

Erstellung und Untersuchung von Langzeitspektren und äquivalenten Rauschsignalen für klassische und populäre Musik in den Dekaden von 1960 bis 2000

Jens-Alrik Adrian, Joerg Bitzer

Institut für Hörtechnik und Audiologie, Jade Hochschule, 26121 Oldenburg, Deutschland, Email: jens-alrik.adrian@jade-hs.de

Einleitung

Mittlere Langzeitspektren vermitteln Erkenntnisse über spektrale Eigenschaften prototypischer Signalklassen. So ist bereits das *Long Term Average Speech Spectrum* (LTASS) [1] bekannt, das als Vorlage für ein sprachähnliches Störsignal mit äquivalenten spektralen Eigenschaften dient. Dieses kann als optimaler Maskierer in Hörversuchen genutzt werden oder um Spracherkennungssysteme auf Robustheit gegenüber Rauschprozessen zu evaluieren. In diesem Beitrag wird eine Erweiterung auch für Algorithmen, welche auf Musiksignale wirken (Bsp: Audiorestoration), vorgestellt. Hierzu wurde das *Long Term Average Music Spectrum* (LTAMS) aus populärer und klassischer Musik erstellt und Störgeräusche mit äquivalentem spektralem Gehalt synthetisiert.

Datenbank

Die spektralen Eigenschaften von Musik sind stark abhängig sowohl von Aufnahmesituation und -qualität als auch vom konkreten Genre und der Entstehungszeit. Um diese Einflüsse zu berücksichtigen, erscheint eine Betrachtung von Musiksignalen in Dekaden und Genres (zunächst populäre und klassische Musik) als sinnvoll. Als Quelle für populäre Musik wurde die Datenbank der SZ-Diskothek [2] gewählt. Hierbei handelt es sich um eine Auswahl von jeweils 20 repräsentativen Titeln aus den Jahren 1955 bis 2004 mit der zum Zeitpunkt der Aufnahme möglichen Qualität, die für jeden einzelnen Jahrgang als ähnlich betrachtet werden kann, da ausschließlich erfolgreiche Stücke der Music-Charts verwendet wurden. Der Vorteil der Datenbank liegt in der Einheitlichkeit bezüglich der Anzahl von Signalen pro Jahr und damit Dekade. Eine weitere Einteilung in Genres der Populärmusik hat nicht stattgefunden. Die Erstellung einer Datenbank gleicher Eigenschaften für klassische Musik hat sich als schwierig herausgestellt. Dabei wurde auf private Musiksammlungen zurückgegriffen. Zum aktuellen Zeitpunkt besitzt die Datenbank keine einheitliche Datenmenge pro Jahr und Dekade und enthält vermehrt symphonische Musik bestimmter Kompositionsarten, wie z. B. Opern und Konzerte. Aus diesem Grund werden in diesem Beitrag nur die Eigenschaften populärer Musik untersucht und Störgeräusche mit äquivalentem spektralem Gehalt synthetisiert. Die Zwischenergebnisse für Klassik liegen vor und können auf der zu diesem Beitrag korrespondierenden Web-Seite eingesehen und die generierten Daten genutzt werden.

Vorverarbeitung

Zur Vorbereitung der Analyse und Synthese wurde gemäß Abb. 1 jedes Signal zu Beginn und Ende um 5% der jeweiligen Signaldauer gekürzt, um etwaige Fading-Prozesse oder Pausen zu entfernen. Weiterhin wurden die Signale auf einen Pegel von -30 dB FS RMS normiert. Zur Bestimmung der Signalenergie kam die Methode der ITU-T P.56 Empfehlung [3] zum Einsatz.

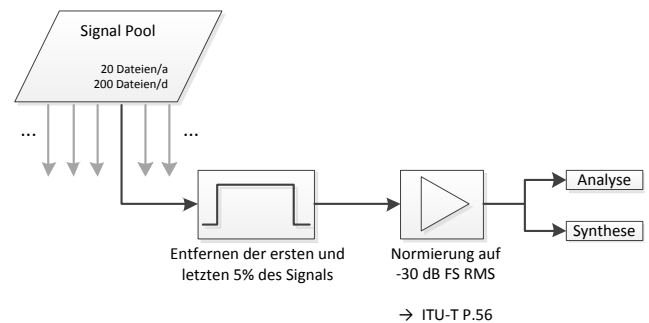


Abbildung 1: Vorverarbeitung der Musiksignale, um einheitliche Daten für Analyse und Synthese hinsichtlich des Signalpegels zu erhalten.

Analyse

Auf Grundlage der normierten Daten wurde für jedes Musiksignal innerhalb einer betrachteten Dekade die spektrale Leistungsdichte geschätzt und daraus das mittlere Spektrum bestimmt. In Abb. 2 sind für die vier Dekaden der Jahre 1960 bis 2000 die mittleren Spektren aufgetragen. Dabei ist zwischen 100 Hz und 10 kHz ein spektraler Abfall von ca. 6 dB/oct zu erkennen, wobei im Bereich von 200 Hz bis 4 kHz die typische tonale Struktur der westlichen Musik mit 12 Halbtonen/oct deutlich wird. Es fällt auf, dass die spektrale Bandbreite mit fortschreitender Dekade deutlich zugenommen hat. Als weitere Analyse wurde die Modulationstiefe bestimmt, indem für jedes der normierten Zeitsignale die Hilbert-Einhüllende berechnet und diese vom Mittelwert befreit wurde. Die modifizierten Periodogramme der Einhüllenden einer Dekade wurden gemittelt und skaliert, wodurch ein mittleres Leistungsdichtespektrum der Amplitudenmodulation für jede Dekade resultiert. Abb. 3 illustriert diese Verarbeitung. Die mittleren Modulationsspektren sind in Abb. 4 dargestellt. Es fällt dabei auf, dass die spektrale Energie mit zunehmender Dekade abnimmt. Dies lässt auf einen Verlust der Modulationstiefe im Zeitbereich mit fortschreitendem Aufnahmejahr schließen. Eine mögliche Ursache könnte der zunehmend verstärkte Einsatz von Lautheitskompression im Mastering-Prozess von populärer

rer Musik sein, welche zu einem Verlust von Modulationstiefe im Zeitbereich [4] führen würde.

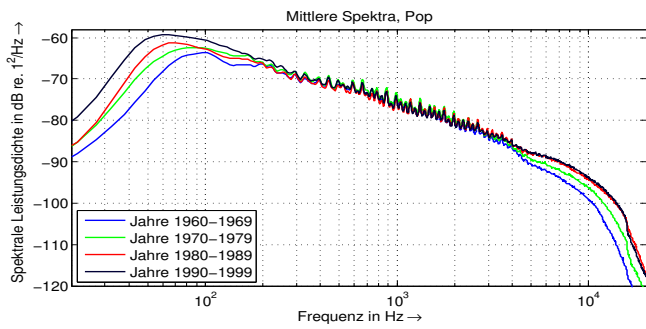


Abbildung 2: Mittlere Spektren der Dekaden zwischen den Jahren 1960 und 2000 für populäre Musik.

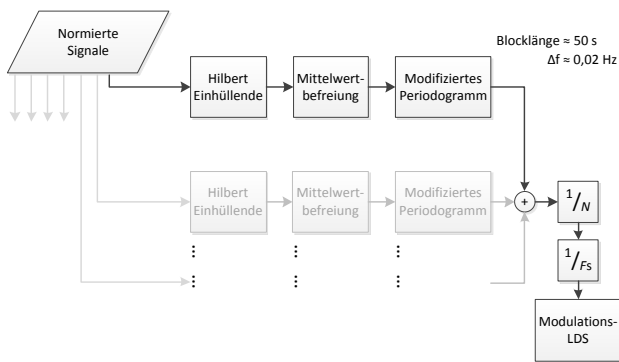


Abbildung 3: Berechnung des mittleren Modulationsspektrums der Signale einer Dekade.

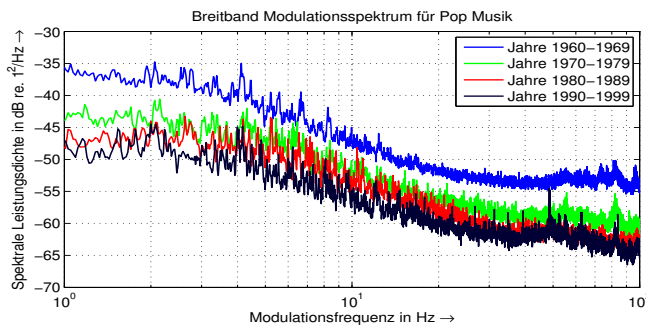


Abbildung 4: Mittlere Modulationsspektren der Zeitsignale in den Dekaden der Jahre 1960 bis 2000.

Synthese

Um Störgeräusche mit LTAMS-äquivalenten Eigenschaften zu synthetisieren, wurden die Zeitsignale jeder Dekade gemäß Abb. 5 in Blöcke eingeteilt (200 ms Länge, 100 ms Überlappung, Hann-Fenster), die pro Signal durchpermutiert wurden. Der Beginn des anschließend per *Overlap-Add*-Verfahrens rekonstruierten Zeitsignals wurde ebenfalls permutiert, um etwaige periodische Instationaritäten an den Blockgrenzen zu vermeiden. Die so von einer kohärenten Semantik befreiten Signale wurden aufsummiert, so dass ein rauschartiges Signal mit LTAMS-äquivalenten spektralen Eigenschaften pro Dekade resultiert. Abb. 6 zeigt die Leistungsdichtespektren für

die vier LTAMS-Störgeräusche der Jahre 1960 bis 2000. Diese bestätigen auch den Eindruck im Hörvergleich, dass die Bandbreite mit fortschreitender Dekade zunimmt.

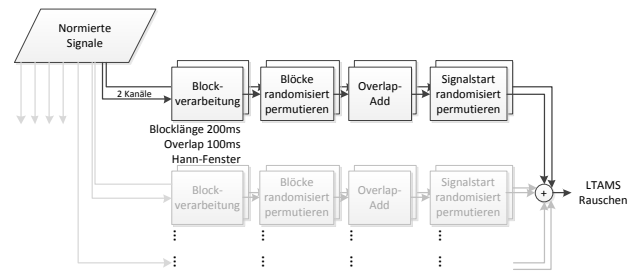


Abbildung 5: Vorgehen zur Synthese der Störsignale mit LTAMS-äquivalenten spektralen Eigenschaften.

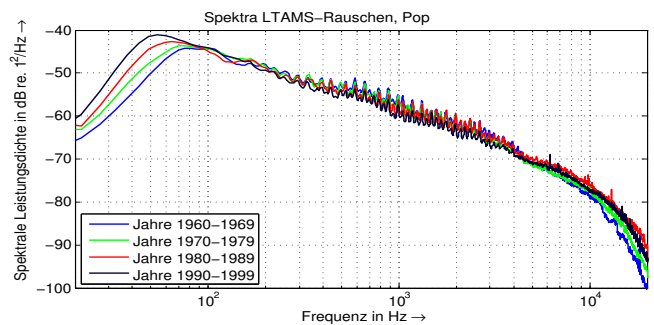


Abbildung 6: Spektren der synthetisierten Störgeräusche.

Zusammenfassung

In diesem Beitrag wurden Langzeitspektren von Musik (Schwerpunkt Pop.) zwischen den Jahren 1960 und 2000 verglichen. Diese weisen auf eine zunehmende Bandbreite mit fortschreitender Zeit hin. Ebenfalls konnte der Verlust an Modulationstiefe mit zunehmender Dekade beobachtet werden. Die Synthese mit Hilfe der verwendeten Datenbank lieferte Störsignale mit LTAMS-äquivalenten spektralen Eigenschaften. Letztere können unter http://tgm.jade-hs.de/Adrian_2014_LTAMS bezogen werden.

Literatur

- [1] Byrne, D., et al.: An international comparison of long term average speech spectra, *The Journal of the Acoustical Society of America* 96.4 (1994): 2108-2120.
- [2] Süddeutsche Zeitung Magazin: SZ Diskothek, <http://sz-magazin.sueddeutsche.de/shop/weitere/szdiskothek>
- [3] ITU-T P.56-Objective Measurement of Active Speech Level, International Telecommunications Union (2011)
- [4] Pleasurize Music Foundation: <http://www.pleasurizemusic.com/de/de/wie-ist-der-loudness-war-entstanden> (Abgerufen: 05. März 2014)

Diese Arbeit wurde gefördert durch das BMBF (Förderkennzeichen 03FH030PX2). Den Inhalt und die Erkenntnisse verantworten ausschließlich die Autoren.