
Filterbänke und Audiocodierung

Gerald Schuller

Filterbänke und Audiocodierung

Komprimierung von Audiosignalen mit
Python

Gerald Schuller
Technische Universität Ilmenau
Ilmenau, Deutschland

ISBN 978-3-031-19989-9 ISBN 978-3-031-19990-5 (eBook)
<https://doi.org/10.1007/978-3-031-19990-5>

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

Springer

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert an Springer Nature Switzerland AG 2023

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: Caroline Strunz

Springer ist ein Imprint der eingetragenen Gesellschaft Springer Nature Switzerland AG und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Gewerbestrasse 11, 6330 Cham, Switzerland

Vorwort

Dieses Buch ist aus unserer Vorlesung „Audiocodierung“ hervorgegangen, die wir zusammen mit Prof. Karlheinz Brandenburg an der Technischen Universität Ilmenau, Deutschland, gehalten haben.

Die Entwicklung der Audiocodierung wurde vor mehreren Jahrzehnten mit dem Ziel begonnen, Musik über die damals viel langsameren Computernetzwerke, das digitale ISDN-Telefonnetz mit 64 kb/s zu übertragen, und für die sich damals entwickelnden digitalen Radio- und TV-Rundfunkstandards in den 1980er- und 1990er-Jahren.

Einige der treibenden Kräfte waren damals die Gruppe von K. Brandenburg am Fraunhofer IIS in Erlangen [1] und die Gruppe von J. Johnston an den damaligen AT&T Bell Labs in Murray Hill, New Jersey, USA [2]. Sie arbeiteten zusammen, und die deutsche Seite beteiligte sich an dem EU-Projekt EUREKA 147 zur Entwicklung eines digitalen Radio- oder Audioübertragungssystems [3].

Eine sich damals entwickelnde beliebte Anwendung war die Speicherung von Musik auf den viel kleineren Computerfestplatten, für die eine Komprimierung unerlässlich war, um sie praktikabel zu machen. Dies führte zu tragbaren Musikgeräten mit noch begrenzterem Speicherplatz, den MP3-Playern. In den 2000er-Jahren wurden Smartphones entwickelt, die drahtlose Telefone mit MP3-Playern kombinierten, insbesondere das iPhone im Jahr 2007 und die iTunes-Musikbibliothek, die es ermöglichte, Musik drahtlos zu kaufen und herunterzuladen, ohne dass eine Verbindung zu einem Computer mit fester Internetverbindung erforderlich war. Etwa zur gleichen Zeit wurden das Internet und die drahtlosen Verbindungen so schnell, dass Internetradio und Streaming immer beliebter wurden. Die Geschwindigkeit der drahtlosen Verbindungen war zuvor ein Engpass für das Musikstreaming, was wiederum die Notwendigkeit einer guten Audiokomprimierung zeigt. Dank der Geschwindigkeit der drahtlosen und kabelgebundenen Internetverbindungen ist das Streaming heute die beliebteste Art des Musikhörens, die nach wie vor auf eine gute Audiokompression angewiesen ist. Dies gilt insbesondere deshalb, weil je beliebter das Musikstreaming wird, desto mehr Netzwerkbandbreite wird verbraucht, und umso wichtiger wird, dass jeder Hörer so wenig Bandbreite wie möglich verbraucht, um eine gute Audioqualität zu erhalten.

Mit Hochgeschwindigkeitsnetzen mit geringer Verzögerung und erhöhter Rechenleistung kamen auch Anwendungen wie Telekonferenzen mit hoher Audioqualität. Hier wird die Verzögerung bei der Codierung/Decodierung wichtig. Die herkömmliche Audiocodierung/-decodierung hat eine zu große Verzögerung; daher wurden in den 2000er-Jahren Audiocoder mit geringer Verzögerung für diesen Zweck entwickelt.

Dieses Buch konzentriert sich auf die Grundlagen der Audiocodierung und verwendet Beispiele in der Programmiersprache Python, auch als algorithmische Beschreibung. Python ist eine Open-Source-Hochsprache, die sich als algorithmische Beschreibungssprache und als Grundlage für eigene Experimente eignet. Ganz nach dem Motto „nur was man erschaffen kann, versteht man auch“. Python ist auch eine vollwertige Programmiersprache, mit der sich funktionierende Audiocodecs erstellen lassen. Sie können auch eine Grundlage oder ein Beispiel für Implementierungen für Ingenieure sein. Dies bereitet den Leser auf fortge-

schrittenere Techniken der Audiocodierung vor (z. B. parametrische Bandbreitenerweiterung oder parametrische räumliche Audiocodierung).

Die Grundlagen der Audiocodierung sind auf höchster Ebene die beiden Prinzipien der „Redundanzreduktion“ und der „Irrelevanzreduktion“. Unter Redundanz versteht man die statistischen Abhängigkeiten eines Signals, die durch die Verwendung von Modellen des Signals ausgenutzt werden. Irrelevanz bedeutet die Informationen, die der Empfänger, hier das menschliche Ohr, nicht erkennen kann. Sie wird durch die Verwendung von Modellen des Gehörs, psychoakustischen Modellen, ausgenutzt.

Das Buch ist so aufgebaut, dass es zunächst die Signalmodelle in Form von Zeit-/Frequenzerlegungen und Filterbänken, wie die Diskrete Cosinus-Transformation, die Modifizierte Diskrete Cosinus-Transformation oder die Pseudo-Quadraturspiegel-Filterbank, sowie Prädiktoren behandelt.

Dann geht es um Hörmodelle mit einem psychoakustischen Modell und einer Beispielimplementierung.

Schließlich werden mehrere vollständige Kodierer mit Python-Implementierungen vorgestellt. Zunächst ein perzeptiver Subband-Codierer, dann ein prädiktiver verlustfreier Codierer, ein skalierbarer verlustfreier Audiocodierer und ein Audiocoder mit geringer Verzögerung unter Verwendung eines psychoakustischen Vor- und Nachfilters.

Die Codebeispiele sind so gestaltet, dass sie sowohl mit Python2 als auch mit Python3 funktionieren. Die Software, die dieses Buch begleitet, kann von <https://github.com/TUilmenauAMS/Python-Audio-Coder> heruntergeladen werden, zum Beispiel in einer Terminal-Shell mit `git clone https://github.com/TUilmenauAMS/Python-Audio-Coder`.

Ilmenau, Deutschland
Mai 2020

Gerald D.T. Schuller

Literatur

- K. Brandenburg, Evaluation of quality for audio encoding at low bit rates, in *82nd AES Convention* (1987)
- J. Johnston, Transform coding of audio signals using perceptual noise criteria. *IEEE J. Sel. Areas Commun.* **6**(2), 314–323 (1988)
- “EUREKA-147 – Digital Audio Broadcasting”. <https://pdfs.semanticscholar.org/96ba/d214344f3ae6b43ac7293115bcf274f3f7af.pdf>. Zugegriffen am 06.01.2023

Inhaltsverzeichnis

1	Filterbänke	1
1.1	Einleitung	1
1.2	Frequenzbereichstransformationen und Notation	2
1.3	Lineare Filter	5
1.4	Abtastung eines analogen Signals	6
1.5	Abwärtstasten eines zeitdiskreten Signals	6
1.5.1	Python Beispiel Abtastung mit Einheits-Impulsfolge	7
1.5.2	Entfernen der Nullen	9
1.5.3	Python-Beispiel	10
1.6	Upsampling	14
1.7	Die z-Transformation und Effekte im z-Bereich	16
1.8	Nicht-Ideale-Filter	18
1.9	Filterbankentwurf für perfekte Rekonstruktion	18
1.9.1	Analyse Filterbank	19
1.9.2	Synthese-Filterbank	19
1.9.3	Blocktransformationen	20
1.10	Verbindung von Blocktransformationen mit Filterbänken	20
1.10.1	DFT-Analyse Transformation	20
1.10.2	DFT-Synthese-Transformation	21
1.10.3	Diskrete Kosinustransformation	21
1.10.4	Kurzzeit-Fourier-Transformation (STFT)	22
1.10.5	Blockverarbeitung bei längeren Filtern	22
1.11	Polyphasen Matrizen	24
1.11.1	Konvertierung von Polyphasen Darstellungen	25
1.11.2	Zeitumkehr in Polyphasen Komponenten	26
1.11.3	Polyphasendarstellung für Korrelationskoeffizienten	26
1.11.4	Parseval's Theorem für die Polyphasen Darstellung	27
1.12	Perfekte Rekonstruktion	27
1.12.1	MDCT	28
1.12.2	Python-Beispiel	33
1.13	Weitere Verlängerung der Länge der Impulsantwort	45
1.13.1	Filterbänke mit geringer Verzögerung	47
1.13.2	Python-Beispiel für Filterbänke mit geringer Verzögerung	49
1.14	Pseudo-Quadraturspiegel-Filterbänke (PQMF)	59
1.15	Zeitvariable und schaltbare Filterbänke	74
1.15.1	Mit einer konstanten Anzahl von Teilbändern	74
1.15.2	Beispiel: MDCT-Fensterumschaltung	74
1.15.3	Mit einer sich ändernden Anzahl von Teilbändern	74
1.15.4	Beispiel: MDCT und Umschaltung der Filterbank mit geringer Verzögerung	75
	Literatur	80

2	Quantisierung	81
	Literatur	82
3	Prädiktive Kodierung	83
3.1	Einleitung	83
3.2	Die Lösung für den mittleren quadratischen Fehler	83
3.3	Online-Adaption	85
3.3.1	LPC-Codierer	85
3.3.2	Adaption nach dem Prinzip der kleinsten mittleren Quadrate (LMS)	87
3.3.3	Quantisierung, Decoder in Encoder	88
3.3.4	Python Beispiel LMS	89
3.3.5	Prädiktion für verlustfreie Kodierung	92
3.3.6	Python-Beispiel für prädiktive verlustfreie Kodierung	92
3.3.7	Gewichtete kaskadierte LMS (WCLMS) Prädiktion	93
	Literatur	94
4	Psycho-Akustische Modelle	95
4.1	Einleitung	95
4.2	Die Bark-Skala	95
4.2.1	Die Schroeder-Approximation	95
4.2.2	Bark Skala Abbildung	95
4.2.3	Abbildung von der Bark Skala Zurück zu Gleichförmig	96
4.3	Hörschwelle in Ruhe	97
4.4	Die Spreizfunktion	97
4.5	Nichtlineare Überlagerung	98
4.6	Das vollständige psychoakustische Modell	100
4.7	Wahrnehmungsbezogene Bewertung	102
4.8	Psychoakustische Modelle und Quantisierung	103
	Literatur	111
5	Entropie-Kodierung	113
5.1	Einleitung	113
5.2	Huffman-Kodierung	113
5.3	Golomb-Rice-Kodierung	114
	Literatur	115
6	Der Python Perceptual Audio Coder	117
6.1	Einleitung	117
6.2	Der Kodierer	117
6.3	Der Decoder	119
	Literatur	121
7	Prädiktive verlustlose Audiocodierung	123
7.1	Einleitung	123
7.2	Der prädiktive verlustlose Encoder	123
7.3	Der prädiktive verlustlose Decoder	125
	Literatur	127
8	Skalierbare verlustlose Audiocodierung	129
8.1	Einleitung	129
8.2	Die Ganzzahl-zu-Ganzzahl-MDCT	129
8.3	Der verlustfreie Encoder	136
8.4	Der verlustfreie Decoder	137
	Literatur	138

9	Psycho-Akustisches Vorfilter	139
9.1	Einleitung	139
9.2	Python-Beispiel	139
9.3	Getrennte Dateien für Vor- und Nachfilter	143
9.3.1	Das Vorfilter	143
9.3.2	Das Post-Filter	144
9.4	Ein Audio-Codierer mit extrem niedriger Verzögerung	146
	Literatur	146

Über den Autor

Gerald D. T. Schuller ist seit 2008 ordentlicher Professor am Institut für Medientechnik der Technischen Universität Ilmenau. Seine Forschungsinteressen liegen in den Bereichen Filterbanken, Audiocodierung, Musiksinalverarbeitung und Deep Learning für Multimedia. Von 1998 bis 2001 war er „Member of Technical Staff“ der Bell Laboratories, Murray Hill, New Jersey, USA. Dort arbeitete er im Multimedia Communications Research Laboratory unter anderem an der Audiocodierung für Satellitenradio.

Danach kehrte er nach Deutschland zurück und leitete von Januar 2002 bis 2008 die Gruppe Audiocodierung für spezielle Anwendungen des Fraunhofer-Instituts für Digitale Medientechnologie in Ilmenau, Deutschland, wo er unter anderem an MPEG AAC-ELD, Ultra Low Delay Audiocodierung und skalierbarer verlustfreier Audiocodierung arbeitete.