

Abschlussbericht

Zuwendungsempfänger :	Jade Hochschule Wilhelmshaven / Oldenburg / Elsfleth (ehemals Fachhochschule Oldenburg / Ostfriesland / Wilhelmshaven)
Förderkennzeichen :	17N3008
Vorhabensbezeichnung :	AURUM: Automatische Restauration von Audiosignalen in Medienarchiven
Laufzeit des Vorhabens :	01.06.2008 bis 31.12.2011
Berichtszeitraum :	01.06.2008 bis 31.12.2011
Partnerunternehmen :	Cube-Tec International GmbH, Bremen

Zusammenfassung

In diesem Projekt wurden Verfahren zur automatischen Restauration von Audiodaten entwickelt und evaluiert. Für die drei Hauptstörungen Brummen, Rauschen und Impulse (Clicks und Crackles) konnten spezifische Lösungen gefunden werden. Im Verlauf des Projektes zeigte sich, dass eine vollständige Automation nicht gewünscht ist, sondern mindestens ein Geschmacksparameter notwendig ist. Für die Brummdetektion wurde ein effizientes Verfahren gefunden und evaluiert. Die Reduktion erfolgt durch ein proprietäres Verfahren des Projektpartners. Für Rauschen als die dominante Störung über alle Daten eines Medienarchives konnte eine sehr gut klingende Lösung gefunden werden. Bei den Impulsstörungen liegt ein neues Verfahren vor, das zunächst eine automatische Restauration von Filmmaterial ermöglicht, aber noch nicht allgemeingültig funktioniert. Zur Evaluation der entwickelten Algorithmen wurden Standard-Hörtests nach ITU-Richtlinien in Flash umgesetzt, um so flexibel und schnell große Gruppen an bereits in Hörtests trainierte Studierende für die Evaluation zu nutzen. Die wesentlichen Erkenntnisse wurden und werden noch international auf Konferenzen und in Fachzeitschriften veröffentlicht.

Inhaltsverzeichnis

Voraussetzungen und Randbedingungen	6
Aufgabenstellung	6
Voraussetzungen für das Vorhaben	7
Planung und Ablauf des Vorhabens	7
Stand der Wissenschaft und Technik	7
Kontinuierliche Störungen:	8
Impulshafte Störungen:	9
Determinierte Störungen:	10
Bekannte Konstruktionen, Verfahren, Schutzrechte	10
Im Projekt verwendete Fachliteratur	11
Zusammenarbeit mit anderen Stellen	11
Ergebnisse	13
Kontinuierliches Rauschen	13
Grundlagen	13
Problemformulierung	13
Optimierungen	14
Spektralglättung der Gewichtungsfunktion	15
Der Decision-Directed-Ansatz	15
Entrauschung in zwei Schritten	17
Schätzung der Rauschleistungsdichte	19
Ermittlung von Abschnitten mit stationärem Rauschboden	20
Blinde Bandbreitenerweiterung	22

Evaluation	22
Impulsstörungen	24
Problembeschreibung	24
Erreichte Lösung	24
Detektion	24
Interpolation	26
Evaluation	27
Determinierte Störungen	28
Grundlagen	28
Lösungsstrategien	29
Evaluation	31
Nutzen und Verwertung	36
Verwertung aus Sicht der Hochschule und der Hochschullehrer	36
Wirtschaftliche Verwertung	36
Gesellschaftlicher Nutzen	36
Fortschritte auf dem Gebiet bei anderen Stellen	37
Produkte	37
Wissenschaftliche Veröffentlichungen	37
Patente	37
Erfolgte und geplante Publikationen der Ergebnisse	38
2009 - EUSIPCO	38
2010 – IWAENC	38
2011 – 130. AES Convention	38
2012 – 132. AES Convention	39
2012 – IWAENC	39
Weitere geplante Publikationen	39
Mit welchen Partnern hat eine Kooperation stattgefunden?	39
Wie viele Drittmittel wurden eingeworben?	40
Verbesserung der Drittmittelfähigkeit	40

Anregungen für das Programm Forschung an Fachhochschulen 40

Voraussetzungen und Randbedingungen

Aufgabenstellung

Audiodateien, dazu zählen Sprachaufnahmen, Geräusche und Musik, sind ein wesentlicher Teil multimedialer Bibliotheken. Große Teile dieser bisher auf Einzelmedien gespeicherten Tondokumente wurden in den letzten Jahren mit viel Aufwand digitalisiert und in große Massenspeicher übertragen. Diese digitale Archivierung ist besonders für viele historische Aufnahmen wichtig, da die bisherigen analogen Datenträger wie Magnetband und Platten (Schellack und Vinyl) einem Alterungs- und Zerfallsprozess unterliegen, der die Signalqualität stark einschränken kann. Gleichzeitig ist der Zugang zu den vielfach kulturell wertvollen Aufnahmen in analoger Form nur sehr begrenzt möglich, da immer nur eine Person den Datenträger nutzen kann. Bei digitaler Distribution ist diese Beschränkung aufgehoben. Viele der historischen Aufnahmen weisen aber bereits heute eine derart schlechte Signalqualität auf, dass die gespeicherten Kopien in der vorhandenen Qualität heutigen Hörgewohnheiten nicht mehr gerecht werden. Dies schränkt zum einen die kommerzielle Nutzung z. B. für Rundfunkanstalten stark ein. Zum anderen sind aber auch kulturelle und forschungsrelevante Aufgaben behindert, da teilweise wesentliche Nutzsignalanteile von Störsignalen maskiert werden oder die Höranstrengung so hoch ist, dass immer nur sehr kurze Passagen des Audiomaterials angehört werden können. Für einige wenige besonders wertvolle Aufnahmen kann eine manuelle Restauration mit Hilfe von Algorithmen der digitalen Signalverarbeitung erfolgen. Hierzu wurden in den letzten Jahren verschiedene kommerzielle Produkte auf den Markt gebracht. Die vorhandenen Werkzeuge verlangen aber eine fachgerechte Bedienung, um qualitativ hochwertige Ergebnisse zu erzielen. Weniger erfahrene Nutzer haben häufig die Schwierigkeit, die von den Algorithmen eingebrachten neuen Störungen (Artefakte) wahrzunehmen, bzw. sie sind nicht in der Lage, die geeigneten Restaurationswerkzeuge und die beste Anwendungsreihenfolge der einzelnen Werkzeuge je nach vorliegendem Problem auszuwählen. Die automatische Restauration führt bisher nur zu suboptimalen Ergebnissen, da die vorhandenen Algorithmen auf die Anwendung mit Nutzereingriff optimiert wurden.

Das Hauptziel dieses Projektes ist die Entwicklung von Signalverarbeitungsalgorithmen zur automatischen Restauration von bereits digitalisierten historischen Audioaufzeichnungen. Hierbei soll eine deutliche Qualitätsverbesserung gegenüber den bestehenden Lösungen erreicht werden. Insbesondere soll die bisher notwendige Nutzerinteraktion möglichst minimiert werden. Falls ein direkter Eingriff notwendig ist, soll das neue System möglichst viele Hilfestellungen durch vorherige Extraktion von Metaparametern geben, um so den Arbeitsprozess zu optimieren. Die neu entwickelten Algorithmen sollen anschließend an das bisher nur suboptimal funktionierende System des beteiligten KMUs zur automatischen Restauration angepasst werden.

Voraussetzungen für das Vorhaben

Die Jade Hochschule hat durch den Projektleiter und seinen Kollegen ein erhebliches Wissen im Bereich Akustik und digitale Signalverarbeitung eingebracht, um bessere Algorithmen zu entwickeln. Gleichzeitig wurde Expertise in den Bereichen Hörtestentwicklung und statistische Auswertung, sowie zur instrumentellen Bewertung von Audioqualität zur Verfügung gestellt, um die zu entwickelnden Algorithmen zu bewerten.

Die eigentliche Evaluation, aber auch die Spezifikation fand durch den Projektpartner Cube-Tec statt, der in diesem Bereich seine große Erfahrung mitbrachte. Die Zusammenführung dieser Expertisen war eine notwendige Bedingung für ein erfolgreiches Projekt.

Planung und Ablauf des Vorhabens

Vergleicht man die Planung und der tatsächliche Ablauf, so ist vor allem zu Beginn eine nicht aufzuholende Verspätung entstanden, da der für die Promotion vorgesehen Mitarbeiter zunächst seine Diplomarbeit beenden musste. Diese Verzögerung von ca. einem halben Jahr konnte nicht kompensiert werden und erforderte gegen Ende des Projektes eine kostenneutrale Verlängerung. Durch das Thema der Masterarbeit wurde auch die Reihenfolge der einzelnen Arbeitspakete leicht geändert. Da die drei angedachten Restaurationswerkzeuge und deren Entwicklung unabhängig sind, war dies aber ohne Abstriche in den Gesamtzielen möglich.

Insgesamt zeigte sich aber, dass einige der Probleme leicht unterschätzt wurden, so dass etwas mehr Zeit für die Entwicklung und für die Evaluation benötigt wurde. Dies führte zu Verlängerungen bei der Hum-Reduktion und einem leicht verspäteten Start bez. der Click-Reduktion.

Die Idee, zunächst eine Befragung von Nutzern von Restaurationssystemen durchzuführen, um ein Optimalergebnis zu erhalten, erwies sich als nicht durchführbar und auch nicht sinnvoll. Die entscheidenden Personen nutzen sehr unterschiedliche Algorithmen, die vom Klang oft nicht verglichen werden können. Aber selbst die Einschränkung auf Nutzer des Systems vom Projektpartner hatte bei einem kurzen Test eine so hohe Variabilität, dass keine Aussage möglich war. Hier ist vor allem die kulturelle Heterogenität ein Grund, da Cube-Tec international vertreibt und im asiatischen Raum andere Geschmacksvorgaben gelten. Dies führte dazu, dass von Anfang an die Idee einer Null-Parameter Lösung aufgegeben wurde und statt dessen jeder Algorithmus am Ende genau einen Parameter haben darf, der die Stärke der Reduktion und damit auch die Stärke der eingebrachten Artefakte festlegt. Dies bedeutet häufig immer noch, dass eine signaladaptive Lösung vorliegt, da die Algorithmen auf gutes bzw. schlechtes Eingangs-Signal-Rauschverhältnis unterschiedlich reagieren.

Stand der Wissenschaft und Technik

Die Störungen bei Audiosignalen lassen sich bezüglich ihrer Herkunft in verschiedene Klassen einteilen. Eine grobe Einteilung ist in Abbildung 1 gezeigt. Man unterscheidet in zufällige und determinierte Störungen, wobei der Schwerpunkt bei der Restauration auf den zufälligen Störungen liegt. Diese können weiter in die beiden Klassen kontinuierliches Rauschen und impulshafte Störungen, z.B. Knackser, unterteilt werden. Bei den determinierten Störungen sind insbesondere die nicht mit dem Nutzsignal korrelierten Störungen ein wesentliches Gebiet der Audiorestauration. Zusätzlich ist die digitale Übersteuerung (Clipping) eine signifikante Störung, die aber zu den korrelierten determinierten Störungen zählt. Im Weiteren wird der Stand der Wissenschaft und Technik für diese drei Teilgebiete genauer erläutert und darauf eingegangen, inwieweit

Verfahren zur automatischen Parameterwahl bekannt sind. Die automatische Detektion und Klassifikation der Störungsart wurde bisher nach dem Kenntnisstand des Antragstellers weder in der wissenschaftlichen Literatur, noch in kommerziellen Produkten behandelt und definiert somit eine eigene neue Fragestellung.

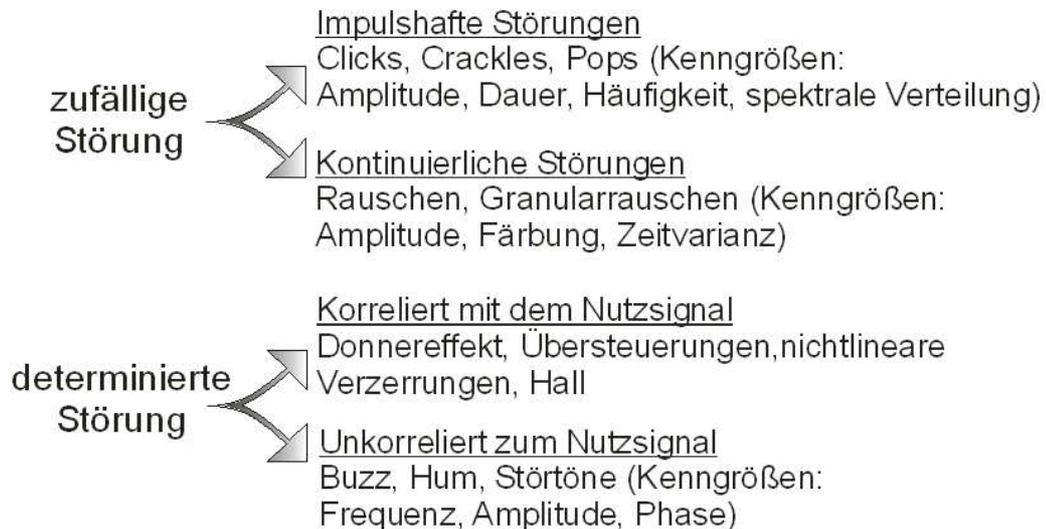


Bild 1: Eine mögliche Einordnung der Störungen bei Musiksignalen

Kontinuierliche Störungen:

Zur Klasse der typischen kontinuierlichen Störungen zählt das breitbandige Rauschen, wie es bei früheren analogen Aufzeichnungsverfahren verursacht wurde. Jedes Verfahren hat dabei eine sehr eigene Rauschcharakteristik, die sich durch die Färbung und den Signal-Rauschabstand (SNR) beschreiben lässt. Häufig kommt noch eine leichte Instationarität hinzu, die klangbildend ist. Ein Beispiel hierfür ist das typische cyclo-stationäre Rauschen einer einseitig verschmutzten Schellack-Platte. Das Rauschen ist sehr laut, weist also einen schlechten SNR auf, ist tiefpassgefärbt und variiert leicht mit der Umdrehung der Scheibe. Andere Quellen für breitbandiges Rauschen können akustische Störquellen wie Klimaanlage sein, die durch die Aufnahme mit Mikrofonen auf das Medium gespeichert wurden.

Für die Restauration dieser Arten von Störungen werden zur Zeit insbesondere Algorithmen zur Geräuschreduktion aus der Sprachsignalverarbeitung genutzt, deren Parameter auf die besonderen Eigenschaften von Musik angepasst werden. Während der letzten 20 bis 30 Jahre wurden eine Vielzahl unterschiedlicher Ansätze vorgestellt. Im Zeitbereich gibt es Lösungen, die auf der Anwendung von Kalman-Filtern zur Glättung der Signale beruhen [GBW98, Pud03]. Häufiger werden aber Verfahren in Frequenzbändern verwendet, wobei die Wahl der Methode zur Frequenzbereichs-Transformation ein eigenständiges Gebiet der Optimierung von Geräuschreduktionsalgorithmen ausmacht. Als Standard haben sich Verfahren basierend auf der schnellen Fourier-Transformation (FFT) etabliert [BSM79, Cro80]. Aber auch Wavelet-Filterbänke [SVD93, GEH98] oder die Karhunen-Loewe Transformation wurden getestet [MP00]. Die Ergebnisse zeigten aber bisher keinen signifikanten Gewinn gegenüber dem Standard-Verfahren. Ein wesentlicher Baustein für eine gut funktionierende Geräuschreduktion ist die Schätzung des Hintergrundrauschens und des aktuell vorliegenden SNRs. Hierfür wurde von Martin das auf der Minimum Statistik basierende Verfahren vorgestellt und erweitert [Mar94, Mar01]. Aufbauend auf diesen Verfahren wurden weitere Verbesserungen entwickelt, die eine aufwandsreduzierte Schätzung des Hintergrundrauschens ermöglichen [Dob95, Coh03]. Als SNR-Schätzer wurden verschiedene Verfahren vorgestellt, wobei sich das rekursive Schätzverfahren nach Ephraim und Malah [EM84, Cap94] als Grundlage durchgesetzt hat und in neueren Veröffentlichungen verbessert wurde [SV96, Coh04a, Coh04b].

Die eigentliche Geräuschreduktion erfolgt durch eine zeitvariante Abschwächung einzelner Teilbänder. Das grundsätzliche Verfahren wurde dabei bereits theoretisch von N. Wiener vorgestellt [Wie49]. Moderne Verfahren erweitern das bekannte Wiener-Filter und dessen Implementierung in Form der spektralen Subtraktion [BSM79] durch eine andere Berechnung der Verstärkungen. Besonders oft genutzt wird dabei das Verfahren nach Ephraim und Malah, das die optimale Verstärkungsregel unter der Bedingung berechnet, dass die logarithmischen Amplituden des Nutzsignals in den einzelnen Frequenzkanälen im Sinne des kleinsten quadratischen Fehlers minimiert werden sollen [EM85, Cap94]. Erweiterungen, die sich von der bisherigen Annahme der Gauß-Verteilung der Amplitudenwerte in den einzelnen Frequenzkanälen trennen, wurden kürzlich vorgestellt [LV03, Mar05]. Ob die Annahme einer nicht Gauß-förmigen Verteilung auch für Musiksignale zu halten ist, wurde bisher nicht gezeigt. Eine weitere Klasse an Dämpfungsregeln orientiert sich an der Wahrnehmung des Menschen, um eine klanglich verbesserte Regel zu entwerfen. Hierbei wird versucht, Erkenntnisse aus der psychoakustischen Forschung zum Maskierungsverhalten von Nutz- und Störsignalen auszunutzen [Vir99, GJV98, WG00]. Anwendungen der oben beschriebenen Ansätze für die Restauration von Musiksignalen finden sich u. a. in [CL95, CN06].

Geräuschreduktion von breitbandigen Störungen in Audio-Signalen ist Stand der Technik, und es gibt viele Lösungen auf dem Markt, wobei die professionellen Produkte (CEDAR Cambridge, Sonic NONOISE, DigiDesign DINR, Sony Oxford DeNoise, Cube-Tec SpectralDeHiss, Waves DeNoiser, TC Electronic DeNoiser) alle auf gut abgestimmte Standardverfahren, basierend auf der FFT, setzen. Einige bieten eine automatische Hintergrundgeräuschschätzung, aber die meisten benötigen einen sog. Fingerprint des Hintergrundgeräusches, in dem kein Nutzsignal vorhanden ist. Dieser Fingerprint muss vom Nutzer ausgewählt und dem Algorithmus in einer Lernphase zugeführt werden. Die Verfahren müssen alle in vielen weiteren Parametern von den Nutzern angepasst werden, um zu verhindern, dass durch die Geräuschreduktion hörbare Artefakte wie Nutzsignaldämpfung oder sog. Musical Tones auftreten. Es ist bisher nur das Verfahren von Cedar als vollautomatisches Verfahren bekannt, wobei aber keine Aussage bez. der Qualität bei einer breiten Auswahl an Eingangssignalen möglich ist.

Impulshafte Störungen:

Impulshafte Störungen sind für alte Platten- und Zylinderaufnahmen typische Störungen. Die Klassifikation erfolgt über die Dauer und Amplitude der Störung. Die Hauptklassen bilden dabei die lauten und längere Signalabschnitte verdeckenden Clicks, verursacht durch große Kratzer, und sog. Crackles, die eine hohe Dichte aufweisen und durch Staub und weitere Ablagerungen in der Spurrille verursacht wurden [KB98, Vas96].

Die Restauration solcher Störungen erfolgt in zwei Schritten; der Detektion der temporären Störung und der Interpolation der gestörten Bereiche. Bei der Detektion werden unterschiedliche Verfahren angewandt, wobei besonders modellbasierte Algorithmen erfolgreich genutzt werden. Diese beruhen darauf, dass die Statistik der Störung nicht der Statistik und dem Modellverhalten von Sprache oder Musik entspricht [Vas96, GR98]. Für die Interpolation werden ebenfalls modellbasierte Verfahren vorgeschlagen, da diese für kleinere Störungen eine fast unhörbare Interpolation ermöglichen [Vas96]. Bei größeren Lücken sind Verfahren im Frequenzbereich erfolgreich angewendet worden [God93]. Neuere Verfahren nutzen zusätzlich eine vorherige Filterbankzerlegung und die Interpolation auch in nicht-kausaler Richtung, um sehr gute Ergebnisse auch bei größeren Lücken zu erzielen [EB06].

Kommerzielle Lösungen sind wie für die breitbandige Geräuschreduktion am Markt fest etabliert. Die professionellen Lösungen (Cedar Cambridge, Sonic NoNoise, Waves DeClicker, Sony deClick, Cube-tec DeClicker, Cube-Tec DeCrackler) basieren auf den oben genannten Prinzipien mit gut angepassten internen Parametern. Auch bei diesen Algorithmen ist eine direkte Nutzerinteraktion notwendig, um den richtigen Schwellwert zur Detektion zu definieren, da sonst das Nutzsignal angegriffen wird. Insgesamt sind aber im Vergleich

zur breitbandigen Geräuschreduktion weniger Parametereinstellungen notwendig, um befriedigende Ergebnisse zu erzielen. Als vollautomatische Lösung sind die Monte Carlo Markov Chain (MCMC)-basierten Algorithmen von Godsill und Rayner bekannt [GR98]. Die benötigte Rechenleistung ist bei diesen rein statistischen Verfahren sehr hoch und die Ergebnisse zeigen bisher keine signifikanten Vorteile gegenüber gut eingestellten Standard-Verfahren.

Determinierte Störungen:

Determinierte Störungen treten insbesondere durch Fehler in der Überspielung oder bei der Produktion auf. Typische Beispiele sind tieffrequente Einstreuungen aus dem Stromnetz durch mangelnde Erdung, im Englischen als Hum bezeichnet. Durch eine schlechte Abschirmung, insbesondere bei Live-Aufzeichnungen können die Phasenanschnittsteuerungen der Lichtdimmer regelmäßige kleine Pulse in Mikrofonsignale einstreuen und so zu einer hochfrequenten Brummstörung führen, die als Buzz bezeichnet wird.

Bei der Überspielung auf digitale Medien kann es vorkommen, dass die Dynamik des Eingangssignals unterschätzt wurde und so digitale Übersteuerungen entstehen (Clipping).

Für alle drei Klassen von Störungen sind professionelle Lösungen bekannt, aber die jeweiligen Verfahren sind nicht veröffentlicht, sondern nur in Produkten realisiert (Cedar DeBuzz/DeClip, Cube-Tec DeBuzz/ DeClip, Sony Oxford DeBuzz/DeClip). Die Produktbezeichnungen sind hierbei verwirrend, da die vorgestellten DeBuzz Algorithmen eigentlich DeHum-Algorithmen sind. Üblicherweise lassen sich Buzz-Störungen durch DeCrackler bereinigen, da die Störung durch kleine Clicks in regelmäßigem Abstand erzeugt werden. Es ist kein Verfahren bekannt, das die Regelmäßigkeit als a-priori Wissen ausnutzt. Zur Restauration von übersteuertem Material können speziell angepasste modellbasierte Interpolatoren verwendet werden, wie sie für die impulshaften Störungen bereits vorgestellt wurden.

Die als Hum bezeichnete Störung entsteht, wie oben erwähnt, über Einstreuung aus dem Stromnetzes. Aus diesem Grund besitzen Hum-Störungen meist eine Grundfrequenz, die der Netzfrequenz entspricht, also 50Hz in Europa bzw. 60Hz in den USA. Diese kann durch Gleichlaufschwankungen bei analogen Aufzeichnungsmedien aber zeitlich variieren. Aus diesem Grund haben alle professionellen Restaurationslösungen adaptive Grundfrequenzverfolger, wobei durch Nutzeraktion die Grundfrequenz grob eingestellt werden muss. Zur zeitvarianten Feinanpassung können adaptive Notch-Filter verwendet werden [Reg91]. Da die Störung zusätzlich aus vielen Obertönen besteht, besitzen die Restaurations-Algorithmen zusätzlich eine Vielzahl von Notch-Filtern, die in ihrer Bandbreite und Anzahl vom Nutzer gesteuert werden. Auch für diese Verfahren sind bisher keine halb- oder vollautomatischen Lösungen bekannt.

Bekannte Konstruktionen, Verfahren, Schutzrechte

Die Patentrecherche in der US-Patentdatenbank und der europäischen Datenbank Espacenet ergab folgende in der Summe nur wenig relevante Patente. Sie stehen einer Verwertung nicht im Wege, da die Firma Cube-Tec andere technische Lösungen nutzt:

- US6795559, Kenichi Taura et al, "Impulse noise reducer detecting impulse noise from an audio signal"
- US4630304, David E. Borth et al, "Automatic background noise estimator for a noise suppression system"
- US6965860, David Llewellyn Rees et al, "Speech processing apparatus and method measuring signal to noise ratio and ..."

- US6032114, Joseph Chan , "Method and apparatus for noise reduction by filtering based on a maximum..."
- WO200608149, Y Shinichi; S. Tetsu, N. Yoshihisa, "Audio restoration apparatus and audio restoration method"

Bei der Vielzahl von Patenten (>500), die Geräuschreduktion für Sprache und Signalverbesserung zum Thema haben, kann nie ausgeschlossen werden, dass relevante Patente übersehen wurden. Gleichzeitig ist das angestrebte Projekt aber in einem Nischenmarkt angesiedelt, so dass die direkte Anwendung nicht betroffen sein dürfte.

Im Projekt verwendete Fachliteratur

- Fachzeitschriften:
 - KEYS
 - Sound on Sound
- Wissenschaftliche Zeitschriften:
 - Computer Music Journal
 - Journal of the AES
 - IEEE Signal Processing Magazine
 - IEEE Transactions on Signal Processing
 - IEEE Transactions on Speech and Audio Processing
- Tagungsbände von jährlich oder halbjährlich stattfindenden Konferenzen:
 - Convention of the AES
 - Digital Audio Effects (DAFX)
 - International Symposium on Music Information Retrieval (ISMIR)
- Informations- und Dokumentationsdienste: An Informations- und Dokumentationsdiensten wurden die elektronischen Bibliotheken der AES, der IEEE ("Xplore") sowie Suchdienste wie Google Scholar verwendet. Die Patentrecherche erfolgte über die Web-Angebote des deutschen und des US-amerikanischen Patentamts.

Zusammenarbeit mit anderen Stellen

Die Zusammenarbeit mit den Projektpartnern war sehr konstruktiv und harmonisch. Die Firma Cube-Tec brachte vor allem ihr immenses Know-How im Bereich automatischer Bearbeitung großer Datenmengen auch über nicht verlässliche Netzwerke ein. Zusätzlich konnte die Jade Hochschule jederzeit auf die Experten zur Evaluation der erzielten Klangqualität zugreifen.

Die Zusammenarbeit mit der Universität Oldenburg verlief ebenfalls sehr gut und führt noch zu einem gemeinsamen Promotionsverfahren, das hoffentlich im nächsten Jahr abgeschlossen werden kann. Insbesondere die zusätzliche Betreuung durch Prof. Dr. Simon Doclo führte zu neuen Erkenntnissen und zu einer

wiss. Fokussierung. Die regelmäßigen Treffen der Arbeitsgruppe an der Universität helfen ebenfalls, die Kultur einer wissenschaftlichen Einrichtung mit ihren diversen Einflüssen aufzunehmen und für die eigene Entwicklung zu nutzen.

Ergebnisse

Kontinuierliches Rauschen

In diesem Abschnitt werden die gewonnenen Erkenntnisse zur automatischen Reduktion breitbandiger Rauschstörungen in Tonaufnahmen dargestellt. Aufgrund des Umfangs der Arbeiten wird an verschiedenen Stellen auf im Rahmen dieses Projektes entstandene Artikel und Abschlussarbeiten verwiesen.

Grundlagen

Grundlage des entwickelten Verfahrens zur Entrauschung ist die *Kurzzeit-Spektraldämpfung* (Short-Time Spectral Attenuation – STSA) in Kombination mit einem Algorithmus zur Schätzung der im Signal enthaltenen Rauschstörung, der auf der Minimum-Statistics Methode [Mar01] basiert. Es hat sich herausgestellt, dass die meisten der bekannten Verfahren zur Reduzierung einer additiven Rauschstörung für die Anwendung auf Sprachsignale spezialisiert und nicht ohne weiteres für die Entrauschung von Musikaufnahmen geeignet sind. Aus diesem Grund wurden an verschiedenen Stellen Erweiterungen und Modifikationen der bekannten Verfahren untersucht und umgesetzt.

Problemformulierung

Wie im Arbeitsplan vorgesehen wurden der Entwicklung des neuen Algorithmus eine Anzahl von Referenzsignalen, die von einem Experten mit Hilfe eines vorhandenen kommerziell erfolgreichen Algorithmus restauriert wurden, zugrundegelegt. Die Einstellung der Parameter bekannter Algorithmen stellt offensichtlich stets einen Kompromiss zwischen dem erreichten Grad der Reduzierung der Rauschstörung einerseits und einer Natürlichkeit des verbleibenden Restrauschens und der Unversehrtheit des Nutzsignals andererseits dar. In Bezug auf die Unverzerrtheit des Nutzsignals gelten für einige der verantwortlichen Algorithmenparameter bestimmte „Regeln“ – zum Beispiel sollten Glättungszeiten stets so kurz wie möglich gewählt werden, um ein spektrales „Verschmieren“ zu minimieren und eine daraus resultierende, unnatürliche Verhallung zu vermeiden. Zu kurze Glättungszeiten führen hingegen zu einer starken Fluktuation der Übertragungsfunktion des Entrauschfilters, resultierend in einer Reststörung, die oft als *Musical Noise* bezeichnet wird – eine Art „Plätschern“ des Restrauschens. In Bild 2 ist ein exemplarischer Verlauf einer Entrauschfilter-Übertragungsfunktion über der Zeit dargestellt. Es ist zu erkennen, dass sich deren Werte an zeitlich sowie spektral benachbarten Stellen in der Regel stark unterscheiden. Im Ausgangssignal kann diese Tatsache zu einem unregelmäßigen, schnellen Ein- und Ausschalten einzelner Sinustöne führen, resultierend in Musical Noise.

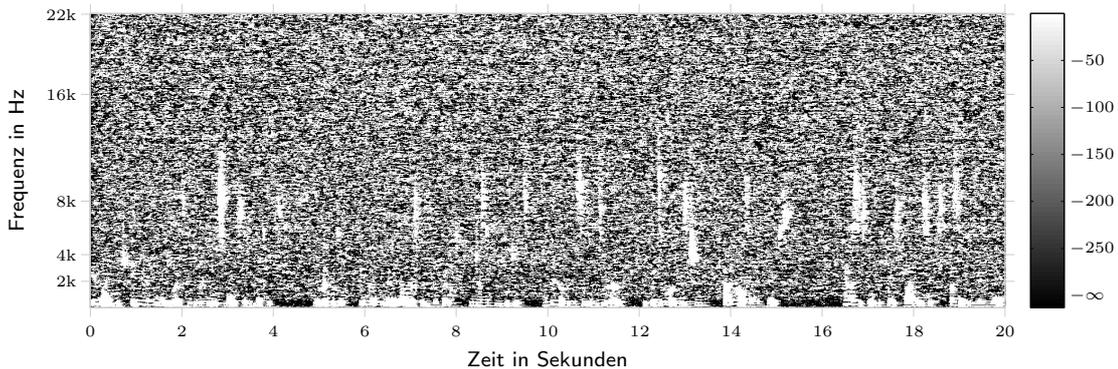


Bild 2: Darstellung der Musical Noise-Störung im Frequenzbereich. Abgebildet ist der Verlauf der logarithmierten Gewichtungsfunktion $20 \cdot \log_{10}(H[n])$. Es ist deutlich zu erkennen, dass sich deren Werte an zeitlich sowie spektral benachbarten Stellen stark unterscheiden. Für dieses Beispiel ist die Gewichtungsfunktion weder zeitlich noch spektral geglättet worden.

Optimierungen

Um in Hinblick auf ein gut klingendes Ausgangssignal nicht auf kurze Glättungszeiten verzichten zu müssen, und trotzdem eine Musical Noise-Störung zu vermeiden, wurden verschiedene Ansätze verfolgt:

- Begrenzung der maximalen Dämpfung
- Einführung eines (automatisch ans Eingangssignal angepassten) Rausch-Überschätzfaktors
- spektrale Glättung der Übertragungsfunktion des Entrauschfilters

Als wirksame Maßnahme gegen die Musical Noise-Störung hat sich eine Kombination der genannten Methoden herausgestellt (siehe auch Abbildung 12), wobei die spektrale Glättung der Übertragungsfunktion des Entrauschfilters besonders effizient ist. In diesem Zusammenhang sind zwei verschiedene Ansätze betrachtet worden:

- Spektralgättung mit konstanter Bandbreite und
- Spektralgättung mit frequenzabhängiger Bandbreite

Wichtig bei der Spektralgättung der Entrausch-Gewichtungsfunktion ist die korrekte Wahl der für die Glättung verwendeten Bandbreite. Bei korrekter Wahl dieses Parameters kann eine wahrnehmbare Verzerrung des Nutzsignals verhindert werden. Die Suche nach eben diesen sinnvollen Parametereinstellungen hat – unter anderem durch umfangreiche Hörtests – die Erkenntnis geliefert, dass geeignete Parameterwerte großenteils nicht von der Art des Eingangssignals (beispielsweise Musik, Sprache, etc.) abhängen, sondern dass es offensichtlich allgemeingültig optimale Einstellungen gibt, die im Rahmen von informellen und formellen Hörtests ermittelt werden konnten. Eine Adaption auf verschiedene Eigenschaften des Eingangssignals ist somit in vielen Fällen nicht erforderlich. Die in diesem Zusammenhang gemachten Erkenntnisse wurden auf der internationalen Fachkonferenz „EUSIPCO“ vorgestellt (siehe auch [BB09]). Die dort veröffentlichten Ergebnisse werden im nächsten Abschnitt kurz zusammengefasst.

Spektralglättung der Gewichtungsfunktion

Das Prinzip der Spektralglättung der Entrauschfilter-Übertragungsfunktion ist in Bild 3 dargestellt. Im oberen Teil des Bildes ist die ursprüngliche, zum Beispiel ausgehend von (3) berechnete, ungeglättete Übertragungsfunktion. Durch Faltung mit einer Fensterfunktion entsteht die geglättete Ausgangs-Übertragungsfunktion (unten im Bild).

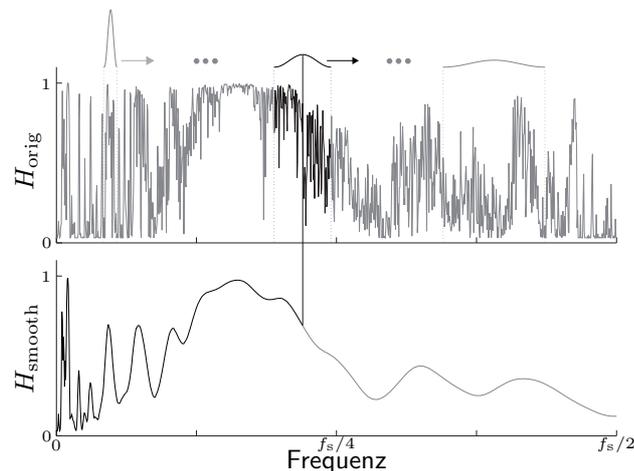


Bild 3: Spektralglättung durch Faltung im Frequenzbereich. Eine bestimmte, in diesem Beispiel frequenzabhängige, Anzahl von Spektralwerten der original Übertragungsfunktion H_{orig} wird mit einer Fensterfunktion gewichtet und dann aufsummiert um einen Spektralwert des geglätteten Spektrums H_{smooth} zu ergeben. Es ist die erste Hälfte des Spektrums (positive Frequenzen) dargestellt.

Um diejenige Art der Spektralglättung – und insbesondere die zu verwendende Bandbreite – mit der höchsten erzielten Ausgangssignalqualität zu ermitteln, sind in [BB09] objektive Maße sowie subjektive Hörtests verwendet worden. Da sich gezeigt hat, dass die technischen, objektiven Maße nicht in der Lage sind, diejenigen Effekte, die zur Klangqualitätsbeeinträchtigung führen, abzubilden, werden ausschließlich die Ergebnisse der Hörversuche in der Beurteilung berücksichtigt. Zusammenfassend konnte festgestellt werden, dass Spektralglättung mit mittleren Bandbreiten (zum Beispiel 200Hz . . . 500Hz oder 1/4oct) einen positiven Einfluss auf die Klangqualität des Ausgangssignals hat. Zu große Bandbreiten allerdings führen zu einer als schlechter empfundenen Qualität des entrauschten Signals. Andersherum führen zu geringe Bandbreiten zu einer ungenügenden Reduktion der Musical Noise Artefakte. Bild 4 zeigt die Ergebnisse der Hörtests, die in Form von Paarvergleichen durchgeführt und mit Hilfe des Bradley-Terry-Luce (BTL) Modells ([BT52], [Luc59]) ausgewertet wurden.

Der Decision-Directed-Ansatz

Ein weit verbreiteter Ansatz die Musical Noise-Störung zu reduzieren, besteht in einer zeitlichen Glättung der in die Berechnung der Gewichtungsfunktion eingehenden Größen. In diesem Abschnitt wird dieser sogenannte *Decision-Directed-Ansatz* vorgestellt.

Neben einer speziellen Gewichtungsfunktion, die hier nicht weiter betrachtet werden soll, wird in [EM84] unter anderem eine Möglichkeit gezeigt, die Musical Noise-Störung durch geschickte zeitliche Glättung zu reduzieren. Das vorgestellte Verfahren basiert auf der folgenden, rekursiv geglätteten Schätzung des a priori

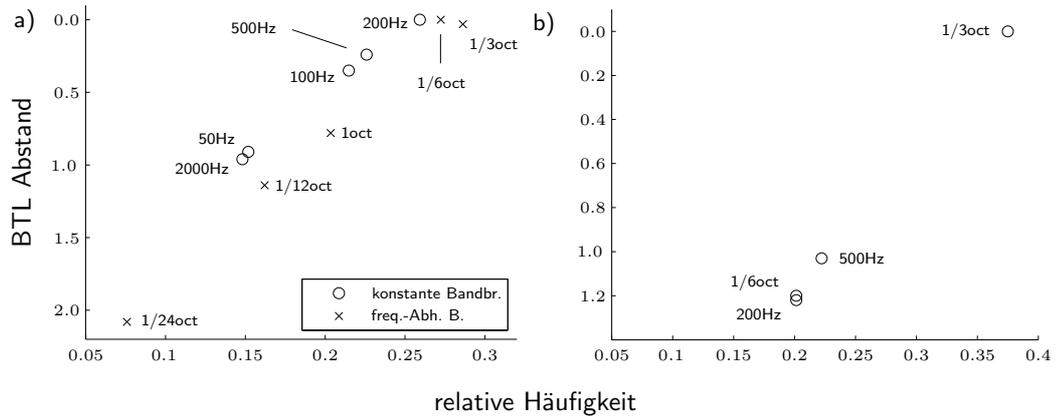


Bild 4: Berechnete BTL-Modell Abstände aufgetragen über der relativen Häufigkeit der Probandenurteile. Die Ergebnisse von zwei Hörtestreihen, um die bevorzugte Konstante- und Oktav-Bandbreite zu ermitteln sind in Bild a) zu sehen. Die Ergebnisse eines zweiten Hörtestdurchlaufs, um die insgesamt bevorzugte Art der Spektralglättung zu ermitteln, werden in Bild b) dargestellt.

Signal-Rausch-Abstands [EM84]:

$$\hat{\text{SNR}}_{\text{prio, DD}}^{(p)}[n] = \alpha_{\text{SNR}} \cdot \frac{|\hat{S}^{(p-1)}[n]|^2}{\hat{P}_{nn}[n]} + (1 - \alpha_{\text{SNR}}) \cdot \text{R} \left\{ \hat{\text{SNR}}_{\text{post, inst}}^{(p)}[n] - 1 \right\}, \quad (1)$$

worin

$$\text{R} \{x\} = \max(x, 0) = \begin{cases} x & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

der Operator einer Halbwellegleichrichtung und

$$\hat{\text{SNR}}_{\text{post, inst}}[n] = \frac{|X[n]|^2}{\hat{P}_{nn}[n]} \quad (2)$$

eine instantane Schätzung des bereits erwähnten a posteriori Signal-Rausch-Abstands ist. Mit der Definition (1) des a priori SNR kann das Wiener-Filter wie folgt ausgedrückt werden:

$$H_{\text{DD}}[k] = \frac{\text{SNR}_n^s[k]}{\text{SNR}_n^s[k] + 1} \quad (3)$$

beziehungsweise

$$H_{\text{DD}}[n] = \frac{\hat{\text{SNR}}_{\text{prio, DD}}[n]}{\hat{\text{SNR}}_{\text{prio, DD}}[n] + 1}. \quad (4)$$

In [Cap94] wird das Verhalten des auf diese Weise modifizierten Wiener-Filters nach (4) wie folgt anhand des Ergebnisses einer experimentellen Untersuchung interpretiert:

- Wenn der instantane a posteriori Signal-Rausch-Abstand (2) gering ist (\approx unter 0dB) entspricht die Schätzung des a priori Signal-Rausch-Abstands einer geglätteten Version des a posteriori Signal-Rausch-Abstands. Die Varianz der a priori SNR-Schätzung ist deutlich geringer als die des a posteriori SNR.
- Wenn auf der anderen Seite der a posteriori Signal-Rausch-Abstand viel größer ist als 0dB, folgt die a priori SNR-Schätzung mit einer Verzögerung, die von der Größe des Glättungsfaktors α_{SNR} abhängt.

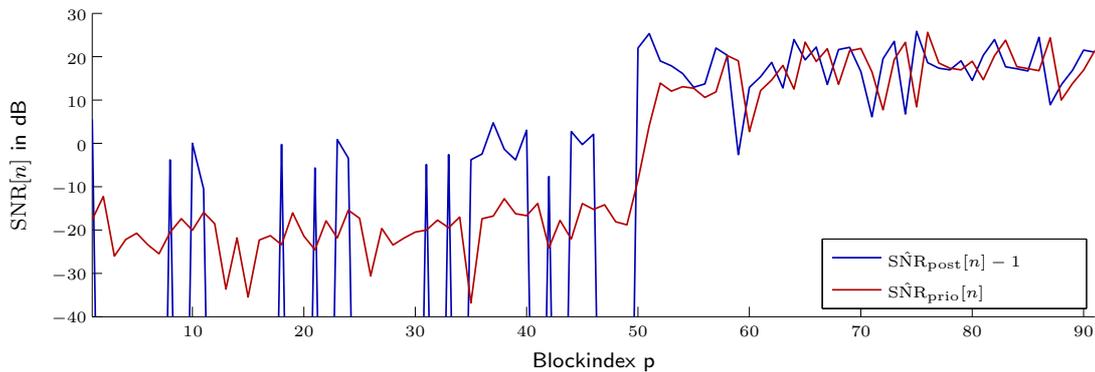


Bild 5: Verlauf der Schätzwerte für *a posteriori* und *a priori* Signal-Rausch-Abstand mit dem *Decision-Directed*-Ansatz. Bei Block 50 steigt die Leistung des Nutzsignals plötzlich an. Um die Kurven besser vergleichen zu können, ist $\hat{\text{SNR}}_{\text{post}}[n] - 1$ dargestellt. Der Frequenzindex ist $n = 50$, der Glättungsfaktor beträgt $\alpha_{\text{SNR}} = 0.98$.

Durch dieses glättende Verhalten (vergleiche auch Bild 5) und eine daraus resultierende geringe Varianz der in die Berechnung der Gewichtungsfunktion einfließenden Schätzung des *a priori* SNR, kann dessen Fluktuation, und damit die Musical Noise-Störung, stark reduziert werden.

In Bild 5 ist zu erkennen, dass die Schätzung des *a priori* Signal-Rausch-Abstands bei ansteigender Nutzsignalleistung nur mit einer gewissen Verzögerung dem tatsächlichen Verlauf des SNRs folgt. Als Folge kann es passieren, dass der Dynamik des Nutzsignals nicht mehr gefolgt werden kann – Transienten, also kurze und schnelle Veränderungen im Zeitsignal, wie zum Beispiel die Einschwingphase einer Trommel oder das Anschlaggeräusch einer Saite, werden herausgefiltert, weil die SNR-Schätzung nicht schnell genug dem Signalverlauf folgen kann.

Im nächsten Abschnitt wird ein Verfahren vorgestellt, das die positiven Effekte des *Decision-Directed*-Ansatzes – nämlich die Glättung der SNR-Schätzung – beibehält, die negativen Effekte – die zeitliche Verzögerung der SNR-Schätzung – jedoch vermeidet. Hierzu wird die Tatsache ausgenutzt, dass das vollständige Signal zur Verfügung steht, und eine *Entrauschung in zwei Schritten* wird durchgeführt.

Entrauschung in zwei Schritten

Der im letzten Abschnitt vorgestellte *Decision-Directed*-Ansatz ermöglicht eine effektive Reduzierung der Musical Noise-Störung bei ausreichend großen Glättungsfaktoren α_{SNR} (siehe (1)). Große Glättungsfaktoren allerdings bringen einen Nachteil mit sich: Die Schätzung des *a priori* Signal-Rausch-Abstands kann nicht schnell genug dem tatsächlichen SNR-Verlauf folgen, und als Konsequenz werden impulshafte Anteile des Nutzsignals durch das Wiener-Filter aus dem Signal entfernt. Die Ursache des Problems liegt offenbar darin begründet, dass zur Schätzung des *a priori* SNRs des aktuellen Blocks, das Spektrum des *letzten* Ausgangsblocks $\hat{S}^{(p-1)}[n]$ eingeht.

Es hat sich gezeigt, dass besser auf plötzlich ansteigende Signalpegel reagiert werden kann, wenn die *Entrauschung in zwei Schritten* durchgeführt wird. Der erste Schritt wird lediglich zur Bestimmung von $\hat{S}_{\text{ES}}^{(p)}[n]$ benötigt, welches daraufhin im zweiten Schritt zur Berechnung einer schnell folgenden Schätzung

des a priori Signal-Rausch-Abstands verwendet wird. (1) nimmt dann die folgende Form an:

$$\hat{\text{SNR}}_{\text{prio}}^{(p)}[k] = \alpha_{\text{SNR}} \cdot \frac{|\hat{S}_{\text{ES}}^{(p)}[k]|^2}{\hat{P}_{nn}[k]} + (1 - \alpha_{\text{SNR}}) \cdot \text{R} \left\{ \hat{\text{SNR}}_{\text{post}}^{(p)}[k] - 1 \right\}, \quad (5)$$

wobei

$$\hat{S}_{\text{ES}}^{(p)}[k]$$

die Schätzung des Nutzspektrums des ersten Schrittes (daher der Index „ES“) ist. Durch informelle Hörversuche konnte festgestellt werden, dass die Entrauschung in zwei Schritten deutliche Vorteile gegenüber der Entrauschung in einem Schritt aufweist. Indem Transienten im Signal erhalten bleiben, wirkt der Klang „präziser“ gegenüber dem als „dumpher“ empfundenen Entrauschungsergebnis in einem Schritt. Spektrogramme der Ausgangssignale der Entrauschung in einem (siehe Bild 6) und zwei Schritten (siehe Bild 7) verdeutlichen die Unterschiede. Beim Vergleich der Spektrogramme in Bild 6 und 7 werden nur

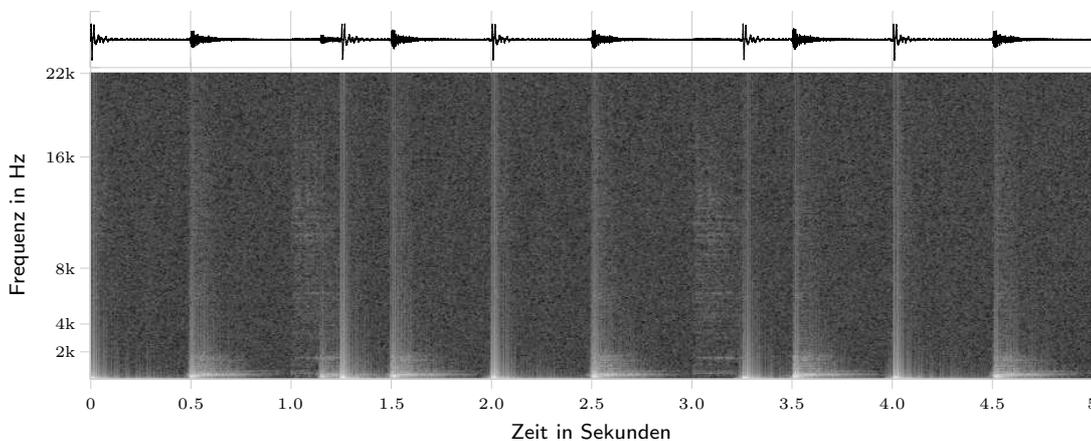


Bild 6: Spektrogramm des Ausgangssignals einer Entrauschung in einem Schritt. Das Nutzsignal ist ein Ausschnitt aus einer Schlagzeugaufnahme. Dargestellt sind die ersten 256 Frequenzbins des rekursiv geglätteten ($\tau = 2\text{ms}$) Periodogramms ($N_{\text{DFT}} = 512$).

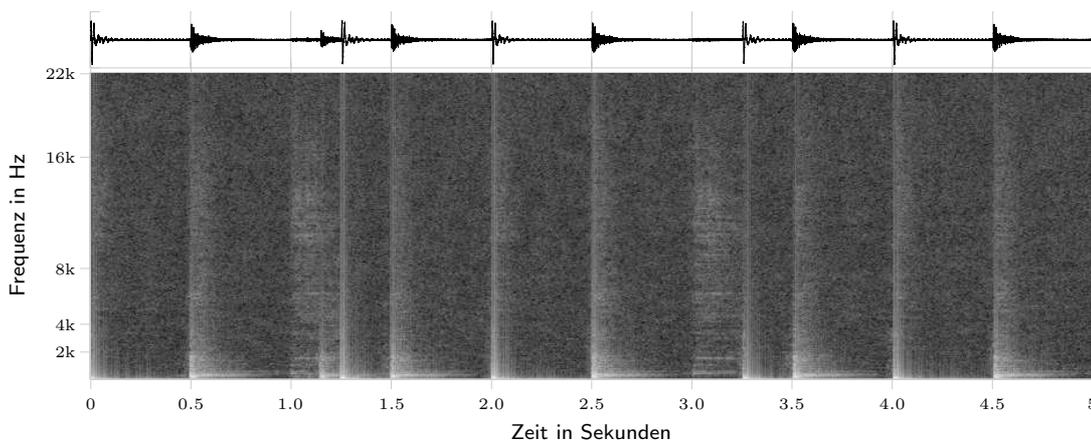


Bild 7: Spektrogramm des Ausgangssignals einer Entrauschung in zwei Schritten. Das Nutzsignal ist ein Ausschnitt aus einer Schlagzeugaufnahme. Dargestellt sind die ersten 256 Frequenzbins des rekursiv geglätteten ($\tau = 2\text{ms}$) Periodogramms ($N_{\text{DFT}} = 512$).

geringe Unterschiede deutlich. Es ist jedoch zu erkennen, dass die Transienten durch die Entrauschung in zwei Schritten breitbandig deutlicher hervortreten.

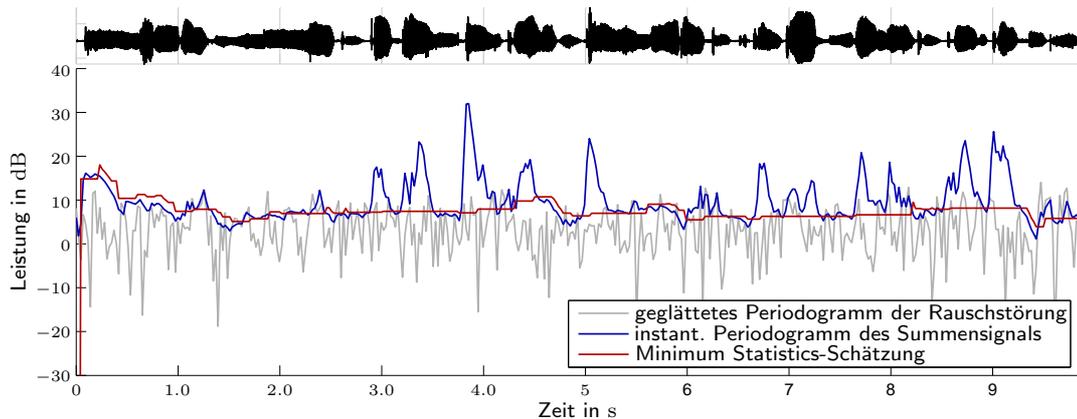


Bild 8: *Minimum Statistics-Rauschbodenschätzung bei einem Gesangssignal.* Eingangssignal ist ein Ausschnitt aus dem Stück „Tom’s Diner“ von Suzanne Vega; weißes Rauschen wurde hinzugefügt, um ein globales SNR von 10dB zu erhalten. Dargestellt sind die jeweiligen Leistungsdichten bei einer Frequenz von ungefähr 1,3kHz.

Durch die Aufteilung in zwei Schritte wird die Möglichkeit eröffnet, die Parameter jedes Schrittes unabhängig voneinander zu optimieren.

Schätzung der Rauschleistungsdichte

Bislang ist das Problem der Rauschbodenschätzung in Audiosignalen in der Literatur fast ausschließlich im Zusammenhang mit einer Anwendung auf Sprachsignale behandelt worden. Bei der Anwendung dieser Verfahren auf Musiksignale ergibt sich aufgrund deren Funktionsweise das generelle Problem, den Rauschboden stark zu überschätzen. Dieser Sachverhalt ist in den Bildern 8 und 9 dargestellt. Der im Zwischenbericht über das Jahr 2008 dargestellte Ansatz, dieses Problem zu überwinden geht – wie sich nun herausgestellt hat – von in in der Praxis nicht vorliegenden Bedingungen aus. Daher mussten weitere Untersuchungen angestellt werden, das Problem der Überschätzung zu verringern.

Als erster Schritt ist in diesem Zusammenhang ein Rahmenprogramm in MATLAB implementiert worden, um das Verhalten verschiedener bekannter Verfahren zur Rauschbodenschätzung zu evaluieren und deren Schwachstellen untersuchen zu können.

Im Vergleich zu alternativen Verfahren, wie zum Beispiel dem MCRA-Verfahren [CB02] oder quantilbasierten Verfahren [SFB00] bot das „Minimum Statistics“-Verfahren [Mar01] die besten Schätzergebnisse. Allerdings konnte dessen prinzipielle Neigung, Rauschböden bei Musiksignalen zu überschätzen, für eine Vielzahl von Testsignalen bestätigt werden. Wie sich gezeigt hat, liegt die Überschätzung der Rauschleistung in zwei Ursachen begründet:

- zu wenige bzw. keine Pausen im Nutzsignal, in denen die Gesamtleistung auf die Rauschleistung abfällt und
- zu wenige bzw. keine „Lücken“ im Spektrum.

Da die prinzipielle Funktionsweise dieses Algorithmus als sinnvoll betrachtet wurde ist versucht worden, beide Schwächen durch geeignete Maßnahmen zu verringern. Indem das Zeitfenster, in dem Minima in der Leistungsdichte gesucht werden, vergrößert wurde, konnte die Überschätzung der Rauschleistung in

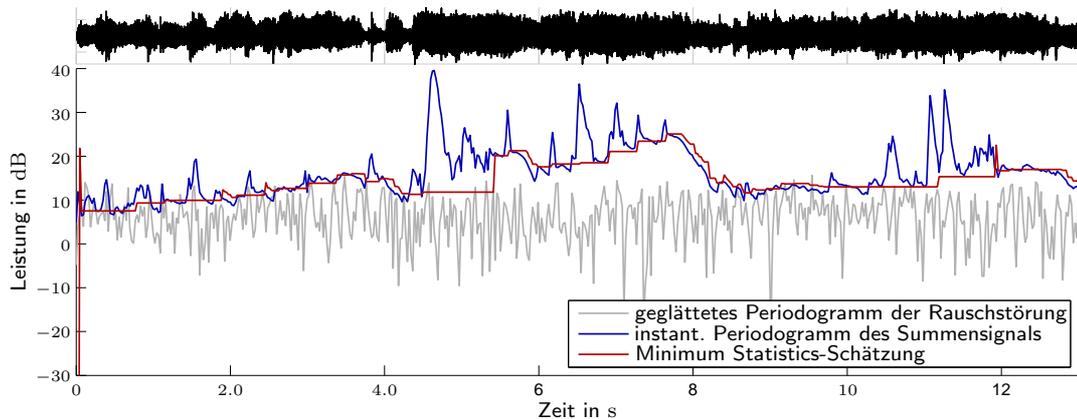


Bild 9: *Minimum Statistics-Rauschbodenschätzung bei einem Musiksignal.* Eingangssignal ist ein Ausschnitt aus dem Stück „St. Teresa“ von Joan Osborne; weißes Rauschen wurde hinzugefügt, um ein globales SNR von 10dB zu erhalten. Dargestellt sind die jeweiligen Leistungsdichten bei einer Frequenz von ungefähr 1,3kHz.

gewissem Umfang verringert werden. Aufgrund der internen Funktionsweise des ursprünglichen „Minimum Statistics“-Algorithmus war es hierzu notwendig, dessen Aufbau nachzuvollziehen und an mehreren Stellen Änderungen zu implementieren, um auch längere Zeitfenster verwenden zu können.

Um das Problem, in Musikaufnahmen nur unzureichende „Lücken“ in den Kurzzeit-Leistungsdichtespektren für die Rauschbodenschätzung auffinden zu können, zu verringern, sind zwei unterschiedliche Ansätze verfolgt worden. Zum einen ist der Einfluss verschiedener Fensterfunktionen zur Verringerung des Leckeffektes untersucht worden (siehe Bild 10), zum anderen konnten durch die Verwendung von unterschiedlichen Blocklängen in Verbindung mit entsprechenden Längen der diskreten Fouriertransformation (DFT) weitere Informationen aus dem Leistungsdichtespektrum des Eingangssignals extrahiert werden. In Bild 10 ist zu erkennen, dass die Wahl der Fensterfunktion nur geringen Einfluss auf das Ergebnis der Rauschbodenschätzung hat. Aufgrund der leichten Vorteile gegenüber dem üblicherweise verwendeten von-Hann-Fenster wird im entwickelten Algorithmus das Blackman-Fenster verwendet.

Naturgemäß bieten kurze Block- und DFT-Längen den Vorteil einer erhöhten zeitlichen Auflösung mit entsprechend verringerter spektraler Auflösung. Bei großen Blocklängen verhält es sich umgekehrt. Dieser Sachverhalt wird in Bild 11 dargestellt. Indem das Eingangssignal zweimal analysiert wird – einmal mit großer Blocklänge, einmal mit kleiner, kann durch die Bildung des Minimums beider Rauschbodenschätzungen der Grad der Überschätzung verringert werden.

Die Rauschstörung vieler Medien – seien es Tonband, Schallplatte, etc. – ist systembedingt als stationär anzunehmen. Um diese Tatsache zugunsten einer genaueren Schätzung der Rauschleistungsdichte auszunutzen, wird das Leistungsdichtespektrum der Rauschstörung einmalig für bestimmte Abschnitte ermittelt. Diese Abschnitte zeichnen sich dadurch aus, dass der Rauschboden konstant bleibt. Dies hat den Vorteil, dass – gegenüber eines üblicherweise verwendeten Minimum Statistics Algorithmus – nicht nur wenige Sekunden des gestörten Signals für die lokale Schätzung des Rauschbodens zu berücksichtigen, sondern ein längerer Signalabschnitt insgesamt. Um eben diese Abschnitte mit konstant anzunehmendem Rauschboden zu ermitteln, ist eine Vorverarbeitung erforderlich, die im nächsten Abschnitt beschrieben wird.

Ermittlung von Abschnitten mit stationärem Rauschboden Für Aufnahmen, die aus mehreren, hintereinander gereihten Einzelaufnahmen bestehen, kann nicht davon ausgegangen werden, dass das Lei-

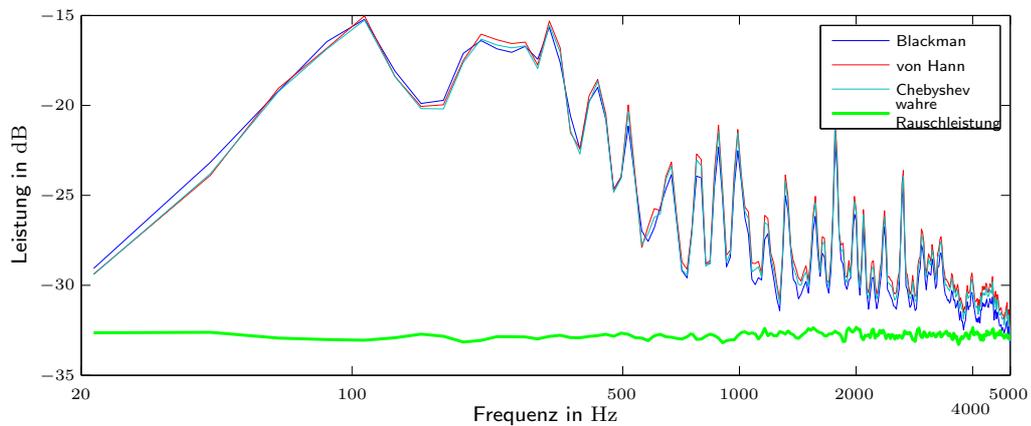


Bild 10: Einfluss verschiedener Fensterfunktionen auf den Grad der Überschätzung des Rauschbodens. Es zeigt sich, dass die Wahl der Fensterfunktion einen geringen Einfluss auf die Überschätzung hat, das prinzipielle Problem wird allerdings von keinem Fenster ausreichend verringert.

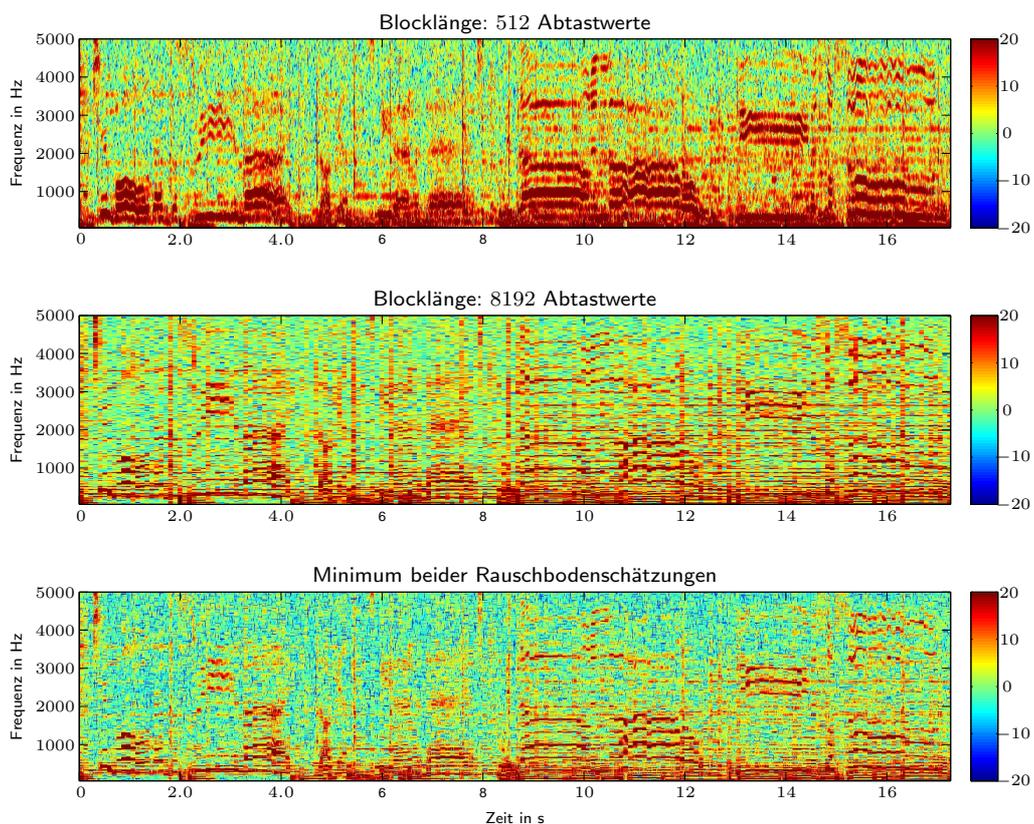


Bild 11: Reduzierung des Grades der Rauschüberschätzung durch Kombination von Schätzungen mit unterschiedlichen Blockgrößen. Die Spektrogramme zeigen jeweils den Unterschied des Schätzwertes zum wahren Wert der Rauschleistungsdichte. Der Wertebereich ist auf $[-20\text{dB} \dots 20\text{dB}]$ beschränkt, um die Übersichtlichkeit zu erhöhen.

stungsdichtespektrum der Rauschstörung während der gesamten Dauer der Aufnahme konstant bleibt. Beispielhaft seien Mitschnitte von Radiosendungen genannt, in denen Sprachbeiträge aus dem Rundfunkstudio mit Musikeinspielungen und Außenaufnahmen abwechseln. Für solche Aufnahmen ist davon auszugehen, dass sich die Rauschböden der einzelnen Abschnitte stark unterscheiden. Jeder Abschnitt für sich ist aber systembedingt näherungsweise konstant. Um auf diese wechselnden Signaleigenschaften reagieren zu können böte sich eine kontinuierlich erneuerte Rauschbodenschätzung an, die aber aus oben erläuterten Gründen vermieden werden sollte. Damit weitaus vorteilhaftere Verfahren, die die Stationarität innerhalb der Abschnitte berücksichtigen, zum Einsatz kommen können, müssen zunächst die Grenzen dieser Abschnitte ermittelt werden.

Das hierzu entwickelte Verfahren basiert in einem ersten Schritt – ähnlich wie das Minimum Statistics-Verfahren – auf einer Betrachtung der Minimalwerte einer Kurzzeit-Leistungsdichteschätzung innerhalb eines bestimmten Zeitfensters der Länge T . Entscheidend hierbei ist, ausschließlich Frequenzen oberhalb einer bestimmten Grenzfrequenz zu berücksichtigen. Dies hat das Ziel, Nutzsignaleinflüsse in der Schätzung zu minimieren. In unserem Verfahren wird das Frequenzband zwischen 14kHz und 18kHz betrachtet. Um die aus der Funktionsweise der Minimaverfolgung resultierenden Verzögerungen bei ansteigender Leistungsdichte zu umgehen, werden die Schätzungen zweier unabhängiger Vorwärts- und Rückwärtsverarbeitungen mittels Maximum-Operator kombiniert.

Um die Varianz der Schätzwerte zu reduzieren, wird eine Glättung über mehrere aufeinander folgende Blöcke mittels Medianfilter vorgenommen.

Nach dem in [NHK04] beschriebenen Verfahren wird anschließend nach Wechseln in der Breitbandleistung gesucht.

Blinde Bandbreitenerweiterung

Es wurde festgestellt, dass Aufnahmen, die mit einem Entrauschalgorithmus bearbeitet wurden, zu dumpfem Klang neigen. Dies beruht auf der Tatsache, dass hohe Frequenzen zu stark gedämpft werden. Um dieses Problem zu beseitigen, wurden Versuche zur sogenannten *blinden Bandbreitenerweiterung* angestellt. Dieser Begriff bezeichnet Verfahren, um nicht oder nur sehr schwach vorhandene hohe Frequenzanteile zu erzeugen beziehungsweise zu verstärken. Ziel ist, den dumpfen Klang von bearbeiteten Signalen zu vermeiden, indem die hohen Frequenzen vor oder nach der Bearbeitung mit einem Entrauschalgorithmus verstärkt werden. Unabhängig vom Einsatz im Rahmen von Entrauschalgorithmen können Verfahren zur blinden Bandbreitenerweiterung verwendet werden, um den Klang von typischen, bandbegrenzten historischen Aufnahmen durch Synthetisierung hoher Frequenzen zu verbessern.

Erkenntnis der umfassenden Beschäftigung mit dem Thema im Rahmen der Bachelorarbeit war schließlich, dass eine ausreichend hohe Klangqualität der Verarbeitung nicht erreicht werden konnte. Dies liegt hauptsächlich daran, dass die Genauigkeit, mit der hochfrequente Signalanteile synthetisiert und zum ursprünglichen Signal gemischt werden, nicht ausreicht. Aufgrund der Fähigkeit des menschlichen Ohres, auch sehr geringe Frequenzabweichungen einzelner Oberschwingungen in harmonischen Tonkomplexen festzustellen und als unangenehmen Klang wahrzunehmen, konnte der ursprüngliche Einsatz – als Erweiterung von Rauschreduktionssystemen – nicht realisiert werden.

Evaluation

Im Rahmen von informellen und formellen Hörtests konnte die Effizienz des entwickelten Verfahrens bestätigt werden. Es hat sich gezeigt, dass für eine Vielzahl unterschiedlicher Eingangssignale eine robuste

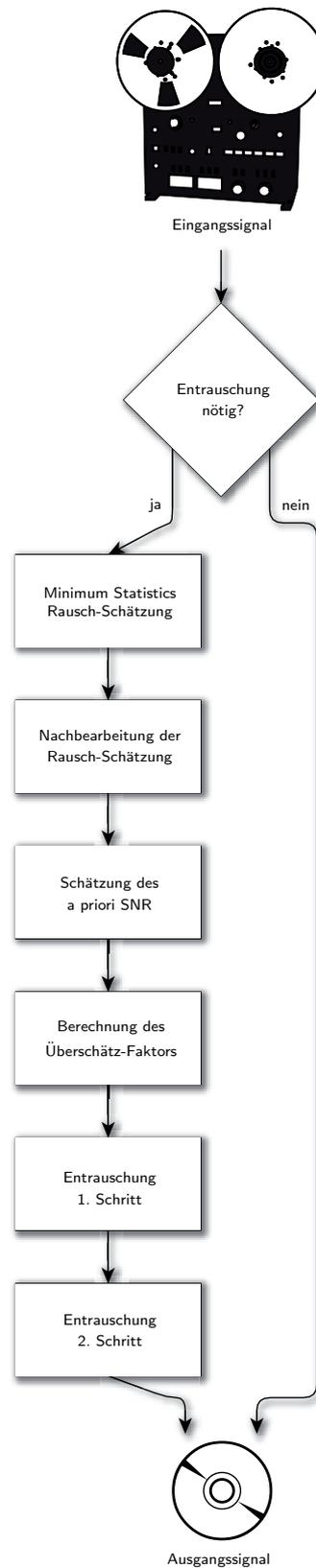


Bild 12: Aufbau des Systems zur Entrauschung.

Entrauschung ohne manuelles Justieren von Parameterwerten möglich ist.

Impulsstörungen

Eine zweite wesentliche Störung für altes Audiomaterial sind Impulsstörungen. Diese lassen sich durch unterschiedliche Signalmodelle beschreiben. Wesentlich ist aber immer, dass die Störung nur für kurze Momente hörbar ist. Die Lösung erfolgt deshalb immer in zwei Schritten, einer Detektion der Störungen im Zeitsignal und einer anschließenden geeigneten Interpolation zur Erzeugung eines ungestörten Signals [Vas96].

Problembeschreibung

Ausgehend von dem grundsätzlichen Signalmodell

$$y(k) = x(k) + n(k) \quad (6)$$

mit $x(k)$ als dem ungestörten Nutzsignal und $n(k)$ als Störung ist es möglich, die unterschiedlichen Impulsstörungen durch eine geeignete Wahl eines Rauschmodells zu modellieren. Üblicherweise wird für größere Störungen (Clicks) eine binäre Folge $i(k)$ verwendet, um die Anwesenheit einer Störung zu simulieren. Diese wird mit einem Gaußrauschen multipliziert und anschließend mit einem linearen, zeitinvarianten Modell $h(k)$ gefaltet.

$$y(k) = x(k) + (i(k)n(k)) * h(k) \quad (7)$$

Eine Alternative, die besonders für Impulsstörungen mit kleiner Amplitude und häufigem Auftreten (Crackles) oder dem granularen Rauschen bei Filmtonaufnahmen entspricht, ist das Rauschen mit einer geeigneten Verteilungsdichte zu generieren und anschließend mit dem typischen Übertragungsfiter $h(k)$ zu falten. Als Verteilungsdichten für $n(k)$ bieten sich dabei supergaußische Verteilungen wie Laplace oder Cauchy-Verteilungen an.

$$y(k) = x(k) + (n(k)) * h(k) \quad (8)$$

Erreichte Lösung

Die entscheidende Schwierigkeit für eine gute Restauration ist die Detektion, da für die Interpolation viele gute Verfahren bekannt sind [Vas96]. Das bekannteste Verfahren mittels Least-Square Autoregressive (LSAR)-Interpolation wird kurz vorgestellt. Die im Projekt neu entwickelte Detektion basiert auf der Idee des sog. Binary-Maskings mit anschließender auto-regressiver Filterung.

Detektion

Die Detektion ist in zwei Phasen unterteilt. Zunächst wird im Frequenzbereich ein möglichst großer Teil des Nutzsignals vom Störsignal getrennt. In einem zweiten Schritt wird durch eine AR-Vorfilterung das Signal so aufbereitet, dass kleinere Impulsstörungen deutlicher hervortreten. Der gesamte Algorithmus ist in Abbildung 13 gezeigt.

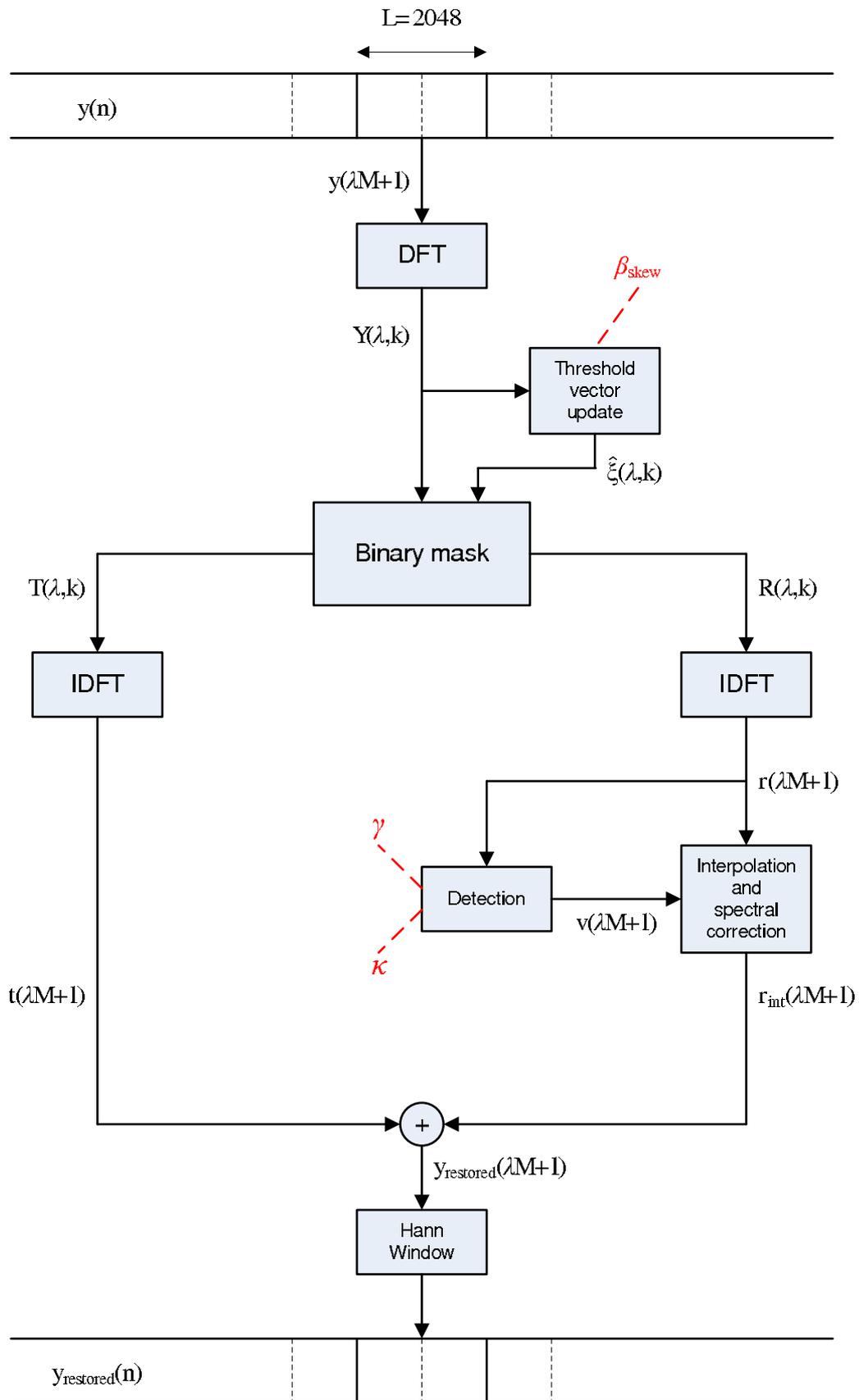


Bild 13: Blockschaltbild zur Verdeutlichung des Algorithmus zur Impulsreduktion

Die überlappende Blockzerlegung des gestörten Signals ist als erster Schritt notwendig, um eine Kurzzeit-Fourier-Transformation durchführen zu können. Das resultierende Betragsspektrum $|Y(\lambda, n)|$ mit dem Blockindex λ und dem Frequenzindex n wird anschließend durch eine sog. binäre Maske in ein Zielsignal und ein Restsignal zerlegt. Die Bestimmung dieser Maske ist der entscheidende Faktor. Bisher bekannt waren Lösungen, die mit einer festen vom Nutzer eingestellten Schwelle arbeiten, um das Ziel vom Restsignal zu trennen [Czy97]. In diesem Projekt wurde eine neue adaptive Schwelle entwickelt, die so eine Automation ermöglicht. Diese Schwelle $\hat{\xi}(\lambda, n)$ wird durch ein geglättetes rekursives Welch-Periodogramm geschätzt:

$$\hat{\xi}(\lambda, n) = \alpha(\lambda, n)\hat{\xi}(\lambda - 1, n) + (1 - \alpha(\lambda, n))|Y(\lambda, n)|^2 + \beta(n) \quad . \quad (9)$$

mit einem zeitabhängigen $\alpha(\lambda, n)$

$$\alpha(\lambda, n) = \begin{cases} \alpha_{\text{attack}} & \text{if } |Y(\lambda, n)|^2 > \hat{\xi}(\lambda - 1, n) \\ \alpha_{\text{release}} & \text{otherwise} \end{cases} \quad (10)$$

und

$$\alpha_{\text{attack}} = \exp\left(\frac{-M}{\tau_{\text{attack}} \cdot f_{\text{sampling}}}\right) \quad (11)$$

und

$$\alpha_{\text{release}} = \exp\left(\frac{-M}{\tau_{\text{release}} \cdot f_{\text{sampling}}}\right) \quad , \quad (12)$$

wobei f_{sampling} die Samplingfrequenz repräsentiert und M den Blockvorschub. Die Zeitkonstanten werden so gewählt, dass ein schneller Rückgang möglich ist, aber nur ein langsamer Anstieg. Dies führt dazu, dass die Schwelle dem Rauschboden folgt. Der additive Faktor $\beta(n)$ ist ein zusätzlicher Kompensationsfaktor, der es ermöglicht, Nutz- und Störsignal besser zu trennen. Ein guter Wert ist eine Addition eines Wertes, der mit ca. 5db/Oktave zunimmt.

Das Residualspektrum wird mit der Originalphase in den Zeitbereich zurück transformiert und dient als Basis für den finalen Detektionsschritt. Hierzu werden aus dem Restsignal die Koeffizienten für eine lineare Prädiktion bestimmt und der Prädiktionsrestfehler berechnet. Dies entspricht einer weiteren Entfärbung des Signals. Für die meisten Signale ist die Funktion eine angepasste Hochpass-Filterung die insbesondere die Störungen verstärkt und das Nutzsignal herausfiltert.

Vom Prädiktionsfehlersignal wird ein Histogramm gebildet, und eine gewisse vom Nutzer einstellbare Anzahl an Extremwerten wird in der abschließenden Interpolation restauriert. Das neu entstandene Signal kann abschließend mit dem nicht verarbeiteten Zielsignal gemischt werden, um so das entstörte Signal zu erhalten.

Interpolation

Für die Interpolation wird davon ausgegangen, dass sich fast alle natürlichen Signale durch ein autoregressives Modell beschreiben lassen

$$x(k) = \sum_{p=1}^P a_p x(k-p) + e(k) \quad (13)$$

mit der Ordnung des AR-Systems P und den AR-Koeffizienten a_p und einem Anregungssignals oder Restrauschens $e(k)$, das sich nicht durch das AR-Modell beschreiben lässt.

Das ungestörte Signal $x(k)$ liegt aber nicht vor, deshalb müssen zunächst aus dem gestörten Signal $y(k)$ die AR-Koeffizienten bestimmt werden. Hierfür sind verschiedene Methoden bekannt. Diese sind auch ziemlich robust gegenüber den hier eher als klein angenommenen Störungen. Aus den AR-Koeffizienten

und dem Detektionsvektor kann nun eine Prädiktionsmatrix erstellt werden (siehe Abbildung), aus der die Untermatrizen für den eigentlichen LSAR-Interpolator

$$\mathbf{y}_u^{LS} = -(\mathbf{A}_u^T \mathbf{A}_u)^{-1} \mathbf{A}_u^T \mathbf{A}_k \mathbf{y}_k \quad . \quad (14)$$

gebildet werden.

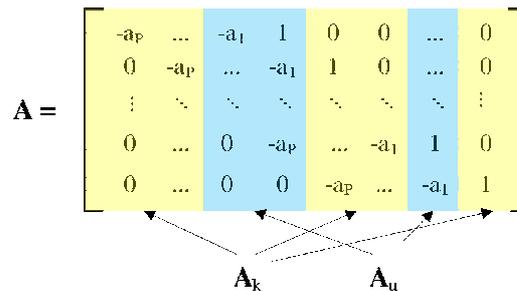


Bild 14: Partitionierung der AR Koeffizientenmatrix in die Matrix der unbekanntem \mathbf{A}_u und bekannten Teile \mathbf{A}_k

Evaluation

Um den neu entwickelten Algorithmus klanglich zu evaluieren, wurde ein Hörtest¹ mit zehn normalhörenden jungen Studierenden durchgeführt, die alle mit Musik zu tun haben (Musiker, Tontechniker). Alle Testpersonen hatten sechs Mal alle neun Klangbeispiele auf einer sog. Mean Opinion Score (MOS) Skala einzustellen. Diese reicht von 1 = schlecht bis 5 = exzellent. Die neun Klangbeispiele ergaben sich aus den drei getesteten Musik-Genres und den drei getesteten Algorithmen. Neben dem neu entwickelten Algorithmus wurden zwei andere professionelle DeNosing Verfahren getestet. Die Auswertung erfolgte so, dass von den sechst Durchläufen die ersten beiden als Training galten und nur die letzten vier zur Bewertung herangezogen wurden.

Abbildung 15 zeigt die gemittelten Ergebnisse über alle Testpersonen. Die Ergebnisse zeigen eine klare Bevorzugung des neuen Algorithmus für alle drei Genres.

¹Alle verwendeten Hörtest sind Flash-basiert und an der Jade Hochschule implementiert worden und bilden ein weiteres Ergebnis dieses Projektes.

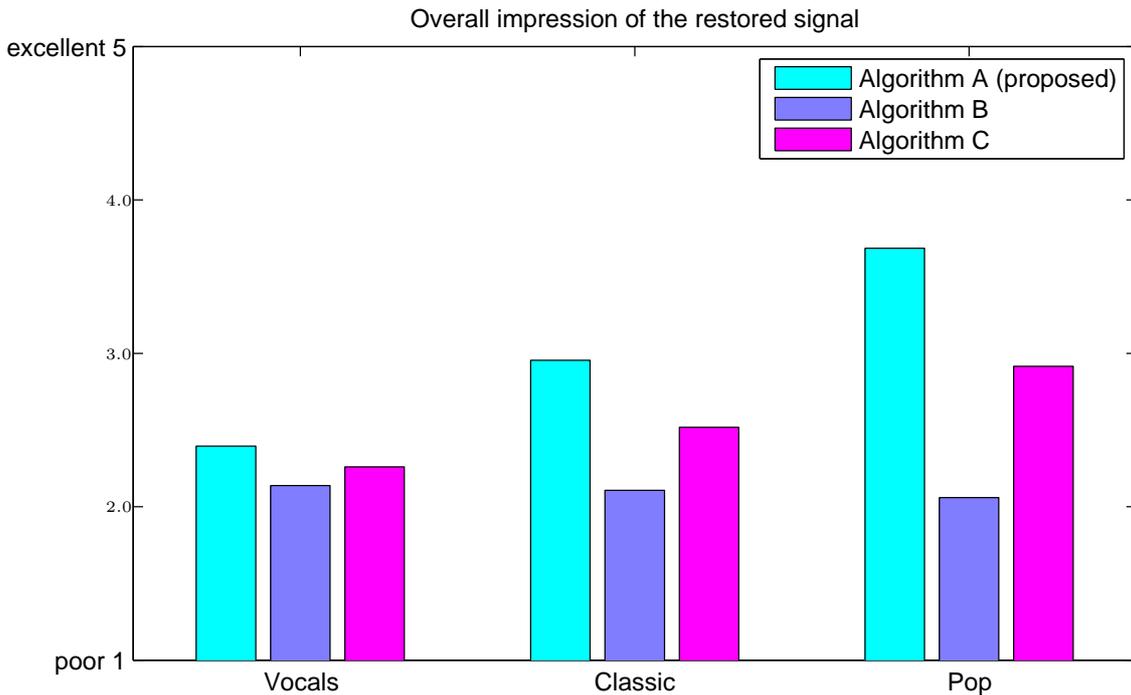


Bild 15: Subjektive Bewertung für verschiedene Algorithmen und Musikrichtungen auf einer MOS-Skala.

Determinierte Störungen

Das zu entwickelnde System zur vollautomatischen Reduktion von Brummstörungen muss aus zwei Komponenten bestehen:

1. einem Detektionsverfahren und
2. einem Verfahren zur Reduktion der zuvor detektierten Störungen.

Es wurde festgestellt, dass das kooperierende KMU zum zweiten Punkt bereits über Lösungen verfügte, die im Rahmen dieses Projektes genutzt werden konnten. Die Firma Cube-Tec International hat ein Verfahren im Produktsortiment, das die hochqualitative manuelle Entfernung von Brummstörungen erlaubt. Aus verschiedenen Gründen wurde daher entschieden, auf die Entwicklung eines weiteren Algorithmus zur Entfernung von Brummstörungen zu verzichten, sondern den Schwerpunkt auf die Entwicklung eines – bislang auch in Konkurrenzprodukten nicht vorhandenen – Brummdetektors zu legen.

Grundlagen

Grundlage zur Detektion von Brummstörungen in Audiosignalen ist die Feststellung, dass die Stationarität von im Signal vorhandenen Brummtönen sich stark von derjenigen eines typischen Nutzsignals unterscheidet. Die grundlegenden Eigenschaften (zum Beispiel Frequenz und Leistung) von Brummstörungen ändern sich üblicherweise nur wenig über einen längeren Zeitabschnitt hinweg – die Fluktuation der Leistungsverteilung eines Nutzsignals ist in der Regel deutlich stärker (siehe Bilder 16 und 17).

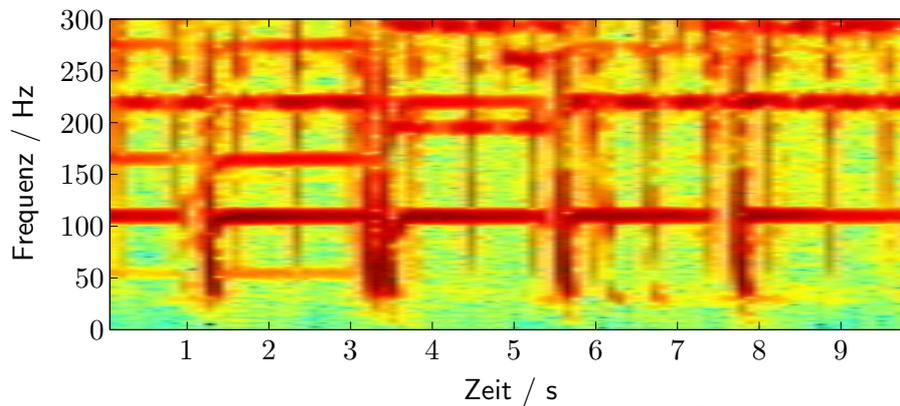


Bild 16: Spektrogramm einer Musikaufnahme ohne Brummstörung. Beim dargestellten Signal handelt es sich um einen Ausschnitt aus „St. Teresa“ von Joan Osborne.

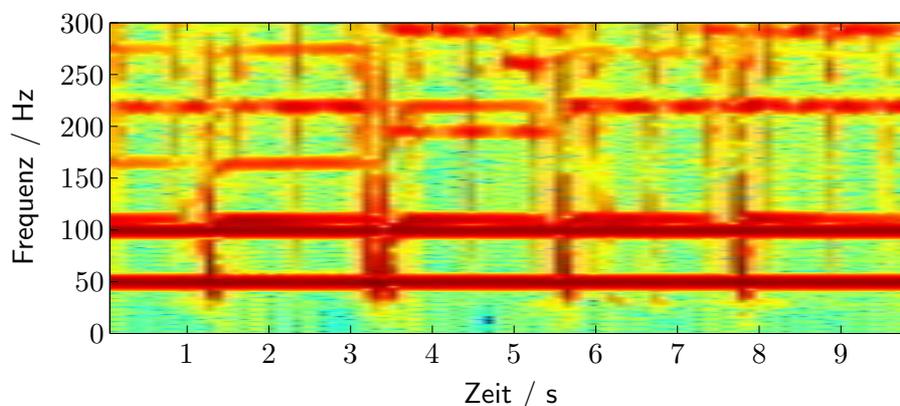


Bild 17: Spektrogramm einer Musikaufnahme mit Brummstörung. Beim dargestellten Signal handelt es sich um einen Ausschnitt aus „St. Teresa“ von Joan Osborne, die Brummstörung wurde künstlich hinzugefügt und besteht aus zwei Sinustönen mit $f_1 = 49,8\text{Hz}$ und $f_2 = 99,7\text{Hz}$ – das Signal-Rausch-Verhältnis beträgt $\text{SNR} = -30\text{dB}$.

Neben dem Hervortreten von Brummstörungen als horizontal verlaufende Linien im Spektrogramm in Bild 17 ist im Bild 16 zu erkennen, dass es passieren kann, dass auch bei Frequenzen, die keine Brummstörung enthalten, über einen längeren Zeitabschnitt hinweg Leistung vorhanden ist. Deutlich wird dies zum Beispiel bei einer Frequenz von ca. 110Hz in Bild 16.

Lösungsstrategien

Das während des Projektes entwickelte Verfahren zur Brummdetektion untersucht das Verhältnis des 10% Perzentil zum 55% Perzentil der Kurzzeit-Leistungsdichtespektren des Eingangssignals innerhalb eines bestimmten Zeitfensters. Dies ist dadurch zu begründen, dass die auf eine bestimmte Frequenz bezogene Leistung des Summensignals

$$\text{Nutzsignal} + \text{Brummstörung} = \text{gestörtes Summensignal}$$

prinzipiell immer wieder auf die Leistung der reinen Brummstörung absinkt. Dies passiert immer dann, wenn das Nutzsignal zu einem Zeitpunkt keine Leistung bei eben dieser Frequenz aufweist. Die Betrachtung der 10% Perzentile – und nicht der reinen Minimalwerte innerhalb eines Zeitfensters – erhöht die Robustheit gegenüber Schwankungen der Brummeistung. Grund für die Betrachtung des Verhältnisses der beiden

Perzentilwerte ist die Feststellung, dass die Leistung für Frequenzen, die eine Brummstörung enthalten, vergleichsweise wenig schwankt verglichen mit Frequenzen, in denen keine Brummstörung vorliegt. Für eben diese Schwankungen ist das Quantilverhältnis als geeignetes und robustes Maß ermittelt worden.

Es hat sich während der Bearbeitung dieser Aufgabe herausgestellt, dass die entscheidenden Punkte bei der Entwicklung des Detektionsverfahrens zum einen dessen Robustheit gegenüber für längere Zeit im Nutzsignal vorhandenen Tönen (siehe zum Beispiel Bild 16 bei $\approx 110\text{Hz}$) ist und zum anderen die Fähigkeit, auch Brummfrequenzen mit sehr geringer Leistung zu detektieren. Informell durchgeführte Hörversuche haben bestätigt, dass das menschliche Ohr in der Lage ist, Brummstörungen mit vergleichsweise geringer Leistung (\rightarrow großes SNR) wahrzunehmen. Die Herausforderung war es daher, einen Kompromiss zwischen diesen beiden Anforderungen, also

- Robustheit gegen Fehldetektionen und
- zuverlässige Detektion von Brummstörungen mit geringer Leistung

zu finden.

Die Analyse verschiedener durch das KMU zur Verfügung gestellter Testsignale aus der täglichen Praxis hat gezeigt, dass Brummstörungen trotz ihrer gegenüber dem Nutzsignal erhöhten Stationarität folgenden Schwankungen unterliegen:

- Amplitudenschwankungen
- (kurze) Aussetzer
- periodische Frequenzmodulation
- Frequenzdrift

Ein Detektor muss auf diese Schwankungen also robust reagieren. Die letzten beiden Punkte, also periodische Veränderungen der Frequenz (beispielsweise aufgrund von Gleichlaufschwankungen von Bandmaschinen) und Frequenzdrift (zum Beispiel verursacht durch den Spannungsabfall schwächer werdender Batterien mobiler Aufnahmegeräte) können durch Funktionen des bereits vorhandenen Verfahrens zur Brummreduktion kompensiert werden. Was bleibt ist, einen gegenüber den ersten beiden Punkten unempfindlichen Detektor zu entwickeln.

Es hat sich herausgestellt, dass ein Zwei-Schritt-Verfahren gute Ergebnisse ermöglicht. In einem ersten Schritt erfolgt eine statistische Analyse der Leistungsverteilung über der Frequenz. Frequenzen, an denen die Leistung während eines bestimmten Zeitfensters auf sehr niedrige Werte abfällt, sind unwahrscheinliche Kandidaten für Brummfrequenzen. Ist das Verhältnis der Perzentile bei einer Frequenz jedoch stets vergleichsweise hoch, so ist die Wahrscheinlichkeit, dass es sich um eine Brummfrequenz handelt, sehr viel größer. Der erste Schritt liefert also eine Aussage, bei welchen Frequenzen mit hoher Wahrscheinlichkeit eine Brummstörung vorliegt. Die prinzipielle Funktionsweise dieses ersten Detektionsschrittes ist in den Bildern 21 und 22 dargestellt.

In einem zweiten, nachgeschalteten Schritt werden die „Kandidatenfrequenzen“ in gewisser Weise „gesiebt“. Die Vorgabe verschiedener Anforderungen an Brummstörungen, nämlich

- Mindestdauer
- maximale Pausendauer
- maximale Frequenzdrift

trägt stark zur Erhöhung der Robustheit des Detektors bei, indem im ersten Schritt fälschlicherweise detektierte Frequenzen herausgefiltert werden.

Ein Prinzipschaltbild des Gesamtalgorithmus ist in Bild 18 dargestellt.

Evaluation

Für eine Vielzahl von aus der Praxis stammenden Testsignalen konnte die Effizienz des entwickelten Verfahrens bestätigt werden. Es hat sich gezeigt, dass für eine Vielzahl unterschiedlicher Eingangssignale eine robuste Detektion der benötigten Parameter ohne manuelles Justieren möglich ist.

Um die Detektionsleistung des Algorithmus unter klar definierten Bedingungen zu untersuchen und um die Ergebnisse umfassend auswerten zu können, sind umfangreiche Testläufe mit künstlich erzeugten Eingangssignalen vorgenommen worden. Damit war es möglich, das Verhalten der Methode in Abhängigkeit des Signal-Rausch-Abstands – in diesem Zusammenhang handelt es sich dabei um das Verhältnis von Nutzsignal- zu Brummleistung – zu untersuchen. Bei den verwendeten Testsignalen handelt es sich konkret um:

- weißes, Gaußsches Rauschen
- zufällige Ausschnitte aus einer Sprachaufnahme ([Wik08])
- zufällige Ausschnitte aus einer Klassikmusikaufnahme ([Van60])
- zufällige Ausschnitte aus einer Popmusikaufnahme ([Osb95])
- zufällige Ausschnitte elektronischer Musik
- zufällige Ausschnitte einer Feldaufnahme aus der Passagierkabine eines Flugzeugs

Bei den künstlich hinzugefügten Störungen handelt es sich um einzelne Sinustöne unterschiedlicher Frequenz (zufällig ausgewählt) zwischen 30Hz und 1kHz. Unter Verwendung von 11 verschiedenen SNR sind Nutzsignal und Störung additiv überlagert worden, resultierend in insgesamt 1980 Testsignalen. Jedes Signal hat eine Länge von zwei Minuten – die Brummstörung wurde nach 40s ein- und nach 80s ausgeschaltet. Um die Fehlalarmrate des entwickelten Verfahrens zu untersuchen, sind zusätzlich 1980 Signale ohne jegliche Brummstörung vom Algorithmus verarbeitet worden. In den Bildern 19 und 20 sind die Treffer- und Fehlalarmrate und die Genauigkeit der Brummfrequenzschätzung in Abhängigkeit des Eingangs-SNR dargestellt.

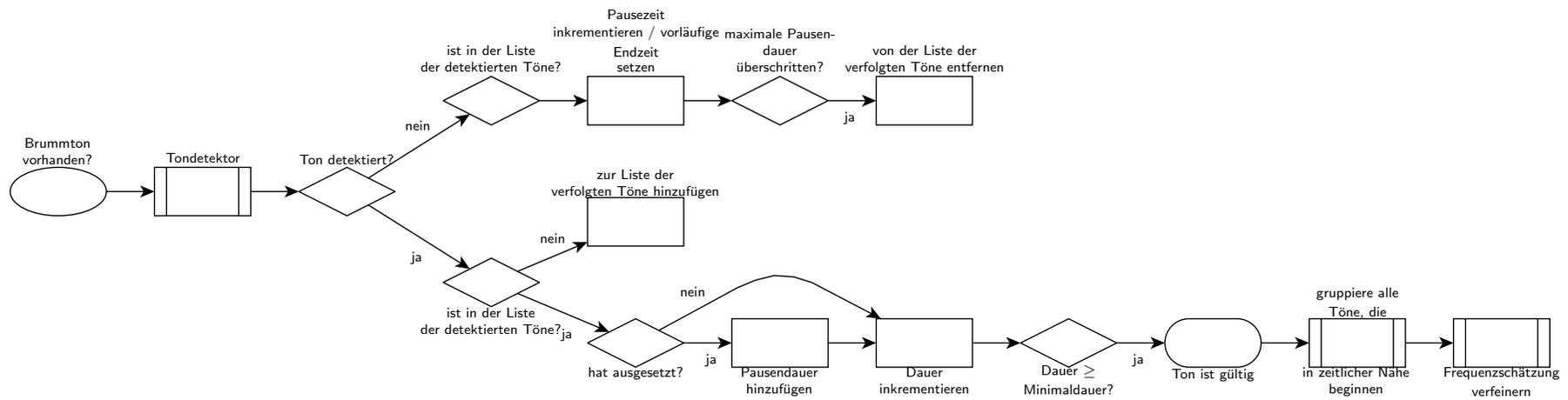


Bild 18: Blockschalbild des Algorithmus zur Detektion von Brummtörungen. Die implementierte Erweiterung befindet sich im Teil rechts vom Block „Tondetektor“.

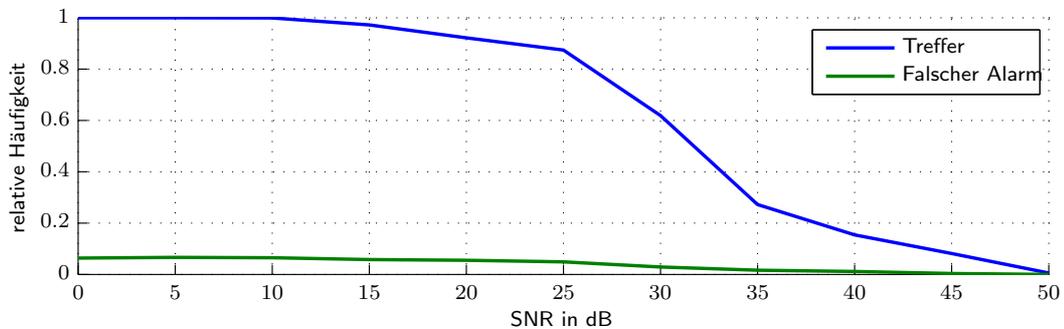


Bild 19: Treffer- und Falsch-Alarm-Rate in Abhängigkeit des Eingangs-SNR. Die Treffer-Rate ist das Verhältnis der Zeit der korrekt detektierten Brummstörungen zur Gesamtdauer der Brummstörungen. Die Falsch-Alarm-Rate ist in dieser Darstellung diejenige Zeit, in der fälschlicherweise Brummen detektiert wurde obwohl kein Brummen vorlag. Nicht enthalten in dieser Darstellung ist die Rate der Fehldetektionen bei Signalen, die keine Brummstörung enthalten haben – sie beträgt Null.

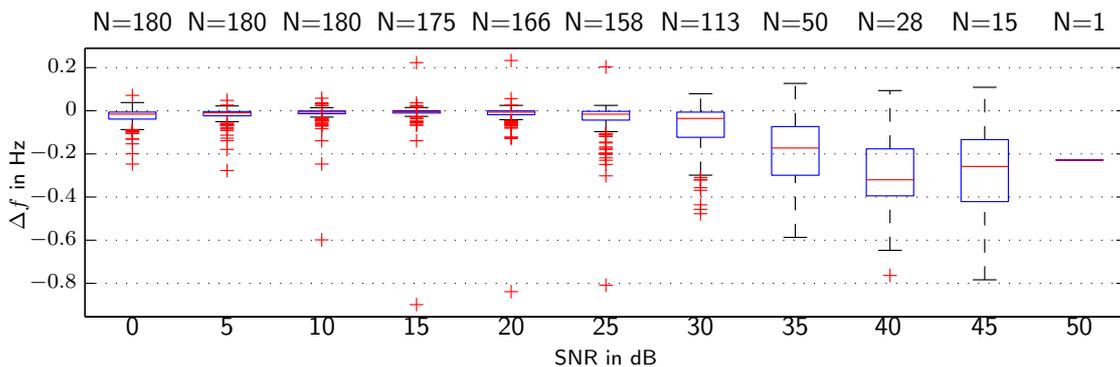


Bild 20: Fehler der Frequenzschätzung in Abhängigkeit des Eingangs-SNR. Bei dieser Box-Plot-Darstellung gibt die rote, horizontale Linie in der Mitte des blauen Kastens den Medianwert der Daten an. Die oberen und unteren Grenzen des Kastens sind die 25%- und 75%-Perzentile. Die gestrichelten, schwarzen Linien oberhalb und unterhalb des blauen Kastens reichen bis zu den extremsten Datenpunkten, die nicht als Ausreißer betrachtet werden. Die Ausreißer werden unabhängig, als rote Kreuze, dargestellt. (Zu Box Plots vergleiche auch [MTL78].) N ist jeweils die Anzahl der bei der Darstellung berücksichtigten Datenwerte.

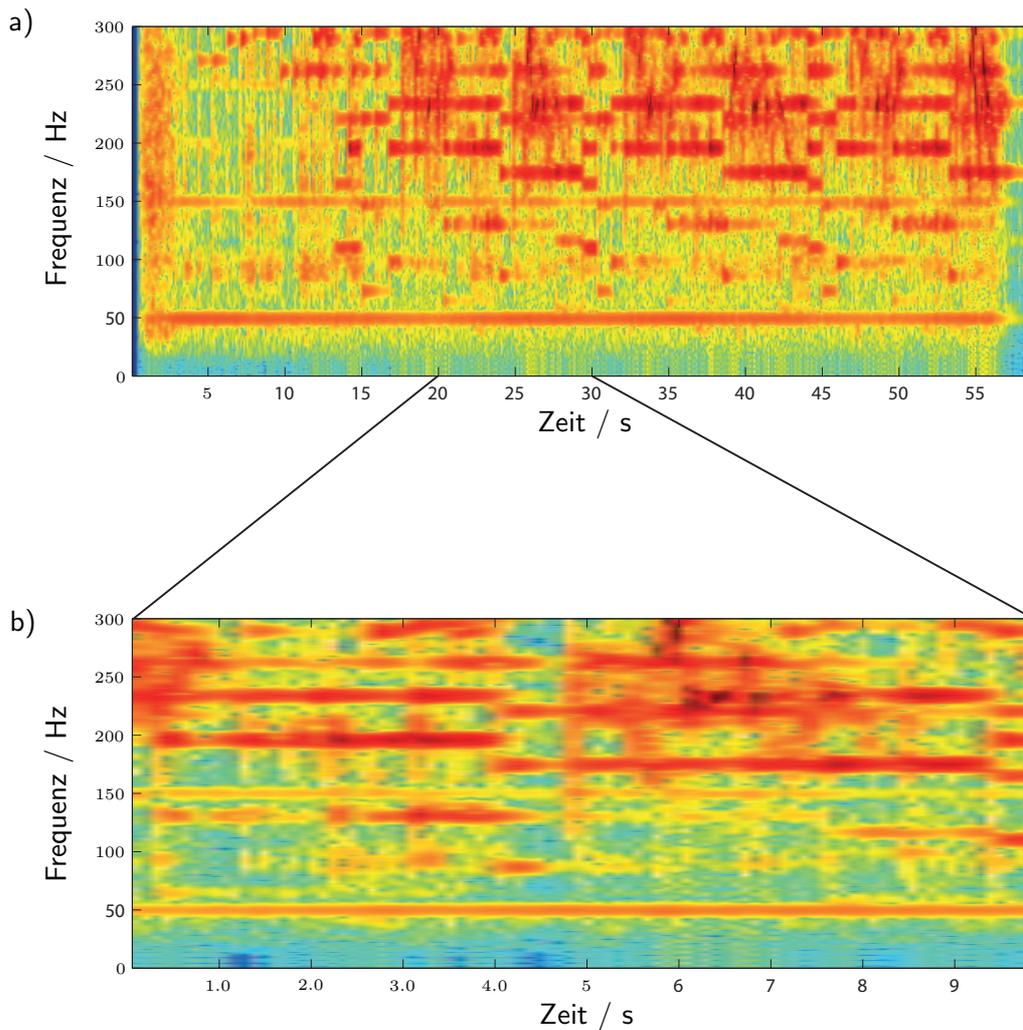


Bild 21: Zur Funktionsweise des Brummdetektors 1. Das Leistungsdichtespektrum des Eingangssignals (a) wird über einen bestimmten Zeitraum (hier: $T_{\text{Analyse}} = 10\text{s}$) beobachtet (b). In diesem Beispiel enthält es eine Brummstörung mit der Grundfrequenz $f_0 \approx 50\text{Hz}$ und deren zweiter Oberschwingung $f_2 \approx 150\text{Hz}$. Bei dem gezeigten Signal handelt es sich um eine Live-Aufnahme des Liedes „Sorry seems to be the hardest word“ von Elton John aus dem Jahre 1976.

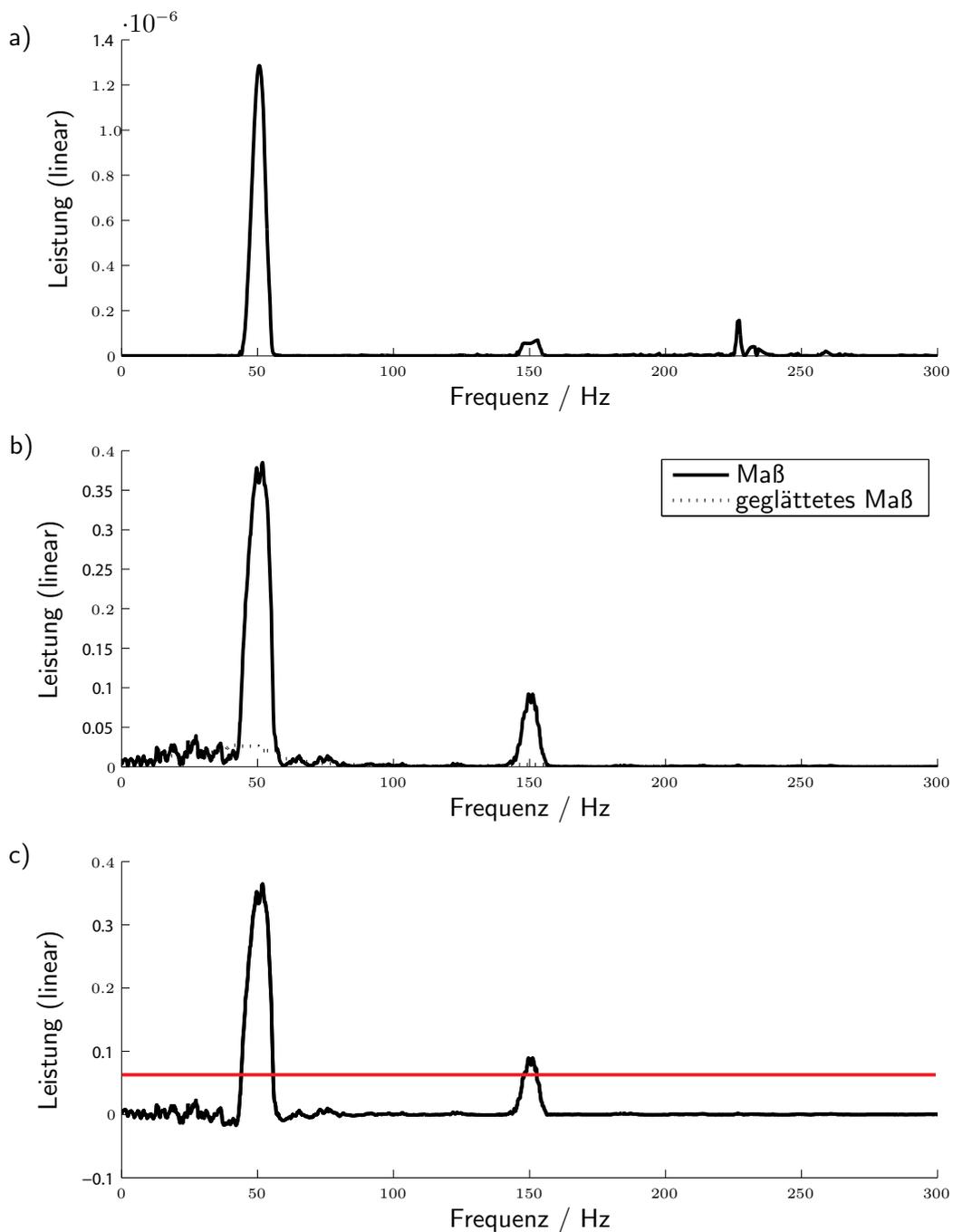


Bild 22: Zur Funktionsweise des Brummdetektors 2. Anstatt von den Minima des Leistungsdichtespektrums während eines Zeitfensters (vgl. Bild 21) – dargestellt in a) – auszugehen, wird ein robusteres Maß für die Wahrscheinlichkeit eines vorhandenen Brummtons berechnet (b). Das Maß ergibt sich aus dem Verhältnis vom 5% Perzentil zum 95% Perzentil. Der Vergleich von Teilbild a) und b) zeigt, dass auf diese Art und Weise Brummstörungen deutlicher zutage treten und die Robustheit gegenüber kurzen Einbrüchen der Störleistung erhöht wird. Indem dies Maß von breitbandig(er) vorhandener Leistung befreit wird (c) kann anschließend ein Schwellwert (rote Linie in Teilbild c) eingeführt werden, der, sobald er überschritten wird, den Verdacht auf eine vorhandene Brummfrequenz anzeigt.

Nutzen und Verwertung

Verwertung aus Sicht der Hochschule und der Hochschullehrer

Die Ergebnisse des Projektes fließen direkt in die Lehre ein, da in den Kursen Audiotechnik und DAFX (Digital Audio-Effects) auch Audio-Restoration als Thema behandelt wird. Weiterhin hat sich das Algorithmen Portfolio der Hochschulen stark erweitert, so dass bei neuen Fragestellungen schnellere Lösungen gefunden werden. Dies zeigt sich auch in der Kooperation mit der Fraunhofer Gesellschaft, wo mit dem neu geschaffenen Wissen sehr schnell spezielle Forensik-Lösungen angeboten werden konnten.

Insgesamt erhöhen die bisherigen und noch geplanten Veröffentlichungen die Sichtbarkeit der Jade Hochschule. Dies kann und sollte kurz- und mittelfristig zu neuen Projekten führen. Die Hoffnung ist weiterhin, dass im Nachgang noch einige hochwertige Zeitschriftenveröffentlichungen entstehen, so dass eine DFG Antragsfähigkeit mittelfristig (3-5 Jahre) für den Bereich Restoration / Geräuschreduktion erreicht werden kann.

Wirtschaftliche Verwertung

Das im Projekt neu geschaffene Wissen fließt direkt in die Produkte des Projektpartners Cube-Tec ein. Dieses führt dazu, dass Cube-Tec der Vision einer automatischen Aufbereitung großer Archive näher kommt, sobald eine Umsetzung der Ideen in die eigenen Frameworks und Architekturen erfolgt. Dies ist für einige Teilprobleme auch bereits geschehen und hat zu fertig vermarktbareren Produkten geführt.

Die wirtschaftlichen Verwertungsmöglichkeiten für die Hochschule sind eher begrenzt, da die Nutzergruppen dieser speziellen Techniken von wenigen Firmen abgedeckt werden und der benötigte Support durch eine Hochschule nicht leistbar ist. Das Wissen und die algorithmische Basis kann natürlich für Projekte in anderen Feldern genutzt werden, die keine direkte Konkurrenz zum Projektpartner darstellen. Hier fehlen im Moment aber konkrete Kontakte.

Gesellschaftlicher Nutzen

Der gesellschaftliche Nutzen der im Projekt entwickelten Ideen und Techniken ist nach einer Nutzbarmachung durch den Projektpartner immens, da viele Archive durch die neuen Technologien überhaupt erst zugänglich werden. Die Nutzung von Bibliotheken und Audio-Archiven wird langfristig einfacher.

Bezüglich des zweiten wichtigen Förderungsziels der Weiterbildung junger Menschen auf allen Ausbildungsebenen hat das Projekt natürlich auch einen gesellschaftliche Einfluss, da es einigen jungen Menschen

Gelegenheit gegeben hat, in einem faszinierenden Bereich ihre Ausbildung abzuschließen und anschließend einer adäquaten Beschäftigung nachzugehen, u.a. als neue Mitarbeiter beim Projektpartner.

Fortschritte auf dem Gebiet bei anderen Stellen

Das Gebiet der automatischen Restauration ist eher speziell und so sind nur wenige Fortschritte publik geworden. Inwieweit andere Firmen an Gesamtlösungen arbeiten, ist nicht bekannt.

Produkte

Die Firma Cedar-Audio aus England hat während der Laufzeit eine neue Reihe von Algorithmen zur voll-automatischen Restauration entwickelt. Die Lösung ist speziell für eine Zielplattform entwickelt und kann nicht für ganze Archive eingesetzt werden. Auffällig ist noch, dass für den DeNoising Prozess weiterhin mehrere Parameter eingestellt werden müssen und kein Vollautomat vorliegt.

Wissenschaftliche Veröffentlichungen

Im Bereich der wiss. Veröffentlichungen sind einige Arbeiten im Bereich der automatischen Sprachverarbeitung publiziert worden [GM09, HHJK09], die aber für einen Einsatz in der Musikproduktion nicht geeignet sind, da sie explizit das apriori Wissen nutzen, dass Sprache das Nutzsignal ist.

Für das Projekt relevant war außerdem die Dissertation von Garcia [Gar10], der sich mit der Rauschbenschätzung und damit auch mit einem Teilgebiet der automatischen Restauration beschäftigt hat. Die Ergebnisse sind in die aktuellen Arbeiten mit eingeflossen.

Patente

Als relevante Patente sind hinzugekommen das US Patent No: 7,787,975, "Restoring audio signals". Der folgende Abstract aus der US Patentdatenbank beschreibt ein Verfahren zur Impulsreduktion, das in dieser Form genau dem bekannten Stand der Technik entspricht und somit eigentlich nicht hätte patentiert werden dürfen.

"Methods, systems, and apparatus, including computer program products, for restoring audio signals. A data sequence of samples representing an audio signal is received. Multiple filter coefficients are defined for a filter, and a current sample in the data sequence is selected to be processed. The filter coefficients are updated based on a previous sample preceding the current sample in the data sequence and a filtered value determined by the filter for the previous sample. A filtered value for the current sample is determined using the filter with the updated filter coefficients. The filtered value of the current sample is used to determine whether the current sample has been corrupted by impulsive noise, for example, a crackle."

Das im Projekt entwickelte Verfahren arbeitet vollständig anders und ist nicht betroffen.

Im Verlauf des Projektes wurden keine Patente angemeldet. Dies ist auch damit zu begründen, dass viele Entwicklungen rein algorithmisch sind und somit schwer patentierbar.

Erfolgte und geplante Publikationen der Ergebnisse

Im Rahmen dieses Projektes sind vier Veröffentlichungen entstanden, deren Inhalt in diesem Abschnitt kurz dargestellt werden.

2009 – Konferenzartikel für die European Signal Processing Conference (EUSIPCO)

Der erste der veröffentlichten Artikel trägt den Titel „Optimal Spectral Smoothing in Short-Time Spectral Attenuation (STSA) Algorithms: Results of Objective Measures and Listening Tests“[BB09].

Es wird ein wichtiger Aspekt moderner Verfahren zur Reduktion breitbandiger Rauschstörungen untersucht: Die spektrale Glättung der Entrauschfilter-Übertragungsfunktion. Es ist bekannt, dass ein grundlegendes Problem derartiger Algorithmen eine bestimmte Fluktuation des Restrauschens ist, die vielfach als *Musical Noise* bezeichnet wird. Eine effiziente Methode, diese Artefakte zu beseitigen, ist, die Gewichtungsfunktion spektral zu glätten, um die Fluktuation benachbarter Spektralwerte zu reduzieren. Wichtige Parameter bei diesem Verfahren sind zum einen die Art der Spektralgättung (feste oder frequenzabhängige Bandbreite) und zum anderen die Bandbreite der Spektralgättung, die direkt den Grad der Glättung beeinflusst. In diesem Artikel werden diejenigen Parameter ermittelt, die zur höchsten subjektiv empfundenen Klangqualität des bearbeiteten Ausgangssignals führen.

2010 – Konferenzartikel für den International Workshop on Acoustic Echo and Noise Control (IWAENC)

Der im Jahr 2010 vorgestellte Artikel trägt den Titel „Detection of Hum in Audio Signals“[BB10].

In diesem Artikel wird ein neues Verfahren vorgestellt, um Brummstörungen in Audiosignalen zu detektieren. Es wird kurz erläutert, was Brummstörungen (in technischer Hinsicht) auszeichnet und wie sie entstehen. Ausgehend von diesen Überlegungen wird ein Signalmodell für das gestörte Signal aufgestellt. Es wird ein Verfahren beschrieben, das es ermöglicht, ohne jegliche a priori Information festzustellen, ob eine Tonaufnahme eine Brummstörung enthält oder nicht. Im Falle einer detektierten Brummstörung werden die Frequenzen der Störtöne, die Start- und Endzeiten und das Signal-Rausch-Verhältnis zuverlässig geschätzt. Durch eine Evaluation mit künstlich gestörten Signalen wird die Effizienz des beschriebenen Verfahrens belegt.

2011 – Konferenzartikel für die 130. AES Convention

Der Titel des im Jahre 2011 in London vorgestellten Artikels lautet: „Evaluation of a New Algorithm for Automatic Hum Detection in Audio Recordings“[BSB11].

Dieser Artikel hat die umfassende Evaluation des im Jahre 2010 veröffentlichten Verfahrens zur Detektion von Brummstörungen zum Gegenstand. Während die Testläufe in [BB10] ausschließlich mit künstlich erzeugten Signalen durchgeführt wurden, wird hier eine große Anzahl von Signalen aus der täglichen Praxis verwendet, um die Leistungsfähigkeit des Detektionsalgorithmus zu untersuchen.

2012 – Konferenzartikel für die 132. AES Convention

Der vorgestellte Artikel trägt den Titel „Hum Removal Filters: Overview and Analysis“[BB12].

Es werden verschiedene Methoden untersucht, um Brummstörungen aus Tonaufnahmen zu entfernen. Verschiedene Filterstrukturen bzw. -algorithmen bieten unterschiedliche Parameterkombinationen, um für verschiedene Störungen einen akzeptablen Kompromiss zwischen dem Grad der erreichten Reduktion der Störung und der Unversehrtheit des Nutzsignals zu erreichen. Zum Beispiel ist eine hohe Frequenzselektivität – also geringe Filterbandbreite – wünschenswert, um das Nutzsignal weitgehend unbeeinträchtigt zu lassen. Allerdings hat eine geringe Bandbreite – erklärt durch die Zeit/Bandbreit-Unschärferelation der Nachrichtentechnik – eine lange Impulsantwort zur Folge, was in hörbarem Ausschwingen des Filters resultieren kann. Daher ist die Wahl des optimalen Algorithmus ein Kompromiss aus hoher Frequenzselektivität und akzeptablem Zeitbereichsverhalten. In diesem Zusammenhang weisen unterschiedliche Filterstrukturen unterschiedliche Eigenschaften auf. Um deren Auswirkungen auf Brummstörung und Nutzsignal zu untersuchen haben wir drei verschiedene Algorithmen mittels objektiver Maße evaluiert und zeigen Vor- und Nachteile der einzelnen Ansätze auf.

2012 – Konferenzartikel für den International Workshop on Acoustic Echo and Noise Control (IWAENC) (accepted)

Der vorgestellte Artikel trägt den Titel „A New Approach For Reduction Of Supergaussian Noise using Autoregressive Interpolation and Time-Frequency Masking“[BB12].

In dem Artikel wird gezeigt, dass mit Hilfe von Interpolationsalgorithmen eine Geräuschreduktion erreicht werden kann. In einer analytischen Untersuchung zeigt sich, dass die erzielte Geräuschreduktion von der Kurtosis abhängt und größer ist, je super-gaußisch das Signal ist.

Weitere geplante Publikationen

Es sind noch weitere Zeitschriftenartikel geplant. Folgende Publikationen sind zur Zeit in Arbeit und werden noch dieses Jahr bei internationalen Fachzeitschriften eingereicht

- Automatic Removal of Hum in Audio Signals Part 1: Detection. Eine Langfassung der oberen Konferenzartikel mit weitergehenden Analysen.
- Automatic Removal of Hum in Audio Signals Part 2: Reduction. Ein Artikel, in dem die unterschiedlichen wissenschaftlichen und kommerziellen Verfahren zur Hum Reduktion evaluiert werden.
- Reduction of Gaussian, Supergaussian and Impulsive Noise by Interpolation of the Binary Mask Residual: Langfassung des IWAENC 2012 papers mit dem deutlichen Schwerpunkt auf Audio Restoration.

Mit welchen Partnern hat eine Kooperation stattgefunden?

Im Verlaufe des Projektes wurde mit beiden im Antrag genannten Partnern intensiv zusammengearbeitet. Mit Cube-Tec International gab es regelmäßige Treffen (mindestens alle 5 Wochen) mit allen Projektmitarbeitern im weiteren Sinne (Studenten während der Abschlussarbeiten, Firmenmitarbeiter (3),

Geschäftsführer, Projektleiter und wiss. Mitarbeiter). Zusätzlich gab es einen intensiven eMail-Kontakt zwischen den beiden wesentlichen Entwicklern und Evaluatoren auf Seiten der Firma und der Jade-Hochschule.

Der wiss. Mitarbeiter war außerdem in der Arbeitsgruppe von Prof. Dr. Simon Doclo an der Universität Oldenburg eingebunden. Der Übergang vom im Antrag genannten Prof. Dr. Dr. Kollmeier wurde wegen der größeren inhaltlichen Nähe zur Signalverarbeitung nach der Neuberufung von Prof. Doclo von allen beteiligten als bessere Lösung bevorzugt. In der Arbeitsgruppe musste der wiss. Mitarbeiter regelmäßig über seinen wiss. Fortschritt berichten, so dass eine fundierte Beratung und Unterstützung durch den zweiten Betreuer an der Universität erfolgen konnte.

Wie viele Drittmittel wurden eingeworben?

In diesem Projekt wurden neben dem vorher vertraglich zugesicherten Anteil des KMU-Partners keine weiteren Drittmittel eingeworben.

Verbesserung der Drittmittelfähigkeit

Durch das regelmäßige Auftreten mit neuen Forschungsergebnissen auf Messen und Konferenzen (u.a. finanziert durch AURUM) ist die Sichtbarkeit weiter erhöht worden. Dies führte mit dazu, dass der Antragssteller nun als wissenschaftlicher Berater beim Fraunhofer Institut für digitale Medientechnologien, Projektgruppe Hör-, Sprach- und Audiotheologie von der Jade Hochschule entsandt wird und dies zu regelmäßigen Drittmitteleinnahmen führt (8.000 Euro per Anno). Zusätzlich konnte ein Drittmittelprojekt im Bereich Forensik durch einen weiteren Konferenzkontakt eingeworben werden (ca. 50.000 Euro).

Anregungen für das Programm Forschung an Fachhochschulen

Gerade für die Förderlinie IngenieurNachwuchs wäre eine Förderung des Abschlussjahres zur Beendigung einer Promotion sehr hilfreich, da durch die Nähe zu den Firmen häufig zu wenig Zeit für grundlegende theoretische Überlegungen bleibt. Zusätzlich ist es in technischen Fächern ausgesprochen selten, eine Promotion in 3 Jahren abschließen zu können. Üblicherweise werden 4-5 Jahre benötigt. Eine spezielle Abschluss-Förderung unter bestimmten Randbedingungen z.B. es wurde bereits peer-reviewed veröffentlicht, beantragt werden kann, könnte hier Abhilfe schaffen.

Insgesamt zeigen die benötigten Punktzahlen bei den Anträgen, dass insgesamt zu wenig Geld für die Forschung an Fachhochschulen zur Verfügung steht.

Literaturverzeichnis

- [BB09] Matthias Brandt und Joerg Bitzer. Optimal Spectral Smoothing in Short-Time Spectral Attenuation (STSA) Algorithms: Results of Objective Measures and Listening Tests. *European Signal Processing Conference (EUSIPCO)*, 2009.
- [BB10] Matthias Brandt und Joerg Bitzer. Detection of Hum in Audio Signals. *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2010.
- [BB12] Matthias Brandt und Joerg Bitzer. Hum removal filters: Overview and analysis. In *Proceedings of the 132nd AES Convention*, 2012.
- [BSB11] Matthias Brandt, Thorsten Schmidt und Joerg Bitzer. Evaluation of a new algorithm for automatic hum detection in audio recordings. In *Audio Engineering Society Convention 130*, May 2011.
- [BSM79] M. Berouti, R. Schwartz und J. Makhoul. Enhancement of speech corrupted by acoustic noise. In *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, pages 208–211, 1979.
- [BT52] Ralph A. Bradley und Milton E. Terry. Rank Analysis of Incomplete Block Designs - Part 1: The Method of Paired Comparisons. *Biometrika*, 39(3-4):324–345, 1952.
- [Cap94] O. Cappe. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE Transactions on Speech and Audio Processing*, 2(4):345–349, April 1994.
- [CB02] Israel Cohen und Baruch Berdugo. Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement. *IEEE Signal Processing Letters*, 9(1):12–15, 2002.
- [CL95] O. Cappe und J. Laroche. Evaluation of short-time spectral attenuation techniques for the restoration of musical recordings. *IEEE Transactions on Speech and Audio Processing*, 3(1):84–93, January 1995.
- [CN06] A. Cohen und I. Neoran. Pyramidal Algorithm for the Restoration of Audio Signal Corrupted by Wideband Noise. In *Proceedings of the 120th Convention of the AES*, Paris, 2006. preprint 6771.
- [Coh03] I. Cohen. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. *IEEE Transactions on Speech and Audio Processing*, 11(5):466–475, September 2003.
- [Cro80] R. E. Crochiere. A Weighted Overlap-Add Methode of Short-Time Fourier Analysis/Synthesis. *IEEE Transactions on Acoustics, Speech, and Signal processing*, 28(1):99–102, feb 1980.
- [Czy97] A. Czyzewski. Learning Algorithms for Audio Signal Enhancement: Part 1 Neural Network Implementation for the Removal of Impulse Distortions. *Journal of the Audio Engineering Society*, 45(10):815–831, 1997.

- [Dob95] G. Doblinger. Computationally efficient speech enhancement by spectral minima tracking in subbands. In *Proceedings of EUROSPEECH*, volume 2, pages 1513–1516, 1995.
- [EB06] P.A.A. Esquef und L.W.P. Biscainho. An efficient model-based multirate method for reconstruction of audio signals across long gaps. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4):1391–1400, jul 2006.
- [EM84] Yariv Ephraim und David Malah. Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(6):1109–1121, 1984.
- [EM85] Y. Ephraim und D. Malah. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(4):443–445, apr 1985.
- [Gar10] G. Garcia. *Automatic denoising for musical audio restoration*. PhD thesis, Stanford University, 2010.
- [GBW98] S. Gannot, D. Burshtein und E. Weinstein. Iterative and sequential Kalman filter-based speech enhancement algorithms. *IEEE Transactions on Speech and Audio Processing*, 6:373–385, jul 1998.
- [GEH98] T. Gülzow, A. Engelsberg und U. Heute. Comparison of a Discrete Wavelet Transformation and a Nonuniform Polyphase Filterbank Applied to Spectral-Subtraction Speech Enhancement. *Elsevier Signal Processing*, 6(1):5–19, jan 1998.
- [GM09] T. Gerkmann und R. Martin. On the statistics of spectral amplitudes after variance reduction by temporal cepstrum smoothing and cepstral nulling. *Signal Processing, IEEE Transactions on*, 57(11):4165–4174, 2009.
- [God93] S. J. Godsill. *The Restoration of Degraded Audio Signal*. PhD thesis, Cambridge University, Dept. Eng., dec 1993.
- [GR98] S. J. Godsill und P. J. W. Rayner. *Digital Audio Restoration – A Statistical Model Based Approach*. Springer-Verlag, London, U.K., 1998.
- [HHJK09] R.C. Hendriks, R. Heusdens, J. Jensen und U. Kjems. Low complexity dft-domain noise psd tracking using high-resolution periodograms. *Eurasip Journal on Advances in Signal Processing*, 2009:55, 2009.
- [KB98] Mark Kahrs und Karlheinz Brandenburg, editors. *Applications of Digital Signal Processing to Audio and Acoustics*, chapter S. Godsill, P. Rayner, O. Cappe: Digital audio restoration. Kluwer Academic Publishers, Massachusetts, U.S.A., 1998.
- [Luc59] R. Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.
- [LV03] T. Lotter und P. Vary. Noise reduction by maximum a posteriori spectral amplitude estimation with supergaussian speech modeling. In *Proceedings of the International Workshop on Acoustic Echo and Noise Control*, pages 83–86, 2003.
- [Mar94] R. Martin. Spectral subtraction based on minimum statistics. In *Proceedings of the European Signal Processing Conference*, pages 1182–1185, 1994.
- [Mar01] Rainer Martin. Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics. *IEEE Transactions on Speech and Audio Processing*, 9(5):504–512, 2001.

- [Mar05] R. Martin. Speech Enhancement Based on Minimum Mean-Square Error Estimation and Supergaussian Priors. *IEEE Transactions on Speech and Audio Processing*, 13(5):845–856, sep 2005.
- [MP00] U. Mittal und N. Phamdo. Signal/Noise KLT Based Approach for Enhancing Speech Degraded by Colored Noise. *IEEE Transaction on Speech and Audio Processing*, 8(2):159–167, mar 2000.
- [MTL78] Robert McGill, John. W. Tukey und Wayne. A. Larsen. Variations of box plots. *American Statistician*, pages 12–16, 1978.
- [NHK04] Naoki Nitanda, Miki Haseyama und Hideo Kitajima. Audio-Cut Detection and Audio-Segment Classification Using Fuzzy C-Means Clustering. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 325–328, 2004.
- [Osb95] Joan Osborne. Relish. *Blue Gorilla Records*, 1995.
- [Pud03] H. Puder. *Geräuschreduktionsverfahren mit modellbasierten Ansätzen für Freisprecheinrichtungen in Kraftfahrzeugen*. Number 721. Darmstädter Dissertation D17, Fortschritt-Berichte VDI, 2003.
- [Reg91] P. A. Regalia. An improved lattice-based adaptive IIR notch filter. *IEEE Transactions on Signal Processing*, 39:2124–2128, sep 1991.
- [Rol12] Christian Rollwage. *Blinde Bandbreitenerweiterung mit Hilfe sinusoider Modellierung*. *Masterarbeit Universität Oldenburg*, 2012.
- [SFB00] V. Stahl, A. Fischer und R. Bippus. Quantile Based Noise Estimation for Spectral Subtraction and Wiener Filtering. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 1875–1878, 2000.
- [SVD93] L. Simon, J.C. Valiere und C. Depollier. New Contribution on Noise Reduction Using Wavelet Techniques: Application to the Restoration of Old Recordings Techniques. In *Proceedings of the 94th Convention of the AES*, Berlin, 1993. preprint 3461.
- [Van60] Johann Baptist Vanhal. *Symphony in c minor – Menuetto Moderato*. 1760.
- [Vas96] Saeed V. Vaseghi. *Advanced Signal Processing and Digital Noise Reduction*. John Wiley & Sons Ltd., Chicester, England und B.G. Teubner, Stuttgart, Deutschland, 1996.
- [Wie49] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. Wiley, Ney York, 1949.
- [Wik08] Wikipedia spoken article. Bird. 2008.