Vertrauenswürdigkeit von Audiodaten – Digitale Wasserzeichen und Verifikation der semantischen Integrität

Sascha Zmudzinski, Martin Steinebach

Fraunhofer Institut für
Integrierte Publikations- und Informationssysteme (IPSI)
Dolivostr. 15
64293 Darmstadt
sascha.zmudzinski@ipsi.fraunhofer.de
martin.steinebach@ipsi.fraunhofer.de

Zusammenfassung: Existierende Systeme und Protokolle zum Schutz der Echtheit und Unverfälschtheit von Audio-Aufzeichnungen ermöglichen es nicht, erlaubte Signalveränderungen von verbotenen Manipulationen zu unterscheiden oder den Ort und die Art der Manipulation genauer zu charakterisieren. Wir stellen experimentelle Ergebnisse zum Schutz der semantischen Integrität von Audio-Aufzeichnungen vor. Wir untersuchen dazu einige spezielle Audio-Merkmale auf ihre Eignung, einen Nachweis für die semantische Integrität einer Audio-Aufzeichnung im Kontext von digitalen Wasserzeichen zu liefern.

1 Motivation und Problemstellung

In zunehmenden Maße werden Archivdaten, Aufzeichnungen von Ereignissen durch Überwachungs-Kameras, Nachrichten-Reportagen oder polizeiliche Vernehmungen in digitaler Form produziert oder ausschließlich auf digitalen Medien aufgezeichnet und bereitgestellt. Moderne Computer-Hardware und -Software ermöglichen einen einfachen Zugriff auf diese digitalen Medien, deren Nachbearbeitung, (Ver-) Fälschung und Verbreitung. Ein Beispiel: Sehr leicht kann bei einer digitalen Sprachaufzeichnung durch Schneiden des Materials oder Ändern der Tonhöhe der Satz „*Ich bin unschuldig!“* oder die Frage „*Ich bin schuldig?*“ manipuliert werden. Schnell entsteht dabei eine veränderte Bedeutung, beispielsweise: „*Ich bin schuldig!*“. Daher kommt der Frage der Verifikation von Echtheit und Unverfälschtheit digitaler Medien eine zunehmende Bedeutung zu.

Hierzu existieren verschiedene Verfahren und Protokolle, die auf Verschlüsselung und unterstützenden Sicherheitsmechanismen basieren, beispielsweise digitale Signaturen und kryptographische Hash-Funktionen etc. Eine Unterscheidung zwischen Veränderungen, die bloß die digitale Darstellung der Daten verändern (z.B. eine Formatwandlung einer Audio-Aufzeichnung nach mp3) und unerlaubten Manipulationen, die die semantische Integrität verändern (z.B. Löschen von Passagen, Hinzufügen von Geräuschen oder Stimmen), ist hierbei nicht möglich, ebenso wenig die

Lokalisierung solcher unerlaubten Manipulationen. Das fragliche Medium selbst trägt hierbei keinerlei Informationen, die eine Verifikation der Integrität erlauben. Der Nachweis macht also ein Sicherheits-Protokoll und eine funktionierende Infrastruktur notwendig, die die notwendigen Authentifizierungsmechanismen o.ä. bereitstellt und erreichbar macht.

Einen vielversprechenden Ansatz für diese Herausforderungen stellen digitale Wasserzeichen dar, wenn sie mit der gezielten Extraktion von Merkmalen kombiniert werden, welche sehr stark mit dem semantischen Inhalt der geschützten Medien verknüpft sind. Wir geben dazu in Kapitel 2 einen kurzen Überblick über die Technologie digitaler Wasserzeichen. In Kapitel 3 stellen wir die von uns untersuchten inhaltsabhängigen Audio-Merkmale vor. In Kapitel 4 untersuchen wir deren Eignung zum Inhaltsschutz und geben in Kapitel 5 eine kurze Zusammenfassung und einen Ausblick.

2 Inhaltsfragile Audio-Wasserzeichen

Der Schutz des Inhalts digitaler Audiodaten bleibt bei der Erforschung digitaler Wasserzeichen im Vergleich zum Schutz der Urheberrechte relativ wenig betrachtet. Zweierlei grundsätzliche Ansätze sind hier bekannt: *Inhaltsfragile Wasserzeichen* (siehe z.B [St96] [Di00], [Ca02]) sind zum robusten Einbetten von Inhaltsinformationen und zum späteren Erkennen von Unterschieden zwischen eingebetteten und vorliegenden Inhalten konzipiert. *Invertierbare Wasserzeichen* (siehe z.B. [Fr01]) hingegen bieten ein hohes Sicherheitsniveau durch den Einsatz kryptographischer Algorithmen, überstehen aber keinerlei Manipulationen.

Inhaltsfragile Wasserzeichen dienen ausschließlich dem Identifizieren der Inhaltsänderungen bzw. dem Schutz der semantischen Integrität. Veränderungen, die nur die binäre Repräsentation der Informationen betreffen (z.B. Formatwandlung ohne extrem verlustbehaftete Kompression) oder die nur minimale, nicht wahrnehmbare Variationen der Inhalte erzeugen (z.B. Wandlung der Abtastrate von 48 kHz auf 44,1 kHz), werden bewusst ignoriert.. Wir gehen sogar davon aus, dass eine Verwendung digitaler Wasserzeichen zum Inhaltsschutz in vielen Fällen nur dann von Vorteil gegenüber kryptographischen Ansätzen sind, wenn die Inhaltsinformationen durch den gesamten inhaltsbelassenden Verarbeitungsprozess in den Audiodaten verbleiben, ohne dabei gelöscht zu werden oder fälschlicher Weise eine Inhaltsänderung anzuzeigen.

2.1 Digitale Wasserzeichen

Generell verstehen wir unter einem digitalen Wasserzeichen ein transparentes, nicht wahrnehmbares Muster, welches in das Datenmaterial (Bild, Video, Audio, 3D-Modelle) mit einem Einbettungsalgorithmus unter Verwendung eines geheimen Schlüssels eingebracht wird [Di00]. Jeder Wasserzeichenalgorithmus besteht aus einem Einbettungsprozess und einem Abfrage- oder Ausleseprozess. Das eingebettete Muster

repräsentiert die eingebrachte Information. Als Wasserzeichen können beispielsweise Informationen in das Material eingebracht werden, die erlauben festzustellen, ob das Datenmaterial manipuliert worden ist oder ob bestimmte Zusatzinformationen zum Datenmaterial korrekt sind.

2.2 Selbst-verifizierende Medien – Inhaltsfragile digitale Wasserzeichen

Ein Ansatz zum Schutz der Audiodaten ist, aus dem zu schützenden Medium inhaltsrelevante Audio-Merkmale abzuleiten und verbotene Manipulationen über Veränderung dieser Merkmale zu detektieren [Go03]. Die Vorgehensweise zum Nachweis der Integrität vollzieht sich dabei in folgenden Arbeitsschritten:

Merkmalsextraktion und Einbettung: Die zu schützende Audiodatei wird in kurze Abschnitte unterteilt. Jeder dieser Abschnitte wird auf Signal-Merkmale untersucht, die besonders eng mit der empfundenen akustischen Wahrnehmung und somit mit der inhaltlichen Bedeutung der Audio-Aufzeichnung verknüpft sind (*Fingerprint,* digitaler Fingerabdruck). Der extrahierte digitale Fingerabdruck wird anschließend vermittels eines *robusten* Wasserzeichenverfahrens in die Audiodatei eingebettet (siehe Abbildung 1). Er übersteht daher bspw. Format-Konvertierungen oder DA/AD-Wandlung [St03].

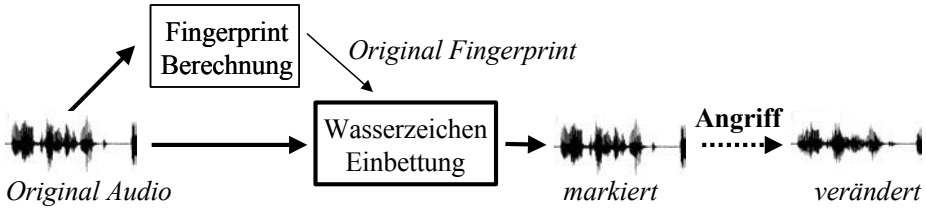

Abbildung 1: Extraktion des digitalen Fingerabdrucks aus der Originaldatei und dessen anschließendes Einbetten in das Audiosignal, nach [Ca02]

Verifikation: Um zu einem späteren Zeitpunkt die Integrität zu verifizieren, wird der eingebettete Original-Fingerabdruck ausgelesen und mit dem *aktuellen* Wert des Fingerabdrucks verglichen (siehe Abbildung 2). Bei einer Übereinstimmung wird die Datei als unverfälscht akzeptiert.

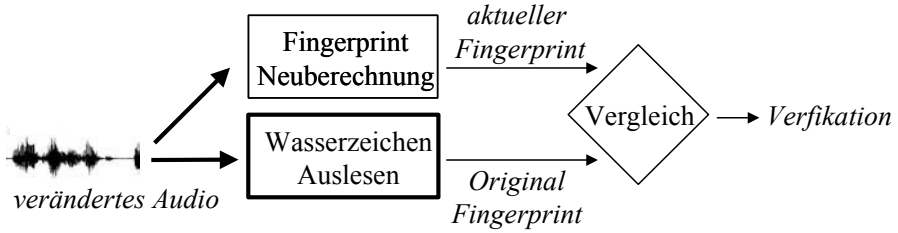

Abbildung 2: Auslesen des eingebetteten Original-Fingerabdrucks, Vergleich zwischen eingebettetem und aktuellem Fingerabdruck zur Verifikation

Für die Auswahl geeigneter Audiomerkmale gelten nach [Ca02] folgende Anforderungen:

- **Trennschärfe**: Dateien, die sich in ihrer empfundenen akustischen Wahrnehmung tatsächlich unterscheiden, sollen stets unterschiedliche digitale Fingerabdrücke besitzen
- **Hohe Kompaktheit**: der digitale Fingerabdruck soll eine möglichst kompakte/ komprimierte Darstellung der Klangeigenschaften bieten, da robuste Wasserzeichen-Verfahren nur eine geringe Datenrate bieten.
- **Robustheit**: der digitale Fingerabdruck soll unverändert bleiben durch erlaubte Signalveränderungen, die den semantischen Inhalt nicht verändern (z.B. moderate verlustbehaftete Kompression, schwaches additives Rauschen durch DA/AD-Wandlung, Dynamik-Kompression). Er darf sich insbesondere durch die Wasserzeichen-Markierung nicht ändern, da bei der anschließenden Verifikation sonst falsche Alarme detektiert würden
- **Sicherheit**: der digitale Fingerabdruck soll Angriffe überstehen, die gezielt auf den Algorithmus gerichtet sind, jede unerlaubte Manipulation der zu schützenden Datei soll zu einer Änderung des digitalen Fingerabdrucks führen.
- **Rechenzeit**: Insbesondere für Echtzeit-Anwendungen (z.B. den Schutz einer Live-Übertragung) ist die Rechenzeit ein wichtiger Faktor.

3 Extraktion inhaltsabhängiger Audiomerkmale

Ausgangspunkt für die robuste Detektion von Manipulationen sind die ausgewählten Audio-Merkmale und hieraus berechneten digitalen Fingerabdrücke. In diesem Kapitel werden wir einige ausgewählte Merkmale genauer vorstellen. Wir nehmen hierbei das in Kapitel 1 motivierte Beispiel einer Sprachaufzeichnung zum Anlass, unerlaubte Veränderungen wie beispielsweise Änderung der Intonation oder Hinzufügen/ Löschen von einzelnen Passagen zu detektieren.

3.1 Grundfrequenz-Bestimmung/ Tonhöhenerkennung

Ein wesentliches Merkmal für den empfundenen Klangeindruck und die inhaltliche Bedeutung einer Audio-Aufzeichnung, insbesondere zum Schutz der Intonation von Sprachaufzeichnungen, ist die Grundfrequenz f_0. und die *empfundene* Tonhöhe (engl. *Pitch*). Im Folgenden werden wir nur die Bestimmung der Grundfrequenz untersuchen, da bei den Schallpegeln und Grundfrequenzen der betrachteten Audiodateien die und die beide Begriffe synonym verwendet werden können [ZF90]. Im Folgenden betrachten wir einige Verfahren genauer:

Harmonisches Produkt-Spektrum (HPS): Dieses Verfahren nach [Qu02] setzt im Fourier-Spektrum der Audiodaten an und hat zur Voraussetzung, dass zusätzlich zur Energie auf der Grundfrequenz auch dominante Frequenzanteile bei den ersten Harmonischen (Oberwellen) vorhanden sind: Im Spektrum seien also äquidistante lokale Maxima (Peaks) beobachtbar. Dies ist bei Klängen von Instrumenten oder stimmhaften Lauten in Sprachaufzeichnungen oftmals gegeben**.**

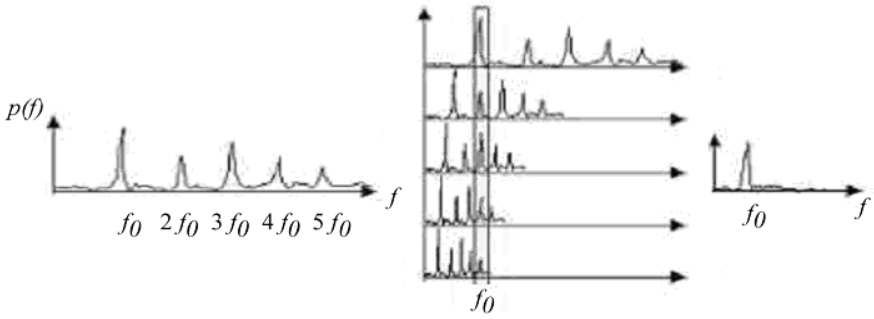

Abbildung 3: Bestimmung der Grundfrequenz f_0 mittels HPS Spektrum: Leistungs-Spektrum , unterabgetastete Leistungs-Spektren, Produkt-Spektrum (v.l.) [Qu02]

Das Verfahren basiert auf einem fortgesetzten Unterabtasten des Leistungs-Spektrums *p(f)*. Führt man beispielsweise eine Unterabtastung um den Faktor 2,3,...,*n* durch, so fällt der Peak der (*n*-1)-ten Harmonischen in dieser gestauchten Darstellung des Leistungs-Spektrum auf den Peak der Grundfrequenz bei f_0 (siehe hierzu Abbildung 3). Multipliziert man nun die verschiedenen unterabgetasteten Spektren miteinander, lässt sich die Grundfrequenz als Peak im Produkt-Spektrum erkennen.

Phasenraum-Darstellung: Ausgangspunkt dieses Verfahrens ist die Darstellung des Signals *s(t)* im sog. *Phasenraum*. Dessen einfachste Ausprägung ist dadurch definiert, zu jedem Zeitpunkt *t* den Wert des Signals *s(t)* und seine erste zeitliche Ableitung zusammenzufassen zu einer neuen Größe

$$\phi(t) := (s(t), \dot{s}(t)) \in \mathbb{R}^2 \quad ,$$

einem Punkt im zweidimensionalen *Phasenraum* [Mi99] [Ge99].

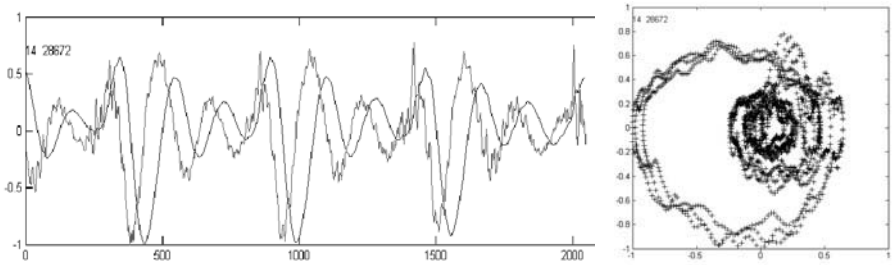

Abbildung 4: Zeitsignal s(t), 2048 Abtastwerte und dessen zeitliche Ableitung (links); zugehörige Phasenraum-Darstellung normiert auf [+1,-1] (rechts)

Ein leicht verrauschtes periodisches Signal der Periode *T* durchläuft dabei eine annähernd zirkulare Phasenraum-Trajektorie (siehe hierzu Abbildung 4, links). In der zugehörigen Phasenraum-Darstellung (Abbildung 4, rechts) fallen daher jeweils Punkte-Paare *φ(t')* und *φ(t'+T)* für beliebige Zeitpunkte *t'* annähernd an die selbe Stelle im Phasenraum.

Die Bestimmung der im Allgemeinen unbekannten Grundfrequenz basiert nun darauf, zu untersuchen, für welchen zeitlichen Abstand *d* das Abstandsmaß

$$h(d) := \sum_i |\phi(t_i + d) - \phi(t_i)|^2$$

minimal wird. Sehr anschaulich darstellen lässt sich dies im *Periodogramm* (siehe Abbildung 5): Für jede mögliche zeitliche Verschiebung *d* wird aufgetragen, wie oft der euklidische Abstand jedes Punktes im Phasenraum zu allen seinen Nachbarn unterhalb einer festgelegten Schwelle liegt. Im betrachteten Beispiel erkennt man im Periodogramm deutlich die Periode T=550 Samples und somit die Grundfrequenz $f_0 = 1/T$.

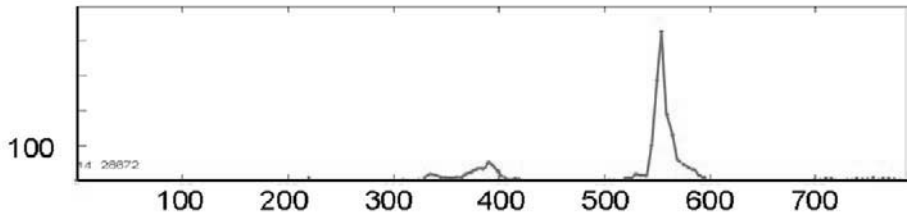

Abbildung 5: Periodogramm - Abszisse: zeitl. Verschiebung *d*; Ordinate: Anzahl von paarweise Abständen im Phasenraum unter einer gegebenen Schwelle

Das beschriebene Verfahren lässt sich auf Phasenraumdarstellung höherer Dimension (d.h. mit Zeitableitungen höherer Ordnung *n*) erweitern [Mi99].

3.2 Tonale Komponenten

Dazu wurde auf das psychoakustische Modell des MPEG-1-Audio Standards [MP92] zurückgegriffen. Das beschriebene Verfahren basiert darauf, im Leistungs-Spektrum hinreichend markante lokale Maxima, sog. *tonale Komponenten* zu erkennen. Die Motivation für das Erkennen von inhaltsverändernden Manipulationen anhand tonaler Komponenten liegt darin, dass deren Existenz und Lage sehr charakteristisch für die wahrgenommenen Klangeigenschaften ist. Unser Verfahren ist eine kompaktere Alternative zu dem in [Ra02], wo ebenfalls Eigenschaften der tonalen Komponenten und der davon abgeleiteten Maskierungsschwelle als Merkmal extrahiert werden.

Die Detektion der tonalen Komponenten wird wie in [MP92] vorgenommen. Als inhaltsfragiles Merkmal werden die Indizes der identifizierten tonalen Frequenzbänder definiert.

4 Ergebnisse

In diesem Kapitel untersuchen wir die vorgestellten Merkmale auf ihre Eignung zur Verifikation der semantischen Integrität von Audiodaten. Dazu diskutieren wir zuerst die Simulation erlaubter Veränderungen und verbotener Manipulationen.

4.1 Simulierte Angriffe

Für die Trainingsphase haben wir eine Reihe von erlaubten und verbotenen Signalveränderungen definiert:

- **erlaubte Signalveränderungen**: Einbettung von Wasserzeichen, MP3-Kompression (128kBit/s und 64kBit/s), schwaches additives Rauschen (Maximalpegel -45 dB)
- **verbotene Manipulationen:** Ersetzen durch Stille/ Hintergrundrauschen der selben Datei (entspricht Löschen eines Lautes), Neu-Anordnung der Reihenfolge (Vertauschen von Lauten), starkes additives Rauschen (Maximalpegel -25 dB)

Zu den erlaubten Veränderungen zählt also insbesondere die Markierung der Datei mit einem robusten Wasserzeichen-Verfahren. Es wurde hierbei exemplarisch das Verfahren nach [St03] gewählt.

4.2 Trainingsphase

Hier wurde untersucht, wie sich die betrachteten Merkmale für das gewählte Trainings-Set verhalten, wenn das Material mit erlaubten/verbotenen Angriffen verändert wurde. Ausgangspunkt war ein Trainings-Set von Sprachaufzeichnungen, beispielsweise aus dem *SQAM Sound Quality Assessment Material* der EBU (*European Broadcasting Union*), aus dem deutschen Rundfunkarchiv, Hörbuchaufnahmen sowie Sprache aus Filmaufnahmen zum Einsatz (Spieldauer: 10 Minuten). Es werden die extrahierten Merkmale Frame für Frame miteinander verglichen.

Tonale Komponenten: Aus Abbildung 6 wird deutlich, dass die Extraktion von tonalen Komponenten eine gute Trennung zwischen erlaubten und verbotenen Angriffen erlaubt: Für erlaubte Angriffe ist die Verteilung der prozentualen Übereinstimmung stark rechtslastig, umgekehrt für die verbotenen Angriffe stark linkslastig.

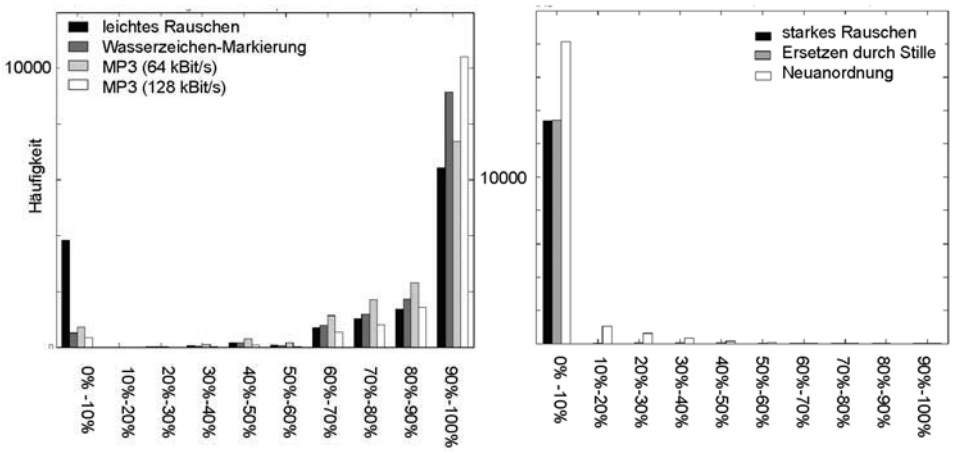

Abbildung 6: Histogramm der prozentualen Übereinstimmung tonaler Komponenten für erlaubte Veränderungen (links) und verbotene Manipulationen (rechts)

Setzt man nun den Schwellwert für die Entscheidung, ob eine Datei einer verbotenen Manipulation ausgesetzt war beispw. auf „25%“, so erhält man folgende Erkennungsraten und Fehler erster sowie zweiter Art (siehe Tabelle 1):

Erlaubte Veränderung	korrekt akzeptiert	falsche Zurückw.
Leichtes Rauschen (-45 dB)	71,96%	28,04%
Wasserzeichenmarkierung	96,21%	3,79%
MP3 (64kBit/s)	94,2%	5,28%
MP3 (128kBit/s)	97,53%	2,47%

Tabelle 1:Tonale Komponenten - Anteil korrekt akzeptierter und fälschlich zurückgewiesener erlaubter Veränderungen

In Tabelle 1 fällt auf, dass selbst ein leichtes Rauschen in fast jedem dritten Frame leider fälschlicherweise als unerlaubte Manipulation erkannt wird. Bei obiger Festsetzung der Schwelle werden die anderen erlaubten Veränderungen mit hoher Erkennungsrate als solche verifiziert. In Tabelle 2 fällt auf, dass die unerlaubten Veränderungen durchweg mit hoher Erkennungsrate korrekt erkannt wurden.

Unerlaubte Manipulation	korrekte Zurückw.	falsche Akzeptanz
Starkes Rauschen (-25 dB)	99,51%	0,49%
Vergleich mit Stille	98,51%	1,49%
Neusortieren der Reihenfolge	96,23%	3,77%

Tabelle 2: Tonale Komponenten – Anteil korrekt zurückgewiesener und fälschlich akzeptierter verbotener Manipulationen

Grundfrequenzbestimmung:

Bevor wir die Grundfrequenz bestimmen, werden die Sprachdaten in stimmhafte und stimmlose Anteil segmentiert. Wir verwenden hierzu u.a. die Nulldurchgangsrate (*zero crosing rate*, ZCR) und die Lage des 95%-Quantils des Leistungsspektrums (*spectral roll off point*), siehe hierzu [Tz99]. Nur für die stimmhaften/ tonalen Anteile wird die Grundfrequenz bestimmt. Die Grundfrequenzen der getesteten Sprachdateien liegen größenordnungsmäßig im Bereich von 100Hz. Nach [ZF90] ist hierbei eine Frequenz-Variation von 3,5% leicht wahrnehmbar. Setzen wir daher die Entscheidungsschwelle für die Zurückweisung/ Annahme daher restriktiv auf 5 Hz, so erhalten wir folgende Erkennungs- und Fehlerraten:

Erlaubte Veränderung	korrekt akzeptiert	falsche Zurückw.
Leichtes Rauschen	74,03%	25,97%
Wasserzeichenmarkierung	95,12%	4,88%
MP3 (64kBit/s)	93,44%	6,56%
MP3 (128kBit/s)	96,97%	3,03%

Tabelle 3: HPS-Verfahren - Anteil korrekt akzeptierter und fälschlich zurückgewiesener erlaubter Veränderungen

Auffällig ist in Tabelle 3, dass ein leichtes Rauschen abermals zu einer hohen Rate von falschen Alarmen führt. Dies lässt sich dadurch erklären, dass in diesem Fall ein Frame irrtümlich als stimmlos/rauschartig segmentiert wird. Daher wird die Grundfrequenz-Schätzung nicht durchgeführt – und hat damit eine scheinbar große Frequenzänderung erlitten.

Unerlaubte Manipulation	korrekte Zurückw.	falsche Akzeptanz
Starkes Rauschen	99,52%	0,48%
Vergleich mit Stille	99,16%	0,84%
Neusortieren der Reihenfolge	90,6%	9,4%

Tabelle 4: HPS-Verfahren - Anteil korrekt zurückgewiesener und fälschlich akzeptierter verbotener Manipulationen

In Tabelle 4 fällt auf, dass beim Angriff „Neusortieren der Reihenfolge“ ca. 10% fälschliche Akzeptanzen beobachtet werden können. Dies rührt daher, dass bei der von uns (automatisch) durchgeführten Neuanordnung u.U. ein bestimmter Frame ersetzt wurde durch einen anderen, der die gleiche Grundfrequenz besitzt. Dies zeigt die Grenzen der Sicherheit des Verfahrens. Ein Austauschen des selben Frames durch Stille lässt sich hiermit gleichwohl erkennen.

Erlaubte Veränderung	korrekte Akzeptanz	falsche Zurückw.
Leichtes Rauschen	67,82%	32,18%
Wasserzeichenmarkierung	92,89%	7,11%
MP3 (64kBit/s)	92,61%	7,39%
MP3 (128kBit/s)	95,43%	4,57%

Tabelle 5: Phasenraum-Verfahren - Anteil korrekt akzeptierter und fälschlich zurückgewiesener erlaubter Veränderungen

In Tabelle 5 erkennt man abermals die hohe Fehlerrate bei leichtem Rauschen. Weiterhin sieht man, dass das Phasenraum-Verfahren bei den gewählten Einstellungen insgesamt eine etwas niedrigere Erkennungsrate bei eigentlich erlaubten Veränderungen besitzt als das HPS-Verfahren: Einen Grund hierfür erkennt man in Tabelle 6: gleichzeitig ist nämlich die Erkennungsrate für das „Neusortieren der Reihenfolge“ höher als beim HPS-Verfahren. Dies spricht dafür, dass der Schwellenwert von 5Hz im Vergleich etwas zu restriktiv gewählt ist und daher etwas mehr falsche Zurückweisungen *und* gleichzeitig weniger fälschlich akzeptierte Manipulationen produziert: Fehler erster und zweiter Art bedingen sich stets gegenseitig und hängen beide von der gesetzten Schwelle ab.

Unerlaubte Manipulation	korrekte Zurückw.	falsche Akzeptanz
Starkes Rauschen	99,85%	0,15%
Vergleich mit Stille	98,93%	1,07%
Neusortieren der Reihenfolge	94,49%	5,51%

Tabelle 6: Phasenraum-Verfahren - Anteil korrekt zurückgewiesener und fälschlich akzeptierter Manipulationen.

4.2 Erkennungsphase

Zur Demonstration der Erkennungsleistung wurde eine Testdatei (Hörbuchaufnahme, männliche Stimme, Länge ca. 0:30 Min), an einigen Stellen unerlaubt manipuliert (siehe Abbildung 7): und anschließend nach MP3 (128 kBit/s) gewandelt.

- **Manipulation A:** Überlagerung mit einer weiblichen Stimme
- **Manipulation B:** Ersetzen durch Hintergrundgeräusche aus der selben Datei
- **Manipulation C:** Anheben der Grundfrequenz / Ändern der Betonung

Man erkennt in Abbildung 7, dass sich die vorgenommen Manipulationen A-C mit den vorher identifizierten Schwellwerten sowohl anhand der tonalen Komponenten als auch im Differenzbild der geschätzten Grundfrequenzen deutlich erkennen lassen. Die vereinzelten Fehldetektionen der Grundfrequenzen rühren von der MP3-Kompression und treten hierbei nur als zeitlich isolierte Ausreißer auf. Nur an den tatsächlich manipulierten Datei-Positionen werden Frequenz-Veränderungen auch über mehrere aufeinanderfolgende Frames hinweg detektiert. Auf diese Weise ließen sich falsche Detektionen nachträglich erkennen und verhindern. Die Rate von fälschlichen

Detektionen vermeintlicher Manipulationen kann weiterhin durch Merkmalskombination der oben untersuchten und Hinzuziehung weiterer – noch zu untersuchende – Audiomerkmale verringert werden.

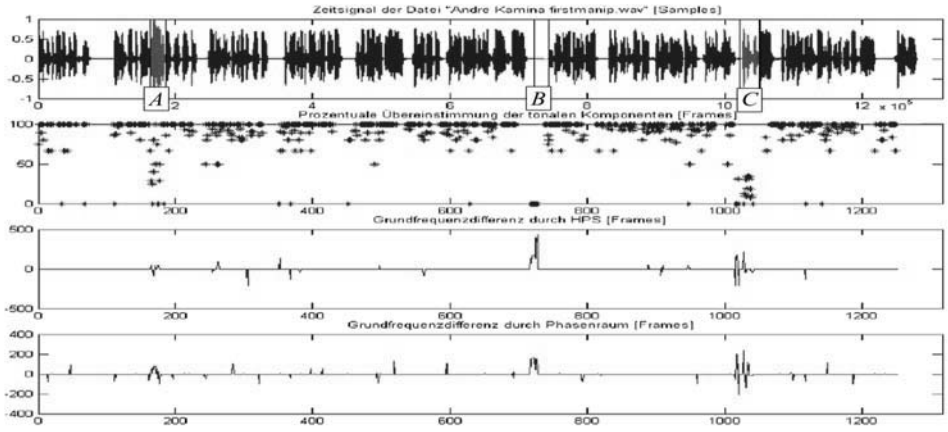

Abbildung 7: Erkennen der unerlaubten Manipulationen A,B,C: (v.o.) Zeitsignal der Testdatei, prozentuale Übereinstimmung tonaler Komponenten, Differenz der Grundfrequenzbest. mittels HPS und Phasenraumdarstellung. Gesamtdauer: ca. 30s

5 Zusammenfassung und Ausblick

Digitale Medien sind leicht zu bearbeiten und inhaltlich zu manipulieren, was gleichzeitig eine leichte Handhabe und eine Gefährdung der Vertrauenswürdigkeit mit sich bringt. Wir haben uns daher mit der Frage beschäftigt, wie mittels inhaltsfragiler Audio-Wasserzeichen die semantische Integrität von Audiodaten effizient geschützt werden kann. Wir haben die Eignung einiger Merkmale von Audiodateien nachgewiesen, zwischen verbotenen Manipulationen und erlaubten Nachbearbeitungen zu unterscheiden. Gleichzeitig sind diese Merkmale robust gegenüber dem Einbetten eines Wasserzeichens. Damit ist eine wichtige Voraussetzung zur Integration dieser Merkmale in ein inhaltsfragiles Wasserzeichen-System erfüllt. In einem nächsten Schritt ist es notwendig, die untersuchten Merkmale auf eine kompaktere Darstellung abzubilden und die Fehlerraten durch geeignete Nachverarbeitungsschritte, beispielsweise zeitliche Glättung oder Merkmalskombination zu verbessern.

Literaturverzeichnis

[Ca02] P. Cano and E. Batlle and E. Gomez and L. de and C. Gomes and M. Bonnet, *Audio fingerprinting: concepts and applications* In: 1st International Conference on Fuzzy Systems and Knowledge Discovery, Singapore, November 2002.

[Di00] Dittmann, Jana, *Digitale Wasserzeichen*, Springer Verlag, Berlin, ISBN 3-540-66661-3, 2000.

[Fr01] Fridrich, Goljan, Du; *Invertible authentication*, Proceedings of SPIE: Security and Watermarking of Multimedia Contents III, S. 197-208, San Jose, California, USA, Vol. 4314, 2001.

[Ge99] Gerhard, David, *Audio Visualization in Phase Space,* 2nd Annual Conference of BRIDGES: Mathematical Connections in Art, Music, and Science, Southwestern College, Winfield, Kansas, July 30 - August 1 1999, pp. 137ff., 1999.

[Go03] Gomes, L. and Cano, P. and Gómez, E. and Bonnet, M. and Batlle, E., *Audio Watermarking and Fingerprinting: For Which Applications?*, Journal of New Music Research", vol. 32, no. 1, 2003.

[MP92] ISO/IEC 11172-3, *Coding of moving pictures and associated audio for digital storage media at up to about 1.5 MBit/s – Part 3: Audio*, ISO/IEC JTC 1/SC,1993.

[Mi99] Mierswa, Ingo, *Beatles vs. Bach: Merkmalsextraktion im Phasenraum von Audiodaten*, GI-Workshop "Lehren – Lernen – Wissen – Adaptivität", GI-Arbeitskreis Knowledge Discovery (FGML/AKKD), Karlsruhe, 6.-8.Oktober 2003.

[Qu02] Quast, Holger; Schreiner, Olaf; Schroeder, Manfred Robert: *Robust Pitch Tracking in the Car Environmen.* In: IEEE (Veranst.) : Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2002 (ICASSP 2002 Orlando May 13 - 17, 2002). 2002, S. 353-356 (on CD: 3087.pdf).

[Ra02] R. Radhakrishnan and N. Memon. *Audio Content Authentication Based on Psycho-Acoustic Model.* SPIE Security and Watermarking of Multimedia Contents, San Jose, CA, February 2002.

[St03] Martin Steinebach, *Digitale Wasserzeichen für Audiodaten*, Dissertationsschrift, Shaker Verlag Aachen, ISBN 3-8322-2507-2 .

[St96] D. Storck, *A New Approach to Integrity of Digital Images*. IFIP World Conference on Mobile Communications 1996: 309-316.

[Tz99] Tzanetakis G., Cook P.: *Multifeature Audio Segmentation for Browsing and Annotation.* Computer Science Department, Princeton University, Princeton New Jersey, USA 1999.

[ZF90] Zwicker, E. and Fastl, H., Psychoacustics – Facts and Models, Springer, Berlin, 1990.