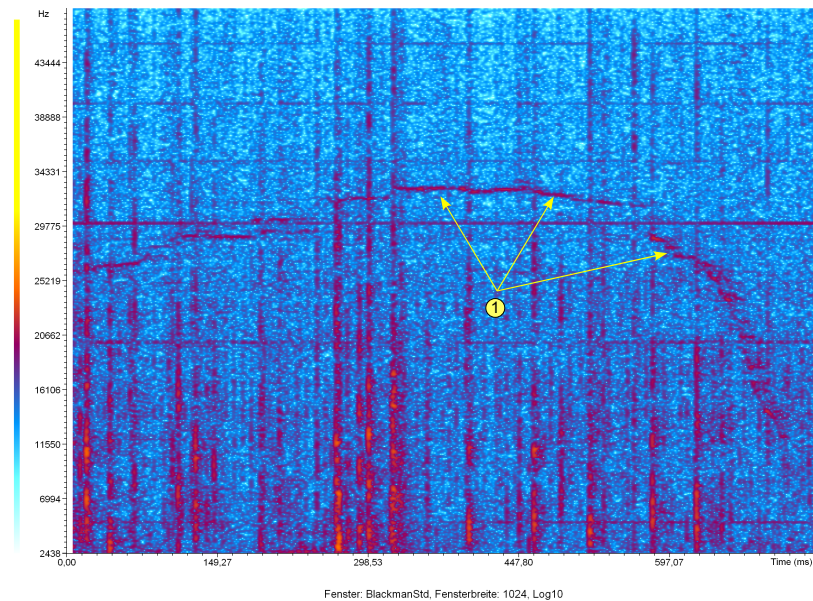


Abbildung 4.9: (1) Spektrogramm eines Pfifflauts mit max. 32,8 kHz.

Sie weisen bisweilen eine klar abgesetzte Obertonstruktur auf. Die grundlegenden Charakteristika dieses Vokalisationstypus sind in Abbildung 4.10 grafisch dargestellt.

3. **Niederfrequente pulsierte Laute:** NFP-Laute basieren auf einer pulsierten Schwingungsform. Die einzelnen Pulse treten in schneller Abfolge hintereinander auf. Die Amplitude liegt meist im unteren bis mittleren Bereich. Die dominanten Frequenzen liegen typischerweise unterhalb von 3 kHz, können aber auch höher liegen (siehe Abbildung 4.12). NFP-Laute zeichnen sich durch eine Obertonstruktur aus, die an Formanten der menschlichen Sprache erinnern kann. Sie weisen manchmal feine Modulationen auf. Die grundlegenden Charakteristika dieses Vokalisationstypus sind in Abbildung 4.11 grafisch dargestellt. NFP-Laute und Burst-Pulsed Laute (s.u.) sind nicht immer klar voneinander zu unterscheiden.
4. **Burst-Pulsed-Laute:** BP-Laute basieren auf einer pulsierten Schwingungsform. Die einzelnen Pulse treten in sehr schneller Abfolge hintereinander auf. Die Amplitude liegt im mittleren Bereich. Die spektrale Energie verteilt sich typischerweise ungleichmäßig über einen breiten Frequenzbereich von ca. 0.5-48 kHz. Die grundlegenden Charakteristika von BP-Lauten sind in Abbildung 4.13 grafisch dargestellt.
5. **Niederfrequente Klicklaute:** NF-Klicklaute basieren auf einer Klick-Schwingungsform. Die Amplitude liegt im mittleren Bereich. Die spektrale Energie ist typischerweise auf einen engen Frequenzbereich unterhalb von 3 kHz kon-

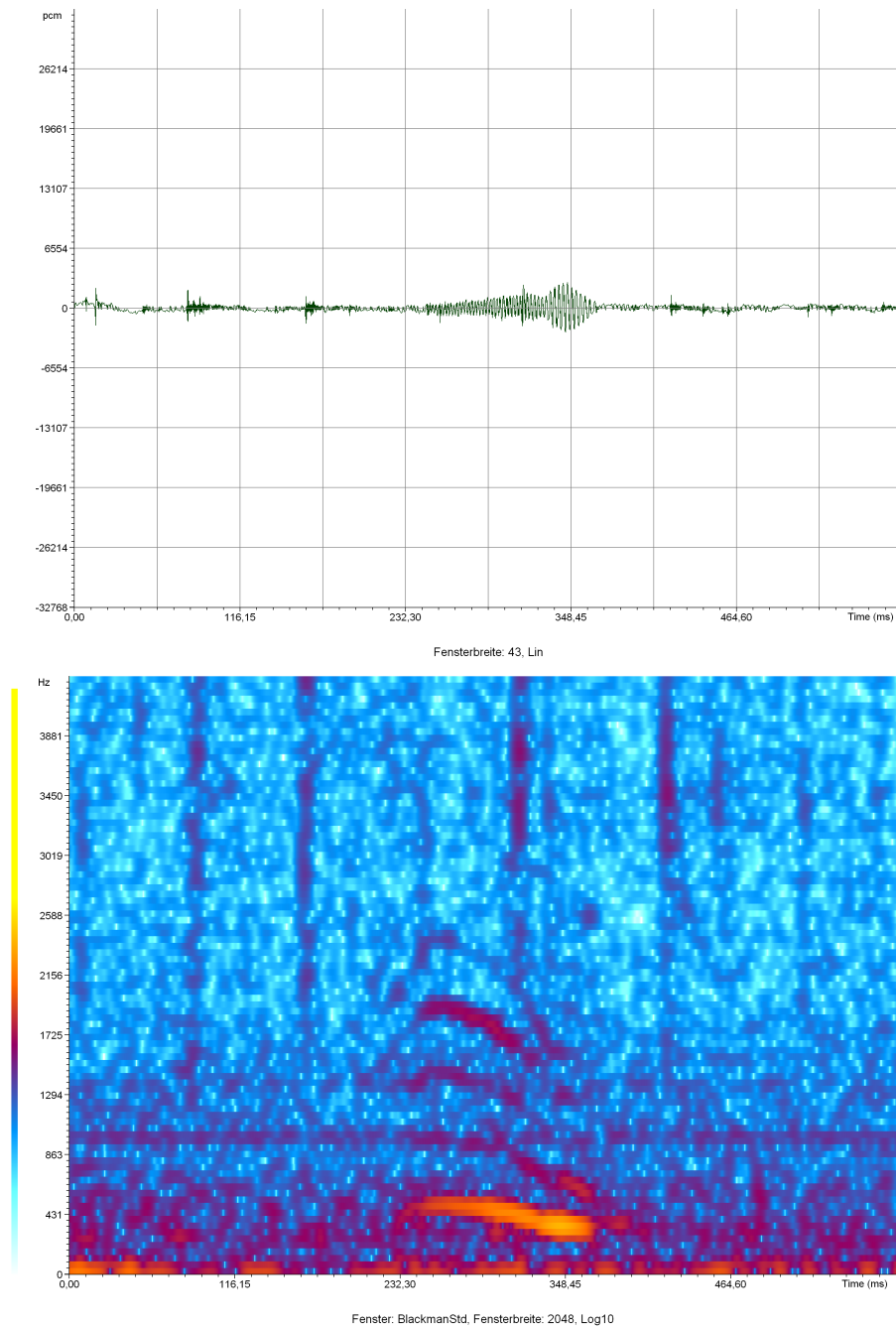


Abbildung 4.10: Oben: Oszillogramm eines niederfrequenten nichtpulsierten Lautes.
Unten: Spektrogramm desselben Lautes.

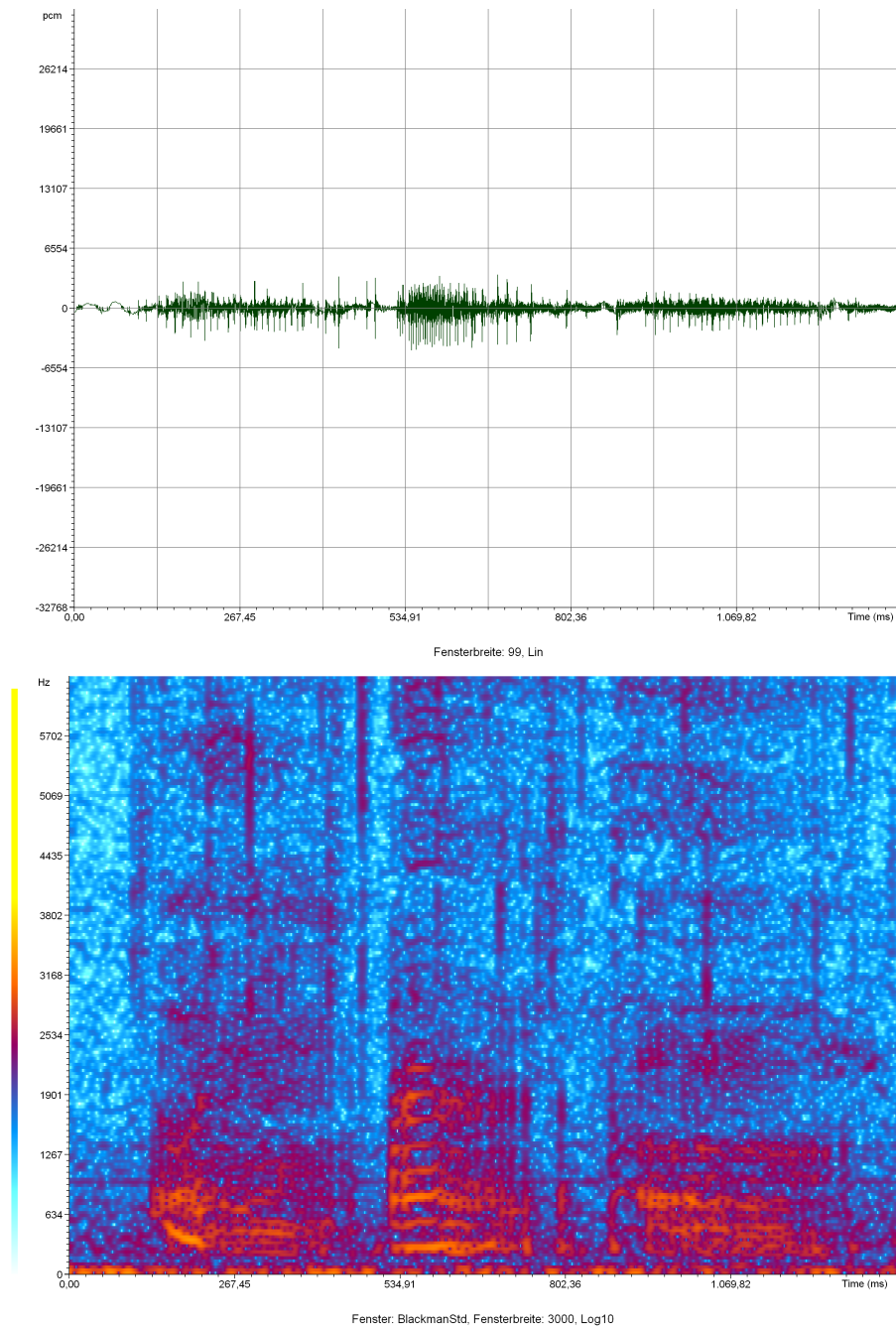


Abbildung 4.11: Oben: Oszillogramm einer Sequenz von niederfrequenten pulsierten Lauten. Unten: Spektrogramm derselben Sequenz.

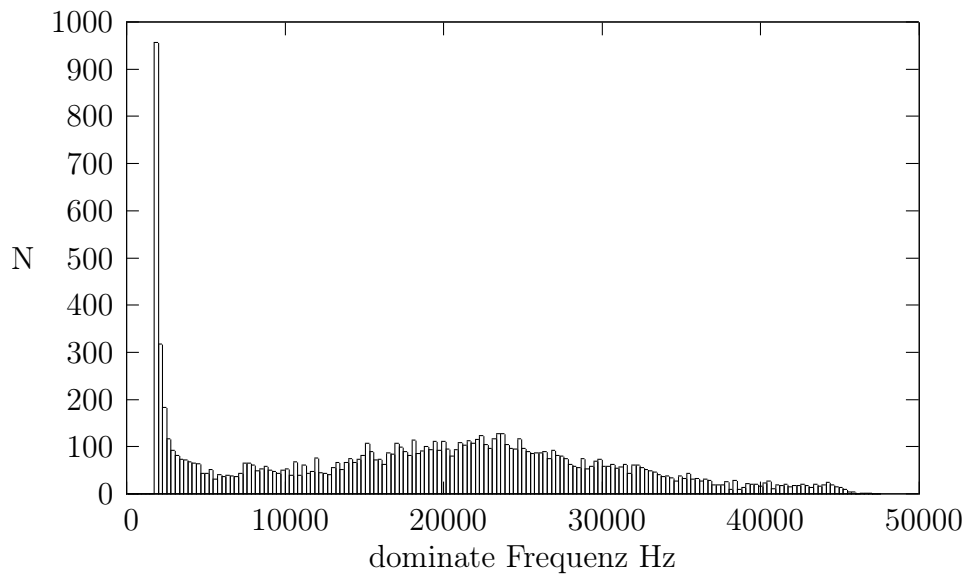


Abbildung 4.12: Gemessene dominante Frequenzen von NFP-Lauten. Die Grafik basiert auf einer Spektralanalyse von 10970 NFP-Lauten. NFP-Laute lassen sich nicht immer von Burst-Pulsed-Lauten unterscheiden.

zentriert. NF-Klicklaute treten immer in einer Abfolge von klar voneinander abgesetzten Einzelpulsen auf. Die Puls-Wiederholrate (*Pulse Repetition Rate*) ist variabel. Die grundlegenden Charakteristika dieses Vokalisationstypus sind in Abbildung 4.14 grafisch dargestellt.

6. **Mittelfrequente Klicklaute:** MF-Klicklaute basieren auf einer Klick-Schwingungsform. Die Amplitude liegt im oberen Bereich. Die spektrale Energie verteilt sich typischerweise sehr ungleichmäßig über einen breiten Frequenzbereich von ca. 5-48 kHz (und darüber). MF-Klicklaute treten fast immer in einer Sequenz von klar voneinander abgesetzten Einzelpulsen auf. Die Puls-Wiederholrate ist variabel. Die spektrale Energie der Pulse ist ungleichmäßig über den Frequenzraum verteilt. Die grundlegenden Charakteristika dieses Vokalisationstypus sind in Abbildung 7.4 grafisch dargestellt.

4.4.3 Zusammenfassung

In Abbildung 4.16 ist die Klassifikation in Schwingungstypen und Grundtypen abgebildet. Tabelle 4.1 gibt eine Übersicht über die Vokalisations-Grundtypen im Bereich von 0-48 kHz. Laute, die nicht mit dem Vokalisationstrakt erzeugt werden, sind in dieser Tabelle nicht berücksichtigt.

Die oben beschriebenen Vokalisations-Grundtypen werden von *Tursiops truncatus* auf vielfältige Art und Weise moduliert. Der nächste Abschnitt widmet sich den verschiedenen Arten der phonetischen Modulation.

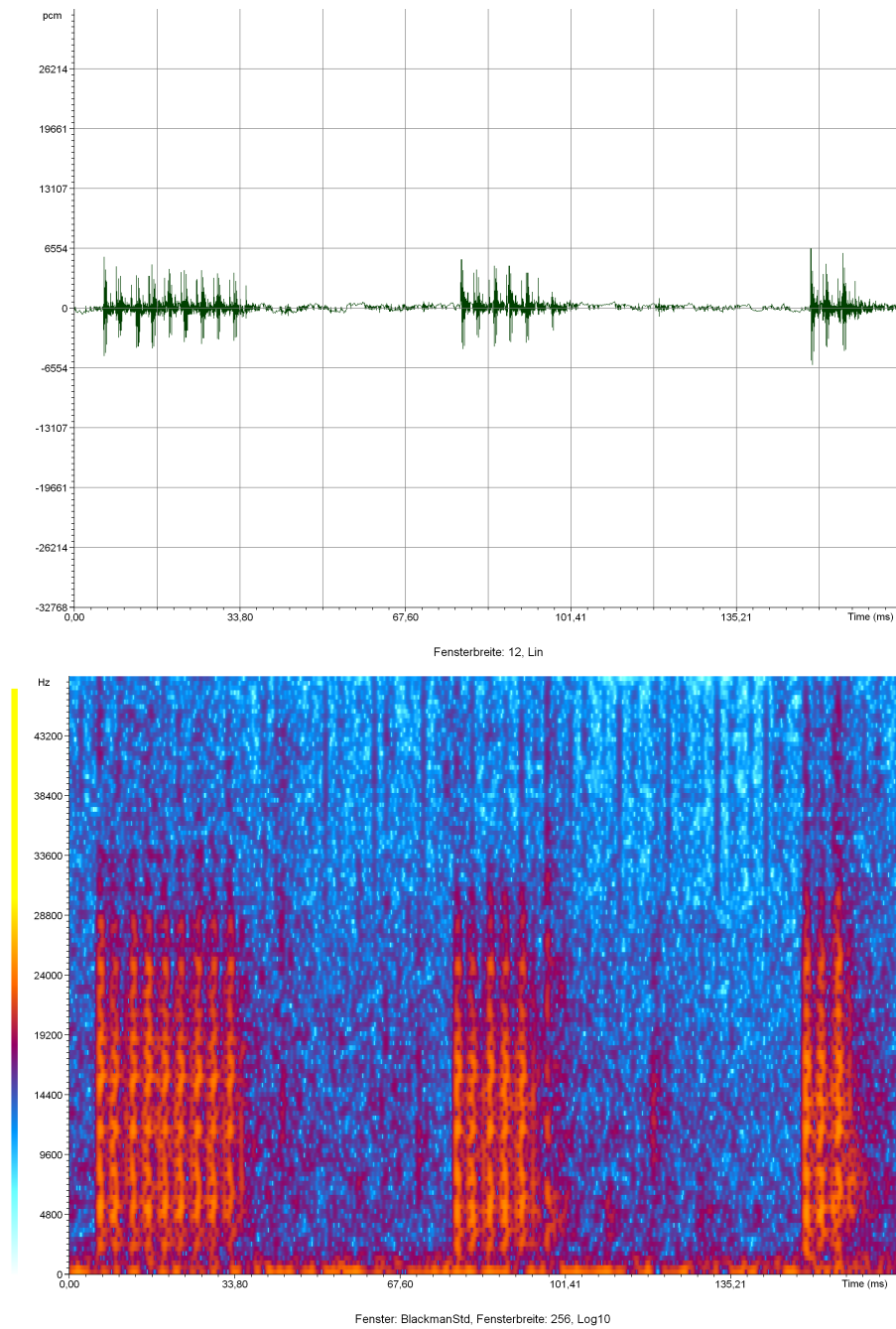


Abbildung 4.13: Oben: Oszillogramm einer Sequenz von Burst-Pulsed-Lauten. Unten: Spektrogramm derselben Sequenz.

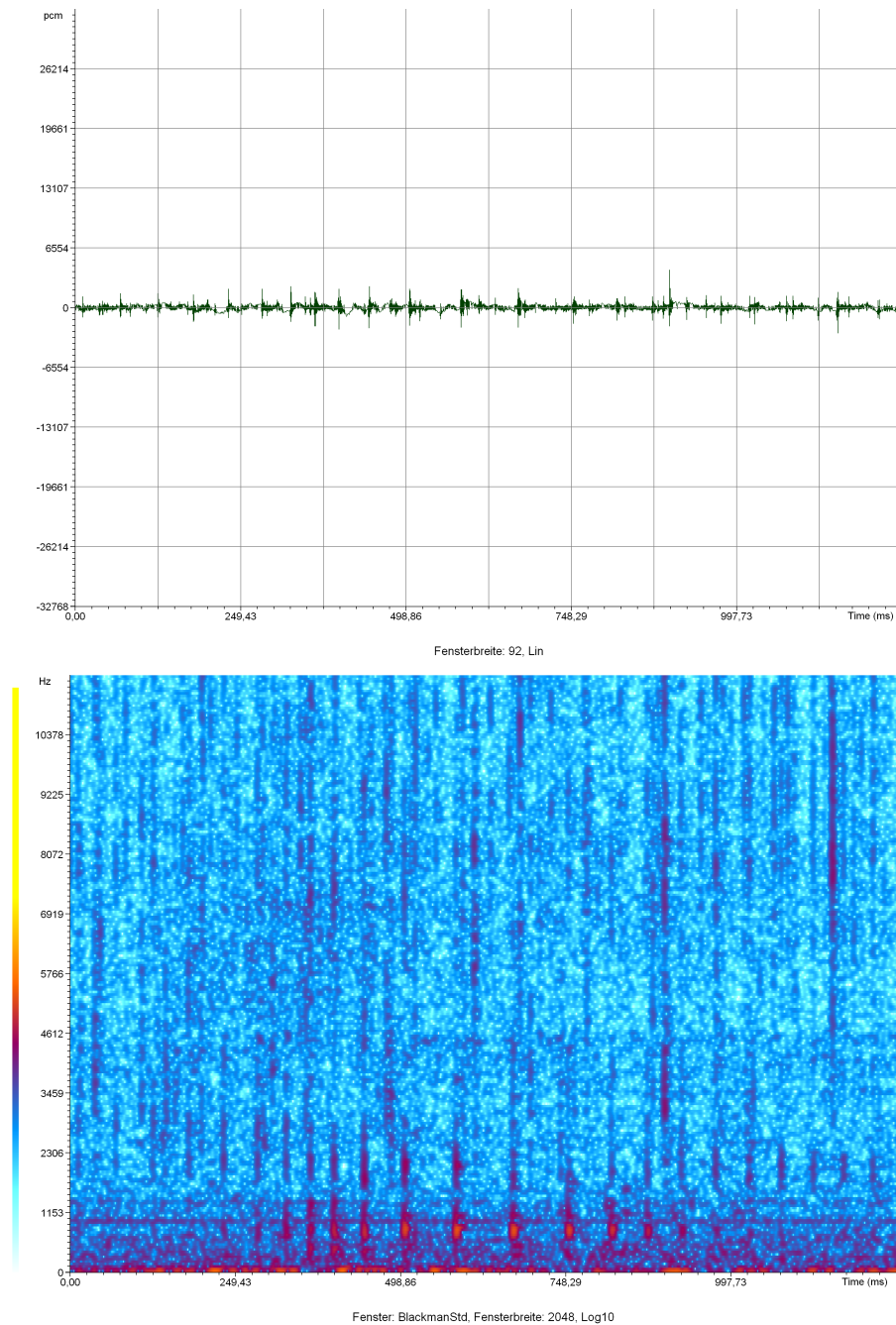


Abbildung 4.14: Oben: Oszillogramm einer Sequenz von niederfrequenten Klicklauten. Unten: Spektrogramm derselben Sequenz.

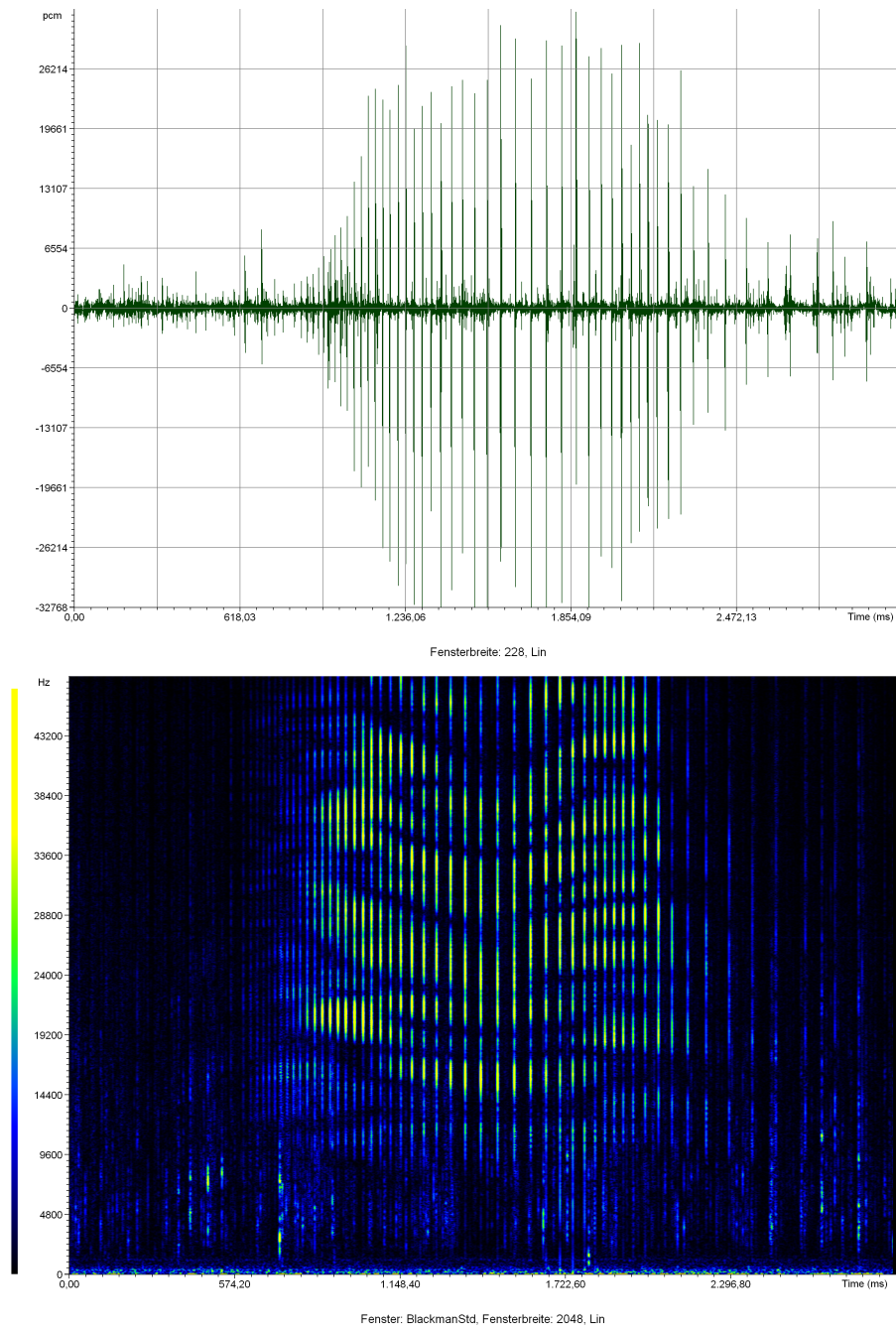


Abbildung 4.15: Oben: Oszillogramm einer Sequenz von mittelfrequenten Klicklauten. Unten: Spektrogramm derselben Sequenz.

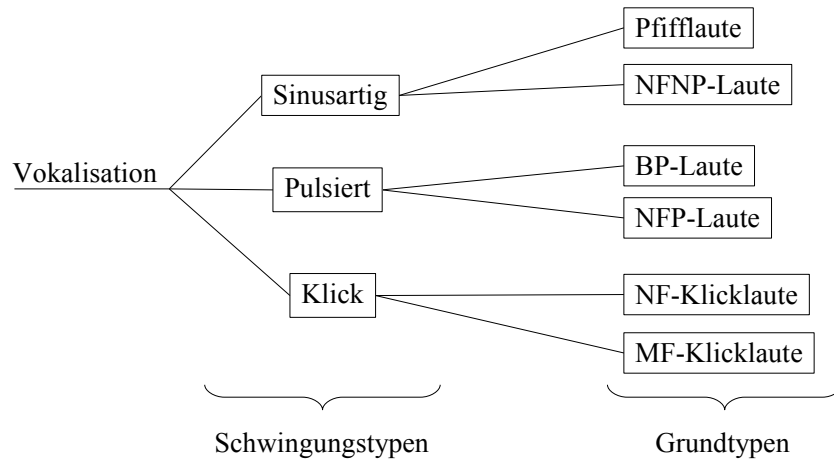


Abbildung 4.16: Klassifikation grundlegender akustischer Phänomene.

Typ	Schwingungstyp	Grundfrequenz	Dauer
Pfifflaute	sinusartig	1-32 kHz	<2000ms
NFNP-Laute	sinusartig	< 1 kHz	<60ms
NFP-Laute	pulsiert	< 3 kHz	<600ms
BP-Laute	pulsiert	0.5-48 kHz	<300ms
NF-Klicklaute	Klick	< 3 kHz	<2ms
MF-Klicklaute	Klick	5-48 kHz	<2ms

Tabelle 4.1: Vokalisations-Grundtypen im Bereich von 0-48 kHz.

4.5 Phonetische Modulation

Im vorausgehenden Abschnitt wurde eine Klassifikation in Vokalisations-Grundtypen vorgenommen. Diese treten im Eilat-Korpus in zahlreichen Variationen auf. Jeder der Grundtypen weist ein breites Spektrum an Modulationsphänomenen auf. Ziel des folgenden Abschnittes ist, eine Übersicht über diese Phänomene zu geben.

4.5.1 Allgemeine Prinzipien

Der Begriff *Modulation* wird in der Literatur uneinheitlich verwendet. Um ein Maximum an Objektivität in diesem Bereich zu erhalten, soll daher der Modulationsbegriff mit Hilfe von Definitionen konkretisiert werden, die in der Nachrichtentechnik üblich sind. In Zusammenhang mit der Übertragung elektromagnetischer Signale versteht man unter Modulation die „*planmäßige Beeinflussung eines Trägersignals durch ein Modulationssignal (= zu übertragendes Nutzsignal)*“ [40].

Die unmittelbare Verwendung der Atmosphäre als Übertragungsstrecke ist für die meisten Nutzsignale (z.B. für Audiosignale) aus physikalischen Gründen nicht möglich. Man nutzt daher ein hochfrequentes Trägersignal, welches der vorhandenen Übertragungsstrecke angepasst ist. Mit dessen Hilfe wird das Nutzsignal im Hückepackverfahren transportiert. Modulation ist also ein Verfahren der gezielten Beeinflussung von Trägersignalen zum Zwecke der Übertragung von nicht direkt übertragbaren Signalen. Abbildung 4.17 veranschaulicht diesen Zusammenhang.

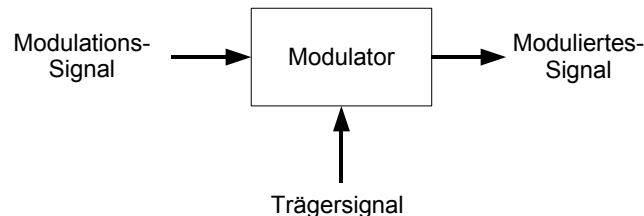


Abbildung 4.17: Modulation in der Nachrichtentechnik.

Als Trägersignale kommen in der Praxis meist Sinusschwingungen und Rechteck-Pulse zur Anwendung, da diese vergleichsweise einfach erzeugt werden können. Man unterscheidet dementsprechend zwischen Verfahren der *Schwingungsmodulation* und Verfahren der *Pulsmodulation*. Hierbei wird unter dem Begriff Schwingungsmodulation die Modulation eines sinusförmigen Trägersignals und unter dem Begriff Pulsmodulation die Modulation eines pulsförmigen Trägersignals verstanden.

Bei der Schwingungsmodulation ist das Trägersignal definiert durch:

$$u_T(t) = \hat{u}_T \cdot \sin(\omega_T t \pm \varphi_T) \quad (4.1)$$

Das Modulationssignal ist definiert durch:

$$u_M(t) = \hat{u}_M \cdot \sin(\omega_M t \pm \varphi_T) \quad (4.2)$$

Im Bereich der Schwingungsmodulation sind als modulierbare Parameter die Amplitude u_T , die Frequenz ω_T und der Phasenwinkel φ_T des Signals denkbar. Entsprechend dieser Parameter wird zwischen Amplitudenmodulation (AM), Frequenzmodulation (FM) und Phasenmodulation (PM) unterschieden.

- **Amplitudenmodulation:** Bei der AM wird die Amplitude des Trägersignals durch das Modulationssignal verändert. Phasenwinkel und Frequenz des Signals spielen eine untergeordnete Rolle:

$$u_{AM}(t) = k \cdot u_M(t) \cdot \sin(\omega_T t \pm \varphi_T) \quad (4.3)$$

- **Frequenzmodulation:** Bei der FM wird die Frequenz des Trägersignals durch das Modulationssignal verändert. Phasenwinkel und Amplitude des Signals spielen eine untergeordnete Rolle:

$$u_{FM}(t) = \hat{u}_T \cdot \sin(k \cdot u_M(t) \cdot t \pm \varphi_T) \quad (4.4)$$

- **Phasenmodulation:** Bei der PM wird der Phasenwinkel des Trägersignals vom Modulationssignal gesteuert. Frequenz und Amplitude des Signals spielen eine untergeordnete Rolle:

$$u_{PM}(t) = \hat{u}_T \cdot \sin(\omega_T t \pm k \cdot u_M(t)) \quad (4.5)$$

Im Bereich der Pulsmodulation sind als modulierbare Parameter die Pulsamplitude, die Pulsfrequenz, die Pulsphase und die Pulsdauer denkbar. Dementsprechend wird zwischen Pulsamplitudenmodulation (PAM), Pulsfrequenzmodulation (PFM), Pulsphasenmodulation (PPM) und Pulsdauermodulation (PDM) unterschieden.

- **Pulsamplitudenmodulation:** Bei der PAM wird die Amplitude der Pulse vom Modulationssignal gesteuert.
- **Pulsfrequenzmodulation:** Bei der PFM wird die Pulsfrequenz (d.h. die Anzahl der Pulse pro Zeiteinheit) im Trägersignal durch das Modulationssignal gesteuert.
- **Pulsphasenmodulation:** Bei der PPM wird durch das Modulationssignal die Lage der Pulse in Bezug auf konstante Zeitintervalle gesteuert.
- **Pulsdauermodulation:** Bei der PDM wird die Dauer der Pulse durch das Modulationssignal gesteuert.

In der Bioakustik ist eine Unterscheidung von Träger- und Modulationssignal scheinbar nicht sinnvoll, da bioakustische Signale direkt übertragen werden. Dennoch ist es erforderlich, die verschiedenen Phänomene der phonetischen Modulation bioakustischer Signale exakt zu charakterisieren. Dies wird möglich, indem das

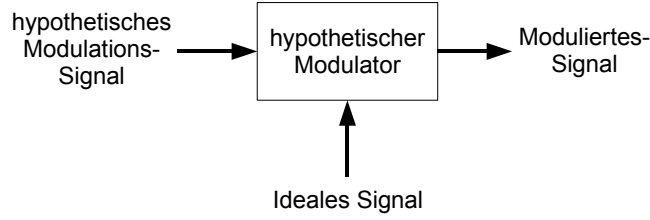


Abbildung 4.18: Phonetische Modulation in der Bioakustik

Konzept des Trägersignals durch das Konzept eines *idealen Signals* ersetzt wird [55]. Dieser Gedanke wird in Abbildung 4.18 veranschaulicht.

Phonetische Modulation kann als Beeinflussung eines hypothetischen idealen Signals durch ein hypothetisches Modulationssignal verstanden werden. Diese Sichtweise ermöglicht eine neutrale Beschreibung von Modulationsphänomenen in der Bioakustik.

Da *Tursiops truncatus* und viele andere Tierarten sinusartige und/oder pulsierte Signale produzieren, ist es zweckmäßig, auch hier zwischen Schwingungsmodulation und Pulsmodulation zu unterscheiden.

Die nachrichtentechnische Auffassung von Schwingungsmodulation geht von sinusförmigen Trägersignalen aus. Die obigen Formeln lassen sich daher fast ohne Änderung auch zur Beschreibung der Schwingungsmodulation in der Bioakustik verwenden.

Das hypothetische ideale sinusförmige Signal ist definiert durch:

$$u_I(t) = \hat{u}_I \cdot \sin(\omega_I t \pm \varphi_I) \quad (4.6)$$

Das hypothetische Modulationssignal ist definiert durch:

$$u_M(t) = \hat{u}_M \cdot \sin(\omega_M t \pm \varphi_T) \quad (4.7)$$

Es ergeben sich drei mögliche Prinzipien Schwingungsmodulation in der Bioakustik zu charakterisieren:

- **Amplitudenmodulation:** Bei der AM wird die Amplitude des idealen Signals durch ein hypothetisches Modulationssignal verändert:

$$u_{AM}(t) = k \cdot u_M(t) \cdot \sin(\omega_I t \pm \varphi_I) \quad (4.8)$$

- **Frequenzmodulation:** Bei der FM wird die Frequenz des idealen Signals durch ein hypothetisches Modulationssignal verändert:

$$u_{FM}(t) = \hat{u}_I \cdot \sin(k \cdot u_M(t) \cdot t \pm \varphi_I) \quad (4.9)$$

- **Phasenmodulation:** Bei der PM wird der Phasenwinkel des idealen Signals durch ein hypothetisches Modulationssignal verändert:

$$u_{PM}(t) = \hat{u}_I \cdot \sin(\omega_I t \pm k \cdot u_M(t)) \quad (4.10)$$

Das ideale diskrete pulsierte Signal ist ein Signal mit konstanter Pulsfrequenz, minimaler Pulsbreite und homogenem Spektrum. Eine mögliche Definition ist

$$u_I(t) = \begin{cases} \hat{u}_I & \text{if } (t \bmod k = 0) \\ 0 & \text{sonst} \end{cases} \quad (4.11)$$

Der konstante Parameter k definiert hierbei die Pulsfrequenz (*Pulse-Repetition-Rate*). Abbildung 4.19 zeigt ein FFT-Spektrogramm eines idealen pulsierten Signals.

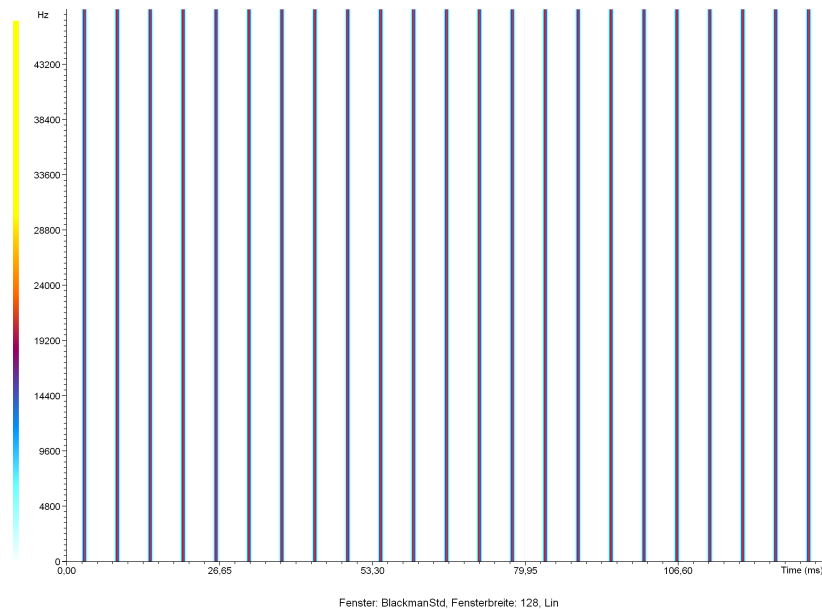


Abbildung 4.19: Spektrogramm eines idealen diskreten pulsierten Signals.

Legt man ein ideales pulsiertes Signal zugrunde, so ergeben sich vier mögliche Prinzipien Pulsmodulation in der Bioakustik zu charakterisieren:

- **Pulsamplitudenmodulation:** PAM bezeichnet die Änderung der Amplitude eines idealen pulsförmigen Signals durch ein hypothetisches Modulationssignal.
- **Pulsfrequenzmodulation:** PFM bezeichnet die Änderung der Pulsfrequenz eines idealen pulsförmigen Signals durch ein hypothetisches Modulationssignal.
- **Pulsphasenmodulation:** PPM bezeichnet die durch ein hypothetisches Modulationssignal verursachte Änderung der Lage der Pulse eines idealen pulsförmigen Signals in Bezug auf konstante Zeitintervalle.

- **Pulsdauermodulation:** PDM bezeichnet die Änderung der Dauer der Pulse eines idealen pulsförmigen Signals durch ein hypothetisches Modulationssignal.

In der Bioakustik kommen ideale pulsierte Signale ebenso wenig vor, wie ideale sinusförmige. Es soll an dieser Stelle daher noch ein weiteres Modulationsprinzip genannt werden, welches dem Umstand Rechnung trägt, dass die spektrale Energie in den pulsierten Lauten und Klicklauten von *Tursiops truncatus* immer inhomogen verteilt ist. Es handelt sich um die Pulsspektrum-Modulation (PSM).

- **Pulsspektrum-Modulation:** PSM bezeichnet die Änderung des Spektrums eines idealen pulsförmigen Signals durch ein hypothetisches Modulationssignal.

Wie die verschiedenen Arten der phonetischen Modulation in den Vokalisationen des Großen Tümmlers in Erscheinung treten, wird in Abschnitt 4.5.4 dargestellt. Da jedoch nicht jede Abweichung vom Idealtyp mit einer biologisch relevanten Modulation gleichzusetzen ist, muss zuvor eine Unterscheidung von beiläufiger und biologisch relevanter kodierender Modulation getroffen werden.

4.5.2 Modulation und Kodierung

Bioakustische Signale, so wie sie von uns aufgezeichnet werden, entsprechen niemals idealen Signaltypen. Sie weichen immer von der idealen Sinusschwingung oder – im Falle der Klicklaute der Delphine – von der idealen Pulsfunktion ab. Nach den obigen Definitionen ist Modulation aber als Abweichung von Idealtypen definiert. Es folgt, dass bioakustische Signale immer moduliert sind. Die Abweichung aufgezeichneter bioakustischer Signale von den zwei oben beschriebenen Idealtypen kann dabei folgende Ursachen haben:

1. Zufällige Fluktuationen während der Signalproduktion;
2. Durch morphologische Strukturen eingeschränkte Freiheitsgrade bei der Signalproduktion;
3. Akustische Reflexionen im Körpergewebe oder in der Umwelt;
4. Noise, Meßfehler und Meßungenauigkeiten;
5. Kodierung biologisch relevanter Informationen in den Signalen;

Eine wichtige Aufgabe der Bioakustik besteht darin, herauszufinden, welche biologisch relevanten Informationen in Signalen enthalten sind und wie diese kodiert werden. Bevor man diesen Schritt tun kann, muss geklärt werden, nach welchen Prinzipien biologisch relevante Informationen in den Signalen kodiert werden. Hierzu muss untersucht werden, in welcher Art und Weise die Signale von bestimmten Idealtypen abweichen und welche der Abweichungen **nicht** auf morphologische Strukturen, akustische Reflexionen, Meßfehler usw. zurückzuführen sind. Um diese Unterscheidung zu vereinheitlichen, soll im Folgenden zwischen beiläufiger Modulation und kodierender Modulation unterschieden werden:

- **Beiläufige Modulation:** Unter dem Begriff beiläufige Modulation werden alle Abweichungen von den Idealtypen zusammengefasst, die durch zufällige Fluktuationen in den Signalen, morphologische Randbedingungen, akustische Reflexionen, Noise, Meßfehler und Meßungenauigkeiten verursacht werden und keinerlei biologisch relevante Informationen kodieren.
- **Kodierende Modulation:** Unter dem Begriff kodierende Modulation werden alle Abweichungen von den Idealtypen zusammengefasst, die biologisch relevante Informationen kodieren.

Diese Unterscheidung wird in Abbildung 4.20 veranschaulicht. Modulationsprinzipien, die sich in den Vokalisationen des Großen Tümmlers zuverlässig beobachten lassen, sind grün umrandet.

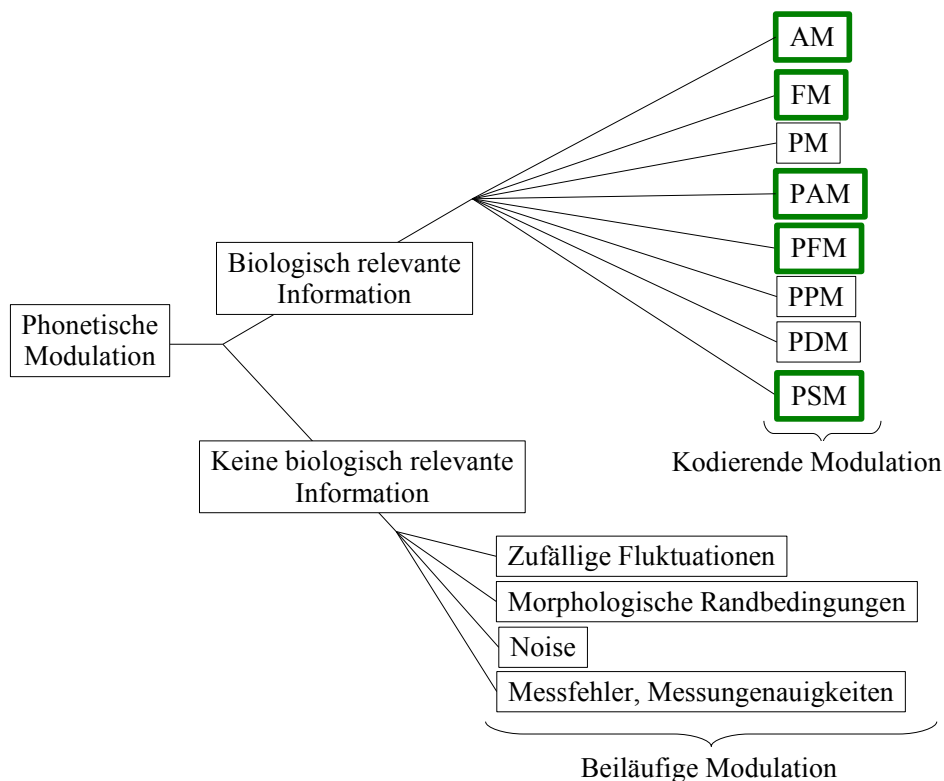


Abbildung 4.20: Allgemeine Prinzipien der phonetischen Modulation in der Bioakustik. Grün umrandet sind die Modulationsprinzipien, welche beim Großen Tümmler beobachtet werden können.

In vielen Fällen ist das Vorliegen kodierender Modulationen offensichtlich. In anderen Fällen können nur detaillierte Studien Klarheit darüber schaffen, ob ein Modulationsphänomen auf eine Kodierung biologisch relevanter Informationen zurückzuführen ist, oder ob es sich um beiläufige Modulation handelt, denn diese kann kodierender Modulation täuschend ähnlich sein.

4.5.3 Spezielle Prinzipien

Die oben dargestellten Prinzipien der phonetischen Modulation beschreiben grundlegende Möglichkeiten, akustische Signale zu modulieren. Neben diesen elementaren Prinzipien existieren im Tierreich komplexere Formen der phonetischen Modulation. Einige dieser speziellen Modulationsarten werden in diesem Abschnitt dargestellt. Die Auswahl beschränkt sich auf Prinzipien, die für die Spezies *Tursiops truncatus* von Bedeutung sind. Hierfür bedarf es zunächst der Konkretisierung einiger in der Bioakustik üblicher Begriffe. Die Bezeichnungen orientieren sich an Tembrock (1977) [109].

- **Ton:** Unter dem Begriff *Ton* wird eine akustische Schwingung in Sinusform mit genau einer Frequenz verstanden. Die Tonhöhe entspricht dabei der Frequenz.
- **Komplexer Ton:** Unter dem Begriff *komplexer Ton* wird eine akustische Schwingung verstanden, die zugleich mehrere Frequenzen enthält.
- **Grundton:** Unter dem Begriff *Grundton* wird die Komponente eines komplexen Tons verstanden, welche die tiefste Frequenz aufweist.
- **Oberton:** Unter dem Begriff *Oberton* wird eine Komponente eines komplexen Tons verstanden, welche über dem Grundton liegt.
- **Harmonische:** Unter dem Begriff *Harmonische* wird eine Komponente eines komplexen Tons verstanden, deren Frequenz in einem ganzzahligen Verhältnis zum Grundton steht.

Für die Spezies *Tursiops truncatus* lassen sich eine Reihe von speziellen Prinzipien der phonetischen Modulation benennen, welche in der folgenden Auflistung beschrieben werden. Die Begriffe *Glissando*, *Formanten* und *Vibrato* sind dem Werk von Tembrock (1977) entlehnt. Der Begriff *Seitenband* ist eine Übersetzung aus dem Englischen. Die Begriffe *Pulsfrequenz-Glissando*, *Frequenzsprünge*, *Zischen* und *Pulsspektrum-Drift* sind Definitionen des Autors.

- **Glissando:** Unter dem Begriff *Glissando* wird eine kontinuierliche Tonhöhenänderung bei tonalen Lauten verstanden. Diese kann als Form der FM verstanden werden.
- **Formanten:** Unter dem Begriff *Formanten* wird ein Resonanzmuster verstanden, in dem Harmonische mit Amplitudenmaximum existieren. Es handelt sich um eine Mischform aus AM und FM.
- **Vibrato:** Unter dem Begriff *Vibrato* werden periodische Fluktuationen der Frequenz bei tonalen Lauten verstanden. Es handelt sich um eine Form der FM.

- **Seitenband:** Unter dem Begriff *Seitenband* wird das lokal begrenzte Auftreten einer Formantenstruktur bei Pfifflauten verstanden. Es handelt sich um eine Mischform aus PFM und FM.
- **Frequenzsprünge:** Unter dem Begriff *Frequenzsprünge* wird das Auftreten von sprunghaften Änderungen der Frequenz in Pfifflauten verstanden. Es handelt sich um eine Form der FM.
- **Zischen:** Unter dem Begriff *Zischen* wird das Auftreten von Zischlauten in Pfiffen verstanden. Zischlaute ähneln im Spektrogramm dem Rauschen. Sie können im Pfiff lokal begrenzt sein oder sich über ganze Pfifflaute hinweg ausdehnen. Es handelt sich um eine Form der FM.
- **Pulsfrequenz-Glissando:** Unter dem Begriff *Pulsfrequenz-Glissando* wird eine kontinuierliche Änderung der Pulsfrequenz von Klicklauten oder pulsierenden Lauten verstanden. Ist die Pulsfrequenz hoch genug entsteht der akustische Eindruck eines tonalen Lautes. Im Spektrogramm entsteht gleichzeitig bei entsprechender Einstellung der Spektrogramm-Parameter die Illusion einer dichten Formantenstruktur. Pulsfrequenz-Glissando kann als Form der PFM verstanden werden.
- **Pulsspektrum-Drift:** Unter dem Begriff *Pulsspektrum-Drift* wird eine kontinuierliche Änderung des Spektrums von Klicklauten oder Einzelpulsen in Sequenzen pulsierter Laute oder Klicklaute verstanden. Die Änderung ist unabhängig von der Pulsfrequenz. Das Phänomen kann als Form der PSM verstanden werden.

4.5.4 Phonetische Modulation bei *Tursiops truncatus*

In den folgenden Abschnitten wird gezeigt, wie die allgemeinen und speziellen phonetischen Modulationsprinzipien in den Vokalisationen des Großen Tümmlers in Erscheinung treten. Die Darstellung orientiert sich an den Grundtypen aus Abschnitt 4.4.2. An dieser Stelle wird keine Einteilung in beiläufige und kodierende Modulation vorgenommen.

Phonetische Modulation in Pfifflauten:

Die vorwiegend sinusartige Schwingungsform der Pfifflaute (PL) wird in vielfältiger Art und Weise moduliert. Zu beobachten sind die folgenden Formen der phonetischen Modulation:

1. **Glissando:** Alle Pfifflaute weisen dieses Modulationsprinzip auf. Abbildung 4.21 zeigt ein Beispiel.

2. **Harmonische:** Pfiffaute weisen dieses Modulationsprinzip häufig auf. Abbildung 4.21 zeigt ein Beispiel. Abbildung 4.22 zeigt ein Beispiel für einen Pfiff laut ohne Harmonische.
3. **Vibrato:** Pfiffaute weisen dieses Modulationsprinzip bisweilen auf. Abbildung 4.23 zeigt ein Beispiel.
4. **Seitenband:** Pfiffaute weisen dieses Modulationsprinzip bisweilen auf. Abbildung 4.24 zeigt ein Beispiel.
5. **Zischen:** Pfiffaute weisen dieses Modulationsprinzip bisweilen auf. Abbildung 4.25 zeigt ein Beispiel.
6. **Frequenzsprünge:** Pfiffaute weisen dieses Modulationsprinzip häufig auf. Abbildung 4.26 zeigt ein Beispiel.

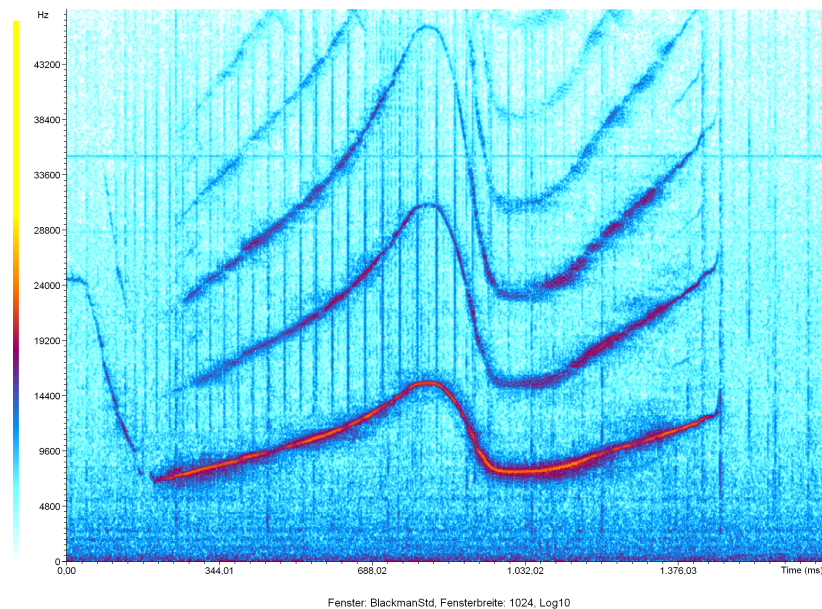


Abbildung 4.21: Glissando und Harmonische in einem Pfiff laut.

Phonetische Modulation in niederfrequenten nichtpulsierten Lauten:

Niederfrequente nichtpulierte Laute (NFNP) weisen die folgenden Formen der phonetischen Modulation auf:

1. **Glissando:** Alle niederfrequenten nichtpulsierten Laute weisen dieses Modulationsprinzip auf. Abbildung 4.27 zeigt ein Beispiel.
2. **Formanten:** Niederfrequente nichtpulierte Laute weisen dieses Modulationsprinzip gelegentlich auf. Abbildung 4.27 zeigt ein Beispiel. Abbildung 4.28 zeigt ein Beispiel für einen Laut ohne Formanten.

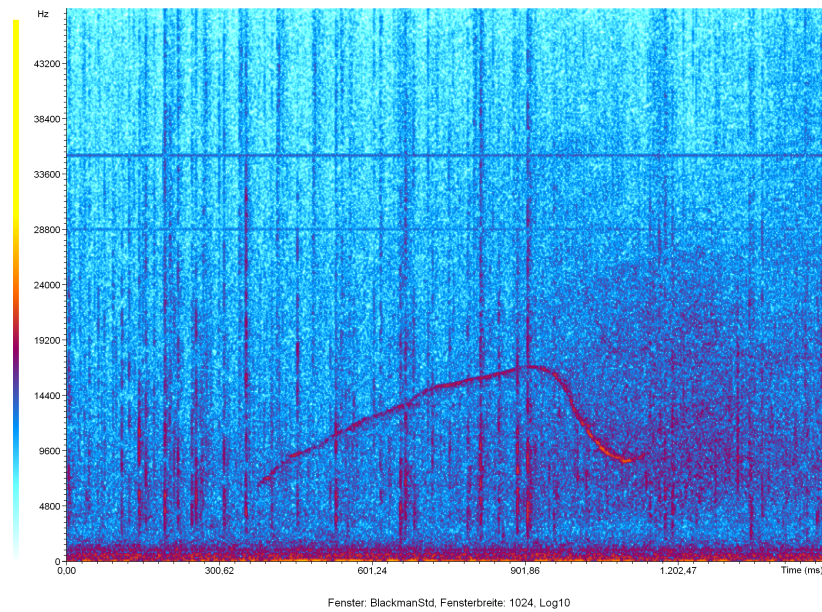


Abbildung 4.22: Pfifflaut ohne Harmonische.

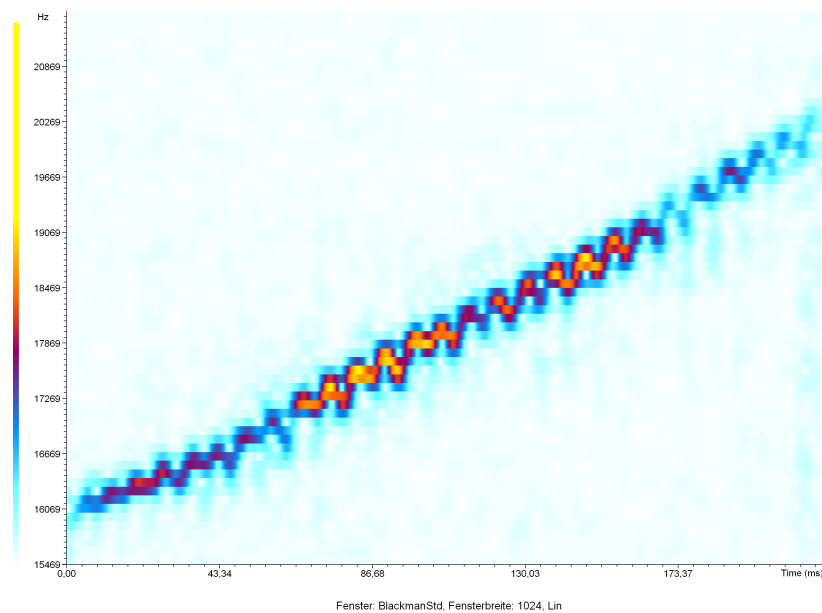


Abbildung 4.23: Vibrato in einem Pfifflaut (Detail).

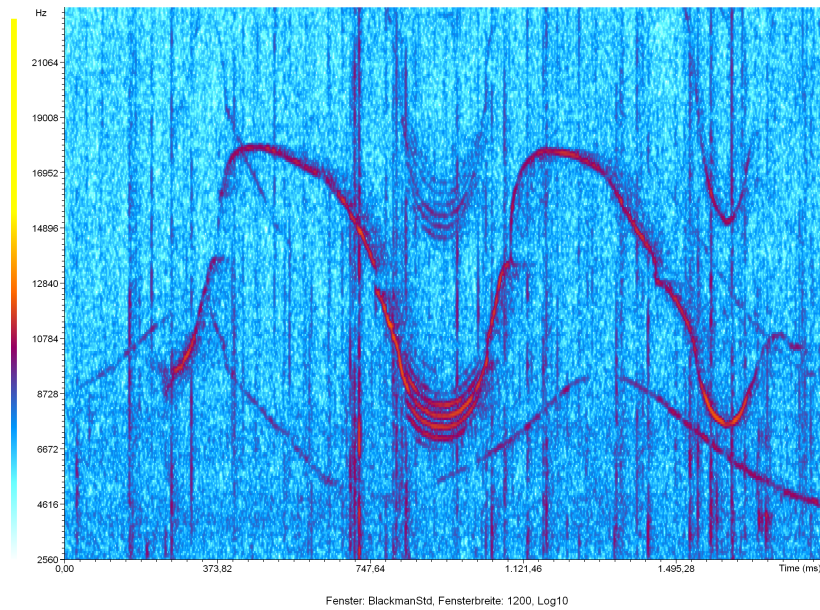


Abbildung 4.24: Seitenband in einem Piffflaut.

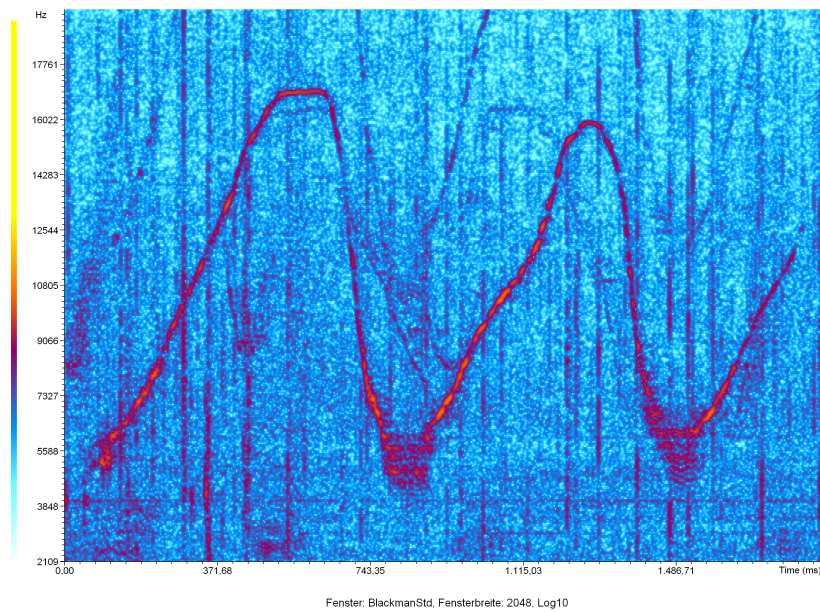


Abbildung 4.25: Zischen in einem Piffflaut.

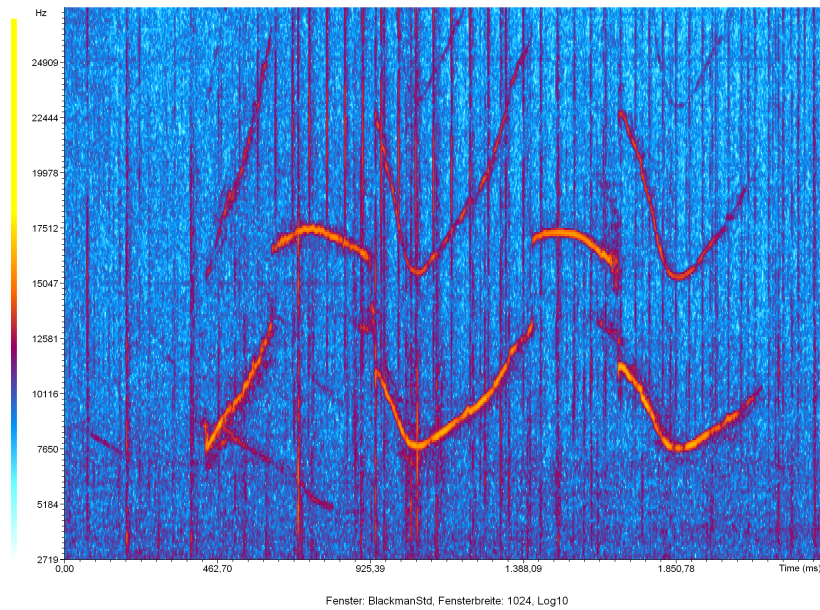


Abbildung 4.26: Frequenzsprünge in einem Pfiff laut.

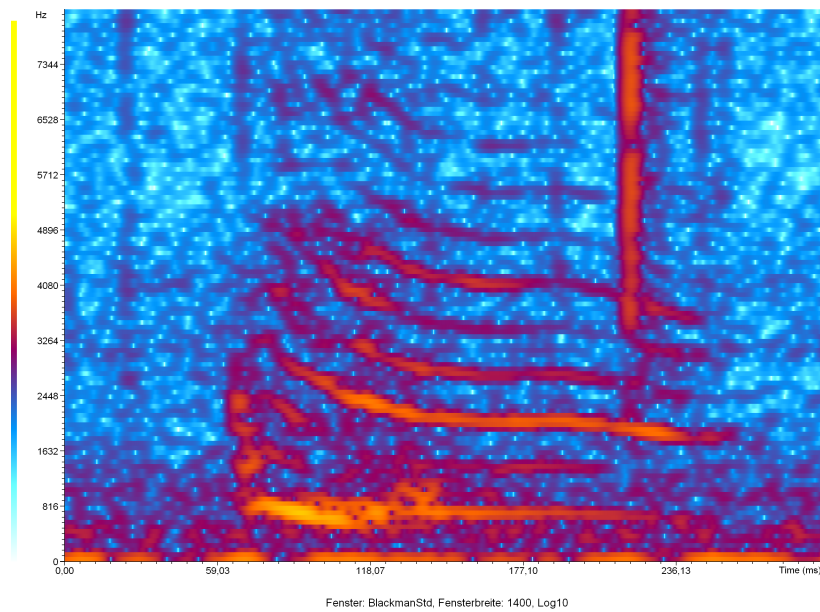


Abbildung 4.27: NFNP-Laut mit Formanten.

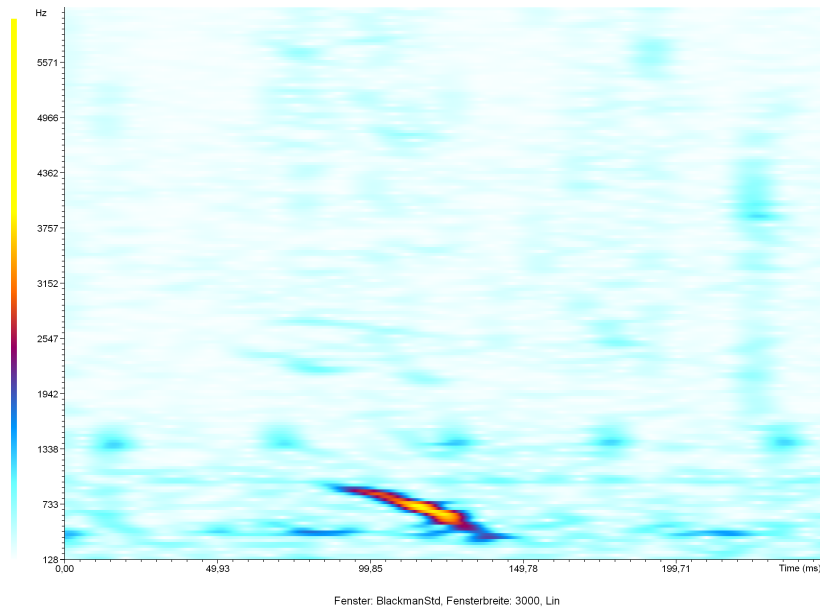


Abbildung 4.28: NFNP-Laut ohne Formanten.

Phonetische Modulation niederfrequenter pulsierter Laute:

Die vorwiegend pulsierte Schwingungsform der niederfrequenten pulsierten Laute (NFP) ist in vielfältiger Art und Weise moduliert. Es entstehen sehr unterschiedliche Geräusche mit mehr oder weniger gut sichtbarer Formantenstruktur. Die Abbildungen 4.30 und 4.29 zeigen Sequenzen von niederfrequenten pulsierten Lauten. Niederfrequente pulsierte Laute weisen die folgenden Formen der phonetischen Modulation auf:

1. **Formanten:** Niederfrequente pulsierte Laute weisen dieses Modulationsprinzip oft gut sichtbar auf – siehe Abbildung 4.30.
2. **Pulsfrequenz-Glissando:** Alle niederfrequenten pulsierten Laute weisen dieses Modulationsprinzip auf. Sie stehen dadurch im Gegensatz zu den Burst-Pulsed-Lauten. Im Spektrogramm ist dieses Modulationsprinzip kaum von einem nichtpulsierten Glissando zu unterscheiden – siehe Abbildung 4.31.
3. **PFM, PAM und PSM:** Alle niederfrequenten pulsierten Laute weisen diese Modulationsprinzipien auf.

Phonetische Modulation in Burst-Pulsed-Lauten:

Burst-Pulsed (BP) Laute sind unterschiedlich lange Sequenzen von Einzelpulsen. Diese wiederum können in Sequenzen angeordnet sein, wodurch der Eindruck eines schnellen „Morsecodes“ entsteht. Abbildung 4.32 zeigt den Ausschnitt einer solchen

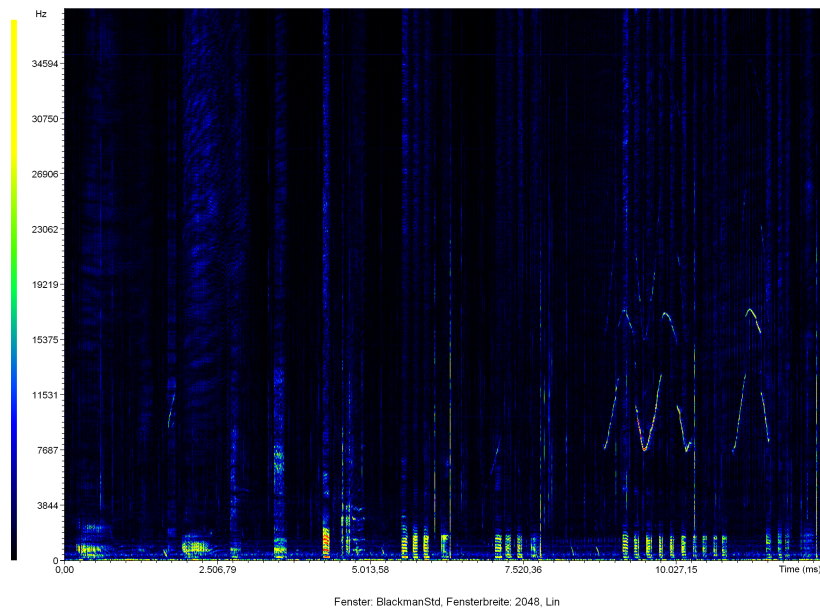


Abbildung 4.29: Sequenz von NFP-Lauten (Übersicht).

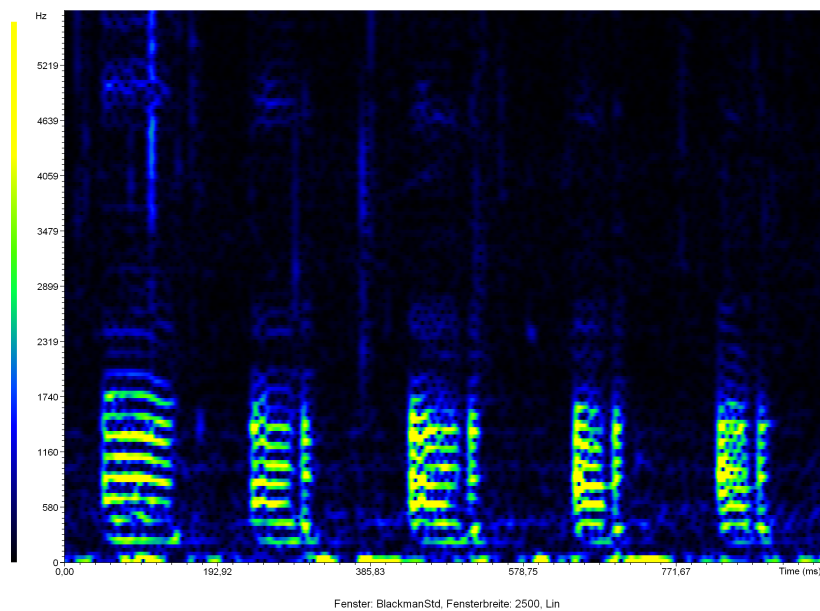


Abbildung 4.30: Sequenz von NFP-Lauten (Detail) mit Formanten.

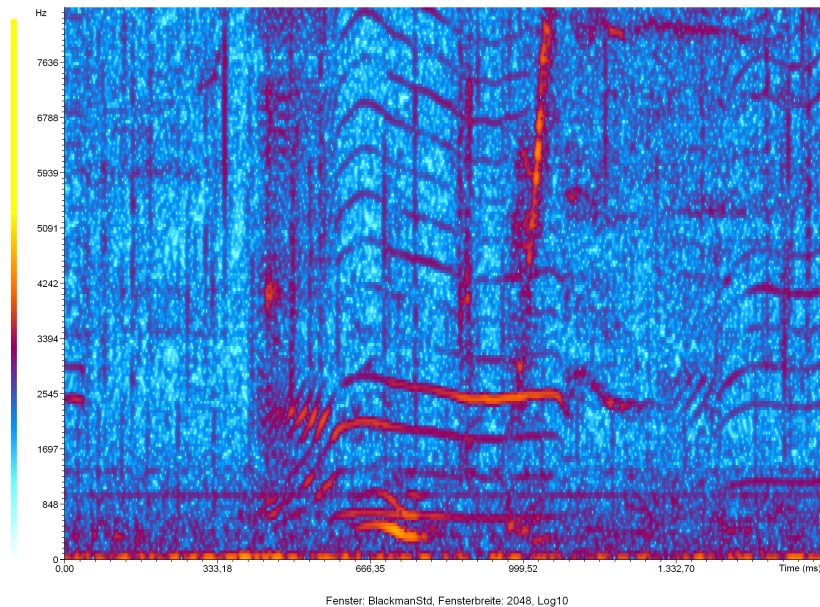


Abbildung 4.31: Pulsfrequenz-Glissando bei NFP-Laut.

Sequenz von Burst-Pulsed-Lauten. Der Ausschnitt ist weniger als eine Sekunde lang und beinhaltet 15 Burst-Pulsed-Laute. Burst-Pulsed-Laute weisen die folgenden Formen der phonetischen Modulation auf:

1. **PFM:** Burst-Pulsed-Laute weisen dieses Modulationsprinzip oft auf. Die Pulsfrequenz ändert sich zwischen Anfang und Ende einer einzelnen Pulssequenz nur wenig, variiert aber stark im Eilat-Korpus.
2. **Pulsspektrum-Drift:** Alle Burst-Pulsed-Laute weisen dieses Modulationsprinzip auf. Das Spektrum der Pulse ändert sich zwischen Anfang und Ende einer Pulssequenz aber nur wenig.
3. **PAM und PSM:** Alle Burst-Pulsed-Laute weisen diese Modulationsprinzipien auf.

Phonetische Modulation in niederfrequenten Klicklauten:

Niederfrequente Klicklaute (NFK) sind in Sequenzen angeordnet. Sie unterscheiden sich von den mittelfrequenten Klicklauten dadurch, dass die spektrale Energie der einzelnen Klicklaute in charakteristischer Art und Weise auf einen sehr engen Frequenzbereich konzentriert ist. Niederfrequente Klicklaute weisen die folgenden Formen der phonetischen Modulation auf:

1. **PFM:** Niederfrequente Klick-Trains weisen dieses Modulationsprinzip oft auf. Abbildung 4.33 zeigt ein Beispiel.

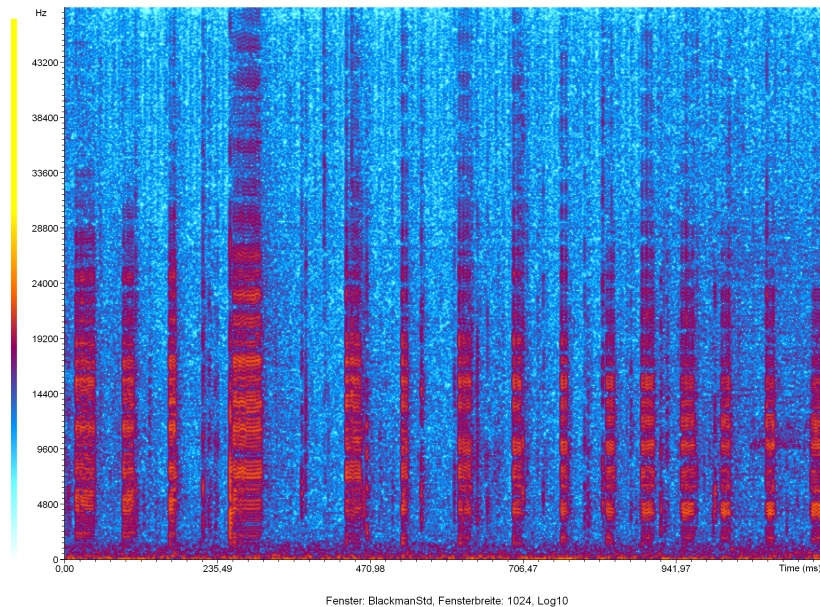


Abbildung 4.32: Sequenz von Burst-Pulsed-Lauten.

2. **PAM:** Niederfrequente Klick-Trains weisen dieses Modulationsprinzip oft auf. Die Unterschiede in den Pulsamplituden sind nicht immer evident.
3. **Pulsspektrum-Drift:** Niederfrequente Klick-Trains weisen dieses Modulationsprinzip oft auf – siehe Abbildungen 4.33 und 4.34.

Phonetische Modulation in mittelfrequenten Klicklauten:

Mittelfrequente Klicklaute (MFK) sind in Sequenzen angeordnet. Sie unterscheiden sich von den niederfrequenten Klicklauten dadurch, dass die spektrale Energie der einzelnen Klicklaute ungleichmäßig über einen großen Ausschnitt des Frequenzraums verteilt ist. Mittelfrequente Klicklaute weisen die folgenden Formen der phonetischen Modulation auf:

1. **PFM:** Alle mittelfrequenten Klick-Trains weisen dieses Modulationsprinzip auf. Die Abbildungen 4.35, 4.36 und 4.37 zeigen verschiedene Beispiele.
2. **PAM:** Alle mittelfrequenten Klick-Trains weisen dieses Modulationsprinzip auf.
3. **Pulsspektrum-Drift:** Alle mittelfrequenten Klick-Trains weisen dieses Modulationsprinzip auf. Die Abbildungen 4.35, 4.36 und 4.37 zeigen verschiedene Beispiele.
4. **Pulsfrequenz-Glissando:** Mittelfrequente Klicklaute weisen dieses Modulationsprinzip oft auf. Im Spektrogramm erzeugt dieses Modulationsphänomen

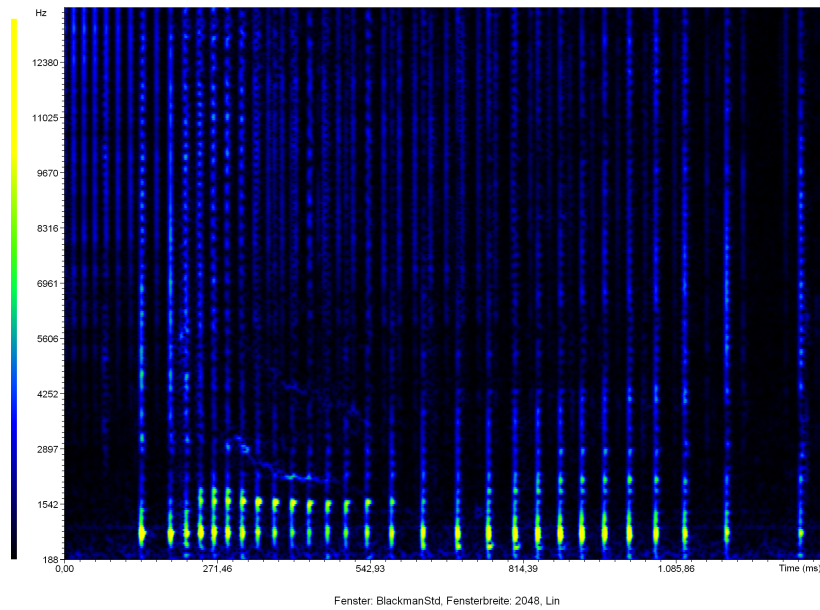


Abbildung 4.33: NF-Klicklaute mit PFM und Pulsspektrum-Drift.

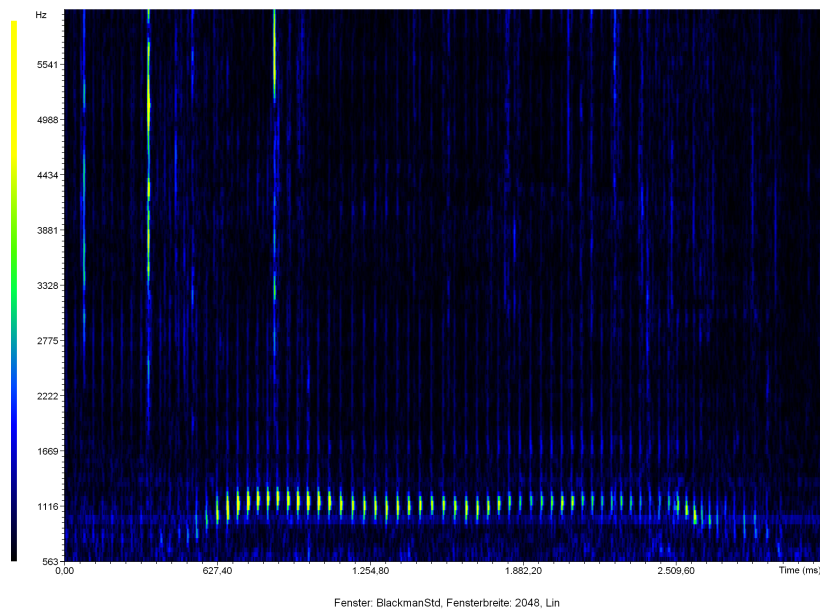


Abbildung 4.34: NF-Klicklaute mit Pulsspektrum-Drift.

bei hoher Pulswiederholrate ein charakteristisches Streifenmuster [116] – siehe Abbildung 4.38.

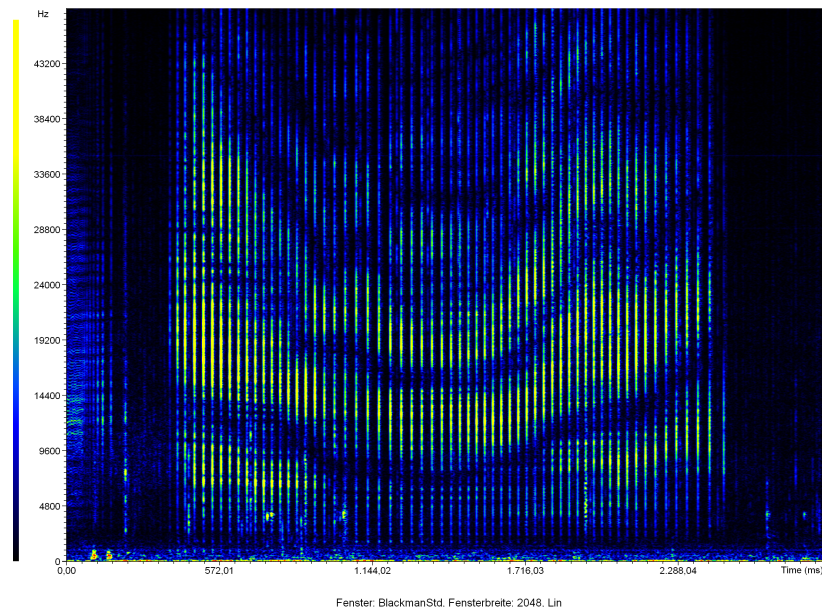


Abbildung 4.35: MF-Klicklaute mit PFM und Pulsspektrum-Drift.

4.5.5 Zusammenfassung

Die Ausführungen in diesem Kapitel können die mannigfaltigen Modulationsphänomene in den akustischen Signalen von *Tursiops truncatus* nur in sehr komprimierter Form wiedergeben. Dies liegt in der Natur der Sache: Das Eilat-Korpus enthält *Millionen* von Lauten, von denen keiner zu hundert Prozent irgendeinem anderen gleicht.

Das dargestellte phänomenologische Klassifikationssystem erhebt keinen Anspruch auf Vollständigkeit. Erst durch viele detaillierte Analysen wirklich aller Vokalisationsphänomene wird es in Zukunft möglich sein, ein umfassendes wissenschaftliches Klassifikationssystem für die akustischen Signale dieser Spezies zu erstellen.

Tabelle 4.2 zeigt die verschiedenen Arten der phonetischen Modulation und ihr Auftreten in den Vokalisations-Grundtypen im Überblick.

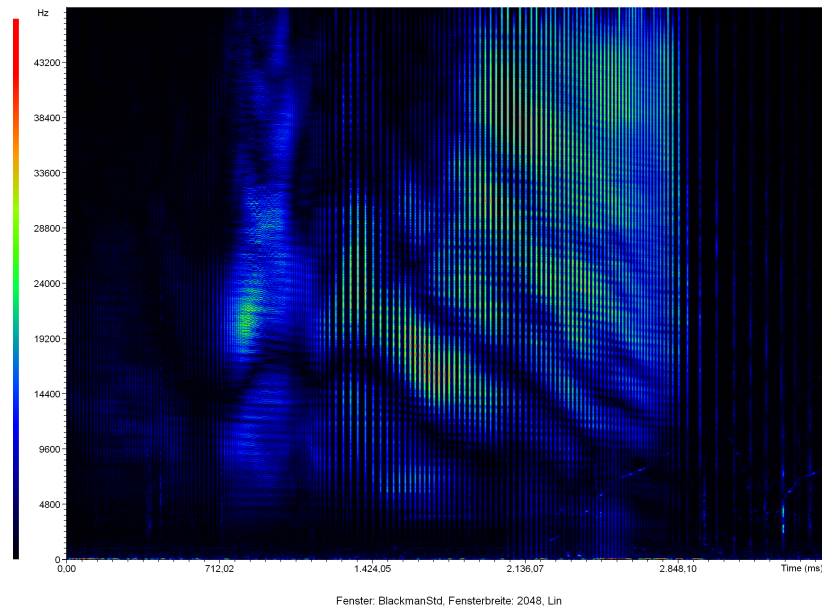


Abbildung 4.36: MF-Klicklaute mit PFM und Pulsspektrum-Drift.

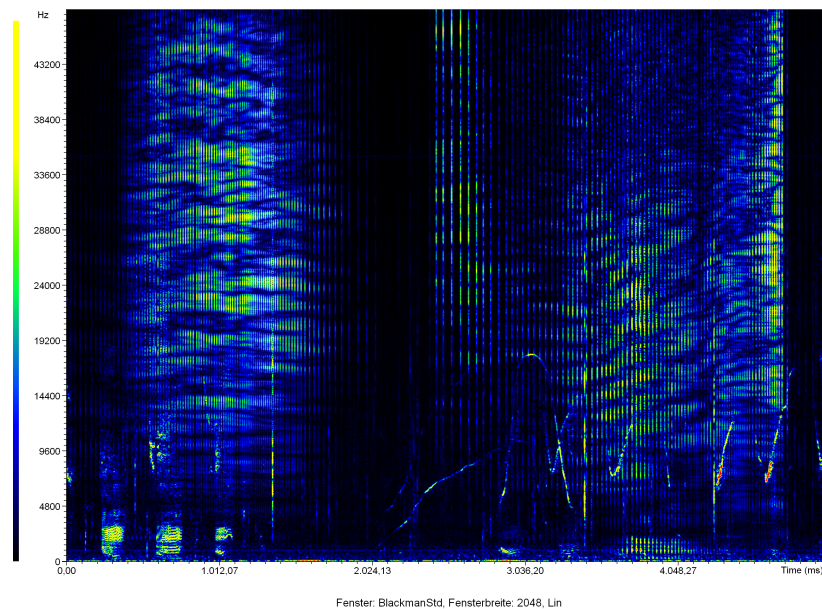


Abbildung 4.37: MF-Klicklaute mit PFM und Pulsspektrum-Drift.

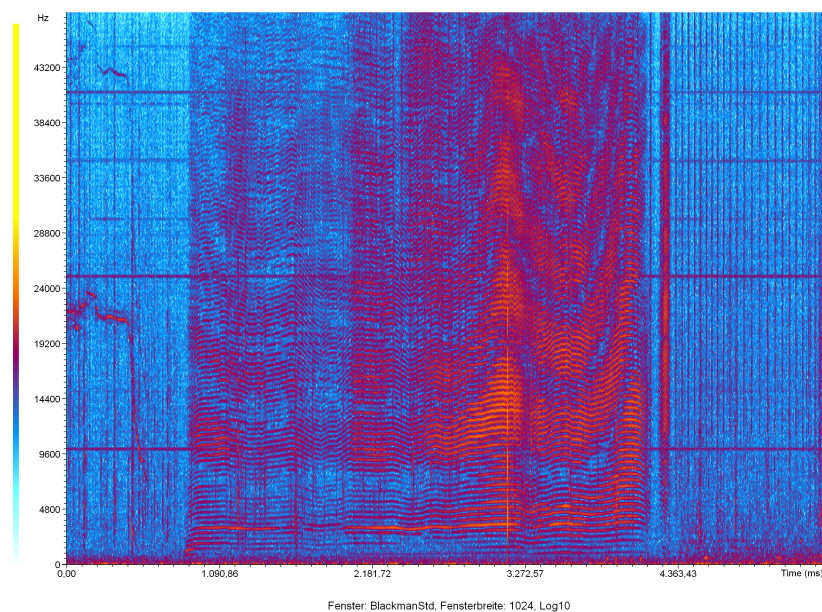


Abbildung 4.38: MF-Klicklaute mit Pulsfrequenz-Glissando. Die hohe Pulswiederholrate erzeugt im Spektrogramm ein charakteristisches Streifenartefakt [116].

	P	NFNP	NFP	BP	NFK	MFK
Glissando	+	+	—	—	—	—
Harmonische	+	—	—	+	—	—
Formanten	—	+	+	+	—	—
Vibrato	+	—	—	—	—	—
Seitenband	+	—	—	—	—	—
Frequenzsprünge	+	—	—	—	—	—
Zischen	+	—	—	—	—	—
PFM	—	—	+	+	+	+
PAM	—	—	+	+	+	+
PSM	—	—	+	+	+	+
Pulsfrequenz-Glissando	—	—	+	—	—	+
Pulsspektrum-Drift	—	—	+	+	+	+

Tabelle 4.2: Phonetische Modulation im Bereich von 0-48 kHz.

Kapitel 5

Klassifikation synthetischer Pfiffklaute

Alle komplexen Verfahren zur KDD-basierten Auswertung von Audiodaten basieren auf First-Level-ASK (FLC). Dieses Kapitel enthält eine Evaluation der in Teil I beschriebenen Methoden zur Modellierung von FLC. Die Evaluation erfolgt anhand von synthetischen Pfiffklauten, welche denen von *Tursiops truncatus* nachempfunden sind. Es wird der Frage nachgegangen, ob sich mit Hilfe der in Abschnitt 3.3 dargestellten Verfahren FLC modellieren lassen, welche unter kontrollierten Bedingungen zu sinnvollen Ergebnissen führen. Hierbei stehen Fragen der Performanz und Robustheit im Vordergrund.

5.1 Die Signaturpfiiff-Hypothese

Die Signaturpfiiff-Hypothese oder engl. *Signature-Whistle-Hypothesis* (SWH) von Caldwell und Caldwell (1965, 1968) besagt, dass jedes Individuum in einer Gruppe von Großen Tümmlern einen charakteristischen Pfiffklat mit stereotyper Zeit-Frequenz-Kontur, einen sogenannten *Signaturpfiiff*, besitzt. Caldwell und Caldwell gehen davon aus, dass es sich bei ca. 90% der Pfiffklauten von *Tursiops truncatus* um Signaturpfiiffe handelt [13, 15]. Viele Forschungen an freilebenden und in Gefangenschaft gehaltenen Tieren bestätigen diese Hypothese:

1. Mimikry von Signaturpfiiffen ist nur selten zu beobachten [113, 62, 107, 92].
2. In Gefangenschaft bilden Kälber Signaturpfiiffe im Alter von ca. 6 Monaten aus [16]. In der freien Natur scheinen alle Tiere ab einem Alter von einem Jahr Signaturpfiiffe zu besitzen [100]. Die Signaturpfiiffe bleiben über einen Zeitraum von bis zu 12 Jahren stabil [100].
3. Signaturpfiiffe von männlichen Jungtieren ähneln mehr denen ihrer Mütter, als die von weiblichen. Dies lässt sich darauf zurückführen, dass weibliche Tiere öfter als männliche in der Gruppe ihrer Mütter verbleiben [100, 101].

Die Funktion von Signaturpfiiffen besteht nach der SWH in der (Selbst-) Kennzeichnung von Individuen. Für ein soziales Tier, wie den Großen Tümmler, welches

in einer komplexen maritimen Lebensumwelt existiert, bedeutet eine individuelle akustische Kennzeichnung vermutlich einen überlebenswichtigen Vorteil [81], denn auf diese Art und Weise können Individuen sich über größere Entfernungen, bei Nacht oder schlechten Sichtverhältnissen gegenseitig lokalisieren und erkennen.

Nach Janik und Slater (1998) [62] dienen Signaturpiffe dem Zusammenhalt von Gruppenmitgliedern. In einer experimentellen Studie an vier Tieren stellten diese Autoren fest, dass sie vor allem dann produziert werden, wenn einzelne Tiere spontan von ihrer Gruppe getrennt werden. Im ungestörten Kontext konnten solche Piffe dagegen kaum beobachtet werden. Sayigh et al. (1990, 1995) zeigten, dass wildlebende Tiere stärker auf die Signaturpiffe von Gruppenmitgliedern, als auf fremde reagieren [102] (siehe auch [61]). Weiterhin zeigen Studien an freilebenden Tümmlern in Australien, dass Signaturpiffe dem Zwecke der Zusammenführung von Kälbern und Muttertieren dienen [107].

Die SWH ist allerdings nicht unumstritten. Tatsächlich ist es so, dass kein natürlicher Pfiff laut zu 100% einem anderen gleicht, auch wenn die Zeit-Frequenz-Konturen im Spektrogramm einander sehr ähnlich sind. Eine verlässliche Überprüfung der SWH an freilebenden Tieren wird durch große praktische Probleme erschwert, denn die Tiere sind sehr mobil und Pfiff laute können nur mit aufwendiger Technik bestimmten Individuen zugeordnet werden [112].

Individualisierte Pfiff laute eignen sich im Prinzip auch dazu, kontextspezifische Informationen zu transportieren: Janik, Dehnhardt und Todt (1994) zeigten an Tieren in Gefangenschaft, dass Variationen von Frequenz und zeitlichen Charakteristika mit Kontextänderungen in Diskriminierungs-Tasks korreliert sind [60]. Auch McCowan und Reiss (1995) fanden empirische Evidenzen, welche der SWH entgegenstehen: Sie kategorisierten die Pfiff laute von drei in Gefangenschaft gehaltenen Tieren und stellten fest, dass jedes der Tiere ein individuell unterschiedliches Pfiff repertoire mit nur wenigen gemeinsamen Pfiff lauten besaß [73].

In Zukunft können Verfahren zur KDD-basierten Klassifikation großer Mengen von Audiodaten sicherlich dazu beitragen, die SWH neu zu überprüfen. In diesem und in den folgenden Kapiteln werden einige neue Ansätze hierzu dargestellt.

5.2 Fragestellung

In den folgenden Abschnitten werden die in den Kapiteln 2 und 3 beschriebenen Methoden zur Modellierung von FLC experimentell evaluiert. Hierzu werden Audiodateien verwendet, welche künstlich generierte Pfiff laute in verschiedenen Variationen enthalten. Im einzelnen werden folgende Fragen untersucht:

1. Wie robust sind FLC für Pfiff laute in Bezug auf den Signal-Stör-Abstand in Audiodaten?
2. Wie robust sind FLC für Pfiff laute in Bezug auf Überlagerungen von Pfiff lauten in Audiodaten?

3. Wie robust sind FLC für Pfiff-laute in Bezug auf das Fehlen von Teilen der Zeit-Frequenz-Kontur von Pfiff-laute?
4. Wie robust sind FLC für Pfiff-laute in Bezug auf Verschiebungen der Grundfrequenz von Pfiff-laute?
5. Läßt sich eine Menge von FLC für Signaturpfiffe aus einem Audiokorpus mit Signaturpfeifen mit Hilfe des in Abschnitt 3.3.11 beschriebenen Verfahrens erlernen?

Jeder der oben genannten Fragen entspricht im Folgenden ein Experiment.

5.3 Methodische Vorbemerkungen

Die Performanz von FLC hängt von verschiedenen Parametern ab: (1) Parameter, die der spektrographischen Darstellung zuzuordnen sind (vgl. Abschnitt 3.2), (2) Parameter, die der Extraktion von elementaren Signaturen zuzuordnen sind (vgl. Kapitel 2 und Abschnitt 3.3.1) sowie (3) Parameter, die das Verhalten des AA (vgl. Abschnitt 3.3.3) bestimmen. Für die weiter unten beschriebenen Experimente wurden alle diese Parameter im Rahmen des FLCD-Prozesses manuell so justiert, dass möglichst optimale Ergebnisse erzielt werden konnten. Das manuelle Parameter-Fitting ist interaktiv und beruht im Wesentlichen auf Erfahrungswerten.

Um die Performanz von Klassifikatoren zu messen werden in dieser Arbeit die Maße *Accuracy* und *Precision* verwendet. Diese Maße basieren auf den Variablen in Tabelle 5.1.

	positiv klassifiziert	nicht erkannt	
tatsächlich positiv	a	c	a+c
tatsächlich negativ	b	d	b+d
	a+b	c+d	n

Tabelle 5.1: Klassifikationsvariablen

Die Treffergenauigkeit $Accuracy = (a + d)/n$ gibt den Anteil der korrekt klassifizierten Beispiele an. Die Präzision $Precision = a/(a + b)$ gibt den Anteil der positiv klassifizierten Beispiele an, die tatsächlich positiv sind.

5.4 Audiomaterialien

Für diese Untersuchung wurden verschiedene Mengen von Audiodateien mit künstlichen Pfiff-laute generiert. In diesen werden die tatsächlichen Gegebenheiten im Eilat-Korpus nachempfunden:

1. Die Menge künstlicher Pfiff-laute A^P : Mit einem Softwaresynthesizer wurden 10 Pfiff-laute generiert, die in Gestalt, Lautstärke, Dauer und Frequenzbereich denen der Eilat-Delphine so ähneln, dass ein Laie sie nicht von echten Delphinlauten unterscheiden kann (siehe Abbildung 5.1). Die Dauer der künstlichen Pfiff-laute beträgt zwischen 570ms und 960ms. Die Bandbreite der einzelnen Pfiff-laute liegt zwischen 6 und 11 kHz. Der Pfiff-laut mit der tiefsten Frequenz reicht von 3,5 bis 10,7 kHz. Der Pfiff-laut mit der höchsten Frequenz reicht von 8,3 bis 19,1 kHz. Die Pfiff-laute besitzen alle eine konstante Amplitude. Jeder Laut weist eine Grundkontur auf, die durch einen Anstieg und ein Abfallen der Frequenz bis zum Ausgangspunkt charakterisiert ist. Eine analoge Grundkontur besteht auch bei den meisten der Signaturpiffe der Eilat-Delphine. Die Kontur beinhaltet 10 zufällig eingestreute spontane Änderungen des Frequenzverlaufs pro Sekunde.

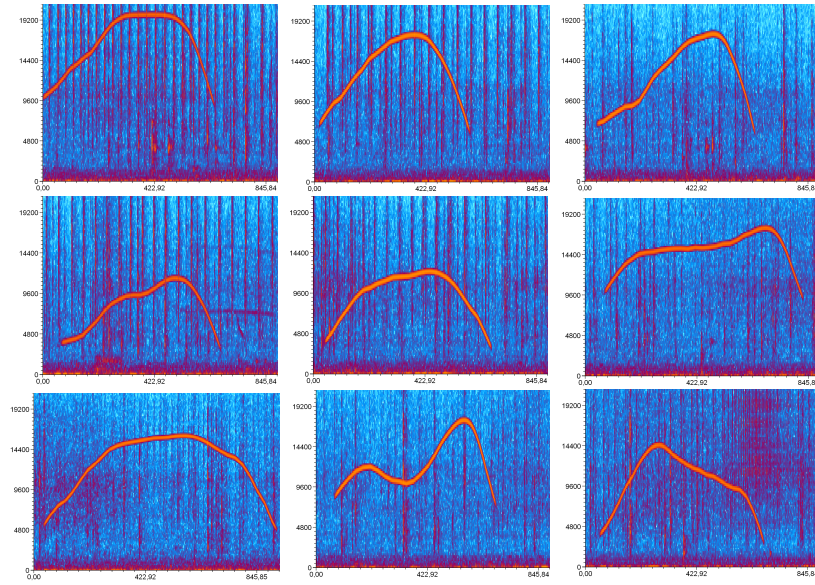


Abbildung 5.1: Neun der zehn Pfiff-laute in A^P .

2. Die Menge von Hintergrund-Audiodateien A^{HG} : Aus dem Eilat-Korpus wurden manuell 10 längere Abschnitte ausgewählt, die keine Pfiff-laute – wohl aber andere Vokalisationen – enthielten. Aus diesen wurden 51 Hintergrund-Audiodateien von je einer Minute Länge generiert, indem aus den Abschnitten zufällig unterschiedlich lange Sequenzen von weniger als 5 Sekunden Dauer ausgewählt und aneinandergefügt wurden.
3. Die Trainingsdatei A^{TR} : Aus einer der Hintergrund-Audiodateien und den künstlichen Pfiff-lauten wurde eine Datei mit Trainingsdaten generiert, indem die Pfiff-laute unverändert nacheinander ohne Überlappungen in die Hintergrund-Datei geschrieben wurden.

4. Die Menge normierter Audiodateien A^{NORM} : Aus 10 zufällig ausgewählten Hintergrund-Audiodateien wurden 10 normierte Audiodateien mit Pfifflauten generiert. Hierzu wurden in jede Hintergrunddatei 40 künstliche Pfiff-laute aus A^P in zufälliger Abfolge ohne Überlappungen eingestreut.
5. Die Menge A^{SNR} zur Evaluation von FLC in Bezug auf den Signal-Stör-Abstand (Signal-to-Noise-Rate, kurz SNR): Aus 50 Hintergrund-Audiodateien wurden 50 Dateien mit Pfifflauten generiert, in denen Varianzen des Signal-Stör-Abstandes auftreten. Hierzu wurden in jede Hintergrunddatei 40 Pfiff-laute aus A^P in zufälliger Abfolge ohne Überlappungen eingestreut. Dabei wurde der Signal-Stör-Abstand variiert. Er ist definiert durch:

$$SNR_{dB} = 10 \lg \frac{E_S}{E_N} \quad (5.1)$$

wobei die durchschnittliche Energie des Signals S gegeben ist durch

$$E_S = \frac{1}{\Delta t} \sum_{t=0}^{\Delta t} S^2(t) \quad (5.2)$$

und die durchschnittliche Energie des Hintergrundsignals N ab einem Anfangszeitpunkt t_0 gegeben ist durch

$$E_N = \frac{1}{\Delta t} \sum_{t=t_0}^{t_0+\Delta t} N^2(t) \quad (5.3)$$

6. Die Menge A^{UEB} zur Evaluation von FLC in Bezug auf gegenseitige Überlagerungen von Pfifflauten: Aus 10 Hintergrunddateien wurde eine Menge von Audiodateien generiert, in der Pfiff-laute aus A^P so eingestreut sind, dass sie sich überlagern. Die Anzahl der durchschnittlichen Überlagerungen pro Pfiff-laute wird dabei von 0 bis 10 variiert. Abbildung 5.2 zeigt einen Ausschnitt von A^{UEB} mit einem durchschnittlichen Überlagerungsfaktor von drei.
7. Die Menge A^{PART} zur Evaluation der FLC in Bezug auf partiell ausgeprägte Pfiff-laute: Aus 50 Hintergrunddateien wurde eine Menge von Audiodateien mit partiell ausgeprägten Pfifflauten generiert. Hierzu wurden in jede Hintergrunddatei aus A^{HG} 40 künstliche Pfiff-laute aus A^P in zufälliger Abfolge ohne Überlappungen eingestreut. Dabei wurden systematisch – beginnend vom Pfiffanfang – Teile der Pfiff-laute weggelassen. Der Grad der Weglassung betrug dabei zwischen 0% und 90%. Abbildung 5.3 zeigt einen Ausschnitt von A^{PART} mit einem durchschnittlichen Weglassungsgrad von 50%.

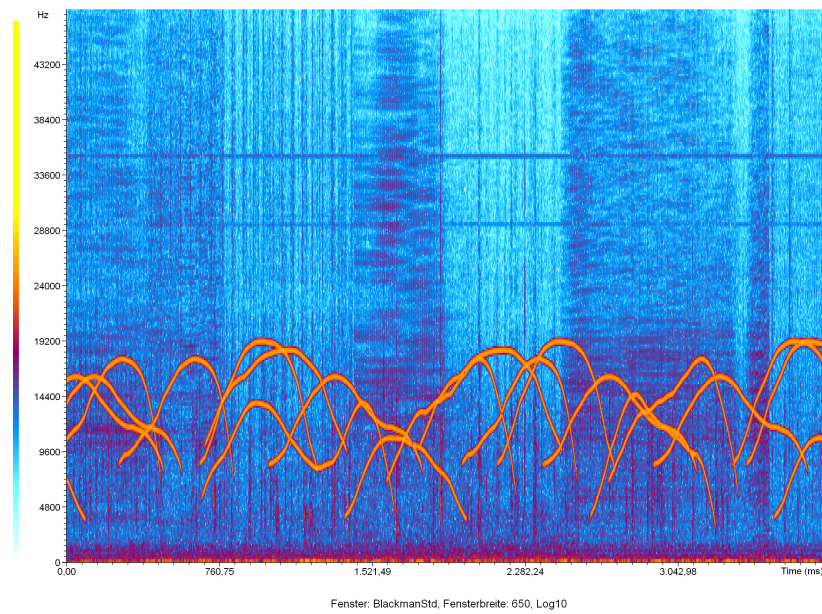


Abbildung 5.2: Überlagerte künstliche Piffflaute. Der durchschnittliche Überlagerungsfaktor beträgt hier 3.

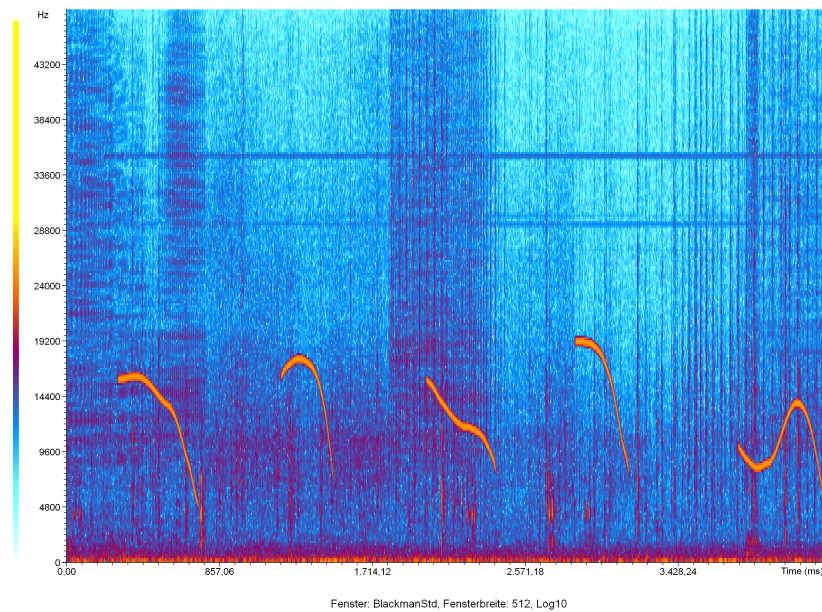


Abbildung 5.3: Weglassungen in synthetischen Piffflauten. Der durchschnittliche Weglassungsgrad beträgt hier 50%.

5.5 Robustheit von FLC und SNR

Im Eilat-Korpus variiert die Lautstärke der einzelnen Pfiff-laute sehr stark, da die Tiere frei umherschwimmen und sich teilweise über 100m vom Hydrophon entfernt befanden. Auch die Störgeräusche variieren mitunter sehr. Hinzu kommt, dass Pfiff-laute oft von lauten Klick-Trains überlagert werden. In diesem Experiment wird der Frage nachgegangen, wie robust spektrographische FLC für Pfiff-laute in Bezug auf diese Faktoren sind. Hierbei wurde wie folgt verfahren:

1. Aus der Trainingsdatei A^{TR} wurden nach dem Verfahren in Abschnitt 3.3.1 manuell 10 FLC generiert. Jedem Pfiff-laut in A^P entspricht dabei genau ein FLC, so dass $|K| = 10$. Die FLC beruhen auf logarithmisierten Spektrogrammen (vgl. Abschnitt 3.2).
2. Die Testmenge A^{SNR} wurde mit Hilfe der FLC in K nach dem Verfahren in Abschnitt 3.3.3 annotiert. Anschließend wurden die Größen *Accuracy*, und *Precision* für F^K berechnet. Die Berechnungen wurden für eine *SNR* von -50dB bis 0dB durchgeführt. Eine *SNR* > 0dB wurde in der Untersuchung nicht berücksichtigt, da solche Verhältnisse im Eilat-Korpus nicht vorkommen.

Das Ergebnis dieses Experiments ist in Abbildung 5.4 dargestellt. Es ist zu sehen, dass die Performanz der FLC erst bei *SNR*-Werten unterhalb von -30dB rapide abnimmt. Falsch positiv klassifizierte Audiodaten fehlen fast völlig, wodurch der hohe *Precision*-Wert auch bei geringer Genauigkeit zu erklären ist.

5.6 Überlagerungen von Pfiff-lauten

Im Eilat-Korpus überlagern sich Pfiff-laute häufig. In der Mehrzahl der Fälle sind allerdings nicht mehr als zwei bis drei Pfiff-laute überlagert. In diesem Experiment wird der Frage nachgegangen, wie robust spektrographische FLC für Pfiff-laute in Bezug auf diesen Faktor sind. Hierbei wurde wie folgt verfahren:

1. Aus der Trainingsdatei A^{TR} wurden nach dem Verfahren in Abschnitt 3.3.1 manuell 10 FLC generiert. Jedem Pfiff-laut in A^P entspricht dabei genau ein FLC, so dass $|K| = 10$. Die FLC beruhen auf logarithmisierten Spektrogrammen (vgl. Abschnitt 3.2).
2. Die Testmenge A^{UEB} wurde mit Hilfe der FLC in K nach dem Verfahren in Abschnitt 3.3.3 annotiert. Anschließend wurden die Größen *Accuracy*, und *Precision* für F^K berechnet. Die Berechnungen wurden für 0 bis 10 Überlagerungen pro Pfiff-laut durchgeführt. Höhere Überlagerungsraten wurden in der Untersuchung nicht berücksichtigt, da Überlagerungen von mehr als 4 Pfiff-lauten im Eilat-Korpus sehr selten sind.

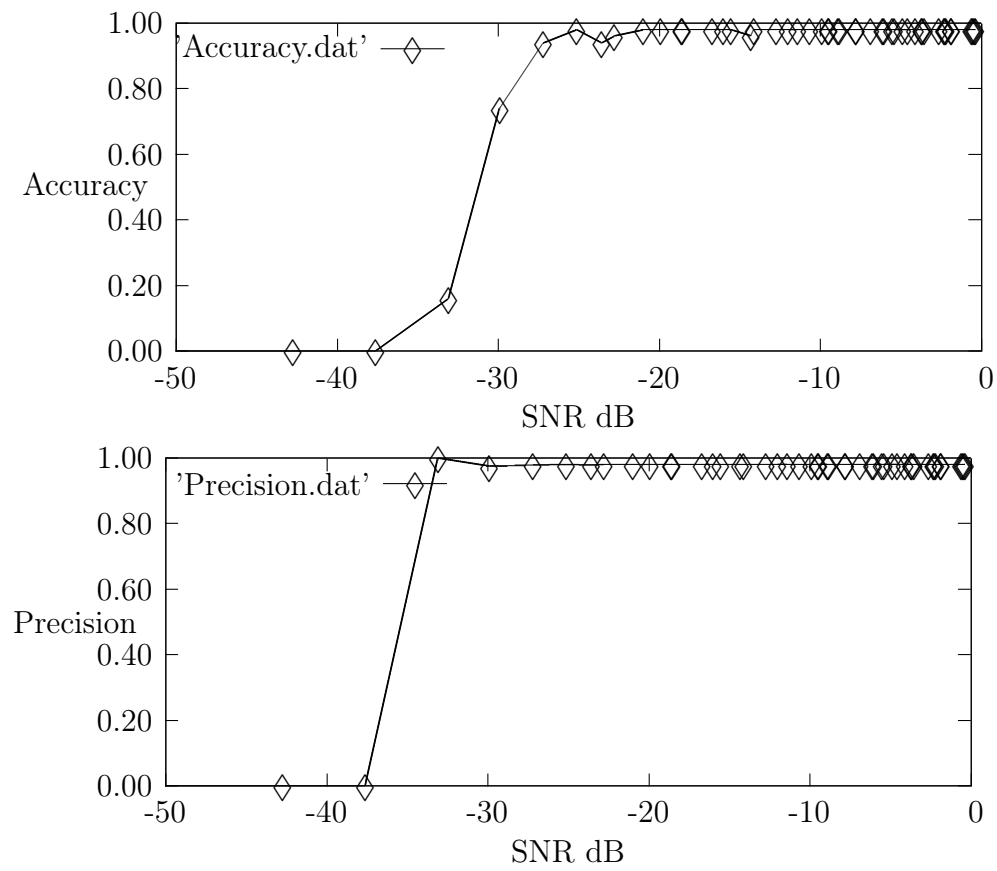


Abbildung 5.4: Performanz von FLC gegenüber Änderungen der SNR

Das Ergebnis dieses Experiments ist in Abbildung 5.5 dargestellt. Es ist zu sehen, dass die Performanz der Klassifikatoren bei einem Überlagerungsfaktor > 2 allmählich abnimmt. Die Genauigkeit beträgt bei durchschnittlich drei Überlagerungen pro Pfiff laut immerhin noch 79% bei einer Präzision von 100%.

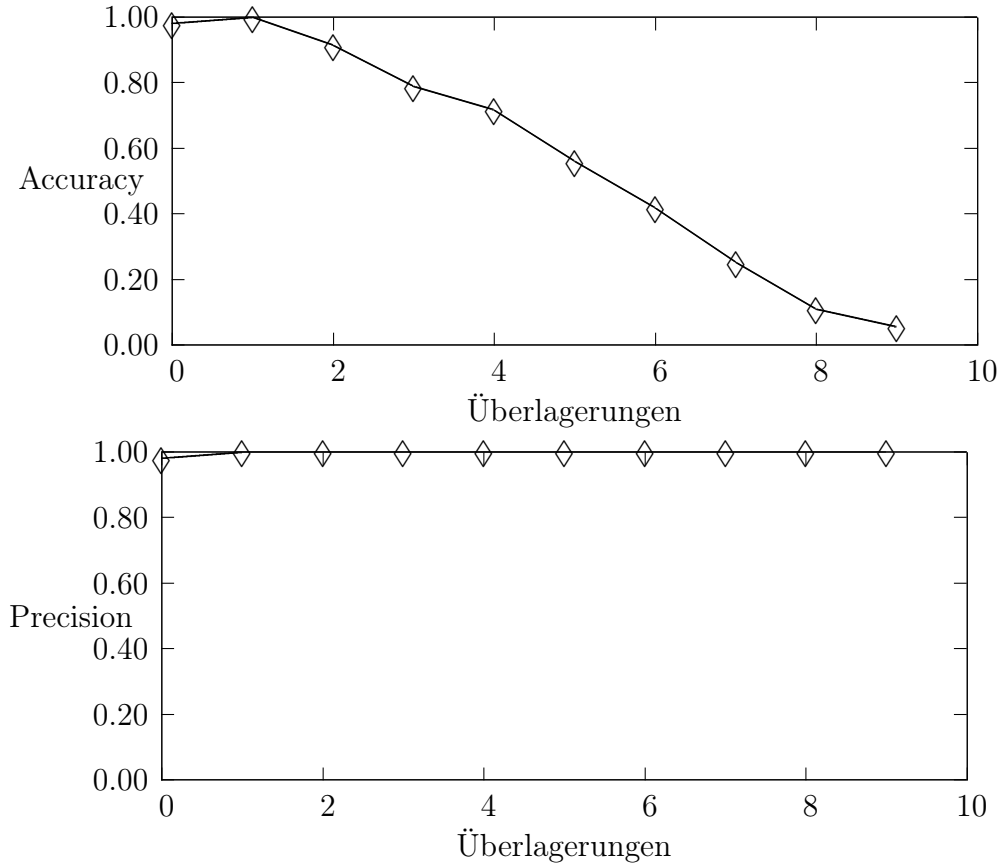


Abbildung 5.5: Performanz von FLC bei Überlagerungen

5.7 Partiiell instanziierte Pfiff laute

Im Eilat-Korpus treten häufig Pfiff laute auf, die nur teilweise dem Idealtyp eines vollständigen Signaturpiffs entsprechen. Teile der Zeit-Frequenz-Kontur fehlen bei diesen Lauten vollständig. In diesem Experiment wird der Frage nachgegangen, wie robust spektrographische FLC für Pfiff laute in Bezug auf diesen Faktor sind. Hierbei wurde wie folgt verfahren:

1. Aus der Trainingsdatei A^{TR} wurden nach dem Verfahren in Abschnitt 3.3.1 manuell 10 FLC generiert. Jedem Pfiff laut in A^P entspricht dabei genau ein FLC, so dass $|K| = 10$. Die FLC beruhen auf logarithmisierten Spektrogrammen (vgl. Abschnitt 3.2).

2. Die Testmenge A^{PART} wurde mit Hilfe der FLC in K nach dem Verfahren in Abschnitt 3.3.3 annotiert. Anschließend wurden die Größen *Accuracy* und *Precision* für F^K berechnet. Die Berechnungen wurden für Weglassungsgrade von 0 bis 90% durchgeführt.

Das Ergebnis dieses Experiments ist in Abbildung 5.6 dargestellt. Es ist zu sehen, dass die Performanz der Klassifikatoren bis zu Weglassungsgraden von 60% und mehr stabil bleibt. Die Präzision ist auch hier selbst bei hohen Weglassungsgraden nicht beeinträchtigt.

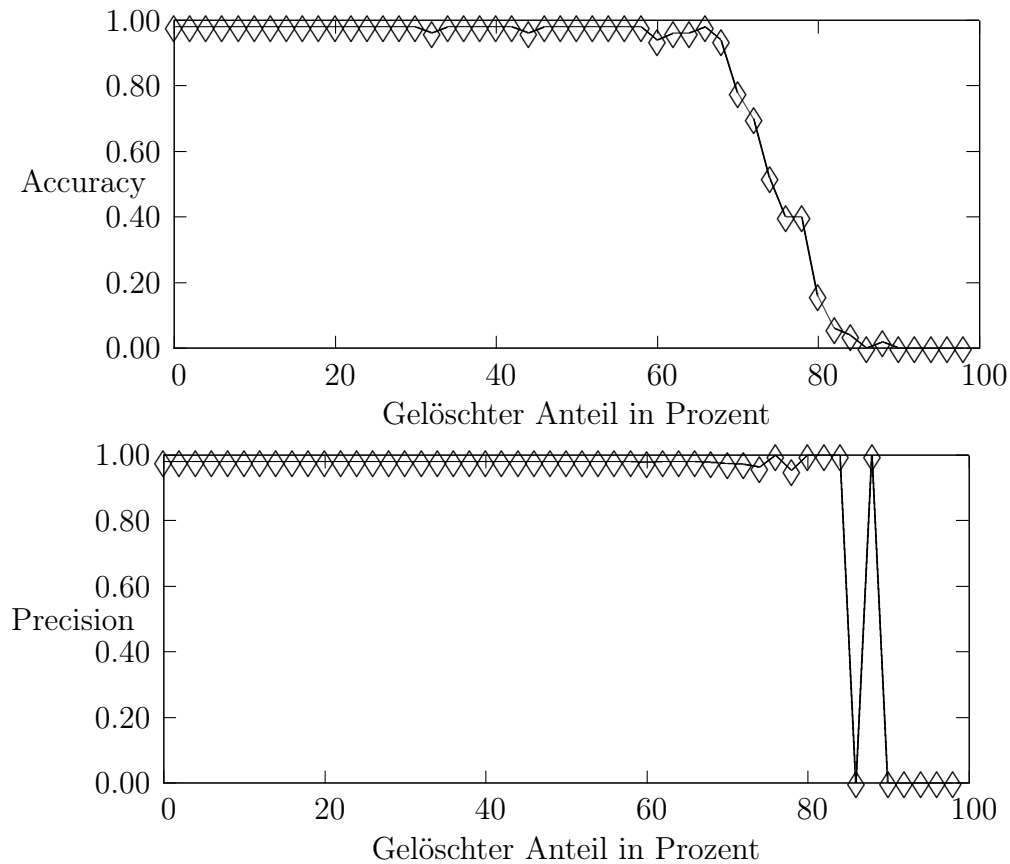


Abbildung 5.6: Performanz bei partiell gelöschten Pfifflauten.

5.8 Verschiebungen der Grundfrequenz

Die Frage, wie stabil die Höhe der Grundfrequenz bei natürlichen Signaturpfeifen ist, kann nicht immer eindeutig beantwortet werden. Bestimmte Pfiffaute haben eine sehr stabile Grundfrequenz, andere wiederum nicht. In diesem Experiment wird der Frage nachgegangen, wie robust spektrographische FLC für Pfiffaute in Bezug auf diesen Faktor sind. Hierbei wurde wie folgt verfahren:

1. Aus der Trainingsdatei A^{TR} wurden nach dem Verfahren in Abschnitt 3.3.1 manuell 10 FLC generiert. Jedem Piff laut in A^P entspricht dabei genau ein FLC, so dass $|K| = 10$. Die FLC beruhen auf logarithmisierten Spektrogrammen (vgl. Abschnitt 3.2).
2. Anschließend wurden die FLC in K nach dem Verfahren in Abschnitt 3.3.7 über einen vorgegebenen Frequenzbereich geklont. Für jedes mögliche Frequenzband wurde genau ein Klon generiert. Der von den Klonen abgedeckte Frequenzbereich reichte von -937.5 Hz bis +937.5 Hz gegenüber dem ursprünglichen. Dies entspricht in positiver und negativer Richtung je 5 Frequenzbändern à 187.5 Hz (entsprechend der DFT-Fensterbreite $w = 512$). Die Menge K wurde so zur Menge K' mit insgesamt 100 FLC erweitert.
3. Die Testmenge A^{NORM} wurde mit Hilfe von K' nach dem Verfahren in Abschnitt 3.3.3 annotiert. Anschließend wurden die Größen *Accuracy* und *Precision* für jeden FLC in $F^{K'}$ berechnet.

Das Ergebnis dieses Experiments ist in Abbildung 5.7 dargestellt. Es ist zu sehen, dass die Performanz der Klassifikatoren bei einer Verschiebung von mehr als einem Frequenzband schlagartig nachläßt. Dies entspricht in diesem Experiment 187.5 Hz.

5.9 Induktion von FLC für Piff laute

In diesem Experiment wird der Frage nachgegangen, ob das in Abschnitt 3.3.11 beschriebene Verfahren zur Induktion spektrographischer FLC für Signaturpiffe geeignet ist. Hierzu wurde wie folgt verfahren:

1. Die Menge A^{NORM} wurde mit Hilfe einer Menge K^1 von fünf geklonten FLC für 10 ms dauernde Sinusschwingungen annotiert (vgl. Abschnitt 3.3.7). Die Frequenzverschiebung der Klone betrug 25 Frequenzbänder und deckte einen den Bereich von 0-24 kHz ab. Aus der Menge der so erhaltenen Annotationen wurde nach dem Verfahren in Abschnitt 3.3.11 eine Menge von FLC auf induktivem Weg generiert. Der Suchraum der FLC in K^1 war so dimensioniert, dass bei der automatischen Extraktion der Signaturen alle Anteile der potentiell zu findenden Piff laute berücksichtigt wurden.
2. Die Menge K_{Tmp} der so erhaltenen FLC wurde im Rahmen des FLCD-Prozesses evaluiert und manuell bereinigt (vgl. Abschnitt 2.5). Es wurden z.B. alle FLC aussortiert, die eindeutig keinen Piff laut en entsprachen oder auf Verschmelzungen von unter 10 automatisch extrahierten Signaturen beruhten.

Das Ergebnis dieses Experiments ist in Abbildung 5.8 dargestellt. Im Rahmen des FLCD-Prozesses ist es leicht möglich, aus A^{NORM} für die 10 verschiedenen künstlichen Piff laute in A^P zehn FLC auf induktivem Weg zu gewinnen. Die Menge

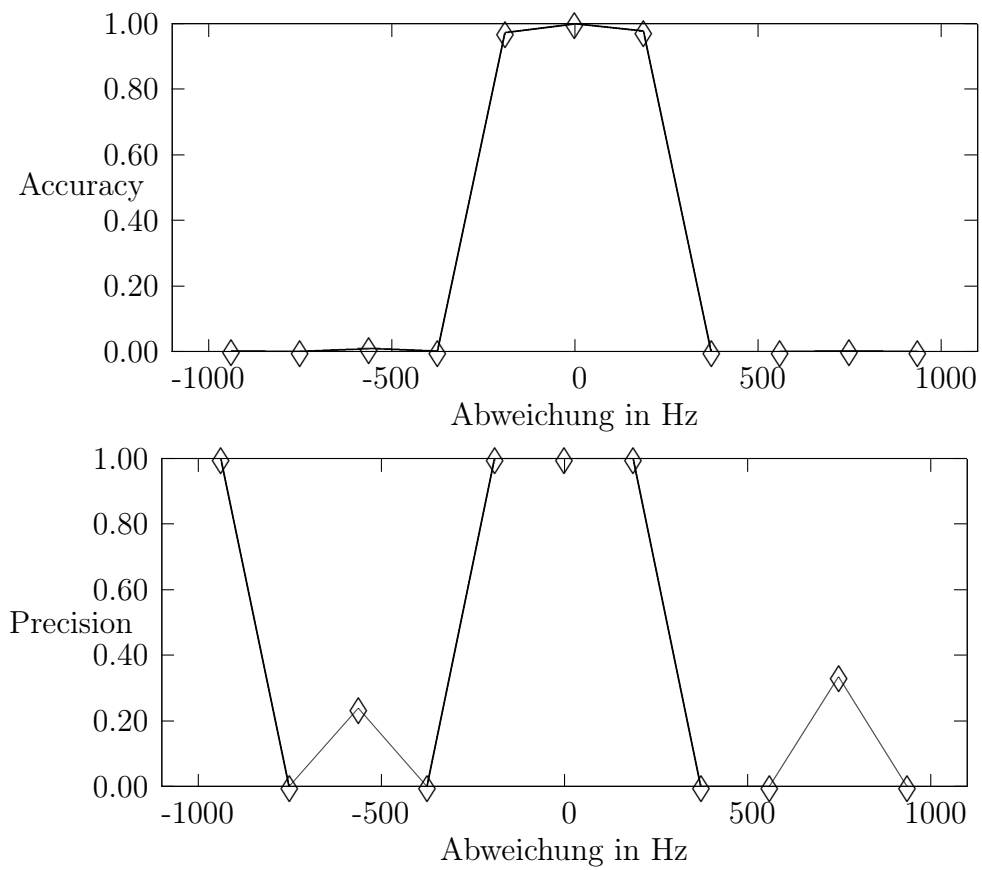


Abbildung 5.7: Performanz bei Frequenzabweichungen. Jedes Frequenzband hat eine Breite von 187.5 Hz. In einem Bereich von ± 187.5 Hz sind die Klassifikatoren stabil.

$K_{T_{mp}}$ muss hierzu allerdings manuell bereinigt werden. Sie umfasste nach der ersten Ausführung des Data-Mining-Schrittes 157 FLC, welche aus 1904 automatisch extrahierten Signaturen gewonnen wurden. Von den 157 FLC konnten 106 sofort gelöscht werden, da sie auf Signaturen mit nom -Werten < 3 beruhten. Der Rest konnte in einem zweiten Durchgang zu 28 FLC verschmolzen werden. Neun der zehn FLC mit den höchsten nom -Werten sind in Abbildung 5.8 dargestellt. Sie entsprechen exakt 9 verschiedenen künstlichen Pffifflauten in A^P .

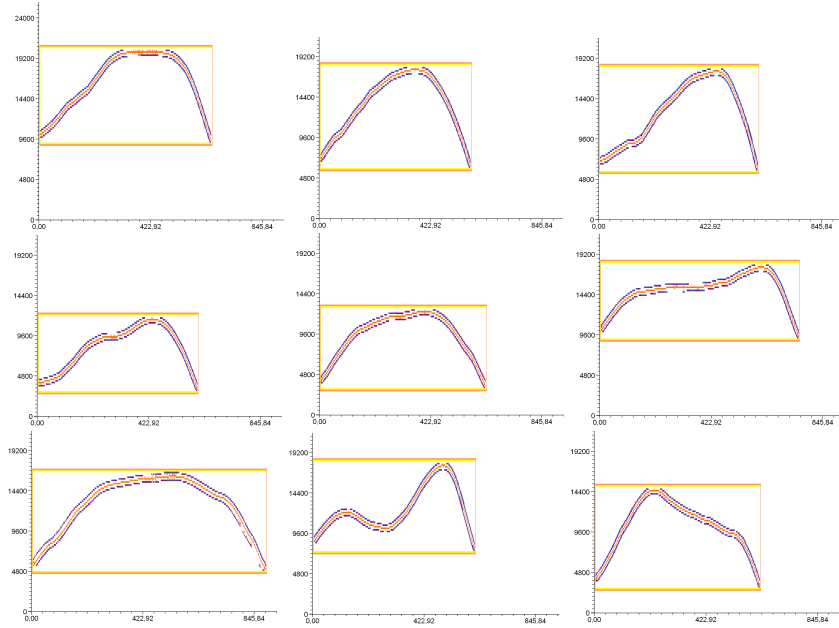


Abbildung 5.8: Neun der zehn FLC mit den höchsten nom -Werten in K . Die Signaturen entsprechen den künstlichen Pffifflauten in Abbildung 5.1

5.10 Diskussion

Die in diesem Kapitel vorgenommene Evaluation kann sicherlich eine Untersuchung an natürlichen Pffifflauten nicht ersetzen. Sie gibt aber einen Eindruck von den Potentialen der KDD-basierten Programmierung von ASK. Sie zeigt vor allem, dass sich mit Hilfe der in Teil I beschriebenen KDD-Methoden FLC modellieren lassen, welche unter kontrollierten Bedingungen zu genau vorhersagbaren Resultaten führen. Im Einzelnen lassen sich folgende Ergebnisse festhalten:

1. FLC für synthetische Pffifflaute sind hinreichend robust in Bezug auf den Signal-Stör-Abstand in Audiodaten.
2. FLC für synthetische Pffifflaute sind hinreichend robust in Bezug auf gegenseitige Überlagerungen von Pffifflauten in Audiodaten.

3. FLC für synthetische Piffflaute sind hinreichend robust in Bezug auf das Fehlen von Teilen der Zeit-Frequenz-Kontur von Piffflauten.
4. FLC für synthetische Piffflaute können empfindlich in Bezug auf Verschiebungen der Grundfrequenz von Piffflauten reagieren.
5. Eine Menge von FLC für synthetische Piffflaute kann aus einem Audiokorpus mit Hilfe des in Abschnitt 3.3.11 beschriebenen Verfahrens erlernt werden. Voraussetzung hierfür ist eine hochgradig stereotype Zeit-Frequenz-Kontur der Piffflaute.
6. Eine explizite Programmierung von FLC für stereotype Piffflaute ist dann sinnvoll, wenn die Lautklassen bekannt sind und konkrete Beispiele zur Extraktion der spektrographischen Signaturen zur Verfügung stehen.

Zusammenfassend kann gesagt werden, dass es mit Hilfe der beschriebenen Verfahren möglich ist, FLC zu modellieren, die hinreichend genau und stabil sind, um auch komplexe KDD-Untersuchungen an empirischen Daten durchzuführen.

Kapitel 6

Pfifftypen im Eilat-Korpus

In diesem Kapitel wird das in Abschnitt 3.4.7 beschriebene Verfahren zur Analyse von Pfifflauten eingesetzt. Das Verfahren wird im Folgenden auch kurz SSDM-Clustering¹ genannt. Ziel der Studie ist es, allgemeine spektrographische Charakteristika und die Häufigkeiten von Pfifftypen im Eilat-Korpus zu bestimmen.

6.1 Fragestellung

Im vorausgehenden Kapitel wurde darauf hingewiesen, dass die Signaturpfiff-Hypothese in der Wissenschaft nicht unumstritten ist. Mit dem Eilat-Korpus liegt eine breite empirische Basis vor, welche geeignet scheint, die Häufigkeiten von Pfifftypen in den Vokalisationen dieser Spezies zu bestimmen. Durch eine Clusteranalyse könnte es gelingen, neue empirische Evidenzen für die Richtigkeit der SWH zu erhalten. Ziel der folgenden Untersuchung ist es, diesen Ansatz mit Hilfe des SSDM-Clustering-Verfahrens zu realisieren. Es stehen dabei folgende Fragen im Mittelpunkt:

1. Welches sind die häufigsten Typen von Pfifflauten im Eilat-Korpus?
2. Lassen sich symbolische Signaturen natürlicher Pfifflaute durch SSDM-Clustering zu Gruppen zusammenfassen, welche die hypothetisierten Signaturpfiffe der untersuchten Delphingruppe modellieren?
3. Lassen sich symbolische Signaturen natürlicher Pfifflaute durch SSDM-Clustering zu Gruppen zusammenfassen, welche Pfifftypen modellieren, die keinen Signaturpfiffen entsprechen?

Die Untersuchung wird parallel an synthetischen und an natürlichen Pfifflauten durchgeführt. Hierdurch wird eine Vergleichsbasis geschaffen, welche die Validität des verwendeten Verfahrens zeigt.

¹Symbolic-Similarity-Driven-Merge

6.2 Materialien

Audiomaterialien:

- A^{AEK} ist ein Korpus, welches dem Eilat-Korpus nachempfunden ist, aber ausschließlich synthetische Pfiffklaute enthält. Es besteht aus allen Dateien in A^{NORM} und jeweils elf Dateien aus A^{SNR} und A^{PART} . Es handelt sich dabei um Dateien, für die in Kapitel 5 *Accuracy*-Werte > 0.9 erzielt werden konnten. Zusätzlich enthält A^{AEK} noch die Dateien aus A^{UEB} mit den Überlagerungsfaktoren null, eins und zwei. A^{AEK} umfasst 403.201.540 Byte in 35 Dateien und enthält insgesamt 1520 synthetische Pfiffklaute.
- A^{EK} ist das gesamte Eilat-Korpus (vgl. Abschnitt 4.2).
- A^{170} ist eine Testdatei mit 170 natürlichen Pfiffklauten aus A^{EK} .

Klassifikatoren:

- κ^{WL} ist ein FLC, der aus einer mit einem Haar-Wavelet gewichteten Sinusschwingung von 20ms Dauer generiert wurde. κ^{WL} ist so optimiert, dass er auf Pfiffklaute anspricht, nicht aber auf Klickklaute. Dies wurde erreicht, indem oberhalb und unterhalb des zentralen Frequenzbereichs von ϕ^+ ein erweitertes Areal für den Umriss ϕ^- definiert wurde. Hierdurch generiert die KKF d^- des Umrisses bei Klickklauten einen geringen Wert für die Klassenähnlichkeit (siehe Abbildung 6.1). Die zeitliche Ausdehnung Δt von κ^{WL} beträgt 20ms. Die Bandbreite Δf von κ^{WL} beträgt 1687.5 Hz. Der FLC beruht auf einer logarithmisierten spektrographischen Signatur der Sinusschwingung.
- Aus κ^{WL} wurde durch Klonen (vgl. Abschnitt 3.3.7) eine Menge K^{WL} von 66 FLC generiert. Die Frequenzverschiebung der Klone beträgt drei Frequenzbänder (562.5 Hz). K^{WL} deckt einen Bereich von 2.6-21.5 kHz ab.

6.3 Methode

Beide Mengen von Audiodateien wurden mit Hilfe des im Folgenden beschriebenen mehrstufigen Verfahrens ausgewertet.

Schritt 1: Annotation der Audiodateien mit Hilfe der FLC in K^{WL} .

1. Die Menge A^{AEK} wurde mit Hilfe der FLC in K^{WL} annotiert, so dass eine Menge F^{AEK} von FLC-Annotationen mit 97458 Elementen entstand. Dies entspricht einem Durchschnitt von ca. 65 FLC-Annotationen pro Pfifflaut.
2. Die Menge A^{EK} wurde mit Hilfe der FLC in K^{WL} annotiert, so dass eine Menge F^{EK} von FLC-Annotationen mit 984033 Elementen entstand.

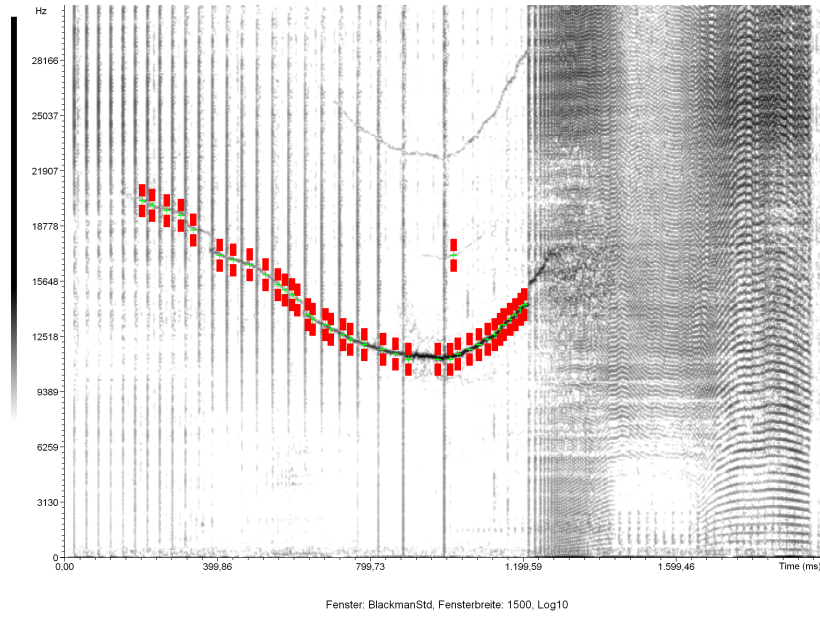


Abbildung 6.1: Annotation unter schwierigen Bedingungen: Die Abbildung zeigt Annotationen der FLC in K^{WL} . Durch den erweiterten Umriss ϕ^- (im Bild rot) wird erreicht, dass die $\kappa_i^{WL} \in K^{WL}$ Klicklaute ignorieren. Die Annotation des Pfiffes erfolgt nur zwischen ihnen. Hierdurch gelingt es, Pfiff-laute auch unter schwierigen Bedingungen zu lokalisieren.

3. Die Menge A^{170} wurde mit Hilfe der FLC in K^{WL} annotiert, so dass eine Menge F^{170} von FLC-Annotationen mit 8806 Elementen entstand.

Schritt 2: Automatische Extraktion von $3 \cdot 2$ Mengen symbolischer Signaturen der Pfiiffaute nach dem Verfahren in Abschnitt 3.4.4.

1. Aus F^{AEK} wurde eine Menge Φ_{10}^{AEK} von symbolischen Signaturen extrahiert. Hierbei wurden für die Extraktion folgende Parameter festgesetzt: $l_{min} := 10$, $\Delta t_{min} := 0.1$, $\Delta t_{max} := 1$, $\Delta f_{min} := 0.1$, $\Delta f_{max} := 0.5$. Alle Parameterangaben sind als Multiplikatoren für die zeitliche Ausdehnung Δt und die Bandbreite Δf der FLC zu verstehen.
2. Aus F^{AEK} wurde eine Menge Φ_{35}^{AEK} von symbolischen Signaturen extrahiert. Hierbei wurden für die Extraktion folgende Parameter festgesetzt: $l_{min} := 35$, $\Delta t_{min} := 0.1$, $\Delta t_{max} := 3$, $\Delta f_{min} := 0.1$, $\Delta f_{max} := 1$.
3. Aus F^{EK} wurde eine Menge Φ_{10}^{EK} von symbolischen Signaturen extrahiert. Hierbei wurden für die Extraktion folgende Parameter festgesetzt: $l_{min} := 10$, $\Delta t_{min} := 0.1$, $\Delta t_{max} := 1$, $\Delta f_{min} := 0.1$, $\Delta f_{max} := 0.5$.
4. Aus F^{EK} wurde eine Menge Φ_{35}^{EK} von symbolischen Signaturen extrahiert. Hierbei wurden für die Extraktion folgende Parameter festgesetzt: $l_{min} := 35$, $\Delta t_{min} := 0.1$, $\Delta t_{max} := 3$, $\Delta f_{min} := 0.1$, $\Delta f_{max} := 1$.
5. Aus F^{170} wurde eine Menge Φ_{10}^{170} von symbolischen Signaturen extrahiert. Hierbei wurden für die Extraktion folgende Parameter festgesetzt: $l_{min} := 10$, $\Delta t_{min} := 0.1$, $\Delta t_{max} := 1$, $\Delta f_{min} := 0.1$, $\Delta f_{max} := 0.5$.
6. Aus F^{170} wurde eine Menge Φ_{35}^{170} von symbolischen Signaturen extrahiert. Hierbei wurden für die Extraktion folgende Parameter festgesetzt: $l_{min} := 35$, $\Delta t_{min} := 0.1$, $\Delta t_{max} := 3$, $\Delta f_{min} := 0.1$, $\Delta f_{max} := 1$.

Die Idee hierbei ist, den Extraktionsalgorithmus auf zwei verschiedene Arten einzustellen: Einerseits werden strenge Kriterien an die Nachbarschaftsbeziehung angelegt und gleichzeitig sehr kurze Sequenzen (mit $l_{min} = 10$) akzeptiert. Andererseits werden weitgefasste Toleranzgrenzen für die Nachbarschaftsbeziehung angelegt aber nur sehr lange Sequenzen (mit $l_{min} = 35$) akzeptiert. Dieses Vorgehen bewirkt, dass zwei verschiedene Typen von Sequenzen extrahiert werden: (1) Sequenzen, die lokale Konturveränderungen innerhalb der Pfiiffaute modellieren und (2) Sequenzen, welche die Kontur des Pfiiffauts in seiner Gesamtheit modellieren.

Schritt 3: Bestimmen der Mengen mit den meisten potentiell interessanten Sequenzen.

1. Aus Φ_{10}^{AEK} wurden mit Hilfe des Verfahrens zum ähnlichkeitsbasierten Verschmelzen symbolischer Signaturen (vgl. Abschnitt 3.4.5) zehn Mengen gebildet, wobei das Kriterium σ^{min} der für die Verschmelzung minimal erforderlichen Anzahl symbolischer Elemente von 10% bis 100% der Länge der Signaturen betrug. Die zeitliche Toleranz Δt_{max} wurde nicht variiert und betrug konstant 0.5. Die Toleranz auf der Frequenzachse Δf_{max} betrug konstant 0.

Anschließend wurde für die entstandenen Mengen

$$\Phi_{10}^{AEK_{M10}}, \Phi_{10}^{AEK_{M20}}, \dots, \Phi_{10}^{AEK_{M100}}$$

die Kompressionsrate und die Anzahl der darin enthaltenen potentiell interessanten Signaturen bestimmt. Die Kompressionsrate CR für die Menge verschmolzener symbolischer Signaturen ist gegeben durch

$$CR_{Mx} = \frac{\sum_{i=0}^{n-1} |\phi_i^S| \in \Phi_{10}^{AEK}}{\sum_{j=0}^{m-1} |\phi_j^S| \in \Phi_{10}^{AEK_{Mx}}}, n = |\Phi_{10}^{AEK}|, m = |\Phi_{10}^{AEK_{Mx}}| \quad (6.1)$$

Potentiell interessante Signaturen werden durch eine Interessantheits-Funktion ermittelt (vgl. Abschnitt 1.1.2), welche die Kompressionsrate cr für jede einzelne symbolische Signatur berechnet und auswertet:

$$I(\phi^s) = \begin{cases} 1 & \text{if } cr(\phi^s) \geq 2 \\ 0 & \text{sonst} \end{cases} \quad (6.2)$$

wobei

$$cr(\phi^s) = \frac{\sum_{i=0}^{n-1} nom(\phi_i^s)}{n}, n = |\phi_i^s| \quad (6.3)$$

ist. nom gibt dabei den nom -Werte eines einzelnen symbolischen Elementes an.

2. Für die Mengen Φ_{35}^{AEK} , Φ_{10}^{EK} und Φ_{35}^{EK} wurden analog zu Φ_{10}^{AEK} jeweils zehn Mengen von verschmolzenen Signaturen gebildet. Auch hier wurde für jede Menge und jede einzelne symbolische Signatur die Kompressionsrate berechnet.
3. Aus den so erhaltenen $4 \cdot 10$ Mengen wurden die vier Mengen mit den jeweils meisten potentiell interessanten Signaturen ausgewählt. Für jede dieser Mengen wurde der σ^{min} -Wert festgehalten.

Schritt 4: Auswahl und Bereinigung der Mengen mit den meisten potentiell interessanten Signaturen und Wiederholung des Data-Mining Schrittes.

1. Aus den vier Ergebnismengen des vorausgehenden Schrittes wurden vier Mengen $\Phi_{10}^{AEK_{PI}}$, $\Phi_{35}^{AEK_{PI}}$, $\Phi_{10}^{EK_{PI}}$ und $\Phi_{35}^{EK_{PI}}$ gebildet, welche ausschließlich potentiell interessante Signaturen enthielten. Aus den in den Mengen enthaltenen Signaturen wurden anschließend symbolische Elemente entfernt, deren nom -Werte < 2 waren.

2. Aus diesen vier Mengen von bereinigten potentiell interessanten Signaturen wurden in einem zweiten Data-Mining-Schritt durch ähnlichkeitsbasierte Verschmelzung vier Mengen $\Phi_{10}^{AEK_M}$, $\Phi_{35}^{AEK_M}$, $\Phi_{10}^{EK_M}$ und $\Phi_{35}^{EK_M}$ von verschmolzenen potentiell interessanten Signaturen generiert. Das Kriterium σ^{min} der für die Verschmelzung minimal erforderlichen Anzahl symbolischer Elemente entsprach dem in Schritt 3.3 festgehaltenen Wert. Die zeitliche Toleranz Δt_{max} betrug konstant 0.5. Die Toleranz auf der Frequenzachse Δf_{max} betrug konstant 0. Abschließend wurden aus den Signaturen in diesen Mengen symbolische Elemente entfernt, deren *nom*-Werte < 3 waren.

Schritt 5: Darstellung und Diskussion der Ergebnisse

Die nach der oben beschriebenen Vorgehensweise erhaltenen Mengen lassen in gewissem Umfang eine Beantwortung der eingangs gestellten Fragen zu. Durch Visualisierung der Ergebnisse mit Hilfe von Streudiagrammen kann die Lage von Clustern im Zeit-Frequenz-Raum im Überblick dargestellt werden. Ein visueller Vergleich der Diagramme zeigt Gemeinsamkeiten und Unterschiede zwischen synthetischen und natürlichen Pfifflauten. Von Interesse sind auch quantitative Parameter zur Kardinalität der verschiedenen Mengen, die Streurate in den Diagrammen, die Kompressionsrate und charakteristische Beispiele für die generierten Cluster.

6.4 Ergebnisse

Entsprechend der Ausführungen in Abschnitt 6.3 wurden $3 \cdot 2$ Mengen symbolischer Signaturen der Pfiffaute aus den FLC-Annotationen extrahiert:

1. Die Menge Φ_{10}^{AEK} enthielt 2795 Signaturen. Die Signaturen hatten eine durchschnittliche Länge von 20.6 Elementen. Pro Pfiffaut in A^{AEK} wurden im Durchschnitt 1.8 Signaturen generiert. Von den 97458 Annotationen in F^{AEK} wurden 57652 oder 59.2 % zur Bildung symbolischer Signaturen verwendet.
2. Die Menge Φ_{35}^{AEK} enthielt 818 Signaturen. Die Signaturen hatten eine durchschnittliche Länge von 47.9 Elementen. Pro Pfiffaut in A^{AEK} wurden im Durchschnitt 0.5 Signaturen generiert. Von den 97458 Annotationen in F^{AEK} wurden 39160 oder 40.2 % zur Bildung symbolischer Signaturen verwendet.
3. Die Menge Φ_{10}^{EK} enthielt 6726 Signaturen. Die Signaturen hatten eine durchschnittliche Länge von 12.7 Elementen. Von den 984033 Annotationen in F^{EK} wurden 85672 oder 8.7 % zur Bildung symbolischer Signaturen verwendet.
4. Die Menge Φ_{35}^{EK} enthielt 1072 Signaturen. Die Signaturen hatten eine durchschnittliche Länge von 39 Elementen. Von den 984033 Annotationen in F^{EK} wurden 42149 oder 4.3 % zur Bildung symbolischer Signaturen verwendet.

5. Die Menge Φ_{10}^{170} enthielt 137 Signaturen. Die Signaturen hatten eine durchschnittliche Länge von 13.6 Elementen. Pro Pfiff laut in der Testdatei wurden im Durchschnitt 0.8 Signaturen generiert. Von den 8806 Annotationen in F^{170} wurden 1868 oder 21.2 % zur Bildung symbolischer Signaturen verwendet.

6. Die Menge Φ_{35}^{170} enthielt 71 Signaturen. Die Signaturen hatten eine durchschnittliche Länge von 41.6 Elementen. Pro Pfiff laut in der Testdatei wurden im Durchschnitt 0.41 Signaturen generiert. Von den 8806 Annotationen in F^{170} wurden 2952 oder 33.5 % zur Bildung symbolischer Signaturen verwendet.

Im Folgenden werden die Mengen Φ_{10}^{AEK} , Φ_{35}^{AEK} , Φ_{10}^{EK} , Φ_{35}^{EK} und ihre Derivate in den Abbildungen 6.2-6.24 nacheinander dargestellt. Die Darstellung erfolgt mit Hilfe von zwei nebeneinander liegenden Streudiagrammen. Auf der linken Seite sind die Signaturen nach ihrem Frequenzmittelpunkt f_{mid} und ihrer Bandbreite Δf in das Diagramm projiziert. Auf der rechten Seite sind die Signaturen nach ihrem Frequenzmittelpunkt und ihrer Steigung m in das Diagramm projiziert. Die Kompressionsrate der Signaturen wird sowohl farbkodiert als auch als Zahl dargestellt. Treffen zwei Signaturen in einem Punkt zusammen, so wird die Kompressionsrate beider Signaturen addiert. Unter jedem Streudiagramm finden sich Angaben zur Anzahl der Signaturen einer dargestellten Menge Φ . Für beide Arten von Streudiagrammen wird ein Streukoeffizient $ST = x/|\Phi|$ angegeben. Der Wert x gibt dabei die Anzahl von Punkten im Diagramm an, denen mindestens eine Signatur zugeordnet wurde. Er ist relativ zu der Größe der Diagramme (400×400 -Punkte).

Am Beispiel der Menge Φ_{10}^{AEK} soll nun die Reihenfolge der Darstellungen veranschaulicht werden. Die Darstellung der Mengen Φ_{35}^{AEK} , Φ_{10}^{EK} , Φ_{35}^{EK} und ihrer Derivate erfolgt analog hierzu: (1) Für die Menge Φ_{10}^{AEK} wird das Verhalten des SSDM-Algorithmus in Abhängigkeit von σ^{min} in % dargestellt. (2) Die Menge Φ_{10}^{AEK} wird im Streudiagramm dargestellt. (3) Die Menge $\Phi_{10}^{AEK_{M50}}$ mit den meisten potentiell interessanten Signaturen wird im Streudiagramm dargestellt. (4) Die bereinigte Menge $\Phi_{10}^{AEK_{PI}}$ der potentiell interessanten Signaturen wird im Streudiagramm dargestellt. (5) Die bereinigte Menge $\Phi_{10}^{AEK_M}$ der verschmolzenen potentiell interessanten Signaturen wird im Streudiagramm dargestellt.

6.4.1 Die Menge Φ_{10}^{AEK}

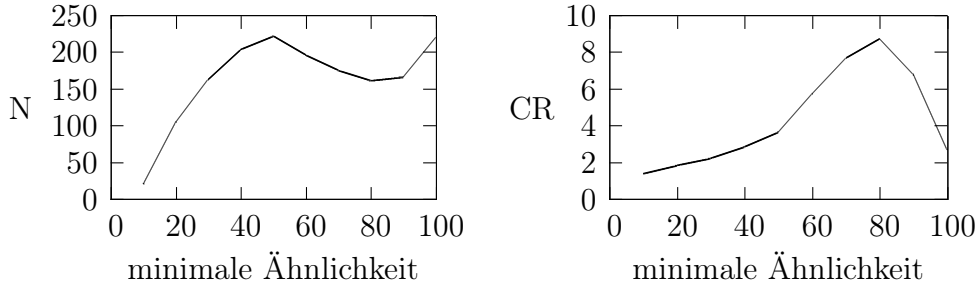


Abbildung 6.2: Verhalten des SSDM-Algorithmus für die Menge Φ_{10}^{AEK} in Abhängigkeit von σ^{min} in %. Links: Anzahl potentiell interessanter Sequenzen in den $\Phi_{10}^{AEK_{Mx}}$; Rechts: Kompressionsrate CR

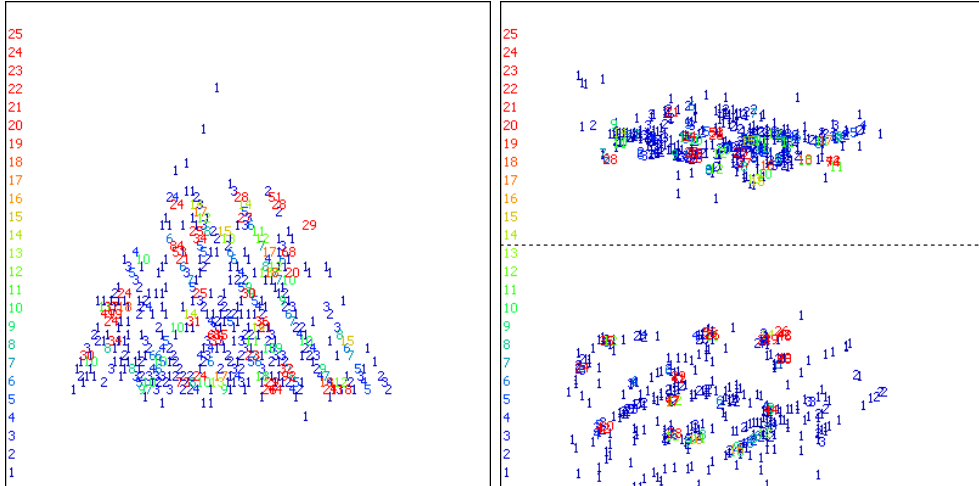


Abbildung 6.3: Streuung der Sequenzen in Φ_{10}^{AEK} mit 2795 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.14$; Rechts: $ST = 0.28$

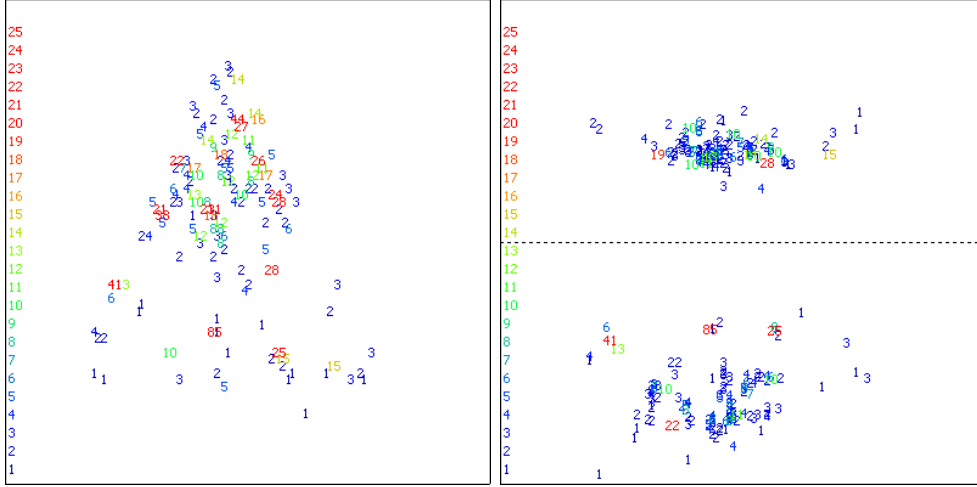


Abbildung 6.4: Streuung der Sequenzen in $\Phi_{10}^{AEK_{M50}}$ mit 283 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.47$; Rechts: $ST = 0.84$

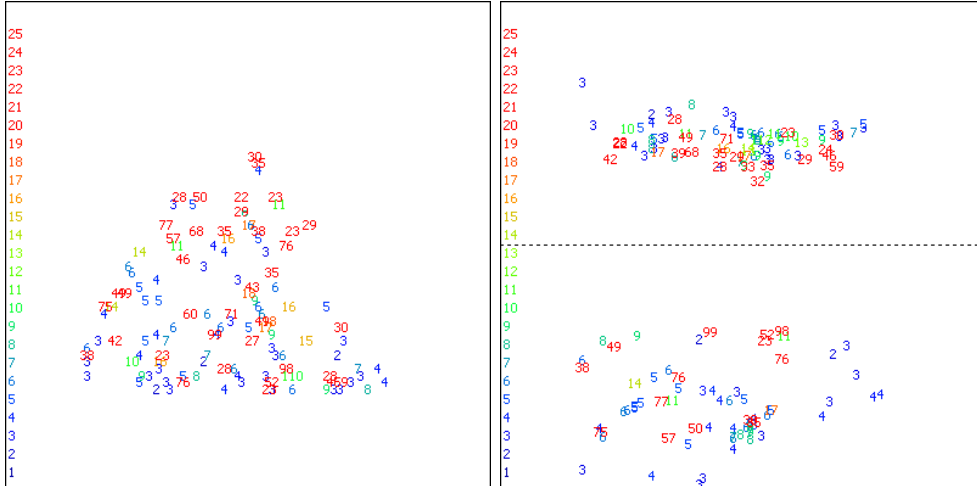


Abbildung 6.5: Streuung der Sequenzen in $\Phi_{10}^{AEK_{PI}}$ mit 161 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.77$; Rechts: $ST = 0.93$

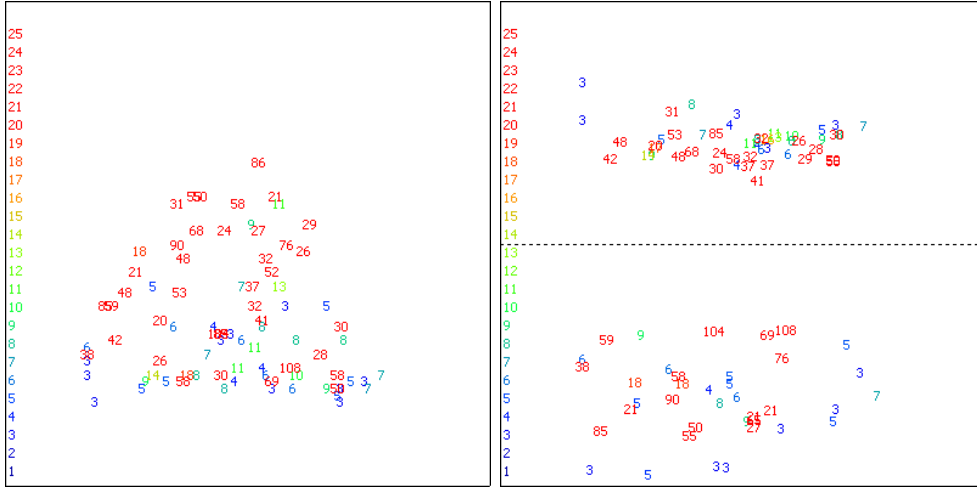


Abbildung 6.6: Streuung der Sequenzen in $\Phi_{10}^{AEK_M}$ mit 87 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.94$; Rechts: $ST = 1$

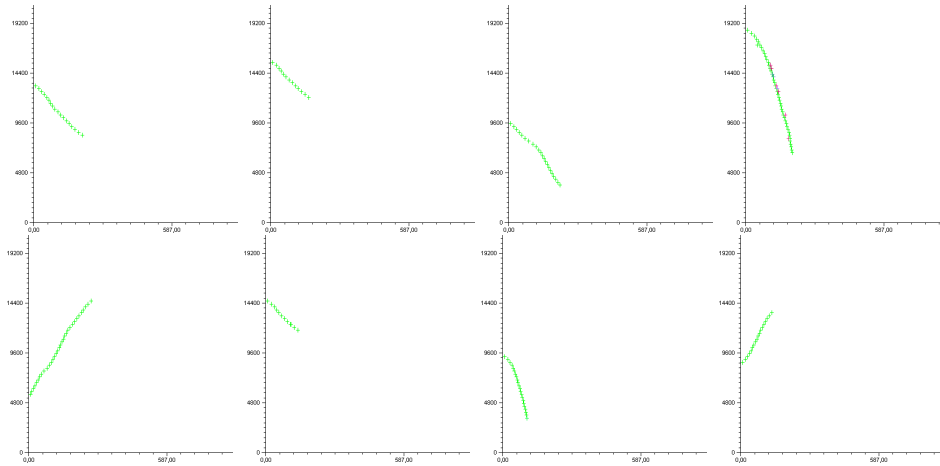


Abbildung 6.7: Elemente der Menge $\Phi_{10}^{AEK_M}$. Die Signaturen bilden Teile synthetischer Pfiffaute aus A^{PL} ab.

6.4.2 Die Menge Φ_{35}^{AEK}

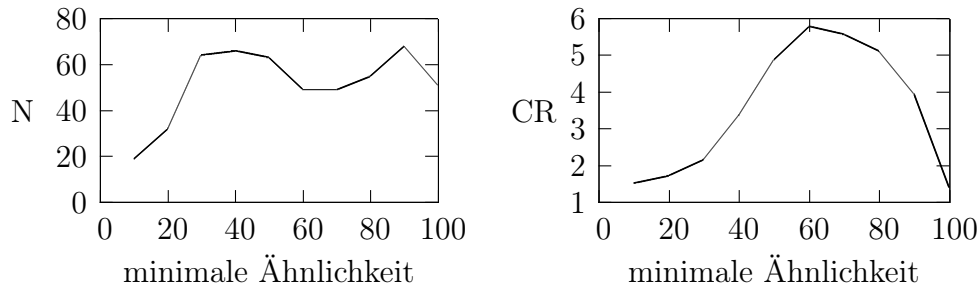


Abbildung 6.8: Verhalten des SSDM-Algorithmus für die Menge Φ_{35}^{AEK} in Abhängigkeit von σ^{min} in %. Links: Anzahl potentiell interessanter Sequenzen in den $\Phi_{35}^{AEK_{Mx}}$; Rechts: Kompressionsrate CR

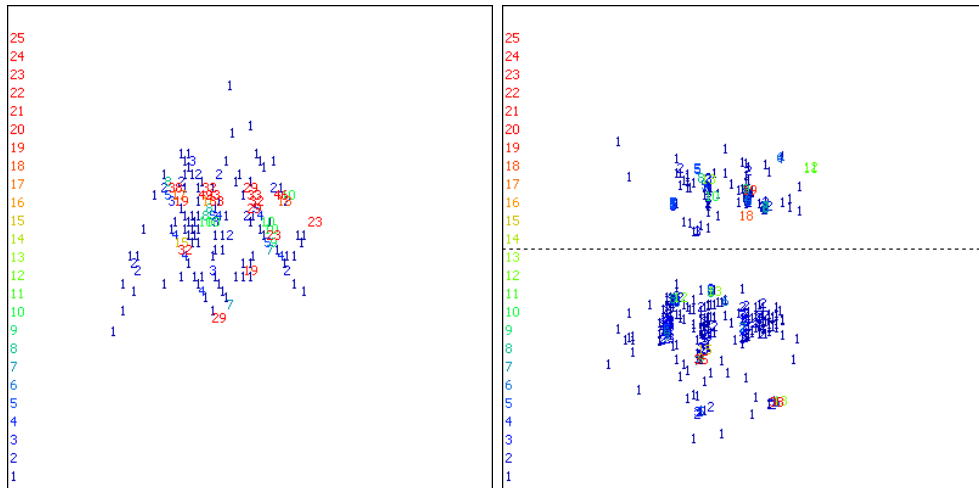


Abbildung 6.9: Streuung der Sequenzen in Φ_{35}^{AEK} mit 818 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.17$; Rechts: $ST = 0.44$

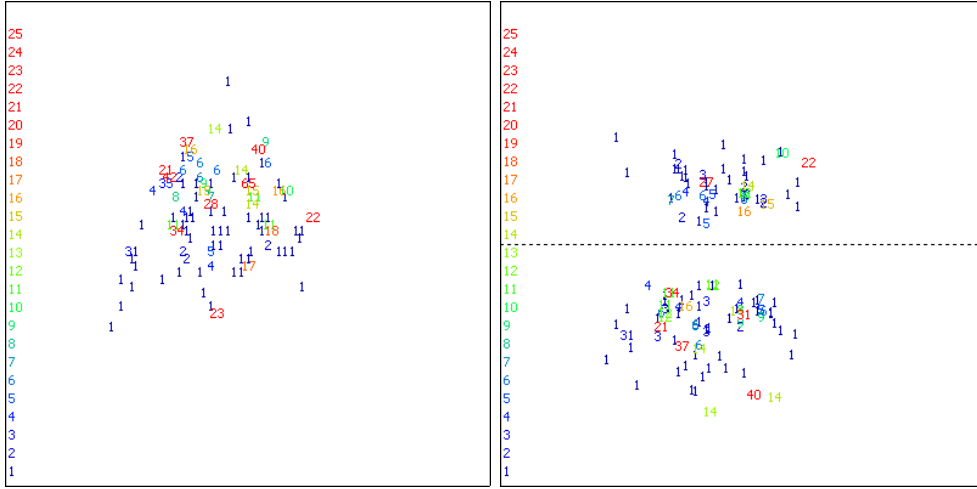


Abbildung 6.10: Streuung der Sequenzen in $\Phi_{35}^{AEK_{M80}}$ mit 136 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.74$; Rechts: $ST = 0.97$

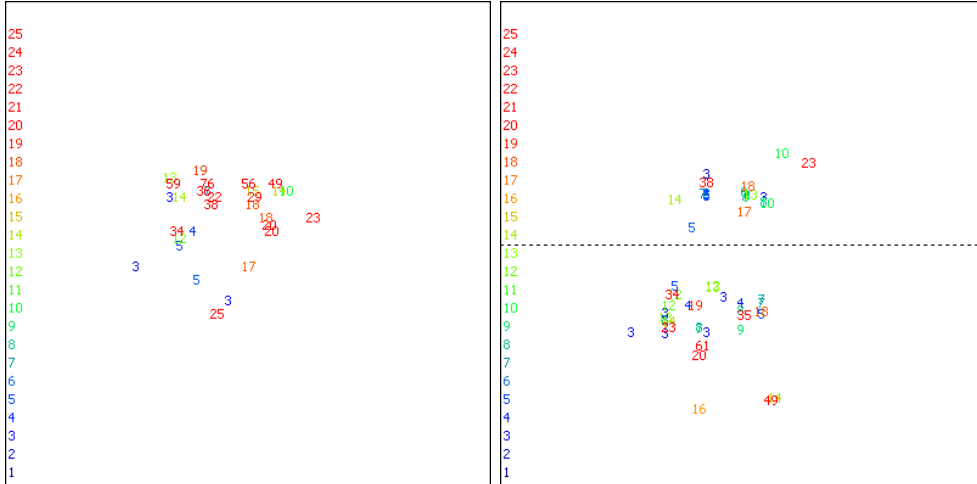


Abbildung 6.11: Streuung der Sequenzen in $\Phi_{35}^{AEK_{PI}}$ mit 55 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.54$; Rechts: $ST = 0.94$

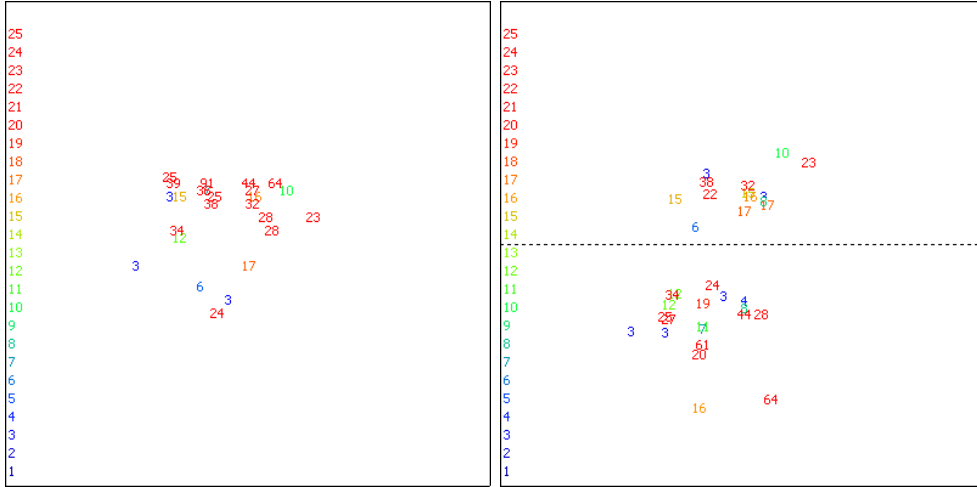


Abbildung 6.12: Streuung der Sequenzen in $\Phi_{35}^{AEK_M}$ mit 35 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.71$; Rechts: $ST = 0.97$

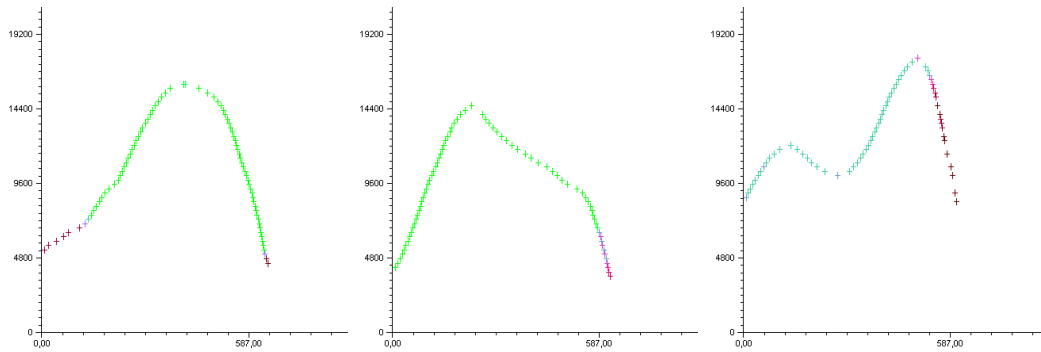


Abbildung 6.13: Elemente der Menge $\Phi_{35}^{AEK_M}$. Die Signaturen bilden synthetische Pfiffhöhe aus A^{PL} ab.

6.4.3 Die Menge Φ_{10}^{EK}

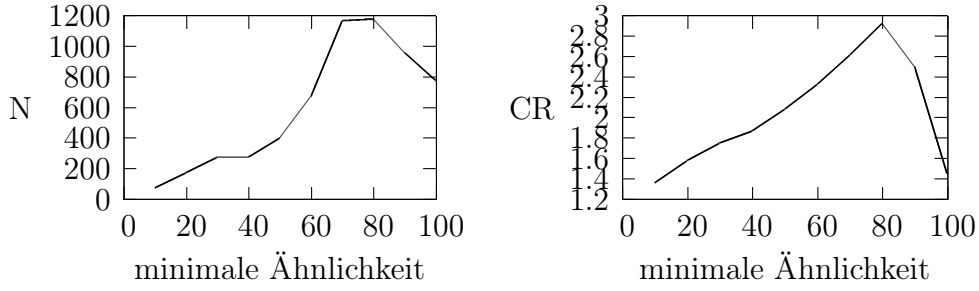


Abbildung 6.14: Verhalten des SSDM-Algorithmus für die Menge Φ_{10}^{EK} in Abhängigkeit von σ^{min} in %. Links: Anzahl potentiell interessanter Sequenzen in den $\Phi_{10}^{EK_{Mx}}$; Rechts: Kompressionsrate CR

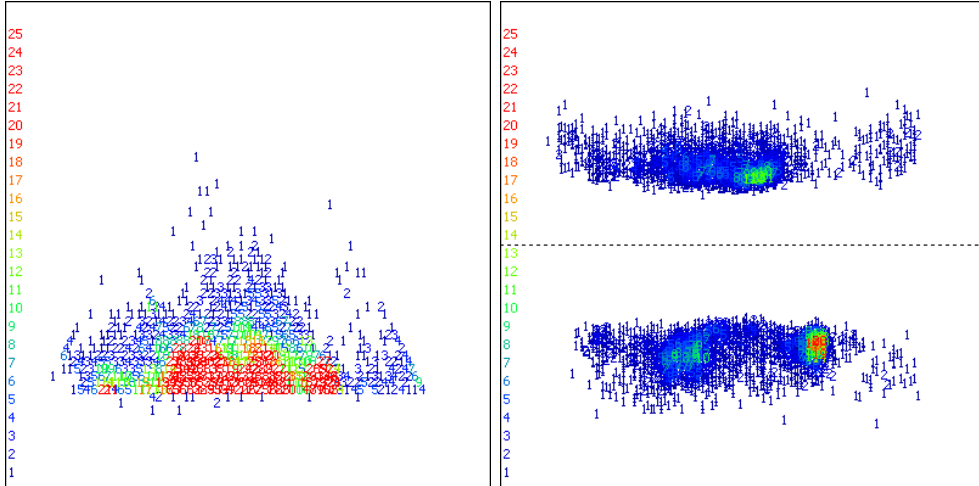


Abbildung 6.15: Streuung der Sequenzen in Φ_{10}^{EK} mit 6726 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.09$; Rechts: $ST = 0.48$

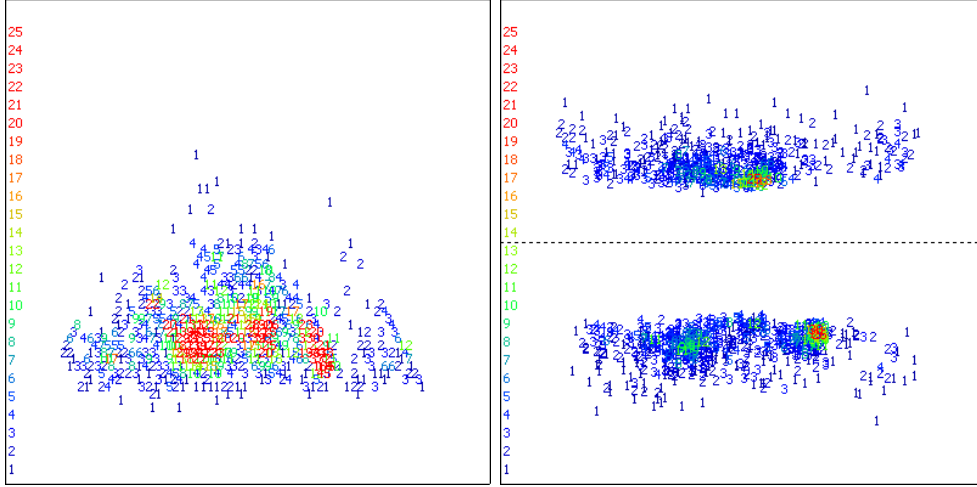


Abbildung 6.16: Streuung der Sequenzen in $\Phi_{10}^{EK_{M80}}$ mit 1508 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.35$; Rechts: $ST = 0.81$

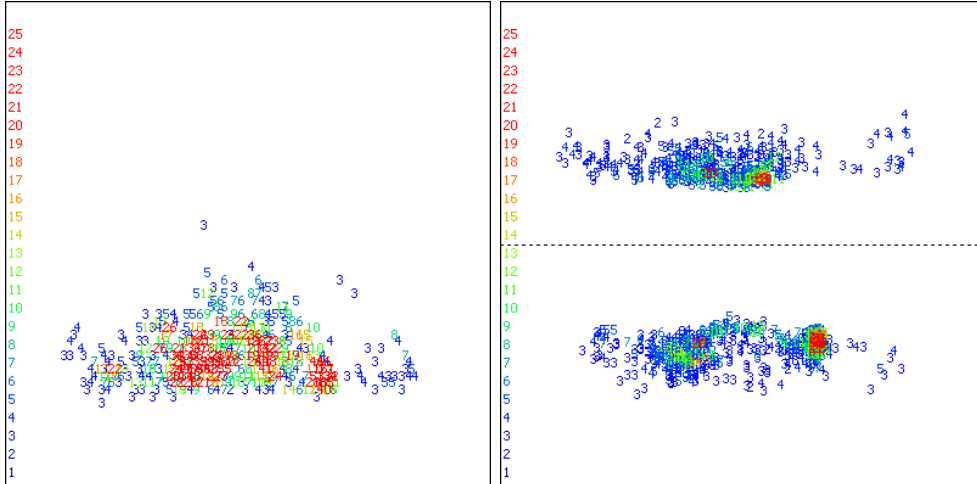


Abbildung 6.17: Streuung der Sequenzen in $\Phi_{10}^{EK_{PI}}$ mit 1181 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.30$; Rechts: $ST = 0.78$

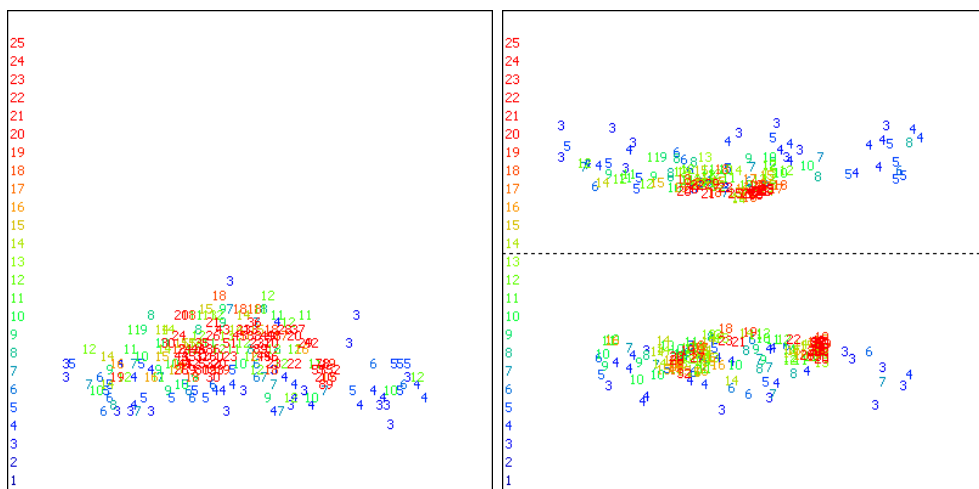


Abbildung 6.18: Streuung der Sequenzen in $\Phi_{10}^{EK_M}$ mit 329 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.63$; Rechts: $ST = 0.93$

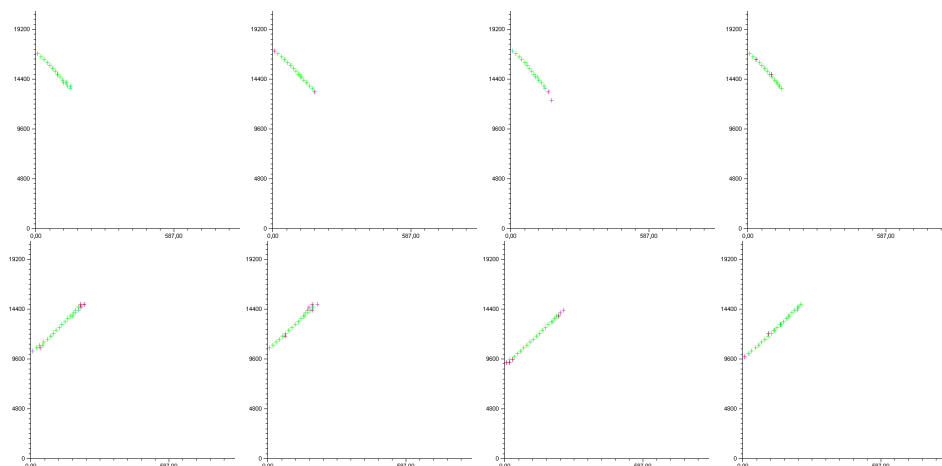


Abbildung 6.19: Elemente der Menge $\Phi_{10}^{EK_M}$. Die Elemente dieser Menge bilden geringfügige Fluktuationen in der Kontur natürlicher Pfifflaute ab. Die Fluktuationen sind erst bei genauerem Hinsehen als solche zu erkennen.

6.4.4 Die Menge Φ_{35}^{EK}

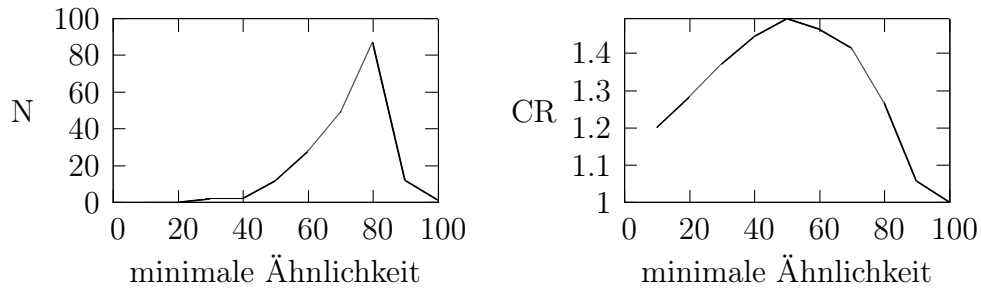


Abbildung 6.20: Verhalten des SSDM-Algorithmus für die Menge Φ_{35}^{EK} in Abhängigkeit von σ^{min} in %. Links: Anzahl potentiell interessanter Sequenzen in den $\Phi_{35}^{EK_{Mx}}$; Rechts: Kompressionsrate CR

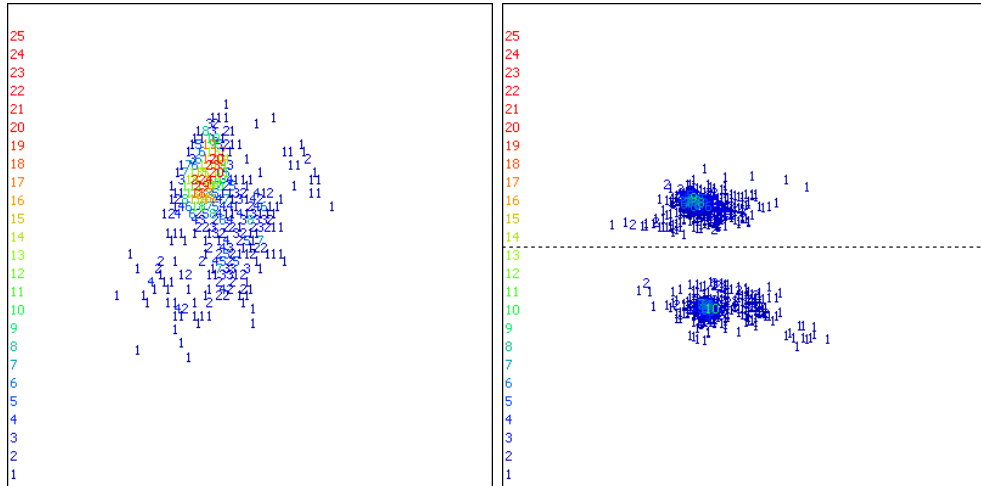


Abbildung 6.21: Streuung der Sequenzen in Φ_{35}^{EK} mit 1072 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.25$; Rechts: $ST = 0.59$

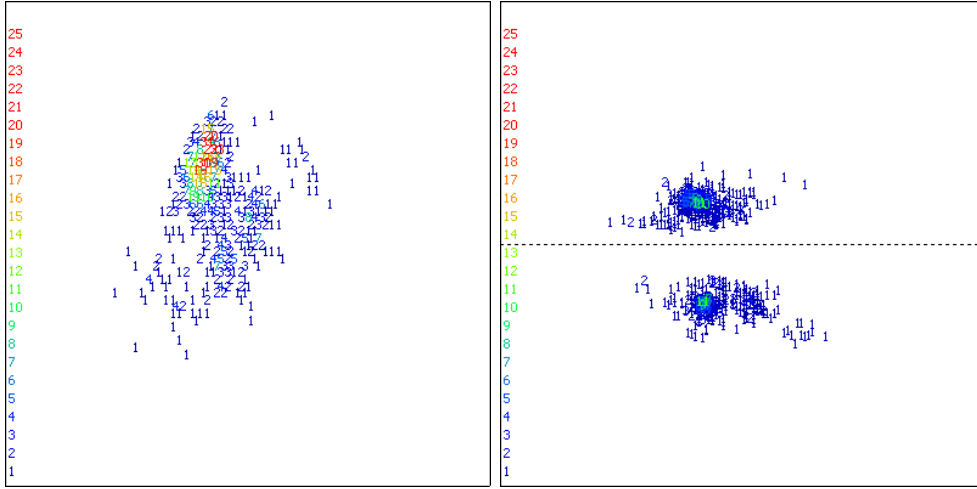


Abbildung 6.22: Streuung der Sequenzen in $\Phi_{35}^{EK_{M80}}$ mit 791 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.33$; Rechts: $ST = 0.70$

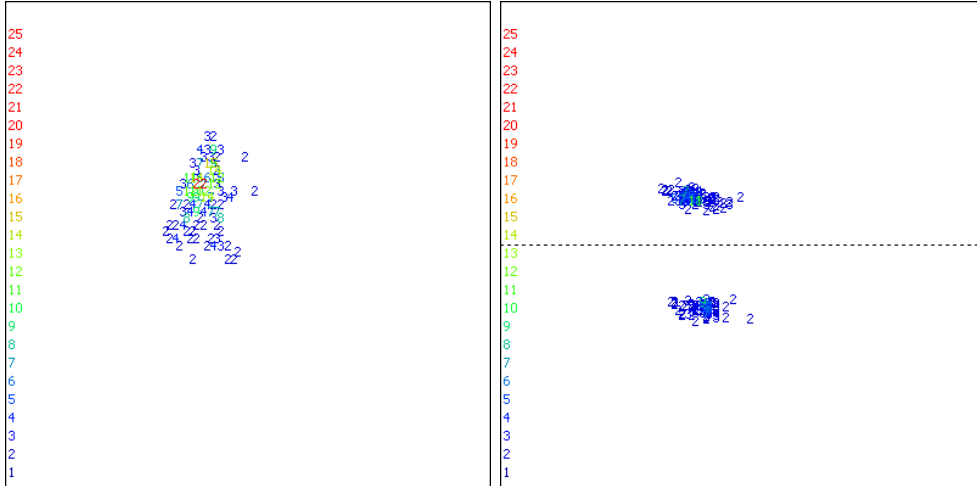


Abbildung 6.23: Streuung der Sequenzen in $\Phi_{35}^{EK_{PI}}$ mit 178 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.48$; Rechts: $ST = 0.83$

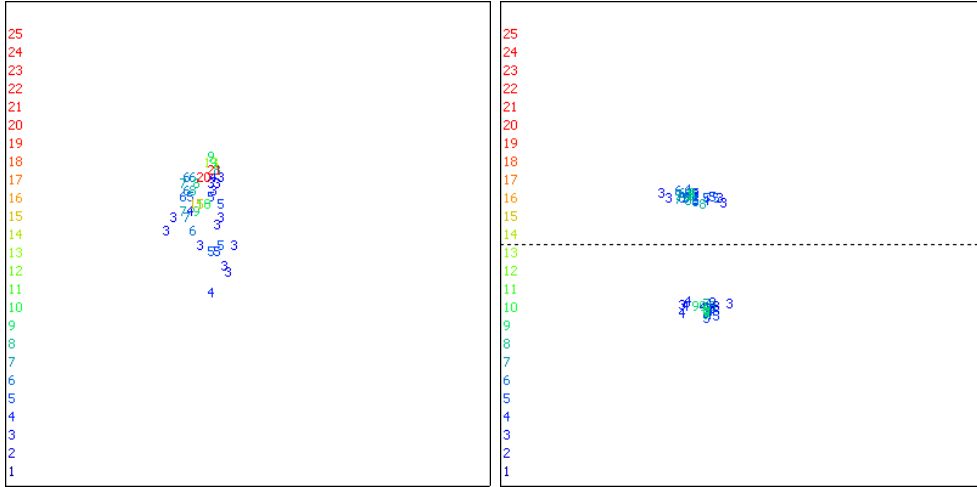


Abbildung 6.24: Streuung der Sequenzen in $\Phi_{35}^{EK_M}$ mit 52 Elementen. Links: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = Δf (2.5–17.5 kHz). Rechts: X-Achse = f_{mid} (2.5–17.5 kHz), Y-Achse = m . Die gestrichelte waagerechte Linie entspricht einer Steigung von 0. Links: $ST = 0.78$; Rechts: $ST = 0.96$

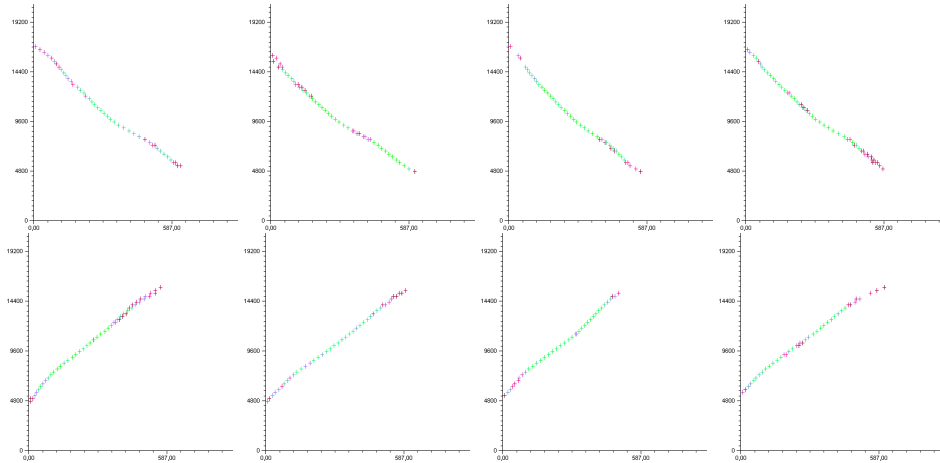


Abbildung 6.25: Elemente in $\Phi_{35}^{EK_M}$. Es finden sich viele Signaturen, die scheinbar eine hohe Ähnlichkeit untereinander aufweisen, vom SSDM-Algorithmus aber nicht verschmolzen werden.

6.4.5 Automatische Extraktion

Der Extraktionsalgorithmus generiert aus den Mengen F^{AEK} und F^{EK} zuverlässig symbolische Signaturen, welche die Konturen der Pfißlaute ganz oder in Teilen modellieren. In der Menge Φ_{35}^{AEK} finden sich Modelle aller Pfißtypen in A^{AEK} . In der Menge Φ_{35}^{EK} finden sich viele Modelle der Signaturpfißtypen, welche den einzelnen Tieren der Gruppe zugeschrieben werden (siehe Abbildung 6.26).

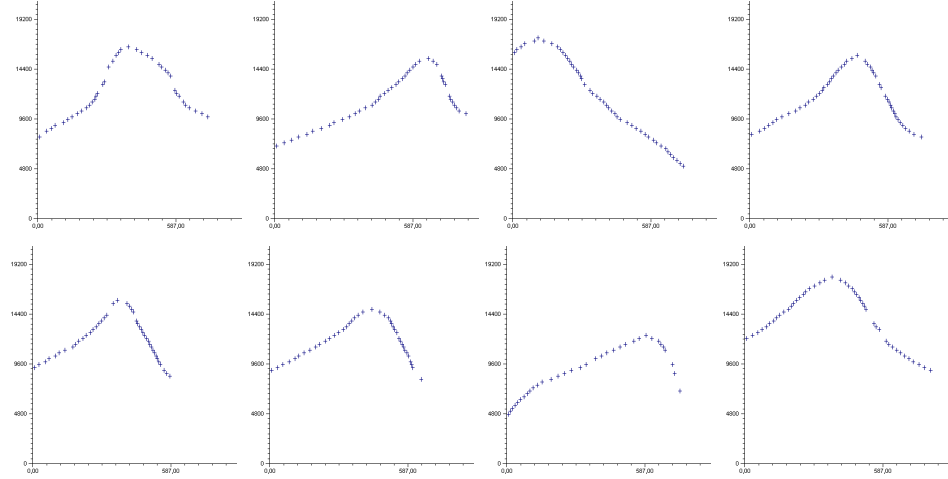


Abbildung 6.26: Symbolische Signaturen verschiedener Signaturpfiße in Φ_{35}^{EK}

Ein Nachteil der beschriebenen Vorgehensweise liegt darin, dass die FLC in K^{WL} nicht unabhängig von der Steigung des Frequenzverlaufs der Pfißlaute sind. Hierdurch werden die FLC-Annotationen unregelmäßig an die Kontur der Laute angepasst. Der Extraktionsalgorithmus ist aber in Abhängigkeit von den maximalen Abstandsparametern Δt_{max} und Δf_{max} gegenüber diesen Unregelmäßigkeiten tolerant. Bei restriktiven Einstellungen der Abstandsparameter werden mehrere Signaturen pro Pfißlaut generiert. Dies ist z.B. bei der Menge Φ_{10}^{AEK} deutlich zu sehen. Eine Herausforderung der Toleranzgrenzen der Abstandsparameter birgt die Gefahr, dass auch solche FLC-Annotationen in eine Signatur eingebunden werden, welche eigentlich nicht zu einem Pfißlaut gehören.

Die starken Unterschiede in der Verwertbarkeit der FLC-Annotationen zwischen synthetischen und natürlichen Daten geht vor allem auf die bessere Qualität der Pfißlaute in A^{AEK} zurück, da hier nur solche Dateien enthalten sind, für die im vorausgehenden Kapitel hohe *Accuracy*-Werte erzielt werden konnten.

Pro Pfißlaut in A^{AEK} sind in Φ_{10}^{AEK} im Durchschnitt 1.84 Signaturen enthalten. Pro Pfißlaut in A^{170} sind unter den gleichen Einstellungen jedoch im Durchschnitt nur 0.80 Signaturen enthalten. Dies kann als Hinweis darauf gesehen werden, dass die Pfißlaute in A^{170} im Durchschnitt *weniger* Änderungen im Konturverlauf besitzen, die unter den gegebenen Toleranzgrenzen das Verhalten des Extraktionsalgorithmus beeinflussen.

6.4.6 Verhalten des SSDM-Algorithmus

Die maximalen Kompressionsraten des SSDM-Algorithmus sind bei kürzeren Signaturen höher als bei längeren. Auch sind die maximalen Kompressionsraten bei synthetischen Daten höher als bei natürlichen. Auffallend ist, dass das Kompressionsverhalten des SSDM-Algorithmus für die Mengen Φ_{35}^{AEK} und Φ_{35}^{EK} ähnlich erscheint (vgl. Abbildungen 6.8 und 6.20). Eine weitere Besonderheit liegt darin, dass die maximale Anzahl potentiell interessanter Signaturen für drei der vier verwendeten Mengen bei einem σ^{min} -Wert von 80-90 % lag. Nur bei der Menge Φ_{10}^{AEK} lag der optimale σ^{min} -Wert bei 50 %.

6.4.7 Eigenschaften verschmolzener Signaturen

Eine Übersicht über das Streuverhalten der vier initialen Mengen und ihrer Derivate ist in den Tabellen 6.1 und 6.2 dargestellt. Auffallend hier ist, dass die aus den natürlichen Pffiflauten extrahierten Signaturen deutlich stärker streuen, als Signaturen synthetischer Laute. Der Streuwert 0.0944 bei Φ_{10}^{EK} ist ein Artefakt, welches durch die große Anzahl der Signaturen und die durch Rasterung der FLC in K^{WL} beschränkte Anzahl der möglichen Bandbreiten und Frequenzmittelpunkte zustandekommt.

Menge	ST Δf	ST m	Menge	ST Δf	ST m
Φ_{10}^{AEK}	0.1499	0.2805	Φ_{10}^{EK}	0.0944	0.4881
$\Phi_{10}^{AEK_{M50}}$	0.4770	0.8480	$\Phi_{10}^{EK_{M80}}$	0.3554	0.8163
$\Phi_{10}^{AEK_{PI}}$	0.7701	0.9378	$\Phi_{10}^{EK_{PI}}$	0.3031	0.7840
$\Phi_{10}^{AEK_M}$	0.9425	1.0000	$\Phi_{10}^{EK_M}$	0.6322	0.9300

Tabelle 6.1: Streuverhalten der Mengen mit $l_{min} = 10$ und ihrer Derivate

Menge	ST Δf	ST m	Menge	ST Δf	ST m
Φ_{35}^{AEK}	0.1760	0.4413	Φ_{35}^{EK}	0.2574	0.5932
$\Phi_{35}^{AEK_{M80}}$	0.7426	0.9705	$\Phi_{35}^{EK_{M80}}$	0.3388	0.7029
$\Phi_{35}^{AEK_{PI}}$	0.5454	0.9454	$\Phi_{35}^{EK_{PI}}$	0.4831	0.8314
$\Phi_{35}^{AEK_M}$	0.7142	0.9714	$\Phi_{35}^{EK_M}$	0.7884	0.9615

Tabelle 6.2: Streuverhalten der Mengen mit $l_{min} = 35$ und ihrer Derivate

Die aus den natürlichen Pffiflauten extrahierten Signaturen zeigen ein auffällig anderes Streumuster, als die Signaturen synthetischer Laute: Die Streudiagramme der Signaturen in Φ_{10}^{AEK} und Φ_{35}^{AEK} zeigen eine geringe Anzahl deutlich sichtbarer eng umrissener Häufungen (rote Bereiche)(vgl. Abbildungen 6.3 und 6.9). Im Gegensatz hierzu stehen die Streudiagramme der Signaturen in Φ_{10}^{EK} und Φ_{35}^{EK} . Hier lassen sich

nur wenige Häufungen erkennen. Diese sind zudem nicht scharf umrissen, sondern zeigen einen fließenden Übergang (orange-gelb-grün) zu den umliegenden Bereichen (vgl. Abbildungen 6.15 und 6.21).

Dieser Eindruck setzt sich bei den Mengen verschmolzener Signaturen (Clustern) fort: Im Falle der Signaturen synthetischer Pfißflaute konvergiert das Verfahren schnell zu einer Menge deutlich voneinander abgesetzter Cluster mit sehr hoher Kompressionsrate. Auffallend hierbei ist, dass die Anzahl der Elemente in $\Phi_{10}^{AEK_M}$ in etwa der Anzahl der in die synthetischen Pfißflaute zufällig eingestreuten spontanen Änderungen des Frequenzverlaufs entspricht (vgl. Abschnitt 5.4). Besonders klar zeigt sich die Konvergenz des Verfahrens auch in Abbildung 6.10, welche die Menge $\Phi_{35}^{AEK_{M50}}$ zeigt: Hier treten genau zehn deutlich voneinander abgesetzte Cluster in Erscheinung. Es verwundert nicht, dass sich in $\Phi_{35}^{AEK_M}$ tatsächlich Signaturen für alle zehn Typen synthetischer Pfißflaute finden (siehe Abbildung 6.10).

Ein gänzlich anderes Bild zeigt sich bei den Mengen verschmolzener Signaturen natürlicher Pfißflaute: Das verwendete Clusteringverfahren kann zwar die Streurate schnell erhöhen, so dass ähnlich hohe Werte erzielt werden, wie bei den Signaturen synthetischer Laute. Die entstehenden Cluster liegen jedoch weiterhin nahe beieinander und fließen ineinander über. Besonders auffallend ist auch die insgesamt geringe Anzahl von Bereichen mit starken Häufungen: Im Falle der Signaturen mit $l_{min} = 10$ lassen sich 4-5 derartige Bereiche ausmachen. Im Falle der Signaturen mit $l_{min} = 35$ lediglich 2-3 (!).

Inspiziert man die Elemente in $\Phi_{35}^{EK_M}$, so finden sich in ihr Signaturen, die scheinbar eine hohe Ähnlichkeit untereinander aufweisen, von dem verwendeten Algorithmus aber nicht mehr verschmolzen werden. Diese Signaturen können als Modelle von maximal vier verschiedenen Signaturpfißflauten gesehen werden (vgl. Abbildung 6.25).

Ein anderes Bild ergibt sich, wenn man die Elemente in $\Phi_{10}^{EK_M}$ inspiziert. Hierin findet sich eine Vielzahl von verschiedenen Elementen mit sehr hoher Kompressionsrate. Diese Elemente bilden zahlreiche geringfügige Fluktuationen in der Kontur natürlicher Pfißflaute ab. Diese Fluktuationen sind oftmals erst bei genauerem Hinsehen als solche zu erkennen. In Abbildung 6.19 ist diese Besonderheit veranschaulicht.

6.5 Diskussion

Janik (1999) verglich die Klassifikations-Performanz von Menschen mit drei maschinellen Verfahren. Dabei wurden 104 Pfißflaute verwendet. Er fand, dass durch manuelles Sortieren spektrographischer Darstellungen bessere Resultate erzielt werden konnten, als mit den maschinellen Verfahren. Deren schlechtes Abschneiden führte er darauf zurück, dass der den Verfahren zugrundeliegende Ähnlichkeitsvergleich alle Abschnitte einer Pfißkontur gleich gewichtete. Hierdurch wurden einerseits unwichtige Details überbewertet, andererseits wurden kurze aber entscheidende Änderungen des Konturverlaufs nicht adäquat gewichtet [59].

In dieser Arbeit werden alle Abschnitte der Konturen ebenfalls gleich gewichtet. Da hier aber eine sehr umfangreiche empirische Basis zugrunde gelegt wurde, sind dennoch Rückschlüsse auf allgemeine Charakteristika und Häufigkeiten der Pfifftypen im Eilat-Korpus möglich. Auch der Vergleich von natürlichen mit synthetischen, also idealisierten Signaturpfeifen zeigt interessante Aspekte dieser Lautklasse auf.

Es bestehen sehr große Unterschiede zwischen idealen Typen von Signaturpfeifen und natürlichen Pfiffauten. Von einem Zufallsmechanismus generierte und in Audiodateien verteilte Instanzen von synthetische Pfifftypen lassen sich mit dem beschriebenen Verfahren schnell und zuverlässig zu Clustern zusammenführen. In diesen können sowohl lokale als auch globale Merkmale der Kontur der verschiedenen Pfifftypen abgebildet werden. Im Gegensatz hierzu führt das SSDM-Verfahren nicht zu einer Bildung von separaten Clustern für *alle* hypothetisierten Signaturpfeife der Eilat-Gruppe. Betrachtet man die Häufigkeiten langer Signaturen ($l_{min} = 35$) natürlicher Pfiffaute, so zeigt sich eine starke Prävalenz von nur zwei Grundtypen. Diese können beide als Instanzen von Signaturpfeifen verstanden werden. Die Häufigkeit von Variationen dieser beiden Typen ist dabei höher als die Häufigkeit von Signaturpfeifen, die ihnen nicht entsprechen. Signaturen von Pfiffauten, die keinen Signaturpfeifen der Gruppe zuzuordnen sind, finden sich oft in der Menge $\Phi^{EK_{10}}$ und in ihren Derivaten. Auch lassen sich hier symbolische Signaturen von Teilen der meisten Signaturpfeife finden. Diese treten in verschiedenen Variationen und Häufigkeiten auf. Der SSDM-Algorithmus in Verbindung mit den hier verwendeten Distanzmaßen trifft viele Entscheidungen, die der menschlichen Intuition widersprechen. Insbesondere unterscheidet das Verfahren oft symbolischen Signaturen, die intuitiv betrachtet eine starke Ähnlichkeit aufweisen.

Betrachtet man symbolische Signaturen, die Details natürlicher Pfiffaute modellieren, so zeigt sich eine Vielzahl an Clustern mit hoher Kompressionsrate. Dies kann als typisches Ergebnis eines KDD-Prozesses gewertet werden. Es ist geeignet die Aufmerksamkeit auf Muster zu lenken, die überdurchschnittlich häufig in den Daten vorkommen, beinhaltet aber keine Aussage über die biologische Relevanz dieser Muster. Es deutet auf das Vorliegen von nicht zufälligen Variationen in den Pfiffauten hin. Die Frage, ob hierin biologisch relevante Informationen abgebildet sind, bedarf weiterführender empirischer Untersuchungen.

In der Repräsentationsform, in der die Beschreibungen der Cluster vorliegen, zeigt sich ein Vorzug des in Kapitel 3 dargestellten Ansatzes: Da es sich um symbolische Signaturen handelt, können sie nicht nur leicht visualisiert werden, sondern es ist möglich, sie *unverändert* für nachfolgende Analysen zu verwenden. Mit ihrer Hilfe kann in neuen Mengen von Audiodaten nach Okkurrenzen der durch sie modellierten Muster gesucht werden.

Kapitel 7

Zur Natur von MF-Klicklauten

In der Literatur finden sich bisher nur wenige Untersuchungen zu Klicklauten im mittleren Frequenzbereich (MF-Klicklaute). Entsprechend der Darstellung in Abschnitt 4.4.2 handelt es sich hierbei um einen der Grundtypen akustischer Phänomene im Eilat-Korpus. Die Laute dieser Klasse weisen zahlreiche Arten der phonetischen Modulation auf. Insbesondere das Phänomen der driftenden Pulsspektrum-Modulation (Pulsspektrum-Drift) wurde bisher in der Literatur noch nicht diskutiert. Lediglich Markov und Ostrovskaya (1990) weisen im Zusammenhang mit der Organisation des Kommunikationssystems von *Tursiops truncatus* auf dieses Phänomen hin [68].

7.1 Fragestellung

Nach Au (2004) variieren die Frequenzcharakteristika von Sonarlauten in Abhängigkeit davon, wo sich die Delphine aufhalten. Im Meer werden Sonarlaute mit signifikant anderen Eigenschaften emittiert, als in einem geschlossenen Becken. Au sieht die Ursache hierfür in einer Koppelung von Energie-Level und dominanter Frequenz des Signals:

Bottlenose dolphins emit short broadband clicks having peak frequencies as high as 120-130 kHz. Signals typically have 4 to 10 positive excursions and durations that vary from 40 to 70 μs . Peak-to-peak source levels between 210 and 227 dB re 1 mPa have been measured (...) Dolphins in tanks naturally emit much lower level signals with lower peak frequency. (...) The frequency spectrum of transmitted signals is coupled to the output level of the signals having higher frequency as the output level increases. Typical bandwidth is between 40 and 60 kHz [3].

Kamminga (1990) weist auf starke individuelle Unterschiede sowohl bei der dominanten Frequenz als auch bei der Bandbreite von Sonar-Klicklauten auf. In einer Studie an den Sonarlauten von sechs verschiedenen Individuen zeigt er Unterschiede in der dominanten Frequenz der Sonarlaute auf. Diese lag je nach Tier bei 80,

67, 80, 49, 55 und 44 kHz. Kamminga postuliert einen funktionalen Zusammenhang von dominanter Frequenz und Bandbreite der Klicklaute [64]. Bemerkenswerterweise zeigte sich auch in dieser Studie der oben beschriebene Zusammenhang: die beiden Tiere, welche sich während der Aufnahme im Meer befanden, waren auch diejenigen, welche Sonarlaute mit den höchsten dominanten Frequenzen (80 kHz) produzierten.

In Abschnitt 4.5.4 wurde darauf hingewiesen, dass MF-Klicklaute einen variablen Frequenzbereich zwischen 0.5-48 kHz abdecken. Die spektralen Charakteristika der einzelnen Klicklaute sowie gesamter Klick-Trains variieren im Zeitverlauf sehr stark (vgl. Abbildungen 4.35 – 4.37). Auch scheinen MF-Klicklaute geringere dominante Frequenzen aufzuweisen, als dies für Sonarlaute von *Tursiops truncatus* bisher angenommen wurde [2, 3, 64]. Es ergeben sich daher folgende Fragen:

1. Handelt es sich bei MF-Klicklauten um Sonarlaute oder um kommunikative Laute?
2. In welchem Bereich liegen die dominanten Frequenzen von MF-Klicklauten?

In den folgenden Abschnitten werden diese beiden Fragen untersucht. Ein besonderes Problem hierbei ist, dass es keine Möglichkeit gibt die Audiodaten mit nicht-akustischen Verhaltensweisen zu korrelieren. Es muss daher eine andere Methode verwendet werden: Deren grundlegende Idee besteht darin, die Emissionsdynamik von MF-Klicklauten mit der von eindeutig kommunikativen Lauten, d.h. von Pfiffen und Burst-Pulsed-Lauten, zu vergleichen. Wenn es sich bei MF-Klicklauten um kommunikative Laute handelt, so sollten sie eine ähnliche Emissionsdynamik aufweisen, wie die Laute der beiden anderen Klassen.

7.2 Materialien

Audiomaterialien:

- A^{EK} ist das gesamte Eilat-Korpus (vgl. Abschnitt 4.2).
- A^{MFK} ist ein kurzer Ausschnitt aus dem Eilat-Korpus, welcher vier Trains von MF-Klicklauten mit insgesamt ca. 500 Einzelklicks enthält.
- A^{BP} ist ein Ausschnitt aus dem Eilat-Korpus, welcher 79 Burst-Pulsed-Laute enthält.
- A^{TypA} ist eine Sammlung von Ausschnitten aus dem Eilat-Korpus, welche ca. 130 Pfiff-Laute enthält. Es handelt sich dabei um einen Pfiff-Typ der in Abbildung 6.24 im rechten Diagramm im Zentrum des Clusters mit *positiver* Steigung liegt (im Folgenden auch Pfiff-Typ A).
- A^{TypB} ist eine Sammlung von Ausschnitten aus dem Eilat-Korpus, welche ca. 100 Pfiff-Laute enthält. Es handelt sich dabei um einen Pfiff-Typ der in Abbildung 6.24 im rechten Diagramm im Zentrum des Clusters mit *negativer* Steigung liegt (im Folgenden auch Pfiff-Typ B).

FLC-Annotationen:

- F^{EK} ist die Menge von FLC-Annotationen von K^{WL} mit 984033 Elementen aus Kapitel 6.

Klassifikatoren:

- κ^{MFK} ist ein FLC, welcher MF-Klicklaute in A^{MFK} in einem Frequenzbereich von 0.5-48 kHz und einer Klick-Wiederholrate bis 1000 Hz zuverlässig annotiert.
- κ^{BP} ist ein FLC, welcher Burst-Pulsed-Laute in A^{BP} in einem Frequenzbereich von 12-48 kHz zuverlässig annotiert.
- κ^A ist ein FLC, welcher Pfifftyp A in A^{TypA} zuverlässig annotiert.
- κ^B ist ein FLC, welcher Pfifftyp B in A^{TypB} zuverlässig annotiert.

Modellierung von κ^{MFK} :

1. Mit Hilfe des in Abschnitt 3.3.1 beschriebenen Verfahrens wurde ein initialer FLC κ^{Init} für einen MF-Klicklaut in A^{MFK} generiert.
2. Mit Hilfe von κ^{Init} wurde A^{MFK} annotiert, so dass eine Menge F^{Init} von FLC-Annotationen entstand.
3. Aus F^{Init} wurden manuell alle korrekten Annotationen ausgewählt. Aus diesen wurde mit Hilfe des Verfahrens in Abschnitt 3.3.8 eine Menge elementarer spektrographischer Signaturen extrahiert. Diese wurden dann mit Hilfe des Verfahrens in Abschnitt 3.3.9 zu einer einzigen Signatur ϕ^M verschmolzen.
4. Aus ϕ^M wurde ein neuer FLC generiert, indem ϕ^M mit der KKF und dem AA von κ^{Init} zu einem neuen FLC κ^{MFK} verknüpft wurde. Abschließend wurde κ^{MFK} manuell so justiert, dass er bei der Annotation von A^{MFK} einen Accuracy-Wert von über 95% erzielte (siehe auch Abbildung 7.1).

Modellierung von κ^A und κ^B :

1. Mit Hilfe des in Abschnitt 3.3.1 beschriebenen Verfahrens wurde manuell aus einem positiven Beispiel ein FLC κ^{Init} für den Pfifftyp A generiert. Der FLC beruht auf einem logarithmisierten Spektrogramm (vgl. Abschnitt 3.2).
2. Mit Hilfe von κ^{Init} wurde A^{TypA} annotiert, so dass eine Menge F^{Init} von FLC-Annotationen entstand. Hierbei wurde die Zeit gemessen, welche zu Annotation benötigt wurde. Anschließend wurden die Eigenschaften von κ^{Init} anhand der Maße in Abschnitt 5.3 bewertet.

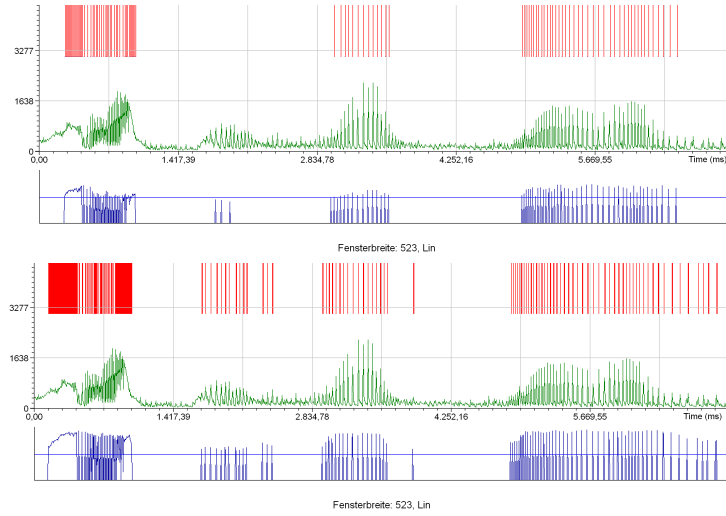
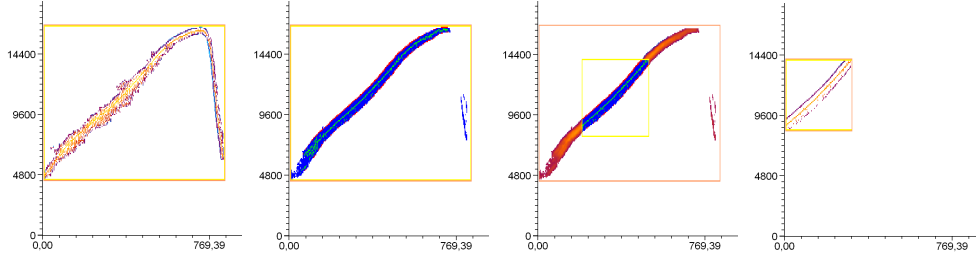


Abbildung 7.1: FLC für MF-Klicklaute. Oben: Performanz von κ^{Init} . Unten: Performanz von κ^{MFK} . Der Energie-Graph des Audiosignals ist grün dargestellt. Der Graph der KKF ist blau dargestellt. Die FLC-Annotationen sind als rote Linien in das Diagramm projiziert. Durch die hohe Klick-Wiederholrate erscheinen sie teilweise als rote Fläche. Nur der optimierte FLC κ^{MFK} erkennt alle vier Trains in dem Ausschnitt.

3. Nun wurde κ^{Init} im Rahmen des FLCD-Prozesses wie folgt optimiert: Aus F^{Init} wurden alle korrekten Annotationen manuell ausgewählt. Aus diesen wurde mit Hilfe des Verfahrens in Abschnitt 3.3.8 eine Menge elementarer spektrographischer Signaturen extrahiert. Die so extrahierten Signaturen wurden mit Hilfe des Verfahrens in Abschnitt 3.3.9 zu einer einzigen Signatur ϕ^M verschmolzen.
4. Aus ϕ^M wurde der zentrale Bereich mit den geringsten Abweichungen in der Zeit-Frequenz-Kontur ausgewählt. Hieraus wurde eine neue Signatur generiert, welche mit der KKF und dem AA von κ^{Init} zu einem neuen FLC κ^A verknüpft wurde
5. Mit Hilfe von κ^A wurde A^{TypA} annotiert, so dass die Menge F^{TypA} entstand. Hierbei wurde die Zeit gemessen, welche zu Annotation benötigt wurde. Anschließend wurde die Eigenschaften von κ^A anhand der Maße in Abschnitt 5.3 bewertet.

Das Ergebnis dieses Prozesses ist in Tabelle 7.1 dargestellt. Der nicht-optimierte FLC benötigt 4,4 Sekunden zur Annotation von A^{TypA} mit 32 MB (System: Pentium 4, 3 GHz, 2 GB RAM). Er findet 77 von 130 Piffen. Der optimierte FLC benötigt für dieselbe Aufgabe 1,2 Sekunden. Er findet 125 von 130 Piffen. Der FLC κ^B wurde auf die gleiche Art und Weise modelliert.



Abbildungung 7.2: Modellierung eines FLC für Pffiffaute. Von links nach rechts: (1) Signatur von κ^{Init} ; (2) Verschmolzene Signatur ϕ^M ; (3) Auswahl des zentralen Bereichs in ϕ^M ; (4) Fertige Signatur von κ^A .

	Zeit	a	Accuracy	Precision
κ^{Init}	4.4 s	77	0.59	1
κ^A	1.2 s	125	0.96	1

Tabelle 7.1: Performanz von nicht-optimiertem und optimiertem FLC für den Pffiff-typ A: a = positiv klassifizierte und tatsächlich positive Laute.

Modellierung von κ^{BP} :

1. Mit Hilfe des in Abschnitt 3.3.1 beschriebenen Verfahrens wurde manuell aus einem positiven Beispiel ein FLC κ^{Init} für einen BP-Laut generiert. Der FLC beruht auf einem logarithmisierten Spektrogramm (vgl. Abschnitt 3.2).
2. Mit Hilfe von κ^{Init} wurde A^{BP} annotiert, so dass eine Menge F^{Init} von FLC-Annotationen entstand.
3. Nun wurde κ^{Init} im Rahmen des FLCD-Prozesses wie folgt optimiert: Aus F^{Init} wurden alle korrekten Annotationen manuell ausgewählt. Aus diesen wurde mit Hilfe des Verfahrens in Abschnitt 3.3.8 eine Menge elementarer spektrographischer Signaturen extrahiert. Die so extrahierten Signaturen wurden mit Hilfe des Verfahrens in Abschnitt 3.3.9 zu einer einzigen Signatur ϕ^M verschmolzen.
4. Aus ϕ^M wurde der zentrale Bereich mit den geringsten Abweichungen der Zeit-Frequenz-Kontur ausgewählt. Hieraus wurde eine neue Signatur generiert, welche mit der KKF und dem AA von κ^{Init} zu einem neuen FLC κ^{BP} verknüpft wurde.
5. Mit Hilfe von κ^{BP} wurde A^{BP} annotiert, so dass die Menge F^{BP} entstand. Anschließend wurde die Eigenschaften von κ^{BP} anhand der Maße in Abschnitt 5.3 bewertet. Dieser Vorgang wurde solange wiederholt, bis der FLC einen

Accuracy-Wert > 0.95 für A^{BP} generierte. Abschließend wurde der Merkmalsvektor nach dem Verfahren in Abschnitt 3.3.6 aufgedünnt.

Der fertige FLC umfasst zwei Bereiche: (1) Einen Bereich unmittelbar vor dem eigentlichen BP-Laut, der keine energiereichen Laute enthalten darf. Diesem entspricht ϕ^- . (2) Einen Bereich für den Beginn des BP-Lautes, der hohe Energiedichten aufweisen muss. Diesem entspricht ϕ^+ . Der FLC generiert nur dort Annotationen, an denen sowohl die KKF von ϕ^- als auch die KKF von ϕ^+ hohe Werte generieren. Hierdurch wird erreicht, dass MFK-Trains mit hoher Pulswiederholrate von dem FLC ignoriert werden. Auch wird so jeder BP-Laut nur maximal einmal annotiert.

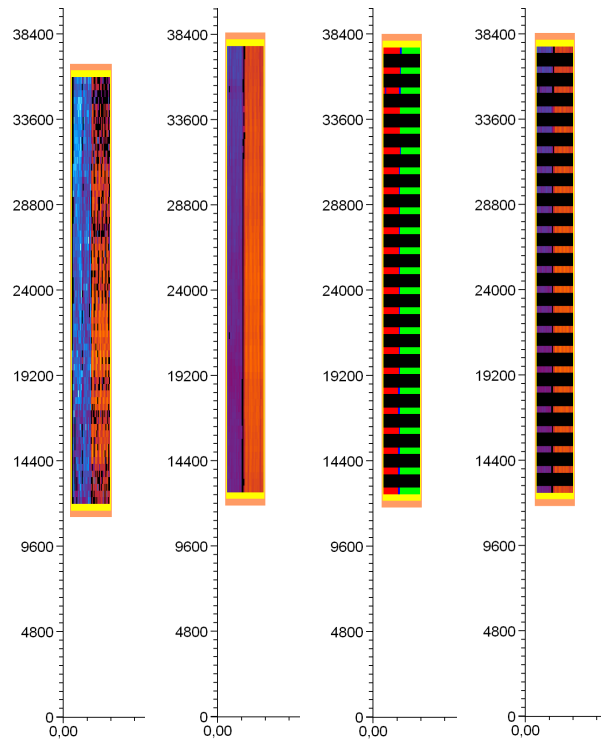


Abbildung 7.3: Modellierung eines FLC für Burst-Pulsed-Laute. Von links nach rechts: (1) Signatur von κ^{Init} ; (2) Verschmolzene Signatur ϕ^M ; (3) κ^{BP} in Kontrastfarben: rot ist ϕ^- und grün ist ϕ^+ ; (4) Fertige Signatur von κ^{BP} .

7.3 Methode

Schritt 1: Generieren der Annotationsmengen

Zunächst wurde mit Hilfe von κ^{MFK} über dem gesamten Eilat-Korpus eine Menge von FLC-Annotationen erzeugt. Es entstand

- die Menge F^{MFK} mit 2556086 Elementen. Jedem Element entspricht dabei genau ein Klicklaut.

Da F^{MFK} sehr viele Annotationen von Garnelen-Klicklauten enthielt, wurden hieraus zwei weitere Mengen abgeleitet:

- Die Menge F'^{MFK} mit 2246845 Elementen, welche nur Annotationen von Klicklauten mit einer dominanten Frequenz > 10.05 kHz enthält. Da die dominante Frequenz der Klicklaute von Garnelen meist unterhalb von 5 KHz liegt, sind in F'^{MFK} fast ausschließlich Annotationen von Delphinlauten enthalten.
- Die Menge F^{MFT} mit 38280 Elementen, welche SLC-Annotationen für Trains von MF-Klicklauten enthält. Diese Menge wurde mit Hilfe des Verfahrens in Abschnitt 3.4.4 zu automatischen Extraktion symbolischer Signaturen generiert. Den SLC-Annotationen entsprechen Sequenzen zeitlich eng benachbarter FLC-Annotationen in F^{MFK} . In F^{MFT} sind nur Trains mit mindestens 12 Einzellauten enthalten. Hierdurch wird ausgeschlossen, dass kurze Sequenzen zufällig benachbarter Klicklaute von Garnelen in der Menge enthalten sind. Die durchschnittliche Anzahl von MF-Klicklauten in einem Train lag bei ca. 59 Einzelklicks.

Anschließend wurde mit Hilfe von κ^{BP} , κ^A , κ^B über dem gesamten Eilat-Korpus drei weitere Mengen von FLC-Annotationen erzeugt:

- Die Menge F^{BP} mit 11015 Elementen. Jedem Element entspricht dabei ein BP-Laut.
- Die Menge F^A mit 2339 Elementen. Jedem Element entspricht dabei ein Pfiff laut vom Typ A.
- Die Menge F^B mit 4470 Elementen. Jedem Element entspricht dabei ein Pfiff laut vom Typ B.

Da in die Statistik die Pfiff laute aller Mitglieder der Gruppe einfließen sollten, wurde aus F^{EK} mit Hilfe des Verfahrens in Abschnitt 3.4.4 eine Menge symbolischer Signaturen generiert, welche Pfiff laute modellieren (siehe auch voriges Kapitel). Hierbei wurden für die Extraktion folgende Parameter festgesetzt: $l_{min} := 15$, $\Delta t_{min} := 0.1$, $\Delta t_{max} := 2$, $\Delta f_{min} := 0.1$, $\Delta f_{max} := 1$. Es entstand eine zweite Menge von SLC-Annotationen:

- Die Menge F^{PL} mit 12280 Elementen.

Allen Elementen in den verschiedenen Mengen entsprechen dabei nur Above-Threshold-Ereignisse. Die absolute Anzahl von Lauten der verschiedenen Lautklassen im Eilat-Korpus wird von ihnen nicht wiedergegeben.

Schritt 2: Erstellen von Verteilungsprofilen

Für jede der oben beschriebenen Mengen wurden vier Verteilungsprofile erstellt:

1. Ein Profil, welches die absolute Anzahl der Ereignisse nach Aufnahmetagen widerspiegelt.
2. Ein Profil, welches die relative Dichte pro Stunde (p.h.) der Ereignisse normiert im Intervall der Gesamtaufnahmedauer des jeweiligen Aufnahmetages nach Aufnahmetagen widerspiegelt.
3. Ein Profil, welches die absolute Anzahl der Ereignisse im 10 Minuten-Intervall im Tagesverlauf von 10-22 Uhr widerspiegelt.
4. Ein Profil, welches die relative Dichte p.h. der Ereignisse normiert im 10 Minuten-Intervall im Tagesverlauf von 10-22 Uhr widerspiegelt.

Schritt 3: Auswertung der Dynamik der dominanten Frequenzen von MF-Klicklauten im Tagesverlauf

Für die oben beschriebenen Mengen F^{MFK} und F'^{MFK} wurde die Dynamik der dominanten Frequenzen von MF-Klicklauten im Tagesverlauf von 10:00 bis 22:00 ausgewertet. Hierdurch wird ersichtlich, in welchem Bereich die dominanten Frequenzen von MF-Klicklauten liegen und ob sich hierbei Unterschiede im Tagesverlauf zeigen.

Schritt 4: Darstellung und Diskussion der Ergebnisse

Die nach der oben beschriebenen Vorgehensweise erhaltenen Mengen lassen in gewissem Umfang eine Beantwortung der eingangs gestellten Fragen zu. Durch Visualisierung der Ergebnisse mit Hilfe von Diagrammen kann die Dynamik der akustischen Aktivität im Bereich der verschiedenen Lautklassen nachvollzogen werden. Weiter unten wird ersichtlich, wie durch einen Vergleich der Profile untereinander Rückschlüsse auf die grundlegende Natur von MF-Klicklauten möglich werden.

7.4 Ergebnisse

Die Ergebnisse dieser Untersuchung sind in den Abbildungen der folgenden Seiten dargestellt.

7.5 Diskussion

Die Verteilungsprofile der MF-Klicklaute zeigen für alle drei Mengen F^{MFK} , F'^{MFK} und F^{MFT} eine deutliche Zunahme der akustischen Aktivität in den Abendstunden

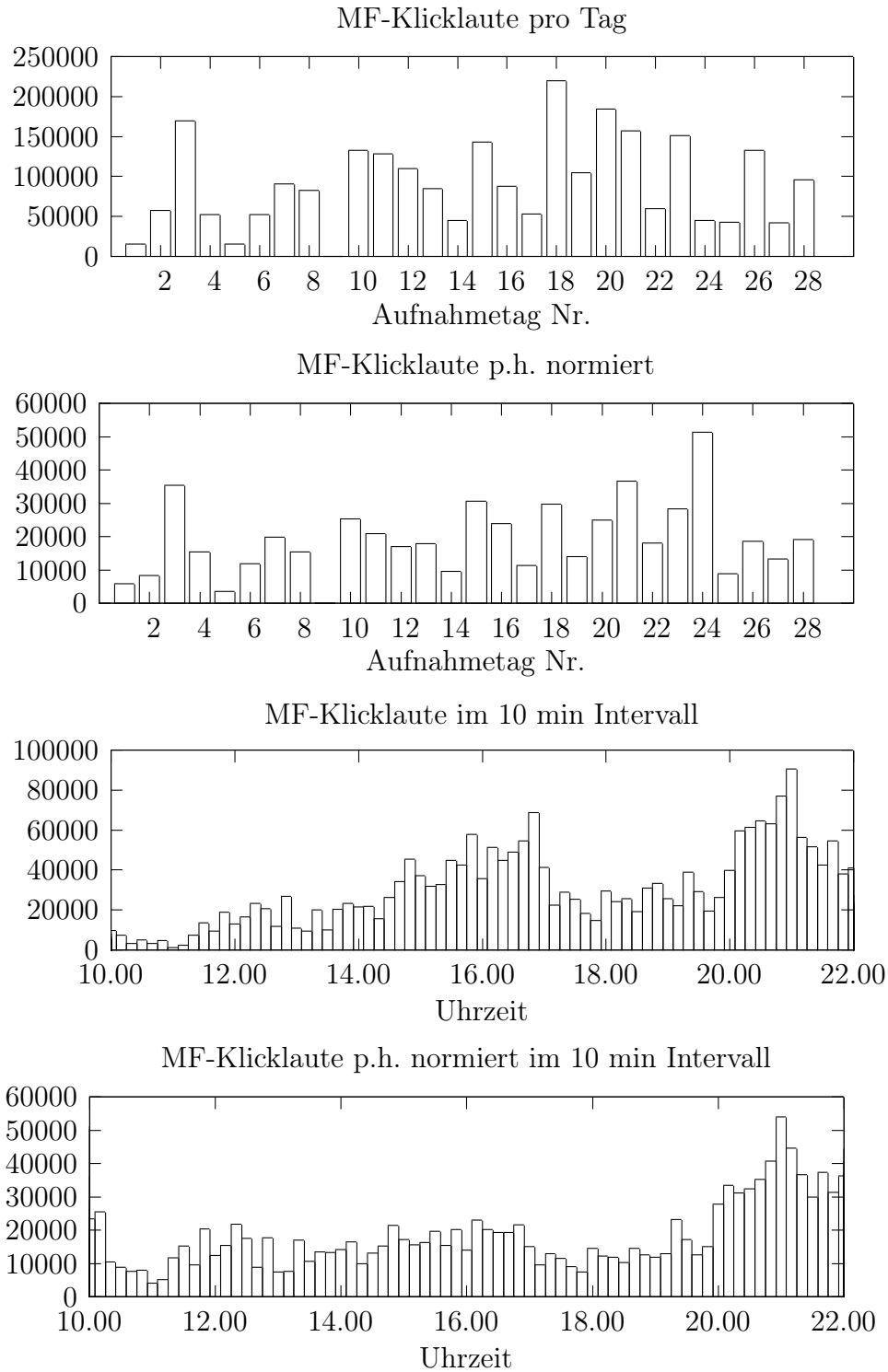


Abbildung 7.4: MF-Klicklaute im Eilat-Korpus entsprechend der Menge F^{MFK} mit 2556086 Elementen

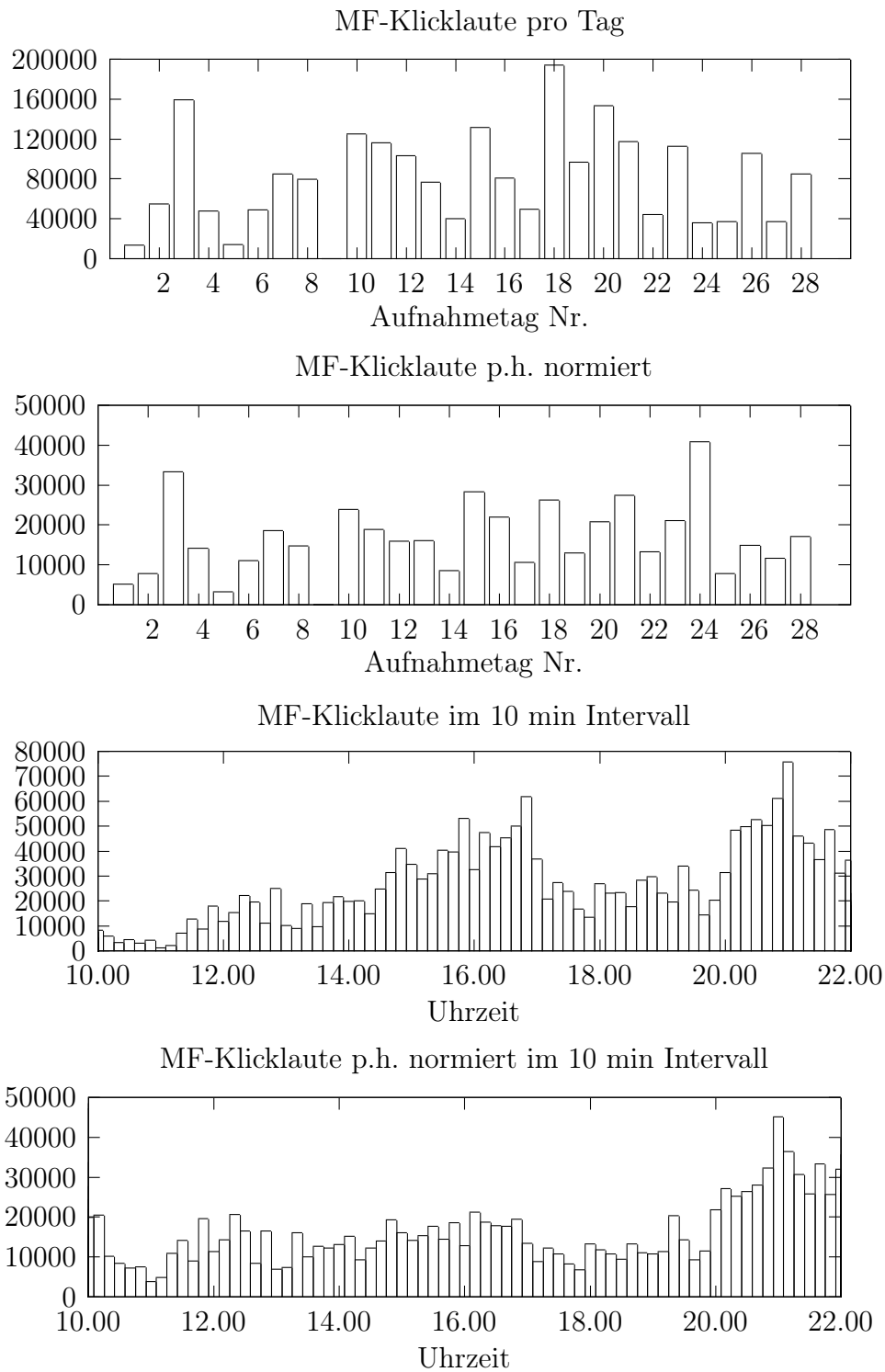


Abbildung 7.5: MF-Klicklaute im Eilat-Korpus entsprechend der Menge F^{MFK} mit 2246845 Elementen

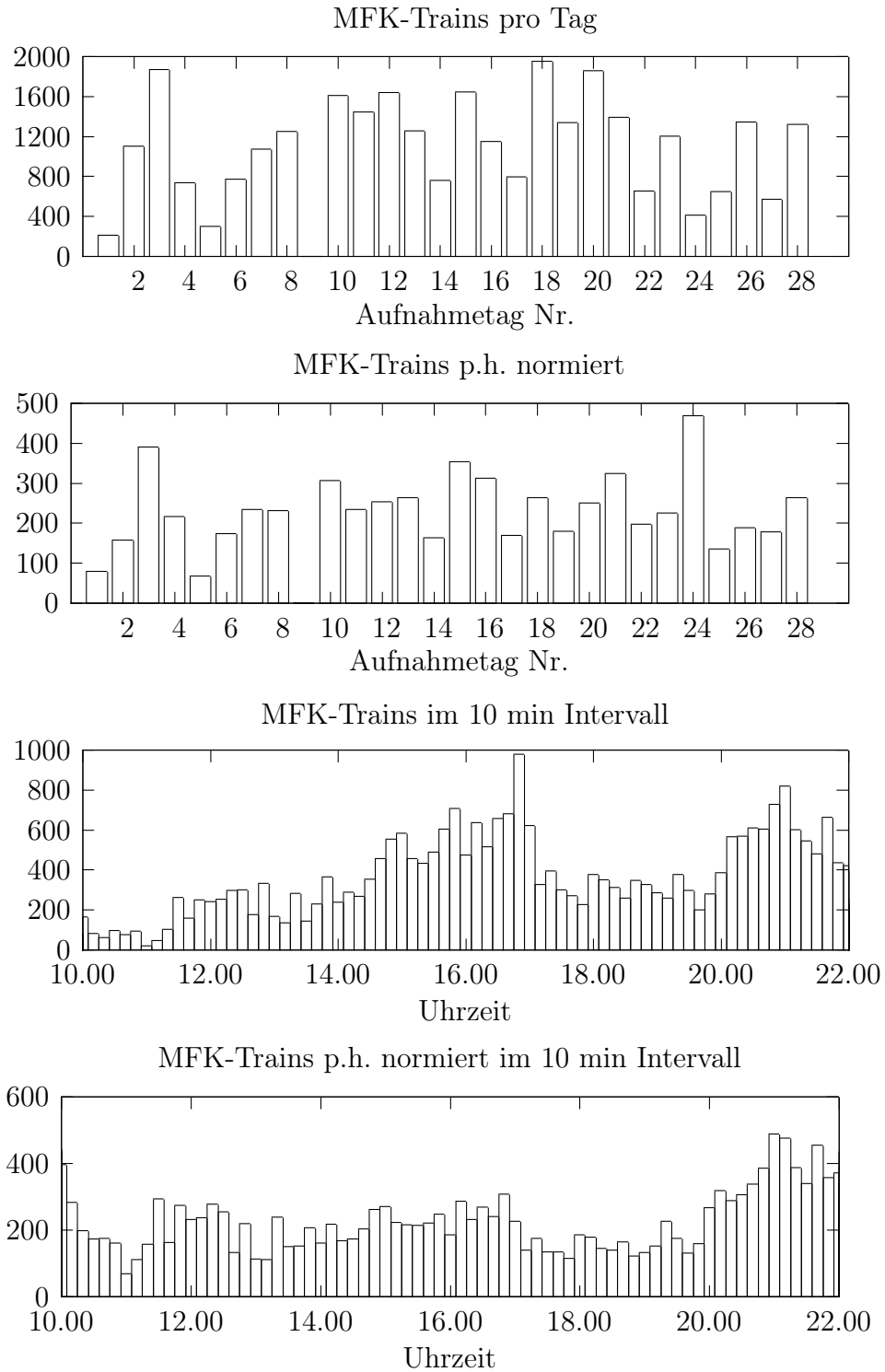


Abbildung 7.6: MFK-Trains im Eilat-Korpus entsprechend der Menge F^{MFT} mit 38280 Elementen

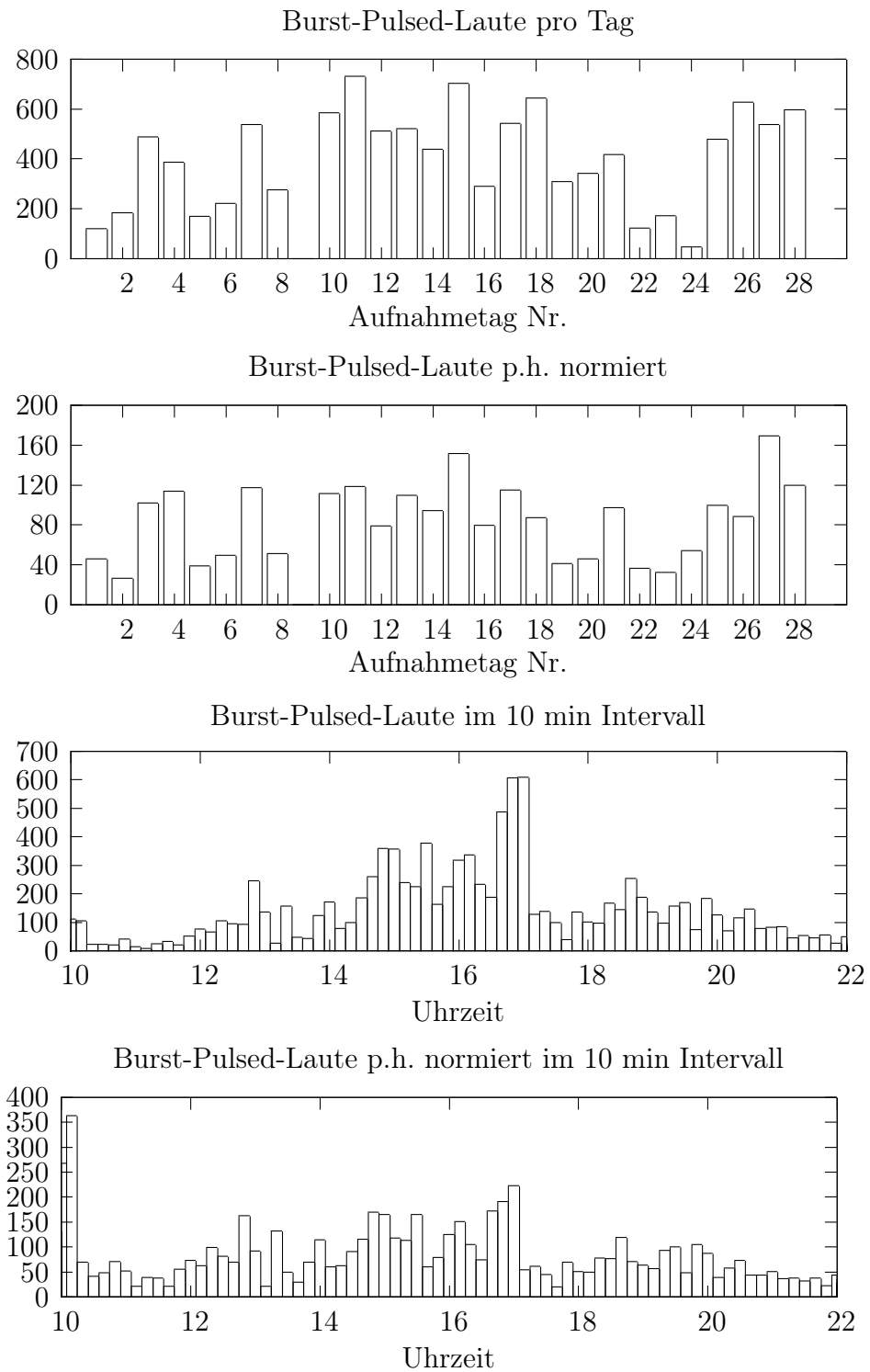
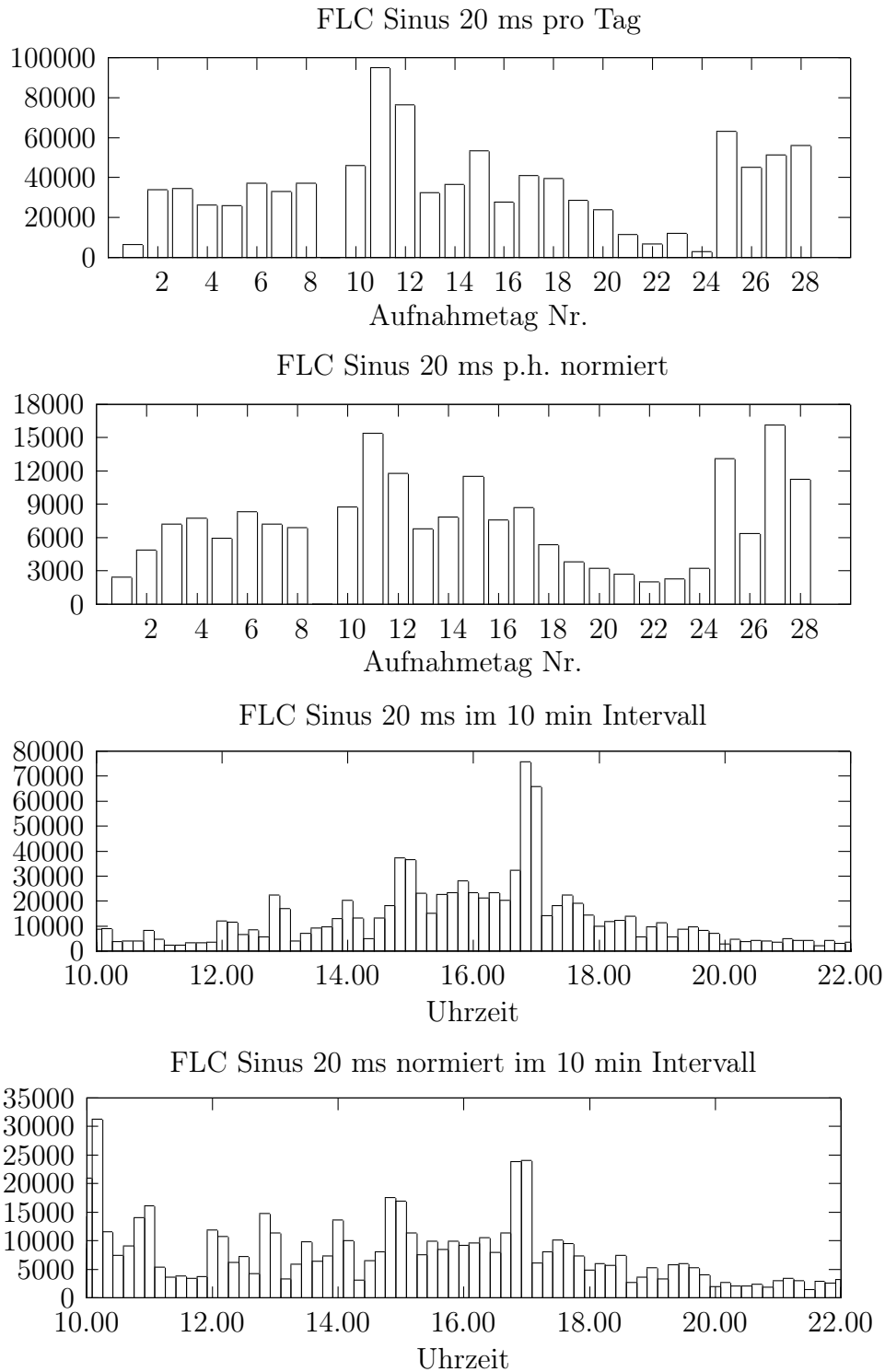


Abbildung 7.7: Burst-Pulsed-Laute im Eilat-Korpus entsprechend der Menge F^{BP} mit 11015 Elementen

Abbildung 7.8: Elemente in F^{EK} mit 984033 Elementen

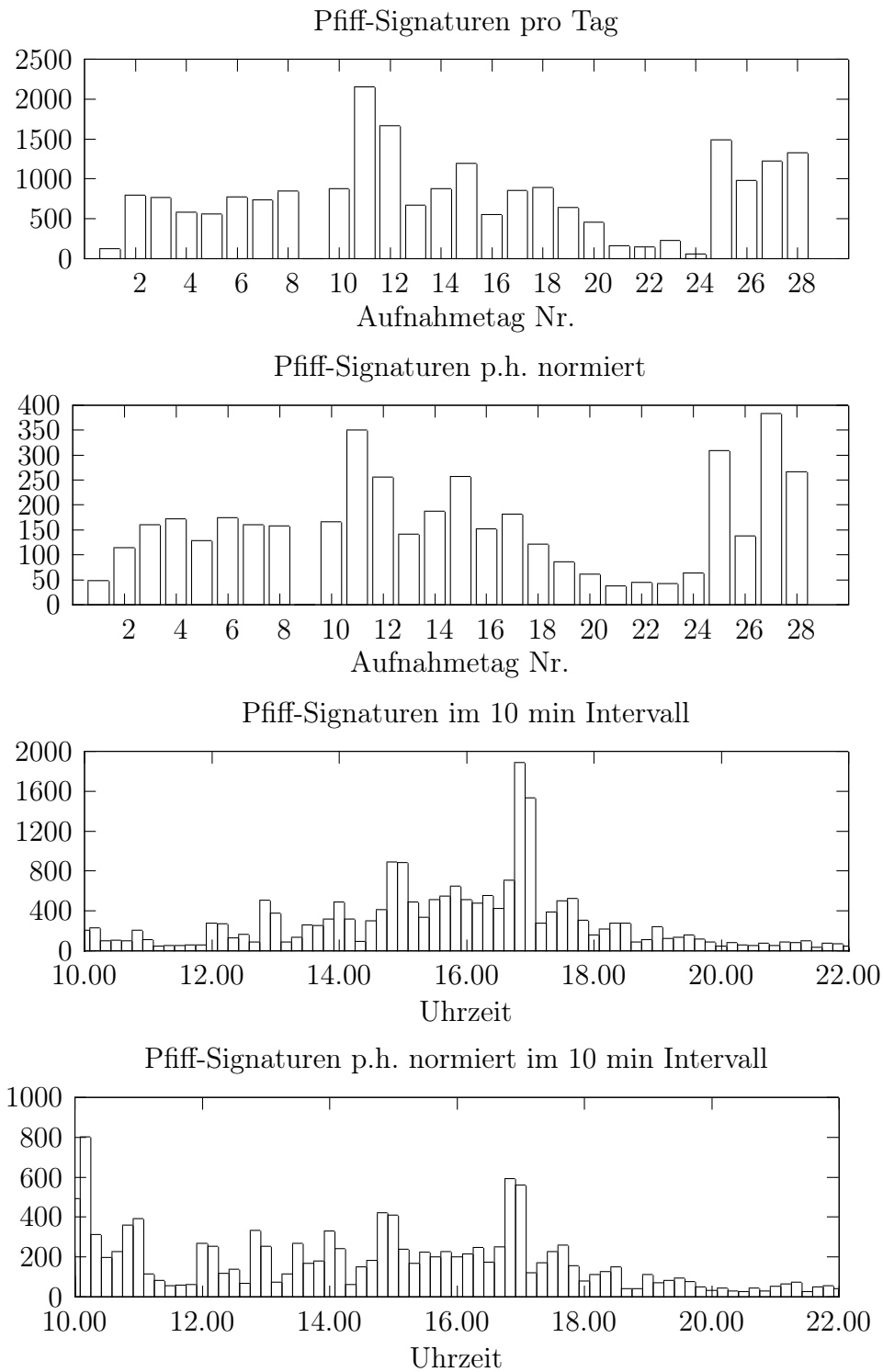


Abbildung 7.9: Pfiffklaute im Eilat-Korpus entsprechend der Menge F^{PL} mit 12280 Elementen

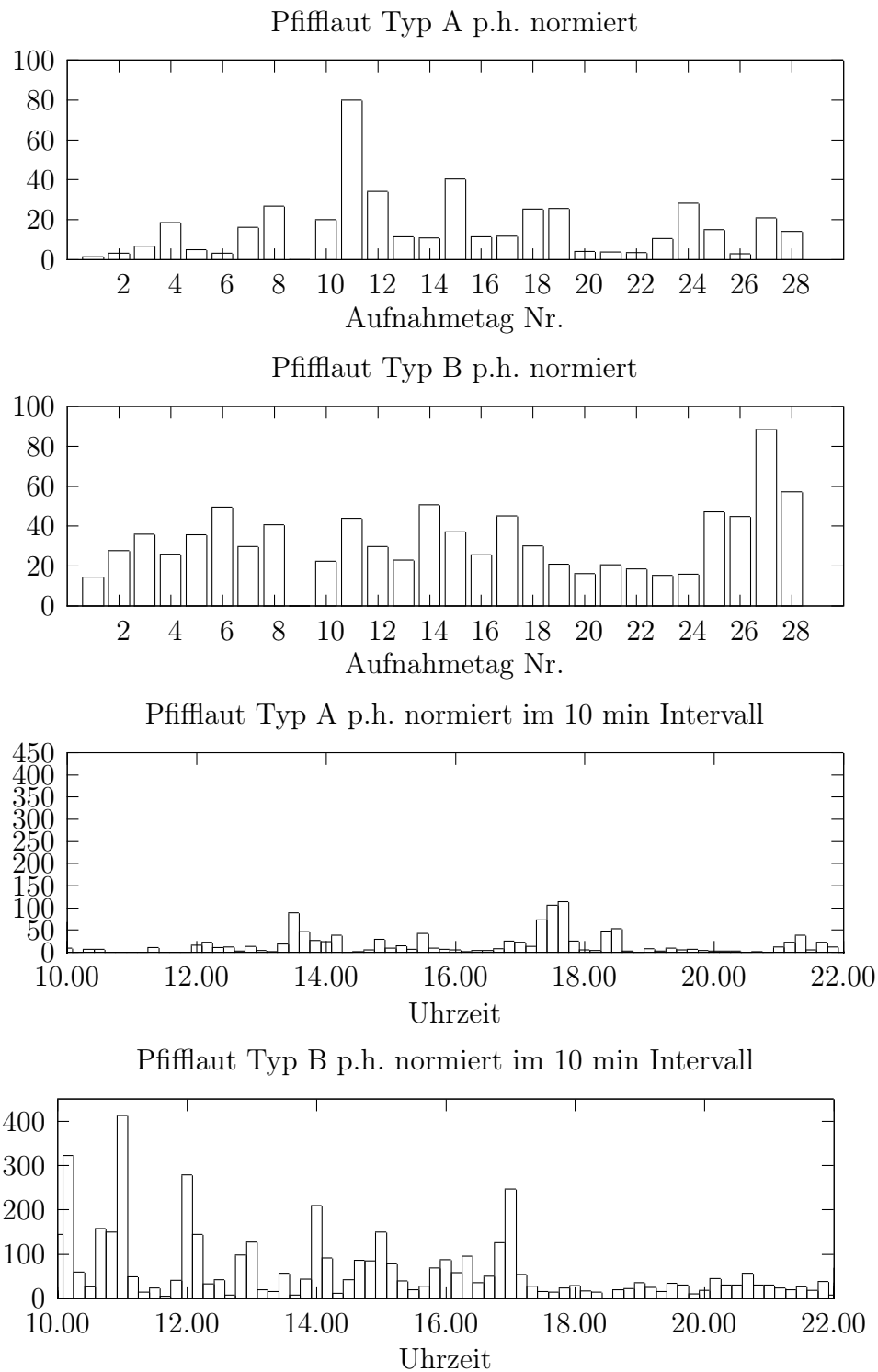


Abbildung 7.10: Vergleich der Okkurrenzen von zwei verschiedenen Pfifflauten entsprechend der Mengen F^A mit 2339 Elementen und F^B mit 4470 Elementen

(vgl. Abbildungen 7.4, 7.5 und 7.6). Die Tiere sind somit auch nach Sonnenuntergang (ca. 18:45 zur Zeit der Aufnahme) akustisch noch aktiv.

Die entsprechenden Verteilungsprofile von Pfifflauten und BP-Lauten weisen in den Abendstunden ein entgegengesetztes Aktivitätsmuster auf (vgl. Abbildungen 7.7, 7.8, 7.9 und 7.10). Kommunikative Interaktionen mittels von Pfifflauten und Burst-Pulsed-Lauten nehmen in den Abendstunden deutlich ab.

Dieser Befund spiegelt sich auch in den Profilen der einzelnen Aufnahmetage wieder: In Tagen, an denen in den Abendstunden viele Aufnahmen durchgeführt wurden, finden sich sehr viele MF-Klicklaute aber relativ wenig Pfiff- und Burst-Pulsed-Laute. In Tagen, an denen in den Abendstunden wenige Aufnahmen durchgeführt wurden, finden sich prozentual mehr Pfiff- und Burst-Pulsed-Laute, aber weniger MF-Klicklaute.

Es kann angenommen werden, dass Sonar-Aktivitäten mit Abnahme des Tageslichts zunehmen, da die Tiere sich bei schlechten Sichtverhältnisse durch Sonar orientieren. Daraus wäre zu folgern, dass es sich bei dem Großteil der MF-Klicklaute um **Sonarlaute** handelt.

Diese Hypothese ist von nicht geringer Tragweite. Einerseits folgert aus ihr, dass das Biosonar von *Tursiops truncatus* mit erheblich niedrigeren Frequenzen arbeiten kann, als bisher angenommen wurde. Andererseits legt die phänomenologische Evidenz des Phänomens des Pulsspektrum-Drifts nahe, dass es sich hier um eine Form von adaptivem Sonar handelt, bei dem spektrale Charakteristika von Sonarsignalen dynamisch im Klick-Train moduliert werden.

Ein möglicher Einwand gegen diese Hypothese besteht in folgender Annahme: Tagsüber bei guten Sichtverhältnissen könnten MF-Klicklaute zu Kommunikationszwecken eingesetzt werden. Diese hätten die oben benannten Charakteristika der phonetischen Modulation. Bei der offensichtlichen Zunahme von MF-Klicklauten in den Abendstunden könnte es sich dennoch um typische Sonarlaute handeln. Diese hätten keine anderen Eigenschaften, als bisher angenommen wurde. Im Eilat-Korpus wäre lediglich der auf der Frequenzskala unterste Teil (< 48 kHz) des Spektrums dieser Laute zu erkennen. Man hätte demzufolge zwei Arten von Klicklauten mit unterschiedlichen spektralen Charakteristika: Kommunikations-Klicklaute und Sonarlaute. Der FLC κ^{MFK} wäre demzufolge übergeneralisierend und hätte beide Typen unterschiedslos positiv klassifiziert.

Um diese Hypothese zu prüfen wurde die Dynamik der dominanten Frequenzen von MF-Klicklauten im Tagesverlauf ausgewertet. In den Abbildungen 7.11 und 7.12 ist das Ergebnis dieser Untersuchung dargestellt.

In beiden Mengen F^{MFK} und F'^{MFK} zeigen sich im Tagesverlauf keine offensichtlichen Unterschiede in der Verteilung der dominanten Frequenzen. Die dominanten Frequenzen weisen allerdings eine große Varianz auf. Häufigkeitsmaxima liegen bei ca. 25 kHz und bei ca. 30 kHz. Oberhalb von 46 kHz konnten kaum noch dominante Frequenzen festgestellt werden. Die Hypothese, dass es sich bei MF-Klicklauten um reine Sonarlaute handelt, wird durch diesen Befund bestätigt.

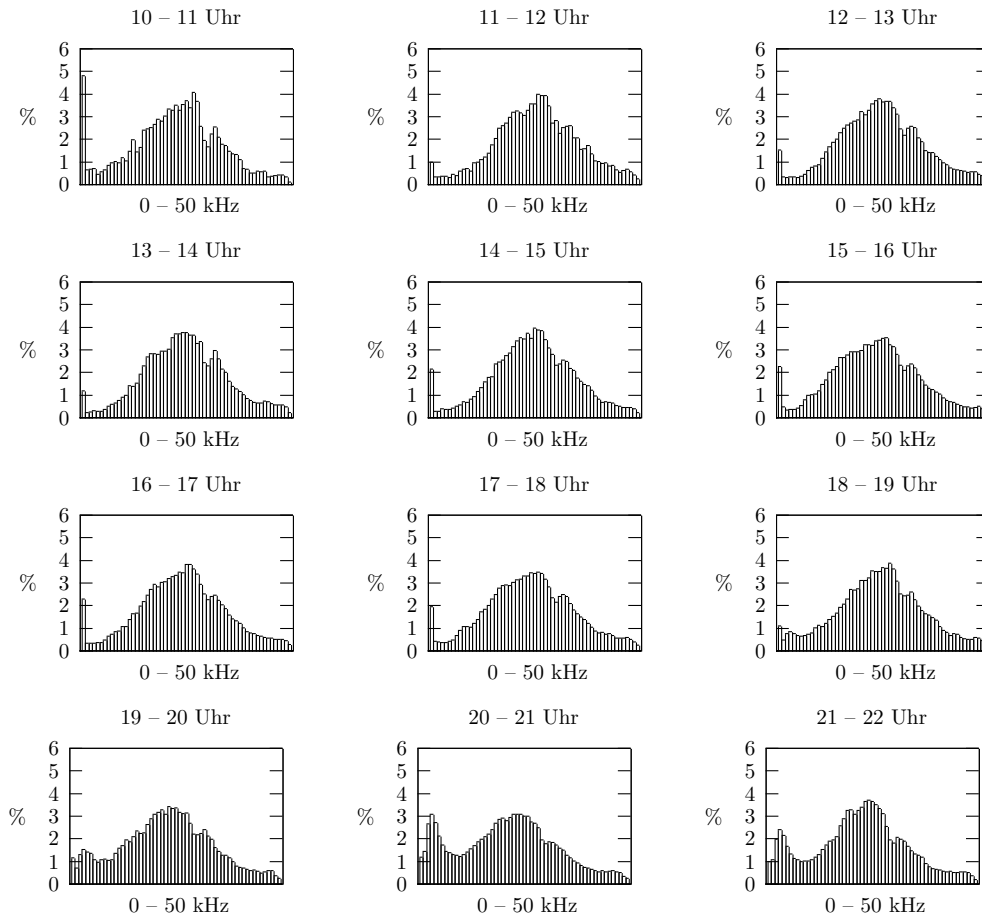


Abbildung 7.11: Dynamik dominanter Frequenzen in MF-Klicklauten aus F^{MFK} im Tagesverlauf. In der Zeit nach 19:00 nimmt die akustische Aktivität der Garnelen stark zu. Hierdurch entsteht temporär ein lokales Maximum bei ca. 4800 Hz. Auffällig ist ein lokales Maximum bei ca. 30 kHz, welches im Tagesverlauf konstant sichtbar ist. Der Peak bei 1000 Hz ist auf Maschinengeräusche zurückzuführen. Untersucht wurden 2556086 MF-Klicklaute von 0-48 kHz. Die x-Achse ist in 64 Frequenzbänder unterteilt. Auf der y-Achse ist der Anteil der Frequenzbänder am gesamten Emissionsvolumen in % abgetragen.

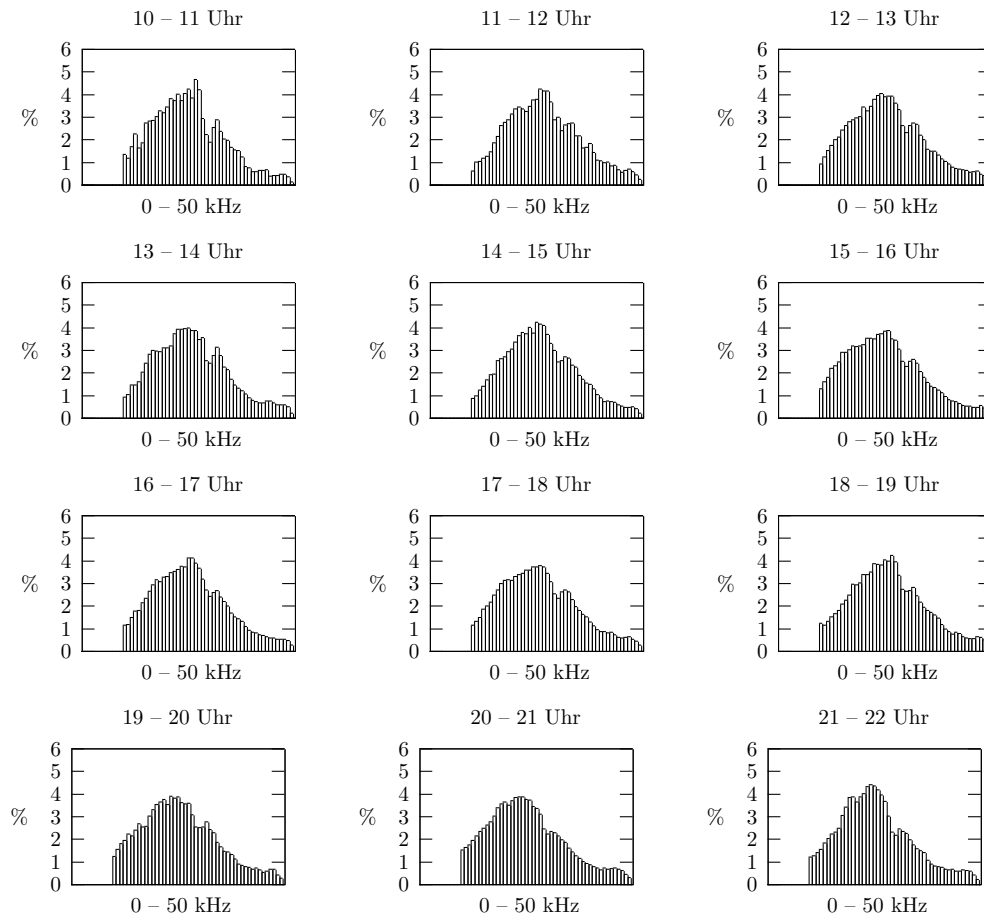


Abbildung 7.12: Dynamik dominanter Frequenzen in MF-Klicklauten aus F^{MFK} im Tagesverlauf. Das lokale Maximum bei ca. 30 kHz ist auch hier im Tagesverlauf konstant sichtbar. Untersucht wurden 2246845 MF-Klicklaute von 10.05-48 kHz. Klicklaute von Garnelen sind vor der Untersuchung herausgefiltert worden. Die x-Achse ist in 64 Frequenzbänder unterteilt. Auf der y-Achse ist der Anteil der Frequenzbänder am gesamten Emissionsvolumen in % abgetragen.

Kapitel 8

Schluss

Ein wichtiges Ziel dieser Arbeit lag darin, eine universelle praktisch anwendbare Methode zur wissensbasierten Modellierung bioakustischer ASK zu entwickeln und algorithmische Werkzeuge zu beschreiben, welche gestatten, alle Einzelschritte dieser Methode auf dem Rechner auszuführen. Der grundlegende Ansatz bestand darin, den komplexen Prozess der Modellierung von ASK durch speziell modifizierte KDD- und Data-Mining-Verfahren maschinell zu unterstützen.

Angestrebt wurde, die Grundlagen für ein offenes System zur interaktiven Modellierung von ASK für die Bioakustik und verwandte Disziplinen zu schaffen. Leitgedanke hierbei war, die Instrumente zur Modellierung von Expertenwissen und die eigentlichen Mustererkennungsalgorithmen konsequent voneinander zu trennen.

Hierzu war es zunächst notwendig, die grundlegenden theoretischen Konzepte eines universellen Systems zur maschinellen Annotation von Audiodaten einer mathematischen Beschreibung zugänglich zu machen (siehe Kapitel 2). Dies wurde durch eine Unterscheidung und Formalisierung von zwei Klassifikatortypen – FLC und SLC – realisiert. Mit Hilfe dieser Konstrukte ist eine Abbildung von zeitdiskreten Audiodaten auf die Ebene symbolischer Beschreibungen elementarer und komplexer akustischer Ereignisse möglich. Die Repräsentationsebene der Audiodaten wird so mit zwei verschiedenen symbolischen Beschreibungsebenen verbunden. Audiodaten, die zunächst in Form von Messreihen vorliegen, können nun auch durch logikbasierte symbolische Data-Mining-Verfahren ausgewertet werden.

Das Vorliegen eines mathematischen Modells warf die Frage auf, wie konkrete FLC und SLC auf Grundlage von Expertenwissen modelliert werden können. Dieses Problem wurde zunächst aus theoretischer Perspektive diskutiert. Es zeigte sich, dass die Modellierung beider Arten von ASK im Rahmen des in der Wissenschaft etablierten KDD-Prozesses erfolgen kann. Es ergaben sich für die beiden ASK-Typen zwei Modellierungsprozesse: der FLCD-Prozess und der SLCD-Prozess. Beide sind durch eine Anzahl eng umrissener methodischer Einzelschritte gekennzeichnet. Sie lassen sich als praktisch anwendbare Methode zur wissensbasierten Modellierung von ASK verstehen.

Auf der Basis der dargestellten formalen Konzepte und methodischen Schritte

war es möglich, eine Menge von grundlegenden Modellierungsproblemen zu benennen. Für jedes dieser Probleme wurde in Kapitel 3 mindestens eine konkrete Lösung im Detail beschrieben. Dabei wurden insgesamt 19 elementare Verfahren dargestellt. Sie sind geeignet, um die zwei Typen von ASK auf der Grundlage spektrographischer Signaturen akustischer Phänomene zu modellieren. Die Darstellung beinhaltete auch Verfahren zur induktiven Programmierung von ASK.

Eine Evaluation der entwickelten methodischen und algorithmischen Werkzeuge erfolgte anhand der akustischen Signale der Spezies *Tursiops truncatus*. Leitgedanke des empirischen Teils war, dass sich die Validität wissenschaftlicher Methoden und Instrumente am besten zeigen lässt, indem durch sie Ergebnisse erzielt werden, die auf empirischen Daten basieren und eine Relevanz für ein bestimmtes Wissenschaftsgebiet besitzen. Die Spezies *Tursiops truncatus* wurde gewählt, da ihre akustischen Signale besonders reich an noch unverstandenen Phänomenen sind.

Eine systematische phänomenologische Inspektion mit Hilfe von kontrastoptimierten Spektrogrammen zeigte eine Fülle von Lauttypen und Modulationsprinzipien in den akustischen Signalen dieser Spezies. Die in Kapitel 4 dargestellte Klassifikation dieser Phänomene veranschaulicht die Mannigfaltigkeit und Komplexität der Fragen, die in diesem Bereich bestehen. Obwohl in den vergangenen vier Jahrzehnten viele Anstrengungen unternommen wurden, um die Struktur und Organisation des Kommunikations- und Sonarsystems des Großen Tümmlers besser zu verstehen, zeigten sich in dem untersuchten Audiomaterial viele Besonderheiten, die bisher in der Literatur noch nicht detailliert beschrieben wurden. Dieser Umstand deutet darauf hin, dass der Einsatz maschineller Analyseverfahren bei dieser Spezies und bei anderen *Odontoceti* besonders angezeigt ist.

Um die Funktionalität und Robustheit der in Teil I beschriebenen Modellierungs- und Klassifikationsverfahren zu prüfen, wurden in Kapitel 5 mit ihrer Hilfe zahlreiche Mengen von halbsynthetischen Audiodaten unter verschiedenen Aspekten untersucht. Dabei wurde anhand von konkreten Fragestellungen gezeigt, wie bioakustische ASK für bestimmte Aufgaben modelliert und praktisch verwendet werden können.

In den beiden abschließenden Kapiteln wurden zwei prinzipiell verschiedene Einsatzmöglichkeiten von ASK in der Bioakustik aufgezeigt: Einerseits eignen sie sich um qualitative und quantitative Eigenschaften von Lautklassen zu untersuchen, andererseits können sie zum mustererkennungsbasierten Monitoring bioakustischer Ereignisse verwendet werden.

In Kapitel 6 wurden mit Hilfe eines symbolischen Lernverfahrens Cluster für Pfiff-laute induziert und anschließend untersucht. Auf Grundlage hunderttausender Annotationen elementarer akustischer Ereignisse zeigten sich deutliche, bisher unbekannte Prävalenzen bestimmter Lauttypen und bevorzugte Frequenzbereiche in den Pfiff-lauten der untersuchten Delphingruppe.

In Kapitel 7 erfolgte eine Untersuchung der Natur von MF-Klick-lauten anhand der Dynamik der Emissionsraten von verschiedenen Lauttypen. Dabei zeigte sich ein unterschiedliches Aktivitätsmuster von kommunikativen Lauten und MF-Klick-lauten: Nach Sonnenuntergang ist die Dynamik der Emissionsraten eindeutig

kommunikativer Laute (z.B. Pfiff-laute) und die Dynamik von MF-Klicklauten offensichtlich gegenläufig. Hierdurch wurde die Hypothese untermauert, dass es sich bei MF-Klicklauten um eine Form von Biosonar handelt. Dieses Ergebnis wurde durch eine Spektralanalyse aller 2.5 Millionen untersuchten Klicklaute gefestigt. Diese Analyse zeigte auch, dass die dominanten Frequenzen von MF-Klicklauten am häufigsten bei ca. 25 kHz liegen. Diese Ergebnisse sind im naturwissenschaftlichen Kontext von nicht unbeträchtlicher Bedeutung: Bei MF-Klicklauten handelt es sich wahrscheinlich um Sonar-laute, die durch eine auffällige Form der Pulsspektrum-Modulation charakterisiert sind. Zudem liegen die dominanten Frequenzen dieses Typus von Sonarlauten weit niedriger als bisher für Sonar-laute von *Tursiops truncatus* angenommen wurde.

Die in Teil I dargestellten Instrumente weisen auf eine Vielzahl von weiterführenden Forschungsfragen hin. Insbesondere im Bereich der algorithmischen Werkzeuge existieren zahlreiche vielversprechende Möglichkeiten, die Menge der beschriebenen Verfahren zu erweitern:

1. FLC-Signaturen können auch auf anderen Arten der Visualisierung, so z.B. auf Oszillogrammen, Energie-Diagrammen, Skalogrammen usw. beruhen.
2. Verbesserte Klassen-Korrelationsfunktionen für FLC können auf die Detektion von Sub-Threshold-Ereignissen ausgerichtet werden.
3. Annotationsalgorithmen können so gestaltet werden, dass sie selektiv Annotationsentscheidungen in Abhängigkeit von Kontextinformationen treffen.
4. Symbolische Elemente in SLC-Signaturen können mit logischen und Klammer-Operatoren zu komplexen Ausdrücken verknüpft werden.
5. Hierarchisch aufgebaute symbolische Signaturen können in Verbindung mit einer dynamischen Instanziierung von FLC-Annotationen die Suche in Audio-daten erheblich beschleunigen.

In dieser Arbeit wurde die Modellierung von ASK *nicht* aus verfahrenstechnischer Sicht gesehen. Die beschriebenen Werkzeuge gestatten, auf interaktive Art und Weise zu einem schrittweisen Erfahrungs- und Wissensgewinn im konkreten Anwendungskontext zu gelangen. Hierdurch wird es möglich, auch solche Modellierungsprobleme in den Griff zu bekommen, die bisher nur schwer lösbar waren. Dazu gehören nicht nur die wenigen in dieser Arbeit aufgezeigten Beispiele. Es handelt sich vielmehr um eine wichtige Problemklasse, die den gesamten Bereich der Bioakustik betrifft: FLCD- und SLCD-basierte Modellierung kann immer dann eingesetzt werden, wenn algorithmische Klassifikatoren hergestellt werden sollen, die geeignet sind, um in Audiodaten mathematisch schwer zu fassende deterministische Ereignisse zu lokalisieren und zu klassifizieren.

Die konsequente Trennung von Signaturen akustischer Phänomene und Verfahren zu deren maschineller Auswertung eröffnet einen Weg, der bisher nicht beschritten

werden konnte: Es ist nun möglich, in Form von Signaturen vorliegendes Expertenwissen unabhängig von konkreten Mustererkennungs-Verfahren zu behandeln. Hierdurch wird erreicht, dass wertvolles formalisiertes Wissen auch dann bewahrt bleibt, wenn Fortschritte in der Informatik verbesserte Möglichkeiten zu dessen Anwendung nahelegen.

Die in Teil II dargestellte Untersuchung der akustischen Signale des Großen Tümmlers weist auf eine Vielzahl von weiterführenden Fragen hin, die unter Zuhilfenahme des dargestellten Ansatzes in Zukunft untersucht werden können:

1. Existieren biologisch relevante Typen von NFP- und BP-Lauten?
2. Existieren biologisch relevante Regeln, welche die zeitliche Abfolge von NFP- und BP-Lauten bedingen?
3. Lassen sich innerhalb der beiden häufigsten Pfifftypen biologisch relevante regelhafte Variationen feststellen?
4. Besitzen andere Delphingruppen dieselben Prävalenzen von Lauttypen und Frequenzbereichen wie die Gruppe in Eilat?
5. Existieren biologisch relevante, voneinander unterscheidbare Modi zur Pulsspektrum-Modulation von MF-Klicklauten?

Die Wissenschaft von den akustischen Signalen der *Cetacea* steckt noch in den Kinderschuhen. Die meisten Kommunikations- und Echolotsysteme der über 80 Spezies dieser Ordnung sind noch nicht beschrieben. Manche ihrer Vertreter, so etwa der Weißwal (*Delphinapterus leucas*) [106, 65], scheinen ein noch umfangreicheres Repertoire akustischer Lauttypen zu verwenden als *Tursiops truncatus*. Durch den Einsatz von verbesserten Verfahren zur maschinellen Auswertung von Audiodaten besteht erstmals eine Chance, diese bioakustischen Signalsysteme und die ungezählten anderer Spezies einer wissenschaftlichen Erforschung zugänglich zu machen. Dies sollte sich lohnen, denn exaktes Wissen über die akustischen Kommunikations- und Sonarsysteme der Tiere wird in der theoretischen Biologie, in den Kognitionswissenschaften, aber auch im praktischen Naturschutz, in Zukunft eine wichtige Rolle spielen. Der Autor hofft mit dieser Arbeit einen kleinen Beitrag auf dem Wege hierhin geleistet zu haben.

Literaturverzeichnis

- [1] R. Agrawal, C. Faloutsos, and A. N. Swami. Efficient Similarity Search In Sequence Databases. In D. Lomet, editor, *Proceedings of the 4th International Conference of Foundations of Data Organization and Algorithms (FODO)*, pages 69–84, Chicago, Illinois, 1993. Springer Verlag.
- [2] W. Au. *The sonar of dolphins*. New York: Springer, 1993.
- [3] W. Au. The dolphin sonar: Excellent capabilities in spite of some mediocre properties. In M. S. M.B. Porter and W. A. Kuperman, editors, *High Frequency Ocean Acoustics*, pages 247–259, American Institute of Physics, New York, 2004.
- [4] W. W. L. Au and A. N. Popper. *Hearing by whales and dolphins*. Springer, New York, 2000.
- [5] A. Whiten, J. Goodall, W. McGrew, T. Nishida, V. Reynolds, Y. Sugiyama, C. Tutin, R. Wrangham, and C. Boesch. Cultures in chimpanzees. *Nature*, 399:682–685, 1999.
- [6] M. Baker. Depauperate meme pool of vocal signals in an island population of singing honeyeaters. *Animal Behaviour*, 51:853–858, 1996.
- [7] J. Batali. Innate biases and critical periods: combining evolution and learning in the acquisition of syntax. In B. Maes, editor, *Proceedings Alife IV workshop*, Cambridge, MIT Press, 1994.
- [8] D. Bickerton. *Language and human behavior*. Seattle: UWP, 1995.
- [9] C. Boesch and H. Boesch. Tool use and tool making in wild chimpanzees. *Folia Primatol*, 54:86–99, 1990.
- [10] J. R. Boisserie, F. Lihoreau, and M. Brunet. The position of hippopotamidae within cetartiodactyla. *PNAS*, 102:1537–1541, 2005.
- [11] G. Bol, C. Breitner, J. Mülle, and J. Schlösser. Wissensgewinnung aus großen datenbasen. Skript der Universität Karlsruhe 1996: <http://www.ubka.uni-karlsruhe.de/cgi-bin/psview?document=ira/1996/8>.

- [12] M. Bunge. *Scientific Research*, volume 1,2. Springer New York, 1967.
- [13] M. C. Caldwell and D. K. Caldwell. Individualized whistle contours in bottlenosed dolphins (*tursiops truncatus*). *Nature*, 207:434–435, 1965.
- [14] M. C. Caldwell and D. K. Caldwell. Intra-specific transfer of information via pulsed sound in captive odontocete cetaceans. In R. G. Busnel, editor, *Animal Sonar Systems*, page 879–836. Laboratoire de Physiologie Acoustique, Jouy-en-Josas, France, 1967.
- [15] M. C. Caldwell and D. K. Caldwell. Vocalization of naive captive dolphins in small groups. *Science*, 159:1121–1123, 1968.
- [16] M. C. Caldwell and D. K. Caldwell. The whistle of the atlantic bottlenosed dolphin (*tursiops truncatus*): Ontogeny. In H. Winn and B. Olla, editors, *Behavior of Marine Animals*, pages 369–410. Plenum Press, New York, 1979.
- [17] O. Chappell, R. Leaper, and J. Gordon. Development and performance of an automated harbour porpoise click detector. *Rep. Int. Whal. Commn.*, 46:587–594, 1996.
- [18] D. Cheney, R. Seyfarth, and R. Palombit. The function and mechanisms underlying baboon 'contact' barks. *Animal Behaviour*, 52,3:507–518, 1996.
- [19] D. Cheney, R. Seyfarth, and J. Silk. The role of grunts in reconciling opponents and facilitating interactions among adult female baboons. *Animal Behaviour*, 50:249–257, 1995.
- [20] D. Chesmore. Automated bioacoustic identification of species. *An. Acad. Bras. Cienc.*, 76:436–440, 2004.
- [21] N. Chomsky. *Lectures on government and binding*. Dordrecht: Foris, 1982.
- [22] N. Chomsky. *Knowledge of language: its nature, origin and use*. New York: Praeger, 1986.
- [23] N. Chomsky. *Language and problems of knowledge — the Managua lectures*. Cambridge: MIT Press, 1988.
- [24] P. Clark and T. Niblett. The cn2 induction algorithm. *Machine Learning*, 3:4:261–283, 1989.
- [25] T. Deacon. *The symbolic species*. New York: Norton, 1997.
- [26] A. Desmond. *The ape's reflexion*. London: Quartet books, 1980.
- [27] A. Desmond. *Grooming, gossip and the evolution of language*. London: Faber and Faber, 1996.

- [28] F. deWaal. Cultural primatology comes of age. *Nature*, 399:635–636, 1999.
- [29] C. Dietrich. *Temporal Sensorfusion for the Classification of Bioacoustic Time Series*. Dissertation an der Universität Ulm, 2003.
- [30] T. Dobzhansky. *Genetics and the Origin of Species*. Columbia University Press, New York, 1951.
- [31] R. Dumke. *Modernes Software Engineering*. Vieweg: Braunschweig/Wiesbaden, 1993.
- [32] R. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavior and Brain Science*, 16:681–735, 1993.
- [33] C. Evans. Referential signals. In D. H. Owings, editor, *Perspectives in ethology*, volume 12, New York: Plenum Press, 1997.
- [34] C. Evans and L. Evans. Chicken food calls are functionally referential. *Animal Behaviour*, 58:307–319, 1999.
- [35] W. E. Evans and J. H. Prescott. Observations of the sound production capabilities of the bottlenosed porpoise: A study of whistles and clicks. *Zoologica*, 47:121–128, 1962.
- [36] U. Fayyad, G. P. Shapiro, P. Smyth, and R. Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, Cambridge, MA, 1996.
- [37] U. Fayyad, G. P. Shapiro, P. Smyth, and R. Uthurusamy, editors. *From Data Mining to Knowledge Discovery: An Overview*, pages 1–34. In Fayyad et al. [36], 1996.
- [38] W. Frawley, G. P. Shapiro, and C. Matheus. *Knowledge discovery in databases: An overview*, pages 1–29. In Shapiro and Frawley [104], 1991.
- [39] W. Frawley, G. P. Shapiro, and C. Matheus. Knowledge discovery in databases an overview. *AI Magazine*, pages 1–34, 1992.
- [40] U. Freyer. *Nachrichten-Übertragungstechnik*. Hanser, Wien, 1981.
- [41] J. Friedman. Multivariate adaptive regression spline. *Annals of Statistics*, 19:1–141, 1989.
- [42] B. für Naturschutz, editor. *Bioakustische Mustererkennung*. Bundesamt für Naturschutz, 2004.
- [43] R. Gardner and B. Gardner. A vocabulary test for chimpanzees (pan troglodytes). *Journal of Comparative Psychology*, 98:381–404, 1984.

- [44] D. Gillespie and O. Chappell. An automatic system for detecting and classifying the vocalisations of harbour porpoises. *Bioacoustics*, 13:37–61, 2002.
- [45] E. Greene and T. Meagher. Red squirrels, *tamiasciurus hudsonicus*, produce predator-class specific alarm calls. *Animal Behaviour*, 55:511–518, 1998.
- [46] G. Grimshaw, A. Adelstein, M. Bryden, and G. MacKinnon. First-language acquisition in adolescence: evidence for a critical period for verbal language development. *Brain and Language*, 63:237–255, 1998.
- [47] M. Hauser. Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations. a learned model of production? *Journal of the Acoustical Society of America*, 91:2175–2179, 1992.
- [48] M. Hauser. *The evolution of communication*. Cambridge: MIT Press, 1997.
- [49] M. Hauser. A primate dictionary? decoding the function and meaning of another species’ vocalizations. *Cognitive Science*, 24:445–475, 2000.
- [50] D. Helweg. Automatic detection and species identification of blue and fin whale calls. *Bioacoustics*, 13:96, 2002.
- [51] L. Herman, D. Richards, and J. Wolz. Comprehension of sentences by bottle-nosed dolphins. *Cognition*, 16:129–219, 1984.
- [52] L. M. Herman and W. N. Tavolga. The communication systems of cetaceans. In L. M. Herman, editor, *Cetacean Behavior: Mechanisms and Functions*, page 149–209. Wiley-Interscience, New York, 1980.
- [53] M. Hetland. A survey of recent methods for efficient retrieval of similar time sequences. In M. Last, A. Kandel, and H. Bunke, editors, *Data Mining in Time Series Databases*, World Scientific, 2004.
- [54] T. Holy and G. Zhongsheng. Ultrasonic songs of male mice. *PLoS Biology*, 3, 2005.
- [55] S. Hopp, M. Owren, and C. Evans, editors. *Animal acoustic communication*. Springer, Berlin, Heidelberg, New York, 1999.
- [56] H. Hultsch and D. Todt. Repertoire sharing and song post distance in nightingales. *Behav Ecol Sociobiol*, 8:182–188, 1981.
- [57] L. Jackson. *Digital Filters and Signal Processing*. Kluwer Academic Publishers, 1989.
- [58] R. Jamal and H. Pichlik. *LabView: Programmiersprache der vierten Generation*. Prentice Hall, 1997.

- [59] V. Janik. Pitfalls in the categorization of behaviour: a comparison of dolphin whistle classification methods. *Animal Behav.*, 57:133–143, 1999.
- [60] V. Janik, G. Dehnhardt, and D. Todt. Signature whistle variations in a bottlenosed dolphin, *tursiops truncatus*. *Behavioural Ecology and Sociobiology*, 35:243–248, 1994.
- [61] V. Janik, L. S. Sayigh, and R. S. Wells. Signature whistle shape conveys identity information to bottlenose dolphins. *P Natl Acad Sci USA*, 103:8293–8297, 2006.
- [62] V. Janik and P. Slater. Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal Behaviour*, 56:829–838, 1998.
- [63] V. M. Janik and P. J. B. Slater. The different roles of social learning in vocal communication. *Animal Behaviour*, 60:1–11, 2000.
- [64] C. Kamminga. Remarks on dominant sonar frequencies from *tursiops truncatus*. *Aquatic mammals*, 16.1:14–20, 1990.
- [65] J. Karlsen, A. Bisther, C. Lydersen, T. Haug, and K. Kovacs. Summer vocalisations of adult male white whales (*delphinapterus leucas*) in svalbard, norway. *Polar Biology*, 25,11:808–817, 2002.
- [66] V. V. Kaznadzei, S. A. Krechi, and E. N. Khakhalkina. Types of dolphin communication signals and their organization. *Sov. Phys. Acoust.*, 22:484–488, 1976.
- [67] D. R. Ketten. The cetacean ear: form, frequency, and evolution. In J. A. Thomas, R. A. Kastelein, and A. Y. Supin, editors, *Marine Mammal Sensory Systems*, page 53–57. Plenum, New York, 1992.
- [68] V. I. Markov and V. M. Ostrovskaya. Organization of communication system in *tursiops truncatus montague*. In J. Thomas and R. Kastelein, editors, *Sensory Abilities of Cetaceans*, pages 599–622. Plenum Press, New York, 1990.
- [69] P. Marler. Animal communication and human language. In L. A. N.G. Jablonski, editor, *The origin and diversification of language*, San Francisco: California Academy of Sciences, 1998.
- [70] A. Marshall, R. Wrangham, and A. Arcadi. Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour*, 58:825–830, 1999.
- [71] A. Martin. Organization of semantic knowledge and the origin of words in the brain. In L. A. N.G. Jablonski, editor, *The origin and diversification of language*, San Francisco: California Academy of Sciences, 1998.

- [72] A. C. McCarthy. Review of armstrong & stokoe & wilcox: 'gesture and the nature of language'. *Lingua*, 99:135–138, 1996.
- [73] B. McCowan and D. Reiss. Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins (delphinidae, tursiops truncatus): A re-evaluation of the signature whistle hypothesis. *Ethology*, 100:194–209, 1995.
- [74] B. McCowan and D. Reiss. Wide-band, low-frequency signals during mother/aunt-infant interactions. *Zoo Biol.*, 14:293–309, 1995.
- [75] L. McNeil. Homo inventans: The evolution of narrativity. *Language and Communication*, 16:331–360, 1996.
- [76] L. Miles. The cognitive foundations for reference in a signing orangutan. In K. G. S.T. Parker, editor, *'Language' and intelligence in monkeys and apes*, San Francisco: California Academy of Sciences, 1990.
- [77] J. Mitani, K. Hunley, and M. Murdoch. Geographic variation in the calls of wild chimpanzees: a reassessment. *American Journal of Primatology*, 47:133–151, 1999.
- [78] M. Nakamichi. Stick throwing by gorillas (gorilla gorilla gorilla) at the san diego wild animal park. *Folia Primatol.*, 69:291–295, 1998.
- [79] D. Nauck, F. Klawonn, and R. Kruse. *Neuronale Netze und Fuzzy-Systeme*. Vieweg Verlag, Wiesbaden, 1996.
- [80] M. Noad, D. Cato, M. Bryden, M. Jenner, and K. Jenner. Cultural revolution in whale songs. *Nature*, 408:537, 2000.
- [81] K. S. Norris. *The Life and Times of the Spinner Dolphin*. Berkeley: University of California Press, New York, 1994.
- [82] N. A. Overstrom. Association between burst-pulsed sounds and aggressive behavior in captive atlantic bottlenosed dolphins. *Zoo Biol.*, 2:93–103, 1983.
- [83] F. Patterson and R. Cohn. Language acquisition by a lowland gorilla: Koko's first ten years of vocabulary development. *Word*, 41:97–144, 1990.
- [84] F. Patterson and E. Linden. The education of koko. <http://www.gorilla.org/gorilla/teok/index.html>, 1981.
- [85] I. Pepperberg. *The Alex studies — cognitive and communicative abilities of Grey Parrots*. Cambridge: Harvard University Press, 1999.
- [86] I. Pepperberg, J. Naughton, and P. Banta. Allospecific vocal learning by grey parrots (psittacus erithacus): a failure of videotaped instruction under certain conditions. *Behav Proc*, 42:139–158, 1998.

- [87] S. Pinker. The evolution of the human language faculty. In L. A. N.G. Jablonski, editor, *The origin and diversification of language*, San Francisco: California Academy of Sciences, 1998.
- [88] J. Placer and C. Slobodchikoff. A method for identifying sounds used in the classification of alarm calls. *Bioacoustics*, 67:87–98, 2004.
- [89] A. N. Popper. Sound emission and detection by delphinids. In L. M. Herman, editor, *Cetacean Behavior: Mechanisms and Functions*, pages 1–52. Wiley-Interscience, New York, 1980.
- [90] S. Qian. *Time-Frequency and Wavelet Transformations*. Prentice Hall, 2002.
- [91] D. Reiss and L. Marino. Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *PNAS*, 10:5937–5942, 2001.
- [92] D. Reiss and B. McCowan. Spontaneous vocal mimicry and production by bottlenose dolphins (*tursiops truncatus*): Evidence for vocal learning. *The Journal of Comparative Psychology*, 107(3):301–312, 1993.
- [93] D. Rendall, R. Seyfarth, D. Cheney, and M. Owren. The meaning and function of grunt variants in baboons. *Animal Behaviour*, 57:583–592, 1999.
- [94] L. Rendell and H. Whitehead. Culture in whales and dolphins. *Behaviour and Brain Science*, 24:309–382, 2001.
- [95] D. Rice. *Marine mammals of the world: systematics and distribution*. Society of Marine Mammalogy Special Publication Number 4, 1998.
- [96] D. Rumbaugh, T. Gill, E. von Glasersfeld, H. Warner, and P. Pisani. Conversations with a chimpanzee in a computer controlled environment. *Biol Psychiatry*, 10:627–641, 1975.
- [97] E. Savage-Rumbaugh, R. Sevcik, D. Rumbaugh, and E. Rubert. The capacity of animals to acquire language: do species differences have anything to say to us? *Phil Trans R Soc London*, 308:177–185, 1985.
- [98] S. Savage-Rumbaugh and R. Lewin. *Kanzi — the ape at the brink of a human mind*. New York: Wiley, 1994.
- [99] S. Savage-Rumbaugh, K. McDonald, R. Sevcik, W. Hopkins, and E. Rubert. Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*pan paniscus*). *J Exp Psych Gen*, 115:211–235, 1986.
- [100] L. Sayigh, P. Tyack, and R. W. M. Scott. Signature whistles of free-ranging bottlenose dolphins, *tursiops truncatus*: mother-offspring comparisons. *Behavioural Ecology and Sociobiology*, 26(4):247–260, 1990.

- [101] L. S. Sayigh, P. L. Tyack, R. S. Wells, M. D. Scott, and A. B. Irvine. Sex difference in signature whistle production of free-ranging bottlenose dolphins, *tursiops truncatus*. *Behavioral Ecology and Sociobiology*, 36:171–177, 1995.
- [102] L. S. Sayigh, P. L. Tyack, R. S. Wells, M. D. Scott, and A. B. Irvine. Individual recognition in wild bottlenose dolphins: A field test using playback experiments. *Animal Behaviour*, 57:41–50, 1998.
- [103] R. Seyfarth, D. Cheney, and P. Marler. Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Animal Behaviour*, 28:1070–1094, 1980.
- [104] G. P. Shapiro and W. Frawley, editors. *Knowledge Discovery in Databases*. AAAI/MIT Press, Cambridge, MA, 1991.
- [105] G. Shuster and P. Sherman. Tool use by naked mole rats. *Animal Cognition*, 1:71–74, 1998.
- [106] B. Sjøre and T. Smith. The vocal repertoire of white whales, *delphinapterus leucas*, summering in cunningham inlet, northwest territories. *Can. J. Zool.*, 64:407–415, 1986.
- [107] R. A. Smolker, J. Mann, and B. B. Smuts. Use of signature whistles during separations and reunions by wild bottlenose dolphin mothers and infants. *Behavioral Ecology and Sociobiology*, 33:393–402, 1993.
- [108] Y. Sugiyama. Tool use by wild chimpanzees. *Nature*, 367:327, 1994.
- [109] G. Tembrock. *Tierstimmenforschung*. Ziemsen, Wittenberg, 1977.
- [110] D. Todt. Social learning of vocal patterns and modes of their application in grey parrots (*psittacus erithacus*). *Zeitschrift für Tierpsychologie*, 39:178–188, 1975.
- [111] D. Todt. From birdsong to speech: a plea for comparative approaches. *An. Acad. Bras. Cienc.*, 76(2):201–208, 2004.
- [112] P. Tyack. Use of a telemetry device to identify which dolphin produces a sound. In K. Pryor and K. S. Norris, editors, *Dolphin Societies: discoveries and puzzles*, pages 319 – 344. Berkeley: University of California Press, 1991.
- [113] P. L. Tyack. Whistle repertoires of two bottlenosed dolphins, *tursiops truncatus*: Mimicry of signature whistles? *Behav. Ecol. Sociobiol.*, 18:251–257, 1986.
- [114] C. van Schaik and C. Knott. Geographic variation in tool use on neesia fruits in orangutans. *Am J Phys Anthro*, 114:331–342, 2001.

- [115] P. Vary, U. Heute, and W. Hess. *Digitale Sprachsignalverarbeitung*. Teubner, Stuttgart, 1998.
- [116] W. A. Watkins. The harmonic interval: Fact or artifact in spectral analysis of pulse trains? *Marine Bioacoustics*, 2:15–42, 1967.
- [117] C. Whaling, M. Solisdagger, A. Doupedagger, J. Soha, and P. Marler. Acoustic and neural basis for innate recognition of song. *Proc Nat Acad Sci*, 94:12694–12698, 1997.
- [118] H. Whitehead. Cultural selection and genetic diversity in matrilineal whales. *Science*, 282:1708–1711, 1998.
- [119] A. Whiten and C. Boesch. The cultures of chimpanzees. *Scientific American*, 01:48–55, 2001.
- [120] L. Wiener. Song learning in birds: possible models for human language acquisition. *Word*, 37:159–175, 1986.
- [121] G. Worthy and J. Hickie. Relative brain size in marine mammals. *Am Nat*, 128(4):445–459, 1986.
- [122] S. Wrobel. Data mining und wissensentdeckung in datenbanken. *Künstliche Intelligenz*, 1, 1998.
- [123] A. Zahavi. A selection for a handicap. *Journal of Theoretical Biologie*, 53:205–214, 1975.

Anhang A

KDSE-Instrumente in DSPROLOG

Komponente	Aufgaben	Tools
KDSE-Shell	Integration aller anderen Komponenten	<ul style="list-style-type: none"> • Projekt-Management • Tools zur Steuerung und Navigation im KDSE • Export und Druck von Diagrammen
Audiodaten-Interface	Wissenschaftliche Visualisierung, Navigation, Selektion, Filtern, Abspielen, Editieren von Audiodaten	<ul style="list-style-type: none"> • Oszillogramme • Energiediagramme • Spektrogramme, Phasendiagramme • Scroll- und Zoomfunktionen • Audioeditor • Bandpassfilter • Verlangsamtes und gefiltertes Abspielen, • Speichern von Phänomenrepräsentationen
Editor für FLC	Design & Test von FLC	<ul style="list-style-type: none"> • Eigenschaftseditor für FLC • Visuelles Editieren von Signaturen auf Pixelebene • Ansteuerung von Programmierverfahren • Visualisierungsaufgaben • Scroll- und Zoomfunktionen • Klonen von Signaturen • Spiegeln von Signaturen • Stapelverarbeitung • Test und Evaluation von FLC
Editor für SLC	Design & Test von SLC	<ul style="list-style-type: none"> • Eigenschaftseditor für SLC • Visuelles Editieren der hierarchischen Struktur • Ansteuerung von Programmierverfahren • Visualisierungsaufgaben • Scroll- und Zoomfunktionen • Stapelverarbeitung • Test und Evaluation von SLC
FLC-Suchmaschine	Suche nach Instanzen von FLC und Annotation von Audiodaten	<ul style="list-style-type: none"> • Suche, Klassifikation und Annotation von Audiodaten • Stapelverarbeitung von Audiodateien • Speichermangement • Status-Callback
SLC-Suchmaschine	Suche nach Instanzen von SLC und Annotation von Audiodaten	<ul style="list-style-type: none"> • Suche, Klassifikation und Annotation von Audiodaten • Stapelverarbeitung von Audiodateien • Speichermangement • Status-Callback
Datenbank für FLC-Annotationen	Interface für FLC-Annotationen	<ul style="list-style-type: none"> • DB-Explorer • SQL
Datenbank für SLC-Annotationen	Interface für SLC-Annotationen	<ul style="list-style-type: none"> • DB-Explorer • SQL

Tabelle A.1: KDSE-Instrumente in DSPROLOG