# Authentifizierung von Audiodaten

# mittels inhalts-fragiler Wasserzeichen

**Dissertation**

zur

Erlangung des akademischen Grades

**Doktor-Ingenieur (Dr.-Ing.)**

der Fakultät für Informatik und Elektrotechnik

der Universität Rostock

vorgelegt von

Michael Gulbis, geboren am 30. September 1980 in Rostock

aus Rostock

Rostock, 7. September 2012

urn:nbn:de:gbv:28-diss2013-0036-2

Als Dissertation genehmigt von der Fakultät für Informatik und Elektrotechnik der Universität Rostock

**Gutachter:**

- Prof. Dr.-Ing. habil. Erika Müller, Universität Rostock
- Prof. Dr. Stefan Katzenbeisser, Technische Universität Darmstadt
- Dr.-Ing. Martin Steinebach, Fraunhofer-Institut für Sichere Informationstechnologie

**Tag der Einreichung:**

07. September 2012

**Tag der öffentlichen Verteidigung:**

07. Februar 2013

# Vorwort

Der wesentliche Anteil der vorliegenden Arbeit entstand während meiner Tätigkeit als Promotionsstipendiat und wissenschaftlicher Mitarbeiter am Institut für Nachrichtentechnik der Fakultät für Informatik und Elektrotechnik der Universität Rostock. Sie wurde durch ein Stipendium der Deutschen Forschungsgemeinschaft im Rahmen des Graduiertenkollegs „Verarbeitung, Verwaltung, Darstellung und Transfer multimedialer Daten - technische Grundlagen und gesellschaftliche Implikationen“ gefördert.

Ich danke meiner Mentorin Frau Prof. Dr.-Ing. habil. Erika Müller für die fachliche als auch menschliche Betreuung, ohne die beständigen Bemühungen um Ihre „Doktoranden“ wäre diese und andere Arbeiten nicht denkbar.

Ich möchte mich weiterhin bei allen Mitarbeitern des Instituts für Nachrichtentechnik für die angenehme und fruchtbare Arbeitsumgebung bedanken. Hervorheben möchte ich hier Ralph Hänsel, welcher immer Zeit und Energie für Diskussionen hatte, und Stephan Lange, unseren Systemadministrator, welcher mich bei der Einrichtung aller technischen Sonderwünsche unterstützte.

Mein besonderer Dank gilt meiner Familie und meiner Frau Cornelia, für die moralische Unterstützung und die häufigen Ermutigungen.

Widmen möchte ich diese Arbeit meinem Vater, Dieter Gulbis, welcher die Fertigstellung nicht mehr erleben durfte.

# Kurzfassung

Die Bedeutung von medialen Daten für die Informationvermittlung nimmt beständig zu. Die rasche Entwicklung in der Digitaltechnik hat zu vielfältigen Möglichkeiten der Produktion, Verbreitung und Bearbeitung von medialen Daten geführt. Diese Vorteile resultieren auch in der Problematik, dass hochqualitative Fälschungen kostengünstig und ohne spezielle Kenntnisse erstellt werden können. Die Integrität von Audiodaten ist grundsätzlich anzuzweifeln.

Die vorliegende Arbeit beschreibt die Entwicklung, Umsetzung und Evaluierung eines inhaltsfragilen Wasserzeichenverfahren zur Authentifizierung von Audiodaten. Um die Integrität von Audiodaten nachweisen zu können, wird eine Beschreibung deren Inhalts mit einem Wasserzeichen in die Daten eingebettet. Das Inhaltsmerkmal wird durch binäre Codierung der Veränderungen in den Frequenzgruppen der Audiodaten gebildet. Über das Merkmal lassen sich inhaltserhaltende von böswilligen Veränderungen unterscheiden. Die Wasserzeichentechnik ist speziell mit der Hinsicht entwickelt, transparent für die Merkmalsextraktion zu sein.

Die hohen Anforderungen an die Wasserzeichenkapazität wird durch eine überlagernde Einbettung von mehreren Wasserzeichen erfüllt. Durch eine hierarchische Struktur der Einbettungsdomain sind die Einbettungen gegenseitig störungsfrei. Aus der beschädigten Wasserzeicheninformation einer gestörten Audiodatei wird eine Zuverlässigkeitsinformation für eine *Soft-Decision*-Fehlerdecodierung generiert. Durch die Einführung einer Totzone kann die Robustheit des Inhaltsmerkmals in stillen Audiobereichen und gegenüber quantisierungsbedingten Fehlern verbessert werden.

Die Leistungsfähigkeit des Inhaltsmerkmals und die Robustheit der Wasserzeichentechnik wird ausführlich analysiert. Zu den Operationen zählen verlustbehaftete Audiokompression, Unterabtastung, Lautstärkeveränderungen, Normalisierung, das Hinzufüge von unterschiedlichen Rauscharten und der Austausch von Audiopassagen. Das entwickelte System wird mit ähnlichen Verfahren anderer Autoren verglichen.

# Abstract

Multimedia data has become an important and widely used carrier of information. The rapid development of technology yields to many ways of production, distribution, archiving and editing of multimedia data. This advancements inhere the problem that high quality forgery is be created at a relatively low cost and no sophisticated knowledge. Hence, the integrity of audio data has to be questioned.

The present work describes the design, implementation and evaluation of a content-fragile watermarking system for the authentication of audio data. To verify the integrity of audio data a description of its content is embedded with a digital watermark within the audio data itself. The feature is constructed by coding the changes in critical bands into a binary representation. This feature allows the distinction between content preserving and malicious modifications of audio data. The watermarking technique is specially designed to be transparent for the content feature embedding.

The problem of high watermark payload requirements are addressed with a stacked embedding of multiple watermarks. The single watermarks do not disturb each other due to an hierarchical embedding structure. The disturbed watermark information of manipulated audio data is used to created a fidelity information for a soft decision error decoding. The introduction of a deadband improves the robustness of the feature in silent audio parts and errors caused by quantization are avoided.

The performance of the content feature and the robustness of the watermarking technique is extensively analyzed by the means of lossy audio compression, subsampling, change of volume, normalization, adding different kinds of noise to the audio and exchange of several audio parts. The main features of the developed system are compared to similar methods proposed by other authors.

# Inhaltsverzeichnis

# Abkürzungen und Symbole

## Abkürzungen

| | |
|---|---|
| ARQ | Rückwärtsfehlerkorrektur (engl. *Automatic Repeat Request*) |
| AWGN | additives weißes Gaußsches Rauschen (engl. *additive white Gaussian noise*) |
| DCT | Diskrete Kosinustransformation (engl. *Discrete Cosine Transform*) |
| DCT-IV | Diskrete Kosinustransformation (engl. *Discrete Cosine Transform*) (Typ IV) |
| DIX | *Disturbance Index* |
| DWT | Diskrete Wavelet-Transformation (engl. *Discrete Wavelet Transform*) |
| FEC | Vorwärtsfehlerkorrektur (engl. *Forward Error Correction*) |
| FFT | Schnelle Fourier-Transformation (engl. *Fast-Fourier-Transform*) |
| FFTW | *Fastest Fourier Transform in the West* |
| LLR | Log-Likelihood Verhältnis (engl . *Log-Likelihood Ratio*) |
| NMR | *Noise-to-Mask Ratio* |
| OASE | *Objective Audio Signal Evaluation* |
| ODG | *Objective Difference Grade* |
| PAQM | *Perceptual Audio Quality Measure* |
| PCM | Puls-Code-Modulation |
| PEAQ | *Perceptual Evaluation of Audio Quality* |
| Perceval | *Perceptual Evaluation* |
| POM | *Perceptual Objective Measure* |
| RIAA | *Recording Industry Association of America* |
| rMAC | *robust Message Authentication Code* |

| | |
|---|---|
| RMS | Effektivwert (engl. *Root Mean Square*) |
| SDG | *Subjective Difference Grade* |
| SNR | s. SRV |
| SPL | Schalldruckpegel (engl. *Sound Pressure Level*) |
| SRV | Signal-zu-Rausch-Verhältnis (engl. *Signal-to-Noise-Ratio* - SNR) |
| ZCR | Nulldurchgangsrate (engl. *Zero Crossing Rate*) |

# Symbole

| | |
|---|---|
| $a$ | Koeffizient der Gruppe $\mathbb{A}$ |
| $\mathbb{A}$ | Erste Koeffizientengruppe für die Wasserzeicheneinbettung |
| $b$ | Koeffizient der Gruppe $\mathbb{B}$ |
| $\mathbb{B}$ | Zweite Koeffizientengruppe für die Wasserzeicheneinbettung |
| $s$ | Koeffizient |
| $\mathbf{c}$ | Vektor von Koeffizienten |
| $\mathbb{C}$ | Menge von Koeffizienten |
| $d$ | Zwischengröße bei der Bestimmung des Inhaltsmerkmals |
| $\Delta d$ | Ein noch nicht binär entschiedener Wert des Inhaltsmerkmals |
| $\varepsilon$ | Distanzwert zwischen den Koeffizientengruppen der Wasserzeicheneinbettung |
| $f$ | Einbettungsstärke der Wasserzeicheneinbettung |
| $Fj$ | Frequenzgruppe mit der Nummer $j$ |
| $\mathbf{g}$ | Generator eines Faltungscodes |
| $k$ | Korrektur-Wert für eine $d$-Wert bei der Merkmalsverstärkung |
| $l$ | Segmentierungslänge |
| $l_M$ | Segmentierungslänge für die Merkmalsextraktion |
| $l_W$ | Segmentierungslänge für die Wasserzeicheneinbettung |
| $m$ | Bitwert des Inhaltsmerkmals |
| $\mathbf{m}$ | Vektor von Bitwerten des Inhaltsmerkmal |
| $\mathbf{M}$ | Matrixdarstellung des Inhaltsmerkmals |
| $s$ | Sample |
| $\mathbf{s}$ | Audiosequenz |
| $\mathbf{S}$ | Matrizendarstellung einer segmentierten Audiosequenz |
| $th$ | Schwellwert |

| | |
|---|---|
| $th_{tot}$ | Schwellwert der Markmalsverstärkung |
| $u$ | Verschlüsselte Nutzinformation |
| $v$ | Kanalcodierte Nutzinformation |
| $w$ | Bit der Wasserzeicheninformation |
| $\mathbf{w}$ | Wasserzeicheninformation |

# Operatoren

| | |
|---|---|
| $n(x)$ | Anzahl der Elemente von $x$ |
| $max(\mathbb{X})$ | Größtes Element der Menge $\mathbb{X}$ |
| $sgn(x)$ | Abbildung des Vorzeichen des Wertes $x$ auf „0“ für negative Werte von $x$ und „1“ für positive Werte von $x$ und Null |

Kapitel 1

# Einleitung

## 1.1. Thematischer Hintergrund

Mit dem Einzug ins Informationszeitalter werden analoge Ton-, Bild- und Videomedien durch digitale ersetzt. Für die Vermittlung von Informationen haben digitale Multimediadaten verstärkt an Bedeutung gewonnen. Entscheidend hierfür sind die fortschreitenden Entwicklungen im Bereich der Digitaltechnik, welche die Produktion, Übertragung, Archivierung und die Bearbeitung multimedialer Daten kostengünstig und auch ohne spezielle Kenntnisse mit handelsüblichen Werkzeugen ermöglichen.

Diese Vorzüge stellen jedoch auch den größten Nachteil dar. Ein vergleichbar hoher zeitlicher und technischer Aufwand und die daraus resultierend hohen Kosten für die Verarbeitung von analogen Daten hielt die Herstellung von Fälschungen in Grenzen und war nur bei einem besonderen Interesse Dritter zu erwarten. Veränderungen digitaler Daten sind vergleichbar einfach und schwer nachweisbar. Bei analogen Kopien kumulieren sich Qualitätsverluste mit jeder zusätzlichen Instanz. Digitale Kopien können schnell und ohne Qualitätsverluste erstellt werden. Die massenhafte Verbreitung von digitalen Daten stellt nicht nur das Problem von nicht genehmigten Kopien durch den Rechteinhaber dar, sondern erschwert auch die Rückverfolgung und Identifizierung des Urhebers bei dezentraler Verbreitung über mehrere Instanzen. Die Integrität von digitalen Daten ist daher grundsätzlich anzuzweifeln.

Gebräuchliche kryptographische Verfahren zum Schutz der Datenintegrität und/oder deren Authentizität wie z.B. Prüfsummen, fehlerkorrigierende Codes, digitale Signaturen etc. besitzen Nachteile im Hinsicht auf die Eigenschaften multimedialer Daten. Im Vergleich zu textuellen Informationen besitzen multimediale Daten eine größere Unabhängigkeit des Informationsgehaltes von ihrer binären Repräsentation. Textdaten erlauben gewöhnlich nur eine binäre Darstellung. Änderungen führen hier folglich zu Fehlern und Veränderungen des Dateninhaltes. Im Gegensatz hierzu kann bei multimedialen Daten der Inhalt selbst nach starken Veränderungen der binären Darstellung erhalten bleiben. Ein typisches Beispiel hierfür die verlustbehaftete Kompression. Die Bedeutung einer Sprachaufnahme verändert sich nicht, ob sie nur in Form von PCM-Daten oder in einem verlustbehafteten Format wie MP3 gespeichert wird. Der Integritätsschutz von multimedialen Daten bezieht sich somit auf den Erhalt des wahrnehmbaren Inhalts der Daten bzw. auf deren Semantik.

Für diese Aufgabe haben sich die digitale Wasserzeichen als geeignete Technologie erwiesen. Digitale Wasserzeichen stellen zusätzliche Informationen dar, welche durch gezielte Modifikation von digitalen multimedialen Daten in die selben einbettet werden. Die Einbettung von digitalen Wasserzeichen ist typischer Weise so gestaltet, dass sich keine wahrnehmbaren Veränderungen der Trägerdaten einstellen. Im Gegensatz zu Meta-Daten, welche in separaten Datenbereichen untergebracht werden und datenformatabhängig sind, stellen digitale Wasserzeichen einen festen Verbund mit den Trägerdaten her. Die mit digitalen Wasserzeichen eingefügten Informationen schränken die gewohnte Verwendung und Verarbeitung der Daten durch einen Anwender nicht ein. Durch die Einbettung von Informationen über Urheber, Rechteinhaber, Käufer, Verwendungsrechten, Inhalt, Meta-Daten, etc. sind eine Vielzahl an Anwendungsszenarien realisierbar. Auf diese Weise kann die Integrität bzw. Authentizität von Daten nachgewiesen werden. Verwertungsrechte können durch die Identifizierung und Rückverfolgung von unerlaubten Kopien durchgesetzt werden. Eine automatische Organisation und Administration von Daten ist mit Einbettung von Annotationen möglich.

## 1.2. Motivation und Zielstellung

Ziel der vorliegenden Arbeit ist es, ein auf digitalen Wasserzeichen basierendes Verfahren zur Verifikation der Echtheit von Audiodaten zu entwickeln. Die einzelnen Teilziele ergeben sich wie folgt:

- Eine Audiodatei kann als echt bezeichnet werden, wenn deren inhaltliche Aussage und Qualität unverändert sind. Das zu entwickelnde Beschreibungsmerkmal der Audiodaten

muss somit sensibel genug sein, um alle wahrnehmbaren und inhaltlichen Veränderung der Audiodaten erkennen zu können.

- Mit dem Bezug auf Sprachdaten ist mit Modifikationen ab dem Umfang einer Silbe mit einer inhaltlichen Veränderung von Audiodaten zu rechnen. Die obere Grenze für die Lokalität eines Inhaltsmerkmals ist mit 200 bis 333 ms (Umfang einer Silbe [Ter98]) festzulegen.

- Der gewöhnliche Umgang mit den geschützten Audiodaten durch den Nutzer darf nicht eingeschränkt werden. Eine übliche Verarbeitung mit Hinsicht auf die Verbreitung und Archivierung der Audiodaten muss ohne Zerstörung der Schutzinformation möglich sein. Hieraus folgt, dass das Inhaltsmerkmal und die Wasserzeichentechnik ausreichend robust sein müssen, um Verarbeitungsschritte ohne qualitative oder inhaltliche Veränderungen der Audiodaten, wie z.B. verlustbehaftete Kompression, ohne Schäden zu überstehen.

- Unmittelbar nach der Integration der Schutzinformation in die Audiodaten dürfen keine Veränderungen der Inhaltsmerkmale trotz Modifikation der Audiodaten durch die Wasserzeicheneinbettung entstehen. Die Wasserzeichentechnik ist so zu gestalten, dass ein eingebettetes Wasserzeichen keinen Einfluss auf die Extraktion des Inhaltsmerkmals ausübt.

- Ein unabsichtliches Entfernen des Wasserzeichens ohne Veränderung des Dateninhaltes ist zu verhindern. Die Wasserzeicheneinbettung hat somit in der gleichen Domain wie die Extraktion des Beschreibungsmerkmales zu erfolgen.

- Die Kapazität des Wasserzeichens muss ausreichend groß sein, um das Beschreibungsmerkmal transportieren zu können.

- Die Datenqualität der Audiodaten darf durch die Wasserzeicheneinbettung nicht wahrnehmbar vermindert werden.

- Die öffentliche Verifizierung der Echtheit geschützter Audiodaten, folglich durch jedermann, soll ermöglicht werden.

- Die Schutzinformation muss manipulationssicher sein, so dass es keinem Angreifer möglich ist, inhaltlich veränderte Audiodaten mit einer gültigen, gefälschten Schutzinformation zu versehen.

## 1.3. Gliederung der Arbeit

Die vorliegende Arbeit ist in sechs Kapitel untergliedert. Die inhaltliche Einordnung der Teilelemente des entwickelten Authentifizierungssystems wird in Abbildung 1.1 dargestellt.

**Kapitel 2** bietet einen Einblick in die Grundlagen der Arbeit. Hier wird zu Beginn ein Überblick über die Aspekte der menschlichen Wahrnehmung von Audiodaten und die Bedeutung der Frequenzgruppen für diese gegeben. Der Schwerpunkt liegt auf der Einführung in die Technologie der digitalen Wasserzeichen. Beide Punkte resultieren aus dem Fokus der Arbeit auf die Entwicklung eines Authentifizierungssystems für digitale Audiodaten auf der Basis eines digitalen Wasserzeichenverfahren. Als weiterer Punkt wird in den Grundlagen die Kanalkodierung betrachtet. Hier wird insbesondere auf die Fehlerkorrektur mittels *Soft-Input*-Decodierung auf Basis des Viterbi-Alorithmus, welche in dieser Arbeit verwendet wird, eingegangen.

Das **Kapitel 3** geht speziell auf die Wasserzeichenverfahren zur Verifizierung der Integrität medialer Daten ein. Die allgemeinen Eigenschaften und die Klassifizierung dieser Verfahren werden vorgestellt. Das in dieser Arbeit entwickelte Wasserzeichenverfahren wird hierauf aufbauend klassifiziert und den inhalts-fragilen Verfahren zugeordnet. Existierende Verfahren der inhalts-fragilen Wasserzeichen für Audiodaten werden vorgestellt und verglichen.

**Kapitel 4** bildet das Kernstück der Arbeit. Die Anforderungen an die zu entwickelnden Systemelemente, das Beschreibungsmerkmal des Audiodateninhalts und die Wasserzeichentechnik zur Integration jener in die Audiodaten, werden definiert. Anschließend werden diese zwei Grundelemente des entwickelten Verfahrens im Detail beschrieben und deren Eigenschaften separat analysiert. Ausgehend von der Analyse werden die günstigen Arbeitsbereiche der Systemelemente identifiziert und die Leistungsanalyse eines Grundsystems aus kombinierter Merkmalsextraktion und Wasserzeicheneinbettung vorgenommen. Den Abschluss des vierten Kapitels bildet die Betrachtung der Sicherheitsaspekte des Systems.

In **Kapitel 5** wird das in Kapitel 4 entwickelte Grundsystem erweitert bzw. modifiziert. Die Robustheit des Inhaltsmerkmals wird zur Reduzierung der Rate falscher Rückweisung gegenüber Störungen mit geringem Einfluss auf die Audiodatenqualität, besonders in leisen und homogenen Audiopassagen, verstärkt. Die robuste Übertragung des Beschreibungsmerkmals mit dem digitalen Wasserzeichen wird durch Fehlerkorrektur in Form von *Soft-Input*-Decodierung der gestörten Wasserzeicheninformation ermöglicht. Des Weiteren wird eine hierarchische Wasserzeichentechnik entwickelt. Diese Technik zeichnet sich durch eine überlagernde sich gegenseitig nicht störende Einbettung von mehreren Wasserzeichen aus. Der Einbettungsbereich der

Wasserzeichentechnik wird hierbei in für die Wahrnehmung weniger relevante Audiodatenbereiche verlagert und reduziert, wodurch die Robustheit bzw. Transparenz der Wasserzeichentechnik positiv beeinflusst wird.

Das **Kapitel 6** fasst die Erkenntnisse der Arbeit zusammen und liefert einen Ausblick auf zukünftige Aufgabenbereiche.

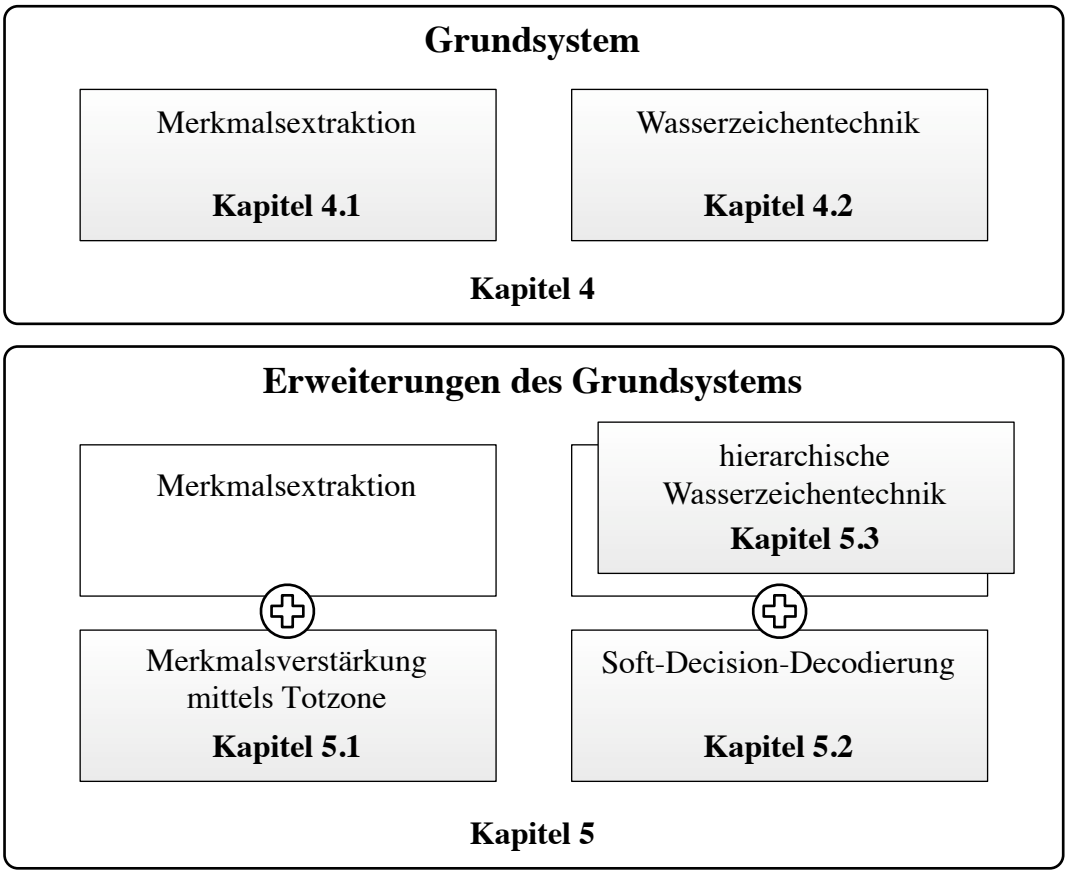


**Abbildung 1.1.:** Einordnung der entwickelten Systemelemente in die Gliederung der Arbeit

# Kapitel 2

# Grundlagen

*Dieses Kapitel beinhaltet wesentliche Grundlagen der in dieser Arbeit betrachteten Thematiken. Zuerst werden ausgewählte Aspekte der menschlichen Wahrnehmung von Schall betrachtet. Dies beinhaltet einen Überblick über den physiologischen Aufbau des menschlichen Ohrs [SLT05], die für die menschliche Wahrnehmung relevanten Frequenzbereiche und eine nichtlineare Frequenzraumunterteilung in Form der Frequenzgruppen. Das Prinzip der Frequenzgruppen wird an ausgewählten Beispielen verdeutlicht.*
*Folgend wird die Kanalcodierung betrachtet. Es wird der Aufbau eines Übertragungskanals und die Einordnung der Kanalcodierung in diesen aufgezeigt. In Bezug auf die Verwendung in der vorliegenden Arbeit wird speziell auf fehlerkorrigierende Maßnahmen in Form der Faltungscodes eingegangen.*
*Der Schwerpunkt dieses Kapitels liegt auf der Einführung in die Technologie der digitalen Wasserzeichen. Das grundlegende Systemkonzept der digitalen Wasserzeichen, deren Eigenschaften und Anwendungsgebiete werden dargelegt.*

## 2.1. Aspekte der menschlichen Wahrnehmung von Schall

### 2.1.1. Das menschliche Ohr

Das menschliche Ohr ist ein komplexes Organ, welches dazu dient, Schallereignisse in für das Gehirn auswertbare neuronale Impulse umzuwandeln. Der Aufbau des menschlichen Ohres ist in Abbildung 2.1 dargestellt. Schall wird durch die Ohrmuschel aufgefangen. Der äußere Gehörgang fungiert als relativ breitbandiger Hohlraumresonator mit einer Resonanzfrequenz

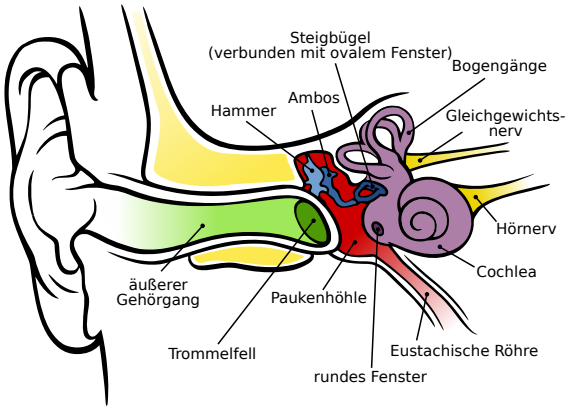


**Abbildung 2.1.:** Aufbau des menschlichen Ohres. [CB05]

von etwa 3,3 kHz. Hierdurch bedingt sich eine große Empfindlichkeit für die Wahrnehmung von Schall zwischen etwa 2 bis 5 kHz. Das Trommelfell wird zum Schwingen angeregt. Die Schwingungen des Trommelfells werden durch die Gehörknöchelchen (Hammer, Amboss und Steigbügel) am ovalen Fenster auf die mit Flüssigkeit gefüllte Cochlea übertragen. Die Gehörknöchelchen übernehmen hierbei die Aufgabe eines Impedanzwandlers. Beim Übergang einer Schallwelle zwischen zwei Medien wird die Schallwelle umso stärker reflektiert, je größer die Differenz der Schallkennimpedanzen ist. Im Falle des menschlichen Ohres werden etwa 60% der Schallenergie übertragen. Ohne Impedanzwandlung werden beim Übergang von Luft zur Flüssigkeit der Cochlea etwa 98% der Schallenergie reflektiert [SLT05, S. 339]. Die Cochlea besteht aus vier übereinander liegenden Strukturen, drei mit Flüssigkeit gefüllten Kanälen (sog. Skalen) und dem Corti-Organ. Die vereinfachte Struktur der abgerollten Cochlea ist in Abbildung 2.2 dargestellt. Die *Scala vestibuli* und *Scala tympani* sind an der Spitze der Cochlea durch das Helikotrema (Schneckenloch) verbunden. Die *Scala media* ist zur *Scala vestibuli* durch die Reissner-Membran und zur *Scala tympani* durch die Basilarmembran abgegrenzt. Auf der Basilarmembran befindet sich das Corti-Organ. An der Basis ist die Cochlea über das ovale Fenster mit der Fußplatte des Steigbügels verbunden. Die Flüssigkeiten in den Skalen sind inkompressibel. Wird eine Schallwelle über das ovale Fenster in die *Scala vestibuli* eingekoppelt, weicht die Flüssigkeit dieser aus. Dabei werden die darunter liegenden Strukturen, Reissner-Membran, *Scala media*, Corti-Organ und Basilarmembran, nach unten gedrückt. Zum Druckausgleich wölbt sich die flexible Membran des runden Fensters ins Mittelohr aus. Entsprechend der Schwingung am ovalen Fenster schließt sich im weiteren Verlauf eine entgegengesetzte Bewegung an. Die Membranen und das Corti-Organ geraten in Schwingungen. Die Amplitude der Schwingung wird nur an den Stellen der Struktur aus Membranen und Corti-Organ stark genug, um die Rezeptorzellen (Haarzellen) zu stimulieren, an denen sie Resonanz mit der Schallwellenfrequenz aufweist. Die Struktur aus Membranen und Corti-Organ zeigt an der Basis Resonanz

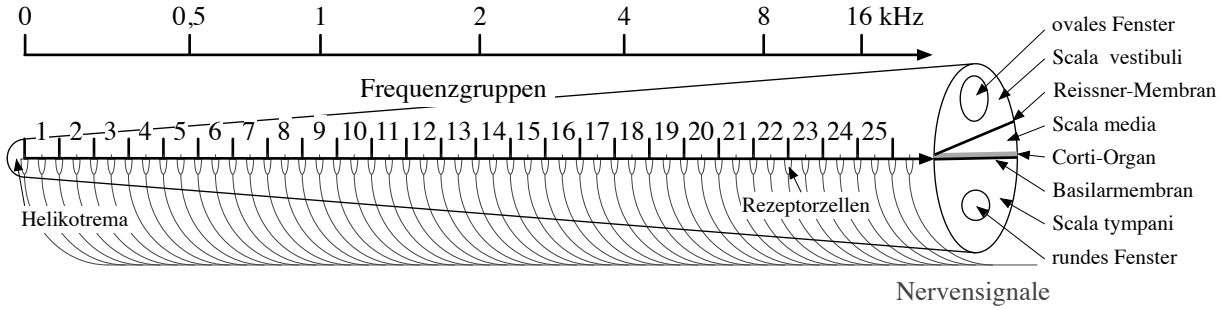


**Abbildung 2.2.:** Schematische Darstellung der abgerollten menschlichen Cochlea. (nach [Eic08, Esk97, EH93])

für hohe Frequenzen. Zur Spitze hin verliert sie an Steifheit und Dicke. Die Resonanz verschiebt sich zu tiefen Frequenzen.

### 2.1.2. Hörfläche

Das menschliche Ohr kann Schall nur in einem Frequenzspektrum von ca. 20 Hz bis 16 kHz wahrnehmen. Die obere Grenze kann mit zunehmenden Alter auf etwa 10 kHz absinken. Um ein Schallereignis innerhalb dieses Frequenzbereiches wahrzunehmen, muss dieses über einen Mindestschalldruck verfügen. Diese Schwelle, auch Hörschwelle genannt, ist stark frequenz-

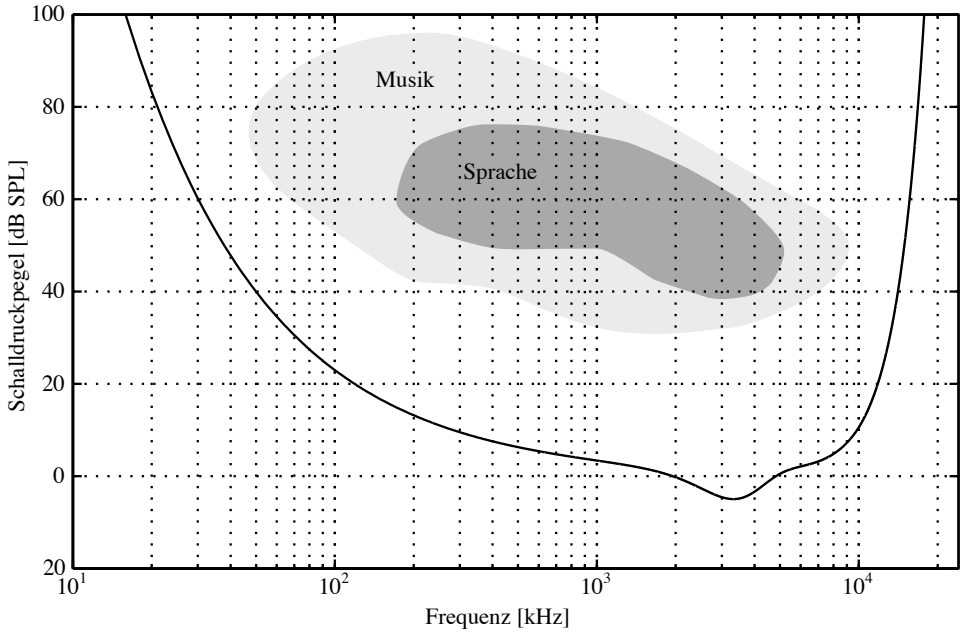


**Abbildung 2.3.:** Hörfläche des menschlichen Gehörs. [Zwi82]

abhängig. Das Minimum der Hörschwelle, folglich der Bereich der höchsten Empfindlichkeit, liegt bei etwa 3,3 kHz. Im Bereich oberhalb von 10 kHz nimmt die Empfindlichkeit rapide ab. Eine Approximation der Hörschwelle [Ter79] für junge Hörer ohne Hörschäden ist durch Gleichung 2.1 bestimmt.

$$T(f) = 3,64\left(\frac{f}{1000}\right)^{-0,8} - 6,5e^{0.6\left(\frac{f}{1000}-3,3\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4 \qquad (dB) \qquad (2.1)$$

Die Hörschwelle sowie der Haupthörbereich für Sprache und Musik sind in der Abbildung 2.3 dargestellt. Der Frequenzbereich für Sprache liegt zwischen 100 Hz und 10 kHz [LSS+07, S.51].

### 2.1.3. Frequenzgruppen

Für jede Frequenz ist ein minimaler Frequenzbereich zu beobachten, in dem es zu Interferenzen mit anderen Signalen kommt. Innerhalb dieses Bereiches werden alle Reize eines Schallereignisse vom menschlichen Gehör zusammen ausgewertet. Dieser Frequenzbereich wird als Frequenzgruppe (engl. *critical band*) [ZF99] bezeichnet. Die Bandbreite $\Delta f_G$ der Frequenzgruppen ist abhängig von ihrer Mittenfrequenz $f_0$. Unter einer Mittenfrequenz von 500 Hz liegt die Bandbreite $\Delta f_G$ konstant bei etwa 100 Hz. Ab einer Mittenfrequenz von 500 Hz wächst die

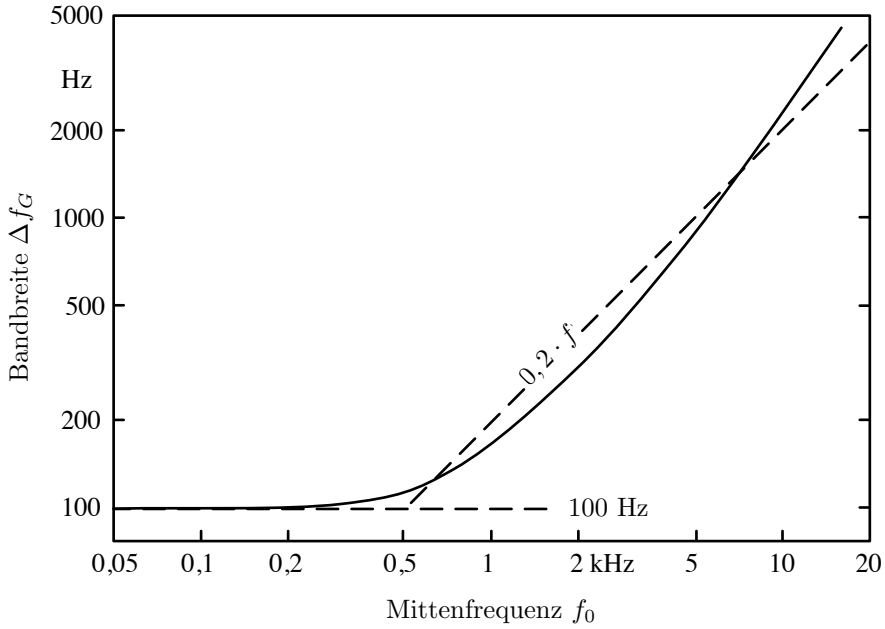


**Abbildung 2.4.:** Verlauf der Bandbreite $\Delta f_G$ in Abhängigkeit der Mittenfrequenz $f_0$ einer Frequenzgruppe. (nach [Zwi82])

**Tabelle 2.1.:** Idealisierte Einteilung der Frequenzgruppen ([PS00, ZF99])

| Nr. | Mittenfrequenz $f_0$ [Hz] | Frequenzbereich [Hz] | Bandbreite $\Delta f_G$ [Hz] |
|---|---|---|---|
| 1 | 50 | 0-100 | 100 |
| 2 | 150 | 100-200 | 100 |
| 3 | 250 | 200-300 | 100 |
| 4 | 350 | 300-400 | 100 |
| 5 | 450 | 400-510 | 110 |
| 6 | 570 | 510-630 | 120 |
| 7 | 700 | 630-770 | 140 |
| 8 | 840 | 770-920 | 150 |
| 9 | 1000 | 920-1080 | 160 |
| 10 | 1175 | 1080-1270 | 190 |
| 11 | 1370 | 1270-1480 | 210 |
| 12 | 1600 | 1480-1720 | 240 |
| 13 | 1850 | 1720-2000 | 280 |
| 14 | 2150 | 2000-2320 | 320 |
| 15 | 2500 | 2320-2700 | 380 |
| 16 | 2900 | 2700-3150 | 450 |
| 17 | 3400 | 3150-3700 | 550 |
| 18 | 4000 | 3700-4400 | 700 |
| 19 | 4800 | 4400-5300 | 900 |
| 20 | 5800 | 5300-6400 | 1100 |
| 21 | 7000 | 6400-7700 | 1300 |
| 22 | 8500 | 7700-9500 | 1800 |
| 23 | 10,500 | 9500-12000 | 2500 |
| 24 | 13,500 | 12000-15500 | 3500 |
| 25 | 19,500 | 15500-(24000) | 8500 |

Bandbreite $\Delta f_G$ proportional mit dem Faktor von $0,2$ zur Mittenfrequenz $f_0$ an. Der frequenzabhängige Verlauf der Bandbreite $\Delta f_G$ ist in Abbildung 2.4 dargestellt. Durch die Aneinanderreihung von sich nicht überlappenden Frequenzgruppen lässt sich die Frequenzskala für die Tonheit $z$ in Einheiten von Bark[1] (engl. *critical band rate*) konstruieren [ZF99]. Für einen Frequenzbereich bis 48 kHz wird diese durch 25 Frequenzgruppen gebildet. Jede Frequenzgruppe umfasst 1 Bark. Eine idealisierte Frequenzgruppeneinteilung ist in der Tabelle 2.1 aufgeführt. Der Zusammenhang zwischen der Tonheit $z$ in Bark und der Frequenz in Hz ergibt sich nach (2.2).

$$z(f) = 13\ arctan(0.76 \times 10^{-3} f) + 3,5\ arctan\left[\left(\frac{f}{7,5 \times 10^3}\right)^2\right] \quad \text{[Bark]} \tag{2.2}$$

Die Frequenzgruppen haben in der Psychoakustik eine große Bedeutung bei der Modellierung des Systems der menschlichen Wahrnehmung von Schall. Das Prinzip der Frequenzgruppen ist

[1] Benannt nach dem deutschen Physiker Heinrich Georg Barkhausen.

Ursache bzw. wesentlicher Bestandteil vieler psychoakustischer Vorgänge, wie die Wahrnehmung von Tonhöhen, Lautheit und Verdeckungseffekten.

### Wahrgenommene Tonhöhe

Im Gegensatz zur objektiv messbaren physikalischen Frequenz eines Tones ist dessen wahrnehmbare Tonhöhe eine subjektiv wahrgenommene Größe. Die psychoakustische Größe für die wahrgenommene Tonhöhe ist die Tonheit $z$, gemessen in Mel bzw. Bark. Eine Verdopplung der Tonheit bedeutet eine Verdopplung der wahrgenommenen Tonhöhe. Der Zusammenhang zwischen der Frequenz eines Tones und seiner Tonheit ist nichtlinear. Die Skala für die Tonheit in Mel geht auf die Untersuchungen von Stevens et.al. [SVN37] zurück. Zwicker [Zwi82] definierte später die Tonheit auf Basis der Bark-Skala. Hier entspricht 1 Bark gleich 100 mel. Als Referenzwert diente Zwicker der Ton C mit einer Frequenz $f = 131$ Hz, welcher einer Tonheit $z$ = 131 mel entspricht. Für die Mel-Skala nach Stevens et.al. entspricht eine Frequenz von $f = 1000$ Hz einer Tonheit von $z = 1000$ mel.

Sind zwei Töne gleichzeitig präsent, hängt die Anzahl der wahrgenommen Töne und die wahrgenommene Tonhöhe von der Differenz der Tonfrequenzen ab. Abbildung 2.5 zeigt die Tonempfindung am Beispiel zweier reiner Töne gleicher Amplitude mit den Frequenzen $f_1$ und $f_2$, wobei die Frequenz $f_1$ konstant gehalten wird und die Frequenz $f_2$ variiert. Besitzen beide Töne

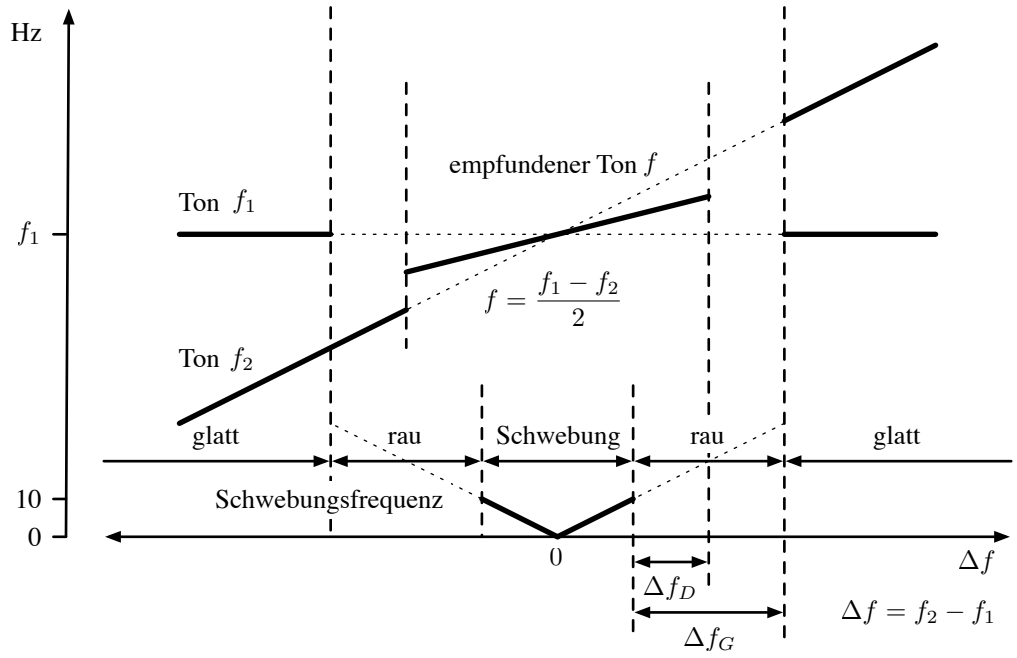


**Abbildung 2.5.:** Schematische Darstellung der Empfindung zweier gleichzeitig präsenter Töne in Abhängigkeit ihres Frequenzabstandes. (nach [Roe00])

die gleiche Frequenz, ist ein Ton mit der Frequenz $f_1 = f_2$ zu hören. Die Lautstärke (Amplitude) hängt von der Phasendifferenz der beiden Töne ab. Wird die Frequenz $f_2$ verringert bzw. erhöht, ist für geringe Frequenzunterschiede $\Delta f = f_2 - f_1$ weiterhin nur ein Ton wahrnehmbar. Es kommt zu einer sogenannten Schwebung. Die Frequenz des empfundenen Tones entspricht $f = \frac{f_1+f_2}{2}$. Seine Lautstärke ändert sich mit der Frequenz $\Delta f$, sie fängt an zu „schweben“. Bei einem Frequenzunterschied $\Delta f$ von etwa 10-15 Hz geht die Empfindung einer Schwebung in eine Rauigkeit des Tones über. Überschreitet $\Delta f$ die Frequenzunterscheidungsgrenze $\Delta f_D$ [Plo64], werden zwei Töne mit den Frequenzen $f_1$ und $f_2$ wahrgenommen. Die Empfindung der beiden Töne ist zunächst noch rau. Erreicht der Frequenzunterschied die Bandbreite $\Delta f_G$ der Frequenzgruppe, zugehörig zur mittleren Frequenz der beiden Töne mit den Frequenzen $f_1$ und $f_2$, verschwindet die Rauigkeit.

### Wahrgenommene Lautstärke

Der Schalldruckpegel (engl. *Sound Pressure Level* - SPL) $L_p$ eines Schallereignisses ist eine physikalisch messbare Größe. Die wahrgenommene Lautstärke eines Schallereignisses ist jedoch subjektiv und wird in der Psychoakustik durch den Lautstärkepegel mit der Einheit Phon und der Lautheit mit der Einheit Sone beschrieben. Der Lautstärkepegel ist ein Vergleichsmaß und basiert auf einem 1 kHz-Ton. Für diesen stimmen der Schalldruckpegel $L_P$ in Dezibel und der Lautstärkepegel $L_N$ in Phon überein. Die Skala der Lautheit entspricht der subjektiven Wahrnehmung der Lautstärke. Verdoppelt sich die wahrgenommene Lautstärke, so verdoppelt sich die Lautheit. Ein Lautstärkepegel $L_N$ eines 1 kHz-Sinuston von 40 phon dient als Referenzwert für die Lautheit $N$ von 1 sone. Oberhalb von 1 sone bzw. 40 phon entspricht eine Verdopplung der Lautheit einer Zunahme des Lautstärkepegels um 10 dB.

Die Abbildung 2.6 stellt den Schalldruckpegel $L_P$ eines 1 kHz-Sinus-Tones in Abhängigkeit

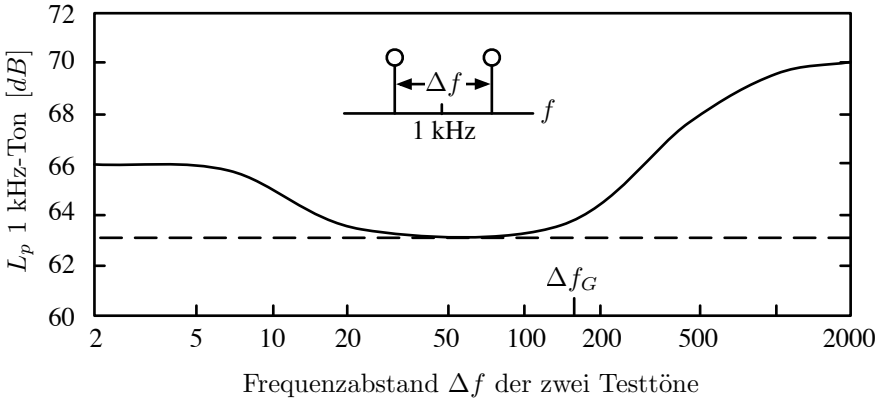


**Abbildung 2.6.:** Schematische Darstellung der empfundenen Lautstärke zweier gleichzeitig präsenter Töne in Abhängigkeit des Abstandes ihrer Frequenzen. (nach [Zwi82])

des Frequenzabstandes $\Delta f_G$ von zwei Sinus-Tönen dar. Die Empfindung der Lautstärke des 1kHz-Tones entspricht der empfundenen Lautstärke der gleichzeitig auftretenden Sinus-Töne. Diese besitzen je einen Schalldruckpegel von 60 dB und haben eine Mittenfrequenz von 1 kHz.

Die Erläuterungen zum Verlauf des Schalldruckpegels $L_p$ sind wie folgt [VHH98]. Für einen Frequenzabstand kleiner 10 Hz kommt es zu einer Schwebung. Die Amplituden der beiden Sinus-Töne werden hier vom Gehör addiert, welches einer Pegelsteigerung von +6 dB entspricht. Im Bereich eines Frequenzabstandes von 20 Hz bis zur Bandbreite der Frequenzgruppe $f_G = 160$ Hz (s. Tabelle 2.1) werden die Leistungen der beiden Sinus-Töne addiert, welches einer Pegelsteigerung von +3 dB entspricht. Steigt der Frequenzabstand weiter, werden die beiden Sinus-Töne vom Gehör separat verarbeitet. Bei einem Frequenzabstand $\Delta f = 2$ kHz bildet das Gehör offensichtlich die Lautheit für jeden der beiden Sinus-Töne und summiert die empfundenen Lautstärken auf. Eine Verdopplung der Lautheit entspricht einer Pegelsteigerung von +10 dB.

## 2.2. Kanalcodierung

Die fehlerfreie Übertragung von Informationen ist eine zentrale Forderung an Informationsübertragungssysteme. In der Praxis existiert jedoch kein ungestörter und somit fehlerfreier Übertragungskanal. Dieses gilt auch für die Informationsübertragung mit der Technologie der digitalen Wasserzeichen. Um Informationen dennoch fehlerfrei übertragen zu können, wurden Codierungsverfahren zum Schutz der Information gegenüber Übertragungsfehlern des Übertragungskanals entwickelt. Die Gesamtheit dieser Verfahren werden in der Nachrichtentechnik unter dem Begriff der Kanalcodierung geführt. Die grundlegende Funktionsweise der Kanalcodierung ist das Hinzufügen von Redundanz zu der zu übertragenden Information, um Übertragungsfehler erkennen oder sogar korrigieren zu können.

### 2.2.1. Übertragungskanal

Abbildung 2.7 zeigt das Blockschaltbild eines digitalen Übertragungssystems. Dabei werden vier große Verarbeitungsblöcke der Quelleninformation unterschieden, nämlich die Quellencodierung, Chiffrierung, Kanalcodierung und die Modulation.

- **Quellencodierung** (auch: Entropiecodierung)**:** Der Zweck der Quellencodierung ist die Reduktion der zu übertragenden Datenmenge durch Entfernung von redundanten und

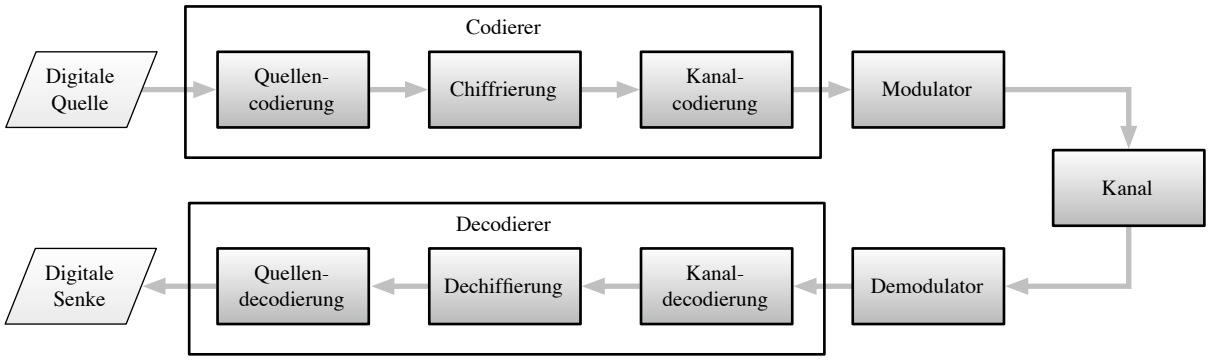


**Abbildung 2.7.:** Blockschaltbild eines digitalen Übertragungssystems.

irrelevanten Datenbestandteilen der Quelleninformation. Als redundant werden die Informationen betrachtet, welche dem Empfänger bereits bekannt sind. Irrelevant sind die Informationen, welche vom Empfänger nicht verwendet bzw. wahrgenommen werden können.

- **Chiffrierung** (auch: Kryptocodierung)**:** Aufgabe der Chiffrierung ist der Schutz der Information vor unautorisiertem Zugriff auf die Information oder deren Fälschung

- **Kanalcodierung:** Schutz der Information vor Übertragungsfehlern durch Hinzufügen von Redundanz

- **Modulation:** Anpassung des Signals an das Übertragungsmedium

### 2.2.2. Verfahren der Kanalcodierung

#### Rückwärtsfehlerkorrektur

Die Verfahren der Rückwärtsfehlerkorrektur (engl. *Automatic Repeat Request* (ARQ)) beschränken sich bei dem Schutz der Information auf die Erkennung von Übertragungsfehlern. Wird ein Fehler detektiert, erfolgt an den Sender die Anforderung einer erneuten Übertragung der Information. Dieses Verfahren setzt im Allgemeinen die Existenz eines Rückkanals voraus. Mögliche Anwendungsszenarios der Rückwärtsfehlerkorrektur ohne Rückkanal liegen vor, wenn eine Schätzung (Fehlerverschleierung) der gesendeten Information oder das Verwerfen von fehlerhaften Informationen hinreichend ist. Die zusätzliche Redundanz, welche zur Information hinzugefügt werden muss, ist im Vergleich zu Verfahren mit Vorwärtsfehlerkorrektur geringer. Eine

hohe Redundanz stellt jedoch die wiederholte Übertragung einer fehlerhaft empfangenen Information dar. Folglich kommt die Rückwärtsfehlerkorrektur vorwiegend bei Übertragungskanälen mit geringen Fehlerraten zum Einsatz.

**Tabelle 2.2.:** Beispiel eines *Even-Parity-Check*-Codes

| Informationswort | Codewort |
|---|---|
| 000 | 0000 |
| 001 | 0011 |
| 010 | 0101 |
| ... | ... |
| 111 | 1111 |

Einer der einfachsten Vertreter der Rückwärtsfehlerkorrektur ist das Paritäts-Bit. Zur Bildung eines Codewortes wird die Information um ein weiteres Paritäts-Bit erweitert. Der Wert dieses Paritäts-Bits wird durch die Anzahl der Bits mit dem Wert 1 in der Information bestimmt. Bei einer *Even-Parity* ist das Paritäts-Bit so zu wählen, dass die Information und das Paritäts-Bit zusammen eine gerade Anzahl an Bits mit dem Wert 1 besitzen. Wird eine *Odd-Parity* verwendet, so ist eine ungerade Anzahl an Bits mit dem Wert 1 herzustellen. Mittels eines Paritäts-Bits können alle Fehler erkannt werden, welche die Parität des Codewortes verändern. Dieses sind alle Bitfehler mit ungerader Anzahl.

### Vorwärtsfehlerkorrektur

Verfahren der Vorwärtsfehlerkorrektur (engl. *Forward Error Correction* (FEC)) ermöglichen die Korrektur von Fehlern ohne Informationen vom Sender anzufordern. Ein Rückkanal, wie bei den Verfahren der Rückwärtsfehlerkorrektur, ist somit nicht notwendig. Die zusätzliche Redundanz ist bei geringer Anzahl an Übertragungsfehlern höher im Vergleich zur Rückwärtsfehlerkorrektur. Jedoch ist der Datendurchsatz, also die Menge der übertragenen Information pro Zeit, konstant. Bei wiederholter Sendung der Information reduziert sich der Datendurchsatz und schwankt somit bei Verfahren mit Rückwärtsfehlerkorrektur. Während bei der Rückwärtsfehlerkorrektur die Information beliebig oft angefordert werden kann, verbleibt für die Vorwärtsfehlerkorrektur ein Restfehler, sobald die Fehleranzahl der Übertragung die Korrekturfähigkeit des Codes übersteigt und die Information nicht mehr fehlerfrei rekonstruiert wird.

**Tabelle 2.3.:** Beispiel eines Wiederholungscodes mit zwei Wiederholungen

| Informationsbit | Codewort |
|---|---|
| 0 | 000 |
| 1 | 111 |

Eine Vorwärtsfehlerkorrektur kann mit einem Wiederholungscode realisiert werden. Hierbei wird jedes Bit einer Nachricht wiederholt übertragen. Liegt während der Übertragung der Information ein Fehler vor, so können die Bits der Nachricht aus einem Mehrheitsentscheid der jeweils zusammengehörenden Bits rekonstruiert werden.

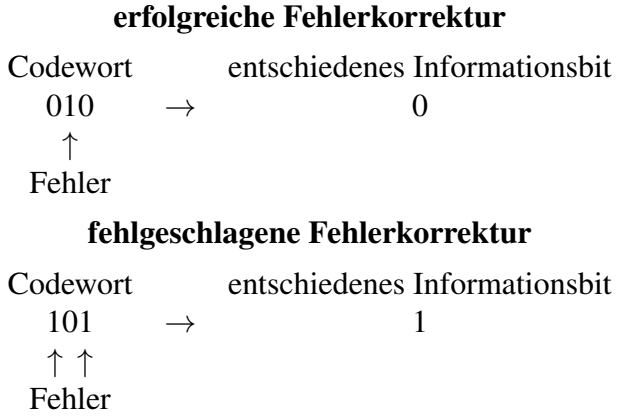


**Abbildung 2.8.:** Decodierung von fehlerhaften Codewörtern für das Beispiel des Wiederholungscodes

### 2.2.3. Faltungscodes

Faltungscodes sind Verfahren zur Vorwärtsfehlerkorrektur. Diese stellen in der Kanalcodierung die zweite wesentliche Codeklasse neben den Blockcodes dar. Wie bei der Codierung mit einem Blockcode, wird auch bei der Codierung mit einem Faltungscode eine Informationssequenz $\mathbf{u}$ in Blöcke $\mathbf{u_i}$ von je $k$ Bits unterteilt. Die Blöcke $\mathbf{u_i}$ werden mit dem Codierer auf Codeblöcke $\mathbf{v_i}$ von je $n$ Bits unterteilt. Das Verhältnis (2.3) der Abbildung von $k$ Informationsbits auf $n$ Codebits wird als Coderate $R$ bezeichnet. Bei einem Blockcode werden die einzelnen Blöcke $\mathbf{u_i}$ unabhängig voneinander codiert. Die Codierung von zwei identischen Informationsblöcken liefert auch zwei identische Codeblöcke. Bei der Faltungscodierung ist ein Codeblock $\mathbf{v_i}$ nicht nur von dem entsprechenden Informationsblock $\mathbf{u_i}$ abhängig, sondern auch noch von $m$ vorhergehenden Informationsblöcken. Man spricht hierbei von einem Codierer mit Gedächtnis, welcher eine Gedächtnisordnung (engl. *memory*) von $m$ besitzt.

$$R = \frac{k}{n} \tag{2.3}$$

$$\mathbf{u} = ((u_1^{(1)} u_1^{(2)} \dots u_1^{(k)}), (u_2^{(1)} u_2^{(2)} \dots u_2^{(k)}), \dots) = (\mathbf{u_1}, \mathbf{u_2}, \dots) \tag{2.4}$$

$$\mathbf{v} = ((v_1^{(1)} v_1^{(2)} \dots v_1^{(n)}), (v_2^{(1)} v_2^{(2)} \dots v_2^{(n)}), \dots) = (\mathbf{v_1}, \mathbf{v_2}, \dots) \tag{2.5}$$

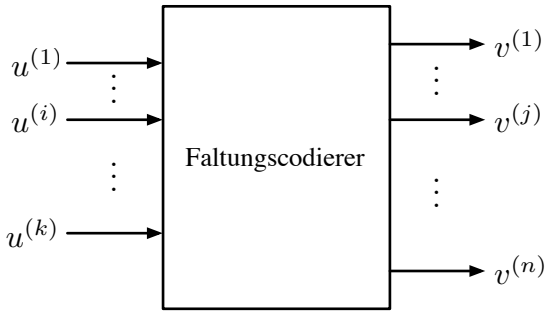


**Abbildung 2.9.:** Faltungscodierer als LTI-System [Bos98].

Die Bezeichnung Faltungscode stammt aus der Beschreibung eines Faltungscodierers als zeitdiskretes lineares zeitinvariantes System (engl. *Linear Time-Invariant* (LTI-) *System*). Die Abbildung 2.9 zeigt einen Faltungscodierer als LTI-System mit Belegung eines *k*-dimensionalen Eingangs mit den Eingangssequenzen aus den Elementen der Informationsblöcke $\mathbf{u_i}$ und einem *n*-dimensionalen Ausgang mit Abgang der Ausgangssequenzen aus den Elementen der Codeblöcke $\mathbf{v_i}$. Die Codesequenz $\mathbf{u^{(j)}}$ ergibt sich durch die mathematische Faltung der Eingangssequenzen mit der Impulsantwort $g^{(j)}$ des LTI-Systems.

$$v^{(j)} = u^{(1)} \cdot g_1^{(j)} + u^{(2)} \cdot g_2^{(j)} + \cdots + u^{(k)} \cdot g_k^{(j)} = \sum_{i=1}^{k} u^{(i)} \cdot g_i^{(j)} \tag{2.6}$$

Ein Faltungscode lässt sich durch die Impulsantworten (auch: Generatorsequenzen oder Generatoren genannt) seines LTI-Systems beschreiben. Diese sind theoretisch unendlich lang, jedoch ist für eine Betrachtung nur der von Null verschiedene, endlich lange Teil am Anfang der Impulsantworten relevant. Für die folgenden Betrachtungen wird ein Faltungscode mit den Generatoren $\mathbf{g^1} = (1,1,1) = 7_8$ und $\mathbf{g^2} = (1,0,1) = 5_8$ verwendet.

#### 2.2.3.1. Codierung

Die Funktionsweise eines Faltungscodierers lässt sich in Form eines Schieberegisters realisieren. In Abbildung 2.10 ist ein exemplarisches Beispiel eines Faltungscodierers in Form eines Schieberegisters dargestellt. Das Beispiel beruht auf den zuvor benannten Generatoren. Die Speicherelemente, also das Gedächtnis, des Registers sind mit *D* bezeichnet. Die Verknüpfungen des Einganges und der Speicherelemente ergeben sich aus den Generatoren des Faltungscodes. Die Ausgangssequenz wird durch das Durchschalten aller Ausgänge des Schieberegisters mit jedem Eingangssymbol aufgebaut. Die Symbolfrequenz am Ausgang des Schieberegisters ergibt sich aus der Coderate des Faltungscodes und ist bei dem ausgewählten Beispiel mit R=1/2 doppelt so hoch wie die Symbolfrequenz am Eingang des Schieberegisters.

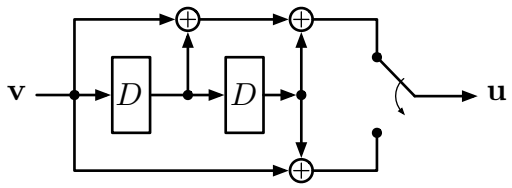


**Abbildung 2.10.:** Realisierung eines Faltungscodierers im Form eines Schieberegisters [Bos98].

#### 2.2.3.2. Graphische Darstellung

### Zustandsdiagramm

Ausgehend vom Gedächtnis eines Faltungscodierers bzw. der Anzahl der Speicherelemente des Schieberegisters kann dieser $2^m$ verschiedene Zustände annehmen. Ein Faltungscodierer kann somit auch als endlicher Zustandsautomat beschrieben werden. Die Darstellung von Zustandsautomaten erfolgt über ein Zustandsdiagramm. Dieses besteht aus Knoten, welche die Zustände des Automaten repräsentieren. Die Pfeile zwischen den Knoten repräsentieren einen Zustandswechsel in Abhängigkeit der Eingangswerte $\mathbf{u_i}$ des Automaten und der daraus resultierenden Ausgabewerte $\mathbf{v_i}$. Die Abbildung 2.11 zeigt das Zustandsdiagramm des beispielhaften Faltungscodierers.

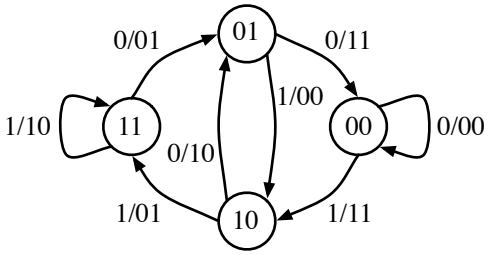


**Abbildung 2.11.:** Zustandsdiagramm des in Abbildung 2.10 beschriebenen Faltungscodierers.

### Trellis-Diagramm

Diese Darstellung des Zustandsdiagramms nimmt keinen Bezug auf die zeitliche Abfolge der Zustände des Zustandsautomaten. Ist dies von Interesse, wird ein Trellis-Diagramm verwendet. Ein Trellis-Diagramm sind aneinandergereihte Zustandsdiagramme, wobei ein Zustandsdiagramm keine Übergänge in sich selbst, sondern immer nur auf das nächste Zustandsdia-

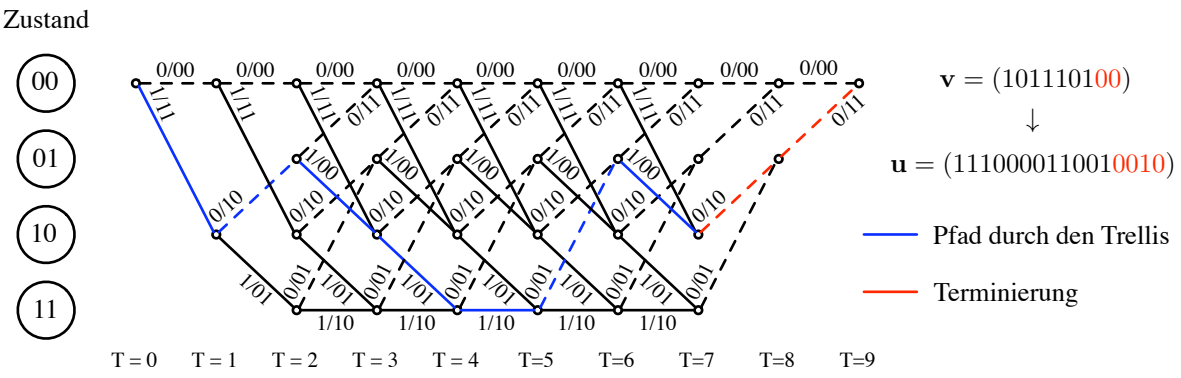


**Abbildung 2.12.:** Trellis-Diagramm des in Abbildung 2.10 beschriebenen Faltungscodierers.

gramm besitzt. Jedes Zustandsdiagramm repräsentiert einen bestimmten Zeitpunkt T des Codierungsvorganges. Die Abbildung 2.12 zeigt das Trellis-Diagramm für den in Abbildung 2.10 beschriebenen Faltungscodierer. Der Codierer startet zum Zeitpunkt T = 0 mit einem Ausgangszustand, typischerweise (00). In Abhängigkeit des Eingangswertes $u_1$ können zwei verschiedene Zustände angenommen werden. Nach einer Anzahl vom $m$ Eingabewerten, welche dem Gedächtnis des Faltungscodierers entspricht, können alle möglichen Zustände erreicht werden. Ab hier ist jedes Segment des Trellis-Diagramms identisch. Das Überführen des Faltungscodieres in seinen Ausgangszustand wird als Terminierung bezeichnet. Hierbei wird solange eine „0“, also eine bekannte Information, übertragen bis der Ausgangszustand (00) erreicht ist. Durch das Gedächtnis hat ein Eingangsbit Einfluss auf den aktuellen Ausgabewert und den folgenden $m$ Ausgabewerten. Dieses gilt auch im Umkehrschluss. Für die Decodierung einer Codesequenz ist die Terminierung nicht zwingend erforderlich, jedoch sinken die Korrekturfähigkeiten des Faltungsdecodierers für die letzten Informationsbits, für die die Einflusslänge nicht beachtet wurde. Bei der Berechnung der Coderate $R$ ergibt sich bei einer Terminierung die Notwendigkeit von $m$ zusätzlich zu codierenden Bits. Für lange Informationssequenzen ist der Einfluss der Terminierung auf die Coderate zu vernachlässigen.

$$R = \frac{k+m}{n} = \frac{k}{n} + \frac{m}{n} \tag{2.7}$$

#### 2.2.3.3. Decodierung

Für die Nachrichtentechnik sind Faltungscodes von besonderem Interesse, da deren Decodierungsverfahren die Verwendung von „weichen“ noch nicht auf binäre Werte entschiedene (engl. *soft-input*) Codewortsequenzen erlauben. Außerdem existieren Verfahren, die eine sehr effizi-

ente Decodierung ermöglichen. Der bekannteste Vertreter ist der von Andrew J. Viterbi vorgestellte Viterbi-Algorithmus [Vit67]. Die Decodierung eines Faltungscodes lässt sich am Beispiel des Trellis-Digramm erläutern. In Abhängigkeit einer bestimmten Eingangssequenz **u** wird ein Pfad durch das Trellis-Diagramm und somit eine Codesequenz **v** erzeugt. Eine Codesequenz **v** und der beschrittene Pfad sind eineindeutig für jede Eingangssequenz **u**. Bei der Decodierung kann anhand einer empfangenen Codesequenz ein Pfad durch das Trellis-Diagramm und somit die ursprüngliche Eingangssequenz **u** bestimmt werden. Durch die möglichen Übergänge zwischen den Zuständen des Faltungscodierers ist auch die Anzahl der möglichen Pfade begrenzt. Ist eine empfangenen Codesequenz fehlerhaft, so entsteht ein ungültiger Pfad durch das Trellis-Diagramm. Bei der Decodierung werden alle möglichen Pfade durch das Trellis-Diagramm bestimmt. Der Pfad, welcher den geringsten Abstand zu dem durch die empfangenen Codesequenz erzeugten Pfad besitzt, wird zur Bestimmung der ursprünglichen gesendeten Informationssequenz **u** verwendet. Eine fehlerhafte Codesequenz, welche einen gültigen Pfad durch das Trellis-Diagramm erzeugt, kann nicht korrigiert werden.

## 2.3. Digitale Wasserzeichen

Digitale Wasserzeichen sind zusätzliche Informationen, die nicht-wahrnehmbar in digitale Medien (z.B. Audio, Bilder, Video) eingebettet sind. Im Gegensatz zu Metadaten, welche neben dem Dateninhalt in eigenständigen Bereichen einer Datei untergebracht sind, werden digitale Wasserzeichen direkt mit dem Dateninhalt verbunden. Hierin liegt die große Stärke der digitalen Wasserzeichen. Unabhängig vom Datenformat können zusätzliche Informationen in digitale Medien eingebettet werden. Das Einfügen von digitalen Wasserzeichen nimmt keine Änderungen am Datenformat vor. Die Daten können weiterhin wie die originalen Daten ohne Einschränkungen verwendet werden. Wird das Wasserzeichen stark genug an den Dateninhalt gebunden, gehen die eingebettete Informationen auch bei Datenformatkonvertierung nicht verloren.

Neben den digitalen Wasserzeichen beschäftigen sich eine Vielzahl weiterer Technologien mit der Einbettung von zusätzlichen Informationen. Die Einordnung und Abgrenzung der Technologie der digitalen Wasserzeichen in diesem Themengebiet wird in der Literatur unterschiedlich vorgenommen. In diesem Zusammenhang ist häufig eine unterschiedliche Auffassung der Begriffe „*Watermarking*“[2], „*Steganography*“ und „*Information (Data) Hiding*“ zu finden. Cox et al. [CMB02] unterteilen *Information Hiding*, das Einfügen von Informationen in Trägerdaten, in die Disziplinen *Stenographic Watermarking*, *Non-stenographic Watermarking*, *Covert Communication* und *Overt Embedded Communications*. Nach Katzenbeisser [Kat00] werden *Covert*

[2] Der englische Begriff für das Einbetten von digitalen Wasserzeichen.

*channels*, *Anonymity*, *Steganography* und *Watermarking* als Hauptdisziplinen des *Information Hiding* eingeordnet. Pan et al. [PHJ04] im Gegensatz hierzu definiert *Information Hiding*, *Robust Watermarking*, *Semi-Fragil Watermarking* und *Fragil Watermarking* als Disziplinen der *Steganography*. Zusammenfassend kann man sagen, dass digitale Wasserzeichen in weitgehender Übereinstimmung darüber definiert werden, dass die Wasserzeicheninformation in Beziehung mit dem Träger steht und deren Existenz nicht verborgen ist.

### 2.3.1. Systemkonzept

Das grundlegende Systemkonzept der digitalen Wasserzeichen (siehe Abbildung 2.13) besteht aus zwei wesentlichen Komponenten, dem Einbettungs- und dem Detektionsprozess. Im Einbettungsprozess erfolgt die Verknüpfung der Wasserzeicheninformation mit dem digitalen Trägermedium. Im Detektionsprozess wird die Wasserzeicheninformation aus dem markierten Träger ausgelesen bzw. die Existenz eines Wasserzeichens detektiert. Das System der digitalen Wasserzeichen kann als Form der Informationsübertragung angesehen werden. In Anlehnung an die Informationstechnik wird für das Systemkonzept der digitalen Wasserzeichen häufig das Modell eines digitalen Übertragungssystem verwendet. Die Wasserzeicheninformation stellt hierbei die Informationsquelle dar. Das Trägermedium wird als Übertragungskanal betrachtet. Der Einbettungs- und Detektionsprozess werden als Sender bzw. Empfänger modelliert. Alle Verarbeitungsschritte des markierten Trägermedium vor der Wasserzeichendetektion werden als Störungen des Übertragungskanals betrachtet. Neben dem Übertragungskanal in Form des Trägermedium wird teilweise ein zusätzlicher Seitenkanal verwendet.

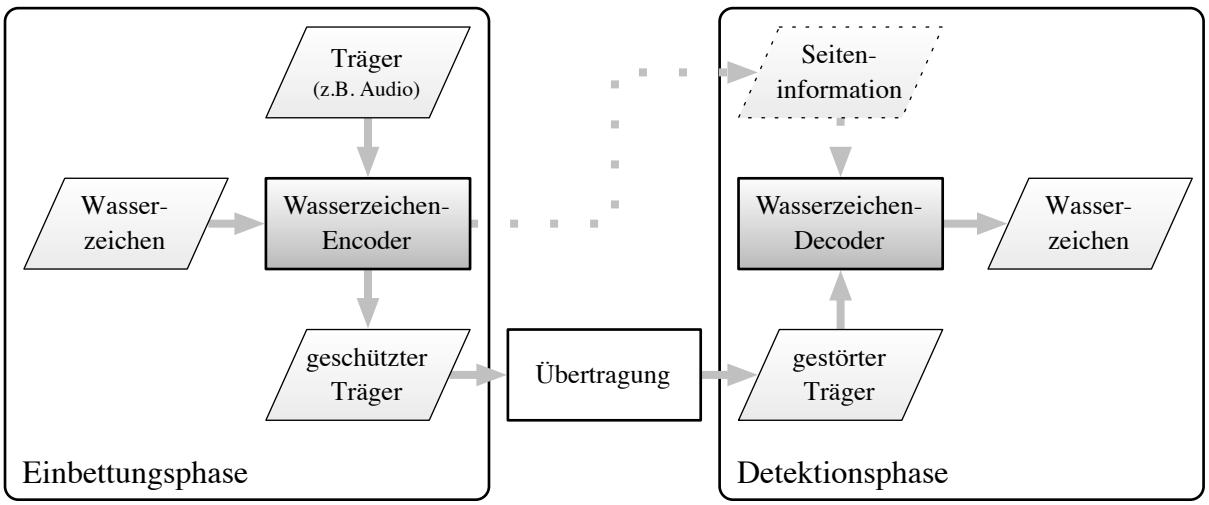


**Abbildung 2.13.:** Genereller Ablaufplan des Systemkonzepts der digitalen Wasserzeichen

### 2.3.2. Eigenschaften von digitalen Wasserzeichen

#### Kapazität

Die Kapazität eines Wasserzeichens bestimmt die Datenmenge, welche in das Trägermedium eingebettet werden kann. Die Angabe der Kapazität erfolgt im Allgemeinen in Bit. Die Bezugsgröße ist abhängig von dem jeweiligen Trägermedium. Bei Audio-Wasserzeichen sind Bit pro Sekunde oder Bit pro Sample gängige Angaben. Anhand der Kapazität werden Techniken unterschieden, die nur die Präsenz des Wasserzeichens im Trägermedium nachweisen (*zero-bit watermark* [CMB02]), Techniken, die eine konstante Kapazität besitzen und Techniken, die inhaltsabhängig eine variable Kapazität erzielen.

#### Transparenz

Das Einbringen einer Wasserzeicheninformation erfolgt über gezielte Veränderungen der Trägerdaten. Die Transparenz (auch Qualität oder Nicht-Wahrnehmbarkeit) beschreibt den Einfluss der Wasserzeicheneinbettung auf die Qualität der Trägerdaten.

#### Robustheit

Die Robustheit eines Wasserzeichens ist die Widerstandsfähigkeit gegenüber Veränderungen der Trägerdaten. Sie bestimmt die Art und Stärke der Operationen, nach denen das Wasserzeichen noch erfolgreich ausgelesen bzw. detektiert werden kann. Im Idealfall bleibt das Wasserzeichen solange erhalten, wie der Dateninhalt der Trägerdaten nicht zerstört bzw. unbrauchbar geworden ist. Die Anforderungen an die Wasserzeichenrobustheit sind anwendungsabhängig. Ein weiterer Aspekt der Wasserzeichenrobustheit ist eine gewollte Zerbrechlichkeit des Wasserzeichens. Diese ist vorwiegend bei Wasserzeichenverfahren zur Überprüfung der Echtheit von Daten anzufinden.

#### Komplexität

Die Komplexität beschreibt die für die Einbettung bzw. für die Detektion eines Wasserzeichens notwendigen Ressourcen und Eingabedaten. Der Ressourcenbedarf (Rechenzeit, Speicher, etc.) ist bedeutend bei ressourcenarmen Umgebungen (z.B. mobile Anwendungen) oder bei Echtzeitanwendungen (z.B. *Media-Streaming*). Im Detektionsprozess werden neben den markierten Trägerdaten in Abhängigkeit der Wasserzeichentechnik weitere Eingabedaten benötigt. Es wird zwischen Techniken unterschieden, welche zusätzlich zu den markierten Trägerdaten die originalen Trägerdaten verwenden (nicht-blinde Techniken, engl. *non-blind*, *non-oblivious*) und Techniken, die ohne die originalen Trägerdaten auskommen (blinde Techniken, engl. *oblivious*).

Techniken, die ohne die originalen Trägerdaten auskommen, jedoch die originale Wasserzeicheninformation benötigen, werden teilweise als semi-blind (engl. *semi-blind*) [Kat00] bezeichnet. Die Übertragung der zusätzlich verwendeten Eingabedaten erfolgt über einen separaten Seitenkanal (vgl. Abbildung 2.13).

#### Sicherheit

Der Aspekt der Sicherheit bezieht sich auf gezielte Angriffe unter Kenntnis des Wasserzeichenalgorithmus. Die Kenntnis über geheime Schlüssel wird hier jedoch ausgeschlossen. Die Sicherheit einer Wasserzeichentechnik ist durch folgende Fragen zu bewerten. Ist es möglich, die Wasserzeicheninformation zu fälschen oder nach Belieben zu ändern? Ist es möglich, das Wasserzeichen ohne Minderung der Trägerdatenqualität zu entfernen?

Die Sicherheit von digitalen Wasserzeichen wird in den überwiegenden Anwendungen durch eine oder eine Kombination der folgenden Maßnahmen erzeugt. Sie ist jedoch nicht hierauf beschränkt.

- Die Selektion erfolgt auf Basis eines geheimen Schlüssels. Digitale Wasserzeichen sind symmetrische Systeme. Bei der Verarbeitung der Trägerdaten müssen vom Einbettungs- und Detektionsprozess dieselben Datenbestandteile selektiert werden.
- Die Wasserzeicheninformation wird auf Basis eines geheimen Schlüssel generiert. Semiblinde Techniken benötigen die Wasserzeicheninformation im Detektionsprozess (siehe Komplexität).
- Die Wasserzeicheninformation wird verschlüsselt.

### 2.3.3. Anwendungsgebiete

Die Eigenschaften der digitalen Wasserzeichen können in den unterschiedlichsten Anwendungsszenarien zum Einsatz gebracht werden. In einer groben Einteilung lassen sich drei große Bereiche festlegen. Die Kontrolle und Durchsetzung von Urheberrechten, die Erweiterung von Funktionalitäten und die Überprüfung der Echtheit von Daten. Aufgrund der Vielzahl an Anwendungsbereichen und deren Problemstellungen werden in der Literatur [CMB02], [Dit00], [Kat00] und Industrie [Dig] weiterführende Einteilungen vorgenommen.

#### Kontrolle und Durchsetzung von Urheberrechten

Dieses Anwendungsgebiet stellt das vielseitigste und größte Gebiet der digitalen Wasserzeichen dar. Die Musik- und Filmindustrie sucht beständig nach Möglichkeiten, die missbräuchliche

Nutzung ihrer Produkte zu unterbinden. Die Technologie der digitalen Wasserzeichen zeigte sich hier als ein wertvolles Instrument. Aufgrund des hohen kommerziellen Interesses wurden die Einsatzmöglichkeiten von digitalen Wasserzeichen ausgiebig erforscht und eine Vielzahl an Applikationen entwickelt.

Digitale Wasserzeichen werden eingesetzt, um Urheberrechte von Daten nachzuweisen (*Robust Authentication Watermark, Proof of Ownership*). Hierzu werden eindeutige Authentifizierungsmerkmale im Form eines Wasserzeichen in die Daten eingebettet. Ein ständiges Problem der Industrie ist die unerlaubte Verbreitung und Wertschöpfung ihrer Werke. Vor der Auslieferung eines Werkes kann dieses mit einem Wasserzeichen versehen werden. Wird für jeden Kunden ein anderes eindeutiges Wasserzeichen in das Werk eingebracht (*Fingerprinting*), können illegale Kopien zurückverfolgt werden (*Traitor Tracing*).

Die genannten Szenarien stellen nur einen Teil der Anwendungen dar. Die Wasserzeichentechniken, die in diesem Anwendungsgebiet eingesetzt werden, erfordern eine hohe Robustheit und Sicherheit. Typische Verarbeitungsschritte der Trägerdaten, die das Wasserzeichen überstehen muss, sind bei illegalen Kopien u.a. starke verlustbehaftete Kompression, Formatkonvertierung, Skalierung, *Cropping* und Digital-Analog/Analog-Digital-Wandlung (DA/AD-Wandlung). Ein Entfernen oder Fälschen des Wasserzeichen muss verhindert werden.

#### Erweiterung von Funktionalitäten

Bei der Erweiterung der Funktionalität von bestehenden Systemen ist die Abwärtskompatibilität ein großes Problem. Mit digitalen Wasserzeichen können zusätzliche Informationen und somit auch Funktionalitäten in die Trägerdaten integriert (*Caption Watermark, Annotation Watermark*) werden, ohne Änderung des Datenformats der Trägerdaten. Die Daten bleiben für ältere Systeme (engl. *legacy devices*) verarbeitbar. Das Anwendungsgebiet wird in Bezug hierauf auch als *Legacy Channel* bezeichnet. Die Wasserzeichen enthalten z.B. Beschreibungen des Datenmaterials, Aufnahmeinformationen wie Ort, Zeit oder Person, aber auch Steuerinformationen für Geräte.

Die Integration der Informationen fordert hohe Wasserzeichenkapazitäten. Die Anforderungen an die Sicherheit ist in diesem Anwendungsgebiet eher gering. Die Nutzungsszenarien zeigen kein oder nur geringes Interesse zur Entfernung oder Fälschung des Wasserzeichens. Die Anforderungen an die Robustheit der eingesetzten Wasserzeichentechniken schwankt sehr stark. Einige Anwendungsszenarien sehen keine oder nur geringe Veränderungen der Trägerdaten vor. Soll die Wasserzeicheninformation z.B. auch bei entwickelten Bildern ausgelesen werden können, erfordern Verarbeitungsschritte wie DA/AD-Wandlung, Rotation, Verzerrung eine hohe Robustheit des Wasserzeichens.

**Verifikation der Unversehrtheit/Echtheit von Daten**

Dieses Anwendungsgebiet stellt im Vergleich zu den Vorhergenannten neue Anforderungen an die digitalen Wasserzeichen. Ist für die zuvor genannten Anwendungsbereiche eine möglichst robuste Wasserzeicheneinbettung von Interesse, werden hier Wasserzeichen mit gezielter Zerbrechlichkeit, sogenannte fragile Wasserzeichen, gefordert. Wasserzeichen gehen einen Verbund mit den Trägerdaten ein. Veränderungen der Trägerdaten wirken sich somit auch auf das eingebettete Wasserzeichen aus. Wasserzeichen können so gestaltet werden, dass Schäden des Wasserzeichen Rückschlüsse auf die Existenz, den Umfang oder auch die Art der Veränderungen der Trägerdaten erlauben.

Dieses Anwendungsgebiet stellt die höchsten Anforderungen an die Kapazität von Wasserzeichen. Der Aspekt der Sicherheit konzentriert sich hier auf den Schutz vor Fälschungen.

Die fragilen Wasserzeichen stellen die ersten Ausprägungen der Wasserzeichenverfahren zum Nachweis der Echtheit von medialen Daten dar. Der Integritätsnachweis ist jedoch nicht auf zerbrechliche Wasserzeichen beschränkt. Die unterschiedlichen Wasserzeichenverfahren zur Verifizierung der Integrität medialer Daten werden im folgenden Kapitel dargelegt.

Kapitel 3

# Wasserzeichenverfahren zur Verifizierung der Integrität medialer Daten

*Dieses Kapitel gibt einen Überblick über Wasserzeichenverfahren zur Verifizierung der Integrität medialer Daten. Die allgemeinen Eigenschaften und Klassifikationen für diese Verfahren werden dargelegt. Das in dieser Arbeit entwickelte Verfahren zählt zu den inhalts-fragilen Wasserzeichenverfahren. Zur Bestimmung des Forschungsbedarfs werden inhalts-fragile Verfahren für Audiodaten vorgestellt und verglichen.*

## 3.1. Einleitung

Unter dem Begriff mediale Daten werden hier im wesentlichen Kommunikations- und Informationsmittel in Form von Schrift, Bild und Ton verstanden. Mediale Daten, besonders Audio, Bild und Video, nehmen bei der täglichen Informationsbeschaffung eine wichtige Rolle ein. Klassische Informationsquellen, wie Zeitung, Rundfunk und Fernsehen, besitzen ein hohes Maß an Vertrauen. Das Vertrauen entsteht dadurch, dass die Quelle der Information bekannt und etabliert ist und Publikationen einen redaktionellen Anspruch erfüllen. Es gibt wenige große Agenturen und Massenmedien. Die Verbreitung erfolgt über zentrale Verbreitungsstellen (Satellit, Kabel, Terrestrik, Druck). Ein Austausch, Erweiterung oder Veränderung der Informationen auf Zwischenwegen ist unwahrscheinlich.

In den letzten Jahren gewinnt das Internet als Informationsquelle gegenüber den klassischen Medien stark an Bedeutung [Sch10, Eck11]. Besonders für jüngere Nutzer besitzt das „Internet das größte Gewicht für die Informations- und Meinungsbildung“ [Eck10]. Ein wichtiger Aspekt hierbei ist die multimediale Präsentation und Interaktivität von Informationen. Die Informationsvermittlung der klassischen Informationsquellen verschmelzen miteinander. Die Informationsbeschaffung ist unabhängig von Sendezeiten. Informationen können komfortabel gesucht und zusammengestellt werden.

Neben dem Angebot an professionellen Informationsdiensten nimmt der Anteil von nutzergenerierten Inhalten, wie Blogs, Podcasts, Videoportale, Webforen, Wikis, etc., beständig zu. Ausschlaggebend hierfür sind vor allem sinkende Kosten für PC-Hardware und Breitband-Internetzugängen, sowie die technischen Entwicklungen im Bereich der Produktion, Verarbeitung und Distribution multimedialer Daten. Jedes aktuelle Handy verfügt mittlerweile über Funktionen zur Aufzeichnung von Bild-, Ton- und Videodokumenten. Auch qualitativ hochwertige Aufzeichnungsgeräte sind für den privaten Bereich erschwinglich. Eine Nachbearbeitung der Daten ist ohne weitreichende Kenntnisse mit handelsüblicher, teilweise freier Software möglich oder wird sogar teilautomatisiert von Internetanwendungen übernommen.

Diese Entwicklung führte nicht nur zu einer positiven Bereicherung der Informations-, Meinungs- und kulturellen Vielfalt. Im Gegensatz zu den professionellen Informationsdiensten werden nutzergenerierte Inhalte meist anonym oder unter Pseudonym veröffentlicht und nutzen kaum eigene Verteilungsstrukturen. Die technische Entwicklung erleichterte nicht nur die Produktionsmöglichkeiten medialer Daten, sondern auch deren Manipulation und Fälschung. Neben dem Engagement von einzelnen Personen und Gruppen zu informieren und zu bilden gibt es genauso das Engagement, gezielt falsche Informationen zu streuen, eigene Ansichten zu verbreiten, Interessen durchzusetzen oder einfach nur zu verleumden. Bei der Masse an Inhalten ist eine Unterscheidung von echten und glaubwürdigen Informationen von Zweifelhaften ohne Hilfsmittel, Fachkenntnis oder hohem zeitlichen und finanziellen Aufwand nicht mehr möglich.

Die Sicherstellung der Integrität von Daten ist ein beständiges Problem der Datenverarbeitung. Zahlreiche Verfahren, wie Prüfsummen, Hash-Funktionen, digitale Signaturen, *Message Authentification Codes*, symmetrische und asymmetrische Verschlüsselung, etc., wurden unter anderem zu diesem Zweck entwickelt. Diese Verfahren haben jedoch Nachteile bezüglich des Schutzes medialer Daten. Die Übertragung der Schutzinformation erfolgt bei diesen Technologien getrennt von den Nutzdaten oder in Form von Metadaten. In beiden Fällen ist eine Sicherstellung der Verfügbarkeit der Schutzinformation über den gesamten Verbreitungs- und Verarbeitungsweg nur unter Einhaltung und Nutzung entsprechender Übertragungsprotokolle, -systeme und Datenformate möglich. Die digitalen Wasserzeichen mit Ihrer Eigenschaft eines nicht

wahrnehmbaren Informationskanals innerhalb der Nutzdaten stellt für den Integritätsschutz medialer Daten eine geeignete Technologie dar.

## 3.2. Eigenschaften

Wasserzeichenverfahren zum Integritätsschutz von medialen Daten beschränken sich nicht nur auf die Verifikation der Integrität. Folgende Eigenschaften sind im Zusammenhang als Anforderungen an effiziente und sichere Verfahren zu finden.

### Integrität

Die Verifikation der Integrität stellt die Hauptfunktionalität der Verfahren dar. Das Wort Integrität ist abgeleitet aus dem Lateinischen „*in tangere*“ unberührbar. Unter der Integrität von Daten versteht man, dass diese vollständig und unverändert vorliegen.

Der Informationsgehalt von medialen Daten besitzt eine große Unabhängigkeit von der binären Repräsentation der Daten. Die Integrität von medialen Daten muss immer mit dem Hintergrund des jeweiligen Anwendungsszenarios betrachtet werden. Der Dateninhalt von medialen Daten ist eine subjektiv wahrgenommene Information. Je nach Art der Datennutzung wird die Integrität medialer Daten unterschiedlich verstanden. Das Spektrum der Integrität reicht von der binären Datenrepräsentation bis hin zur Datensemantik. Im Folgenden werden mögliche Interpretationen der Datenintegrität dargelegt.

- **strikt:** Bei einer strikten Verifikation der Integrität werden keine Veränderungen der Daten zugelassen. Die Integrität der binären Repräsentation der Daten wird sichergestellt. Ändert sich auch nur ein Bit, schlägt die Verifikation der Daten fehl. Diese Form der Verifikation gewährleistet die höchste Sicherheit der Datenintegrität, schränkt jedoch die Nutzungsmöglichkeiten der Daten am meisten ein.

- **qualitativ:** Eine strikte Verifizierung von medialen Daten ist häufig nicht zweckmäßig. Mediale Daten vermitteln subjektiv wahrgenommene Informationen. Ein Teil der Informationen, die durch mediale Daten vermittelt werden, wie z.B. Farb- und Helligkeitswerte (Bilder) oder die Auslenkung einer Schwingung (Audio), sind für einen Nutzer nicht wahrnehmbar bzw. irrelevant. Der Aspekt von irrelevanten Datenbestandteilen kommt besonders bei der verlustbehafteten Kompression zu tragen. Mediale Daten, die rein subjektiv wahrgenommen identisch sind, können sich in ihrer binären Repräsentation stark unterscheiden. Für einen Nutzer ist die Integrität von medialen Daten gegeben, solange sich das Original und die veränderten Daten qualitativ nicht unterscheiden lassen. Die

Übertragung oder Archivierung von medialen Daten schließen fast grundsätzlich Operationen wie verlustbehaftete Kompression oder Formatkonvertierung ein. Diese Operationen sind darauf ausgelegt, die Qualität der Daten zu erhalten, verändern jedoch die binäre Repräsentation. Bei einer qualitativen Verifizierung ist im Gegensatz zu einer strikten Verifizierung die Integrität der Daten auch nach geringfügigen Veränderungen gegeben. Die Nutzbarkeit der Daten wird bei dieser Form der Verifizierung erhöht. Eine exakte Trennung zwischen nicht wahrnehmbaren Veränderungen und störenden Veränderungen ist nicht möglich. Die Sicherheit der qualitativen Verifikation der Integrität nimmt im Vergleich zur strikten Verifizierung ab.

- **inhaltlich:** Auch nach wahrnehmbaren Qualitätsverlusten sind mediale Daten weiterhin nutzbar. Qualitätsverluste werden in bestimmten Anwendungsszenarien gewollt in Kauf genommen. Ein Beispiel hierfür sind ressourcenarme Umgebungen, wie mobile Anwendungen. Verminderte Speicherkapazitäten, Rechenleistung und Übertragungsbandbreite werden meist mit einer Reduktion des Datenumfangs durch verringerte Auflösung und starke Kompression der Daten kompensiert. Von einer inhaltlichen Integrität kann man sprechen, wenn die Daten wahrnehmbar verändert wurden, die Veränderungen aber nicht den grundsätzlichen Inhalt der Daten verändern. Die Störungen der Daten dürfen keine eigene inhaltliche Bedeutung besitzen. Zulässige Störungen bei dieser Form der Verifikation lassen sich am besten als rauschartige Störungen beschreiben.

- **semantisch:** Eine semantischen Verifikation bezieht sich auf die Integrität der Bedeutung der Daten. Mediale Daten vermitteln Informationen in einer für den Menschen leicht verarbeitbaren Form. Die Informationsaufnahme in Form von Hören und Sehen bietet einen besseren Lerneffekt im Vergleich zum Lesen von textuellen Informationen allein. Neben der eigentlichen Information, die mit den Daten vermittelt werden soll, weisen mediale Daten eine Vielzahl weiterer Informationen auf, welche nicht zur primären Bedeutung der Daten oder nur am Rande dazu beitragen. Der Integritätsschutz der Datensemantik erfordert, dass zwischen Sender und Empfänger der Daten klar definiert ist, welche Zeichen und Objekte mit einem Sinn bzw. Bedeutung belegt sind. Angenommen wird das Beispiel einer Audiodatei, in der eine Person einen Satz spricht. Bei der Bedeutung der Audiodatei ist zu vermuten, dass der gesprochene Satz die zu übertragende Information darstellt. Es kann aber auch sein, dass die Identität des Sprechers in Form seiner Stimme die eigentliche Information darstellt und der Satz von seinem Inhalt her uninteressant ist. Die semantische Integrität von Daten ist gegeben, wenn alle Objekte, welche mit einer Bedeutung belegt sind, erkennbar erhalten bleiben und keine neuen Objekte mit einer zwischen Sender und Empfänger definierten Bedeutung hinzukommen. Eine Sprachdatei, in der die gesprochenen Wörter als Semantik definiert sind, behält ihre semantische Integrität,

auch wenn diese mit Hintergrundmusik und -geräuschen belegt wird, oder Sprechpausen, solange diese nicht in der Semantik definiert sind, gekürzt werden.

Mit dem Übergang von einer strikten hin zur einer semantischen Verifikation nimmt der Umfang an zulässigen Störungen zu. Eine exakte Trennung zwischen zulässigen und unzulässigen Störungen der Daten ist praktisch nicht möglich. Die Sicherheit der Verfahren nimmt potenziell ab, wobei die Nutzungs- und Verarbeitungsmöglichkeiten zunehmen.

### Authentizität

Daten können erst als vertrauenswürdig angesehen werden, wenn sie unverändert und vollständig vorliegen (Integrität) und die Quelle der Daten eindeutig identifiziert werden kann. Die Authentizität[1] von Informationen ist gegeben, wenn die Identität des Urhebers und die Echtheit der Information nachgewiesen sind. Authentizität bedeutet nicht, dass die Informationen unverändert vorliegen. Wird neben der Integrität auch die Authentizität von Daten nachgewiesen, spricht man von einer Authentifizierung.

### Lokalisation

Die Lokalität beschreibt die kleinste zu verifizierende Einheit bzw. Struktur. Eine Beschränkung auf die reine Detektion von Veränderungen führt auch bei der kleinsten detektierbaren Veränderungen zu einer kompletten Rückweisung der Datenintegrität und somit zum Totalverlust an verwendbaren Informationen. Die Lokalisation von Störungen ermöglicht es, veränderte Datenbestandteile zu detektieren und die Integrität von unveränderten Datenbestandteilen weiterhin zu verifizieren. Je nach Anwendungsszenario und Sicherheitsanforderungen ist eine Weiternutzung der unveränderten Datenbestandteile möglich. Position und Umfang der Störungen können auch Rückschlüsse auf den Angreifer und dessen Intention (z.B. Zensur) liefern. Wie genau Position und Umfang einer Störung eingegrenzt werden können, wird durch die Lokalität eines Verfahrens angegeben.

### Rekonstruktion

Unter der Eigenschaft der Rekonstruktion wird eine teilweise bis vollständige Wiederherstellung der Trägerdaten nach einem störungsbedingten Informationsverlust verstanden. Werden Datenbestandteile verschoben oder entfernt, liegt in der Wiederherstellung der originalen Positionen der verschobenen bzw. noch vorhandenen Datenbestandteilen eine Rekonstruktion vor.

[1] Aus dem Griechischen für: Glaubwürdigkeit, Echtheit.

Eine Inhaltsabhängigkeit des Wasserzeichens von den Trägerdaten ist für eine Rekonstruktion nicht zwingend notwendig. Eine Rekonstruktion wird durch eine hohe Redundanz begünstigt. Aufgrund der Wechselwirkung zwischen Wasserzeichenkapazität und -robustheit ist eine Rekonstruktionsfähigkeit vorwiegend bei eher fragilen Verfahren zu finden. Die Rekonstruktion ist nicht mit der Invertierbarkeit, der Möglichkeit, die Veränderungen der originalen Trägerdaten in Folge der Wasserzeicheneinbettung rückgängig zu machen, zu verwechseln.

### Inhaltsabhängigkeit

Ist die Wasserzeicheninformation nicht von den Trägerdaten abhängig, können die Trägerdaten nicht nachweisbar verändert werden, solange die Wasserzeicheninformation erhalten wird. Dieses Sicherheitsrisiko kann umgangen werden, indem die Wasserzeicheninformation in Abhängigkeit der Trägerdaten generiert wird.

### Verifizierung

Die Verifikation von Wasserzeichen kann geheim oder öffentlich gestaltet werden. Bei einer geheimen Verifikation ist es nur dem Erzeuger bzw. einer eingeschränkten Gruppe möglich, die Wasserzeicheninformation auszulesen. Digitale Wasserzeichentechniken sind symmetrische Verfahren. Wird die Wasserzeicheneinbettung abhängig von einem geheimen Schlüssel (engl. *secret key*) konstruiert, ist dieser auch zum Auslesen des Wasserzeichens notwendig. Das Auslesen des Wasserzeichens ist nur mit Kenntnis des geheimen Schlüssels möglich. Bei einer öffentlichen Verifizierung ist der Schlüssel öffentlich bekannt. Jedem ist es möglich, das Wasserzeichen auszulesen. Diese Eigenschaft ist für die meisten Anwendungsszenarien des Integritätsschutzes erforderlich. Die öffentliche Verifizierung stellt jedoch ein Sicherheitsproblem dar. Auf Grund der symmetrischen Arbeitsweise ist mit Bekanntgabe des Schlüssel neben dem Auslesen auch die Einbettung und somit eine beliebige Manipulation der Wasserzeicheninformation möglich. Auch wenn das Wasserzeichen nicht vor Veränderungen geschützt werden kann, ist es mittels asymmetrischer Verschlüsselung möglich, die Generierung von gültigen Wasserzeicheninformationen zu verhindern. Wird die Wasserzeicheninformation mit dem geheimen Schlüssel eines asymmetrischen Verschlüsselungsverfahrens verschlüsselt, kann die Information mit dem öffentlichen Schlüssel wiedergewonnen werden. Ist es einem Angreifer möglich, die Wasserzeicheninformation beliebig zu verändern, benötigt er den geheimen Schlüssel des asymmetrischen Schlüsselpaares, um die Wasserzeicheninformation zu generieren, die entschlüsselt eine sinnvolle Information darstellt. Bis jetzt ist es noch nicht gelungen, eine sichere öffentliche Wasserzeichentechnik zu entwickeln, die sowohl den Schutz des Wasserzeichens und der Wasserzeicheninformation gewährleistet.

## 3.3. Klassifizierung

In der Literatur werden Wasserzeichenverfahren zum Integritätsschutz von medialen Daten unter dem Begriff der „fragilen Wasserzeichen“ zusammengefasst. Durch die unterschiedlichen Anforderungen an den Integritätsschutz und Sicherheitsaspekt haben sich verschiedene Wasserzeichenverfahren entwickelt. In [RD02, Ste04, ZS03] wird die Klassifizierung in fragile, semi-fragile und inhalts-fragile (engl. *content-fragile*) Wasserzeichenverfahren, sowie invertierbare Wasserzeichenverfahren [Ste04] vorgenommen.

#### Fragile Wasserzeichen

Die fragilen Wasserzeichen sind drauf ausgelegt, eine strikten Schutz der Trägerdaten zu gewähren. Jegliche Veränderungen der Trägerdaten führen zur Beschädigung der Wasserzeicheninformation. Das erfolgreiche Auslesen bzw. die Detektion der Wasserzeicheninformation verifiziert die Integrität der Trägerdaten.

Invertierbare Wasserzeichen haben die Eigenschaft, dass die Veränderungen der Trägerdaten, die im Zuge der Wasserzeicheneinbettung erfolgten, rückgängig gemacht werden können. Ein ursprüngliches Einsatzgebiet der invertierbaren Wasserzeichen sind z.B. medizinische Bilder, in denen durch Wasserzeichen verursachte Artefakte zu Fehldiagnosen führen können. Für den Einsatz im Bereich des Integritätsschutzes bedient man sich der Eigenschaft, dass ein Wasserzeichen nur invertiert (ausgelesen) werden kann, solange es nicht beschädigt worden ist. Das erfolgreiche Invertieren (Auslesen) des Wasserzeichens verifiziert hier die Integrität der Trägerdaten. Da invertierbare Wasserzeichen keine Robustheit aufweisen, werden sie bei den fragilen Wasserzeichenverfahren eingeordnet.

#### Semi-fragile Wasserzeichen

Ziel dieser Verfahren ist es, bestimmte Verarbeitungsschritte der Trägerdaten ohne Integritätsverlust zuzulassen. Ein qualitativer bis inhaltlicher Integritätsschutz der Daten soll gewährt werden. Das Wasserzeichen ist bei diesen Verfahren dahingehend entwickelt, dass es zulässige Verarbeitungsschritte unbeschadet überstehen kann, jedoch in Folge anderer Veränderungen zerbricht. Fast grundsätzlich, aber nicht zwingenderweise, ist bei diesen Verfahren eine Robustheit gegenüber verlustbehafteter Kompression gefordert.

#### Inhalts-fragile Wasserzeichen

Im Vergleich zu anderen fragilen Wasserzeichenverfahren nehmen die inhalts-fragilen Wasserzeichen am meisten Bezug auf die Charakteristiken medialer Daten. Die Verifikation der

Integrität erfolgt nicht, wie bei den anderen Verfahren, anhand von Beschädigung des Wasserzeichens, sondern anhand von inhaltlichen Beschreibungen der Trägerdaten. Diese werden mit dem Wasserzeichen robust in die Trägerdaten eingebettet.

Die Einteilung in fragil, semi-fragil und inhalts-fragil deckt die meisten existierenden Verfahren ab. Die Inhaltsabhängigkeit, welche bei den inhalts-fragilen Verfahren obligatorisch ist, wird bei den fragilen und semi-fragilen Verfahren nicht berücksichtigt. Verfahren mit robusten Wasserzeichentechniken ohne Bezug auf den Dateninhalt, wie dem Verfahren von Olsen et al. [OQV05], sind in dieser Klassifikation ausgeschlossen. Bei dem Verfahren von Olsen et al. werden Audiodaten in Blöcke unterteilt. Mit einer robusten Wasserzeichentechnik werden Indexnummern in diese Blöcke eingebettet. Alle Operationen, welche die Reihenfolge bzw. Vollständigkeit der Indexnummern verändern, wie z.B. das Löschen, Einfügen oder Versetzen von Audiopassagen, können erkannt werden.

Des Weiteren werden die Begrifflichkeiten und Definitionen in der Literatur unterschiedlich verwendet. Inhalts-fragile Verfahren sind unter den Bezeichnungen *content-fragile* [Dit01], *feature-based watermark* [RD02], *semi-fragile* [WF10], *semi-fragile signature watermarking* [FKK04] oder *asymmetric signature scheme* [WLC07] zu finden. Fragile Verfahren bei einer Inhaltsabhängigkeit vom Trägermaterial als (*self-embedding*) Verfahren [CHW08] bezeichnet. Rey et al. [RD02] klassifizieren neben fragilen und semi-fragilen Verfahren die *signature* Verfahren, welche inhalts-fragile Verfahren in Kombination mit asymmetrischen Signaturen darstellen.

In dieser Arbeit wird eine Einteilung der Verfahren anhand der Robustheit des Wasserzeichens und dessen Inhaltsabhängigkeit von den Trägerdaten (siehe Abbildung 3.1) vorgeschlagen. Diese deckt sich mit der Klassifikation nach [RD02, Ste04, ZS03], berücksichtigt jedoch durchgehend Inhaltsabhängigkeit und robuste Verfahren. Die Einteilung nach der Robustheit des Wasserzeichens erfolgt in fragil, semi-fragil und robust. Als fragile gelten dabei Wasserzeichentechniken, deren Wasserzeicheninformation nach jeglichen Veränderungen der Trägerdaten Beschädigungen aufweisen sollen. Semi-fragile Techniken besitzen Robustheit gegen bestimmte Veränderungen der Trägerdaten, sind aber ansonsten fragil. Robuste Wasserzeichentechniken überstehen sämtliche Veränderungen der Trägerdaten unbeschadet. Eine vollständige Robustheit kann in der Praxis nicht erfüllt werden. Alle Techniken, die keine gewollte Zerbrechlichkeit aufweisen, werden als robust betrachtet. Es können somit auch Wasserzeichentechniken als robust eingestuft werden, die eine geringere Robustheit als semi-fragile Techniken besitzen. Als inhalts-fragil werden hier jene Verfahren verstanden, bei denen die Wasserzeicheninformation vom Inhalt der Trägerdaten abhängig und Wasserzeichentechnik robust ausgelegt ist.

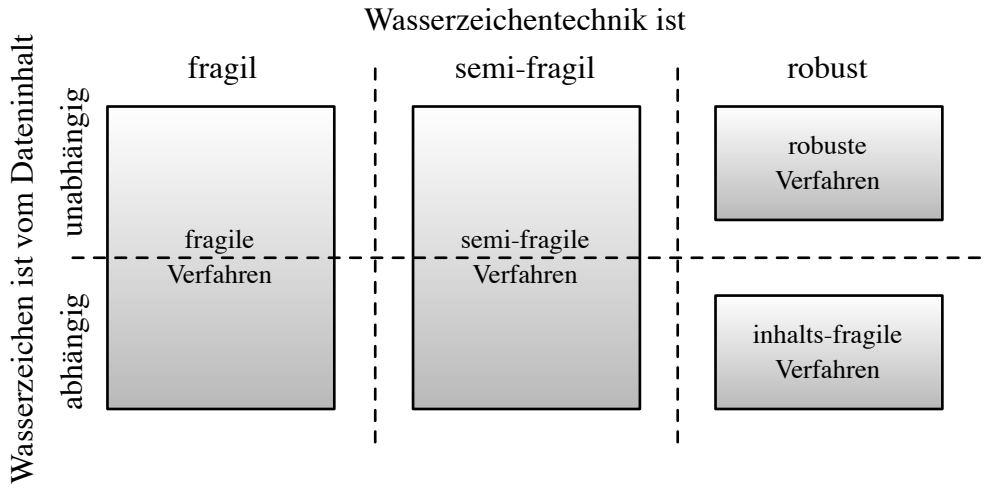


**Abbildung 3.1.:** Klassifikation von Wasserzeichenverfahren zum Integritätsschutz

Eine weiterführende Klassifizierung ist generell über die oben beschriebenen Eigenschaften möglich.

## 3.4. Inhalts-fragile Wasserzeichenverfahren für Audiodaten

### 3.4.1. Systemkonzept der inhalts-fragilen Wasserzeichen

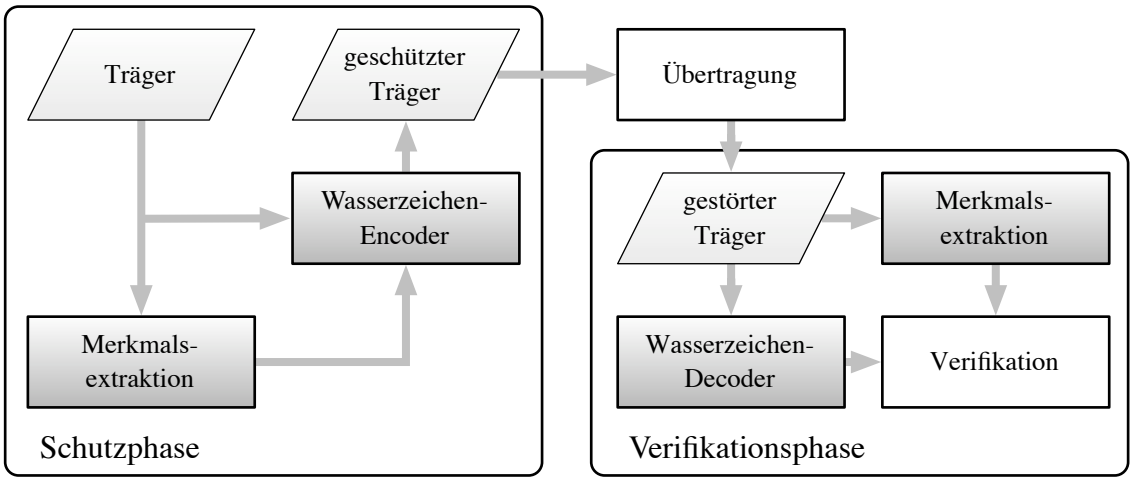


**Abbildung 3.2.:** Allgemeines Systemkonzept der inhalts-fragilen Wasserzeichenverfahren

Die Struktur der inhalts-fragilen Wasserzeichen besteht in Anlehnung an die digitalen Wasserzeichen aus einem Wasserzeichen-Encoder und einem Wasserzeichen-Decoder. Die grundlegende Struktur der inhalts-fragilen Wasserzeichen ist in Abbildung 3.2 dargestellt. Das Verfahren gliedert sich in eine Schutzphase, die Übertragung der Daten und eine Verifikationsphase.

In der Schutzphase wird eine Beschreibung des Dateninhaltes, im Weiteren als Inhaltsmerkmal (engl. *feature*) bezeichnet, aus den Trägerdaten extrahiert. Das Inhaltsmerkmal wird mit einer robusten Wasserzeichentechnik in die zu schützenden Trägerdaten eingebettet. Die mit einem Wasserzeichen markierten Trägerdaten werden jetzt sinnbildlich über einen unsicheren Übertragungskanal geleitet. Der Prozess der Übertragung der Audiodaten ist nicht nachvollziehbar, da generell keine Informationen über Verarbeitungsschritte während der Verbreitung der Trägerdaten mit übertragen werden. Auf der Empfängerseite wird aus den potenziell manipulierten Trägerdaten die Wasserzeicheninformation und somit das originale Inhaltsmerkmal extrahiert. Analog zur Merkmalsextraktion aus den originalen Daten, werden für die empfangenen Trägerdaten erneut die Inhaltsmerkmale bestimmt. Über den Vergleich, der mit dem Wasserzeichen übertragenen originalen Inhaltsmerkmale und der erneut extrahierten Inhaltsmerkmale, wird der Nachweis der Datenintegrität erbracht. Die Integrität der Trägerdaten ist bei Übereinstimmung beider Inhaltsmerkmale gegeben. Unterschiede zwischen den Inhaltsmerkmalen zeigen unzulässige Veränderungen an.

Im Folgenden werden bekannte Vertreter der inhalts-fragilen Wasserzeichenverfahren für Audiodaten vorgestellt. Die Beschreibung der Verfahren soll dazu dienen, eine Vorstellung für das verwendete Beschreibungsmerkmal und die Wasserzeichentechnik zu bekommen. Ein Vergleich der Verfahren wird über eine Auflistung der wesentlichen Verfahrensparameter geführt. Bei Interesse an der detaillierten Funktionsweise der Verfahren werden die zu den jeweiligen Verfahren genannten Quellen empfohlen.

### 3.4.2. Verfahren von Steinebach et al.

Eines der ersten inhalts-fragilen Verfahren zum Integritätsschutz von Audiodaten wurde von Steinebach et. al. [SD03, Ste04] vorgestellt. Die Autoren untersuchen verschiedene Metriken auf ihre Eignung als inhaltsrelevantes Audiomerkmal. Vorgeschlagen werden der Effektivwert (engl. *Root Mean Square* - RMS) in Bezug auf die mittlere empfundene Lautstärke und die Nulldurchgangsrate (engl. *Zero Crossing Rate* - ZCR ) in Bezug auf die Helligkeit der Audiodaten. Des Weiteren wird der Einsatz der Spektralinformation verschiedener Frequenzbänder untersucht. Drei Bänder mit den Frequenzbereichen von 500 bis 4 000 Hz, 4 000 bis 8 000 Hz und 8 000 bis 16 000 Hz werden vorgeschlagen. Um die Merkmalsextraktion in Form des Effektivwertes und der Nulldurchgangsrate auf für die menschliche Wahrnehmung relevante Frequenzanteile zu beschränken, werden als Vorverarbeitungsschritt der Merkmalsextraktion unterschiedliche Bandpassfilter angewendet. In der weiterführenden Arbeit von Eickhoff [Eic08] werden diese Merkmale und weitere für ein modellbasiertes Verfahren zur Beschreibung von

Audiomerkmalen verwendet.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}\mathbf{s}[i]^2} \tag{3.1}$$

$$ZCR = \frac{1}{N-1}\sum_{i=0}^{N-2}\left|\frac{sign(\mathbf{s}[i]) - sign(\mathbf{s}[i+1])}{2}\right|, \qquad sign(\mathbf{s}[i]) = \begin{cases} 1 & \text{für } \mathbf{s}[i] \geq 0 \\ 0 & \text{für } \mathbf{s}[i] < 0 \end{cases} \tag{3.2}$$

Die Eignung der Merkmale wird anhand verschiedener Operationen, u.a. Einbetten von Wasserzeichen, MP3-Kompression, Equalizer, Bandpass, Hall, Veränderung der Dynamik und Tonhöhenänderung, untersucht. Die Operationen werden subjektiv in „unhörbar“, „leicht“, „mittel“ und „stark“ hörbar unterteilt. Die Autoren kommen zu dem Ergebnis, dass der Effektivwert „mittel“ hörbare, die Nulldurchgangsrate „stark“ hörbare, die spektralen Eigenschaften in dem Band von 500 bis 4 000 Hz „stark“ hörbare und in dem Band von 4 000 bis 8 000 Hz „mittel“ hörbare Veränderungen erkennen können.

Die Datenrate der Inhaltsmerkmale ist von der Dauer des Audiosegmentes, aus dem sie extrahiert werden, und ihrer Auflösung abhängig. Bei sinnvollen Einstellungen dieser Parameter übersteigt die Datenrate der Inhaltsmerkmale die Kapazität von robusten Wasserzeichensystemen um ein Vielfaches. Zur Verringerung der Datenrate untersuchen die Autoren Hash-Funktionen und Prüfsummen. Der Einsatz dieser Funktionen führt bei minimalen Veränderungen der Merkmale auf stark unterschiedliche Hash-Werte bzw. Prüfsummen. Die Toleranz gegenüber leichten Veränderungen der Trägerdaten geht hierbei verloren. Eine Unterscheidung von leichten und starken Störungen ist nicht mehr möglich, lediglich eine Unterscheidung von unveränderten und veränderten Daten. Um dem entgegnen zu wirken, werden die Merkmale vor Berechnung der Hash-Wert bzw. Prüfsummen quantisiert. In Folge der Quantisierung zeigt sich nur noch der Effektivwert als geeignetes Inhaltsmerkmal. Die Leistungsfähigkeit des Wasserzeichenverfahrens wird mit folgenden Parametern dargelegt. Als Inhaltsmerkmal wird der Effektivwert genutzt. Die Merkmalsextraktion erfolgt in dem Frequenzbereich von 2 000 bis 6 000 Hz. Die Prüfsumme besteht aus 4 Bit und wird über 48 Audiosegmente mit einer Länge von je 2 048 Samples gebildet. Es liegt somit ein Inhaltsmerkmal mit einer Auflösung von 4 Bit pro 2,23 Sekunden Audio vor. Neben dem Inhaltsmerkmal wird eine Synchronisationsinformation von 2 Bit eingebettet. Die Nutzbitrate der Wasserzeichentechnik muss wenigstens 2,6917 Bit pro Sekunde betragen. In ausgewählten Beispielen wird das Inhaltsmerkmal mit einer 8 Bit Auflösung pro 1,49 Sekunden Audio gebildet [Ste04, Seiten 91-92]. Die Einbettung des Merkmals wird im Frequenzbereich von 10 bis 14 kHz vorgenommen. Als Testdaten dient eine Auswahl von 125 Sprach- und Musikdateien (mono, 16 Bit, 44,1 kHz) mit wechselnder Qualität, entnommen von CDs, Hör- und Rundfunk.

Die Verifikation der Datenintegrität wird anhand eine Schwelle für die zulässige Fehlerrate des

Beschreibungsmerkmales vorgenommen. Unmittelbar nach dem Einbetten des Wasserzeichens und nach unhörbaren bzw. leicht hörbaren Verarbeitungsschritten kann die Wasserzeicheninformation nicht korrekt ausgelesen werden. Die Grundfehlerrate der Wasserzeichentechnik wird mit etwa 10% angegeben. Bei der Verifikation von Daten muss diese Grundfehlerrate berücksichtigt werden. Die Fehlerrate gibt die unterste Schwelle vor, für die eine erfolgreiche Verifikation der Datenintegrität vorgenommen werden kann. Durch Anpassung des Schwellwertes und der Quantisierung kann die Toleranz des Verfahrens gegenüber Störungen festgelegt werden. Der Integritätsschutz bewegt sich im Bereich eines qualitativen bis inhaltlichen Verständnisses der Datenintegrität.

### 3.4.3. Audio Authentifizierung mittels robuster Hash-Funktionen

In den Veröffentlichungen [ZS08b, ZS08a, ZS09a, ZS09b] stellen Zmudzinski et al. ein Authentifizierungverfahren auf Basis von robusten *Message Authentication Codes* (rMAC) vor. Die Autoren bauen bei der Generierung der Merkmalsinformation auf der Arbeit von Haitsma et al. [TK01, HK02], welche sich mit der Entwicklung von robusten Hash-Funktionen zur Identifikation von Audiomaterial beschäftigen, auf. In dem Vordergrund stellen die Autoren die sichere schlüsselabhängige Generierung der Merkmalsinformation. Für die Merkmalsgenerierung wird die Audiodatei in Blöcke von mehreren Sekunden unterteilt. Für jeden Block wird ein rMAC, bestehend aus 128 Bit, berechnet. Ein Block wird in Segmente von je 1 024 Samples unterteilt. Die Segmente werden mittels der Fast-Fourier-Transformation (FFT) in den Frequenzbereich überführt. Um ein Bit des rMAC zu generieren, werden mit Hilfe eines Schlüssels pseudozufällig vier Koeffizienten gewählt. Über Differenzenbildung der gewählten Koeffizienten (3.3) und anschließender Vorzeichenentscheidung (3.5) erfolgt die Bestimmung des Bitwerts $H$. In [ZS09a] werden nur noch zwei Koeffizienten für die Bestimmung eines Bits verwendet (3.4). Dieser Prozess wird solange wiederholt, bis die gewünschte Länge an Bitwerten des rMAC erreicht ist.

$$d_4 = (e(n_1,t_1) - e(n_2,t_2)) - (e(n_3,t_3) - e(n_4,t_4)) \tag{3.3}$$

$$d_2 = e(n_1,t_1) - e(n_2,t_2) \tag{3.4}$$

$$H = \begin{cases} 1 & \text{für } d \geq 0 \\ 0 & \text{für } d < 0 \end{cases} \tag{3.5}$$

Der eben beschriebene Algorithmus stellt nur die grundlegende Funktionsweise der Merkmalsextraktion dar. Die Autoren führen als weitere Prozesse vor der Selektion der Koeffizienten eine ungleichförmige Quantisierung der Koeffizienten auf Basis des psychoakustisches Modells des

MPEG Audio Layer 2 [Int98] und eine Normalisierung des Frequenzspektrums ein. Als Wasserzeichentechnik wird der PCM-Algorithmus von [Ste04] verwendet.

Die Verifikation der Eigenschaften des rMAC erfolgt anhand verlustbehafteter MP3-Kompression und dem Ersetzen von Audiopassagen. Als Testdaten (mono, 16 Bit, 44,1 kHz) werden Musik, Hörbücher, Radiomitschnitte unterschiedlicher Genre und Produktionsqualitäten mit einer Gesamtspieldauer von mehreren Stunden verwendet. Der Extraktionsbereich für einen rMAC liegt in ersten Veröffentlichungen [ZS08a, ZS08b] im Frequenzbereich von 150 bis 7 000 Hz und erstreckt sich über 20 Sekunden Audio. In [ZS09b] erweitern die Autoren den Frequenzbereich auf 100 bis 10 000 Hz. Die Größe der Audiopassagen wird auf 5 Sekunden, in einem Beispiel auch auf 3 Sekunden, verringert. Die Länge des rMAC ist frei wählbar und lässt sich leicht an die Kapazität der Wasserzeichentechnik anpassen. Die Autoren verwenden einen Wert von 128 Bit und geben diesen als untere Grenze für Sicherheit gegenüber *Brute-Force*-Angriffen an. Die Fehlerraten der rMAC sind für leichte Kompression gering und steigen mit zunehmender Kompressionsstärke deutlich an. Nach dem Ersetzen von Audiopassagen liegt die Fehlerrate der rMAC bei etwa 50%. Die Verifizierung der Integrität erfolgt, wie im Verfahren von Steinebach [Ste04], schwellwertbasiert. Für die Bestimmung einer geeigneten Schwelle definieren die Autoren die MP3-Kompression bei 160 kBit/s und 128 kBit/s sowie die Wasserzeicheneinbettung als “zulässige„ Operationen. MP3-Kompression mit Bitraten von 16 kBit/s und 8 kBit/s sowie das Ersetzen von Audiopassagen werden als “böswillige„ Operationen festgelegt. Die Entscheidungsschelle liegt in [ZS08a] bei einer Fehlerrate des rMAC von etwa 18,5% und in [ZS08b] bei etwa 9,4%.

Wie bei [Ste04] liegt auch hier eine Grundfehlerrate unmittelbar nach Einbettung des Wasserzeichens vor. Die Wasserzeicheneinbettung verwendet Koeffizienten, die auch in der Merkmalsextraktion verwendet wurden. Die Modifikationen der Koeffizienten im Einbettungsprozess führen in [ZS08a] zu einer Fehlerrate von 4,84%. In [ZS08b, ZS09b] verhindern die Autoren eine überschneidende Selektion der Koeffizienten und vermindern die Fehlerrate auf 2,2%.

Eine Schwäche der Merkmalsgenerierung könnte in der Verwendung nur minimaler Datenbestandteile liegen. Eine Blockgröße von 5 Sekunden besteht bei einer Abtastrate von 44,1 kHz aus 220 500 Samples. Bei der Generierung des rMAC werden jedoch nur 2 mal 128 Koeffizienten verwendet. Dies entspricht einem Anteil von 0,23% des gesamten Datenumfangs. Ein signifikanter Beitrag der verwendeten Koeffizienten zum wahrnehmbaren Dateninhalt ist zu überprüfen. Verändert man den Dateninhalt und belässt die vom rMAC verwendeten Koeffizienten, sind keine stark störenden Effekte im veränderten Material zu erwarten. Dies setzt Kenntnis über die Positionen der Koeffizienten voraus. Aber auch ohne deren Kenntnis können etwa 9,4% der Koeffizienten verändert werden ohne, dass die Verifizierung der Daten fehlschlägt.

Bei einem Block von 5 Sekunden und einer zeitlichen Gleichverteilung der Koeffizienten in der Audiopassage entspricht dies einer veränderbaren Audiopassage von etwa 0,27 Sekunden, welches ungefähr der Dauer einer Silbe entspricht [Ter98]. Die Betrachtungen erfolgen unter der Annahme eines robusten Wasserzeichens.

### 3.4.4. Schwerpunkt-basierte Authentifizierung

Wang & Fan [WF10] präsentieren ein Authentifizierungsverfahren, welches inhaltliche Veränderung aufdecken soll, jedoch tolerant gegenüber qualitativen, wahrnehmbaren Veränderungen ist. Die folgende Beschreibung ist vereinfacht und begrenzt sich auf wesentliche Aspekte der Berechnung des Inhaltsmerkmals, für Details wird auf die Arbeit [WF10] verwiesen. Die Verarbeitung der Audiodaten erfolgt segmentweise. Für jedes Segment wird eine Art Schwerpunkt ermittelt. Ein Segment wird in $M$ Teilabschnitte $\mathbf{x}$ unterteilt. Die Teilabschnitte werden mit der Schnellen Fourier-Transformation (engl. *Fast-Fourier-Transform* - FFT) in die Frequenzraumdarstellung überführt. Für jeden Teilabschnitt $\mathbf{x}_j$ wird mittels 3.6 ein Wert $D(j)$ berechnet.

$$D(j) = \sqrt{\frac{\sum_{i=1}^{N} log_2(\left|\mathscr{F}_{FFT}\left\{\mathbf{x}_j(i)\right\}\right|^2 + 1{,}01)}{N}} \tag{3.6}$$

Aus den $D$-Werten der Teilabschnitte eines Segments wird der „Schwerpunkt“ $C$ des Segments nach (3.7) berechnet.

$$C = \left\lfloor \frac{\sum_{j=1}^{M} j \cdot D(j)}{\sum_{j=1}^{M} D(j)} \right\rfloor \tag{3.7}$$

Der Schwerpunkt $C$ wird durch eine Hash-Funktion, gefolgt von XOR-Operationen auf wenige Bits abgebildet, welche das Inhaltsmerkmal des Segments repräsentieren. Die Einbettung der Merkmalsinformation erfolgt in den Teilabschnitten, welche den Schwerpunkt ihres Segments beherbergen. Die Einbettungsdomain wird durch eine aufeinanderfolgende Anwendung der Diskreten Wavelet„=Transformation (engl. *Discrete Wavelet Transform* - DWT) und der Diskreten Kosinustransformation (engl. *Discrete Cosine Transform*) - DCT) gebildet. Die Wasserzeicheninformation wird durch Quantisierung der DCT Koeffizienten eingebracht.

Die Fähigkeiten des Systems werden an zwei Signalen (mono, 16 Bit, 44,1 kHz) mit einer Gesamtspielzeit von etwa 190 Sekunden evaluiert. Das erste Signal besitzt einen erkennbaren Grundschlag und wird mit moderner Musik verglichen. Das zweite Signal hat keinen erkennbaren Grundschlag, vergleichbar mit Jazz oder Klassik. Die Segmentierung der Testsignale wurde in 4 096 Samples große Abschnitte vorgenommen. Diese wurden in 32 Teilabschnitte mit je 128 Samples unterteilt. Das Merkmal eines Segments wird auf 4 Bit abgebil-

det, wofür eine theoretische Wasserzeichenkapazität von etwa 43 Bit pro Sekunde notwendig ist.

Die Toleranz des vorgestellten Inhaltsmerkmals sowie die Robustheit der Wasserzeichentechnik werden gegenüber MP3-Kompression, Tiefpassfilterung, Unterabtastung, Echo, Entfernen und Hinzufügen von Rauschen und Quantisierung demonstriert. Beide Größen zeigen eine hohe Toleranz bzw. Robustheit. Die Fähigkeit ihres Verfahrens, böswillige Manipulationen aufzudecken, zeigen die Autoren an Löschungen, dem Ersetzen mit Rauschen und dem Vertauschen von Audiopassagen.

### 3.4.5. Authentifizierung mittels Signatur der spektralen Komponenten

Wang et al. [WLC07] verwenden als Inhaltsmerkmal spektrale Maxima des Audiosignals. Das Audiosignal wird in Segmente von je 5 Sekunden unterteilt. Die Segmente werden mittels der FFT in den Frequenzraum überführt. Für jedes Segment wird die Frequenz mit der höchsten Amplitude bestimmt. Diese werden in einem Vektor $M$ zusammengefasst. Die Merkmalsinformation $M$ wird einer RSA-Verschlüsselung [RSA78] unterzogen. Die Verschlüsselung dient der Verifikation der Urheberschaft (Authentizität des Signals) und dem Schutz gegenüber Fälschung. Die binäre Merkmalsinformation wird in Form der Präsenz von Sinustönen in das Audiosignal eingebettet. Das Audiosignal wird hierfür in Segment von je 1 Sekunde unterteilt. Es werden zwanzig Signale, bestehend aus Sinustönen von 1 bis 20 Hz erzeugt. Jedem Sinuston wird ein Informationsbit zugeordnet. Das Audiosignal wird mit den Sinustönen, welche eine „1“ repräsentieren, überlagert. Im Vorfeld der Überlagerung wird der Frequenzbereich von 1 bis 20 Hz gelöscht. Hierfür werden die Segmente mittels der FFT in den Frequenzraum transformiert. Die Koeffizienten des Frequenzbereichs von 1 bis 20 Hz werden auf Null gesetzt. Das Signal wird in den Zeitbereich überführt. Die Detektion der Wasserzeicheninformation erfolgt durch einen schwellwertbasierten Test auf die Präsenz der Sinustöne.

Als Testdaten (mono, 16 Bit, 44,1 kHz) werden populäre Musik und Sprachaufnahmen mit einer Gesamtspielzeit von 4 Minuten und 43 Sekunden verwendet. Das Verfahren soll Veränderungen und die Zerstörung eines Audiosignales erkennen können. Der Integritätsschutz bezieht sich augenscheinlich auf inhaltliche Veränderungen. Wang et al. nehmen keine Klassifizierung von zulässigen qualitativen Veränderungen vor.

Eine Praxistauglichkeit der Wasserzeichentechnik ist zu verifizieren. Sehr tiefe Frequenzen werden teilweise bei verlustbehafteter Kompression gefiltert. Die Option, Frequenzen unter 10 Hz herauszufiltern, ist z.B. beim Import von Audiotiteln mit iTunes® vorhanden. Ein Wasserzeichen in diesem Frequenzbereich würde entfernt werden.

### 3.4.6. Zusammenfassung

In der Tabelle 3.1 sind die wesentlichen Verfahrensparameter der vorgestellten Verfahren zusammengefasst. Der Anspruch an die Integrität reicht vom qualitativen bis zum inhaltlichen Schutz von Audiodaten.

Kritisch zu betrachten sind die teilweise relativ großen Extraktionsbereiche für ein einzelnes eigenständiges Beschreibungsmerkmal (Lokalität) von 2,23 bis 5 Sekunden langen Audiobereichen. Manipulationen der inhaltlichen Aussage sind in der Größenordnung von Silben (zeitlicher Umfang von 200 bis 333 ms [Ter98]) zu erwarten. Dies wird nur durch das Verfahren von Wang et al. [WF10] mit einem Extraktionsbereich von 0,1 Sekunde für das Beschreibungsmerkmal erreicht. Eine Erkennung von Störungen, welche zeitlich kürzer sind als der Extraktionsbereich, ist nicht ausgeschlossen. Bei dem rMAC [ZS09b] ist eine anteilige Störung des Merkmals zu erwarten. Ein deutlicher Einfluss von kurzen Störungen auf die spektralen Maxima [WLC07] ist fraglich.

Weiterer Forschungsbedarf ist bei der Variationsbreite des Inhaltsmerkmales gegeben. Der geringe Extraktionsbereich von 0,1 Sekunden bei dem Verfahren von Wang et al. [WF10] ist mit einer geringen Variationsbreite des Beschreibungsmerkmals erkauft. Je geringer die Auflösung des Inhaltsmerkmals, desto größer ist die Wahrscheinlichkeit einer zufälligen Übereinstimmung von Inhaltsmerkmalen der originalen und veränderten Daten. Mit einer Auflösung von 4 Bit (Verfahren von Wang et al. [WF10] und Steinebach et al. [Ste04]) ergeben sich 16 mögliche Werte für das Inhaltsmerkmal und somit eine zufällige Übereinstimmung von 6,25%. Das bedeutet, dass 6,25% aller Veränderungen nicht erkannt werden, weil diese auf den gleichen Wert des originalen Inhaltsmerkmals abgebildet werden. In ausgewählten Beispielen wird das Inhaltsmerkmal mit einer 8 Bit Auflösung pro 1,49 Sekunden Audio gebildet [Ste04, Seiten 91-92]. Das Inhaltsmerkmal kann hier 256 mögliche Zustände annehmen und reduziert die Übereinstimmungswahrscheinlichkeit auf etwa 0,4%. Dies gilt unter der Voraussetzung, dass das Wasserzeichen korrekt ausgelesen werden kann.

Aus der Lokalität und der Variation des Beschreibungsmerkmals leitet sich die notwendige Kapazität ab, welche benötigt wird, um das Beschreibungsmerkmal mit dem Wasserzeichen zu transportieren. Die Wasserzeichenkapazität stellt für alle Verfahren den limitierenden Faktor für die Beschreibungsmerkmale dar.

Problematisch zeigt sich auch eine zu geringe Robustheit der Wasserzeichentechniken. Ausgehend von dem Systemkonzept der inhalts-fragilen Wasserzeichenverfahren erfolgt die Verifikation der Integrität über den Vergleich des originalen Beschreibungsmerkmals aus der Wasserzeicheninformation und des erneut extrahierten Beschreibungsmerkmals. Kann das Wasserzeichen

unmittelbar nach der Einbettung oder nach Störungen, welche das Merkmal nicht beeinflussen, nicht erfolgreich extrahiert werden, ergibt sich ein Grundfehler zwischen den Beschreibungsmerkmalen. Diese Grundfehlerrate, bedingt durch die mangelnde Robustheit der Wasserzeichentechnik, muss bei der Verifikation berücksichtigt werden. Die Verifikation kann in diesem Fall nicht mehr durch die Prüfung auf Identität der Merkmale durchgeführt werden, sondern anhand einer Schwelle für die zulässige Abweichung der Merkmale. Die Grundfehlerrate gibt die unterste Schwelle vor, für die eine erfolgreiche Verifikation der Datenintegrität vorgenommen werden kann. In dem Verfahren von Steinebach [Ste04] liegt die Grundfehlerrate der Wasserzeichentechnik bei etwa 10%.

Die Höhe der Schwelle ist nicht allein durch die Robustheit des Wasserzeichens bestimmt. Eine Störung des Inhaltsmerkmals allein durch die Wasserzeicheneinbettung trägt ebenso zu der Grundfehlerrate bei. Für das Verfahren von Zmudzinski et al. wird eine Störung des Beschreibungsmerkmals von 4,84% [ZS08a] bzw. 2,2% [ZS08b, ZS09b] durch die Wasserzeicheneinbettung angegeben. Es ist kausal bedingt, dass die Einbettung der Inhaltsmerkmale erst nach deren Extraktion erfolgen kann. Die Einbettung eines Wasserzeichens erfordert Modifikationen der Trägerdaten. Die Audiodaten, aus denen die Inhaltsmerkmale extrahiert wurden, werden im Zuge der Wasserzeicheneinbettung verändert. Ohne weitere Maßnahmen ist es wahrscheinlich, dass die geschützten, aber ansonsten unveränderten Audiodaten während der Überprüfung als unzulässig verändert zurückgewiesen werden.

Im Bereich der Audioauthentifizierung mittels inhalts-fragiler Wasserzeichen werden zwei verschiedene Lösungsansätze für dieses Problem verfolgt.

- Ein Ansatz ist die räumliche Trennung der Domain der Inhaltsmerkmal-Extraktion von der Domain der Wasserzeicheneinbettung. Störungen des Inhaltsmerkmals durch die Wasserzeicheneinbettung können hierdurch ausgeschlossen werden. Die Inhaltsmerkmal-Extraktion und die Wasserzeicheneinbettung treten in Konkurrenz um die verfügbare Domain, da beide typischerweise die für die Wahrnehmung relevanten Datenbestandteile als Domain nutzen. Dieser Ansatz findet Verwendung in den Verfahren von Steinebach et al. [SD03] und Wang et al. [WLC07]. Die Trennung der Domain der Inhaltsmerkmal-Extraktion von der Domain der Wasserzeicheneinbettung erfolgt unter Verwendung disjunkter Frequenzbereiche.

- Der zweite Ansatz verhindert Störungen des Inhaltsmerkmales nicht, sondern berücksichtigt diese während der Überprüfung der Audiodaten in der Verifikationsphase. Die Rückweisung von Audiodaten erfolgt hier erst, wenn die Abweichungen zwischen der eingebetteten Wasserzeicheninformation und den erneut extrahierten Inhaltsmerkmalen einen

Schwellwert übersteigen. Die Differenzierungsfähigkeit verschlechtert sich mit steigender Entscheidungsschwelle. Falsche Rückweisungen von geschützten, aber ansonsten unveränderten Audiodaten können nicht ausgeschlossen werden. Die Wasserzeichentechnik und Inhaltsmerkmal-Extraktion sind bei diesem Ansatz weitestgehend unabhängig und austauschbar. Dieser Ansatz findet Verwendung in dem Verfahren von Zmudzinski et al. [ZS08b].

Das Wasserzeichen muss jedoch im Extraktionsbereich des Inhaltsmerkmals liegen. Ein unabsichtliches Entfernen des Wasserzeichens wird hierdurch verhindert. Wichtiger noch, eine Manipulation des Wasserzeichens oder des Inhaltsmerkmals kann bei getrennten Bereichen ohne Einfluss auf die jeweils andere Größe vorgenommen werden.

## 3.5. Forschungsbedarf

Zusammenfassend wird in der nachfolgenden Arbeit die Entwicklung eines Authentifizierungssystems für Audiodaten mit den folgenden Eigenschaften angestrebt.

- Eine Audiodatei kann als echt bezeichnet werden, wenn deren inhaltliche Aussage und Qualität unverändert sind. Das zu entwickelnde Beschreibungsmerkmal der Audiodaten muss somit sensibel genug sein, um alle wahrnehmbaren und inhaltlichen Veränderung der Audiodaten erkennen zu können.

- Mit dem Bezug auf Sprachdaten ist mit Modifikationen ab dem Umfang einer Silbe mit einer inhaltlichen Veränderung von Audiodaten zu rechnen. Die obere Grenze für die Lokalität eines Inhaltsmerkmals ist mit 200 bis 333 ms (Umfang einer Silbe [Ter98]) festzulegen.

- Der gewöhnliche Umgang mit den geschützten Audiodaten durch den Nutzer darf nicht eingeschränkt werden. Eine übliche Verarbeitung mit Hinsicht auf die Verbreitung und Archivierung der Audiodaten muss ohne Zerstörung der Schutzinformation möglich sein. Hieraus folgt, dass das Inhaltsmerkmal und die Wasserzeichentechnik ausreichend robust sein müssen, um Verarbeitungsschritte ohne qualitative oder inhaltliche Veränderungen der Audiodaten, wie z.B. verlustbehaftete Kompression, ohne Schäden zu überstehen.

- Unmittelbar nach der Integration der Schutzinformation in die Audiodaten dürfen keine Veränderungen der Inhaltsmerkmale trotz Modifikation der Audiodaten durch die Wasserzeicheneinbettung entstehen. Die Wasserzeichentechnik ist so zu gestalten, dass ein eingebettetes Wasserzeichen keinen Einfluss auf die Extraktion des Inhaltsmerkmal ausübt.

- Ein unabsichtliches Entfernen des Wasserzeichens ohne Veränderung des Dateninhaltes ist zu verhindern. Die Wasserzeicheneinbettung hat somit in der gleichen Domain wie die Extraktion des Beschreibungsmerkmales zu erfolgen.
- Die Kapazität des Wasserzeichens muss ausreichend groß sein, um das Beschreibungsmerkmal transportieren zu können.
- Die Datenqualität der Audiodaten darf durch die Wasserzeicheneinbettung nicht wahrnehmbar vermindert werden.
- Die öffentliche Verifizierung der Echtheit geschützter Audiodaten, folglich durch jedermann, soll ermöglicht werden.
- Die Schutzinformation muss manipulationssicher sein, so dass es keinem Angreifer möglich ist, inhaltlich veränderte Audiodaten mit einer gültigen, gefälschten Schutzinformation zu versehen.

**Tabelle 3.1.:** Vergleich der Verfahrensparameter der vorgestellten inhalts-fragilen Verfahren

| **Autor** | **Steinebach et al.** [SD03, Ste04] | **Zmudszinski et al.** [ZS09b] | **Wang, Liao & Chen** [WLC07] | **Wang & Fan** [WF10] |
|---|---|---|---|---|
| **Integrität** | qualitativ - inhaltlich | qualitativ | qualitativ oder inhaltlich[2] | inhaltlich |
| **öffent. Verifizierung** | nein | nein | ja | nein |
| **Merkmal** | | | | |
| Art | RMS | rMAC | spektrale Maxima | Schwerpunkt |
| Verschlüsselung | k.A. | **sym. 128 Bit** | RSA (1 080 Bit[3]) | XOR |
| Variation | 4 Bit / 8 Bit [Ste04, S. 91] | **128 Bit** | 20 Bit | 4 Bit |
| Lokalität | 98 304 Samples | | $\sim$ 220k Samples | **4 096 Samples** |
| | 2,23 s | 3 s; 5 s | 5 s | $\sim$ **0,1 s** |
| | $2^{16}$ Samples / 1,49 s [Ste04, S. 91] | | | |
| Extraktionsbereich | 2 - 6 kHz | 0,1 - 10 kHz | >27,7 Hz | |
| **Wasserzeichen** | | | | |
| Transparenz | k. A | k. A | k.A. | ODG < -0,07<br>SNR < 51 dB |
| Kapazität | | | 20 Bit/s | |
| Nutzbitrate | 2,6917 Bit/s | 27,6 Bit/s[4] | 4 Bit/s | $\sim$ **43 Bit/s**[5] |
| Kanalcodierung | k. A. | k. A. | Wiederholung | k. A. |
| Domain | FFT (10 - 14 kHz) | FFT | FFT (1-20 Hz) | DWT/DCT |
| **Testdaten** | | | | |
| Art | breites Spektrum an Musik und Sprache | breites Spektrum an Musik und Sprache | populäre Musik, Sprache | mit/ohne Grundschlag |
| Anzahl | **125** | | 1 | 2 |
| Gesamtdauer | **1,5 - 7,5 h** | k. A. | 283 s | 190 s |
| tech. Daten | mono, 16 Bit, 44,1 kHz | | | |

[2]Die Autoren halten sich bei den Eigenschaften des Inhaltsmerkmals bedeckt.
[3]Theoretischer Maximalwert der Schlüssellänge für die beschriebene Testumgebung.
[4]Bezogen auf einen Extraktionsbereich von 5 Sekunden und einer rMAC-Länge von 128 Bit und 10 Bit Block Index.
[5]Theoretisch notwendige Nutzbitrate.

Kapitel 4

# Inhalts-fragile Audioauthentifizierung

*In diesem Kapitel wird ein Verfahren zur Überprüfung der Echtheit von Audiodaten mittels eines inhalts-fragilen digitalen Wasserzeichenverfahrens entwickelt. Zu Beginn wird das Systemkonzept und dessen Entwicklungskriterien dargelegt. Anschließend werden die beiden Hauptkomponenten des Systemkonzepts, die Extraktion eines Beschreibungsmerkmals der Audiodaten und die robuste Wasserzeichentechnik detailliert beschrieben. Es folgt die Leistungsanalyse beider Elemente, in der günstige Arbeitsbereiche separat für jedes Element ermittelt werden. Anschließend werden Systemmodifikationen vorgenommen, um die beiden Elemente unter Beachtung ihrer Arbeitsbereiche in einem ersten Grundsystem kombinieren zu können. Die Leistungsfähigkeit des Grundsystems wird analysiert. Der Optimierungsbedarf des Systems wird für eine Behandlung im nachfolgenden Kapitel aufgezeigt. Im Abschluss erfolgt eine Diskussion des Sicherheitsaspekts.*

## 4.1. Systemkonzept

Die Grundstruktur des Authentifizierungssystems ist in Abbildung 4.1 dargestellt. Die Verarbeitung von Audiodateien erfolgt in Verarbeitungsrahmen fester Größe. Die Spieldauer von Audiodateien ist variabel. Sie kann wenige Sekunden, aber auch mehrere Stunden umfassen. Audiodateien werden unabhängig von ihrer Länge in aufeinanderfolgende Abschnitte fester Länge zerlegt. Diese Abschnitte, im Weiteren als Rahmen bezeichnet, werden nach demselben Prinzip verarbeitet.

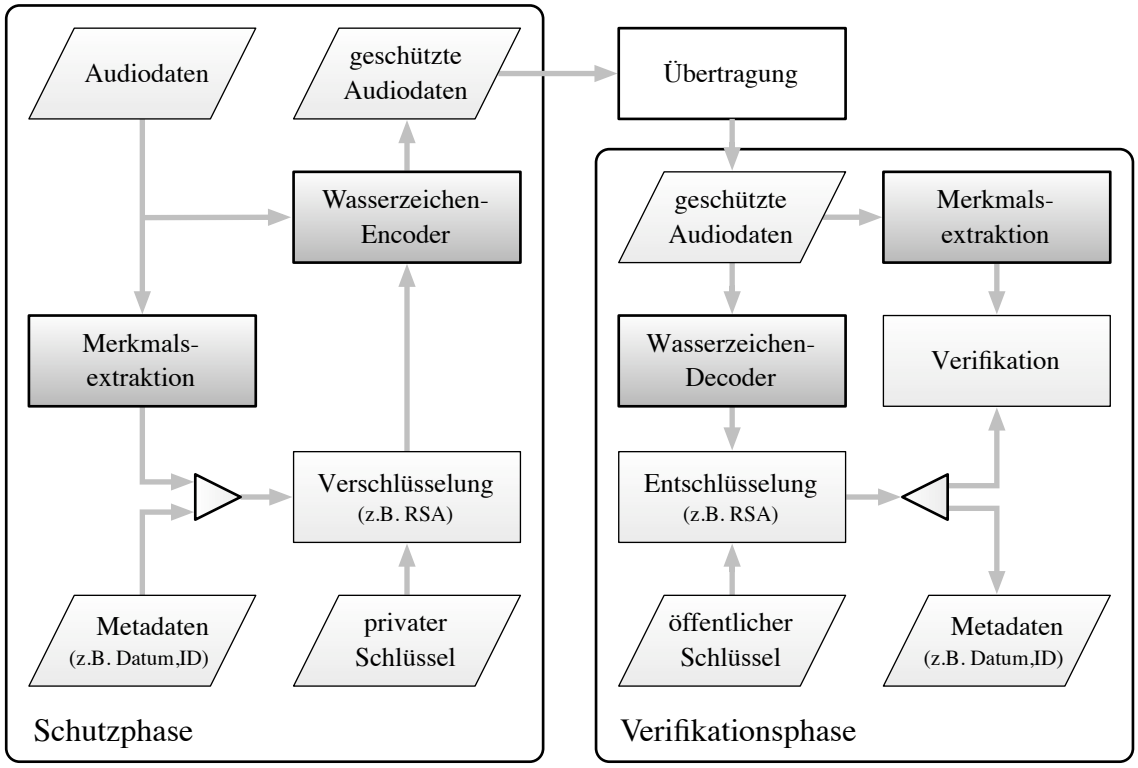


**Abbildung 4.1.:** Grundstruktur des entwickelten Authentifizierungssystems

Das Grundsystem der inhalts-fragilen Wasserzeichen ist in die Schutz- und die Verifikationsphase gegliedert. In der Schutzphase wird der Dateninhalt der Audiodaten in Form von Inhaltsmerkmalen extrahiert. Die Inhaltsmerkmale werden zusammen mit weiteren Informationen (z.B. Datum, Urheber, ID , Rahmennummer) mit dem privaten Schlüssel eines asymmetrischen Verfahrens verschlüsselt. Die verschlüsselte Information wird mittels einer robusten Wasserzeichentechnik in die Audiodaten eingebettet. Aufgrund der öffentlichen Verifizierbarkeit ist die Wasserzeicheninformation durch jedermann auslesbar und somit auch veränderbar. Die verschlüsselte Wasserzeicheninformation kann nur durch den Inhaber des privaten Schlüssels generiert werden. In der Verifizierungsphase wird aus den geschützten Audiodaten die verschlüsselte Wasserzeicheninformation ausgelesen und mit dem öffentlichen Schlüssel des asymmetrischen Verschlüsselungsverfahrens entschlüsselt. Die mit dem Wasserzeichen übertragenen originalen Inhaltsmerkmale werden mit den aus einer erneut durchgeführten Inhaltsmerkmal-Extraktion gewonnenen Inhaltsmerkmale verglichen. Der Dateninhalt ist echt, wenn beide Inhaltsmerkmale identisch sind.

## 4.2. Entwicklung eines inhaltsrelevanten Beschreibungsmerkmals für Audiodaten

### 4.2.1. Anforderungen

Digitale Audiodaten beschreiben die physikalischen Eigenschaften, wie Amplitude und Frequenz, von hörbaren Schallereignissen. Die Darstellung erfolgt typischerweise in Form der Puls-Code-Modulation (PCM), einer zeit- und wertdiskreten Beschreibung der Amplitudenwerte, bzw. einer datenreduzierten Form der PCM-Werte. Während diese Beschreibungsform für die Aufnahme bzw. Wiedergabe von Audiodaten geeignet ist, sind hinsichtlich der Wahrnehmung oder Verarbeitung des Audiodateninhalts andere Beschreibungsformen oder die Beschreibung von speziellen Eigenschaften (Merkmale) sinnvoll. Die Extraktion und Auswertung von Merkmalen findet in verschiedenen Bereichen der Audiodatenverarbeitung Anwendung. Beispiele hierfür sind Psychoakustik [ZF99], Sprach - und Sprechererkennung oder *Music Information Retrieval*. Die Forschungsergebnisse dieser Bereiche lassen sich als Grundlage für die Entwicklung von geeigneten Inhaltsmerkmalen verwenden. Für den Einsatz in inhalts-fragilen Wasserzeichen werden folgende Eigenschaften von den Inhaltsmerkmalen gefordert.

**Differenzierung:** In Abhängigkeit eines Anwendungsszenarios werden Mengen von zulässigen und unzulässigen Veränderungen der Audiodaten definiert. Die Festlegung der beiden Mengen erfolgt typischerweise durch die Zuordnung der für das Anwendungsszenario spezifischen Operationen in die jeweilige Menge oder durch die Definition einer Unterscheidungsgrenze. So läßt sich beispielsweise die verlustbehaftete Audiodatenkompression in die zulässige Menge einordnen, während Operationen wie das Löschen, Einfügen und Ersetzen von Audiopassagen in die Menge der unzulässigen Veränderungen gehören. Eine intuitive Grenze für den zweiten Fall ist die Hörbarkeit. Die Differenzierung zwischen den zulässigen bzw. unzulässigen Störungen über das Inhaltsmerkmal sollte eine hohe Deckungsgleichheit mit den Vorgaben des Anwendungsszenarios besitzen. Die Grenze zwischen den zulässigen und unzulässigen Störungen ist jedoch fließend und nicht exakt abzustecken.

Die Unterscheidung von zulässigen und unzulässigen Veränderungen der Audiodaten während der Verifikationsphase erfolgt über die Robustheit bzw. Zerbrechlichkeit des Inhaltsmerkmals und dessen Variationsbreite.

- **Toleranz:** Eine Möglichkeit, Veränderungen der Audiodaten zu klassifizieren, ist die Unterscheidung zwischen qualitativen und inhaltlichen Veränderungen. Qualitative Veränderungen lassen die inhaltliche Bedeutung der Daten unberührt, führen aber zu einer nicht

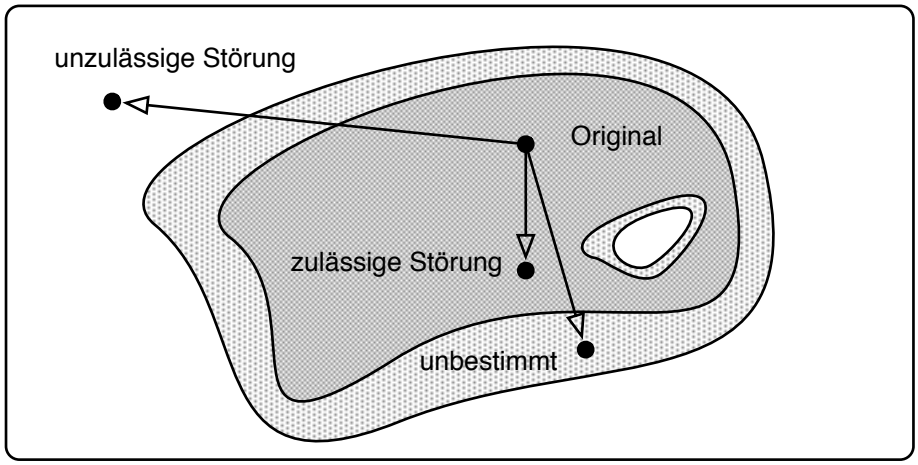


**Abbildung 4.2.:** Sinnbild für die Differenzierung der Menge von zulässigen und unzulässigen Störungen (nach [Sch09, S. 41])

wahrnehmbaren bis stark störenden Beeinflussung des wahrnehmbaren Inhalts. Beispiele hierfür sind verlustbehaftete Kompression, welche eine Verschlechterung der Qualität bewirken kann, oder das Hinzufügen von Rauschen, welches die Qualität der Audiodaten stark mindert. In beiden Fällen handelt es sich um Veränderungen mit ähnlichem Inhalt. Die Unterscheidung zwischen zulässigen und unzulässigen Veränderungen erfolgt im Bereich mit ähnlich wahrgenommenen Inhalt durch die Toleranz (Robustheit / Zerbrechlichkeit) des Inhaltsmerkmals. Das Inhaltsmerkmal sollte robust gegenüber nicht oder kaum wahrnehmbaren Veränderungen der Audiodaten sein und mit zunehmend wahrnehmbar gestörter Datenqualität Veränderungen aufweisen.

- **Variation:** Für die Erkennung von inhaltlichen Veränderungen, wie das Löschen, Ersetzen oder Einfügen von Audiopassagen, ist eine ausreichend hohe Anzahl an möglichen Variationen eines Inhaltsmerkmals notwendig. Audiosegmente mit unterschiedlich wahrnehmbaren Inhalt müssen zu unterschiedlichen Inhaltsmerkmalen führen, damit diese nicht gegeneinander ausgetauscht werden können, ohne das Alarm in der Verifikationsphase ausgelöst wird. Um die Wahrscheinlichkeit einer zufälligen Übereinstimmung unter $10^{-9}$ zu halten, ist für ein Inhaltsmerkmal mit gleichverteilten Ereignissen eine Variationsbreite von etwa 30 Bit notwendig.

In dieser Arbeit wird ein allgemeiner Schutz von Audiodaten angestrebt. Für die Entwicklung eines robusten Inhaltsmerkmals werden die Operationen als erlaubte Störungen angesehen, welche gewöhnlich bei der Verbreitung, Übertragung und Archivierung von Audiodaten anfallen, wie Formatkonvertierung, speziell verlustbehaftete Kompression, Änderung der Samplerate und Lautstärke.

**Lokalität:** Veränderungen, die darauf abzielen, den Inhalt der Audiodaten zu verändern, wie

das Entfernen, Hinzufügen oder Ersetzen von Sequenzen, sind meist lokal begrenzt. Beschränkt man sich auf den reinen Nachweis der Existenz von Veränderungen, führen auch kleinste erkennbare Veränderungen dazu, dass die komplette Audiodatei als unzulässig verändert eingestuft wird, obwohl der Großteil der Daten unverändert und verwertbar ist. Neben dem Nachweis der Existenz von unzulässigen Veränderungen der Audiodaten sind auch deren Umfang und Verbreitung von Interesse. Um eine Lokalisierung von Veränderungen anhand des Inhaltsmerkmals zu ermöglichen, müssen Veränderungen des Inhaltsmerkmales auf lokal begrenzte Veränderungen der Audiodaten zurückzuführen sein. Wahrnehmbare Störengen liegen im Bereich von 20 ms [ZF99], Silben haben einen Umfang von 200 bis 333 ms [Ter98], im Bereich von 5 Sekunden liegen Wortgruppen und kurze Sätze vor. Um einen möglichst großen Einfluss von Störungen auf ein Beschreibungsmerkmal sicherzustellen, sind ähnliche Größenordnungen des Extraktionsbereichs des Beschreibungsmerkmals und der zu erkennenden Störungsdauer anzustreben. Eine übliche Vorgehensweise, um dies zu erreichen, ist die abschnittsweise Extraktion der Inhaltsmerkmale.

**Datenrate:** Die Einbettung der Merkmalsvektoren in die Audiodaten erfolgt mittels robuster Wasserzeichenverfahren. Die größtmögliche Datenrate des Inhaltsmerkmales ist somit durch die Kapazität des Einbettungsverfahrens begrenzt. Die Kapazität dieser Verfahren ist im Bezug auf das Anwendungsszenario der Authentifizierung stark begrenzt. Inhaltsmerkmale mit geringer Datenrate begünstigen eine robuste Einbettung, jedoch zu Lasten anderer Eigenschaften, wie Lokalität und Variation. Die Datenrate wird typischerweise in Bits pro zu schützenden Audiorahmen bzw. Zeiteinheit angegeben.

**Komplexität:** Der Schutz von Audiodaten sollte mit deren Aufnahme oder spätestens vor deren Verbreitung beginnen. Für eine effiziente Integration von Schutzmechanismen in bestehende Aufnahme- und Verarbeitungssysteme ist eine geringe Komplexität gefordert. Der Ressourcenbedarf der Schutzmechanismen, z.B. in Form von Rechenzeit oder Speicherbedarf, ist gering zu halten, um zusätzliche Verzögerungen in den Verarbeitungssystemen zu vermeiden. Generell müssen Inhaltsmerkmale automatisch generierbar und numerisch auswertbar sein.

**Sicherheit:** Die Inhaltsmerkmale müssen robust gegen Angriffe sein, welche direkt auf die Merkmalsextraktion abzielen.

### 4.2.2. Extraktion des Inhaltsmerkmals

Das entwickelte Inhaltsmerkmal stellt eine stark komprimierte Zeit-Frequenz-Repräsentation der Audiodaten dar. Die Verarbeitung von Schallereignissen durch das menschliche Ohr erfolgt frequenzselektiv. Das menschliche Ohr unterteilt den Frequenzraum für die Analyse von

Audiosignalen in nichtlineare Frequenzbereiche, die sogenannten Frequenzgruppen (engl. *critical bands*). In den Frequenzgruppen erfolgt eine eigenständige Informationsauswertung, die zu einem Gesamteindruck des Hörereignisses zusammengefasst wird. Diese Eigenschaft der menschlichen Wahrnehmung dient als Basis der Entwicklung eines wahrnehmungsrelevanten Inhaltsmerkmales.

Der Ablaufplan der Inhaltsmerkmal-Extraktion ist in Abbildung 4.3 dargestellt. Die Konstruktion des Inhaltsmerkmales gestaltet sich wie folgt:

Der originale Audiorahmen $\mathbf{s} = \{s_1, s_2, \ldots, s_N\}$ wird in $n$ Segmente $\mathbf{s}_i^S$ der Länge $l$ unterteilt.

$$\mathbf{s}_i^S = \{s_{(i-1)l+1}, s_{(i-1)l+2}, \ldots, s_{il}\} = \{s_{1,i}, s_{2,i}, \ldots, s_{l,i}\}; \qquad i \in \{1, \ldots, n\}; \qquad n = \frac{N}{l} \tag{4.1}$$

Die Segmente $\mathbf{s}_i^S$ werden in einer zweidimensionalen Matrix $\mathbf{S}$ angeordnet. Die Segmente $\mathbf{s}_i^S$ bilden hierbei die Spalten von $\mathbf{S}$.

$$\mathbf{S} = \begin{Bmatrix} s_1 & s_{l+1} & \cdots & s_{N-l+1} \\ s_2 & s_{l+2} & \cdots & s_{N-l+2} \\ \vdots & \vdots & \ddots & \vdots \\ s_l & s_{2l} & \cdots & s_N \end{Bmatrix} = \begin{Bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,n} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{l,1} & s_{l,2} & \cdots & s_{l,n} \end{Bmatrix} \tag{4.2}$$

Diese Umformung dient hauptsächlich einer blockweisen eindimensionalen Transformation aller Segmente unter Nutzung effizienter Algorithmen (FFTW[1]). Des Weiteren erlaubt die Darstellung als Matrix eine einfachere Abstraktion der Verarbeitung. Die Matrix $\mathbf{S}$ wird mittels einer eindimensionalen Diskreten Kosinustransformation (Typ IV) (engl. *Discrete Cosine Transform (Type IV)* - (DCT-IV)) in den Frequenzraum überführt. Die resultierende Matrix $\mathbf{C}$ stellt eine Art Spektrogramm dar, welches den Verlauf des Frequenzspektrums der Segmente beschreibt (s. Abbildung 4.4).

$$\mathbf{C} = \begin{Bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{l,1} & c_{l,2} & \cdots & c_{l,n} \end{Bmatrix} \tag{4.3}$$

Der Frequenzbereich wird über die Bark-Skala in 25 Frequenzgruppen $\mathbf{c}_i^{F1}, \mathbf{c}_i^{F2}, \ldots, \mathbf{c}_i^{F25}$ unterteilt. Die Beträge der Koeffizienten der einzelnen Frequenzgruppen werden segmentweise

[1] *„Fastest Fourier Transform in the West„* ist eine freie in C implementierte Softwarebibliothek zur Berechnung diskreter Fourier Transformationen. Die in der FFTW verwendeten Algorithmen zählen zu den effektivsten freien Software Implementierungen der schnellen Fourier-Transformation.

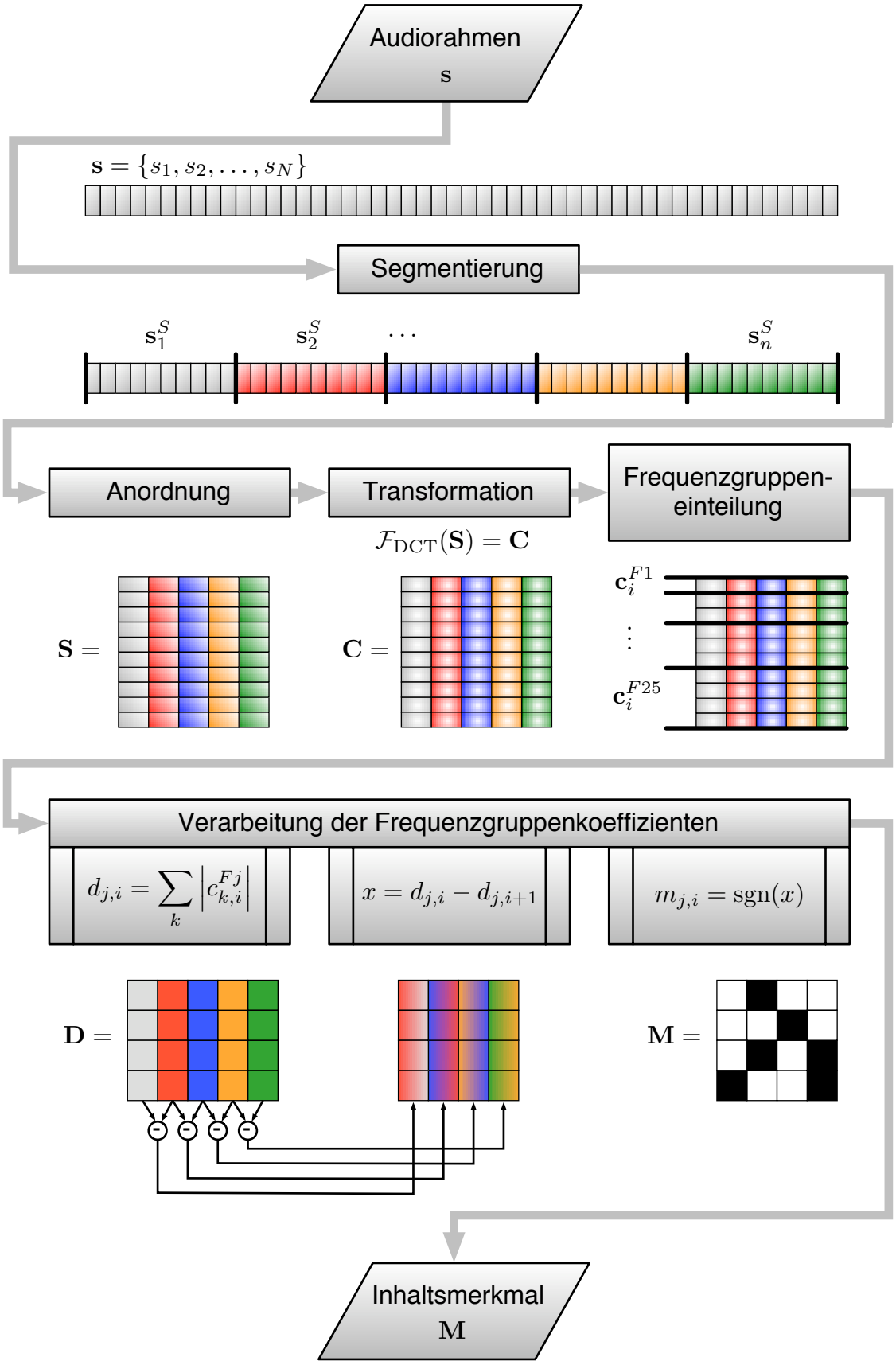


**Abbildung 4.3.:** Ablaufplan der entwickelten Inhaltsmerkmal-Extraktion

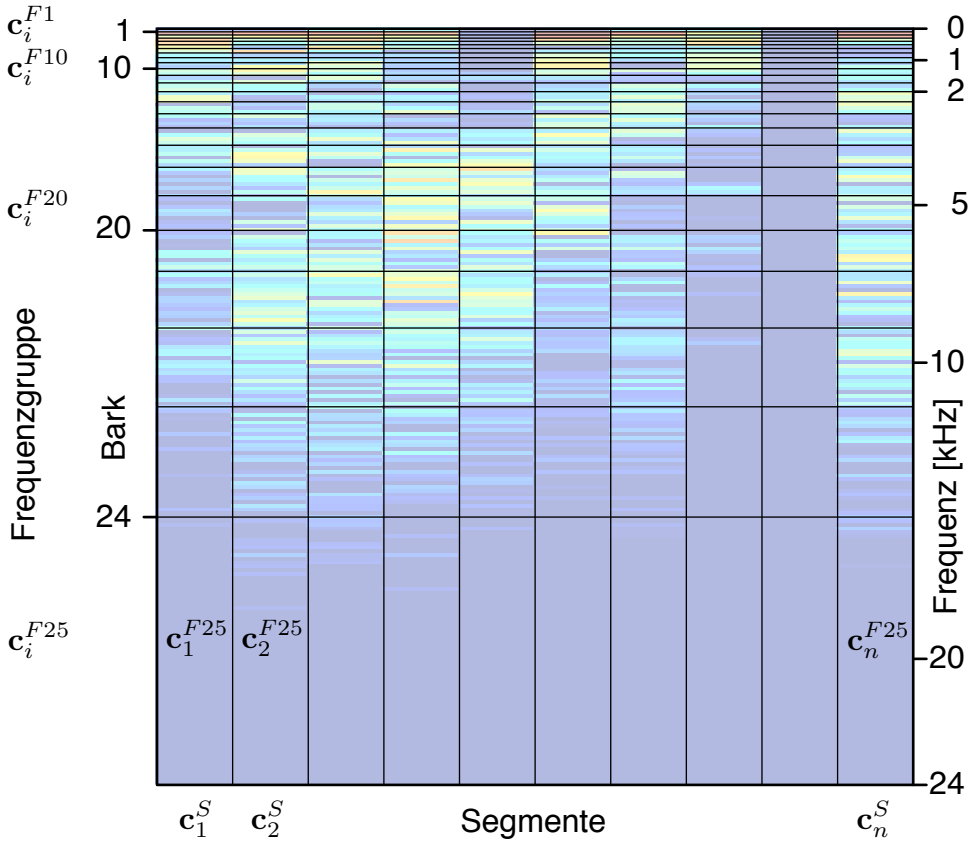


**Abbildung 4.4.:** Spektraler Verlauf von **C** mit Frequenzgruppenunterteilung

summiert und in der $25 \times n$ Matrix **D** zusammengefasst.

$$d_{j,i} = \sum_k \left| c_{k,i}^{Fj} \right| \tag{4.4}$$

$$\mathbf{D} = \left\{ \begin{array}{cccc} \sum_k \left|c_{k,1}^{F1}\right| & \sum_k \left|c_{k,2}^{F1}\right| & \cdots & \sum_k \left|c_{k,n}^{F1}\right| \\ \sum_k \left|c_{k,1}^{F2}\right| & \sum_k \left|c_{k,2}^{F2}\right| & \cdots & \sum_k \left|c_{k,n}^{F2}\right| \\ \vdots & \vdots & \ddots & \vdots \\ \sum_k \left|c_{k,1}^{F25}\right| & \sum_k \left|c_{k,2}^{F25}\right| & \cdots & \sum_k \left|c_{k,n}^{F25}\right| \end{array} \right\} = \left\{ \begin{array}{cccc} d_{1,1} & d_{1,2} & \cdots & d_{1,n} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{25,1} & d_{25,2} & \cdots & d_{25,n} \end{array} \right\} \tag{4.5}$$

Aus der Matrix **D** wird das Inhaltsmerkmal **M** gewonnen. Hierfür wird die Differenz zwischen den $d$-Werten gleicher Frequenzgruppe von benachbarten Segmenten gebildet. Die Differenzwerte werden in Abhängigkeit ihres Vorzeichens auf Binärwerte abgebildet. Die aus Binärwerten bestehende Matrix **M** stellt das Inhaltsmerkmal dar.

$$\Delta d_{j,i} = d_{j,i} - d_{j,i+1}; \qquad \forall i \in \{1,\ldots,n-1\} \quad j \in \{1,\ldots,25\} \tag{4.6}$$

$$m_{j,i} = \text{sgn}(\Delta d_{j,i}); \qquad \forall i \in \{1,\ldots,n-1\} \quad j \in \{1,\ldots,25\} \tag{4.7}$$

$$\text{sgn}(x) = \begin{cases} 1 & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases} \tag{4.8}$$

$$\mathbf{M} = \begin{Bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,n-1} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ m_{25,1} & m_{25,2} & \cdots & m_{25,n-1} \end{Bmatrix} \qquad m_{j,i} \in \{0,1\} \tag{4.9}$$

## 4.3. Entwicklung der Wasserzeichentechnik

### 4.3.1. Anforderungen

Die für die Merkmalseinbettung zu entwickelnde Einbettungstechnik unterliegt den allgemeinen Verfahrensparametern digitaler Wasserzeichen (siehe Abschnitt 2.3.2). Spezielle Anforderungen durch das Einsatzgebiet der inhalts-fragilen Wasserzeichen ergeben sich wie folgt:

#### Kapazität

Die Nutzkapazität des Wasserzeichens muss ausreichend sein, um das Inhaltsmerkmal und zusätzliche Meta„=Daten robust zu übertragen. Für den vorliegenden Anwendungsfall des Integritätsschutzes sind die Kapazitätsanforderungen deutlich höher als im Vergleich zu Anwendungsgebieten wie dem *Copy Control*. Laut Zmudzinski et al. [ZS09a] werden für den Integritätsschutz 128 Bits und mehr für einen Audiorahmen (typischerweise von wenigen Sekunden Länge) benötigt. Diese bezieht sich auf die symmetrische Verschlüsselung der Wasserzeicheninformation. Für eine asymmetrische Verschlüsselung werden Schlüssellängen von bis 3 072 bis 4 096 empfohlen [ECR10, FNI10, BBB$^+$07]. Zum Vergleich, die Kapazitätsspezifikationen der *Recording Industry Association of America* (RIAA) [RIA09] gibt 108 Bit pro Audiodatei an. Die Länge der Audiodateien bewegen sich üblicherweise im Bereich über einer Minute. Die Robustheitsanforderungen für den Integritätsschutz sind jedoch geringer.

Geht man von einem Extraktionsbereich für das Inhaltsmerkmal mit dem Umfang einer Silbe (333 ms) aus und einer Variationsbreite von 30 Bit, folgt eine notwendige Wasserzeichenkapazität von 90 Bit/s.

**Robustheit**

Die Robustheit der Wasserzeicheninformation muss solange gegeben sein, wie das Beschreibungsmerkmal bei zulässigen Störungen der Audiodaten unverändert bleibt. Wird das Beschreibungsmerkmal verändert, ist eine korrekte Übertragung des originalen Beschreibungsmerkmals nicht zwingend erforderlich.

**Transparenz**

Die Transparenz des Wasserzeichen beinhaltet im vorliegenden Anwendungsfall der inhaltsfragilen Wasserzeichenverfahren zwei Aspekte. Der erste Aspekt ist die generelle Anforderung an Wasserzeichenteckniken, transparent für die menschliche Wahrnehmung zu sein. Veränderungen der Audiodaten während Wasserzeicheneinbettung müssen so gestaltet sein, dass diese nicht bzw. kaum wahrnehmbar sind. Der zweite Aspekt gilt speziell für inhaltsfragile Wasserzeichenverfahren. Hier wird nicht nur die Transparenz in Bezug auf die Wahrnehmbarkeit gefordert, sondern auch in Bezug auf die Merkmalsextraktion. Die Wasserzeicheneinbettung darf keinen Einfluss auf das Ergebnis der Merkmalsextraktion nehmen.

### 4.3.2. Entwickelter Wasserzeichenalgorithmus

Der Ablaufplan der Wasserzeicheneinbettung ist in Abbildung 4.5 dargestellt. Die entwickelte Wasserzeichentechnik gehört zu der Gruppe *Patchwork*-Verfahren. Das *Patch-Work*-Verfahren ist ein statistisches Verfahren und basiert auf dem Test von Hypothesen. Das Verfahren geht auf die Arbeit von Bender et al. [BGML96] zurück. Hier wurde das Verfahren für Bild-Wasserzeichen in der Pixel-Domain vorgeschlagen. Für den Bereich der Audio-Wasserzeichen erfolgte eine effiziente Adaption mit den Arbeiten von Arnold [Arn00, Arn04]. Weitere Variationen des Verfahrens für Audiodaten [YK03, Cve04] erfolgten mit dem Ziel, die Transparenz und Robustheit zu verbessern.

Bedingung für die Entwicklung der Wasserzeichentechnik ist eine für das Inhaltsmerkmal transparente Modifikation der Audiodaten. Für den vorliegenden Fall ist es hinreichend, die Elemente $d_{j,i}$ der Matrix $\mathbf{D}$, also die Summen der Koeffizientenbeträge der Frequenzgruppen $\mathbf{c}_i^{Fj}$, zu erhalten. Eine einfache Möglichkeit, dies zu erreichen, ist die Einbettung multipler 1-Bit-Wasserzeichen unter Verwendung der Koeffizienten je eines Vektors $\mathbf{c}_i^{Fj}$ als Population $\mathbb{C} = \{c_1, \ldots, c_n\}$ der einzelnen 1-Bit-Wasserzeichen.

Für die Einbettung eines einzelnen Wasserzeichens werden die Elemente der zugehörigen Population $\mathbb{C}$ pseudozufällig mittels eines Saatwertes $\sigma$ in die Gruppen $\mathbb{A} = \{a_1, \ldots, a_l\}$ und

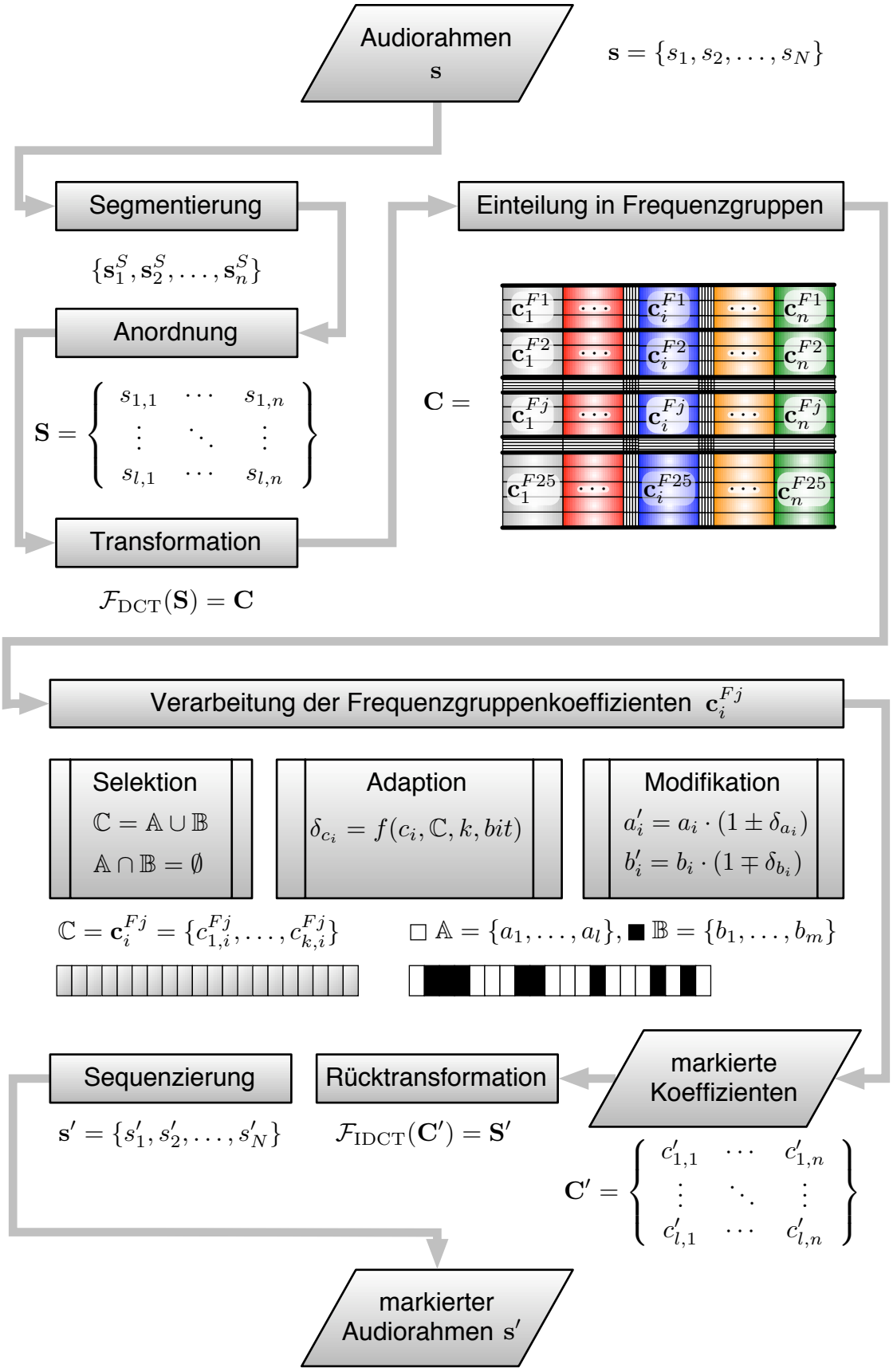


**Abbildung 4.5.:** Ablaufplan der entwickelten Wasserzeicheneinbettung

$\mathbb{B} = \{b_1, \ldots, b_m\}$ unter der Bedingung $\mathbb{A} \cap \mathbb{B} = \emptyset$ selektiert. Die vom Saatwert abhängige Selektion ermöglicht die Generierung identischer Koeffizietengruppen $\mathbb{A}$ und $\mathbb{B}$ in der Schutz- und Verifikationsphase.

Ziel der Einbettung ist die Schaffung einer Distanz $\varepsilon$ entsprechend (4.13) zwischen den Gruppen $\mathbb{A}'$ und $\mathbb{B}'$. Das Vorzeichen des Wertes $\varepsilon$ wird durch das einzubettende Wasserzeichen-Bit $w \in \{0,1\}$ nach (4.14) bestimmt. Der Betrag von $\varepsilon$ ist in Hinsicht auf die Wasserzeichen-Transparenz adaptiv an die Summe der Koeffizientenbeträge der Population $\mathbb{C}$ angepasst. Die Stärke der Einbettung kann mit dem Faktor $f$ (4.15) festgelegt werden.

$$\mathbb{C} = \mathbb{A} \cup \mathbb{B}; \quad \mathbb{A} \cap \mathbb{B} = \emptyset \tag{4.10}$$

$$\mathbb{A} = \{a_1, \ldots, a_l\}; \quad \mathbb{B} = \{b_1, \ldots, b_m\}; \quad \mathbb{C} = \{c_1, \ldots, c_n\} \tag{4.11}$$

$$w \in \{0,1\} \tag{4.12}$$

$$\sum_{i=1}^{l} |a'_i| - \sum_{i=1}^{m} |b'_i| = \varepsilon \tag{4.13}$$

$$\varepsilon = (-1)^w \cdot \sum_{i=1}^{n} |c_i| \cdot f \tag{4.14}$$

$$(0,1] := \{f \in \mathbb{R} \,|\, 0 < f \leq 1\} \tag{4.15}$$

Die Zielwerte der modifizierten Mengen $\mathbb{A}' = \{a'_1, \ldots, a'_l\}$ und $\mathbb{B}' = \{b'_1, \ldots, b'_m\}$ ergeben sich nach (4.16). Der Statistik nach können die Eigenschaften der Mengen $\mathbb{A}$ und $\mathbb{B}$ für die gegebenen Selektionskriterien, eine zufällige Selektion ohne gemeinsame Elemente aus der Population $\mathbb{C}$, für ausreichend große Populationen $\mathbb{C}$ als ähnlich betrachtet werden. Bei einer vollständigen Selektion aller Elemente aus der Population $\mathbb{C}$ kann die Summe der Koeffizientenbeträge einer Menge $\mathbb{A}$ bzw. $\mathbb{B}$ durch die Summe der Koeffizientenbeträge der Population $\mathbb{C}$ nach (4.17) ersetzt werden. Durch die Verwendung der Population $\mathbb{C}$ anstelle der Mengen $\mathbb{A}$ bzw. $\mathbb{B}$ zur Bestimmung der Zielwerte der modifizierten Mengen $\mathbb{A}'$ und $\mathbb{B}'$ in (4.16) ist die Berechnungsvorschrift allgemeingültig für den kompletten Wertebereich (4.15) der Einbettungsstärke $f$.

$$\sum_{i=1}^{l} |a'_i| = \frac{1}{2}\left(\sum_{i=1}^{n} |c_i| + \varepsilon\right) \qquad \sum_{i=1}^{m} |b'_i| = \frac{1}{2}\left(\sum_{i=1}^{n} |c_i| - \varepsilon\right) \tag{4.16}$$

$$\sum_{i=1}^{l} |a'_i| \approx \sum_{i=1}^{m} |b_i| \approx \frac{1}{2}\sum_{i=1}^{n} |c_i| \tag{4.17}$$

Ausgehend von (4.16) ergibt die Summe der Koeffizientenbeträge beider Mengen $\mathbb{A}$ und $\mathbb{B}$ nach der Modifikation die Summe (4.18) der unmodifizierten Koeffizientenbeträge der Population $\mathbb{C}$. Da die Population $\mathbb{C}$ von den Koeffizienten des Vektors $\mathbf{c}_i^{Fj}$ gebildet wird, bleibt somit auch

der aus den Elementen des Vektors $\mathbf{c}_i^{Fj}$ berechnete $d_{j,i}$-Wert unverändert und in weiterer Folge das Inhaltsmerkmal. Solange die Population $\mathbb{C}$ ausschließlich aus den Koeffizienten eines Vektors $\mathbf{c}_i^{Fj}$ gebildet wird, ist die Wasserzeicheneinbettung transparent für das Inhaltsmerkmal.

$$\sum_{i=1}^{l} |a_i'| + \sum_{i=1}^{m} |b_i'| = \frac{1}{2}\left(\sum_{i=1}^{n} |c_i| + \varepsilon\right) + \frac{1}{2}\left(\sum_{i=1}^{n} |c_i| - \varepsilon\right) = \sum_{i=1}^{n} |c_i| \tag{4.18}$$

Die Modifikation der Koeffizienten $a_i$ erfolgt in Abhängigkeit ihres Beitrags zur Summe der Koeffizientenbeträge $\sum_{i=1}^{n} |a_i|$ der zugehörigen Gruppe $\mathbb{A}$. Ein modifizierter Koeffizient $a_i'$ ergibt sich für die Wasserzeicheninformation $w$ nach (4.19).

$$a_i' = a_i \cdot \left(1 + \frac{a_i}{\sum_{i=1}^{l} |a_i|} \cdot \left(\frac{1}{2}\left(\sum_{i=1}^{n} |c_i| + \varepsilon\right) - \sum_{i=1}^{l} |a_i|\right)\right) \tag{4.19}$$

Für die Koeffizienten der Gruppe $\mathbb{B}$ gilt in Analogie:

$$b_i' = b_i \cdot \left(1 + \frac{b_i}{\sum_{i=1}^{m} |b_i|} \cdot \left(\frac{1}{2}\left(\sum_{i=1}^{n} |c_i| - \varepsilon\right) - \sum_{i=1}^{m} |b_i|\right)\right) \tag{4.20}$$

Im Wasserzeichen-Decoder wird die empfangene Wasserzeicheninformation $w'$ über den Vergleich der Summen der Koeffizientenbeträge der Gruppen nach (4.21) gewonnen.

$$w' = \begin{cases} 0 & \text{für} \quad \varepsilon' = \sum |a_i'| - \sum |b_i'| \geqq 0 \\ 1 & \text{für} \quad \varepsilon' = \sum |a_i'| - \sum |b_i'| < 0 \end{cases} \tag{4.21}$$

## 4.4. Leistungsanalyse der Systemelemente

Im Vorfeld der Leistungsanalyse wird die Testmenge an zulässigen und unzulässigen Störungen festgelegt. Die Bewertung der Zulässigkeit erfolgt anhand der Art der Störung und deren Einfluss auf die Audiodatenqualität. Ein geeigneter Frequenzgruppenbereich für die Merkmalsextraktion wird identifiziert. Die Differenzierungsfähigkeit des entwickelten Inhalsmerkmals wird in Abhängigkeit von der Segmentlänge $l$ analysiert. Die Analyse der Robustheit erfolgt gegenüber zulässigen und qualitätsmindernden Störungen. Die Untersuchung der Manipulationssicherheit des Merkmales erfolgt durch Ersetzung von Audiopassagen unterschiedlicher Länge.

## 4.4.1. Testdaten, Datenqualität und Störungen

### 4.4.1.1. Testdaten

Um eine objektive, nachvollziehbare und mit anderen Forschungsergebnissen vergleichbare Evaluierung durchzuführen, ist ein gemeinsamer und öffentlicher Testdatensatz notwendig. In den Bereichen der Video- und Bild-Wasserzeichen haben sich solche Sätze an Testdaten etabliert. Hier sind die *Video Trace Library* [ASU] der *Arizona State University* und die *USC-SIPI Image Database* [USC] der *University of Southern California* zu nennen. Im Bereich der Audio-Wasserzeichen ist kein geeigneter, öffentlicher Audiotestdatensatz bekannt. Im Forschungsbereich der verlustbehafteten Audiokompression wird häufig der „*Sound Quality Assessment Material recordings for subjective tests*"-Testsatz [EBU08, EBU] verwendet. Die Audiodateien des Testsatzes sind sehr kurz und die spektrale Verteilung der Aufnahmen ist begrenzt. Qualitativ hochwertiges Audiomaterial ist in Form von instrumental begleitetem Gesang, rein instrumentalen Stücken und Instrumentenklang mit der RWC *(Real World Computing) Music Database* [Got04, RWC] erhältlich. Der Einsatzbereich der Daten ist auf Forschungszwecke beschränkt und wird nur kostenpflichtig verteilt. Hochqualitative Sprachaufnahmen sind in der *TSP Speech Database* [EBU08, EBU] zu finden, die Nutzungsrechte sind hier jedoch unklar. Eine Alternative ist die Verwendung von Datenbanken, deren Inhalte zur freien Nutzung und Verbreitung freigegeben sind (z.B. *Internet Archive* [ia], *The Freesound Project* [fsp]). Die verfügbaren Aufzeichnungen sind jedoch begrenzt in Anzahl, Genre und Qualität. Eine umfangreiche und frei nutzbare Audiotestdatenbank ist noch immer eine offene Herausforderung für die Audio-Wasserzeichen-Forschungsgemeinschaft sowie für andere Multimedia-Forschungsbereiche.

Der in dieser Arbeit verwendete Testdatensatz besteht aus 130 Audiodateien (48 kHz, 16 bit, mono) mit einer Abspieldauer von je 61,44 Sekunden. Die Abspieldauer der einzelnen Audiodateien orientiert sich an etwa einer Minute, wobei die Anzahl der Samples ein Vielfaches der untersuchten Segmentlängen $l$ darstellt. Die Audiodaten werden hierdurch für alle während der Leistungsanalyse verwendeten Segmentlängen $l$ vollständig verwendet. Die Audiodaten stammen aus 13 unterschiedlichen Quellen (s. Anhang A.1), bestehend aus Hörspielen, Buch- und Lyriklesungen, „Comedy" und Rundfunkbeiträgen.

### 4.4.1.2. Bewertung der Datenqualität

Die Bewertung der Datenqualität hat bei der Evaluation von inhalts-fragilen Wasserzeichenverfahren zwei Aspekte. Der erste Aspekt ist die Bewertung der Beeinträchtigung der Audiodaten durch das Aufbringen des Wasserzeichens. Der zweite Aspekt ist die Klassifikation von

Störungen als zulässig bzw. unzulässig mittels der Bewertung der Datenqualität nach Störungen.

Für die Bewertung der Datenqualität gibt es zwei Ansätze: Die subjektive und die objektive Bewertung. Die subjektive Bewertung erfolgt in Form eines Hörtests. Die Bewertungsergebnisse sind abhängig von Bedingungen des Hörtests, wie Testumgebung, Größe und Zusammenstellung der Testgruppe. Vergleichbare und reproduzierbare Bewertungsergebnisse erfordern einheitliche Bedingungen. Empfehlungen für die Durchführung von Hörtests wurden von der *International Telekommunication Union* [Int03] veröffentlicht. Die Bewertung der Datenqualität wird mit einer 5-stufigen Skala, dem *Subjective Difference Grade* (SDG), vorgenommen. Das SDG entspricht der Differenz aus der Bewertung des Test- und des Referenzsignals. Die Bewertungsskala reicht von 0 (nicht wahrnehmbar) bis 5 (sehr störend). Für das SDG ergibt sich bei richtiger Zuordnung des Referenzsignales ein Wert zwischen 0 und -4 (s. Tabelle 4.1). Die subjektive Bewertung ist die beste Methode, um die Datenqualität zu bewerten, jedoch auch die teuerste und zeitaufwendigste.

Objektive Methoden sind weit weniger teuer und wesentlich schneller. Ein häufig verwendetes Bewertungsmass ist das Signal-zu-Rausch-Verhältnis (engl. *Signal-to-Noise-Ratio* - SNR) (SRV). Das SRV ist eine einfache Methode, nimmt jedoch keinen Bezug auf die menschliche Wahrnehmung. Um bei der objektiven Bewertung einen besseren Bezug auf die menschliche Wahrnehmung zu nehmen, wurden verschiedene Metriken entwickelt: „*Noise-to-Mask Ratio*“ (NMR) [Bra87], „*Perceptual Audio Quality Measure*“ (PAQM) [Bee92], „*Perceptual Evaluation*“ (PERCEVAL) [PMMS92] , „*Perceptual Objective Measure*“ (POM) [CLR$^+$93], „*Disturbance Index*“ (DIX) [Thi96], „*Objective Audio Signal Evaluation*“ (OASE) [Spo97]. Auf Grundlage dieser Metriken wurde „*Perceptual Evaluation of Audio Quality*“ (PEAQ) [TBS$^+$98, peaq] entwickelt. Die Bewertung erfolgt hier, angepasst an das SDG mittels einer 5-stufigen Skala, siehe Tabelle 4.1. In dieser Arbeit wird die Implementierung des PEAQ-Algorithmus [Kab] von der *McGill University* verwendet.

**Tabelle 4.1.:** Bewertungsskala ITU-R Rec. 1284-1 [Int03], *Subjective Difference Grade* (SDG) und *Objective Difference Grade* (ODG)

| ITU-R | SDG/ODG | Bewertung |
|---|---|---|
| 5.0 | 0 | nicht wahrnehmbar |
| 4.0 | -1.0 | wahrnehmbar, aber nicht störend |
| 3.0 | -2.0 | leicht störend |
| 2.0 | -3.0 | störend |
| 1.0 | -4.0 | sehr störend |

#### 4.4.1.3. Zulässige und unzulässige Störungen

Für die Analyse der Leistungsfähigkeit des Inhaltsmerkmals und der Wasserzeichentechnik wurde eine Auswahl von neun Störungen mit je drei Parametereinstellungen festgelegt. Der erste Teil der Störungen, bestehend aus verlustbehafteter Audiodatenkompression, Unterabtastung und Lautstärkeveränderung, repräsentiert die für Verbreitung und Archivierung von Audiodaten typische Operationen. Die verlustbehaftete Audiodatenkompression wird verwendet, um digitale Audiodaten effektiv in ihrer Größe zu reduzieren mit der Prämisse, die Datenqualität zu erhalten. Die Unterabtastung von Audiodaten dient häufig der Anpassung des Datenformates an Verarbeitungssysteme und einer einfachen Form der Datenreduzierung. Die Lautstärkeanpassung von Audiodaten wird verwendet, um einen bestimmten Dynamikumfang der Amplitudenwerte einzustellen (Normalisierung) oder, um die Lautheit von Audiodateien anzugleichen. Die Datenqualität bleibt bei diesen Operationen erhalten. Diese Arten der Störung können allgemein als zulässig betrachtet werden. Mit zunehmender Intensität der Störung kann die Datenqualität wahrnehmbar störend beeinflusst werden. Eine genauere Einstufung dieser Störungen wird neben der Art der Störung im Anschluss über die Degradierung der Audioqualität vorgenommen.

- verlustbehaftete AAC-Kompression (*Advanced Audio Codec*[2])
  Parameter: Qualitätsfaktor $Q = 40, 100, 300$
- verlustbehaftete MP3-Kompression (*MPEG-1 Audio Layer 3*[3])
  Parameter: Bitrate 32 $kBit/s$, 64 $kBit/s$, 128 $kBit/s$
- verlustbehaftete Ogg Vorbis-Kompression (*Ogg Vorbis* (OGG)[4])
  Parameter: Bitrate 32 $kBit/s$, 64 $kBit/s$, 128 $kBit/s$
- Unterabtastung (engl. *subsampling*)
  Parameter: Samplerate 44,1 kHz, 32 kHz, 16 kHz
- Lautstärkeveränderung
  Parameter: Verstärkungsfaktor $f = 0,67$; $1,5$ und normalisiert

Der zweite Teil repräsentiert unbestimmte, qualitätsmindernde Störungen der Audiodaten. Hierfür werden rauschartige Störungen mit unterschiedlicher Charakteristik und Eindruck auf die menschliche Wahrnehmung verwendet. Die verwendeten Rauscharten sind weißes, rosa, braunes und dynamisches Rauschen. Das weiße Rauschen wird trotz konstanter Leistungsdichte des Frequenzspektrums subjektiv als Signal wahrgenommen, dessen Amplitude mit der Frequenz

[2] AAC-Kompression wurde mittels des „*Freeware Advanced Audio Coder*“ [faac] realisiert.
[3] MP3-Kompression wurde mittels „*LAME*“ [lam06] realisiert.
[4] Ogg Vorbis-Kompression wurde mittels „*oggenc*“ (Teil der „*vorbis-tools*“) [vor] realisiert.

zunimmt. Das Rosa Rauschen wird über den Frequenzbereich gleich als laut wahrgenommen. Der Frequenzgang verläuft bei Rosa Rauschen umgekehrt proportional zur Frequenz ( $1/f$). Die Leistungsdichte des Frequenzspektrums nimmt um 3 dB pro Oktave mit steigender Frequenz ab. Braunes Rauschen wird über den Frequenzbereich mit abnehmender Intensität wahrgenommen. Der Frequenzgang verläuft bei Braunem Rauschen umgekehrt proportional zum Quadrat der Frequenz ( $1/f^2$). Die Leistungsdichte des Frequenzspektrums nimmt um 6 dB pro Oktave mit steigender Frequenz ab. Das dynamische Rauschen stellt weißes, mit dem Audiosignal gewichtetes, gleichverteiltes Rauschen dar. Im Gegensatz zu den anderen drei Rauscharten bleibt das Signal-Rausch-Verhältnis konstant.

- Weißes gleichverteiltes Rauschen
  Parameter: Verstärkungsfaktor -30 dB, -40 dB, -50 dB

- Rosa Rauschen, auch $1/f$ Rauschen
  Parameter: Verstärkungsfaktor -30 dB, -40 dB, -50 dB

- Braunes Rauschen, auch $1/f^2$ Rauschen
  Parameter: Verstärkungsfaktor -30 dB, -40 dB, -50 dB

- Dynamisches Rauschen
  Parameter: Vertärkungsfaktor -20 dB, -30 dB, -40 dB

Zur Einstufung der Störungen als zulässig bzw. unzulässig wird als Metrik für die Bewertung der Audioqualität das *Objective Difference Grade (ODG)* verwendet. Störungen mit einem ODG-Wert größer -1 (wahrnehmbar, aber nicht störend) werden als zulässig angesehen. Störungen mit einem Wert kleiner -2 (leicht störend) werden als unzulässig eingestuft. Der restliche Bereich zwischen den zulässigen und unzulässigen Störungen stellt einen Graubereich dar (vgl. Abbildung 4.2 aus Abschnitt 4.2.1), in den die Trennungsschwelle gelegt werden soll. Abbildung 4.6 zeigt die Ergebnisse der Qualitätsbewertung der gestörten Audiodaten. Nach starker Kompression nimmt die Datenqualität deutlich ab. Diese Störungen werden folglich als unzulässig gewertet. Leicht und mittlere Kompression, Unterabtastung und Lautstärkeveränderungen liegen im Bereich von nicht bis leicht störend wahrnehmbar und können somit weiterhin als zulässig betrachtet werden. Die Auswirkungen der rauschartigen Störungen variieren für unterschiedliche Audiodaten stark. Je nachdem, ob Audiodaten schon einen gewissen Rauschanteil besitzen, wirkt sich das Aufbringen von weiterem Rauschen unterschiedlich stark auf die wahrnehmbare Datenqualität aus.

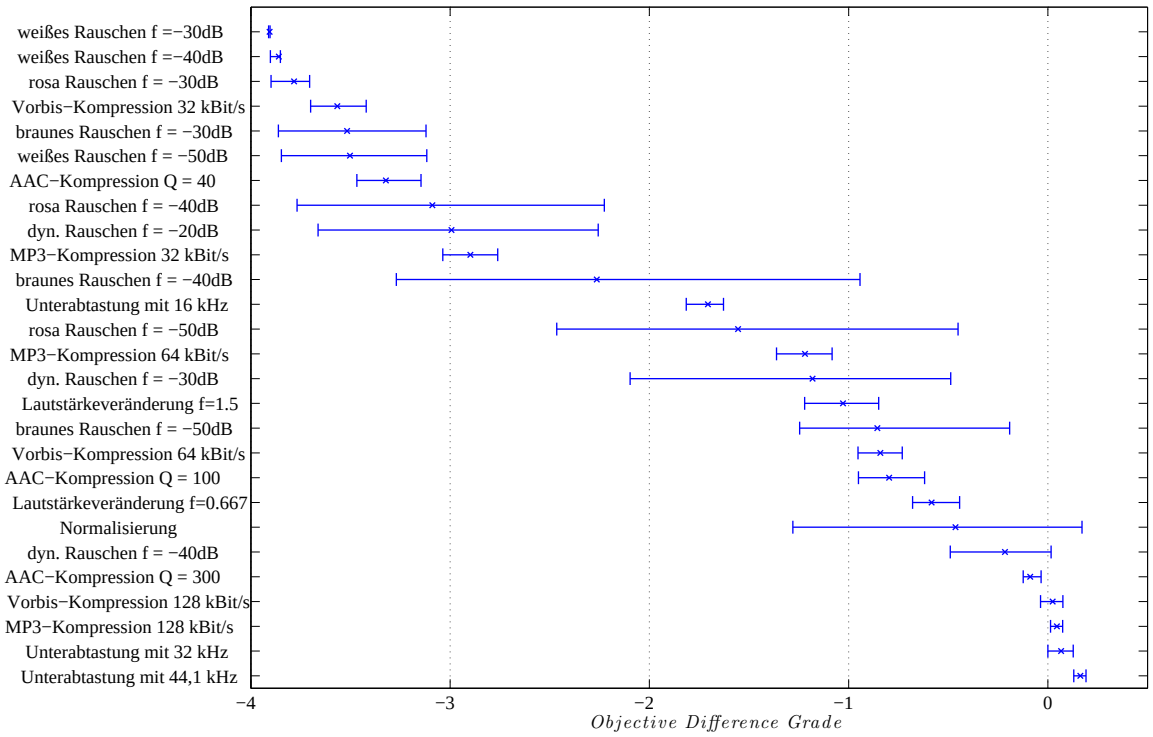


**Abbildung 4.6.:** Degradierung der Audioqualität durch Störungen

### 4.4.2. Robustheit des Inhaltsmerkmals

Der Extraktionsbereich des Inhaltsmerkmals und somit dessen Eigenschaften wird durch die Segmentlänge $l$ und die Wahl der Frequenzgruppen bestimmt. Die Segmentlänge bestimmt die zeitliche Auflösung des Merkmals. Die Auswahl der Frequenzgruppen legt den Frequenzbereich fest, auf den das Inhaltsmerkmal sensibel reagieren soll. Die Abbildung 4.7 zeigt einen Auszug der im Anhang A.2.1 aufgeführten Analyse des Einflusses der verwendeten Frequenzgruppe auf die Robustheit des Inhaltsmerkmals. Die aufgeführten Störungen sind auf zulässige Störungen bzw. Störungen im Graubereich begrenzt. Die dargestellte Bitfehlerrate repräsentiert den Mittelwert über alle Realisierungen mit Segmentlängen von 512 - 32 768 Samples.

Es ist zu erkennen, dass die Robustheit des Inhaltsmerkmals nach Kompression und Unterabtastung aufgrund des Tiefpass-Charakters dieser Störungen für sehr hohe Frequenzen deutlich abnimmt. Auch die erste Frequenzgruppe hebt sich durch eine tendenziell höhere Bitfehlerrate hervor. In Bezug auf die Robustheit des Inhaltsmerkmals gegenüber zulässigen Störungen aber auch in Bezug auf die Datenrate des Merkmals ist eine Eingrenzung der verwendeten Frequenzgruppen sinnvoll. Ausgehend von den Simulationsergebnissen wird für die weitere Analyse der Extraktionsbereich des Inhaltsmerkmals auf die Frequenzgruppen $F2$ bis $F23$ festgelegt. Dies

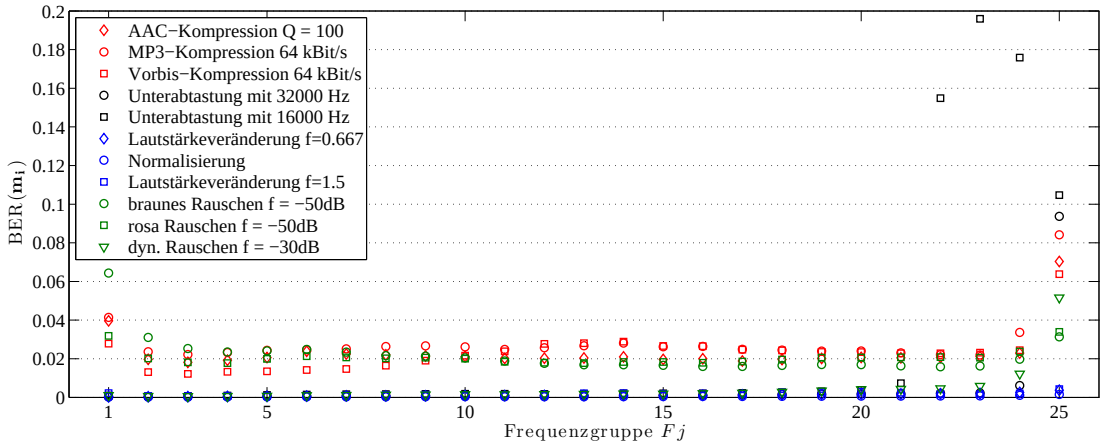


**Abbildung 4.7.:** Bitfehlerrate der Merkmalsinformation $\mathbf{m_i}$ in Abhängigkeit von der Frequenzgruppe $Fj$ nach Störung der Audiodaten. (Auszug der Simulationsergebnisse aus Anhang A.2.1.)

entspricht einem Frequenzbereich von $0,1 - 12$ kHz und deckt somit den Hörbereich für Sprache von $0,1 - 10$ kHz [LSS+07, Seite 51] ab.

Abbildung 4.8 enthält einen Auszug der im Anhang A.2.2 aufgeführten Analyse des Einflusses der Segmentlänge $l_M$ auf die Robustheit des Inhaltsmerkmals. Dargestellt ist die Bitfehlerrate der Merkmalsvektoren $\mathbf{m_i}$ nach zulässigen Störungen bzw. Störungen des Graubereichs. Die Robustheit des Inhaltsmerkmals nimmt mit steigender Segmentgröße $l_M$ zu. Die rauschartigen Störungen sind Zufallsprozesse. Die Eigenschaften zweier gleichartiger Rauschprozesse gleichen sich mit zunehmender Beobachtungsdauer an. Da ein Inhaltsmerkmal aus zwei Segmenten gewonnen wird, nimmt der Einfluss für die rauschartigen Störungen mit steigender Segmentlänge ab. Für die anderen Störungen läßt sich dieses Verhalten dadurch erklären, dass sich die Eigenschaften der Segmente der Audiodatei für kurze Betrachtungszeiträume stark unterscheiden, mit größer werdenden Bertrachtungszeiträumen jedoch angleichen. Da die Störungen durch den Dateninhalt beeinflusst werden, gleichen sich deren Eigenschaften an und der Einfluss auf das Inhaltsmerkmal nimmt ab.

Für ähnliche Störungen zeigt sich ein ähnlicher Einfluss auf das Inhaltsmerkmal. Dies ist besonders deutlich für die unterschiedlichen Kompressionsverfahren oder für das Braune und das Rosa Rauschen. Lautstärkeveränderung, Unterabtastung und dynamisches Rauschen verursachen nur geringe Störungen des Inhaltsmerkmals. Für die Unterabtastung mit 16 kHz ergibt sich ein konstantes Fehlerverhalten. Dies ist durch den Informationsverlust der Bestandteile des Inhaltsmerkmals oberhalb von 8 kHz begründet.

Eine vollständige Robustheit des Inhaltsmerkmals gegenüber zulässigen Störungen wird nicht erreicht. Die Differenzierung von zulässigen und unzulässigen Störungen kann jedoch über eine Schwellwertentscheidung der Bitfehlerrate der Merkmalsvektoren geführt werden. Ab einer

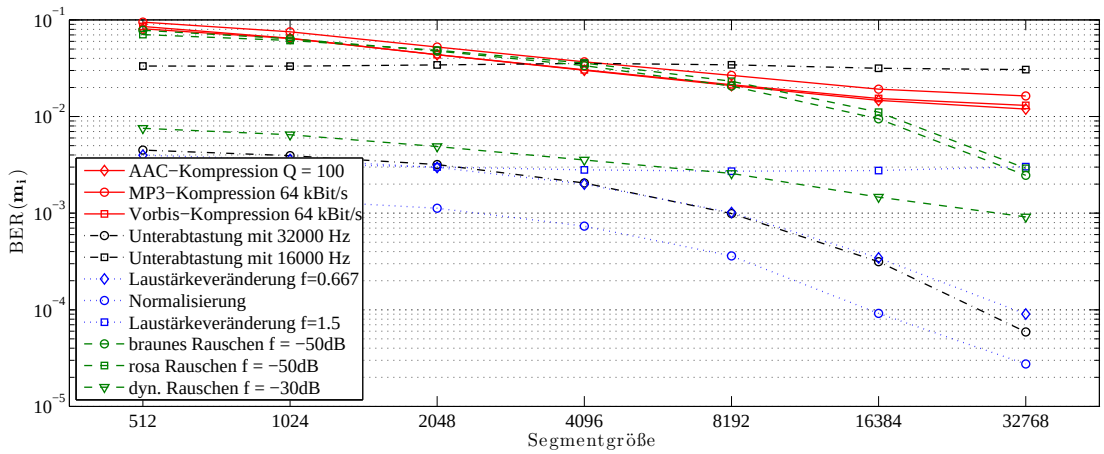


**Abbildung 4.8.:** Bitfehlerrate der Merkmalsvektoren $\mathbf{m_i}$ nach Störung von Audiopassagen. (Auszug der Simulationsergebnisse aus Anhang A.2.2.)

Segmentlänge von 4 096 Samples liegen die zulässigen Störungen unterhalb einer Bitfehlerrate von 5%. Abbildung 4.9 zeigt die Bitfehlerraten der Merkmalsvektoren $\mathbf{m_i}$, aufgeschlüsselt nach zulässigen Störungen, unzulässigen Störungen und Störungen des Graubereichs. Die Störungen grenzen sich gut voneinander ab. Die unzulässigen Störungen weisen höhere Bitfehlerraten als die zulässigen Störungen auf. Eine Ausnahme bildet das dynamische Rauschen, welches auch bei starkem Einfluss auf die Datenqualität nur geringe Störungen des Merkmales bewirkt. Für große Segmentlängen nähern sich die Bitfehlerraten der zulässigen und unzulässigen Störungen weiter an und überlappen sich teilweise. Als Arbeitsbereich für eine schwellwertbasierte Trennung der Störungen sind Segmentlängen unterhalb von 8 192 Samples sinnvoll.

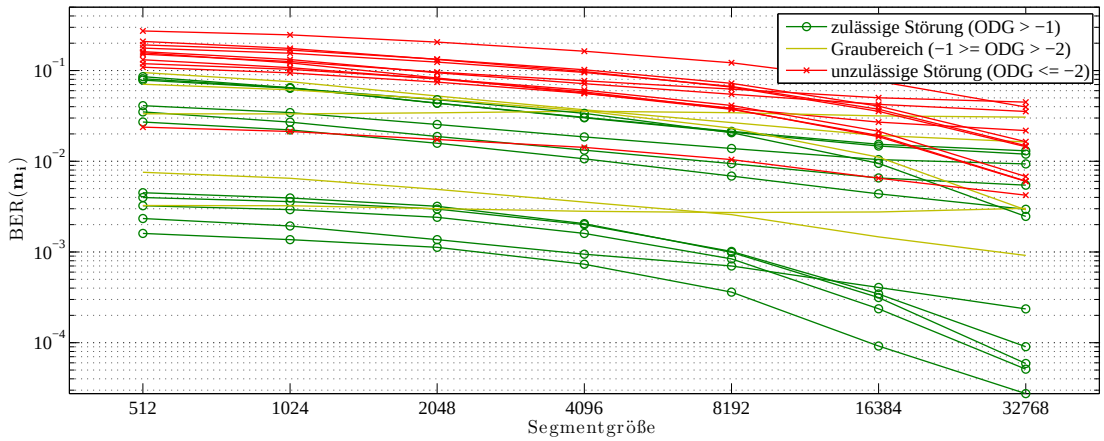


**Abbildung 4.9.:** Vergleich der Bitfehlerraten der Merkmalsvektoren $\mathbf{m_i}$ nach zulässigen und unzulässigen Störungen. (vgl. Simulationsergebnisse Anhang A.2.2.)

### 4.4.3. Manipulationssicherheit des Inhaltsmerkmals

Die Simulation von Veränderungen des Dateninhaltes erfolgt durch zwei Methoden. Die erste Methode ist der Austausch von Audiopassagen. Die zu ersetzenden Audiopassagen werden hierbei zufällig aus derselben Audiodatei gezogen, in der die Ersetzung stattfinden soll, um eine möglichst hohe Ähnlichkeit der Charakteristik der originalen und der ersetzenden Audiopassage zu erzielen. Die zweite Methode stellt die Ersetzung von Audiopassagen mit Stille dar. Beide sollen eine inhaltliche Veränderung der Audiodaten simulieren. Von einer manuelle Ersetzung mit Erstellung eines neuen sinnvollen Inhaltes wurde aufgrund des hohen Aufwandes abgesehen.

Die Störungslängen bewegen sich im Bereich von 1 024 bis 32 768 Samples (21,3 bis 682,67 ms). Laut Zwicker [ZF99] muss ein Reiz bzw. eine Störung 20 ms andauern, um wahrnehmbar zu sein. Die Größenordnung von inhaltsverändernden Störungen liegt höher. Die durchschnittliche Dauer einer Silbe liegt entsprechend einer Silbenfrequenz von 3-5 Hz [Ter98] im Bereich von 200 bis 333 ms. Die Abbildung 4.10 enthält einen Auszug der im Anhang A.2.3 aufgeführten Simulationsergebnisse zur Analyse der Bitfehlerrate des Inhaltsmerkmales nach inhaltsverändernden Störungen. Es ist ersichtlich, dass die Bitfehlerrate des Inhaltsmerkmals für Segmentlängen oberhalb der Störungslänge stark abnimmt. Für eine zuverlässige Erkennung von inhaltlichen Störungen muss die Segmentlänge der Merkmalsextraktion die Störungslänge unterschreiten.

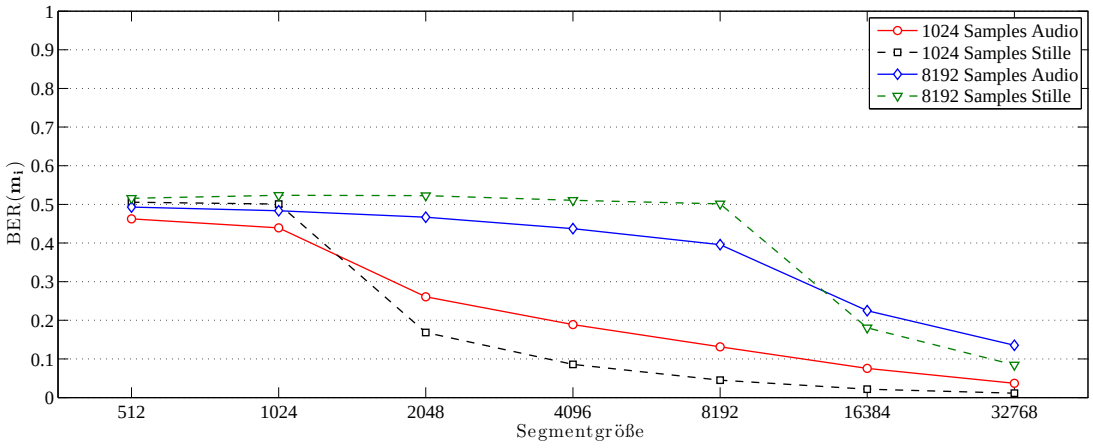


**Abbildung 4.10.:** Bitfehlerraten der Merkmalsvektoren $\mathbf{m_i}$ nach Ersetzen von Audiopassagen mit zufälligen Audiopassagen bzw. Stille. Die Störungslänge beträgt jeweils 1 024 Samples (21,3ms) bzw. 8 192 Samples (170,7ms). (Auszug der Simulationsergebnisse aus Anhang A.2.3.)

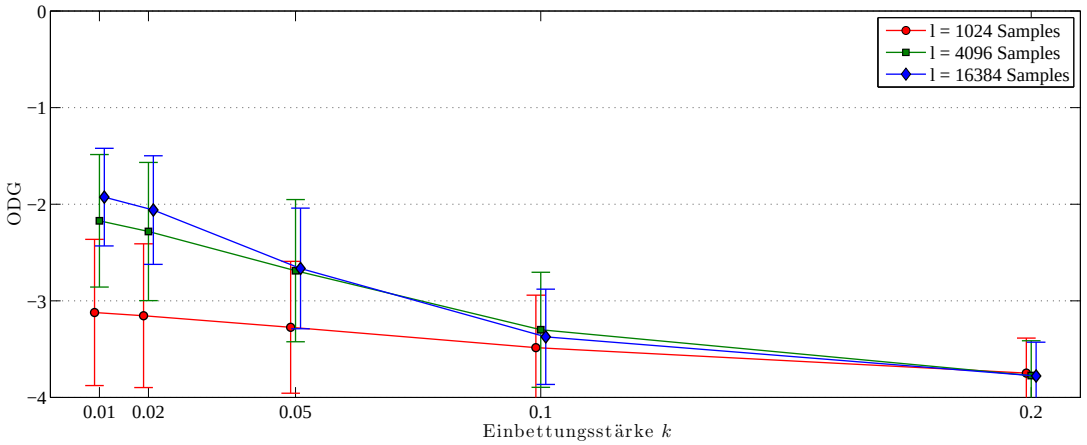


**Abbildung 4.11.:** Transparenz der Wasserzeicheneinbettung unter Einbeziehung aller Frequenzgruppen in die Einbettungsdomain.

### 4.4.4. Robustheit der Wasserzeichentechnik

Für die Leistungsanalyse der Wasserzeichentechnik wird ähnlich wie bei der Leistungsanalyse des Inhaltsmerkmals eine sukzessive Eingrenzung der Wasserzeichendomain auf geeignete Arbeitsbereiche vorgenommen. Bewertungsgrundlage hierfür sind die Wasserzeichen-Transparenz gemessen in ODG-Werten und die Robustheit gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen. Im Anschluss an die Eingrenzung günstiger Arbeitsbereiche wird die Robustheit der Wasserzeichentechnik auf Basis eines festgelegten Arbeitsbereiches analysiert.

#### 4.4.4.1. Analyse geeigneter Frequenzgruppen

Die Einbettungsdomain wird durch die für die Einbettung verwendeten Frequenzgruppen und die für die Segmentierung verwendete Segmentlänge $l_W$ bestimmt. Der limitierende Faktor der Einbettung stellt die Wassserzeichen-Transparenz dar. Im ersten Schritt der Leistungsanalyse wird untersucht, inwieweit die verfügbare Einbettungsdomain unter Betrachtung der Wasserzeichen-Transparenz genutzt werden kann.

Für eine erste Abschätzung wurde die Wasserzeichen-Transparenz unter Nutzung aller 25 Frequenzgruppen für die Wasserzeicheneinbettung ermittelt. Die Betrachtung erfolgt für die Segmentlängen von $1\,024, 4\,096$ und $16\,384$ Samples, sowie für Einbettungsstärken $f$ im Bereich von $0,01$ (schwache Einbettung) bis $0,2$ (starke Einbettung). Die resultierenden Wasserzeichen-Transparenzen der Analyse, gemessen in ODG-Werten, sind in Abbildung 4.11 dargestellt.

Ein vollständige Nutzung der verfügbaren Einbettungsdomain ist in Hinsicht der Wasserzeichen-Transparenz nicht möglich. Auch für geringe Einbettungsstärken $f$ liegen die ODG-Werte des Vergleichs der originalen Audiodaten mit den markierten Daten in einem Bereich, der als störend empfunden wird.

Für die Identifikation von geeigneten Frequenzgruppen wurde die Einbettung der Wasserzeichen separat für jede Frequenzgruppe durchgeführt. Die Betrachtung erfolgte für eine Einbettungsstärke von $f = 0,05$. Die Ergebnisse der Simulation sind in Abbildung 4.12 dargestellt.

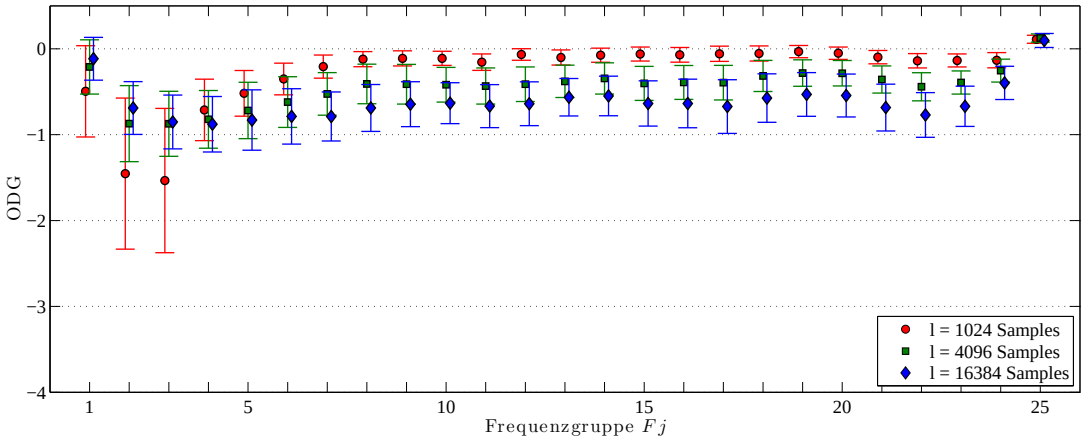


**Abbildung 4.12.:** Transparenz der Wasserzeicheneinbettung unter Einbezug einzelner Frequenzgruppen in die Einbettungsdomain.

Es zeichnen sich hier zwei wesentliche Aspekte ab, die entscheidend für die Transparenz der Wasserzeichen sind.

Mit steigender Segmentlänge nimmt die Transparenz ab. Audiosignale sind instationäre Signale. Die Eigenschaften der Störung, welche durch die Wasserzeicheneinbettung erzeugt werden, leiten sich durch den Einbettungsprozess aus dem Trägersignal ab. Die notwendigen Modifikationen für die Wasserzeicheneinbettung (s. Abschnitt 4.3.2) stellen eine Funktion mit einer Abhängigkeit vom Trägersignal dar. Werden für die Einbettung kleine Segmentlängen verwendet, bedingen sich die Eigenschaften der Störung aus den lokalen Signaleigenschaften des jeweiligen Segmentes. Die Signaleigenschaften und die Eigenschaften der Störung sind hier ähnlich, was zu einer guten Transparenz der Wasserzeichen führt. Mit steigender Segmentlänge nehmen die Eigenschaften der Störung einen globalen Signalcharakter an. Die Transparenz der Wasserzeichen nimmt ab, da die Ähnlichkeit zwischen den lokalen Eigenschaften des Audiosignals und den Eigenschaften der Störung mit globalem Signalcharakter abnimmt.

Der zweite Aspekt ergibt sich aus dem Einbrechen der Transparenz mit sinkender Segmentlänge und der Verwendung der unteren Frequenzgruppen. Die Einbettungsdomain eines Wasserzeichens muss über eine Mindestanzahl an Koeffizienten verfügen. Der Einbettungsprozess

basiert auf der Annahme, dass die Mengen $\mathbb{A}$ und $\mathbb{B}$, die aus der Population der Koeffizienten in der Einbettungsdomain gezogen werden, ähnliche Eigenschaften besitzen. Dies gilt jedoch nur für eine ausreichend große Anzahl an Koeffizienten. Bei einer unzureichend großen Anzahl an Koeffizienten weichen die Eigenschaften der Mengen $\mathbb{A}$ und $\mathbb{B}$ stärker voneinander ab. Die Transparenz der Wasserzeichen nimmt ab, da die für die Wasserzeicheneinbettung notwendigen Modifikationen der Koeffizienten zunehmen.

Ausgehend von den Simulationsergebnissen wird für weitere Untersuchungen die Einbettungsdomain auf die Frequenzgruppen $F15$ bis $F20$ (2 320 Hz bis 6 400 Hz) festgelegt. Dies stellt einen Kompromiss aus Transparenz und Robustheit dar. In Hinsicht der Wasserzeichen-Transparenz werden Frequenzgruppen mit einer möglichst hohen Anzahl an Koeffizienten gewählt. Die oberen Frequenzgruppen werden aus Gründen der Robustheit ausgeschlossen. Diese beinhalten zwar die meisten Koeffizienten, die Wasserzeicheninformation ist hier jedoch sensibler gegenüber Störungen mit Tiefpass-Charakter, wie zum Beispiel verlustbehaftete Kompression oder Unterabtastung (s. Anhang A.3.1).

#### 4.4.4.2. Analyse geeigneter Segmentlängen

Um eine Aussage über die Robustheit der Wasserzeichentechnik bei unterschiedlichen Segmentlängen vornehmen zu können, ist es notwendig, die Wasserzeichen-Transparenz auf einen einheitlichen Wert einzustellen. Die Wasserzeichen-Transparenz ist bei fester Eingbettungsdomain über die Einbettungsstärke $f$ einstellbar. Die Größe der Einbettungstärke $f$ zur Einstellung eines bestimmten ODG-Wertes ist neben der Einbettungsdomain (Frequenzgruppen, Segmentlänge) auch von den Audiodaten abhängig. Je nach Charakteristik der Audiodaten kann die Einbettungstärke $f$ bei gleichem ODG-Wert stark variieren. Die Einstellung eines bestimmten ODG-Wertes ist aufgrund der starken Streuung der Einbettungsstärke in Abhängigkeit der Audiodaten sehr rechenaufwendig. Die Einstellung der Einbettungsstärke $f$ wurde durch die Berechnung von Stützstellen und anschließender linearer Interpolation zwischen den am ODG-Zielwert nächstgelegenen Stützstellen vorgenommen. Als Stützstellen wurden Einbettungsstärken $f$ zwischen 0,3 und 0,001 gewählt. Die mit dieser Methode ermittelten Einbettungsstärken sind in Abbildung 4.13 dargestellt. Die erzielten ODG-Werte der Wasserzeichentransparenz sind in Abbildung 4.14 abgebildet. Abbildung 4.15 zeigt einen Auszug der in Anhang A.3.2 aufgeführten Ergebnisse der Robustheitsanalyse der Wasserzeicheninformation bei einer Wasserzeichentransparenz von ODG=-1 in Abhängigkeit der Segmentlänge $l$ unter Verwendung der Frequenzgruppen $F15$-$F20$.

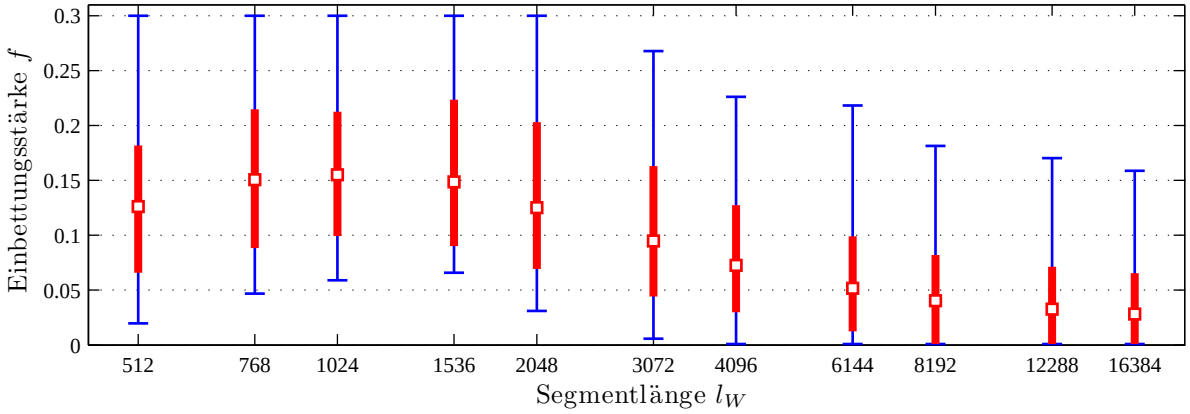


**Abbildung 4.13.:** Wertebereich I, Bereich zw. dem 15,85%- und 84,15%-Fraktil ■ und Mittelwert □ der Einbettungsstärke $f$ zur Einstellung einer Wasserzeichentransparenz von $ODG \approx -1$ in Abhängigkeit der Segmentlänge $l_W$ unter Verwendung der Frequenzgruppen $F15$-$F20$.

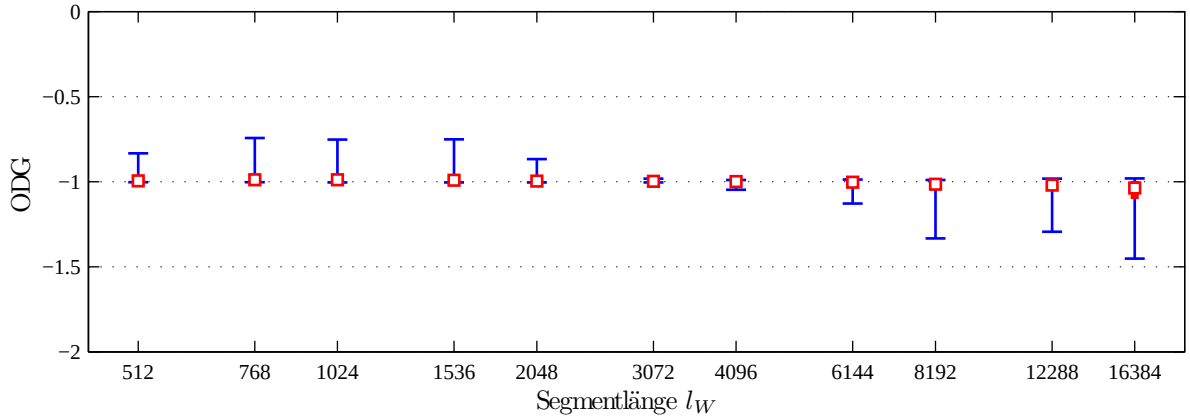


**Abbildung 4.14.:** Wertebereich I, Bereich zw. dem 15,85%- und 84,15%-Fraktil ■ und Mittelwert □ der erzielten ODG-Werte der Wasserzeichentransparenz in Abhängigkeit der Segmentlänge $l_W$ unter Verwendung der Frequenzgruppen $F15$-$F20$.

### 4.4.5. Fazit

- Das Inhaltsmerkmal ist geeignet, zulässige und unzulässige Störungen mittels einer Schwellwertentscheidung zu unterscheiden.
- Die Segmentlänge der Merkmalsextraktion hat direkten Einfluss auf Robustheit/Sensibilität des Inhaltsmerkmals gegenüber Störungen der Audiodaten.
- Für eine zuverlässige Erkennung von Störungen muss die Segmentlänge der Merkmalsextraktion unterhalb der Störungsdauer liegen.
- Die Transparenz der Wasserzeichentechnik nimmt mit größer werdenden Segmentlängen ab.

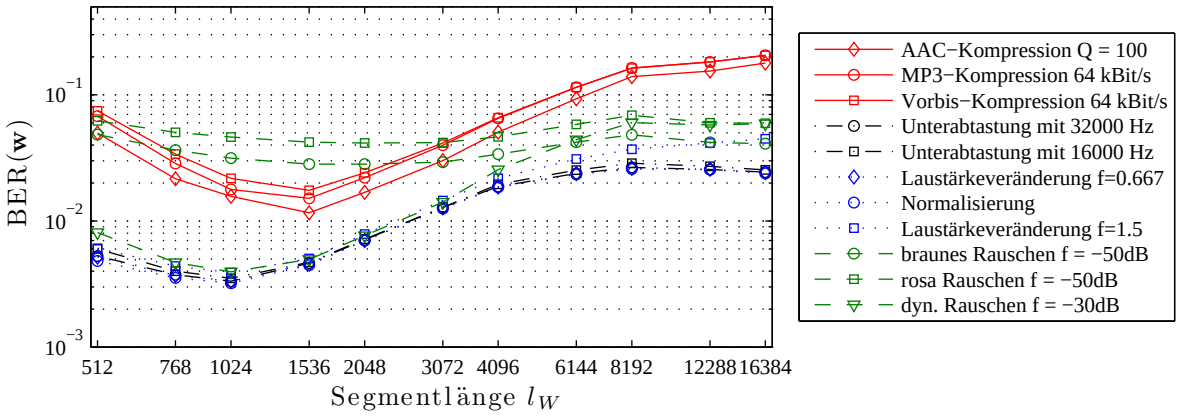


**Abbildung 4.15.:** Bitfehlerrate der Wasserzeicheninformation bei einem ODG von -1 in Abhängigkeit der Segmentlänge $l_w$ unter Verwendung der Frequenzgruppen $F15$-$F20$.

- Eine unzureichend große Anzahl an Koeffizienten im Einbettungsbereich vermindert die Transparenz der Wasserzeicheneinbettung.
- Die Robustheit der Wasserzeichentechnik besitzt ein lokales Minimum im Bereich einer Segmentlänge von 1 024 - 1 536 Samples.

## 4.5. Kombination von Merkmalsextraktion und Wasserzeichentechnik

Ein Entwicklungsziel des Wasserzeichenverfahrens ist eine für das Inhaltsmerkmal störungsfreie Einbettung des Wasserzeichens. Hierfür wurde der Einbettungsalgorithmus der Wasserzeichentechnik so konzipiert, dass dieser bei Nutzung der Extraktionsdomain des Beschreibungsmerkmals als Einbettungsdomain transparent für die Merkmalsextraktion ist. Die vorhergehende Analyse von günstigen Arbeitsbereichen resultierte jedoch für die Extraktion des Beschreibungsmerkmales in Segmentlängen $l_M$ von 4 096 - 8 192 Samples und für die Wasserzeichentechnik in Segmentlängen von $l_W$ von 1 024 - 1 536.

Bei der Verwendung von unterschiedlichen Arbeitsbereichen für die beiden Systemelemente sind zur Erhaltung der für das Inhaltsmerkmal transparenten Wasserzeicheneinbettung Modifikationen des Systemablaufs notwendig.

Die Abbildung 4.16 veranschaulicht die Kombination der Systemelemente. Die Segmentierung des Audiorahmens wird durch die Segmentierungslänge $l_W$ der Wasserzeichentechnik bestimmt. Die Segmente werden in der Matrix $S$ angeordnet, transformiert und die Koeffizienten werden in die Frequenzgruppen $F\,j$ eingeteilt. Die Frequenzgruppenkoeffizienten $\mathbf{c}_i^{F\,j}$

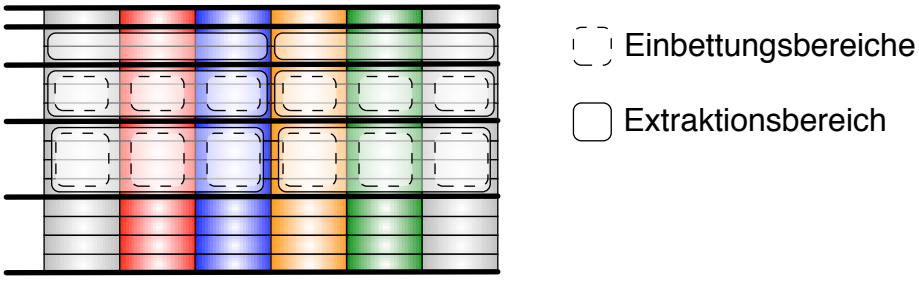


**Abbildung 4.16.:** Kombination der Merkmalsextraktion und Wasserzeicheneinbettung bei unterschiedlichen Segmentlängen $l_M$ und $l_W$

werden, wie in Abschnitt 4.3.2 beschrieben, als Einbettungsdomain für die Wasserzeicheninformation verwendet. An dieser Stelle erfolgt die Anpassung für die Merkmalsextraktion. Der Extraktionsbereich wird auf eine für das Inhaltsmerkmal günstige Segmentlänge $l_M$ vergrößert. Die Frequenzgruppen von $h = \frac{l_M}{l_W}$ aufeinanderfolgende Segmente werden für die Berechnung der Matrix $D$ zusammengefasst. Die Elemente $d_{j,h}$ der Matrix $D$ ergeben sich nach (4.22).

$$d_{j,h} = \sum_{h'} \sum_{k} \left| c^{Fj}_{k,h'} \right|; \qquad h' = (h-1)\frac{l_M}{l_W} + 1, \ldots, h\frac{l_M}{l_W} \tag{4.22}$$

Die Wasserzeicheneinbettung ist für die Merkmalsextraktion transparent, da diese die $d$-Werte unverändert lässt. Die folgenden Schritte entsprechen wieder dem in Abschnitt 4.2.2 beschriebenen Systemablauf. Es werden die Differenzen zwischen den $d$-Werten der gleichen Frequenzgruppen benachbarter Segmente gebildet und in Abhängigkeit ihres Vorzeichens auf Binärwerte abgebildet.

Für die Analyse des Gesamtkonzeptes wird die Extraktionsdomain mit den Frequenzgruppen $F2$ - $F23$ und einer Segmentlänge $l_M = 8\,192$ Samples festgelegt. Die Einbettungsdomain liegt in den Frequenzgruppen $F15$-$F20$ und nutzt eine Segmentlänge $l_W$ von 1024 Samples. Die Transparenz der Wasserzeicheneinbettung wird auf eine ODG von -1 eingestellt. Die Robustheitsanalyse wird mit den in Abschnitt 4.4.1.3 festgelegten Störungen mit einem erweiterten Parametersatz durchgeführt.

Aus den geschützten Audiodaten konnten die Inhaltsmerkmale zu 99,95 Prozent wieder korrekt extrahiert werden. Die Wasserzeicheninformation konnte zu 99,67 Prozent korrekt ausgelesen werden. Eine mögliche Fehlerursache liegt in beiden Fällen in den stillen bzw. sehr leisen Bereichen der Audiodaten. Die Größe der Modifikation hängt proportional von den Koeffizientenwerten im Einbettungsbereich ab. In „Stille“-Bereichen liegen die durch die Wasserzeicheneinbettung verursachten Modifikationen häufig unterhalb der Auflösung (hier 16 Bit) des Audiodatenformats. Im Verlaufe der Quantisierung werden die Modifikationen wieder verworfen, wodurch die Wasserzeicheneinbettung fehl schlägt, oder die Modifikationen werden verstärkt,

wodurch eine merkmalstransparente Einbettung nicht gewährleistet werden kann. Die Häufigkeit dieser Fehler hängt von der Charakteristik der geschützten Audiodaten ab. In Abbildung 4.17 und 4.18 sind die Fehlerraten des Inhaltsmerkmales und der Wasserzeicheninformation für die unterschiedlichen geschützten Testdaten dargestellt.

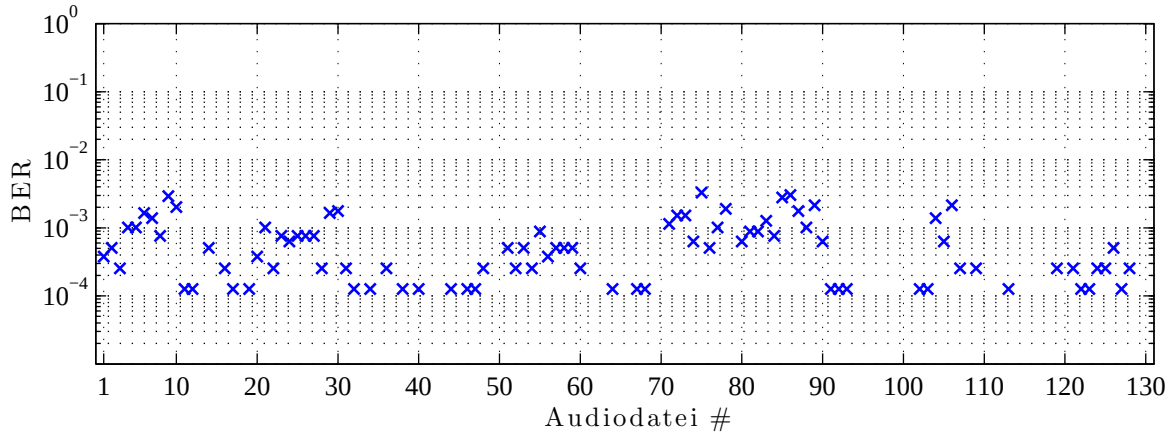


**Abbildung 4.17.:** Fehlerrate des Inhaltsmerkmals nach Schutz der Testdaten. Segmentlänge: $l_M = 8\,192$, $l_W = 1\,024$; Merkmalsdomain: $\mathbf{c}^{F2}$-$\mathbf{c}^{F23}$; Einbettungsdomain: $\mathbf{c}^{F15}$-$\mathbf{c}^{F20}$; Transparenz: ODG $= -1$

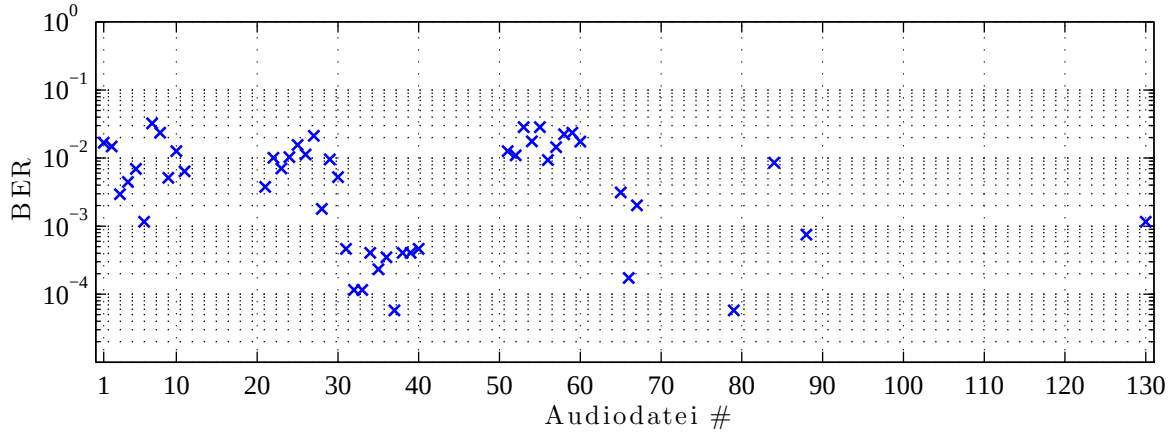


**Abbildung 4.18.:** Bitfehlerrate der Wasserzeicheninformation nach Schutz der Testdaten. Segmentlänge: $l_M = 8\,192$, $l_W = 1\,024$; Merkmalsdomain: $\mathbf{c}^{F2}$-$\mathbf{c}^{F23}$; Einbettungsdomain: $\mathbf{c}^{F15}$-$\mathbf{c}^{F20}$; Transparenz: ODG $= -1$

Die Abbildungen 4.19 und 4.20 zeigen einen Auszug aus den in Anhang A.4.2 aufgeführten Ergebnissen der Robustheitsanalyse des Inhaltsmerkmals und der Wasserzeicheninformation gegenüber dem festgelegten Satz an Störungen.

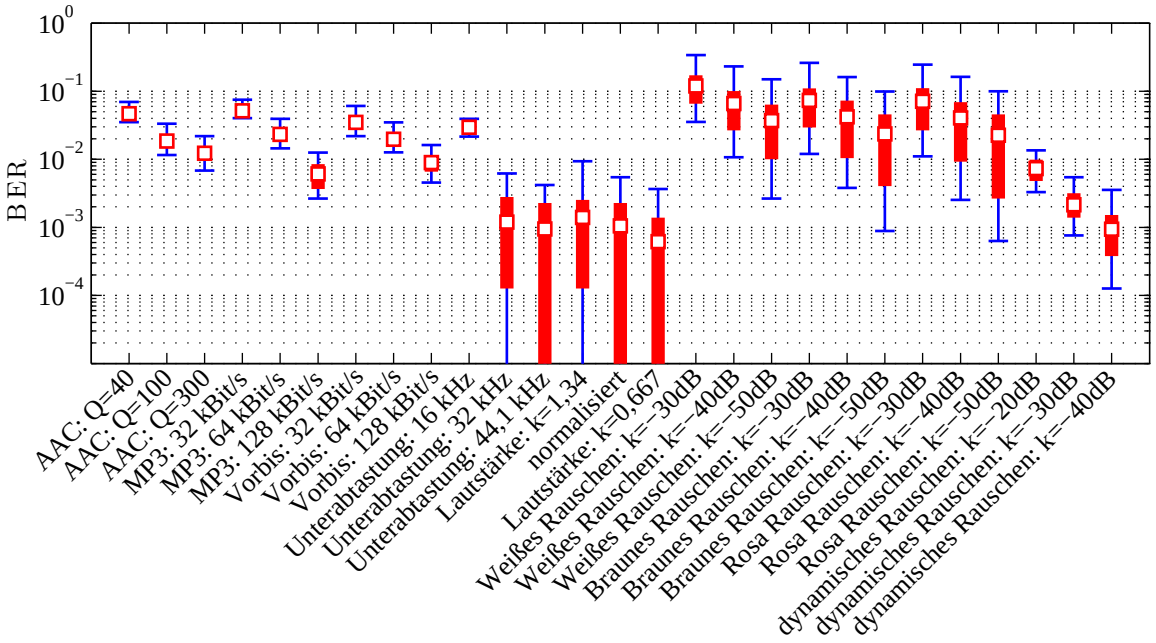


**Abbildung 4.19.:** Fehlerrate des Inhaltsmerkmals nach Schutz der Testdaten. Segmentlänge: $l_M = 8\,192$, $l_W = 1\,024$; Merkmalsdomain: $\mathbf{c}^{F2}$-$\mathbf{c}^{F23}$; Einbettungsdomain: $\mathbf{c}^{F15}$-$\mathbf{c}^{F20}$; Transparenz: ODG $= -1$

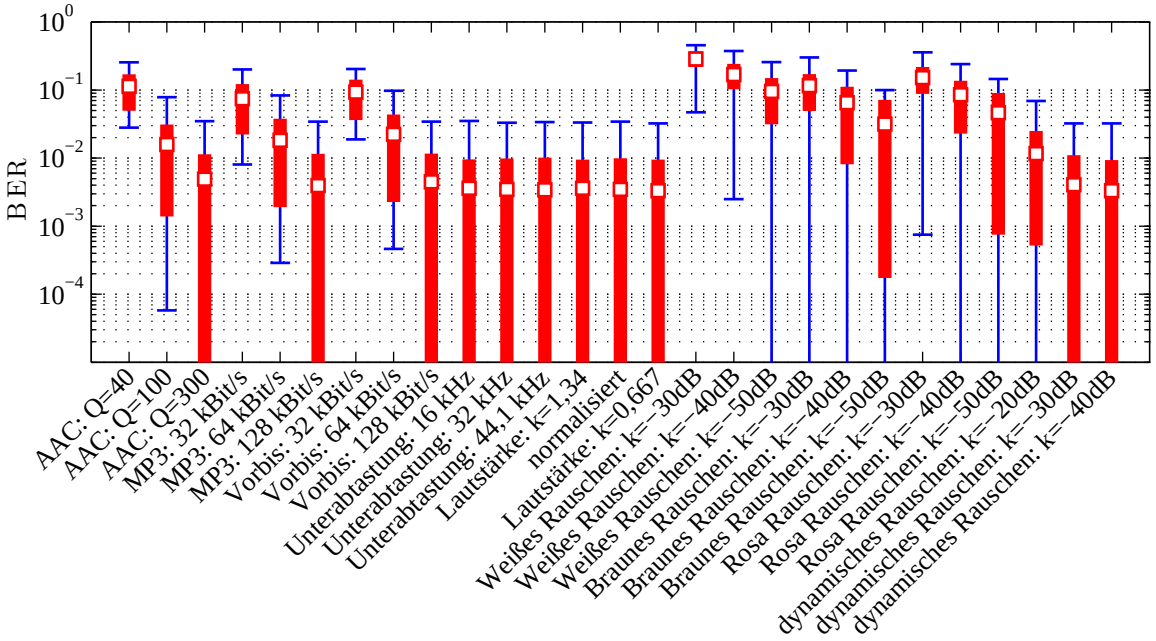


**Abbildung 4.20.:** Bitfehlerrate der Wasserzeicheninformation nach Schutz der Testdaten. Segmentlänge: $l_M = 8\,192$, $l_W = 1\,024$; Merkmalsdomain: $\mathbf{c}^{F2}$-$\mathbf{c}^{F23}$; Einbettungsdomain: $\mathbf{c}^{F15}$-$\mathbf{c}^{F20}$; Transparenz: ODG $= -1$

## 4.6. Sicherheitsaspekt

In der Verifizierungphase werden die erneut extrahierten Inhaltsmerkmale und die Wasserzeicheninformation abgeglichen. Ausgehend von dem Szenario einer öffentlichen Verifizierbarkeit der Audiodaten sind beide Informationen bei Kenntnis der Extraktions- bzw. Wasserzeichentechnik öffentlich zugänglich. Wasserzeichentechniken sind symmetrische Verfahren. Bei Kenntnis des Einbettungsalgorithmus und des Einbettungsschlüssels $K_E$ ist eine Manipulation der Wasserzeicheninformation möglich. Ein Angreifer kann ohne weitere Schutzmechnismen nach Manipulation des Dateninhaltes die Wasserzeicheninformation so ändern, dass diese wieder mit dem manipulierten Inhaltsmerkmal übereinstimmt. Die Schutzstrategie gegen die Fälschung der Datenintegrität beruht auf einer asymmetrischen Verschlüsselung (z.B. RSA [RSA78]) der Wasserzeicheninformation. In der Schutzphase wird die Wasserzeicheninformation vor ihrer Einbettung mittels eines geheimen Schlüssels $K_G$ eines asymmetrischen Schlüsselpaares verschlüsselt. In der Verifikationsphase wird die verschlüsselte Wasserzeicheninformation ausgelesen und mit dem öffentlichen Schlüssel $K_{\text{Ö}}$ des asymmetrischen Schlüsselpaares entschlüsselt. Die entschlüsselten Inhaltsmerkmale werden mit dem erneut extrahierten Inhaltsmerkmal verglichen. Ein Angreifer kann die Wasserzeicheninformation weiterhin beliebig verändern. Er ist jedoch nicht in der Lage, die zum veränderten Inhaltsmerkmal passende verschlüsselte Wasserzeicheninformation zu erzeugen. Die Erzeugung der verschlüsselten Wasserzeicheninformation ist nur unter Kenntnis des geheimen Schlüssels $K_G$ möglich. Um eine hinreichende Sicherheit gegen *„Brute Force“*-Angriffe zu gewährleisten, werden Schlüssellängen von wenigstens 128 Bit empfohlen [ECR10, FNI10, BBB$^+$07]. Asymmetrische Verschlüsselungsverfahren weisen Schwächen in ihrem Algorithmus auf, welche den Aufwand für *„Brute Force“*-Angriffe verringern. Hier werden Schlüssellängen von 3 072 bis 4 096 empfohlen [ECR10, FNI10, BBB$^+$07].

Die Analyse des in Abschnitt 4.2.2 konstruierten Inhaltsmerkmals zeigt gegenüber der Menge von zulässigen Störungen keine vollständige Robustheit. Die Verifikation der Echtheit der Audiodaten erfolgt schwellwertbasiert. Dieses führt zu einer Abschwächung der Verschlüsselung. Man nehme an, der geheime Schlüssel und die Wasserzeicheninformation sind binär und haben eine Länge von $n$. Ist für die Verifikation der Audiodaten die Identität von dem originalen Inhaltsmerkmal, welches in der Wasserzeicheninformation transportiert wird, und dem aus den gestörten Audiodaten erneut extrahierten Inhaltsmerkmal notwendig, muss ein Angreifer genau die eine Wasserzeichensequenz aus $2^n$ möglichen Sequenzen finden, die entschlüsselt mit dem Inhaltsmerkmal der gefälschten Audiodaten übereinstimmt. Werden die Audiodaten auch bei geringen Abweichungen des erneut extrahierten Inhaltsmerkmals vom dem Inhaltsmerkmal aus der Wasserzeicheninformation als echt verifiziert, ergeben sich für einen Angreifer mehrere günstige Wasserzeichensequenzen. Einem Angreifer genügt jede Sequenz, die entschlüsselt

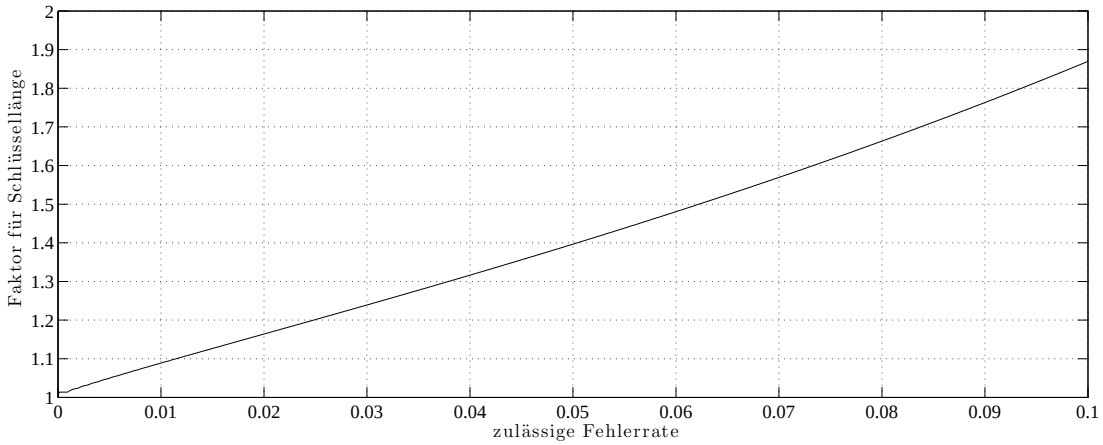


**Abbildung 4.21.:** Schlüssellängenerweiterung in Abhängigkeit einer zulässigen Fehlerrate zwischen gesuchter und erzeugter Sequenz bei konstanten *„Brute Force“*-Aufwand

zum gefälschten Inhaltsmerkmal eine Hamming-Distanz kleiner gleich der für die positive Verifikation zulässigen Fehleranzahl aufweist. Der Aufwand für einen *„Brute Force“*-Angriff verringert sich für eine zulässige Fehlerrate $th$ auf die Suche von einer Sequenz aus $\sum_{k=0}^{\lfloor n \cdot th \rfloor} \frac{n!}{k!(n-k)!}$ günstigen Sequenzen bei $2^n$ möglichen Sequenzen.

$$[0,1] := \{th \in \mathbb{R} \,|\, 0 \leq th \leq 1\} \tag{4.23}$$

In Abbildung 4.6 ist eine numerische Abschätzung des Faktors gegeben, um den die Schlüssellänge bei einer zulässigen Fehlerrate $th$ erhöht werden muss, um die gleiche Sicherheit gegenüber einem *„Brute Force“*-Angriff zu gewährleisten wie im dem Fall, dass nur die Übereinstimmung von dem erneut extrahierten Inhaltsmerkmal und dem Inhaltsmerkmal aus der Wasserzeicheninformation zur positiven Verifikation der Audiodaten führt. Für die Abschätzung wurde nur der größte Summand der Summe $\sum_{k=0}^{\lfloor n \cdot th \rfloor} \frac{n!}{k!(n-k)!}$ betrachtet.

Geht man von einer zulässigen Fehlerrate von 5% aus, erhöht sich die empfohlene Schlüssellänge gegen *„Brute Force“*-Angriffe von 128 Bit auf etwa 180 Bit. Für eine Größenordnung von 192 Bit werden für eine asymmetrische Verschlüsselung eine Schlüssellänge von 7 680 Bit empfohlen [BBB$^+$07]. Es bleibt zu untersuchen, ob sich bei dieser Strategie weitere Möglichkeiten eines kryptografischen Angriffes bieten.

Die Dauer des Audiorahmens erhöht sich bei der verwendeten Segmentlänge $l_M$ von 8 192 Samples auf ungefähr eine Minute. Eine Reduzierung der Schlüssellänge ist durch die Verbesserung der Robustheit des Inhaltsmerkmales gegenüber zulässigen Störungen, und somit der Herabsetzung der Entscheidungsschwelle, möglich.

## 4.7. Zusammenfassung

In dem vorliegenden Kapitel wurde die Entwicklung eines System zur Überprüfung der Echtheit von Audiodaten beschrieben. Das System gehört zu den inhalts-fragilen Wasserzeichenverfahren und ermöglicht einen qualitativen bis inhaltlichen Schutz von Audiodaten. Ein Nutzer ist im gewöhnlichen Umgang mit den geschützten Audiodaten nicht eingeschränkt. Verarbeitungsschritte ohne qualitative Störungen bzw. inhaltliche Veränderungen der Daten sind zulässig.

Die Grundelemente der Entwicklung sind ein Beschreibungsmerkmal des Dateninhalts und eine auf das Merkmal angepasste Wasserzeichentechnik. Das Inhaltsmerkmal stellt eine Beschreibung der Audiodaten in Form der zeitlichen Veränderung der Frequenzgruppen dar. Die entwickelte Wasserzeichentechnik beruht auf dem Prinzip des *Patch-Work*-Algorithmus.

Die Leistungsanalyse des Systems erfolgte anhand eines 130 Dateien umfassenden Audiotestdatensatzes bestehend aus Hörspielen, Buch- und Lyriklesungen, „Comedy“ und Rundfunkbeiträgen. Die Robustheit des Inhaltsmerkmales und der Wasserzeichentechnik wurde gegenüber verlustbehafteter Audiokompression, Unterabtastung und Lautstärkenveränderung sowie verschiedener rauschartiger Störungen bewertet. Die Transparenz der Wasserzeichentechnik wurde auf ein ODG von -1 eingestellt.

Die Entwicklungsziele des Verfahrens wurden weitestgehend erreicht.

- Das System erlaubt eine öffentliche Verifizierung der Integrität und Authentizität von Audiodaten.

- Der Schutz des Systems vor Manipulationen erfolgt auf Basis einer symmetrischen Verschlüsselung der Wasserzeicheninformation.

- Das entwickelte Inhaltsmerkmal ermöglicht eine schwellwertbasierte Differenzierung von zulässigen und unzulässigen Störungen.

- In Abhängigkeit der Größe der Entscheidungsschwelle ergeben sich für einen Angreifer mehrere günstige Sequenzen für eine Wasserzeicheninformation, welche es ermöglichen, gefälschte Daten als echt zu verifizieren. Eine Reduzierung der Entscheidungswelle ermöglicht eine Reduzierung der für die Verschlüsselung notwendigen Schlüssellänge. Soll die Größe eines eigenständig geschützten Audiorahmens verringert werden, besteht Optimierungsbedarf für die Robustheit des Inhaltsmerkmals gegenüber zulässigen Störungen.

- Die Nutzung der Extraktionsdomain des Inhaltsmerkmals für die Wasserzeicheneinbettung ist ausgehend vom Konzept der Wasserzeichentechnik störungsfrei für das Inhaltsmerkmal. Die Ursachen für geringfügige Störungen des Inhaltsmerkmals nach Einbettung des Wasserzeichens sind zu untersuchen.

- Die Kapazität des Wasserzeichens ist ausreichend groß, um neben dem Inhaltsmerkmal weitere Informationen (z.B. Datum, Urheber, ID , Rahmennummer) zu übertragen.

- Auf Grund der verschlüsselten Wasserzeicheninformation ist eine vollständige Robustheit der Wasserzeicheninformation erforderlich. Die Wasserzeichentechnik zeigt gute Robustheitseigenschaften gegenüber zulässigen Störungen, jedoch keine vollständige Robustheit. Für die Wasserzeichentechnik existiert unabhängig von einer vorliegenden Störung eine systematische Fehlerschranke. Entwicklungsbedarf besteht hier in der Reduktion der Bitfehlerrate der Wasserzeicheninformation durch fehlerkorrigierende Maßnahmen oder Meidung von fehleranfälligen Audiopassagen, wie Stille.

Kapitel 5

# Erweiterungen des Grundsystems

*Im vorherigen Kapitel erfolgte die Entwicklung und Leistungsanalyse des Grundsystems eines inhalts-fragilen Wasserzeichenverfahrens. Die Robustheit des Inhaltsmerkmals und der Wasserzeicheninformation des Grundsystems weisen noch Schwächen in leisen bzw. stillen Bereichen der Audiodaten auf. Als Konsequenz daraus werden in diesem Kapitel Systemerweiterungen zur Vermeidung und Korrektur von Fehlern entwickelt.*

## 5.1. Merkmalsextraktion mit Totzone

Die Leistungsanalyse des im vorangehenden Kapitel entwickelten Inhaltsmerkmals zeigt, dass eine gewisse Fehlerrate (siehe Abschnitt A.4.1) nach Störungen nicht unterschritten wird. Diese Fehlerrate zeigt sich auch unmittelbar nach der Schutzphase. Um ein Fehlschlagen der Verifikation von geschützten, aber ansonsten unveränderten Audiodaten zu vermeiden, wird in den folgenden Abschnitten eine Fehleranalyse vorgenommen. Darauf aufbauend wird eine Merkmalsverstärkung in Form einer Totzone entwickelt.

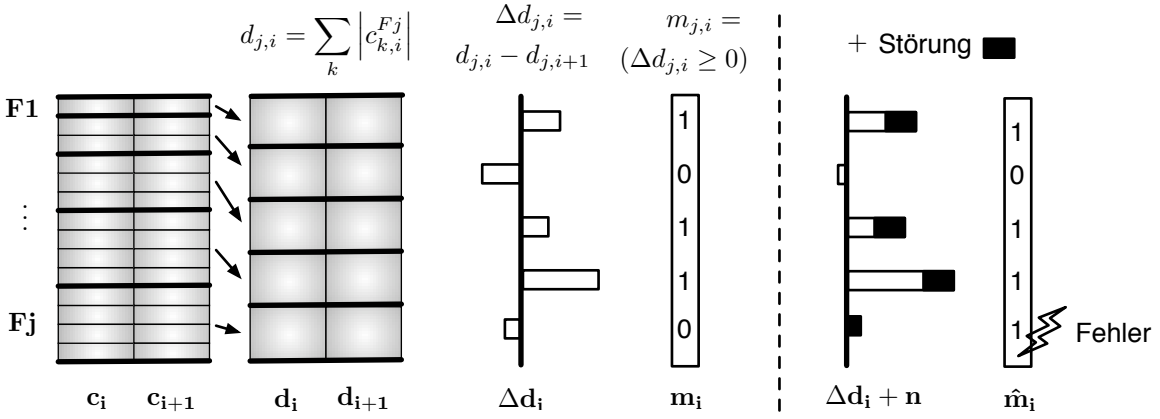


**Abbildung 5.1.:** Veränderung des Inhaltsmerkmals in Folge einer Störung der Audiodaten

### 5.1.1. Analyse der Auswirkung von Quantisierungsfehlern auf das Inhaltsmerkmal

In Erinnerung an Abschnitt 4.2.2 wird das Inhaltsmerkmal **M** konstruiert, indem die $\Delta d$-Werte, die Differenzen zwischen zeitlich aufeinanderfolgenden Frequenzgruppen, anhand des Vorzeichens auf Binärwerte entschieden werden. Um eine Veränderung des Inhaltsmerkmals zu bewirken, muss eine Störung vorliegen, die ausreichend groß ist, um das Vorzeichen eines $\Delta d$-Wertes zu kippen (s. Abbildung 5.1).

Theoretisch sollten unmittelbar nach der Schutzphase keine Veränderungen des Inhaltsmerkmals auftreten. Die einzigen Veränderungen der Audiodaten erfolgen während der Wasserzeicheneinbettung. Der Wasserzeichenalgorithmus ist so konstruiert, dass die Inhaltsmerkmale auch nach Veränderung der Audiodaten durch die Wasserzeicheneinbettung unverändert bleiben. Die mathematische Formulierung des Wasserzeichenalgorithmus (siehe Abschnitt 4.3.2) beruht auf wertkontinuierlichen Signalen. Digitale Audiosignale sind zeit- und wertdiskret. Die Ursache der Störungen der Inhaltsmerkmale sind in der Quantisierung und Begrenzung des Wertebereichs der Audiosignale zu suchen.

Das Ausmaß des Quantisierungfehlers wird üblicherweise durch das Signal-Rausch-Verhältnis (engl. *Signal to Noise Ratio* - SNR), dem Pegelverhältnis von Signalleistung $P_S$ zu Fehlerleistung $P_N$, beschrieben.

$$SNR = 10 \cdot log_{10} \frac{P_S}{P_N} = 10 \cdot log_{10} \left( \frac{\tilde{\mathrm{A}}_{Signal}}{\tilde{\mathrm{A}}_{Rauschen}} \right) \tag{5.1}$$

In vorliegenden Fall werden Audiodaten mit einer Samplingtiefe von 16 Bit pro Sample verwendet. Hierbei ergeben sich pro Sample 65 536 Quantisierungsstufen mit einer konstanten

Intervallbreite von $\Delta q$. Der maximale Quantisierungsfehler eines Samples beträgt eine halbe Intervallbreite $\left|\frac{\Delta q}{2}\right|$. Ist das Eingangssignal am Quantisierer ein Zufallssignal, so ist auch der Quantisierungsfehler ein Zufallssignal. Audiodaten sind informationstragende Signale und können somit als Zufallssignale angesehen werden. Für den Fall, dass eine Quantisierungsstufe sehr viel kleiner ist als der Wertebereich (Samplingtiefe $\geq$ 6 Bit), kann der Quantisierungsfehler als gleichverteiltes Zufallssignal angenommen werden [Rop06]. Die Rauschleistung ist somit unabhängig von der Signalamplitude. Das Signal-Rausch-Verhältnis wird bei festgelegter Samplingtiefe durch die Signalleistung bzw. den Effektivwert bestimmt.

Es ist zu vermuten, dass die Fehlerpositionen des Inhaltsmerkmals vorwiegend in Audiopassagen mit geringer Signalleistung, in stillen bzw. leisen Audiopassagen, zu finden sind. Eine weitere mögliche Fehlerursache liegt in der Begrenzung des Wertebereichs. Überschreitet ein Samplewert den zulässigen Wertebereich, wird dieser auf den nächsten gültigen Wert beschnitten (engl. *clipping*). *Clipping*-Fehler sind potentiell größer als Quantisierungsfehler. Ein Auftreten ist in den Bereichen des Audiosignals wahrscheinlich, in dem die Samplewerte nahe des maximalen Wertebereichs liegen, folglich in lauten Audiopassagen. Neben sehr leisen bzw. sehr lauten Audiopassagen sind Passagen mit geringen Änderungen des Inhalts und somit kleinen $\Delta d$-Werten als fehleranfällig zu betrachten.

Um geeignete Maßnahmen zur Kompensation der anfänglichen Fehler des Inhaltsmerkmals zu entwickeln, werden im folgenden die Effektivwerte der Audiopassagen und die $\Delta d$-Werte an Fehlerpositionen des Inhaltsmerkmals untersucht. Die Parametereinstellungen entsprechen der Leistungsanalyse in Abschnitt 4.5. Die Merkmalsextraktion erfolgt in den Frequenzgruppen $F2$ bis $F23$ bei einer Segmentlänge von $l_M = 8\,192$ Samples. Jedes Informationsbit $m_{j,i}$ des Inhaltsmerkmals $M$ ist somit abhängig von einer Audiopassage der Länge $2l_M = 16\,384$ Samples, bestehend aus den Segmenten $\mathbf{s}_i$ und $\mathbf{s}_{i+1}$.

In Abbildung 5.2 ist die geschätzte Dichte $\hat{p}(\mathrm{RMS}|m_{i,j} \neq m'_{i,j})$ der Effektivwerte der Audiopassagen mit verändertem Inhaltsmerkmal dargestellt. Als Vergleichsgröße ist zusätzlich die geschätzte Dichte $\hat{p}(\mathrm{RMS})$ der Effektivwerte der Audiopassagen der ungestörten Originaldaten aufgeführt. Aufgrund der logarithmischen Darstellung sind Effektivwerte von Null auf $10^{-5}$ abgebildet. Störungen des Inhaltsmerkmals durch die Einbettung des Wasserzeichens (geschützt) treten verteilt über den gesamten Wertebereich der originalen Effektivwerte auf. Die Wahrscheinlichkeit von Störungen nimmt für kleiner werdende Effektivwerte der Audiopassage zu. Die anfängliche Hypothese, dass stille bzw. leise Audiopassagen störanfällig sind, bestätigt sich. Laute Audiopassagen sind als Fehlerpositionen weniger bedeutend. Störungen durch *clipping* sind jedoch nicht auszuschließen. Ein ähnlicher Verlauf tritt auch für Normalisierung und Unterabtastung (32 kHz Abtastfrequenz) auf. Diese Störungen erzeugen eine ähnliche Fehlerrate

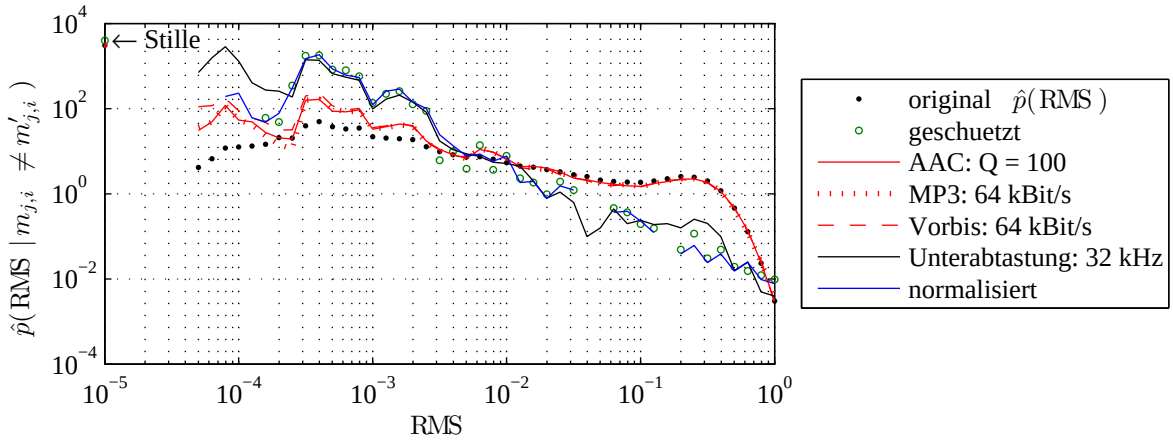


**Abbildung 5.2.:** Geschätzte Dichte $\hat{p}$ des Effektivwerts (engl. *Root Mean Square*) aller Audiosegmente $\{\mathbf{s_i}, \mathbf{s_{i+1}}\}$, für welche gilt, dass das originale Inhaltsmerkmal $m_{j,i}$ ungleich dem gestörten Inhaltsmerkmal $m'_{j,i}$ ist.

wie die Wasserzeicheneinbettung. Da alle Störungen in Kombination mit der Wasserzeichen-einbettung auftreten, sind hier die Störungen im wesentlichen durch die Einbettung bedingt. Ein anderer Verlauf zeigt sich nach verlustbehafteter Kompression. Der Verlauf ist für die verwendeten Kompressionsformate ähnlich und deckt sich weitgehend mit dem Verlauf der Dichte $\hat{p}(\mathrm{RMS})$ der originalen Trägerdaten. Die Störungen treten unabhängig von der „Lautheit“ der Audiopassagen auf. Die Wasserzeicheneinbettung tritt hier als Fehlerquelle in den Hintergrund.

Die Abbildung 5.3 zeigt die geschätzte Dichte $\hat{p}(|\Delta d| \, | m_{i,j} \neq m'_{i,j})$ der Beträge der $\Delta d$-Werte der Audiopassagen mit verändertem Inhaltsmerkmal sowie die geschätzte Dichte $\hat{p}(|\Delta d|)$ der Beträge der $\Delta d$-Werte der Audiopassagen der ungestörten Originaldaten als Vergleichsgröße. Auch hier lassen sich die Störungen ähnlich gruppieren wie bei den Effektivwerten. Die Normalisierung und Unterabtastung (32 kHz Abtastfrequenz) verursachen einen ähnlichen Verlauf der Dichte $\hat{p}(|\Delta d| \, | m_{i,j} \neq m'_{i,j})$ wie die Wasserzeicheneinbettung (geschützt). Die Kompressionsformate zeigen untereinander ebenfalls ähnliche Verläufe. Die Wahrscheinlichkeit einer Änderung des Inhaltsmerkmals nimmt für kleiner werdende $|\Delta d|$ zu. Inhaltsmerkmale, welche auf $\Delta$ d-Werte kleiner 0,003 aufbauen, sind störanfällig gegenüber Quantisierung.

### 5.1.2. Generierung einer Totzone mittels der Wasserzeichentechnik

Die Analyse zeigt, dass ein günstiger Ansatz darin liegt, eine Totzone für $|\Delta d|$ einzufügen. Unter einer Totzone wird hier ein als unzulässig definierter und somit nicht zu besetzender Wertebe-

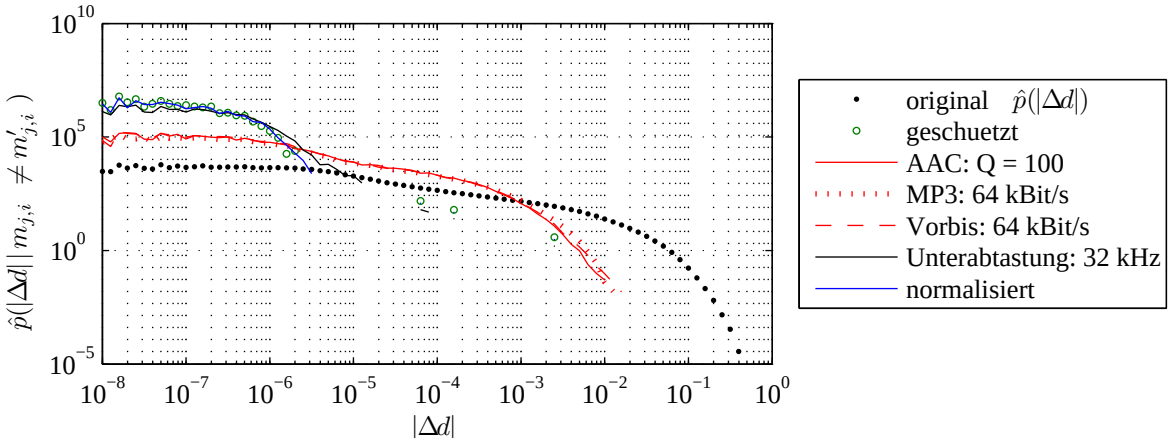


**Abbildung 5.3.:** Geschätzte Dichte $\hat{p}$ der $|\Delta d|$=Werte, für welche gilt, dass das originale Inhaltsmerkmal $m_{j,i}$ ungleich dem gestörten Inhaltsmerkmal $m'_{j,i}$ ist.

reich verstanden. Das Original ist dahingehend zu verändern, dass kein $|\Delta d|$-Wert unterhalb einer Schwelle $Th_{tot}$ liegt.

Für die Generierung einer Totzone existieren verschiedene Räumungsstrategien des der Totzone entsprechenden Wertebereichs.

- Ein Ansatz besteht darin, die Werte, welche in der Tonzone liegen, auf den nächsten zulässigen Wert abzubilden. Der durch diesen Ansatz eingebrachte Fehler (Unterschied zum Original) wird minimal gehalten. Es kommt hierbei jedoch zur Häufung der Werte an den Zonengrenzen. Solche Häufungen können zu störend wahrnehmbaren Artefakten führen. Bei Audiodaten sind insbesondere Häufungen von Koeffizienten-Werten aufgrund der Frequenzselektivität des menschlichen Gehörs zu vermeiden.

- Ein weiterer Ansatz liegt deshalb darin, Werte innerhalb der Totzone nicht allein auf die Grenzen der Totzone abzubilden, sondern auf größere Bereiche an den Totzonengrenzen bis hin zum kompletten zulässigen Wertebereich.

In dem vorliegenden Fall gilt es, eine Totzone für den Wertebereich der $|\Delta d|$-Werte kleiner einer Schwelle $Th_{tot}$ zu schaffen. Ein $\Delta d_{j,i}$-Wert ergibt sich nach (4.6) aus den Koeffizienten der Frequenzgruppe $\mathbf{c}_i^{Fj}$ des Segmentes $\mathbf{s}_i^S$. Die Korrektur der $\Delta d$-Werte erfolgt somit über eine Korrektur der Koeffizienten-Werte. Da kein direkter Zusammenhang zwischen einem $|\Delta d|$-Wert und dem zugehörigen einzelnen Koeffizienten-Wert besteht, wird durch eine Häufung von $|\Delta d|$-Werten keine Häufung von Koeffizienten-Werten erzeugt. Der Räumungsansatz, bei dem die Korrektur von Werten im unzulässigen Wertebereich durch Verschiebung auf die Totzonengrenze erfolgt, ist aufgrund des minimalen Fehlers zu bevorzugen. Die Abbildung 5.4 veranschaulicht die Generierung der Totzone mittels Korrektur der $\Delta d$-Werte am Beispiel zweier

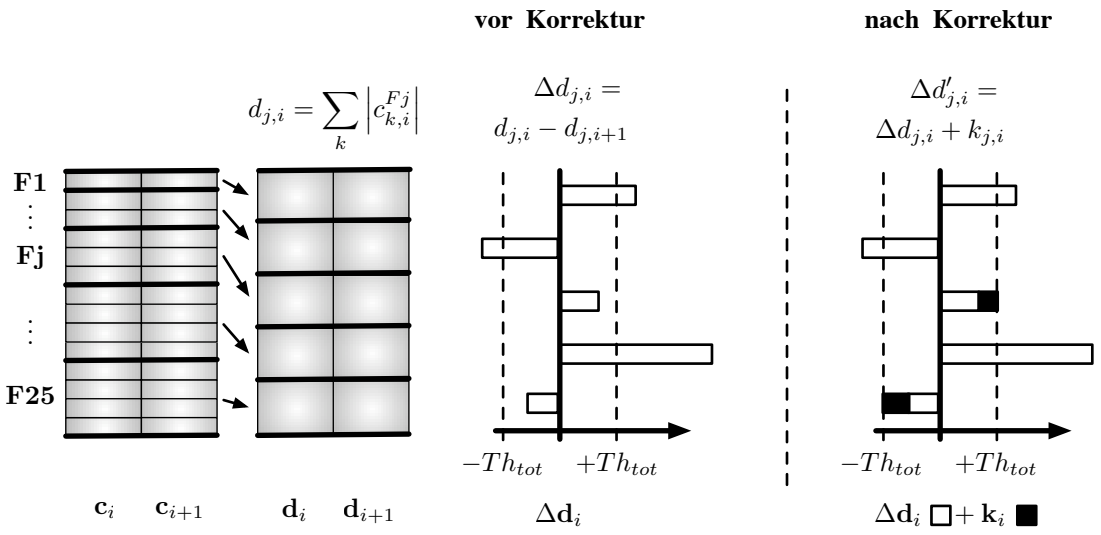


**Abbildung 5.4.:** Generierung der Totzone mittels Korrektur der $\Delta d$-Werte

transformierter Segmente $\mathbf{c}_i$ und $\mathbf{c}_{i+1}$.

Die $\Delta d$-Werte sind eine Metrik der Trägerdaten. Eine Korrektur der $\Delta d$-Werte erfolgt indirekt über die $d$-Werte, welche wiederum über die Koeffizienten modifiziert werden. $\Delta d_{j,i}$ ergibt sich in Abhängigkeit aus $d_{j,i}$ und $d_{j,i+1}$ des $i$-ten bzw. $(i+1)$-ten Segments, (4.6). Eine Korrektur von $\Delta d_{j,i}$ wirkt sich bei Veränderung von $d_{j,i}$ bzw. $d_{j,i+1}$ somit auch auf die Werte $\Delta d_{j,i-1}$ bzw. $\Delta d_{j,i+1}$ der benachbarten Segmente $\mathbf{s}_{i-1}^S$ und $\mathbf{s}_{i+1}^S$ der gleichen Frequenzgruppe aus. Die Korrektur der $\Delta d$-Werte ist sukzessiv mit einer fortlaufenden Verarbeitung der Segmente und Korrektur des jeweils zweiten $d$-Wertes vorzunehmen.

Jede Veränderung der originalen Trägerdaten wirkt sich nachteilig auf deren Qualität aus. Sowohl für die Generierung einer Totzone als auch für die Wasserzeicheneinbettung sind Modifikationen der Koeffizienten der Trägerdaten notwendig. Eine aufeinanderfolgende Modifikation der Koeffizienten, zuerst durch die Generierung der Totzone und anschließend durch die Wasserzeicheneinbettung, birgt die Problematik von sich verstärkenden Auswirkungen auf die Trägerdatenqualität. Die Generierung der Totzone wurde in dieser Hinsicht so konzipiert, dass die Koeffizienten nur einmal während der Wasserzeicheneinbettung modifiziert werden.

Abbildung 5.5 zeigt den Ablaufplan der Schutzphase des erweiterten Systemkonzepts mit Merkmalsverstärkung durch Generierung einer Totzone. Der Ablauf bleibt im Vergleich zum vorhergehenden Systemkonzept 4.1 weitgehend unverändert. Die wesentlichen Veränderungen betreffen die Prozesse der Merkmalsextraktion und der Wasserzeicheneinbettung. Die Merkmalsextraktion wird um die Berechnung der Korrekturwerte erweitert. Neben der Merkmalsinformation $\mathbf{M}$ liefert der Prozess der Merkmalsextraktion zusätzlich die Korrekturwert-Matrix $\mathbf{K}$. Die Korrekturwerte werden während der Wasserzeicheneinbettung zur Generierung der Totzone be-

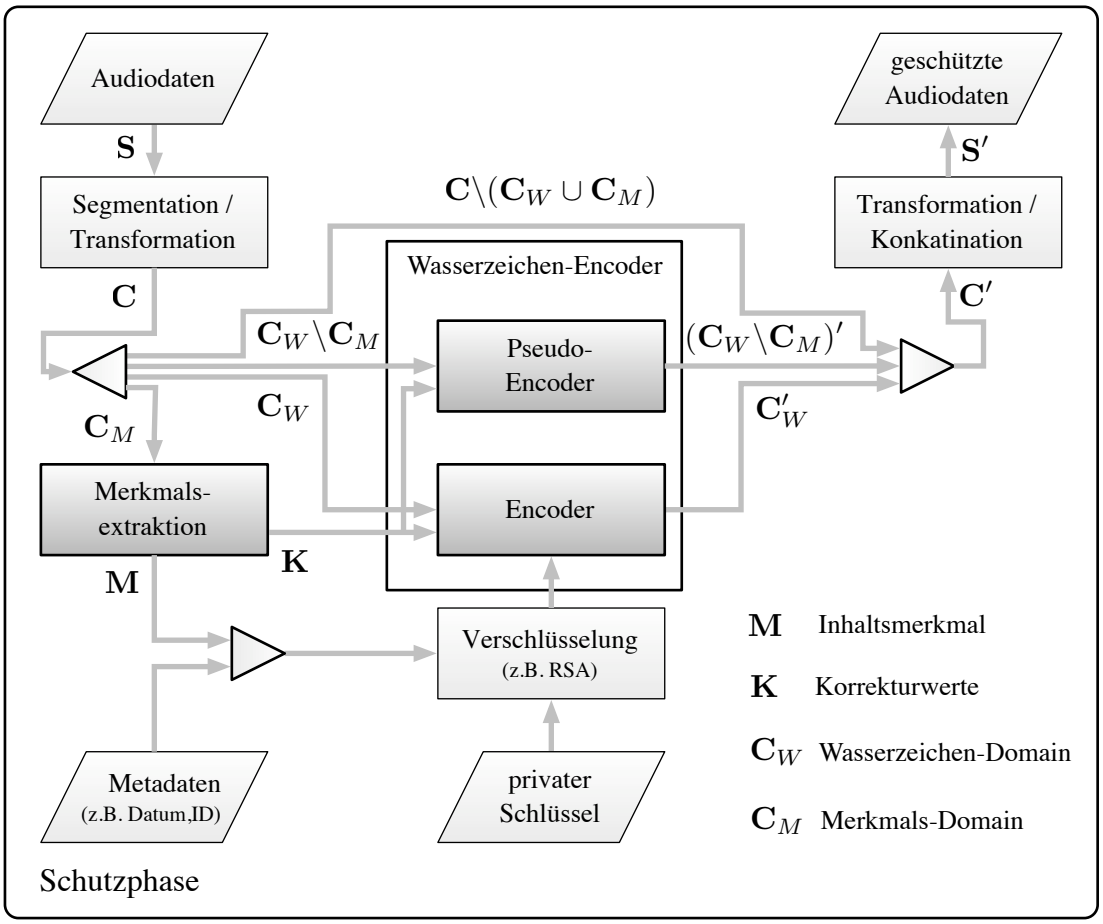


**Abbildung 5.5.:** Ablaufplan der Schutzphase des Systemkonzepts mit Merkmalsverstärkung durch Generierung einer Totzone

rücksichtigt. Neben dem eigentlichen Encoder zur Einbettung der Wasserzeicheninformation kommt ein Pseudo-Encoder hinzu, welcher die Generierung der Totzone außerhalb der Wasserzeichendomain vornimmt. Der Datenstrom der Audiodaten wird zur Verarbeitung aufgeteilt in die Wasserzeichen-Domain $\mathbf{C}_W$, die Merkmalsdomain $\mathbf{C}_M$, den Anteil der Merkmalsdomain, welcher nicht zur Wasserzeichendomain gehört $(\mathbf{C}_M\backslash\mathbf{C}_W)$ und dem Anteil, welcher weder zur Wasserzeichendomain noch zur Merkmalsdomain gehört $(\mathbf{C}\backslash(\mathbf{C}_M \cup \mathbf{C}_W))$ und somit nicht verarbeitet werden muss. Die Verifikationsphase bleibt unverändert.

### Berechnung des Inhaltsmerkmals mit Korrekturwerten

In den ersten Verarbeitungsschritten der Trägerdaten wird, wie im Abschnitt 4.2.2 beschrieben, die Berechnung des Inhaltsmerkmals vorgenommen. Ein Audiorahmen wird segmentiert und transformiert. Die Koeffizienten der einzelnen Frequenzgruppen werden betragsweise zu den $d$-Werten aufsummiert. Durch die Subtraktion der $d$-Werte aufeinanderfolgender Segmente liegen

die $\Delta d$-Werte vor. Bevor an dieser Stelle mit der Berechnung des Inhaltsmerkmals fortgefahren wird, erfolgt die Berechnung der Korrektur-Werte $k$.

Jeder $|\Delta d|$-Wert wird mittels der Hypothese (5.2) dahingehend überprüft, ob dieser innerhalb der Totzone, bestimmt durch einen Mindestwert $Th_{tot}$, liegt und somit korrigiert werden soll. Ist eine Korrektur von $|\Delta d_{j,i}|$ notwendig, erfolgt die Berechnung eines Korrektur-Werts $k_{j,i+1}$ zur Korrektur von $\Delta d_{j,i}$ nach (5.5) unter Beachtung von (5.3) und (5.4). Die Werte $\Delta d'$ und $d'$ stellen hier die korrigierten Werte dar. Die eigentliche Korrektur durch Manipulation der Koeffizienten erfolgt jedoch erst mit der Wasserzeicheneinbettung.

$$H \quad : \quad (\Delta d = d'_{j,i} - d_{j,i+1}) < Th_{tot} \tag{5.2}$$

$$d'_{j,i} = d_{j,1} + k_{j,i}; \qquad k_{j,1} = 0 \tag{5.3}$$

$$\Delta d'_{j,i} = d'_{j,i} - (d_{j,i+1} + k_{j,i+1}) \tag{5.4}$$

$$k_{j,i+1} = \begin{cases} \Delta d'_{j,i} - Th_{tot} & \forall (d'_{j,i} \geq d_{j,i+1}) \\ \Delta d'_{j,i} + Th_{tot} & \forall (d'_{j,i} < d_{j,i+1}) \end{cases} \tag{5.5}$$

Ausschlaggebend für die Bestimmung eines Merkmal-Werts $m_{j,i}$ (5.8) ist das Vorzeichen von $\Delta d_{j,i}$. Die Korrektur von $d_{j,i}$ erfolgt unter dem Aspekt, das Vorzeichen von $\Delta d_{j,i}$ und somit vorliegende Inhaltsmerkmal nicht zu verändern. Dies ist jedoch nur unter der Bedingung $d_{j,i+1} > Th_{tot} | d'_{j,i} \geq d_{j,i+1}$ möglich. Ist der Betrag von $d_{j,i+1}$ zu reduzieren, muss $d_{j,i+1}$ größer als sein Korrektur-Wert $k_{j,i+1}$ sein. Unter der Bedingung $d_{j,i+1} < Th_{tot} | d'_{j,i} \geq d_{j,i+1}$ ist eine ausreichende Reduzierung von $d_{j,i+1}$ zur Generierung der Totzone nicht möglich. Um die Totzone in diesem Sonderfall trotzdem generieren zu können, wird bei der Korrektur von $d_{j,i+1}$ die Beibehaltung des Inhaltsmerkmales missachtet. Die Korrektur von $d_{j,i+1}$ wird nach oben vorgenommen, so dass (5.6) gilt. Das Vorzeichen von $\Delta d$ wird hierbei gekippt und führt zur Änderung des Inhaltsmerkmals. Dieser Sonderfall wird bei der Berechnung der Korrektur-Werte nach (5.7) berücksichtigt.

$$d_{j,i+1} + k_{j,i+1} = d_{j,i} + Th_{tot} \tag{5.6}$$

$$k_{j,i+1} = \begin{cases} \Delta d'_{j,i} - Th_{tot} & \forall (d'_{j,i} \geq d_{j,i+1} \wedge (d'_{j,i} \geq Th_{tot}) \\ \Delta d'_{j,i} + Th_{tot} & \forall (d'_{j,i} < d_{j,i+1}) \vee ((d'_{j,i} \geq d_{j,i+1}) \wedge (d_{j,i+1} < Th_{tot})) \end{cases} \tag{5.7}$$

$$m_{j,i} = \operatorname{sgn}(\Delta d'_{j,i}) \tag{5.8}$$

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases} \tag{5.9}$$

### Wasserzeicheneinbettung unter Einbeziehung von Korrekturwerten

Die Wasserzeicheneinbettung erfüllte bis jetzt den Zweck das in der Merkmalsextraktion bestimmte Inhaltsmerkmal nebst einigen Meta-Information in die Trägerdaten zu integrieren. An dieser Stelle kommt der Wasserzeicheneinbetttung eine weitere Rolle zu. Die in dem Prozess der Merkmalsextraktion bestimmten Korrektur-Werte werden bei den ohnehin notwendigen Modifikationen der Trägerdaten berücksichtigt, um neben der Einbettung der Wasserzeicheninformation die Korrektur der $\Delta d$-Werte zu erzielen. Die Funktionsweise der verwendeten Wasserzeichentechnik wurde in Abschnitt 4.3.2 beschrieben. An dieser Stelle wird auf die Abschnitte eingegangen, die für die Integration der Korrektur-Werte von Bedeutung sind.

Die Einbettung einer Bit-Information $w \in \{0,1\}$ erfolgt in die Koeffizienten einer Frequenzgruppe $\mathbf{c}_i^{Fj}$ (hier: $\mathbf{c} = \{c_1,\ldots,c_n\}$). Diese werden vollständig in zwei disjunkte Gruppen $\mathbb{A} = \{a_1,\ldots,a_l\}$ und $\mathbb{B} = \{b_1,\ldots,b_m\}$ selektiert. Die Modifikation der Koeffizienten der Gruppen erfolgt nach (5.11), (5.12) und (5.10). Das Ergebnis der Modifikation wird durch (5.13) repräsentiert.

$$\varepsilon = \sum_{i=1}^{n} |c_i| \cdot f \tag{5.10}$$

$$a_i' = a_i \cdot \left(1 + \frac{a_i}{\sum_{i=1}^{l} |a_i|} \cdot \left(\frac{1}{2}\left(\sum_{i=1}^{n} |c_i| + (-1)^w \varepsilon\right) - \sum_{i=1}^{l} |a_i|\right)\right) \tag{5.11}$$

$$b_i' = b_i \cdot \left(1 + \frac{b_i}{\sum_{i=1}^{m} |b_i|} \cdot \left(\frac{1}{2}\left(\sum_{i=1}^{n} |c_i| - (-1)^w \varepsilon\right) - \sum_{i=1}^{m} |b_i|\right)\right) \tag{5.12}$$

$$\sum_{i=1}^{l} |a_i'| - \sum_{i=1}^{m} |b_i'| = (-1)^w \varepsilon \tag{5.13}$$

Ein Korrektur-Wert $k_{j,i}$ wurde für die indirekte Korrektur von $\Delta d_{j,i-1}$ über die Korrektur von $d_{j,i}$ ermittelt. Ein Wert $d_{j,i}$ stellt die Summe der Beträge der Koeffizienten der Frequenzgruppe $\mathbf{c}_i^{Fj}$ (5.14) dar. Für einen korrigierten Wert $d'_{j,i}$ gilt somit (5.15). Wird in (5.11) und (5.12) der Ausdruck $\sum \left|c_i^{Fj}\right|$ mit dem Ausdruck $\sum \left|c_i^{Fj}\right| + k_{j,i}$ substituiert, ergeben sich für die Korrektur der $\Delta d$-Werte die Vorschriften (5.18), (5.19) und (5.17). Betrachtet werden nur die Koeffizienten $\mathbf{c}_i^{Fj}$ einer Frequenzgruppe $Fj$ eines Segments $\mathbf{s}_i$. Die vereinfachte Notation für die folgende Betrachtung ist in (5.16) aufgeführt.

$$d_{j,i} = \sum \left|c_i^{Fj}\right| \tag{5.14}$$

$$d'_{j,i} = d_{j,1} + k_{j,i} = \sum \left|c_i^{Fj}\right| + k_{j,i} \tag{5.15}$$

$$\mathbf{c}_i^{Fj} \rightarrow \mathbf{c} = \{c_1, \ldots, c_n\}; \qquad k_{j,i} \rightarrow k \tag{5.16}$$

$$\varepsilon = \left( \left( \sum_{i=1}^{n} |c_i| \right) + k \right) \cdot f \tag{5.17}$$

$$a_i' = a_i \cdot \left( 1 + \frac{a_i}{\sum_{i=1}^{l} |a_i|} \cdot \left( \frac{1}{2} \left( \left( \sum_{i=1}^{n} |c_i| \right) + k + (-1)^w \varepsilon \right) - \sum_{i=1}^{l} |a_i| \right) \right) \tag{5.18}$$

$$b_i' = b_i \cdot \left( 1 + \frac{b_i}{\sum_{i=1}^{m} |b_i|} \cdot \left( \frac{1}{2} \left( \left( \sum_{i=1}^{n} |c_i| \right) + k - (-1)^w \varepsilon \right) - \sum_{i=1}^{m} |b_i| \right) \right) \tag{5.19}$$

Durch die Substitution gilt (5.20) für die Summen der Beträge der modifizierten Koeffizienten der beiden Gruppen $\mathbb{A}' = \{a_1', \ldots, a_l'\}$ und $\mathbb{B}' = \{b_1', \ldots, b_m'\}$. Das Ergebnis der Wasserzeicheneinbettung (5.22) bleibt durch die Generierung der Totzone unverändert.

$$\sum_{i=1}^{l} |a_i'| = \frac{1}{2} \left( \left( \sum_{i=1}^{n} |c_i| \right) + k + (-1)^w \varepsilon \right) \qquad \sum_{i=1}^{m} |b_i'| = \frac{1}{2} \left( \left( \sum_{i=1}^{n} |c_i| \right) + k - (-1)^w \varepsilon \right) \tag{5.20}$$

$$\sum_{i=1}^{l} |a_i'| - \sum_{i=1}^{m} |b_i'| = \frac{1}{2} \left( \left( \sum_{i=1}^{n} |c_i| \right) + k + (-1)^w \varepsilon \right) - \frac{1}{2} \left( \left( \sum_{i=1}^{n} |c_i| \right) + k - (-1)^w \varepsilon \right) \tag{5.21}$$

$$\sum_{i=1}^{l} |a_i'| - \sum_{i=1}^{m} |b_i'| = (-1)^w \varepsilon \tag{5.22}$$

### Modifikation der Frequenzgruppen ohne Wasserzeichen

Wird die Wasserzeicheneinbettung zur Korrektur der $\Delta d$-Werte verwendet, kann die Korrektur nur für Frequenzgruppen vorgenommen werden, welche auch zur Domain der Wasserzeicheneinbettung gehören. Für die Frequenzgruppen, welche zur Domain der Merkmalsextraktion, jedoch nicht zur Domain der Wasserzeicheneinbettung gehören, wird eine gesonderte Modifikation der Koeffizienten notwendig. Die Funktionalität der Wasserzeichentechnik kann hier in Teilen übernommen werden. Da keine Wasserzeicheninformation $w$ einzubetten ist, entfällt die Unterteilung der Koeffizienten $\mathbb{C}$ in die Gruppen $\mathbb{A}$ und $\mathbb{B}$. Die Berechnung des Zielwertes $\sum_{i=1}^{n} |c_i'|$ der Modifikation ergibt sich ohne Einbettung nach (5.23). Die Modifikation eines einzelnen Koeffizienten ergibt sich in Analogie zur Modifikation mit Einbettung eines Wasserzeichen nach (5.24).

$$\sum_{i=1}^{n} |c_i'| = \left( \sum_{i=1}^{n} |c_i| \right) + k \tag{5.23}$$

$$c_i' = c_i \cdot \left( 1 + \frac{c_i}{\sum_{i=1}^{n} |c_i|} \cdot \left( \left( \sum_{i=1}^{n} |c_i| \right) + k \right) \right) \tag{5.24}$$

**Aufteilung der Korrektur-Werte bei $l_M \neq l_W$**

Die obigen Betrachtungen erfolgten unter der Annahme, dass die Segmentlänge $l_W$ der Wasserzeicheneinbettung mit der Segmentlänge $l_M$ der Merkmalsextraktion übereinstimmt. Unter Betrachtung der Leistungsanalyse des Gesamtsystems in Abschnitt 4.5 wird die Segmentlänge $l_M$ ein Vielfaches der Segmentlänge $l_W$ betragen. Die Korrektur-Werte $k_{j,i}$ wurden während der Merkmalsextraktion für eine Segmentlänge $l_M$ berechnet. Für $l_M = l_W \cdot x;\ x > 1,\ x \in \mathbb{N}^*$ ist eine Aufteilung eines Korrektur-Wertes $k_{j,i_M}$ des $i_M$-ten Segmentes $\mathbf{s}^S_{i_M}$ der Merkmalsextraktion in $x$ Korrektur-Werte $k_{j,i_{W1}}, \ldots, k_{j,i_{Wx}}$ erforderlich. Die $x$ Segmente der Wasserzeicheneinbettung, welche mit dem Segment $\mathbf{s}^S_{i_M}$ der Merkmalsextraktion übereinstimmen, werden hier als $\mathbf{s}^S_{i_{Wy}},\ y = \{1, \ldots, x\}$ bezeichnet. Bei der Aufteilung ist zu beachten, dass $d_{j,i_{Wy}}$ einer Frequenzgruppe $\mathbf{c}^{Fj}_{i_{Wy}}$ bei einem negativen $k_{j,i_M}$ nicht kleiner sein darf als der zugeteilte Korrektur-Wert $k_{j,i_{Wy}}$. Hierbei handelt es sich um denselben Sachverhalt, welcher als Sonderfall bei der Berechnung der Korrektur-Werte berücksichtigt wurde. Um dieses Problem zu umgehen, wird eine prozentuale Aufteilung des Korrektur-Werts $k_{j,i_M}$ entsprechend den Anteilen von $d_{j,i_{Wy}}$ an $d_{j,i_M}$ vorgenommen. Ein Korrektur-Wert $k_{j,i_{Wy}}$ ergibt sich somit nach (5.26).

$$d_{j,i_{Wy}} = \sum \left| c^{Fj}_{i_{Wy}} \right| \tag{5.25}$$

$$k_{j,i_{Wy}} = k_{j,i_M} \cdot \frac{d_{j,i_{Wy}}}{d_{j,i_M}} \tag{5.26}$$

### 5.1.3. Leistungsanalyse

Als Grundlage für die Durchführung der Leistungsanalyse der Merkmalsverstärkung dient die Leistungsanalyse des Gesamtkonzepts (Abschnitt 4.5). Die Extraktionsdomain des Inhaltsmerkmals erstreckt sich somit über die Frequenzgruppen $F2$-$F23$ bei einer Segmentlänge $l_M$ von 8 192 Samples. Die Einbettungsdomain des Wasserzeichens liegt in den Frequenzgruppen $F15$-$F20$ und nutzt eine Segmentlänge $l_W$ von 1 024 Samples.

Der erste Schritt der Analyse beschäftigt sich mit der Dimensionierung der Totzone über den Schwellwert $Th_{tot}$. Die limitierende Größe ist an dieser Stelle wiederum die Datenqualität der Trägerdaten. Neben der Transparenz der Wasserzeicheneinbettung, welche über die Einbettungstärke $f$ reguliert wird, ist hier nun zusätzlich die Merkmalsverstärkung mit der Schwelle $Th_{tot}$ zu betrachten. Als Zielwert der Transparenz der Wasserzeicheneinbettung in Kombination mit der Merkmalsverstärkung wird ein ODG, gemessen zwischen den originalen und den geschützten Trägerdaten, von -1 eingestellt. Die Abbildung 5.6 zeigt den Bereich der für gegebene Schwellwerte $Th_{tot}$ resultierenden Einbettungsstärken $f$. Als Vergleichsgröße ist das Ergebnis

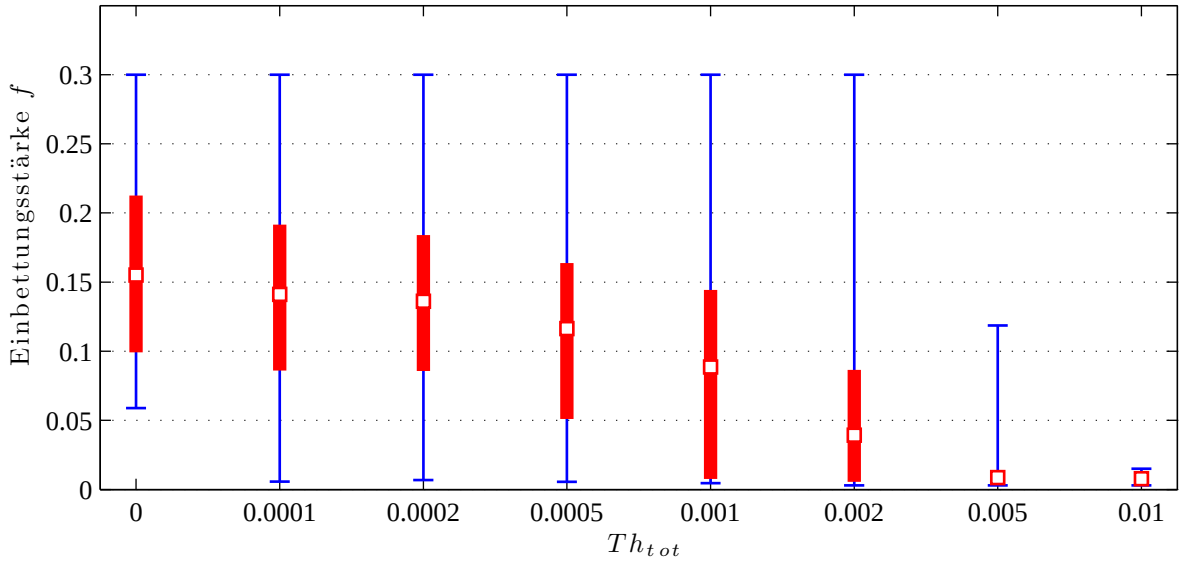


**Abbildung 5.6.:** Wertebereich I, Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■ und Mittelwert □ der Einbettungsstärke $f$ zur Einstellung einer Wasserzeichentransparenz von $ODG \approx -1$ in Abhängigkeit der Merkmalsverstärkung mit der Schwelle $Th_{tot}$.

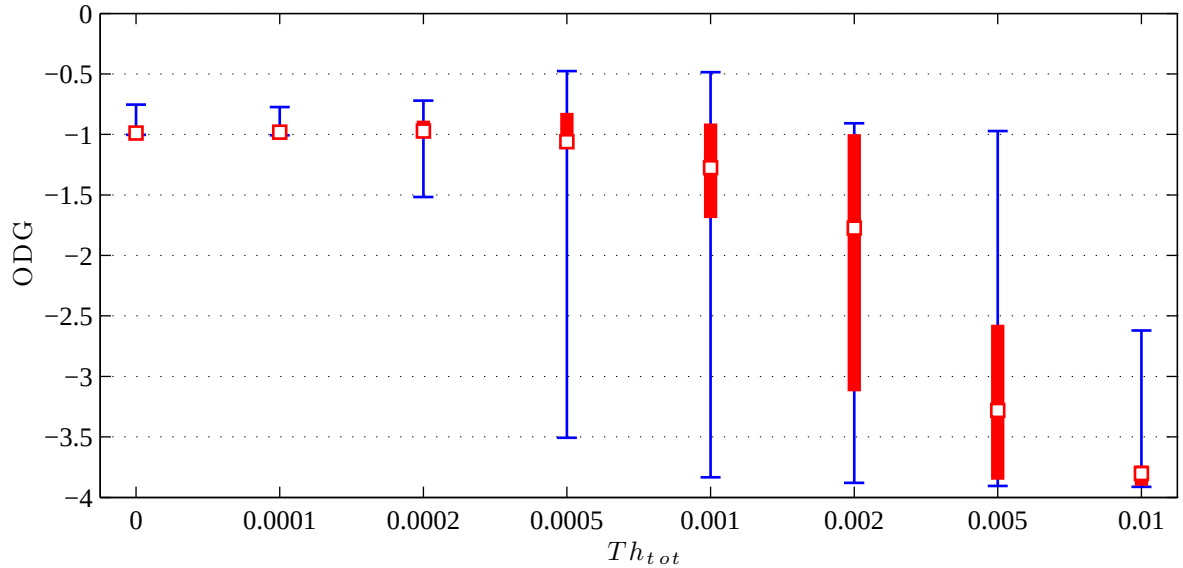


**Abbildung 5.7.:** Wertebereich I, Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■ und Mittelwert □ der erzielten ODG-Werte der Wasserzeichentransparenz in Abhängigkeit der Merkmalsverstärkung mit der Schwelle $Th_{tot}$.

des Gesamtsystems ohne Merkmalsverstärkung ($Th_{tot} = 0$) dargestellt. Die Werte für die kombinierte Transparenz der Wasserzeichentechnik und Merkmalsverstärkung ist in Abbildung 5.7 aufgeführt.

Die Merkmalsverstärkung ist nur für Schwellwerte $Th_{tot} \leq 0.0001$ praktikabel. Hier kann noch

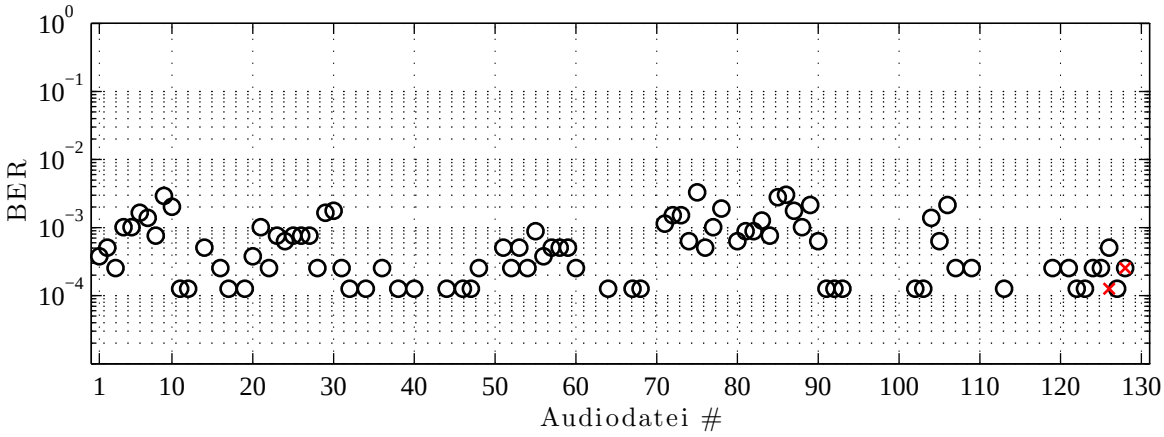


**Abbildung 5.8.:** Fehlerrate × (○) des Inhaltsmerkmals mit (ohne) Merkmalsverstärkung nach Schutz der Testdaten. Schwelle: $Th_{tot} = 0,0001$; Segmentlänge: $l_M = 8192$, $l_W = 1024$ ; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

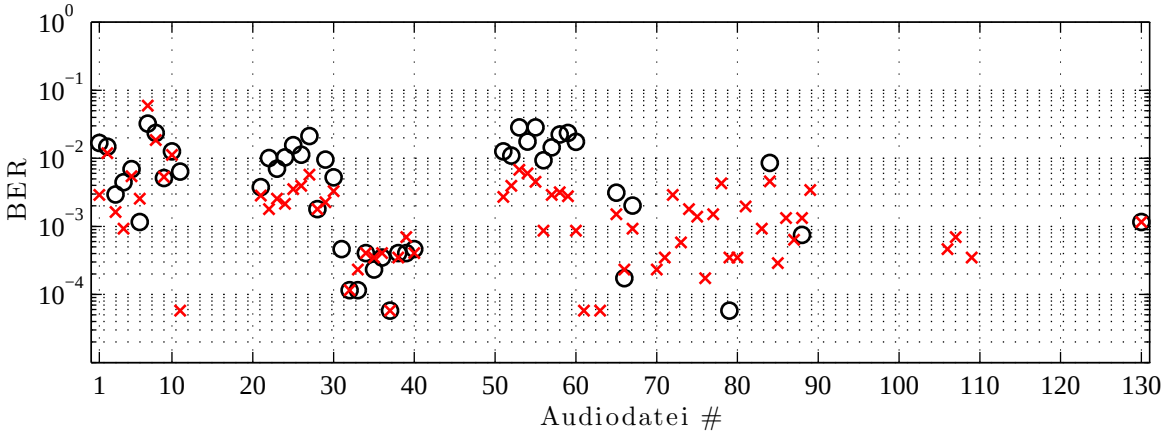


**Abbildung 5.9.:** Bitfehlerrate × (○) der Wasserzeicheninformation mit (ohne) Merkmalsverstärkung nach Schutz der Testdaten. Schwelle: $Th_{tot} = 0,0001$; Segmentlänge: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

eine Transparenz von ODG$\approx -1$ mit geringer Reduzierung der Einbettungsstärke erreicht werden. Die Streuung der Transparenz nimmt mit steigender Schwelle $Th_{tot}$ zu. Für $Th_{tot} = 0,0002$ ist die Reduzierung der Einbettungsstärke noch gering, die Wasserzeicheneinbettung mit Merkmalsverstärkung ist aber teilweise wahrnehmbar und als nicht bis leicht störend einzustufen. Bei weiterer Erhöhung der Schwelle $Th_{tot}$ ist eine Transparenz von ODG$\approx -1$ auch bei Reduktion der Einbettungsstärke $f$ nicht mehr realisierbar. Zur Bewertung der Robustheit des Inhaltsmerkmals und der Wasserzeichentechnik wird die Schwelle $Th_{tot}$ mit 0,0001 festgelegt. Die Robustheitsanalyse wird mit den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz für die einzelnen Störungen durchgeführt. Die Abbildungen 5.8 und 5.9 zeigen die Fehlerraten des Inhaltsmerkmales und der Wasserzeicheninformation für die unterschiedlichen geschützten Testdaten. Als Vergleichsgröße sind die Ergebnisse aus Abschnitt 4.5

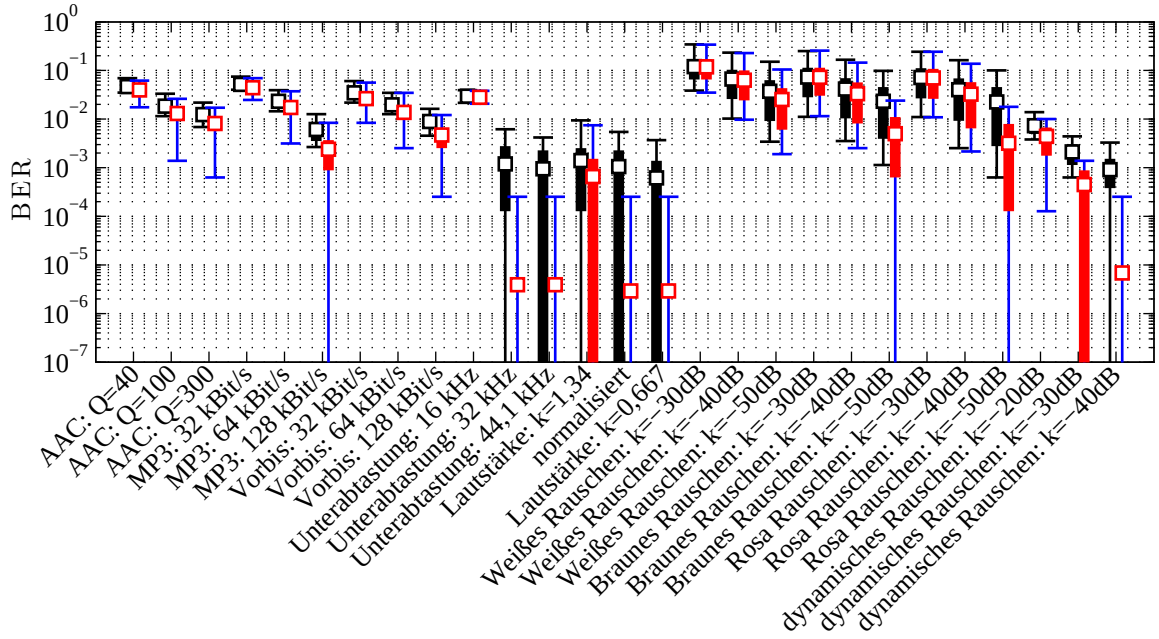


**Abbildung 5.10.:** Wertebereich I(I), Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■(■) und Mittelwert □(□) der Fehlerrate des Inhaltsmerkmals nach Schutz der Testdaten mit (ohne) Merkmalsverstärkung. Schwelle: $Th_{tot} = 0,0001$; Segmentlänge: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

ohne Merkmalsverstärkung (schwarz dargestellt) aufgeführt. Ohne die Merkmalsverstärkung konnten aus den geschützten Audiodaten die Inhaltsmerkmale zu 99,9510 Prozent wieder korrekt extrahiert werden. Mit der Merkmalsverstärkung kann die Extraktion der Inhaltsmerkmale auf eine Erfolgsrate von 99,9997 Prozent erhöht werden. Bis auf die Ausnahme von 2 Testdaten konnten die durch das Quantisierungsrauschen der Wasserzeicheneinbettung bedingten Störungen des Inhaltsmerkmals beseitigt werden. Es bleibt noch zu untersuchen, ob die verbliebenden Störungen durch das Quantisierungsrauschen oder durch andere Audioformat bedingte Effekte, wie das durch die Wertebereichbegrenzung bedingte *clipping*, entstehen. Die Wasserzeicheninformation konnte ohne Merkmalsverstärkung zu 99,6713 Prozent korrekt ausgelesen werden. Die Wasserzeicheninformation kann mit Merkmalsverstärkung trotz Reduzierung der Einbettungsstärke auf 99,8298 Prozent erhöht werden. Im Mittel sinkt die Fehlerrate der Wasserzeicheninformation, jedoch ist zu sehen, dass die Anzahl der gestörten Testdaten steigt.

Die Abbildungen 5.10 und 5.11 zeigen einen Auszug aus den im Anhang A.5 aufgeführten Ergebnissen der Robustheitsanalyse des Inhaltsmerkmals und der Wasserzeicheninformation gegenüber dem festgelegten Satz an Störungen. Erneut sind die Ergebnisse ohne Merkmalsverstärkung (schwarz dargestellt) enthalten. Die Fehlerrate des Inhaltsmerkmals sinkt mit der Merkmalsverstärkung gegenüber allen verwendeten Störungen. Die Robustheit der Wasser-

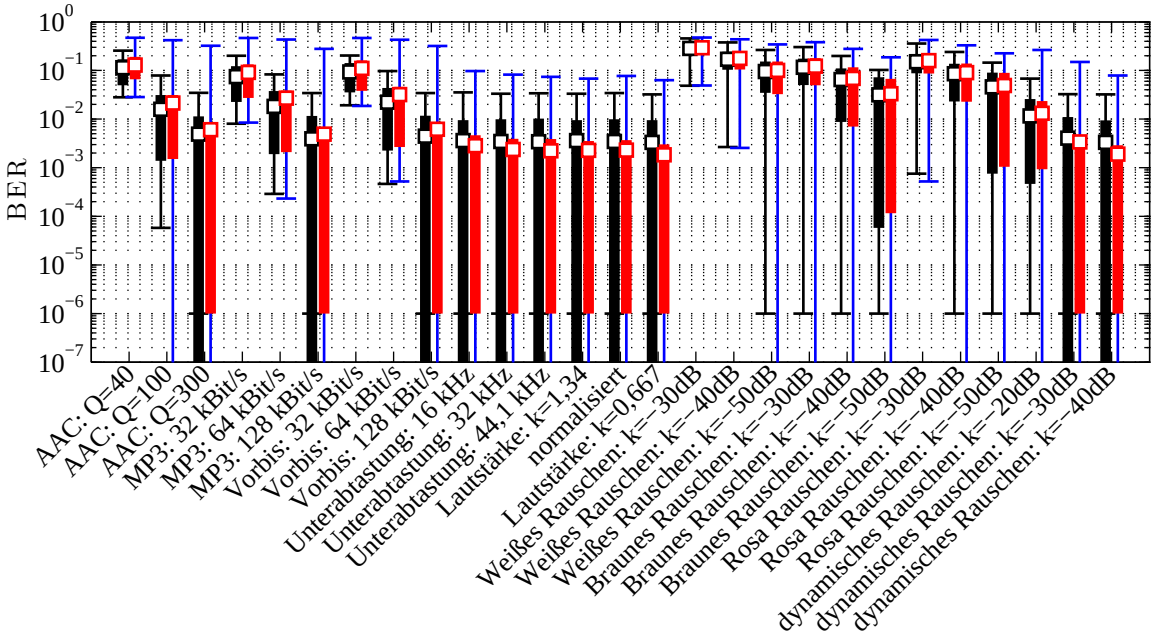


**Abbildung 5.11.:** Wertebereich I(I), Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■(■) und Mittelwert □(□) der Bitfehlerrate der Wasserzeicheninformation nach Schutz der Testdaten mit (ohne) Merkmalsverstärkung. Schwelle: $Th_{tot} = 0,0001$; Segmentlänge: $l_M = 8\,192$, $l_W = 1\,024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

zeicheninformation ändert sich im Mittel kaum. Die Fehlerraten für einzelne Testdaten erreichen jedoch höhere Extremwerte im Vergleich zum Schutz ohne Merkmalsverstärkung. Sind bei einer Audiodatei für eine erhebliche Anzahl an Inhaltsmerkmalen Korrekturen notwendig, muss die Einbettungsstärke $f$ entsprechend stark gesenkt werden, um eine ausreichende Transparenz zu erzielen. Je nach Charakteristik der Audiodaten kann die Umsetzung der Merkmalsverstärkung auch in einer deutlichen Reduzierung der Wasserzeichenrobustheit resultieren.

## 5.2. Fehlerkorrektur mittels *Soft-Input*-Decodierung

Das Systemkonzept des entwickelten Verfahrens sieht vor, dass alle notwendigen Informationen zum Extrahieren der Merkmalsinformation als auch der Wasserzeicheninformation öffentlich zugänglich sind. Aufgrund der symmetrischen Funktion der Einbettungs- und Extraktionsalgorithmen der Wasserzeichentechnik kann somit auch eine neue Wasserzeicheninformation eingebettet werden. Die Sicherheit des entwickelten Verfahrens beruht auf der asymmetrischen Verschlüsselung der Merkmalsinformation. Einem Angreifer ist es möglich, eine beliebige neue Wasserzeicheninformation in die Trägerdaten einzubringen, jedoch ist es ihm

nicht möglich, ohne den privaten Schlüssel die verschlüsselte Merkmalsinformation zu generieren.

Die Verwendung einer Verschlüsselung stellt besondere Anforderungen an die Robustheit von Wasserzeichenverfahren. Ausschlaggebend hierfür ist die Eigenschaft der Verschlüsselung, dass ähnliche Klartext-Informationen nach Verschlüsselung in unterschiedlichen Geheimtext-Informationen resultieren. Diese Eigenschaft gilt natürlich auch für die Entschlüsselung. Jeder noch so kleine Fehler führt zum Totalverlust der Klartext-Information. Für den vorliegenden Anwendungsfall würde dies bedeuten, dass ein falsch decodiertes Bit der Wasserzeicheninformation und somit ein falsches Bit der verschlüsselten Merkmalsinformation zum Totalverlust der übertragenen Merkmalsinformation führt. Eine vollständige Robustheit der Wasserzeichentechnik muss folglich nach zulässigen Störungen (s. Abschnitt 4.4.1.3) vorliegen bzw. solange vorliegen, wie die Fehlerrate des Inhaltsmerkmals die für eine erfolgreiche Verifizierung zulässige Entscheidungsschwelle nicht überschreitet. Um dieser Anforderung gerecht zu werden, ist die Robustheit der entwickelten Wasserzeichentechnik weiter zu steigern.

### 5.2.1. Systemkonzept

Das System der digitalen Wasserzeichen kann in Form eines Übertragungskanals modelliert werden [CMM99, Arn04]. Insofern liegt es nahe, Techniken der Kanalcodierung aus dem Bereich der Nachrichtentechnik auf die digitalen Wasserzeichen zu übertragen.

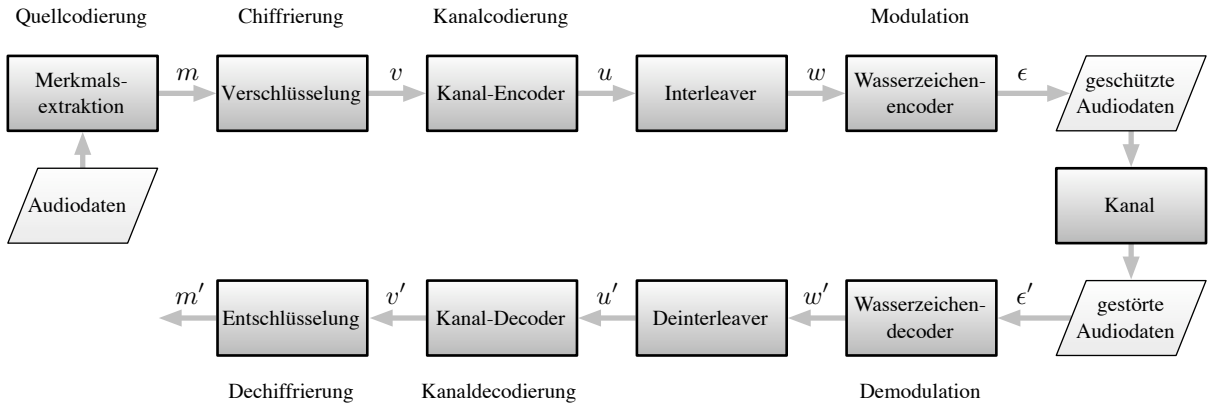


**Abbildung 5.12.:** Darstellung des Authentifizierungssystems in Form eines Übertragungssystems unter Einbeziehung der Kanalkodierung.

Die Abbildung 5.12 zeigt die wesentlichen Elemente des vorliegenden Übertragungssystems unter Einbeziehung der Kanalcodierung. Um die Notation der Größen zu vereinfachen, wird im folgenden anstelle der Matrizendarstellung (5.27) eine Sequenzdarstellung (5.28) der Größen

verwendet. (5.29) liefert die Beziehung zwischen den beiden Darstellungsformen.

$$\mathbf{X} = \begin{Bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{l,1} & x_{l,2} & \cdots & x_{l,n} \end{Bmatrix} \tag{5.27}$$

$$\mathbf{x} = \{x_1, x_2, \ldots, x_i, \ldots, x_N\} \tag{5.28}$$

$$x_i = x_{(i-\lfloor \frac{i}{l} \rfloor \cdot l),(\lceil \frac{i}{l} \rceil)} \qquad l \cdot n = N \tag{5.29}$$

Die Quelle stellt die zu schützende Audiodatei dar. Die Merkmalsextraktion kann als Quellcodierung angesehen werden, welche nach einer Irrelevanzreduktion das Inhaltsmerkmal in Form der Bit-Sequenz **m** liefert. Es folgt die Chiffrierung des Inhaltsmerkmals **m** als Schutzmechanismus gegen Fälschung eines gültigen Inhaltsmerkmals mit dem Ergebnis der verschlüsselten Bit-Sequenz **v**. An dieser Stelle fügt sich die Kanalcodierung in das Übertragungssystem ein. Der Kanal-Decoder fügt der verschlüsselten Sequenz **v** zusätzlich Redundanz zum Zweck der Fehlererkennung und Korrektur zu. Um die Leistungsfähigkeit von Kanal-Decodierern auszunutzen, ist neben der hart entschiedenen Wasserzeicheninformation $\mathbf{w}'$ die Zuverlässigkeit der entschiedenen Bit-Werte $\mathbf{w}'$ notwendig. Man spricht hier auch von einer sogenannten *Soft-Input*-Decodierung. Eine *Soft-Input*-Decodierung ist komplexer, liefert jedoch typischerweise bessere Ergebnisse als eine Decodierung mit *Hard-Input*-Werten [Mey02, Seite 298 ff], [Bos98, Seite 169 ff]. Als Kanal-Code für die weitere Entwicklung einer Fehlerkorrektur wird ein Faltungs-Code verwendet. Die *Soft-Input*-Decodierung läßt sich bei Verwendung von Faltungs-Codes effizient in Form des Viterbi-Algorithmus realisieren. Des Weiteren lassen sich Faltungs-Codes mittels Terminierung auf eine gewünschte Codelänge anpassen. Der Nachteil von Faltungs-Codes ist ihre Anfälligkeit gegenüber Bündelfehlern. Aus diesem Grund kommt nach dem Kanal-Codierer ein Interleaver zur Anwendung, um Bündelfehler der empfangenen Wasserzeichensequenz $\mathbf{w}'$ in Einzelfehler der Eingangssequenz $\mathbf{u}'$ des Kanal-Decodierers zu zerstreuen. Die Ausgangssequenz des Interleavers ist die einzubettende Wasserzeichensequenz **w**. Mittels der Wasserzeichentechnik wird die Wasserzeicheninformation **w** auf die Trägerdaten moduliert. Empfängerseitig erfolgt die Extraktion der Wasserzeicheninformation $\mathbf{w}'$. Diese wird über den Deinterleaver dem Kanal-Decodierer in Form der Sequenz $\mathbf{u}'$ zugeführt. Im Kanal-Decodierer sollen alle Übertragungsfehler korrigiert werden, so dass die decodierte Sequenz $\mathbf{v}'$ erfolgreich entschlüsselt werden kann. Das entschlüsselte Inhaltsmerkmal $\mathbf{m}'$ wird entsprechend dem Konzept des Authentifizierungssystems zur Verifizierung der Audiodatei verwendet. Eine Rekonstruktion der ursprünglichen Audiodatei aus dem empfangenen Inhaltsmerkmal $\mathbf{m}'$ ist im Übertragungssystem nicht vorgesehen.

### 5.2.2. Generierung der Zuverlässigkeitsinformation

Die Zuverlässigkeit von Empfangswerten kann durch das Log-Likelihood Verhältnis (engl. *Log-Likelihood Ratio* - LLR) angegeben werden. Das Log-Likelihood Verhältnis $L(x)$ einer Variablen $x \in \{+1,-1\}$ ist definiert in Form von (5.30). Ist die Variable $x$ abhängig von einer weiteren Variablen $y$, ist das Log-Likelihood Verhältnis $L(x|y)$ definiert in Form von (5.31).

$$L(x) = ln\left(\frac{P(x=+1)}{P(x=-1)}\right) \tag{5.30}$$

$$L(x|y) = ln\left(\frac{P(x=+1|y)}{P(x=-1|y)}\right) \tag{5.31}$$

Das Vorzeichen des Log-Likelihood Verhältnisses von $L(\cdot)$ entspricht hierbei der harten Entscheidung (5.32) von $\hat{x}$. Der Betrag $|L(\cdot)|$ stellt die Zuverlässigkeitsinformation dar.

$$\hat{x} = \begin{cases} +1 & \text{für} \quad L(\cdot) \geqq 0 \\ -1 & \text{für} \quad L(\cdot) < 0 \end{cases} \tag{5.32}$$

$$P(x|y) = \left(\frac{P(y|x)\cdot P(x)}{P(y)}\right) \tag{5.33}$$

Wendet man die Regel von Bayes für bedingte Wahrscheinlichkeiten (5.33) auf (5.31) an, folgt (5.34).

$$L(x|y) = ln\left(\frac{P(y|x=+1)}{P(y|x=-1)}\right) + ln\left(\frac{P(x=+1)}{P(x=-1)}\right) \tag{5.34}$$

Für kontinuierliche Werte $y$ kann das Verhältnis der Wahrscheinlichkeiten $P(\cdot)$ durch das Verhältnis der Dichten $p(\cdot)$ ersetzt werden. Das Log-Likelihood Verhältnis bleibt hierbei unverändert.

$$\begin{aligned} L(x|y) &= ln\left(\frac{p(y|x=+1)}{p(y|x=-1)}\right) + ln\left(\frac{P(x=+1)}{P(x=-1)}\right) \\ &= L(y|x) + L(x) \end{aligned} \tag{5.35}$$

Das Log-Likelihood Verhältnis $L(x|y)$ stellt die Zuverlässigkeitsinformation der Entscheidung am Wasserzeichen-Decoder dar und wird als a posteriori Log-Likelihood Verhältnis bezeichnet. Dies setzt sich aus dem a priori Log-Likelihood Verhältnis $L(x)$ und der Kanalinformation $L(y|x)$ zusammen. Das a priori Log-Likelihood Verhältnis liefert den Anteil der Zuverlässigkeitsinformation, welcher sich aus den Auftretenswahrscheinlichkeiten $P(x=+1)$ bzw. $P(x=-1)$ ergibt. Sind die Wahrscheinlichkeiten gleich groß, liegt keine a priori Information vor. Das a priori Log-Likelihood Verhältnis ergibt sich in diesem Fall gemäß (5.30) mit $L(x)=0$. Die Kanalinformation $L(y|x)$ ist der Anteil der Zuverlässigkeitsinformation, welche durch die

Kanalcharakteristiken bestimmt wird.

Auf die vorliegende Problemstellung bezogen, lassen sich drei Log-Likelihood Verhältnisse $L(\cdot)$ wie folgt interpretieren. Die Variable $x$ stellt die bipolare Repräsentation der zu übertragenen binären Wasserzeicheninformation $w$ dar. Die Abbildung erfolgt über die Vorschrift (5.36). Eine a priori Information liegt nicht vor. Die Wasserzeicheninformation ergibt sich aus dem verschlüsselten und kanalcodierten Inhaltsmerkmal $m$. Aufgrund der Eigenschaften der Verschlüsselung ergibt sich eine Gleichverteilung der Bitwerte der verschlüsselten Sequenz $v$. Die anschließende Faltungscodierung und das Interleaving erzeugen aus Eingangssequenzen mit gleichverteilten Werten wiederum eine Ausgangssequenz mit gleichverteilten Werten. Das a priori Log-Likelihood Verhältnis $L(w)$ kann als Null angenommen werden.

$$\begin{aligned} w = 1 &\leftrightarrow x = +1 \\ w = 0 &\leftrightarrow x = -1 \end{aligned} \tag{5.36}$$

Die Zuverlässigkeitsinformation wird somit durch die Kanalinformation $L(\varepsilon'|w)$ geliefert. Die Kanalinformation lässt sich bei Kenntnis eines Störungsmodells berechnen. In der Literatur findet sich für die analytische Bestimmung der Kanalinformation das Beispiel eines AWGN[1]-Kanals [Bos98]. Da für den Kanal in Form von Audiodaten und für den in Abschnitt 4.4.1.3 festgelegten Testsatz an Störungen keine Modelle vorliegen, erfolgt eine empirische Ermittlung der Zuverlässigkeitsinformation.

Wie zuvor im Abschnitt 4.3.2 über den Wasserzeichenalgorithmus beschrieben, dient ein Wert $\varepsilon'$ als Ausgangswert für die Entscheidung der empfangenen Wasserzeicheninformation $w'$. Der Wasserzeichen-Encoder modifiziert in der Einbettungsdomain die Koeffizienten der Gruppen $\mathbb{A}$ und $\mathbb{B}$ in Abhängigkeit der einzubettenden Wasserzeicheninformation $w$ mit dem Ziel (4.13), eine Differenz $\varepsilon$ (4.14) zwischen den Betragssummen der Koeffizienten der Gruppen $\mathbb{A}$ und $\mathbb{B}$ zu erzeugen.

$$\sum_{i=1}^{l} |a'_i| - \sum_{i=1}^{m} |b'_i| = \varepsilon \tag{4.13}$$

$$\varepsilon = (-1)^w \cdot \sum_{i=1}^{n} |c_i| \cdot f \tag{4.14}$$

Für die Entscheidung der empfangenen Wasserzeicheninformation $w'$ in Form von hart entschiedenen Werten $w' \in \{0,1\}$ mittels (4.21) ist nur das Vorzeichen der am Wasserzeichen-Decoder gebildeten Differenz $\varepsilon$ relevant. Die Größe der Einbettungsdomain $\mathbb{C}$ und die Einbettungsstärke $f$ werden für die Rekonstruktion der gesendeten Wasserzeicheninformation $w$ nicht berücksichtigt. Für eine harte Entscheidung ist nur der Wert $\varepsilon$ von Interesse, da es sich bei beiden anderen

[1] additives weißes Gaußsches Rauschen (engl. *additive white Gaussian noise*)

Größen um positive Faktoren handelt, welche den Betrag des zu entscheidenden Wertes, jedoch nicht dessen Vorzeichen, beeinflussen.

$$w' = \begin{cases} 0 & \text{für} \quad \varepsilon' = \sum |a'_i| - \sum |b'_i| \geqq 0 \\ 1 & \text{für} \quad \varepsilon' = \sum |a'_i| - \sum |b'_i| < 0 \end{cases} \tag{4.21}$$

Bei der Bestimmung der Zuverlässigkeitsinformation für eine weiche Entscheidung sind keine Informationen, welche Rückschlüsse auf die gesendete Wasserzeicheninformation $w$ liefern, zu verwerfen. Ausgehend von (5.37), welche sich aus der Umstellung von (4.14) ergibt, wird für die weiche Entscheidung der Wasserzeicheninformation $w'$ die Größe $\varepsilon'_{norm}$ nach (5.38) verwendet.

$$(-1)^w = \frac{\varepsilon}{\sum_{i=1}^{n} |c_i| \cdot f} \tag{5.37}$$

$$\varepsilon'_{norm} = \frac{\varepsilon'}{\sum_{i=1}^{n} |c_i| \cdot f} \tag{5.38}$$

Die Einbettungsstärke $f$ und die originale Größe der Einbettungsdomain $\sum_{i=1}^{n} |c_i|$ muss für diesen Ansatz nicht am Wasserzeichen-Decoder vorliegen. Nimmt man an, dass der Fehler $e_i$ der Koeffizienten $c_i$ mittelwertfrei (5.39) ist, kann für eine große Anzahl $n$ an Koeffizienten in der Einbettungsdomain $\mathbb{C}$ die originale Größe der Einbettungsdomain $\sum_{i=1}^{n} |c_i|$ durch die Größe der gestörten Einbettungsdomain $\sum_{i=1}^{n} |c'_i|$ ersetzt werden (5.40).

$$\frac{1}{n} \sum_{i=1}^{n} e_i \approx 0 \tag{5.39}$$

$$\sum_{i=1}^{n} |c'_i| = \sum_{i=1}^{n} |c_i + e_i| \approx \sum_{i=1}^{n} |c_i| \tag{5.40}$$

Aus den empfangenen $\varepsilon'$-Werten und der Größe der Einbettungsdomain $\sum_{i=1}^{n} |c'_i|$ lässt sich mit guter Näherung die Einbettungsstärke $\hat{f}$ schätzen. Abbildung 5.13 zeigt das Verhältnis aus der geschätzten Einbettungsstärke $\hat{f}$ und der originalen Einbettungsstärke $f$ unter Verwendung der Schätzfunktion (5.41) nach Störung der Trägerdaten. Die Schätzfunktion ermittelt das Maximum der geschätzten Dichte $\hat{p}(\frac{\varepsilon'}{\sum |c'_i|})$. Für diese und alle weiteren Analysen wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz

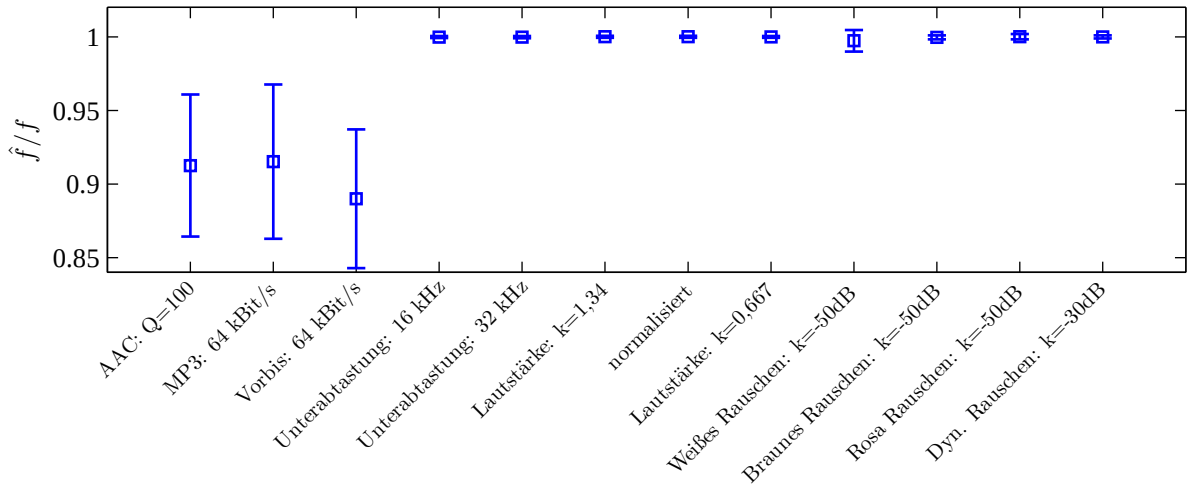


**Abbildung 5.13.:** Mittelwert und Standardabweichung des Verhältnis $\hat{f}/f$ aus der originalen Einbettungsstärke $f$ und der geschätzten Einbettungsstärke $\hat{f}$ unter Verwendung der Schätzfunktion (5.41) nach Störung der Trägerdaten.

der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt.

$$\hat{f} = max\left(\hat{p}\left(\frac{\varepsilon'}{\sum_{i=1}^{n}|c'_i|}\right)\right) \tag{5.41}$$

$$x = max(\mathbb{X}) :\Longleftrightarrow x \in \mathbb{X} \forall y \in \mathbb{X} : y \leq x \tag{5.42}$$

Das Ergebnisbild der geschätzten Einbettungsstärke $\hat{f}$ für die Störungen ist zweigeteilt. Das Bild für Störungen in Form von verlustbehafteter Kompression weicht von dem der anderen Störungen ab. Bei verlustbehafteter Kompression besitzt die Schätzung nicht nur eine größere Streuung, sondern zusätzlich auch einen systematischen Fehler. Diese Abweichung der geschätzten Einbettungsstärke $\hat{f}$ von der originalen Einbettungsstärke $f$ ist durch einen nicht mittelwertfreien Fehler der Größe der gestörten Einbettungsdomain $\sum_{i=1}^{n}|c'_i|$ bedingt. Warum die Kompressionen einen nicht mittelwertfreien Fehler hervorrufen, ist nicht weiter untersucht. Für die weitere Bestimmung der Zuverlässigkeitswerte ist dieser Fehler unerheblich. Die geschätzte Einbettungsstärke $\hat{f}$ stellt eine Zwischengröße zur Berechnung der normierten Distanz $\varepsilon'_{norm}$ nach (5.43) dar. Die Abweichung der Einbettungsstärke $\hat{f}$ durch den Fehler $\sum_{i=1}^{n} e_i$ der Größe der gestörten Einbettungsdomain wird bei der Berechnung der normierten Distanz $\varepsilon'_{norm}$ kompensiert. Die Größe der gestörten Einbettungsdomain und somit deren Fehler geht in (5.43) als Faktor ein, folglich als reziproker Wert im Vergleich zur Berechnung (5.41) der geschätzten

Einbettungsstärke $\hat{f}$.

$$\varepsilon'_{norm} = \frac{\varepsilon'}{\sum_{i=1}^{n} |c'_i| \cdot \hat{f}} \tag{5.43}$$

Die Abbildung 5.14 zeigt eine Auswahl der im Anhang A.6.1 aufgeführten geschätzten Dichten $\hat{p}(\varepsilon'_{norm})$ der geschätzten normierten Distanzen $\varepsilon'_{norm}$ unter der Bedingung, dass die gesendete Wasserzeicheninformation $w$ mit 1 bzw. 0 vorlag. Die aus den Dichten resultierenden Kanalinformationen $L(\varepsilon'_{norm}|w)$ sind in der Abbildung 5.15 dargestellt, bzw. im Anhang A.6.2 aufgeführt. Wie bei der Schätzung der Einbettungsstärke $\hat{f}$ ergeben sich wiederum zwei Fehlerbilder. Die Abbildungen 5.14 (links) und 5.15 (links) zeigen die Dichte bzw. die Kanalinformation für Störungen in Form von verlustbehaftete Kompression am Beispiel der mp3-Kompression mit 64 kBit/s. Die Dichte bzw. Kanalinformation nach additivem weißen Gaußschen Rauschen mit einer Stärke von -50 dB in den Abbildungen 5.14 (rechts) und 5.15 (rechts) ist repräsentativ für die anderen Störungen. Die dargestellten Kanalinformationen sind unter Kenntnis der gesende-

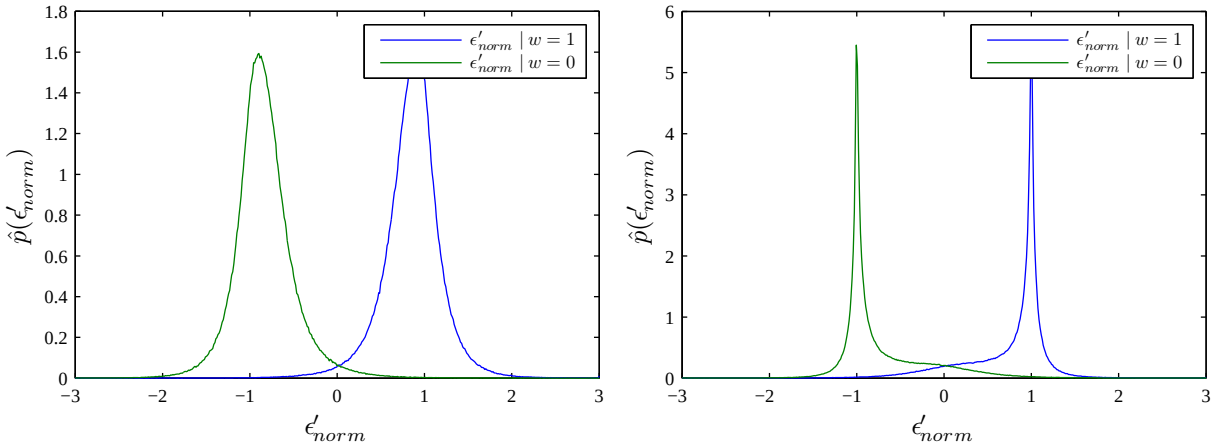


**Abbildung 5.14.:** Geschätzte Dichte $\hat{p}(\varepsilon'_{norm})$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch verlustbehaftete mp3-Kompression mit 64 kBit/s (links) bzw. weißes Gaußsches Rauschen mit einer Stärke von -50 dB (rechts) und unter der Bedingung, dass die gesendete Wasserzeicheninformation $w$ mit 1 bzw. 0 vorlag.

ten Wasserzeicheninformation **w** und Art der Störung ermittelt worden. In der Anwendung des Authetifizierungssystems liegen diese Informationen zur Bestimmung der Zuverlässigkeitswerte aus den empfangenen $\varepsilon'_{norm}$-Werten nicht vor. Die Kanalinformation wird hier in Form einer synthetischen Funktion $L_{syn}$ (5.44) bereitgestellt. Der Verlauf der Funktion $L_{syn}$ stellt einen Kompromiss zwischen den beiden Fehlerbildern der verwendeten Störungen dar und lässt sich über einen Wert $\lambda$ anpassen. Die Funktion $L_{syn}$ ist in Abbildung 5.16 für verschiedene Werte $\lambda$ dargestellt. Für die weitere Analyse wurde der Wert $\lambda$ mit 2, der Faktor $\alpha$ mit 10 festgelegt.

$$L_{syn}(x) = \alpha \left( e^{-\lambda \cdot |x-1|} - e^{-\lambda \cdot |x+1|} \right) \tag{5.44}$$

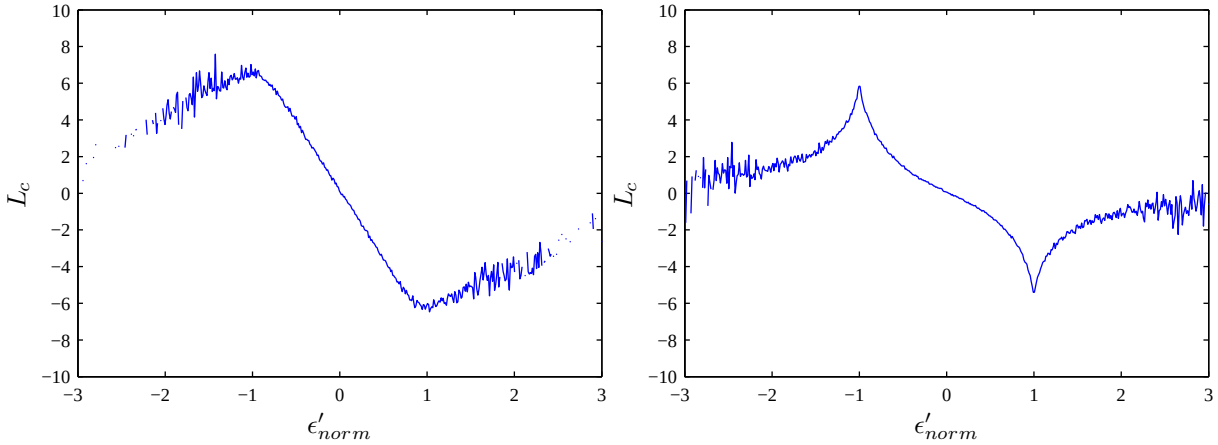


**Abbildung 5.15.:** Kanalinformation $L_c(\hat{p}(\varepsilon'_{norm}))$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch verlustbehaftete mp3-Kompression mit 64 kBit/s (links) bzw. weißes Gaußsches Rauschen mit einer Stärke von -50dB (rechts).

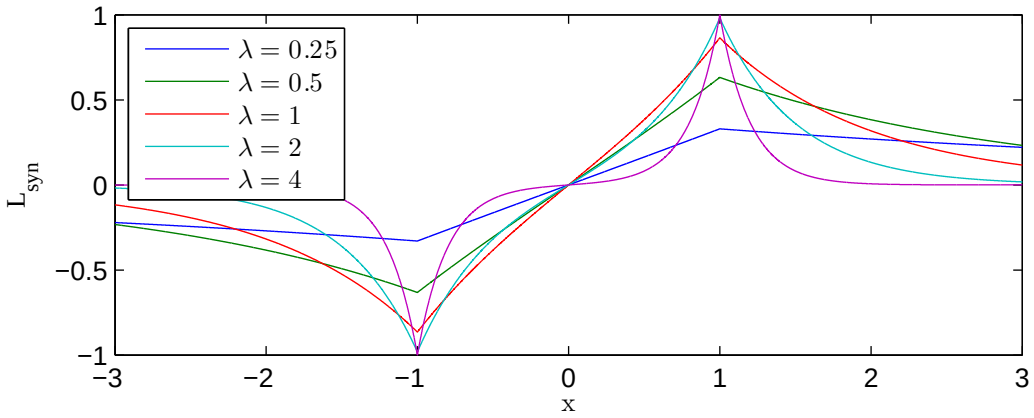


**Abbildung 5.16.:** Synthetische Kanalinformation $L_{syn}$ für $\alpha = 1$

## 5.2.3. Leistungsanalyse

Die Umsetzung des Faltungscodierers erfolgt mit der Matlab-Implementierung eines *Symbol by Symbol Maximum A-Posteriori* (SS-MAP) *Decoding*-Algorithmus von Volker Kühn[2] auf Basis der Arbeit von Bahl et al. [BCJR74].

Die Darstellung des verwendeten Faltungscodes ist in Abbildung 5.17 dargestellt. Die zugehörigen Generatoren sind in (5.45) ersichtlich. Der Faltungscode besitzt eine Code-Rate $R = 1/2$.

[2]Universität Rostock, Institut für Nachrichtentechnik

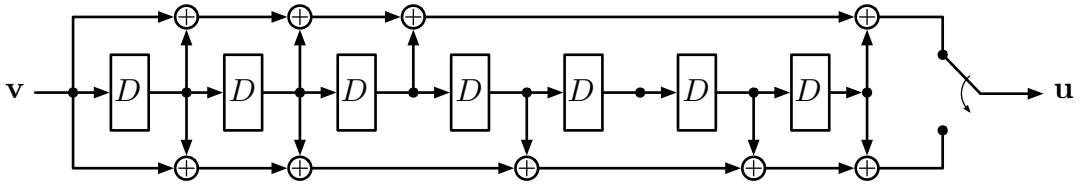


**Abbildung 5.17.:** Realisierung des verwendeten Faltungscodierers im Form eines Schieberegisters

Unter Berücksichtigung der angewandten Terminierung des Codes und der daraus resultierenden Notwendigkeit von weiteren 8 Bit entsprechend der Länge des Code-Gedächtnisses ergibt sich für eine vollständige Nutzung der Wasserzeichenkapazität von 17 280 Bit eine Code-Rate $R = 0,4998$.

$$\begin{aligned} \mathbf{g^1} &= (1,1,1,1,0,0,0,1) \\ \mathbf{g^2} &= (1,1,1,0,1,0,1,1) \end{aligned} \tag{5.45}$$

Durch die Kanalcodierung der Wasserzeicheninformation wird eine Verringerung der Fehlerrate der Nutzinformation auf Kosten der Nutzkapazität erzielt. Eine Vergleichbarkeit der Fehlerrate der uncodierten Wasserzeicheninformation und der Fehlerrate der Nutzinformation nach der Kanaldecodierung ist aufgrund der unterschiedlichen Bit-Anzahl der beiden Sequenzen nicht gegeben. Um eine Bewertung der Leistungsfähigkeit der entwickelten *Soft-Input*-Decodierung vorzunehmen, wird als Referenzgröße die Fehlerrate der uncodierten Wasserzeicheninformation mit einer auf die Hälfte reduzierten Wasserzeichenkapazität bestimmt.

Bei der Reduzierung der Wasserzeichenkapazität ist darauf zu achten, dass Parameter, welche einen Einfluss auf die Robustheit und Transparenz des Wasserzeichenalgorithmus haben, unverändert bleiben. Der Einbettungsalgorithmus, die genutzte Einbettungsdomain und die Einbettungsstärke können zur Reduzierung der Kapazität nicht verändert werden, lediglich die Wasserzeicheninformation kann variieren. Die Reduzierung der Wasserzeichenkapazität wird durch die kombinierte Auswertung zweier Wasserzeichen zur Bestimmung eines Bits Wasserzeicheninformation erzielt. Jedes Bit Wasserzeicheninformation $\mathbf{w}$ wird doppelt mit zwei Wasserzeichen (W1, W2) eingebettet, somit reduziert sich die Anzahl der Nutzbits auf die Hälfte. Im Wasserzeichendecoder werden zur Bestimmung der empfangenen Wasserzeicheninformation $\mathbf{w}'$ die Gruppen $\mathbb{A}_{W1}$, $\mathbb{B}_{W1}$ sowie $\mathbb{A}_{W2}$ und $\mathbb{B}_{W2}$ der beiden Wasserzeichen (W1, W2) nach (5.46) ausgewertet.

$$w' = \begin{cases} 0 & \text{für} \quad \varepsilon'_{W1} + \varepsilon'_{W2} = \sum |a_{W1_i}'| - \sum |b_{W1_i}'| + \sum |a_{W2_i}'| - \sum |b_{W2_i}'| \geqq 0 \\ 1 & \text{für} \quad \varepsilon'_{W1} + \varepsilon'_{W2} = \sum |a_{W1_i}'| - \sum |b_{W1_i}'| + \sum |a_{W2_i}'| - \sum |b_{W2_i}'| < 0 \end{cases} \tag{5.46}$$

Auch ohne Nutzung der Zuverlässigkeitsinformation ist durch die Verwendung eines Faltungscodes eine Reduzierung der Fehlerrate der Nutzinformation zu erwarten. Um die Leistungsfähigkeit der *Soft-Input*-Decodierung analysieren zu können, wird neben der Fehlerrate der uncodierten Wasserzeicheninformation mit halber Kapazität als weitere Bewertungsgrundlage die Fehlerrate der kanalcodierten Wasserzeicheninformation mit *Hard-Input*-Decodierung herangezogen. Die Ergebnisse für die uncodierte Wasserzeicheninformation ist im Folgenden mit „$R = 1$“ und für die Wasserzeicheninformation mit halber Kapazität mit „$R = 1/2$“ gekennzeichnet. Die Kennzeichnung lehnt an die Coderate $R$ an. Im Fall der uncodierten Wasserzeicheninformation stellt die gesamte Wasserzeichenkapazität die Nutzinformation dar. Im zweiten Fall, nach Reduktion der Kapazität und somit der Nutzinformation, liegt diese im Verhältnis eins zu zwei zur ursprünglichen Wasserzeichenkapazität vor. Die Ergebnisse für die faltungscodierten Wasserzeicheninformationen sind mit „$Hard$“ bzw. „$Soft$“ entsprechend der *Hard-Input*-Decodierung bzw. der *Soft-Input*-Decodierung gekennzeichnet.

Die Analyse der Leistungsfähigkeit der entwickelten *Soft-Input*-Decodierung erfolgt für das Grundsystem (Abschnitt 4.5) sowie für das um die Totzone erweiterten Systemkonzepts (Abschnitt 5.1). Die Merkmalsverstärkung erfolgt mit dem Schwellwert $Th_{tot}$=0,0001. Die Abbildungen 5.18, 5.20 und 5.21 zeigen einen Auszug aus den im Anhang A.6.3 aufgeführten Ergebnissen der Fehlerkorrektur für das Grundsystem. Die Abbildungen 5.19, 5.22 und 5.23 zeigen einen Auszug aus den im Anhang A.6.4 aufgeführten Ergebnissen der Fehlerkorrektur für das Grundsystem mit Totzone. Die Abbildungen 5.18 und 5.18 zeigen hierbei die Fehlerrate der Nutzinformation eines Audiorahmens. Die Fehlerrate spiegelt den Anteil der Audiorahmen wieder, für welche die Nutzinformation wenigstens einen Bitfehler aufweist und somit aufgrund der Verschlüsselung ein vollständiger Informationsverlust für den Audiorahmen eintritt. Durch den Einsatz der Faltungscodierung konnte eine Reduzierung der Fehlerrate der Nutzinformation nach Störung der Audiodaten erzielt werden. Eine fehlerfreie Rückgewinnung der Nutzinformation konnte jedoch nur für die *Soft-Input*-Decodierung mit der entwickelten Zuverlässigkeitsinformation erzielt werden. In Abhängigkeit der Störung und deren Intensität versagt jedoch auch die *Soft-Input*-Decodierung, so dass die Nutzinformation nicht mehr fehlerfrei aus den Audiodaten extrahiert werden kann. Mit zunehmender Bitfehlerrate der codierten Wasserzeicheninformation kommt die Eigenschaft der Faltungscodierung zum Tragen, so dass Fehler nicht mehr korrigiert, sondern vermehrt werden. Während die *Soft-Input*-Decodierung durchgängig für alle Störungen die geringste Fehlerrate der Nutzinformation pro Audiorahmen aufweist, überschreitet die Bitfehlerrate der Nutzinformation für die Faltungscodierung bei stärkeren Störungen die Bitfehlerrate der uncodierten Wasserzeicheninformation. Die Abbildung 5.20, 5.21, 5.22, 5.23 zeigen zur Veranschaulichung dieses Sachverhalts die mittlere bzw. maximale Bitfehlerrate der Nutzinformation eines Audiorahmens. Die Abbildung 5.24 zeigt die Differenz zwischen den Fehlerraten der Nutzinformation pro Audiorahmen für das Grundsystem und des Grundsystem

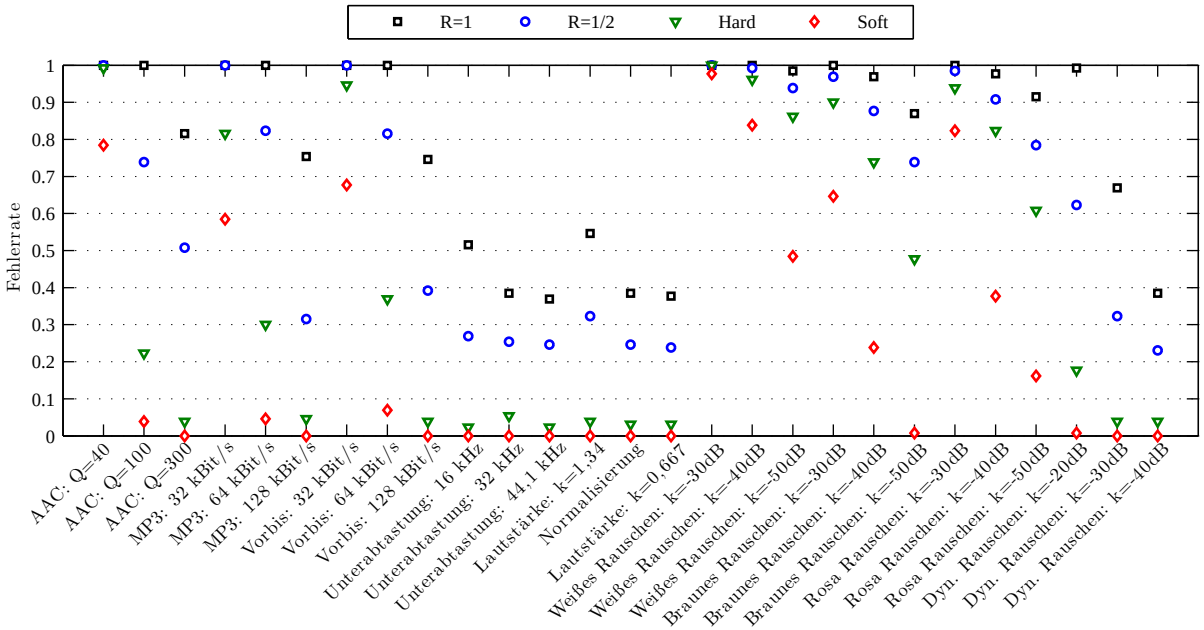


**Abbildung 5.18.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem

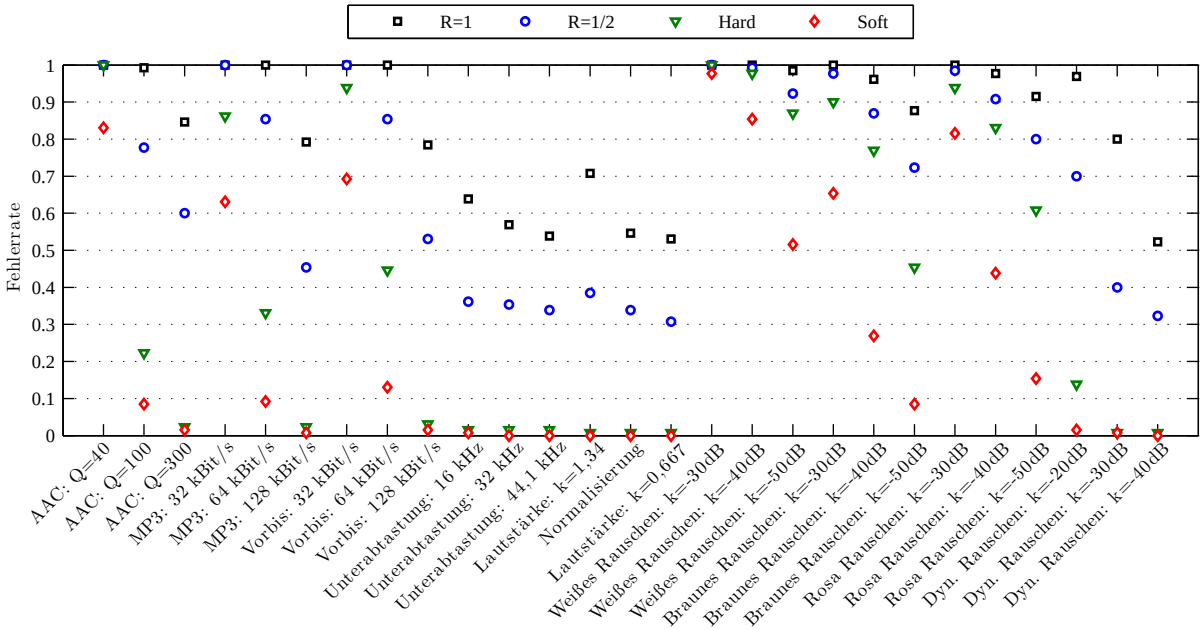


**Abbildung 5.19.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone

mit Totzone für das Inhaltsmerkmal. Wie die Leistungsanalyse der Erweiterung des Grundsystems um eine Totzone für das Inhaltsmerkmal (Abschnitt 5.1.3) zeigte, führt die Durchsetzung der Totzone für den überwiegenden Teil der Testdaten dazu, dass sich die audiodateiabhängigen Bitfehlerraten der Wasserzeicheninformation stärker um ihren Mittelwert konzentrieren. Dies bedeutet, dass der Anteil von Audiorahmen mit fehlerfreier Wasserzeicheninformation

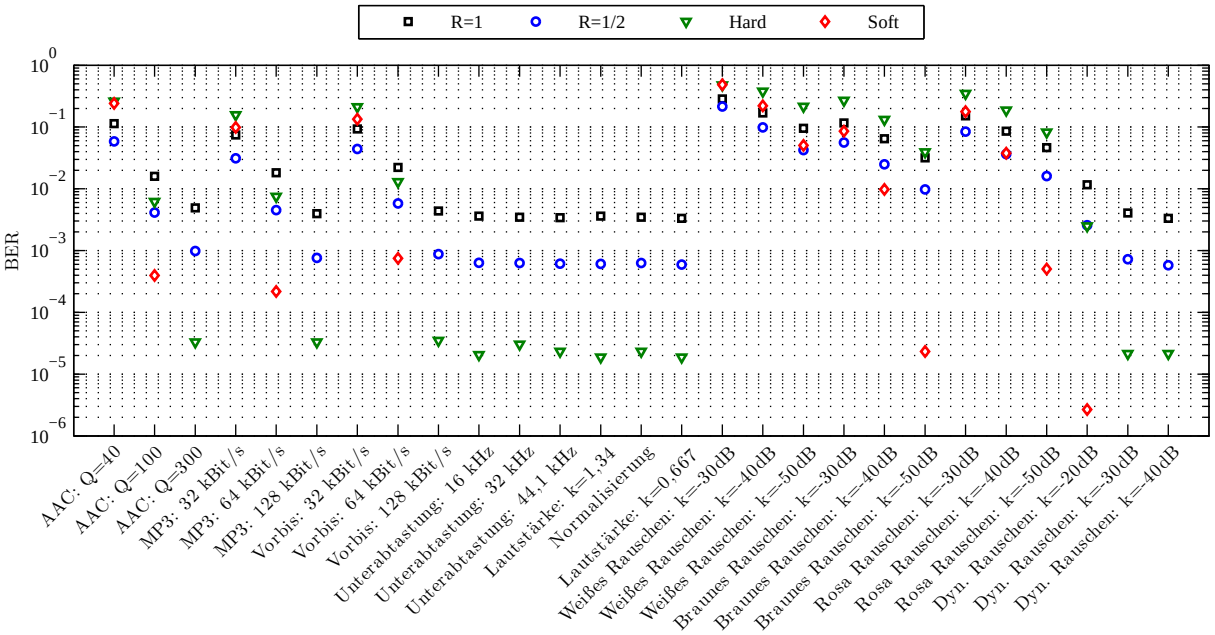


**Abbildung 5.20.:** Bitfehlerrate der Nutzinformation für das Grundsystem

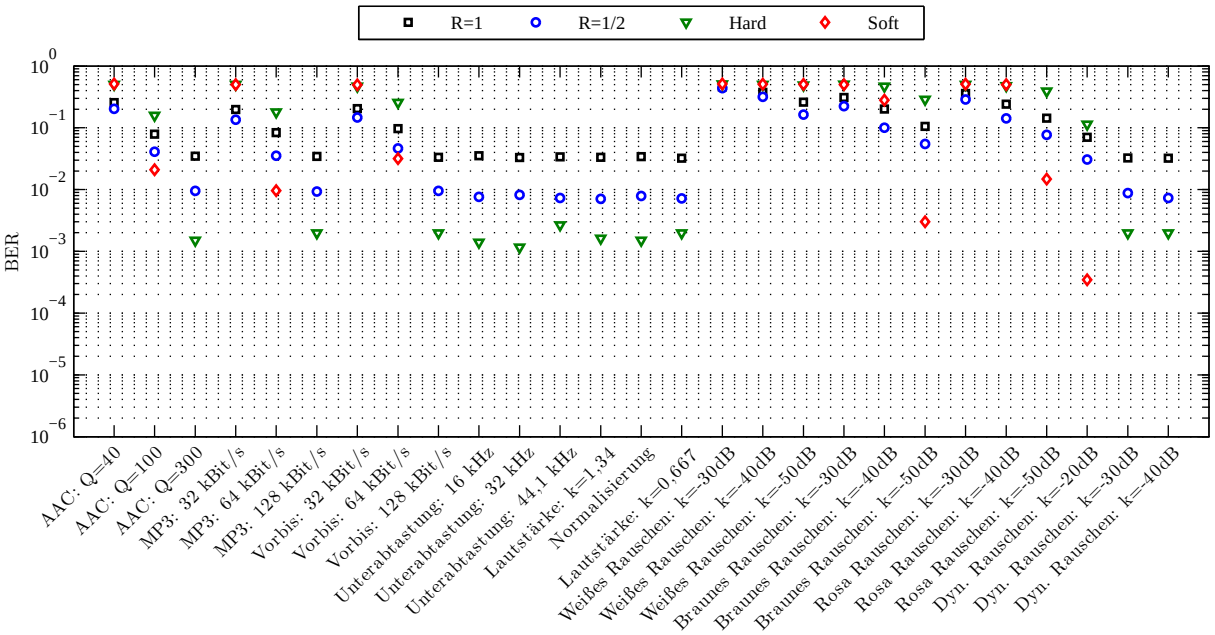


**Abbildung 5.21.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem

bzw. Wasserzeicheninformation mit geringer Bitfehlerrate abnimmt. Die Fehlerraten der Nutzinformation für die uncodierte Wasserzeicheninformation und die Wasserzeicheninformation mit halber Kapazität steigen somit an, während der Faltungscode eine geringe Anzahl an Fehlern korrigieren kann. Auf der anderen Seite sinkt durch die Totzone auch für einen Anteil an

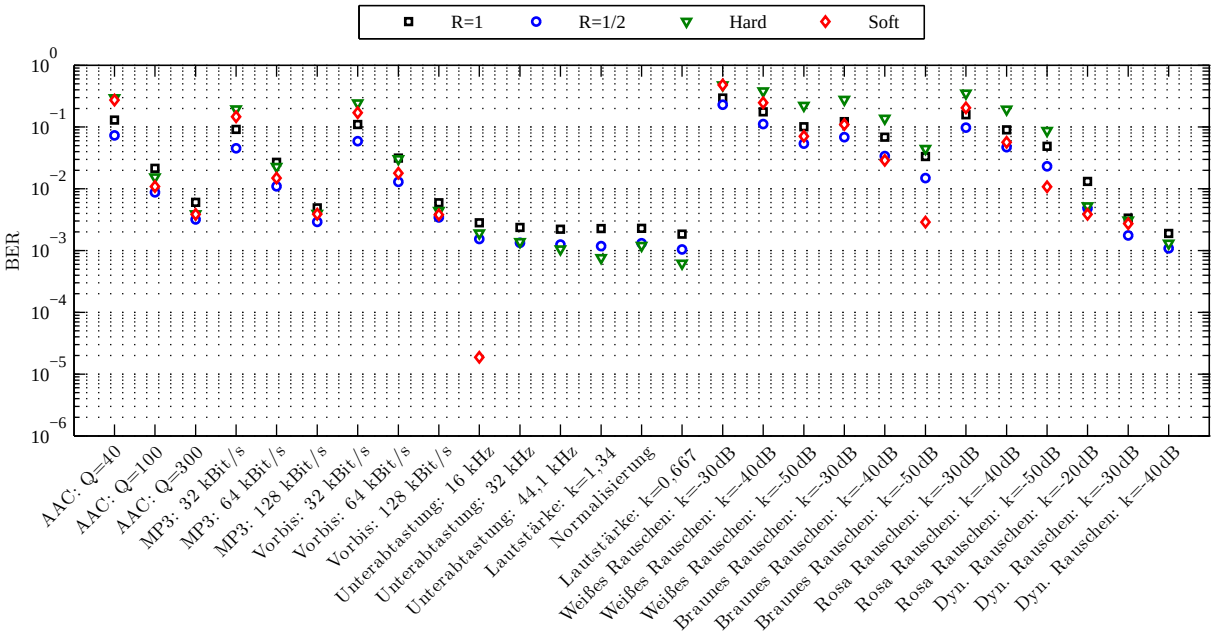


**Abbildung 5.22.:** Bitfehlerrate der Nutzinformation für das Grundsystem mit Totzone

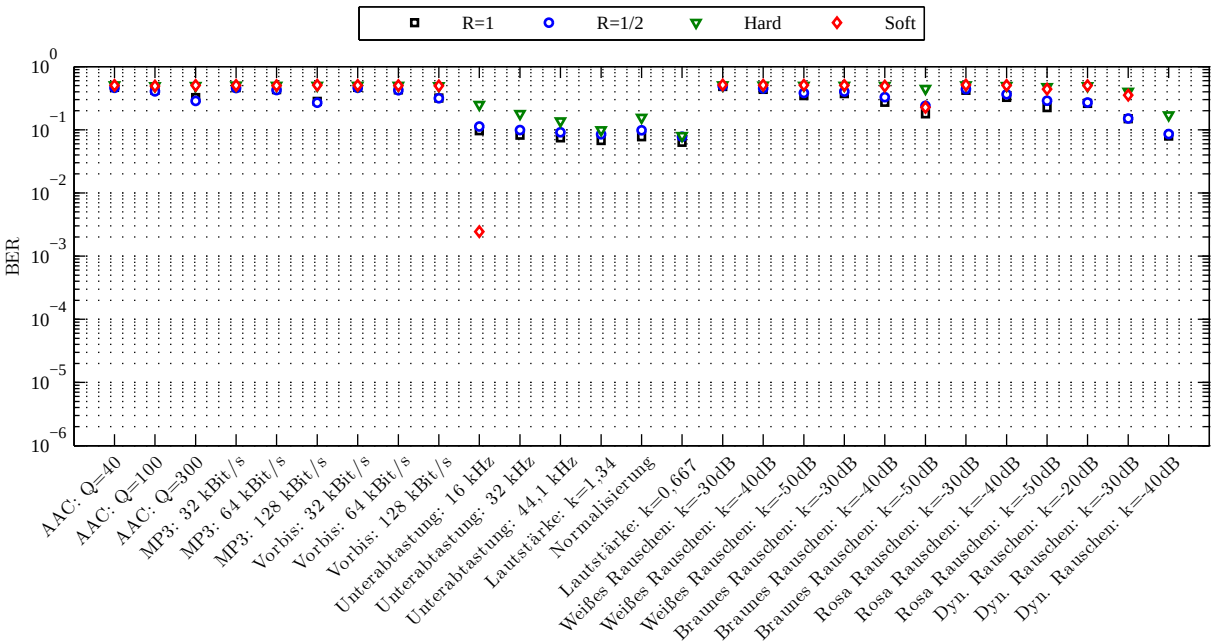


**Abbildung 5.23.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone

Audiorahmen die Bitfehlerrate der Wasserzeicheninformation, wodurch die *Hard-Input*-Decodierung bei Störungen mit geringer Intensität in einer verringerten Fehlerrate der Nutzinformation resultiert. Die weitere Auswirkung der Totzone neben der Konzentration der Bitfehlerraten sind größere Extremwerte der audiodateispezifischen Bitfehlerraten der Wasserzeicheninfor-

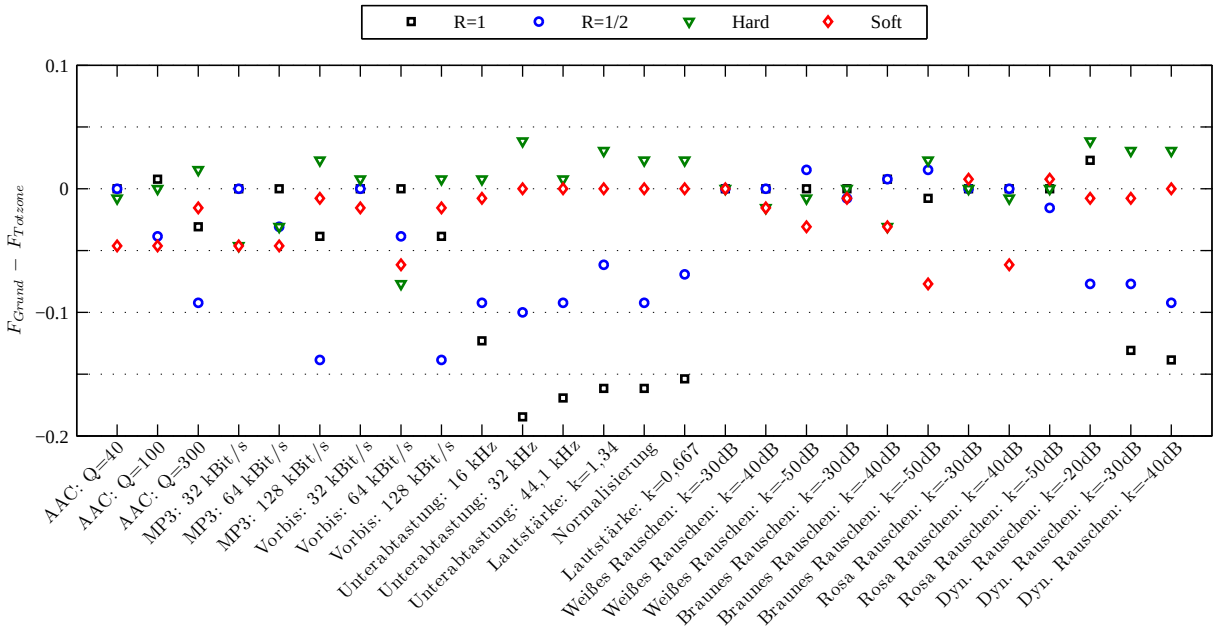


**Abbildung 5.24.:** Differenz zwischen den Fehlerraten der Nutzinformation pro Audiorahmen für das Grundsystem $F_{Grund}$ und dem Grundsystem mit Totzone für das Inhaltsmerkmal $F_{Totzone}$

mation. Die maximalen Bitfehlerraten der Wasserzeicheninformation steigen mit zunehmender Intensität der Störungen im Vergleich zum Grundsystem schneller an. Die *Soft-Input*-Decodierung stößt somit mit zunehmender Intensität der Störungen auch schneller an die Grenzen ihrer Leistungsfähigkeit.

## 5.3. Hierarchische Wasserzeicheneinbettung

Die Leistungsfähigkeit des entwickelten Wasserzeichenalgorithmus ist hauptsächlich durch die Segmentgröße der Wasserzeicheneinbettung $l_W$ und die Anzahl der für die Einbettung verwendeten Frequenzgruppen bestimmt. Die Einbettungsstärke $f$ dient nur dem Zweck, die Wasserzeichentransparenz auf einen bestimmten Wert einzustellen, typischerweise auf ein ODG größer -1. Wie in der Leistungsanalyse im Abschnitt 4.4 dargestellt, bewegt sich der günstige Arbeitsbereich für die Wasserzeicheneinbettung um Segmentlängen $l_W$ von 1024 und 1536 Samples. Ohne grundlegende Veränderungen des Algorithmus kann die Transparenz oder Robustheit des Wasserzeichens nur durch Reduzierung der für die Einbettung verwendeten Frequenzgruppen auf Kosten der Wasserzeichenkapazität erzielt werden.

Dieser Abschnitt behandelt einen Ansatz mit überlagernder Einbettung von mehreren 1-Bit-Wasserzeichen. Die Motivation für den Einbettungsansatz ist die Konzentration der Wasser-

zeichen auf eine geringere Anzahl an Frequenzgruppen und Verlagerung der Einbettungsdomain in einen für die Wahrnehmung weniger relevanten Frequenzbereich. Neben der Verlagerung der Einbettungsdomain stellt auch die Einschränkung der Wasserzeichendomain und somit die Reduzierung der für die Einbettung zu modifizierenden Bestandteile der Audiodaten eine die Wasserzeichentransparenz potentiell begünstigende Maßnahme dar. Der Sicherheitsaspekt des Authentifizierungssystems wird hierdurch nicht betroffen, da die Angriffsszenarien keine absichtliche Entfernung bzw. Zerstörung der Wasserzeicheninformation vorsehen. Für das Angriffsziel der Fälschung der Wasserzeicheninformation bringt die Verlagerung keine Vorteile.

Die Funktionsweise der überlagernden Wasserzeicheneinbettung ist eine wiederholte Anwendung des entwickelten 1-Bit Wasserzeichen Algorithmus auf seine eigenen Koeffizienten-Gruppen. Die Abbildung 5.25 verdeutlicht die Vorgehensweise bei der überlagernden Einbettung mehrerer 1-Bit Wasserzeichen. Der in Abschnitt 4.3.2 entwickelte Wasserzeichenalgorithmus verwendet als Population $\mathbb{C}$ zur Einbettung eines Wasserzeichen-Bits die Koeffizienten einer Frequenzgruppe. Diese wird während der Einbettung vollständig in zwei weitere Gruppen an Koeffizienten $\mathbb{A}$ und $\mathbb{B}$ unterteilt. Diese beiden Gruppen $\mathbb{A}$ und $\mathbb{B}$ dienen bei der überlappenden Einbettung jeweils wieder als Population für weitere Wasserzeichen. Die Zahl der für die Einbettung benötigten Frequenzgruppen und somit die Größe des Einbettungsbereichs kann bei gleichbleibender Kapazität verringert werden. Ein Entwicklungsziel des entwickelten Wasser-

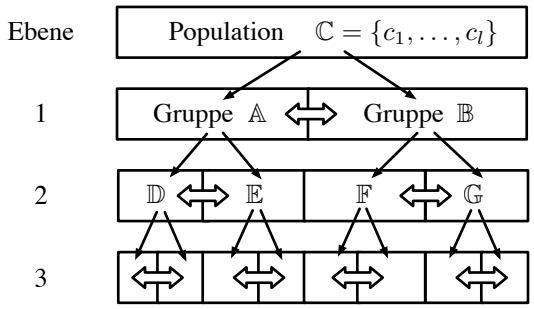


**Abbildung 5.25.:** Hierarchische Einbettungsstruktur für eine Einbettung über drei Ebenen.

zeichenalgorithmus fordert die Erhaltung des Inhaltsmerkmals, also die Summe der Koeffizientenbeträge einer Frequenzgruppe bzw. der Population *C* der Einbettung. Die Einbettung eines Wasserzeichens verändert nur die Summen der Koeffizientenbeträge der Gruppen $\mathbb{A}$ und $\mathbb{B}$ zur Herstellung der Distanz $\varepsilon$. Die Summe der Koeffizientenbeträge der Population $\mathbb{C}$ bleibt unverändert. Ausgehend von dieser Eigenschaft (5.47) ändert sich die Summe der Koeffizientenbeträge einer Gruppe $\mathbb{A}$ bzw. $\mathbb{B}$ durch die Einbettung eines Wasserzeichens in die Gruppe nicht,

da diese die Population des untergeordneten Wasserzeichens darstellt.

$$\sum_{i=1}^{n} |a'_i| + \sum_{i=1}^{m} |b'_i| = \sum_{i=1}^{l} |c'_i| = \sum_{i=1}^{l} |c_i| \tag{5.47}$$

Für die Einbettung eines weiteren Wasserzeichens können die Koeffizienten einer Gruppe vollständig oder auch nur teilweise verwendet werden. Die vollständige Nutzung ist in zweierlei Hinsicht empfehlenswert.

- Der Einbettungsalgorithmus baut auf der Hypothese auf, das die Koeffizienten-Gruppen $\mathbb{A}$ und $\mathbb{B}$ die gleichen Eigenschaften besitzen. Dies ist mit großer Wahrscheinlichkeit für eine große Anzahl an Elementen in den Koeffizienten-Gruppen gegeben. Zur Maximierung der Koeffizientenanzahl in den Gruppen eines untergeordneten Wasserzeichens ist ein vollständige Nutzung der Koeffizienten einer Gruppe des übergeordneten Wasserzeichens notwendig.

- Ein zweiter Aspekt, welcher für die vollständige Nutzung der Koeffizienten spricht, ist der Aufwand für den Einbettungsprozess. Die Einbettung der Wasserzeichen aller Ebenen erfolgt mit der Modifikation der Koeffizienten auf der letzten Ebene. Die notwendige Information für die Modifikation eines Koeffizienten ist die Summe der Koeffizientenbeträge der verwendeten Population $\mathbb{C}$ und die Einbettungsstärke $f$. Diese kann unter Kenntnis der einzubettenden Wasserzeicheninformation, der Größe der Population, welche vom Wasserzeichen auf der obersten Ebene genutzt wird und der Einbettungsstärke $f$ im Vorfeld der Modifikation bestimmt werden. Neben der Bestimmung der Zielwerte für die Summen der Koeffizientenbeträge der einzelnen Gruppen nach der Modifikation ergibt sich für die hierarchische Einbettung kein zusätzlicher Aufwand.

### 5.3.1. Modifizierte Gruppenbildung

Die Einbettung eines Wasserzeichenbits erfolgt über die Reduzierung der Koeffizientenbeträge einer Gruppe, während die Koeffizietenbeträge der anderen Gruppe erhöht werden. Eine Gruppenbildung, wie in Abbildung 5.25 dargestellt, resultiert in einer verminderten Robustheit des untergeordneten Wasserzeichens, welches die Gruppe als Population nutzt, deren Koeffizientenbeträge zur Einbettung des übergeordneten Wasserzeichens reduziert wurden. Die Intensität von Störungen durch die Verarbeitung der Audiodaten sind für alle Wasserzeichen einer Ebene der hierarchischen Einbettungsstruktur ähnlich, da sich alle Wasserzeichen die gleiche Ausgangspopulation an Koeffizienten teilen. Die Robustheit eines Wasserzeichens ist durch die Distanz $\varepsilon$ zwischen dessen Gruppen bestimmt. Solange eine Störung $\varepsilon$ nicht übersteigt, bleibt das

Wasserzeichen intakt. Ausgehend von (5.48) ist der Betrag von $\varepsilon$ einzig durch die Einbettungsstärke $f$ und die Summe der Koeffizientenbeträge der für die Einbettung genutzten Population an Koeffizienten bestimmt. Während die Intensität der Störung auf einer Ebene gleich bleibt, unterscheiden sich die Summen der Koeffizientenbeträge der Gruppen des übergeordneten Wasserzeichens und somit die Signalstärken der Wasserzeichen, welche die Gruppen des übergeordneten Wasserzeichens als Population verwenden. Das Wasserzeichen, welches die Gruppe mit den reduzierten Koeffizientenbeträgen nutzt, besitzt im Vergleich zum Wasserzeichen, welches die andere Gruppe als Population verwendet, ein vermindertes Signal-Rausch-Verhältnis und somit eine verminderte Robustheit.

$$\varepsilon = (-1)^w \cdot \sum_{i=1}^{n} |c_i| \cdot f \tag{5.48}$$

Zur Behandlung dieser Problematik ist eine modifizierte Gruppenbildung entwickelt worden. Ziel der modifizierten Gruppenbildung sind zwei gleich große Summen der Koeffizientenbeträge der beiden Gruppen $\mathbb{A}$ und $\mathbb{B}$ nach Einbettung des Wasserzeichens (s. (5.49)). Die Robustheit des untergeordneten Wasserzeichens ist somit unabhängig davon, welche Gruppe des übergeordneten Wasserzeichens als Population verwendet wird.

$$\sum_{i=1}^{l} |a'_i| = \sum_{i=1}^{m} |b'_i| \tag{5.49}$$

Für die modifizierte Gruppenbildung werden die Koeffizienten der Gruppen $\mathbb{A}$ und $\mathbb{B}$ in positiv ($\mathbb{A}^+$, $\mathbb{B}^+$) und negativ ($\mathbb{A}^-$, $\mathbb{B}^-$) agierende Gruppen unterteilt. Die Selektion von $\mathbb{A}^+$ und $\mathbb{A}^-$ erfolgt, wie bei der Selektion von $\mathbb{A}$ und $\mathbb{B}$ aus der Population $\mathbb{C}$, vollständig und disjunkt aus der Population von Koeffizienten der Gruppe $\mathbb{A}$. Die Selektion der Gruppen $\mathbb{B}^+$ und $\mathbb{B}^-$ erfolgt in Analogie aus der Population $\mathbb{B}$. Die Berechnungsgrundlagen für den Einbettungsalgorithmus bleiben weitgehend unverändert (vgl. Abschnitt 4.3.2). Die während der Einbettung zu erzeugende Distanz $\varepsilon$ ist für die vier Gruppen nach (5.50) definiert. Die Definition (5.51) des Wertes $\varepsilon$ und der mögliche Wertebereich (5.52) der Einbettungsstärke $f$ bleiben im Vergleich zur Einbettung über zwei Gruppen unverändert (vgl. (4.14) und (4.15)).

$$\left(\sum_{i=1}^{o} |a^{+\prime}_i| - \sum_{i=1}^{r} |a^{-\prime}_i|\right) - \left(\sum_{i=1}^{q} |b^{+\prime}_i| - \sum_{i=1}^{p} |b^{-\prime}_i|\right) = \varepsilon \tag{5.50}$$

$$\varepsilon = (-1)^w \cdot \sum_{i=1}^{n} |c_i| \cdot f \tag{5.51}$$

$$(0,1] := \{f \in \mathbb{R} \,|\, 0 < f \leq 1\} \tag{5.52}$$

Die Berechnung der Zielwerte der Summen der Koeffizientenbeträge der Gruppen $\mathbb{A}^{+\prime}$, $\mathbb{B}^{+\prime}$, $\mathbb{A}^{-\prime}$ und $\mathbb{B}^{-\prime}$, welche nach der Modifikation der Koeffizienten vorliegen sollen, ändert sich bis auf einen Faktor von 0,5 im Vergleich zu den Zielwerten (4.16) für zwei Gruppen nicht. Der Faktor von 0,5 ergibt sich daraus, dass die Koeffizienten der Population $\mathbb{C}$ jetzt in vier anstatt wie vorher in zwei Gruppen selektiert werden. Unabhängig von dem einzubettenden Wasserzeichen-Bit $w$ wird ein Teil der Koeffizientenbeträge der Gruppe $\mathbb{A}$ bzw. $\mathbb{B}$ reduziert, während der andere Teil der Koeffizientenbeträge um den gleichen Wert erhöht wird. Die Summe der Koeffizientenbeträge der Gruppen $\mathbb{A}$ und $\mathbb{B}$ haben nach der Einbettung des Wasserzeichens denselben Wert. Für ein untergeordnetes Wasserzeichen ergeben sich keine Nachteile hinsichtlich dessen Robustheit, ob dieses die Gruppe $\mathbb{A}$ oder $\mathbb{B}$ des übergeordneten Wasserzeichens als Population nutzt.

$$\sum_{i=1}^{o} |a^{+\prime}{}_i| = \sum_{i=1}^{p} |b^{-\prime}{}_i| = \frac{1}{4}\left(\sum_{i=1}^{n} |c_i| + \varepsilon\right) \tag{5.53}$$

$$\sum_{i=1}^{q} |b^{+\prime}{}_i| = \sum_{i=1}^{r} |a^{-\prime}{}_i| = \frac{1}{4}\left(\sum_{i=1}^{n} |c_i| - \varepsilon\right) \tag{5.54}$$

Die Modifikation eines einzelnen Koeffizienten erfolgt in Analogie der ursprünglichen Berechnungsvorschrift (4.19) bzw. (4.20). Die Berechnungsvorschriften für die Koeffizienten der Gruppen $\mathbb{A}^{+\prime}$, $\mathbb{B}^{+\prime}$, $\mathbb{A}^{-\prime}$ und $\mathbb{B}^{-\prime}$ sind in (5.55) bis (5.58) dargestellt.

$$a^{+\prime}{}_i = a^{+}{}_i \cdot \left(1 + \frac{a_i^+}{\sum_{i=1}^{o} |a^{+}{}_i|} \cdot \left(\frac{1}{4}\left(\sum_{i=1}^{n} |c_i| + \varepsilon\right) - \sum_{i=1}^{o} |a^{+}{}_i|\right)\right) \tag{5.55}$$

$$a^{-\prime}{}_i = a^{-}{}_i \cdot \left(1 + \frac{a_i^-}{\sum_{i=1}^{r} |a^{-}{}_i|} \cdot \left(\frac{1}{4}\left(\sum_{i=1}^{n} |c_i| - \varepsilon\right) - \sum_{i=1}^{r} |a^{-}{}_i|\right)\right) \tag{5.56}$$

$$b^{+\prime}{}_i = b^{+}{}_i \cdot \left(1 + \frac{b_i^+}{\sum_{i=1}^{q} |b^{+}{}_i|} \cdot \left(\frac{1}{4}\left(\sum_{i=1}^{n} |c_i| - \varepsilon\right) - \sum_{i=1}^{q} |b^{+}{}_i|\right)\right) \tag{5.57}$$

$$b^{-\prime}{}_i = b^{-}{}_i \cdot \left(1 + \frac{b_i^-}{\sum_{i=1}^{p} |b^{-}{}_i|} \cdot \left(\frac{1}{4}\left(\sum_{i=1}^{n} |c_i| + \varepsilon\right) - \sum_{i=1}^{p} |b^{-}{}_i|\right)\right) \tag{5.58}$$

Die Rückgewinnung des Wasserzeichen-Bits $w'$ erfolgt am Wasserzeichendecoder über den Vergleich der vier Gruppen nach (5.59).

$$w' = \begin{cases} 0 & \text{für } \left(\sum |a_i^{+\prime}| - \sum |a_i^{-\prime}|\right) - \left(\sum |b_i^{+\prime}| - \sum |b_i^{-\prime}|\right) \geqq 0 \\ 1 & \text{für } \left(\sum |a_i^{+\prime}| - \sum |a_i^{-\prime}|\right) - \left(\sum |b_i^{+\prime}| - \sum |b_i^{-\prime}|\right) < 0 \end{cases} \tag{5.59}$$

Durch die modifizierte Gruppenbildung nimmt die Komplexität der hierarchischen Einbettungsstruktur zu. Die grafische Darstellung der Struktur für eine Einbettung über zwei Ebenen ist in

Abbildung 5.26 ersichtlich. Die positiv und die negativ agierende Untergruppe dienen zusammen als Population für ein untergeordnetes Wasserzeichen. Die Berechnung der Zielwerte für die Modifikationen der Koeffizienten müssen jedoch getrennt für beide Untergruppen vorgenommen werden. Die Anzahl der Untergruppen verdoppelt sich nochmals mit jeder weiteren Einbettungsebene.

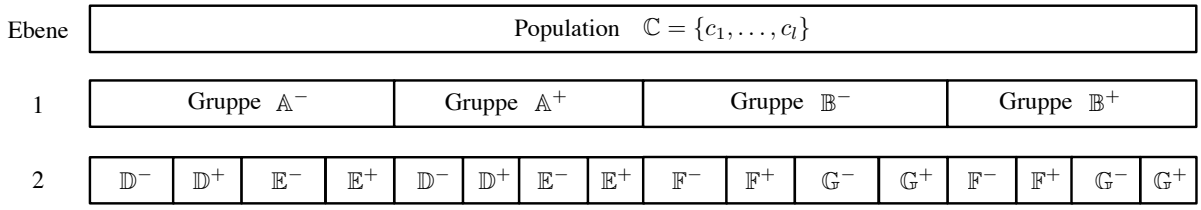


**Abbildung 5.26.:** Hierarchische Einbettungsstruktur mit modifizierter Gruppenbildung für eine Wasserzeicheneinbettung über zwei Ebenen.

### 5.3.2. Leistungsanalyse

Die Leistungsfähigkeit der entwickelten hierarchischen Einbettung erfolgt in Analogie zur Leistungsanalyse des Grundsystems (Abschnitt 4.5) und der Leistungsanalyse der Merkmalsverstärkung (Abschnitt 5.1.3). Die Einbettungsdomain des Wasserzeichens umfasst die beiden Frequenzgruppen $F19$ und $F20$. Die hierarchische Einbettung wird über zwei Ebenen vorgenommen. In eine Frequenzgruppe werden 3 Bits und somit insgesamt 6 Bits pro Wasserzeichensegment eingebettet. Die Kapazität entspricht der Kapazität des Grundsystems bzw. des Grundsystems mit Merkmalsverstärkung, welche ein Bit Kapazität pro Frequenzgruppe aufweisen und die Frequenzgruppen $F15$ - $F20$ verwenden. Die Extraktionsdomain des Inhaltsmerkmales umfasst die Frequenzgruppen $F2$-$F23$ und verbleibt identisch zu der Extraktionsdomain des Grundsystems mit bzw. ohne Merkmalsverstärkung. Die Transparenz der Wasserzeicheneinbettung wird mit dem Zielwert von -1 ODG eingestellt. Die vergleichende Bewertung der Leistungsfähigkeit der hierarchischen Einbettung mit dem Grundsystem bzw. dem Grundsystem mit Merkmalsverstärkung erfolgt bei identischer Kapazität und Transparenz über die Wasserzeichenrobustheit.

Wie bei dem Grundsystem ergibt sich auch für die hierarchische Einbettung ein günstiger Arbeitsbereich für Segmentlängen $l_W$ von 1024 bzw. 1536 Samples. Die Abbildung 5.27 zeigt hierzu die Ergebnisse der Robustheitsanalyse der Wasserzeicheninformation für zulässige Störungen bei einer Wasserzeichentransparenz von ODG = -1 in Abhängigkeit der Segmentlänge $l_W$ unter Verwendung der Frequenzgruppen $F19$-$F20$ bei einer hierarchischen Einbettung über

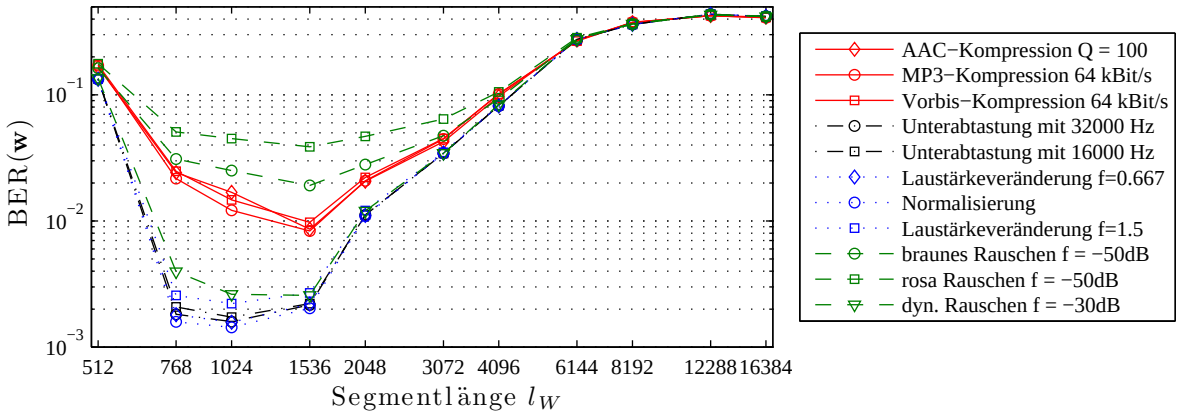


**Abbildung 5.27.:** Bitfehlerrate der Wasserzeicheninformation bei einem ODG von -1 in Abhängigkeit der Segmentlänge $l_W$ unter Verwendung der Frequenzgruppen $F19$-$F20$ bei einer hierarchischen Einbettung über 2 Ebenen.

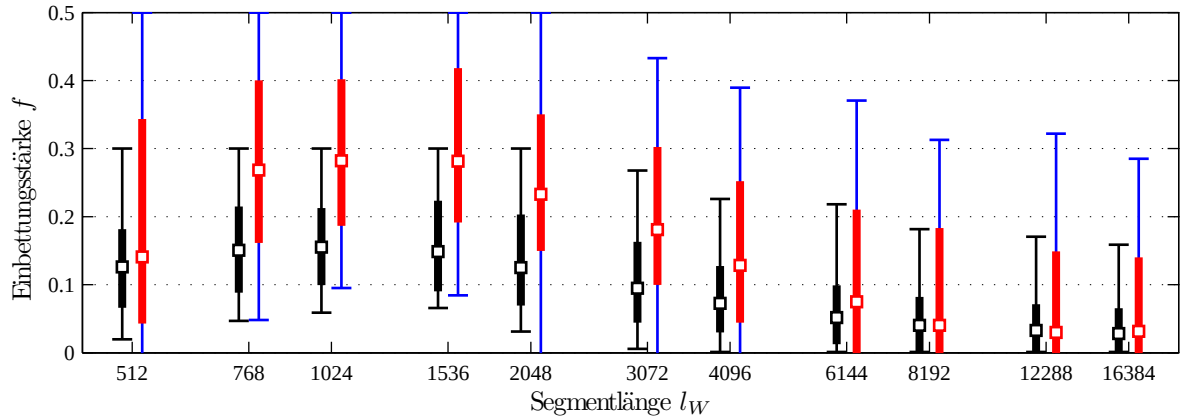


**Abbildung 5.28.:** Wertebereich I(I), Bereich zw. dem 15,85%- und 84,15%-Fraktil ■(■) und Mittelwert □(□) der Einbettungsstärke $f$ zur Einstellung einer Wasserzeichentransparenz von ODG$\approx -1$ in Abhängigkeit der Segmentlänge $l$ unter Verwendung der Frequenzgruppen $F19$-$F20$ bei einer hierarchischen Einbettung über 2 Ebenen (für das Grundsystem).

zwei Ebenen. Wie bei der Leistungsanalyse des Grundsystems wurde die Einbettungsstärke $f$ zur Einstellung einer Ziel-Transparenz von einem ODG = -1 mittels linearer Interpolation zwischen Stützstellen mit fest definierten Werten der Einbettungsstärke $f$ im Bereich von 0.5 und 0.0001 ermittelt. Die mit dieser Methode ermittelten Einbettungsstärken $f$ sind in Abbildung 5.28 dargestellt. Die erzielten ODG-Werte der Wasserzeichentransparenz sind in Abbildung 5.29 ersichtlich. Die für die hierarchische Einbettung ermittelten Werte der Einbettungsstärke $f$ fallen im Mittel größer aus im Vergleich zum Grundsystem. Die anvisierte Transparenz mit einem ODG von -1 wird jedoch nur noch für Segmentlängen $l_W$ von 768, 1 024 und 1 536 erreicht. Mit größer werdenden Segmentlängen $l_W$ gewinnen die Störungen der Wasserzeicheneinbettung einen eher globalen Charakter und unterscheiden sich stärker von den lokal vorherrschenden Eigenschaften der Trägerdaten. Die Abhängigkeit der Einbettungsstärke $f$ und somit der

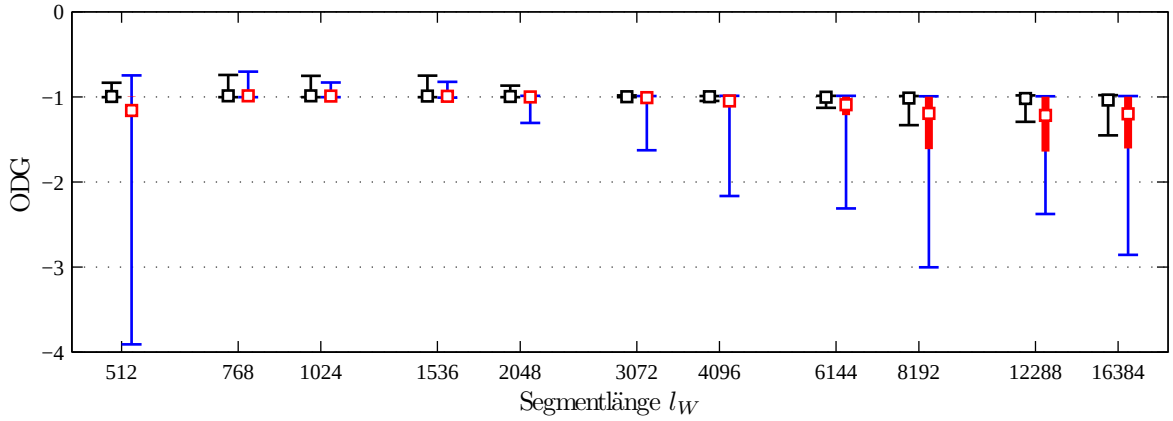


**Abbildung 5.29.:** Wertebereich I(I), Bereich zw. dem 15,85%- und 84,15%-Fraktil ■(■) und Mittelwert □(□) der erzielten Wasserzeichentransparenz gemessen in ODG in Abhängigkeit der Segmentlänge $l$ unter Verwendung der Frequenzgruppen $F19$-$F20$ bei einer hierarchischen Einbettung über 2 Ebenen (für das Grundsystem).

Wasserzeichentransparenz vom Inhalt der jeweiligen Trägerdatei nimmt bei der hierarchischen Einbettung ebenfalls zu. Der Einfluss der Segmentlänge auf die Wasserzeichentransparenz wurde bereits bei der einfachen Einbettung des Grundsystems beobachtet. Die Eigenschaften eines Segments haben bei der hierarchischen Einbettung einen noch stärkeren Einfluss auf die Audioqualität. Nur für die Segmentlänge $l_W$ von 768, 1 024 und 1 536 konnte eine ausreichende Wasserzeichentransparenz erzielt werden. Die Reduzierung und Verlagerung der Einbettungsdomain auf bzw. in den Frequenzbereich der Frequenzgruppen $F19$-$F20$ ermöglicht hier jedoch eine Erhöhung der Einbettungsstärke $f$.

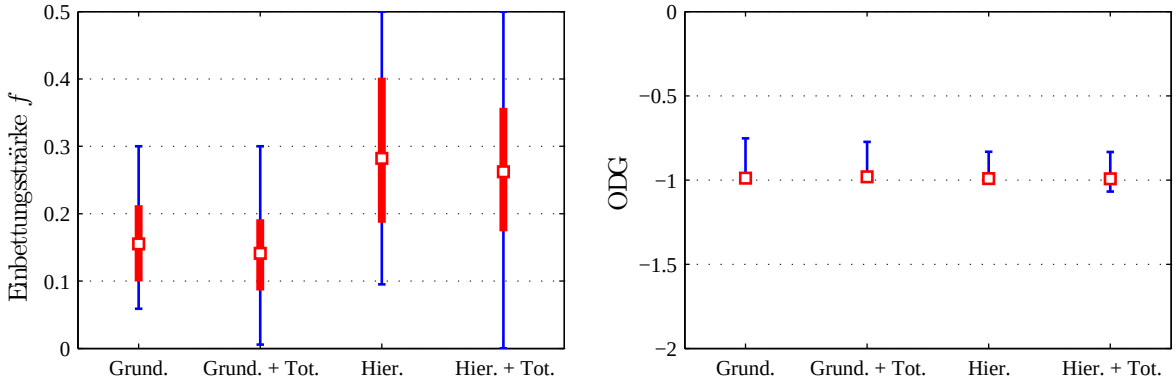


**Abbildung 5.30.:** Gegenüberstellung der Einbettungsstärken $f$ zur Einstellung einer Wasserzeichentransparenz von $ODG \approx -1$ (links) und die erzielte Wasserzeichentransparenz (rechts) des Grundsystems und des Systems mit hierarchischer Einbettung mit/ohne Merkmalsverstärkung bei einer Segmentlänge $l_W = 1\,024$ (Wertebereich I, Bereich zw. dem 15,85%- und 84,15%-Fraktil ■ und Mittelwert □).

Die Analyse der Grundsystems baut auf einer Segmentlänge von $l_W = 1\,024$ Samples auf. Um eine gemeinsame Vergleichsbasis herzustellen, wird die Leistungsanalyse der hierarchischen

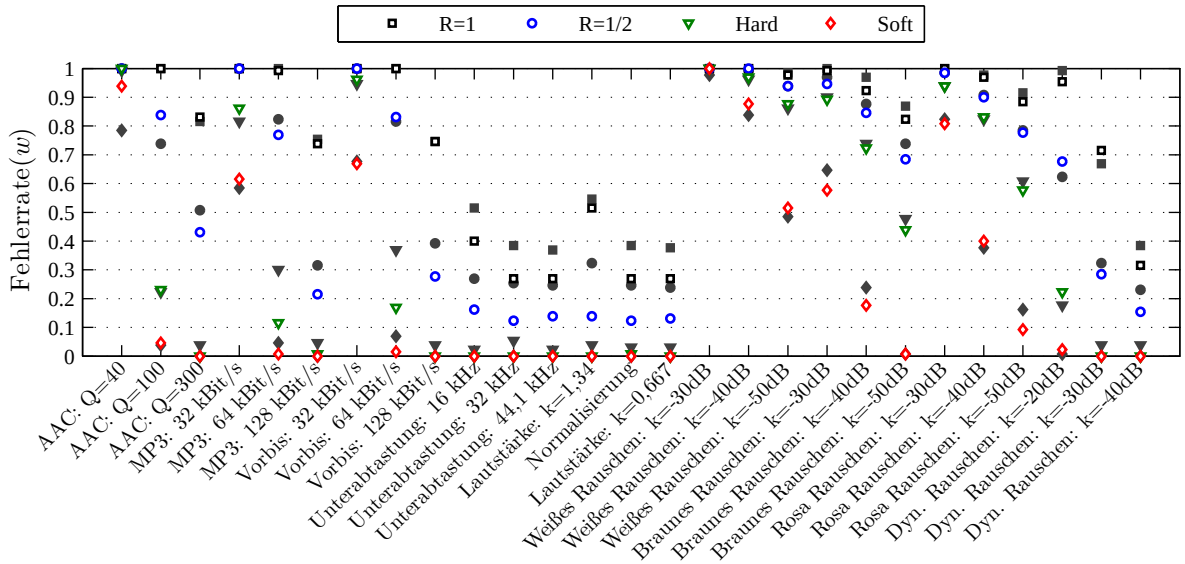


**Abbildung 5.31.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Wasserzeicheneinbettung (zusätzlich für das Grundsystem - graue Symbole)

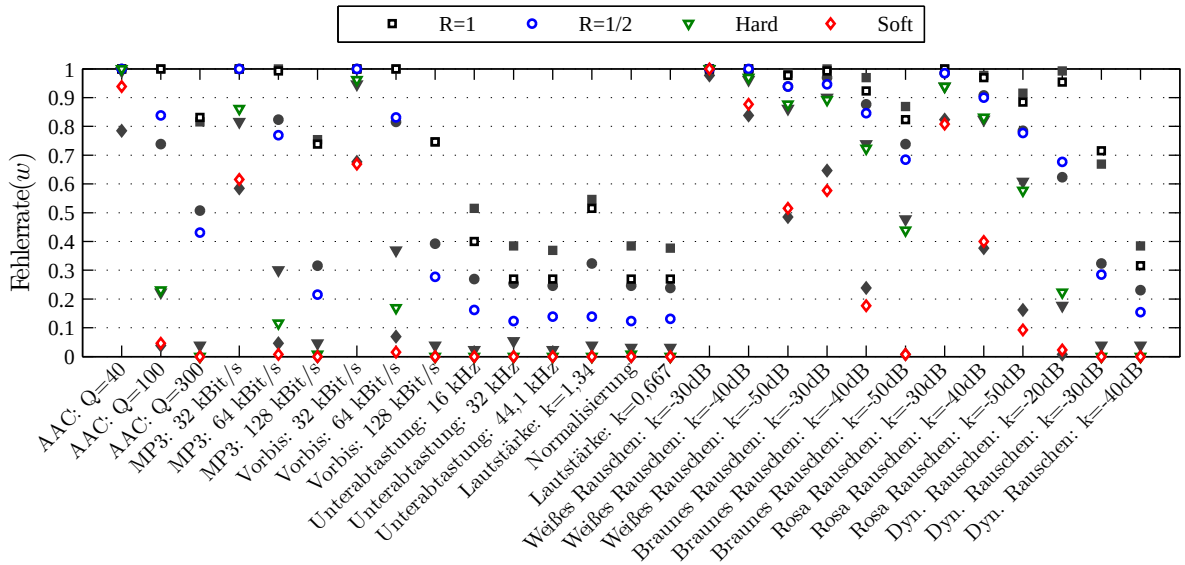


**Abbildung 5.32.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Wasserzeicheneinbettung (zusätzlich für das Grundsystem - graue Symbole) mit Totzone

Einbettung ebenso mit der Segmentlänge $l_W = 1\,024$ Samples durchgeführt. Die Integration der Merkmalsverstärkung erfolgt für die hierarchische Einbettung mit dem Wert 0,0001 für die Schwelle $Th_{tot}$ (vgl. Abschnitt 5.1.3). In der Abbildung 5.30 wird für die Segmentlänge $l_W = 1\,024$ Samples die Variation der Einbettungsstärke $f$ zur Einstellung einer Wasserzeichentransparenz mit einem ODG von etwa -1 und die erzielten Wasserzeichentransparenzen für alle vier Konfigurationen des Systems gegenübergestellt. Auszüge der in Anhang A.7 aufgeführten

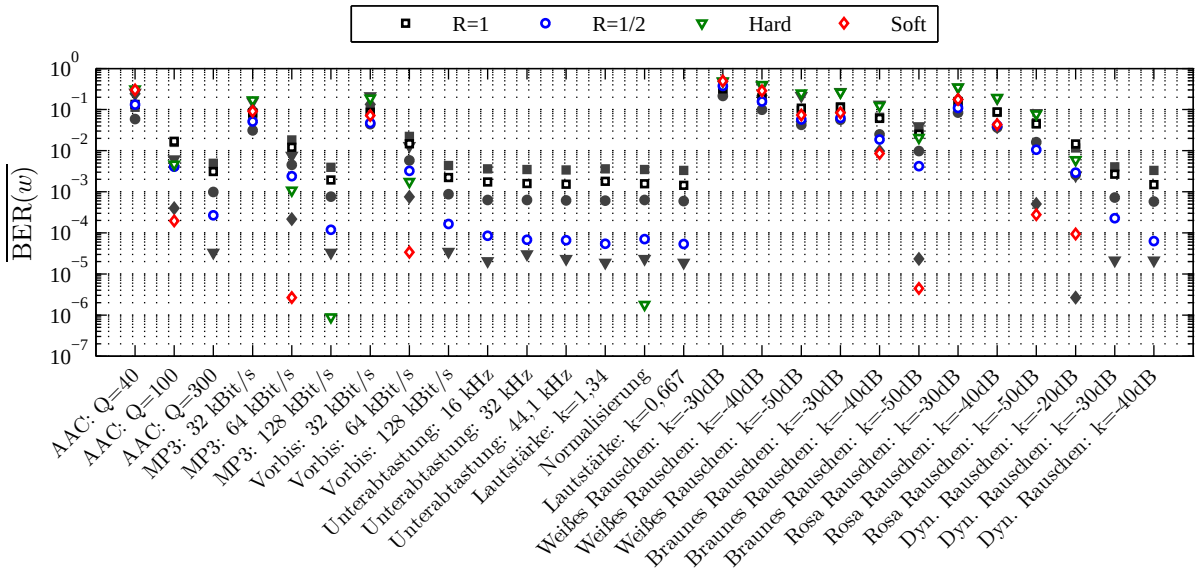


**Abbildung 5.33.:** Bitfehlerrate der Nutzinformation bei hierarchischer Wasserzeicheneinbettung (zusätzlich für das Grundsystem - graue Symbole)

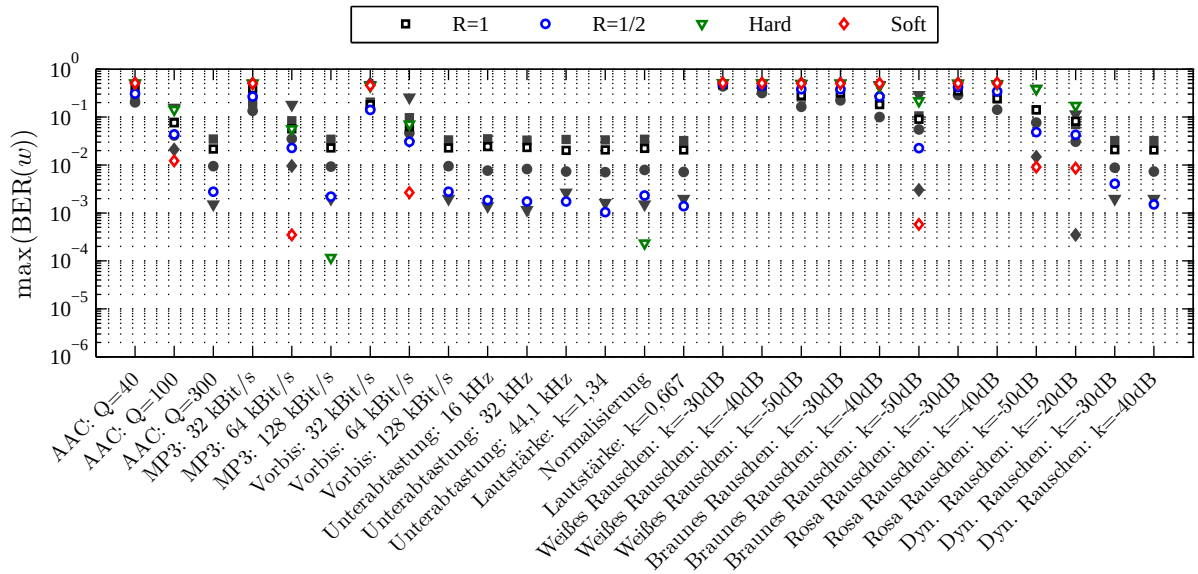


**Abbildung 5.34.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Wasserzeicheneinbettung (zusätzlich für das Grundsystem - graue Symbole)

Ergebnisse der Robustheitsanalyse der hierarchischen Einbettung sind den Abbildungen 5.31 bis 5.37 zu entnehmen.

Mit der hierarchischen Einbettung wird die Robustheit der Wasserzeicheninformation weiter gesteigert. Bei allen Störungen mit starkem Tiefpass-Charakter, wie z.B. starker Kompression oder Unterabtastung, zeigt sich die hierarchische Einbettung durch die Nutzung der höheren Frequenzbereiche der einfachen Einbettung unterlegen. Dies sind jedoch ausschließlich unzu-

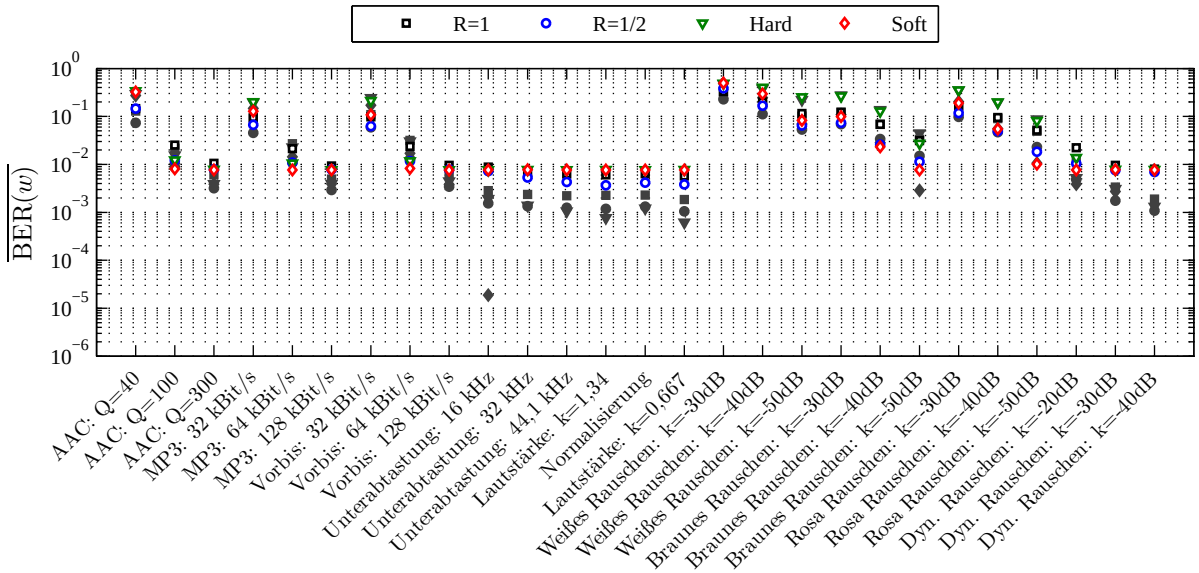


**Abbildung 5.35.:** Bitfehlerrate der Nutzinformation bei hierarchischer Wasserzeicheneinbettung (zusätzlich für das Grundsystem - graue Symbole) mit Totzone

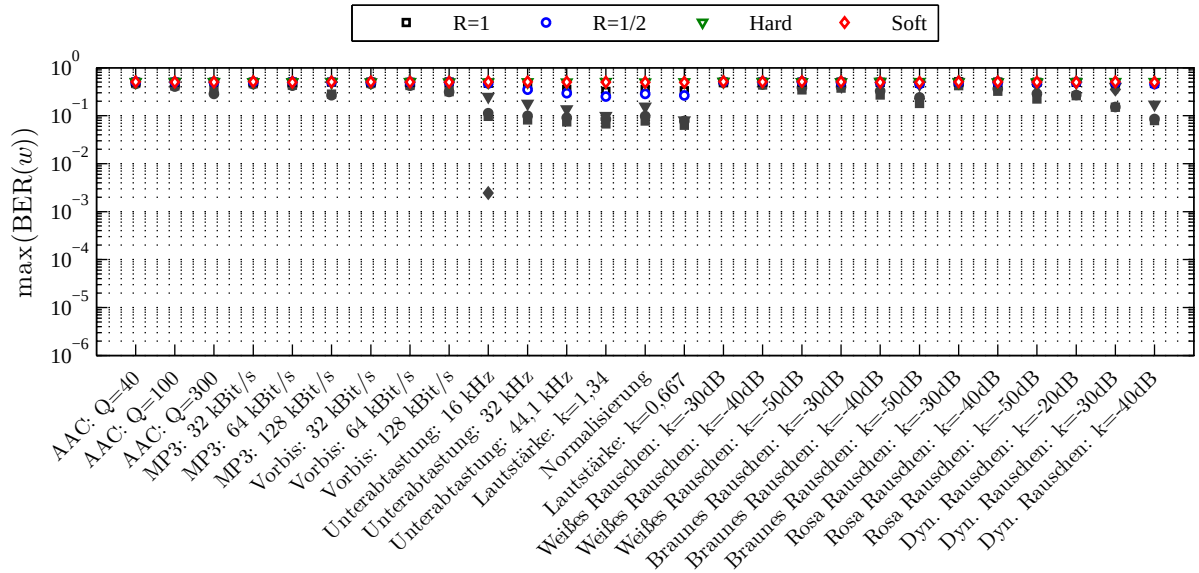


**Abbildung 5.36.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Wasserzeicheneinbettung (zusätzlich für das Grundsystem - graue Symbole) mit Totzone

lässige Störungen. Der Nachteil der Merkmalsverstärkung wird in Kombination mit der hierarchischen Einbettung besonders deutlich. Für eine der Testdateien ist die Einbettungsstärke $f$ so stark abzusenken (s. Abbildung 5.30), dass keine robuste Wasserzeicheneinbettung möglich ist. Die Bitfehlerrate der Wasserzeicheninformation liegt hier bei etwa 0,5 (s. Abbildung 5.36).

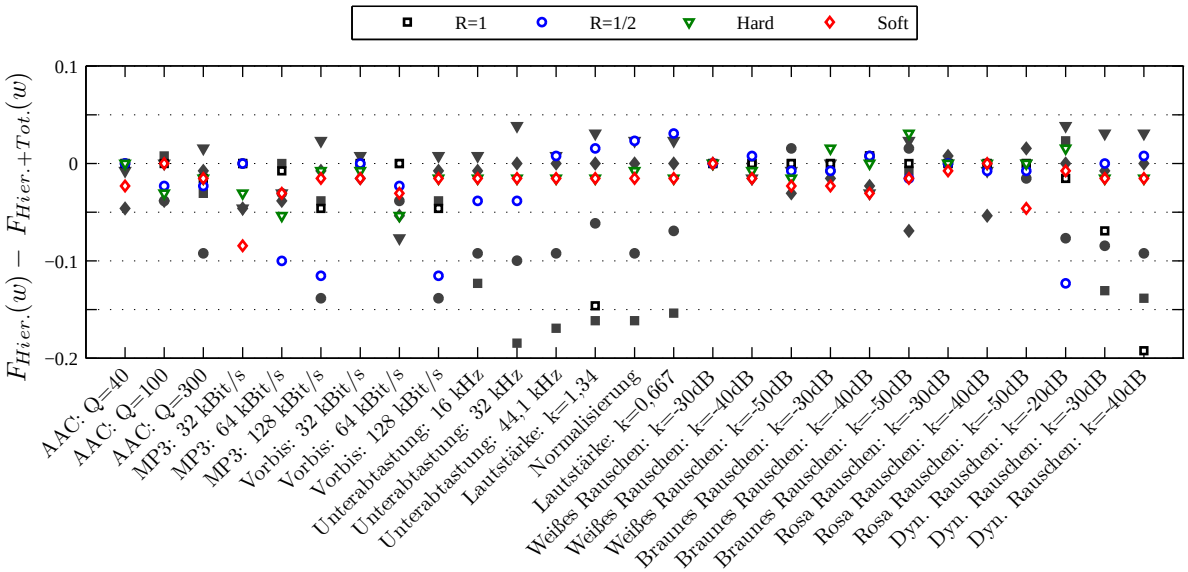


**Abbildung 5.37.:** Differenz zwischen den Fehlerraten der Nutzinformation pro Audiorahmen bei hierarchischer Wasserzeicheneinbettung $F_{Hier.}$ (zusätzlich für das Grundsystem - graue Symbole) und bei hierarchischer Wasserzeicheneinbettung mit Totzone für das Inhaltsmerkmal $F_{Totzone}$

## 5.4. Fazit

Im dem vorliegenden Kapitel wurde das im Kapitel 4 entwickelte Grundsystem durch mehrere Systemerweiterungen verbessert. Alle Erweiterungen können unabhängig voneinander mit dem Grundsystem kombiniert werden.

Die erste Erweiterung befasste sich mit der Verbesserung der Robustheit des Inhaltsmerkmals in Audiopassagen mit geringer Signalamplitude. Stille bzw. leise Bereiche ohne verständlichen Inhalt besitzt wie ein gesprochenes Wort, Musik, Geräusche, etc. eine Bedeutung und stellten somit einen zu schützenden Inhalt von Audiodaten dar. Diese Bereiche sind jedoch kritisch für die Robustheit des entwickelten Beschreibungsmerkmals. Aufgrund der geringen Signalamplituden in stillen Bereichen führen bereits geringe, nicht wahrnehmbare Veränderungen der Audiodaten und somit zulässige Störungen zu Veränderungen des Inhaltsmerkmals. Die Bestimmung des Inhaltsmerkmals beinhaltet eine Schwellwertentscheidung. Damit sich das Ergebnis der Schwellwertentscheidung nach geringen Störungen der Audiodaten nicht ändert, wird bei der Einbettung des Wasserzeichens eine Totzone in der Nachbarschaft der Entscheidungsschwelle erzeugt.

Eine nicht wahrnehmbare Modifikation der Trägerdaten zur Ausprägung eines bereits vorhanden Merkmals ist geeignet, die robuste Wiedererkennung des Beschreibungsmerkmals nach inhalts- und qualitätserhaltenden Störungen zu verbessern. Die Rate der falschen Rückweisung

von Audiodaten im Verifikationsprozess der Integrität wird verringert. Quantisierungsbedingte Fehler des Beschreibungsmerkmals unmittelbar nach der Wasserzeicheneinbettung können ausgeschlossen werden. Fehler durch eine Wertebereichsbegrenzung von Audiodaten (*clipping*) bleiben jedoch bestehen.

Die Verschlüsselung der mit dem Wasserzeichen zu übertragenden Nachricht stellt besondere Anforderungen an die Robustheit der Wasserzeichentechnik. Die Entschlüsselung einer Nachricht ist nur bei einer fehlerfreien Übertragung der verschlüsselten Nachricht möglich. Für den Integritätsschutz der Audiodaten muss das Wasserzeichen jedoch nur solange robust sein, wie keine unzulässigen Störungen der Trägerdaten vorliegen. Wie für das Beschreibungsmerkmal stellen Stille bzw. leise Audiobereiche eine Herausforderung dar. Stille Audiobereiche können mit der entwickelten Wasserzeichentechnik nicht robust markiert werden. Wie jedes Übertragungssystem benötigen auch die digitalen Wasserzeichen einen Träger. Fällt der Träger weg bzw. geht dessen Amplitude gegen Null, ist eine Übertragung nicht mehr möglich bzw. aufgrund des geringen Signal-Rausch-Verhältnis stark fehlerbehaftet. Sprechpausen, also stille Bereiche, sind besonders bei Sprache typisch und auch nicht vermeidbar. Eine robuste Übertragung einer Nachricht mit Wasserzeichen ohne den Ausschluss von Stille kann nicht gewährleistet werden. Die Kanalcodierung stellt mit fehlerkorrigierenden Maßnahmen eine Lösung für dieses Problem dar. Eingesetzt wurde ein Faltungscode mit einer *Soft-Input*-Decodierung. Für eine effektive Fehlerkorrektur wurde die Generierung einer Zuverlässigkeitsinformation aus der gestörten Wasserzeicheninformation entwickelt. Mit der *Soft-Input*-Decodierung konnte für einen Großteil der zulässigen Störungen eine vollständige Robustheit der Nutzinformation erzielt werden. Für die Vergleichsverfahren wurde dies für keine Störung erreicht. Verbesserungsbedarf besteht bei Störungen in Form von verlustbehafteter Kompression. Die entwickelte Zuverlässigkeitsinformation ist bei der Fehlerdecodierung grundsätzlich anzuwenden. Die katastrophalen Korrektureigenschaften bei starken Störungen der Trägerdaten stellen für die *Soft-Input*-Decodierung kein Ausschlusskriterium dar.

Generell gilt für eine möglichst robuste Einbettung eines Wasserzeichens, dieses in die Bestandteile der Audiodaten einzubetten, welche für die Wahrnehmung des Dateninhalts relevant sind. Eine Zerstörung bzw. das Entfernen des Wasserzeichens geht hier mit der Zerstörung des Dateninhalts oder wenigsten mit einer wahrnehmbaren Minderung der Qualität einher. Des Weiteren ist in diesem Bereich mit den geringsten Störungen durch Operationen wie verlustbehafteter Kompression zu rechnen. Für den vorliegenden Anwendungsfall des Integritätsschutzes können jedoch auch weniger relevante Bereiche verwendet werden. Zum einen liegt als Angriffsszenario nicht das Entfernen des Wasserzeichens vor. Das Hauptangriffsszenario bildet das Fälschen einer gültigen Wasserzeicheninformation. Zum anderen ist die Robustheit nur für inhaltserhaltende Operationen notwendig. Für den Integritätsschutz des Inhaltes von Audiodaten stellen

die für die Wahrnehmung weniger relevanten Frequenzbereiche der Trägerdaten eine günstige Einbettungsdomain für ein Wasserzeichen dar. Durch eine überlagernde Einbettung multipler Wasserzeichen wurde der für die Einbettung aller Wasserzeichen verwendete Frequenzbereich verringert und in einen für die Wahrnehmung weniger relevanten Frequenzbereich verlagert. Mit einer hierarchischen Struktur der Einbettungsdomain wurden gegenseitige Störungen der Wasserzeichen verhindert. Die Wasserzeichentransparenz wird durch gezielte Reduzierung und Verlagerung des Frequenzbereichs positiv beeinflusst, wodurch die Einbettungsstärke zugunsten der Wasserzeichenrobustheit erhöht werden kann. Der Transparenzgewinn ist durch die Möglichkeiten bedingt, den Einbettungsbereich auf für die Wahrnehmung weniger relevante Bereiche zu konzentrieren. Für Szenarien mit starker Kompression oder Störungen mit starken Tiefpass-Charakter ist die hierarchische Einbettung ungeeignet.

Kapitel 6

# Zusammenfassung & Ausblick

## 6.1. Zusammenfassung

Das Ziel dieser Arbeit bestand in der Entwicklung eines Authentifizierungssystems für digitale Audiodaten, welches eine öffentliche Verifizierung erlaubt. Das Grundkonzept des Systems basiert auf dem Verfahren der inhalts-fragilen Wasserzeichen. Die Integrität der Audiodaten wird hier über Beschreibungsmerkmale der Audiodaten geführt, welche mit einem digitalen Wasserzeichen in die Audiodaten eingebettet werden.

Anhand der Fragilität/Robustheit des Beschreibungsmerkmals sollen inhaltliche Veränderungen als auch qualitative Störungen der Audiodaten detektiert werden können. Es wurde erkannt, dass ein Beschreibungsmerkmal in Form der zeitlichen Veränderungen der Betragssummen der Frequenzgruppen-Koeffizienten von Audiodaten diesen Anspruch erfüllt. Um die Einbettung des Beschreibungsmerkmals aus Kapazitätsgründen des Wasserzeichens zu ermöglichen, wurde ausschließlich das Vorzeichen der Differenz von den Betragssummen der Frequenzgruppen-Koeffizienten von aufeinanderfolgenden Segmenten in binär codierter Form verwendet.

Die entwickelte Wasserzeichentechnik gehört zu der Gruppe der *Patchwork*-Verfahren. Als Einbettungsdomain werden, wie für die Extraktion des Beschreibungsmerkmals, die Frequenzgruppen-Koeffizienten des DCT-IV transformierten Audiosignals verwendet. Der Einbettungsalgorithmus wurde mit dem Ziel entwickelt, dass die Modifikationen der Trägerdaten, insbesondere auch die Modifikation der für die Extraktion des Beschreibungsmerkmals genutzten Koeffizienten transparent für eine erneute Extraktion des Beschreibungsmerkmals sind. Die Konkurrenz der Systemelemente um die Arbeitsdomain wurde somit eliminiert.

Anhand umfangreicher Simulationen wurden die Frequenzgruppen und Segmentierungslängen der Audiodaten identifiziert, welche günstige Extraktionsbereiche für das Beschreibungsmerkmal bzw. Einbettungsbereiche für das Wasserzeichen darstellen. Die beiden Systemelemente wurden trotz unterschiedlicher Arbeitsbereiche ohne Verlust ihrer einzelnen Leistungsfähigkeit in einen Grundsystem kombiniert.

Durch die Einführung einer Totzone für das Beschreibungsmerkmal konnte dessen Fragiliät gegenüber zulässigen Störungen in homogenen sowie leisen bis stillen Audiobereichen verbessert werden. Ebenfalls wurden durch diesen Ansatz quantisierungsbedingte Fehler des Beschreibungsmerkmals nach Einbettung des Wasserzeichens minimiert.

Zum Schutz gegen Fälschungen des eingebetteten Beschreibungsmerkmales sieht das Systemkonzept eine asymmetrische Verschlüsselung des Beschreibungsmerkmals und zusätzlicher Meta-Informationen vor. Die Verschlüsselung benötigt jedoch eine vollständige Robustheit der Wasserzeicheninformation bei zulässigen Störungen der Audiodaten. Um diese Anforderungen zu erfüllen, wurden fehlerkorrigierende Maßnahmen in Form einer *Soft-Input*-Decodierung entwickelt. Kernpunkt der Entwicklung war die Generierung einer Zuverlässigkeitsinformation aus der gestörten Wasserzeicheninformation für eine effiziente *Soft-Input*-Decodierung. Der Codierungsgewinn durch die Zuverlässigkeitsinformation wurde im Vergleich zu einer *Hard-Input*-Decodierung und dem Fall ohne Fehlerkorrektur nachgewiesen.

Um eine weitere Leistungssteigerung des Systems zu erzielen, wurde eine hierarchische Wasserzeicheneinbettung entwickelt. Durch eine überlagernde Einbettung mehrerer Wasserzeichen konnte der Einbettungsbereich verringert und in einen für die Wahrnehmung weniger störenden Frequenzbereich verlagert werden. Die verbesserte Transparenz dieser Konfiguration konnte zu Gunsten der Wasserzeichenrobustheit ausgenutzt werden.

In der Tabelle 6.1 sind die wesentlichen Eckpunkte vergleichbarer Verfahren aufgeführt. Ein direkter Vergleich der Verfahren untereinander ist aufgrund der unterschiedlichen Parameter der Leistungsanalyse (Testdaten, Störungen, Metriken, etc.) nur bedingt möglich. Bei vergleichbaren Daten wurden die zwei besten Vertreter markiert, wobei das entwickelte Verfahren immer einen dieser Vertreter darstellt.

**Tabelle 6.1.:** Vergleich von Verfahrensparameter des entwickelten Systems mit denen bekannter Verfahren

| **Autor** | **Gulbis** | **Steinebach et al.** [SD03, Ste04] | **Zmudszinski et al.** [ZS09b] | **Wang, Liao & Chen** [WLC07] | **Wang & Fan** [WF10] |
|---|---|---|---|---|---|
| **Integrität** | qualitativ - inhaltlich | qualitativ - inhaltlich | qualitativ | qualitativ oder inhaltlich[1] | inhaltlich |
| **öffent. Verifizierung** | ja | nein | nein | ja | nein |
| **Merkmal** | | | | | |
| Art | | RMS | rMAC | spektrale Maxima | Schwerpunkt |
| Verschlüsselung | **asym. 8192 Bit** | k.A. | **sym. 128 Bit** | RSA (1 080 Bit[2]) | XOR |
| Variation | **44 Bit** | 4 Bit / 8 Bit [Ste04, S. 91] | **128 Bit** | 20 Bit | 4 Bit |
| Lokalität | **8 192 Samples** ∼ **0,18 s** | 98 304 Samples 2,23 s $2^{16}$ Samples / 1,49 s [Ste04, S. 91] | 3 s;5 s | ∼ 220k Samples 5 s | **4 096 Samples** ∼ **0,1 s** |
| Extraktionsbereich | 0,1 - 12 kHz | 2 - 6 kHz | 0,1 - 10 kHz | >27,7 Hz | |
| **Wasserzeichen** | | | | | |
| Transparenz | -1 ODG | k. A | k. A | k.A. | ODG < -0,07 SNR < 51 dB |
| Kapazität | **262,5 Bit/s** | | | 20 Bit/s | |
| Nutzbitrate | **131,25 Bit/s** | 2,6917 Bit/s | 27,6 Bit/s[3] | 4 Bit/s | ∼ **43 Bit/s**[4] |
| Kanalcodierung | Faltungscode mit *Soft-Decision*-Decodierung | k. A. | k. A. | Wiederholung | k. A. |
| Domain | DCT-IV (10 - 14 kHz) | FFT (10 - 14 kHz) | FFT | FFT (1-20 Hz) | DWT/DCT |
| **Testdaten** | | | | | |
| Art | vorwiegend Sprache mit u. ohne Musik | breites Spektrum an Musik und Sprache | breites Spektrum an Musik und Sprache | populäre Musik, Sprache | mit/ohne Grundschlag |
| Anzahl | **130** | **125** | | 1 | 2 |
| Gesamtdauer | **2,22 h** | **1,5 - 7,5 h** | k. A. | 283 s | 190 s |
| tech. Daten | mono, 16 Bit, 48 kHz | mono, 16 Bit, 44,1 kHz | | | |

[1] Der Autor hält sich bei den Eigenschaften des Inhaltsmerkmales bedeckt.
[2] Theoretischer Maximalwert der Schlüssellänge für die beschriebene Testumgebung.
[3] Bezogen auf einen Extraktionsbereich von 5 Sekunden und einer rMAC-Länge von 128 Bit und 10 Bit Block Index.
[4] Theoretisch notwendige Nutzbitrate.

## 6.2. Ausblick auf zukünftige Arbeiten

Stille und Bereiche ohne wahrnehmbaren Inhalt stellen Herausforderungen für das entwickelte Beschreibungsmerkmal und die entwickelte Wasserzeichentechnik dar. Durch die geringen Amplitudenpegel der Audiodaten liegt auch bei leichten, zulässigen Störungen ein geringes Signal-Rausch-Verhältnis vor. Mit der entwickelten Merkmalsverstärkung und der hierarchischen Einbettung wurde eine Verbesserung des Signal-Rausch-Verhältnis angestrebt.

Anstatt das Signal-Rausch-Verhältnis in den stillen Bereich zu verbessern, gilt es zu untersuchen, ob ein Ausschluss von stillen Bereichen sich gewinnbringend für das entwickelte Verfahren zeigt. Für die Merkmalsbestimmung bedeutet dies, dass die Übertragung der Inhaltsmerkmale für die stillen Bereiche entfällt. Für den Integritätsschutz sind jedoch Umfang von Position der stillen Bereiche neben den Inhaltsmerkmalen der übrigen Bereiche mit dem Wasserzeichen zu übertragen. Für die Wasserzeicheneinbettung werden zwei potenziell günstige Ansätze gesehen. In einem Ansatz wird davon ausgegangen, dass die Modifikation der Audiodaten in den Stillebereichen überflüssig ist, da keine zuverlässige Wasserzeicheninformation ausgelesen werden kann. Der hier entstehende Fehler wird über fehlerkorrigierende Maßnahmen behandelt. Die Wasserzeicheneinbettung wird zugunsten der Wasserzeichentransparenz und somit zugunsten einer stärkeren Einbettung in den übrigen Bereichen unterlassen. Auch im zweiten Ansatz wird in stillen Bereichen keine Wasserzeicheneinbettung vorgenommen. Die Bitinformation wird hier jedoch mit dem nächsten Wasserzeichen eingebettet. Fehlerkorrigierende Maßnahmen sind hier nicht gefordert. Eine fehlerhafte Erkennung der stillen Bereiche würde hier zur Desynchronisation der Bitreihenfolge der Wassserzeicheninformation führen.

# Abbildungsverzeichnis

# Tabellenverzeichnis

# Literaturverzeichnis


[Arn00] ARNOLD, Michael: Audio watermarking: Features, applications and algorithms. In: *IEEE Int. Conf. Multimedia and Expo 2000* Bd. 2, 2000, S. Seitren 1013 – 1016

[Arn04] ARNOLD, Michael: *Digital Audio Watermarking*. Logos Verlag Berlin, 2004

[ASU] ARIZONA STATE UNIVERSITY: *Video Trace Library*. `http://trace.eas.asu.edu`. – (10.09.2010)

[BBB+07] BARKER, Elaine ; BARKER, William ; BURR, William ; POLK, William ; SMID, Miles: Recommendation for Key Management, Special Publication 800-57 Part 1, 03/2007. / National Institute of Standards and Technology (NIST). 2007. – Technischer Bericht

[BCJR74] BAHL, L. ; COCKE, J. ; JELINEK, F. ; RAVIV, J.: Optimal decoding of linear codes for minimizing symbol error rate. In: *IEEE Transactions on Information Theory* (1974)

[Bee92] BEERENDS, Jan A. John G.; Stemerdink S. John G.; Stemerdink: A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation. In: *Journal of the Audio Engineering Society* 40 (1992), Nr. 12, 963–978. `http://www.aes.org/e-lib/browse.cfm?elib=7019`

[BGML96] BENDER, W. ; GRUHL, D. ; MORIMOTO, N. ; LU, A.: Techniques for data hiding. In: *IBM Syst. J.* 35 (1996), Nr. 3/4, S. 313–336

[Bos98] BOSSERT, Martin: *Kanalcodierung*. 2. vollständig neubearbeitete und erweiterte Auflage. Stuttgart : Teubner Verlag, 1998. – ISBN 3519161435

[Bra87] BRANDENBURG, Karlheinz: Evaluation of Quality for Audio Encoding at Low Bit Rates. In: *Audio Engineering Society Convention 82*, 1987

[CB05] CHITTKA, Lars ; BROCKMANN, Axel: Perception Space The Final Frontier. In: *PLoS Biol* 3 (2005), April, Nr. 4, e137. http://dx.doi.org/10.1371/journal.pbio.0030137. – DOI 10.1371/journal.pbio.0030137

[CHW08] CHEN, Fan ; HE, HongJie ; WANG, HongXia: A Fragile Watermarking Scheme for Audio Detection and Recovery. In: *Proceedings of the 2008 Congress on Image and Signal Processing, Vol. 5 - Volume 05*. Washington, DC, USA : IEEE Computer Society, 2008. – ISBN 978–0–7695–3119–9, Seiten 135 - 138

[CLR+93] COLOMES, C. ; LEVER, M. ; RAULT, J. B. ; DEHERY, Y. F. ; FAUCON, G.: A Perceptual Model Applied to Audio Bit-Rate Reduction. In: *Audio Engineering Society Convention 95*, 1993

[CMB02] COX, Ingemar J. ; MILLER, Matthew L. ; BLOOM, Jeffrey A.: *Digital Watermarking*. Morgan Kaufmann Publishers, 2002

[CMM99] COX, I.J. ; MILLER, M.L. ; MCKELLIPS, A.L.: Watermarking as communications with side information. In: *Proceedings of the IEEE* 87 (1999), Juli, Nr. 7, S. 1127 –1141. http://dx.doi.org/10.1109/5.771068. – DOI 10.1109/5.771068. – ISSN 0018–9219

[Cve04] CVEJIC, Nedeljko: *Algorithms for Audio Watermarking and Steganography*. 1st Edition. Oulu University Press, 2004

[Dig] DIGITAL WATERMARKING ALLIANCE: *Digital Watermarking Applications*. http://www.digitalwatermarkingalliance.org/applications.asp, . – (09.07.2011)

[Dit00] DITTMANN, Jana: *Digitale Wasserzeichen: Grundlagen, Verfahren, Anwendungsgebiete*. Springer Verlag, 2000

[Dit01] DITTMANN, Jana: Content-fragil Watermarking for Image Authentication. In: *Proceedings of Electronic Imaging 2001 Security and Watermarking of Multimedia Contents III* Bd. Vol. 4314 SPIE, 2001 (Security and Watermarking of Multimedia Contents III), S. Seiten 175 – 184

[EBU] EUROPEAN BROADCASTING UNION: *Sound Quality Assessment Material recordings for subjective tests*. http://tech.ebu.ch/publications/sqamcd. – (10.09.2010)

[EBU08] EUROPEAN BROADCASTING UNION: Sound Quality Assessment Material recordings for subjective tests (Users handbook for the EBU SQAM CD) / EBU TECHNICAL. Geneva, September 2008 (3253). – Technischer Bericht

[Eck10] ECKE, Dr. O. ; TNS Infratest MediaResearch: *Relevanz der Medien für die Meinungsbildung: Empirische Grundlagen zur Ermittlung der Wertigkeit der Mediengattungen bei der Meinungsbildung*. Berlin : `www.blm.de/files/pdf1/Praesentation_Studie_Meinungsmacht_o1.pdf`, 17. März 2010. – (16.07.2011)

[Eck11] ECKE, Dr. O. ; TNS Infratest MediaResearch: *Relevanz der Medien für die Meinungsbildung: Empirische Grundlagen zur Ermittlung der Wertigkeit der Mediengattungen bei der Meinungsbildung*. `www.blm.de/apps/documentbase/data/pdf1/Medienkonzentration.pdf`, 15. Juli 2011

[ECR10] SMART, Nigel (Hrsg.): *ECRYPT II Yearly Reporton Algorithms and Keysizes (2009-2010)*. 30. März 2010

[EH93] EPPINGER, Bernd ; HERTER, Eberhard: *Sprachverarbeitung*. Carl Hanser Verlag Mnchen Wien, 1993 (Reihe Informationstechnik, Nachrichtentechnik). – ISBN 3-446-16076-0

[Eic08] EICKHOFF, Christian A.: *Integritätsschutz von Audiodaten: Entwicklung und Evaluierung von modellbasierten Verfahren zur Beschreibung von Audiomerkmalen*. VDM Verlag Dr. Müller, 2008. – ISBN: 978-3-639-07928-9

[Esk97] ESKA, Georg: *Schall & Klang: Wie und was wir hören*. Birkhäuser Verlag, 1997

[faac] *Freeware Advanced Audio Coder*. `http://www.audiocoding.com/faac.html`,

[FKK04] FEI, Chuhong ; KUNDUR, Deepa ; KWONG, Raymond H.: Analysis and design of authentication watermarking. In: *Security, Steganography, and Watermarking of Multimedia Contents*, 2004, S. Seiten 760 – 771

[FNI10] *Mécanismes cryptographiques - Règles et recommandations, Rev. 1.20*. : *Mécanismes cryptographiques - Règles et recommandations, Rev. 1.20*, Oktober 2010

[fsp] *The Freesound Project*. `http://www.freesound.org/`. – (10.09.2010)

[Got04] GOTO, Masataka: Development of the RWC Music Database. In: *Proceedings of the 18th International Congress on Acoustics (ICA 2004)*, 2004. – Invited Paper

[HK02] HAITSMA, Jaap ; KALKER, Ton: Jaap Haitsma and Ton Kalker. In: *Proceedings of 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, 2002

[ia] *Internet Archive: Audio Archive*. http://www.archive.org/details/audio. – (10.09.2010)

[Int98] INTERNATIONAL STANDARD, ISO/IEC/JTC1/SC29 WG11: *ISO/IEC 13818-3, Information technology – generic coding of moving pictures and associated audio information – Part 3: Audio*. Geneva, Switzerland : International Organization for Standardization, 1998. – 182 S.

[Int03] INTERNATIONAL TELECOMMUNICATIONS UNION (Hrsg.): *ITU-R Recommendation BS.1284-1, General methods for the subjective assessment of sound quality*. Geneva: International Telecommunications Union, 1997-2003

[Kab] KABAL, Peter ; Dept. Electrical & Computer Engineering, McGill University: *Perceptual Evaluation of Audio Quality (PEAQ)*. http://www-mmsp.ece.mcgill.ca/Documents/Downloads/PQevalAudio/. – (20.09.2010)

[Kat00] KATZENBEISSER, Stefan: *Information Hiding, techniques for steganography and digital watermarking*. Artech House, 2000

[lam06] *The LAME Project*. lame.sourceforge.net, September 2006

[LSS+07] LAZARUS, Hans ; SUST, Charlotte A. ; STECKEL, Rita ; KULKA, Marko ; KURTZ, Patrik: *Akustische Grundlagen sprachlicher Kommunikation*. Springer, Berlin, 2007

[Mey02] MEYER, Martin: *Kommunikationstechnik. Konzepte der modernen Nachrichtenübertragung*. Vieweg Verlag, 2002 (2. Auflage)

[OQV05] OLSEN, Adam ; QUINONES, Jose ; VAISVIL, Pat: Using current digital watermarking techniques to authenticate audio Recordings. In: *AES 26$^{th}$ International Conference, Denver, USA* (2005)

[peaq] *www.peaq.org*. http://www.peaq.org/. – (20.09.2010)

[PHJ04] PAN, J.-S. ; HUANG, H.-C. ; JAIN, L. C.: *Intelligent Watermarking Techniques*. World Scientific Publishing Co. Pte. Ltd., 2004. – ISBN 981-238-955-5

[Plo64] PLOMP, R.: The Ear as a Frequency Analyzer. In: *The Journal of the Acoustical Society of America* 36 (1964), Nr. 9, 1628-1636. http://dx.doi.org/10.1121/1.1919256. – DOI 10.1121/1.1919256

[PMMS92] PAILLARD, B. ; MABILLEAU, P. ; MORISSETTE, S. ; SOUMAGNE, Joël: PERCEVAL: Perceptual Evaluation of the Quality of Audio Signals. In: *Journal of the Audio Engineering Society* 40 (1992), Nr. 1/2, 21–31. http://www.aes.org/e-lib/browse.cfm?elib=7061

[PS00] PAINTER, T. ; SPANIAS, A.: Perceptual Coding of Digital Audio. In: *Proc. IEEE* 88 (2000), April, Nr. 4

[RD02] REY, Christian ; DUGELAY, Jean-Luc: A Survey of Watermarking Algorithms for Image Authentication. In: *EURASIP Journal on Applied Signal Processing* 2002 (2002), Nr. 6, S. pp. 613–621

[RIA09] Recording Industry Association of America: *Watermark Payload Specification*. 1025 F ST N.W., 10th Floor, Washington, D.C. 20004 : specifications@riaa.com, 2009

[Roe00] ROEDERER, Juan G.: *Physikalische und psychoakustische Grundlagen der Musik*. Springer, 2000

[Rop06] ROPPEL, Carsten: *Grundlagen der digitalen Kommunikationstechnik: Übertragungstechnik - Signalverarbeitung - Netze*. München Wien : Fachbuchverlag Leipzig im Carl-Hanser-Verlag, 2006

[RSA78] RIVEST, R.L. ; SHAMIR, A. ; ADLEMAN, L.: A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. In: *Communications of the ACM* 21 (1978), S. 120–126

[RWC] GOTO, Masataka ; National Institute of Advanced Industrial Science and Technology (AIST): *RWC (Real World Computing) Music Database*. http://staff.aist.go.jp/m.goto/RWC-MDB/. – (10.09.2010)

[Sch09] SCHLAUWEG, Mathias: *Digitale Wasserzeichen zur Überprüfung der Echtheit von Bildern*. Mensch und Buch Verlag, 2009. – ISBN 978-3-86664-733-6

[Sch10] SCHNELLER, Johannes ; Allensbacher Computer- und Technik-Analyse: *Zukunftstrends im Internet*. http://www.acta-online.de/praesentationen/acta_2010/acta_2010_Internettrends.pdf, 2010. – (23.06.2011)

[SD03] STEINEBACH, Martin ; DITTMANN, Jana: Watermarking-Based Digital Audio Data Authentication. In: *EURASIP Journal on Applied Signal Processing 2003* (2003), S. 1001–1015

[SLT05] SCHMIDT, Robert F. ; LANG, Florian ; THEW, Gerhard: *Physiologie des Menschen: mit Pathophysiologie*. Springer Verlag, 2005 (29. Auflage)

[Spo97] SPORER, Thomas: Objective Audio Signal Evaluation-Applied Psychoacoustics for Modeling the Perceived Quality of Digital Audio. In: *Audio Engineering Society Convention 103*, 1997

[Ste04] STEINEBACH, Martin: *Digitale Wasserzeichen für Audiodaten*. 2. Shaker Verlag, 2004

[SVN37] STEVENS, S. S. ; VOLKMANN, J. ; NEWMAN, E. B.: A Scale for the Measurement of the Psychological Magnitude Pitch. In: *The Journal of the Acoustical Society of America* 8 (1937), Nr. 3, 185-190. http://dx.doi.org/10.1121/1.1915893. – DOI 10.1121/1.1915893

[TBS+98] THIEDE, Thilo ; BITTO, Roland ; SCHMIDMER, Christian ; SPORER, Thomas ; BRANDENBURG, Karlheinz ; TREURNIET, William C. ; BEERENDS, John G. ; COLOMES, Catherine ; KEYHL, Michael ; STOLL, Gerhard ; FEITEN, Bernhard: PEAQ (Perceptual Evaluation of Audio Quality) - der künftige ITU-Standard zur objektiven Messung der wahrgenommenen Audioqualität. In: *Bericht der 20. Tonmeistertagung, International Convention on Sound Design, 20. - 23. Nov. 1998, Karlsruhe*. München : Verlag K. G. Saur, 1998, Kapitel PEAQ (Perceptual Evaluation of Audio Quality) - der künftige ITU-Standard zur objektiven Messung der wahrgenommenen Audioqualität, S. 724 – 766

[Ter79] TERHARDT, Ernst: Calculating virtual pitch. In: *Hearing Research* 1 (1979), Nr. 2, S. 155 – 182. http://dx.doi.org/10.1016/0378-5955(79)90025-X. – DOI 10.1016/0378–5955(79)90025–X. – ISSN 0378–5955

[Ter98] TERHARDT, Ernst: *Akustische Kommunikation: Grundlagen mit Hörbeispielen*. Springer, Berlin, 1998

[Thi96] THIEDE, E. Thilo; K. Thilo; Kabot: A New Perceptual Quality Measure for Bit-Rate Reduced Audio. In: *Audio Engineering Society Convention 100*, 1996

[TK01] TON, Jaap H. ; KALKER, Ton: Robust Audio Hashing for Content Identification. In: *Content-Based Multimedia Indexing (CBMI) 2001*. Brescia Italy, 2001

[USC] UNIVERSITY OF SOUTHERN CALIFORNIA: *The USC-SIPI Image Database*. http://sipi.usc.edu/database/. – (10.09.2010)

[VHH98] VARY, Peter ; HEUTE, Ulrich ; HESS, Wolfgang: *Digitale Sprachsignalverarbeitung*. Stuttgart : B. G: Teubner Verlag, 1998

[Vit67] VITERBI, Andrew J.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. In: *Information Theory, IEEE Transactions on* 13 (1967), April, Nr. 2, S. 260 –269. http://dx.doi.org/10.1109/TIT.1967.1054010. – DOI 10.1109/TIT.1967.1054010. – ISSN 0018–9448

[vor] *vorbis site*. http://www.vorbis.com/,

[WF10] WANG, HongXia ; FAN, MingQuan: Centroid-based semi-fragile audio watermarking in hybrid domain. In: *SCIENCE CHINA Information Sciences* 53 (2010), 619-633. http://dx.doi.org/10.1007/s11432-010-0058-0. – ISSN 1674–733X. – 10.1007/s11432-010-0058-0

[WLC07] WANG, Ching-Te ; LIAO, Chiu-Hsiung ; CHEN, Tung shou: Audio-Signal Authenticating System Based on Asymmetric Signature Schemes. In: *Internationa2007 l Conference on Multimedia and Ubiquitous Engineering (MUE'07)*, 2007, S. 656–661

[YK03] YEO, In-Kwon ; KIM, Hyoung J.: Modified Patchwork Algorithm: A Novel Audio Watermarking Scheme. In: *IEEE Transaction on Speech and Audio Processing* Bd. 11, 2003

[ZF99] ZWICKER, E. ; FASTL, H. ; SCHROEDER, M. R. (Hrsg.): *Psychoacoustics: Facts and Models*. 2nd Edition. Springer, 1999 (Information Sciences)

[ZS03] ZHU, B. B. ; SWANSON, M. D.: Multimedia Authentication and Watermarking. In: FENG, D. (Hrsg.) ; SIU, W. C. (Hrsg.) ; ZHANG, H. (Hrsg.): *Multimedia Information Retrieval and Management*. Springer-Verlag, Berlin, Heidelberg, New York, 2003, Kapitel 7, S. 148–177

[ZS08a] ZMUDZINSKI, Sascha ; STEINEBACH, Martin: Psycho-acoustic model-based message authentication coding for audio data. In: *Proceedings of the 10th ACM workshop on Multimedia and security*. New York, NY, USA : ACM, 2008 (MM&Sec '08). – ISBN 978–1–60558–058–6, 75–84

[ZS08b] ZMUDZINSKI, Sascha ; STEINEBACH, Martin: Robust Audio Hashing for Audio Authentication Watermarking. In: *Proceedings of SPIE Conference on Security, Forensics, Steganography, and Watermarking of Multimedia Contents X* Bd. 6819, 2008

[ZS09a] ZMUDZINSKI, Sascha ; STEINEBACH, Martin: Perception-Based Audio Authentication Watermarking in the Time-Frequency Domain. In: KATZENBEISSER, Stefan (Hrsg.) ; SADEGHI, Ahmad-Reza (Hrsg.): *Pre-Proceedings of 11th Information Hiding*. Darmstadt, Germany, 7-10 June 2009, S. 293–308

[ZS09b] ZMUDZINSKI, Sascha ; STEINEBACH, Martin: Perception-based Authentication Watermarking for Digital Audio Data. In: *IS&T SPIE Electronic Imaging 2009 Conference - Media Forensics and Security XI, San Jose, USA*, 2009

[Zwi82] ZWICKER, E.: *Psychoakustik*. Berlin : Springer Verlag, 1982

# Eigene Veröffentlichungen


[1] GULBIS, M. ; MÜLLER, E. ; STEINEBACH, M.: Audio Integrity Protection and Falsification Estimation by Embedding Multiple Watermarks. In: *Proceedings of the 2006 IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP-2006)*. Pasadena, CA, USA : IEEE Computer Society, Dezember 2006, 469 - 472

[2] GULBIS, M. ; STEINEBACH, M. ; MÜLLER, E.: Combining Multilevel Manipulation Estimation with Content-Based Authentication Watermarking. In: *Proceedings of the 2007 IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP-2007)* Bd. I. Kaohsiung City, Taiwan, November 2007, S. 461 – 464. – Best Paper Award

[3] GULBIS, M. ; MÜLLER, E.: Inhaltsbasierender Integritätsschutz von Audiodaten mittels Digitaler Wasserzeichen. In: *Tagungsband des 12. Symposium Maritime Elektrotechnik, Elektronik und Informationstechnik*. Rostock, Germany : Universität Rostock, Fakultät für Informatik und Elektrotechnik, Oktober 2007, S. 197 – 202

[4] GULBIS, M. ; STEINEBACH, M. ; MÜLLER, E.: Content-based Authentication Watermarking with Improved Audio Content Feature Extraction. In: *Proceedings of the 2008 Fourth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP-2008)*. Harbin, China, August 2008, S. 620 – 623. – ISBN 978-0-7695-3278-3

[5] GULBIS, M. ; MÜLLER, E. ; STEINEBACH, M.: Synchronization Approach for Audio Authentication Watermarking. In: *The 10th IASTED International Conference on Signal and Image Processing (SIP 2008)*. Kailua-Kona, Hawaii, USA, August 2008

[6] GULBIS, M. ; MÜLLER, E. ; STEINEBACH, M.: Content-based Audio Authentication Watermarking. In: *International Journal of Innovative Computing, Information and Control (IJICIC)* 5 (2009), Juli, Nr. 7, S. 1883 – 1892

[7] GULBIS, M. ; MÜLLER, E.: Content-based audio authentication using a hierarchical patchwork watermark embedding. In: *Proceedings of SPIE Conference on Optics, Photonics and Digital Technologies for Multimedia Applications* Bd. 7723. Brussels, Belgium, 2010, S. 77230N–1 to N–12

Anhang A

# Anhang

## A.1. Testdaten

Alle in dieser Arbeit aufgeführten Ergebnisse wurden auf Basis der folgend aufgeführten Testdaten ermittelt.

Der Testdatensatz besteht aus 130 Audiodateien (48 kHz, 16 bit, mono) mit einer Abspieldauer von je 61,44 Sekunden. Die Abspieldauer der einzelnen Audiodateien orientiert sich an etwa einer Minute, wobei die Anzahl der Samples ein Vielfaches der untersuchten Segmentlängen $l$ darstellt. Die Audiodaten werden hierdurch für alle während der Leistungsanalyse verwendeten Segmentlängen $l$ vollständig verwendet. Die Audiodaten stammen aus 13 unterschiedlichen Quellen, bestehend aus Hörspielen, Buch- und Lyriklesungen, „Comedy“ und Rundfunkbeiträgen.

- **Datei 1-10:**
  Das Buch von Eden. CD1, Bastei Lübbe; Auflage: 1 (2004), ISBN: 978-3-7857-1430-0
- **Datei 11-20:**
  Der Gewissenlose Mörder. CD1, Der Audio Verlag, Auflage: 1., September 2006, ISBN: 978-3-8981-3592-4

- **Datei 21-30:**
  Der Katalane. CD1, Random House Audio, Auflage: gekürzte Lesung,(11. August 2008, ISBN: 978-3-8660-4800-3

- **Datei 31-40:**
  Der Medicus von Saragossa. CD1, BMG Wort, 1. April 2000, ISBN: 978-3-8983-0072-8

- **Datei 41-50:**
  Der Steppenwolf. CD1, Universal Music, Auflage: 1., 9. August 2005, ISBN: 978-3-8291-1558-2

- **Datei 51-60:**
  Die Rache der Kreuzfahrer. CD1, Bastei Lübbe, Auflage: 1, 2005, ISBN: 978-3-7857-1472-0

- **Datei 61-70:**
  Deutsch-Frau. CD1, Langenscheidt, September 2004, ISBN: 978-3-4687-3116-7

- **Datei 71-80:**
  Harry Potter und der Orden des Phönix. CD1, Dhv der Hörverlag; Auflage: 1., 19. Februar 2004, ISBN: 978-3-8994-0172-1

- **Datei 81-90:**
  Maria Wimmer spricht Goethe, Hebbel, Racine, Schiller und Shakespeare. CD1, Universal Music, Auflage: 1, 2004, ISBN: 978-3-8291-1468-4

- **Datei 91-100:**
  Urknaller - Physik ist Sexy. CD1, HERBERT Management, Februar 2005, ISBN: 978-3-8218-6303-0

- **Datei 101-110:**
  Sherlock Holmes und Dr. Watson - Die größten Fälle. CD1, Der Audio Verlag, Auflage: 1., 2004, ISBN: 978-3-8981-3309-8

- **Datei 111-120:**
  Stenkelfeld - 2000 rüüührend!. CD1, , September 1999, ASIN: B00004SP6C

- **Datei 121-130:**
  Rundfunkbeiträge von ELF-TV[1]

[1]Das Erste Laager Fernsehen ist ein Medienprojekt der ev.-luth. Kirchgemeinde Laage. `http://www.elf-tv.eu/`

# A.2. Leistungsanalyse des Inhaltsmerkmals

## A.2.1. Eignung der Frequenzgruppen

Die Abbildungen A.1 bis A.5 zeigen die Simulationsergebnisse hinsichtlich der Robustheit des Inhaltsmerkmals in Abhängigkeit der Frequenzgruppen $F1$ bis $F25$ gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen. Für die Analyse wurden Segmentlängen von $l_M = \{512, 1\,024, 2\,048, 4\,096, 8\,192, 16\,384, 32\,768\}$ für die Extraktion der Inhaltsmerkmale verwendet. Die Abbildungen stellen in Bezug auf die Frequenzgruppe $F\,j$ den Mittelwert sowie den Wertbereich der Bitfehlerrate der Merkmalsvektoren $m_i$ bei Nutzung aller Segmentlänge $l_M$ für die Simulationen dar. Ein Merkmalsvektor $m_i$ besteht in diesem Fall nur aus dem Wert $m_{i,j}$ entsprechend der verwendeten Frequenzgruppe $F\,j$.

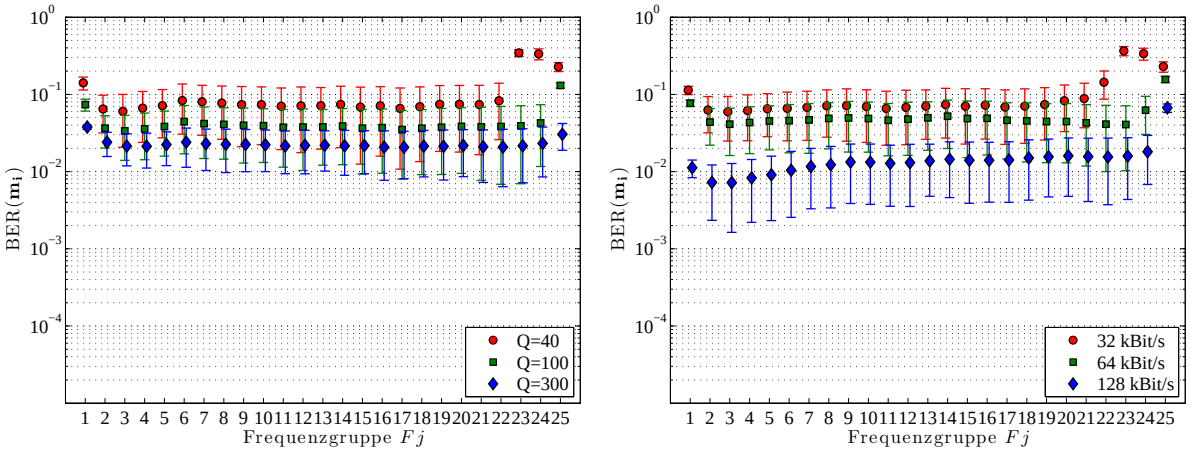


**Abbildung A.1.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Frequenzgruppe $F\,j$ gegenüber AAC-Kompression (links) und MP3-Kompression (rechts).

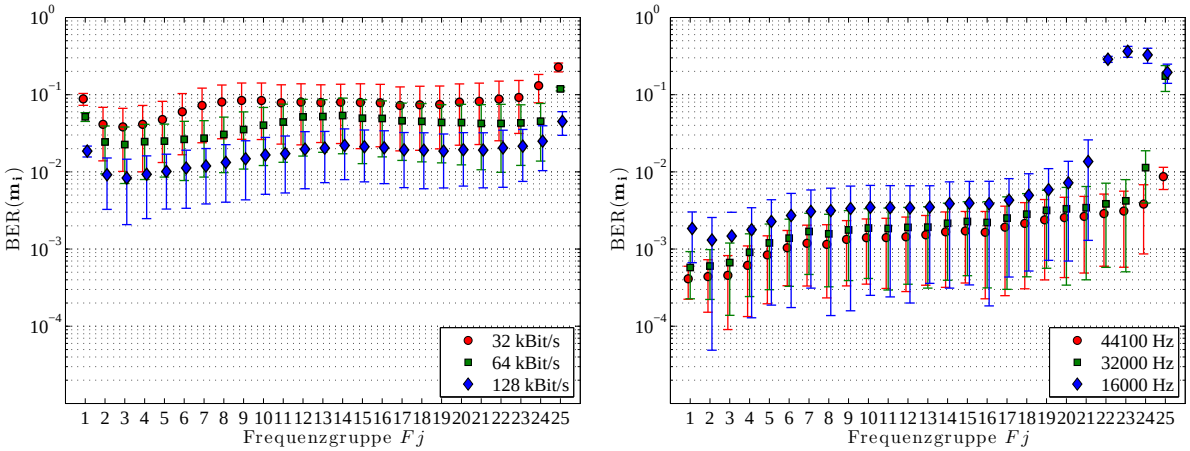


**Abbildung A.2.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Frequenzgruppe $F\,j$ gegenüber Vorbis-Kompression (links) und Unterabtastung (rechts).

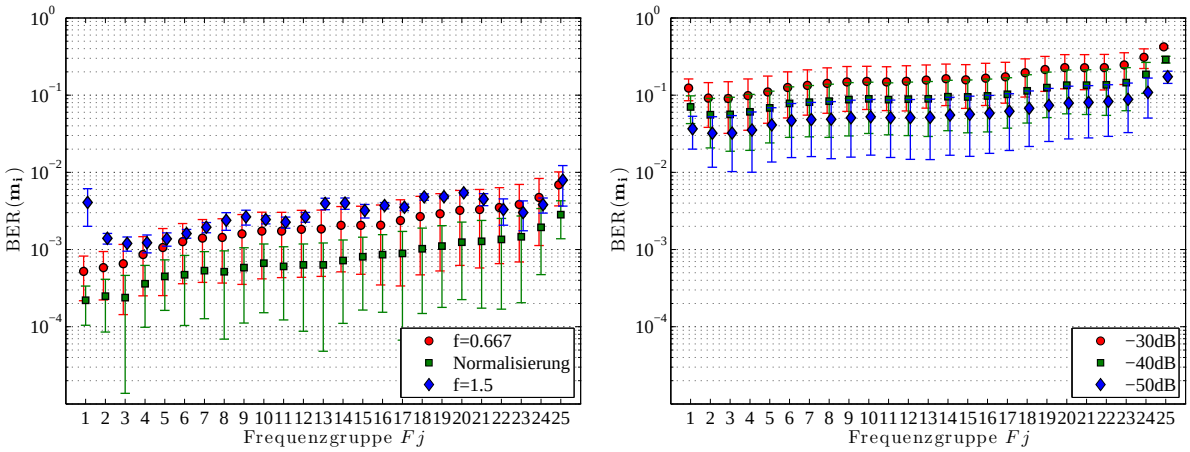


**Abbildung A.3.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Frequenzgruppe $Fj$ gegenüber Laustärkeveränderung (links) und Weißem Rauschen (rechts).

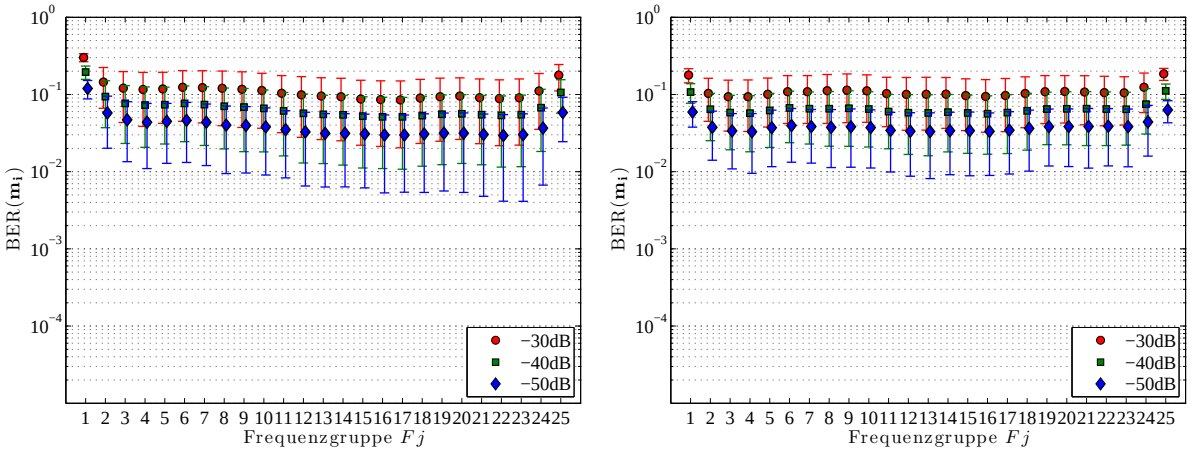


**Abbildung A.4.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Frequenzgruppe $Fj$ gegenüber $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

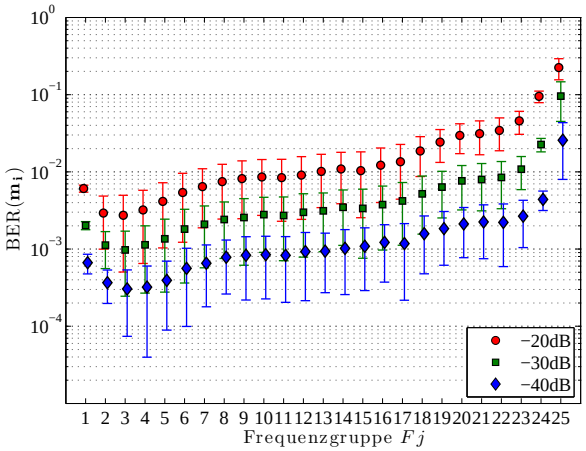


**Abbildung A.5.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Frequenzgruppe $Fj$ gegenüber dynamischen Rauschen (links).

## A.2.2. Robustheit des Inhaltsmerkmals

Die Abbildungen A.6 bis A.10 zeigen die Simulationsergebnisse hinsichtlich der Robustheit des Inhaltsmerkmals gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurden Segmentlängen von $l_M = \{512, 1\,024, 2\,048, 4\,096, 8\,192, 16\,384, 32\,768\}$ für die Extraktion der Inhaltsmerkmale verwendet. Der Extraktionbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ gebildet. Die Abbildungen stellen in Bezug auf die Segmentlänge $l_M$ den Mittelwert sowie den Wertbereich der Bitfehlerrate der Merkmalsvektoren $\mathbf{m_i}$ dar.

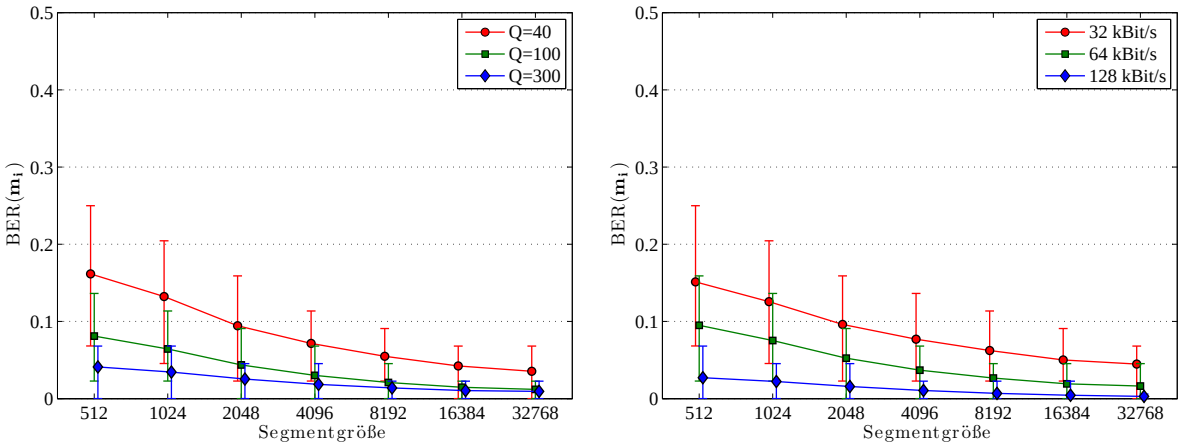


**Abbildung A.6.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Segmentlänge gegenüber AAC-Kompression (links) und MP3-Kompression (rechts).

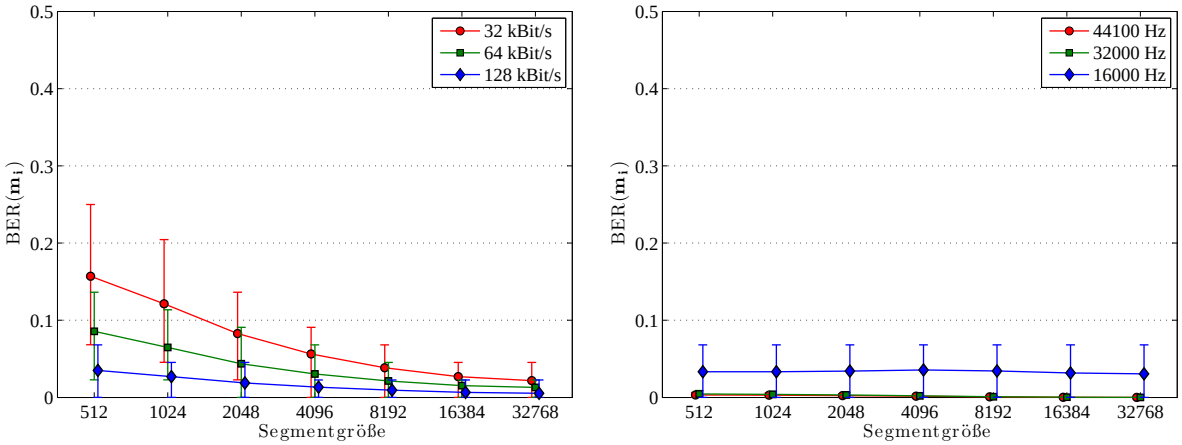


**Abbildung A.7.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Segmentlänge gegenüber Vorbis-Kompression (links) und Unterabtastung (rechts).

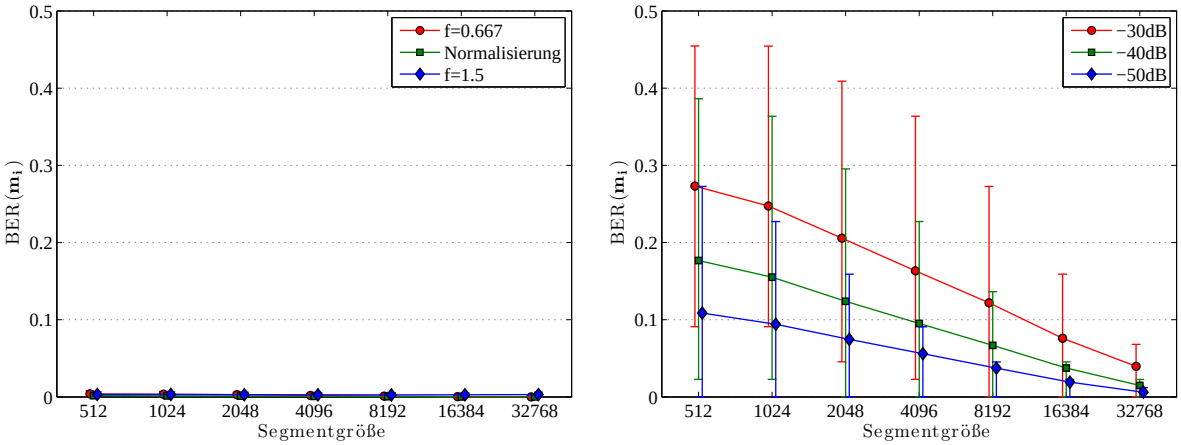


**Abbildung A.8.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Segmentlänge gegenüber Lautstärkeveränderung (links) und Weißem Rauschen (rechts).

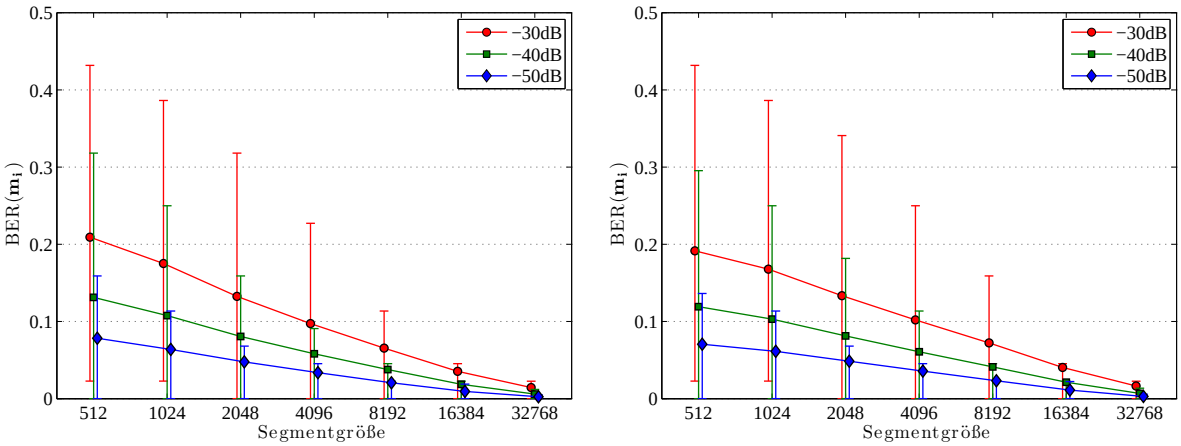


**Abbildung A.9.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Segmentlänge gegenüber $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

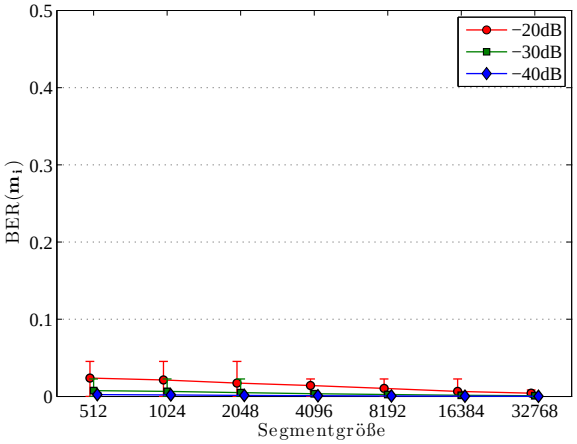


**Abbildung A.10.:** Robustheit des Inhaltsmerkmals $\mathbf{m_i}$ in Abhängigkeit der Segmentlänge gegenüber dynamischen Rauschen (links).

## A.2.3. Manipulationssicherheit des Inhaltsmerkmals

Die Abbildungen A.11 bis A.13 zeigen die Simulationsergebnisse hinsichtlich der Manipulationssicherheit des Inhaltsmerkmals gegenüber Veränderungen des Dateninhalts. Für die Analyse wurden Segmentlängen von $l_M = \{512, 1\,024, 2\,048, 4\,096, 8\,192, 16\,384, 32\,768\}$ für die Extraktion der Inhaltsmerkmale verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ gebildet. Veränderungen das Dateninhaltes werden durch Ersetzen von Audiopassagen mit anderen zufällig gewählten Audiopassagen aus der gleichen Testdatei und durch Ersetzen mit Stille simuliert. Die Länge der ersetzten Audiopassage wird durch die Störungslänge angegeben. Die Abbildungen stellen den Mittelwert sowie den Wertbereich der Bitfehlerrate der Merkmalsvektoren $\mathbf{m_i}$ dar.

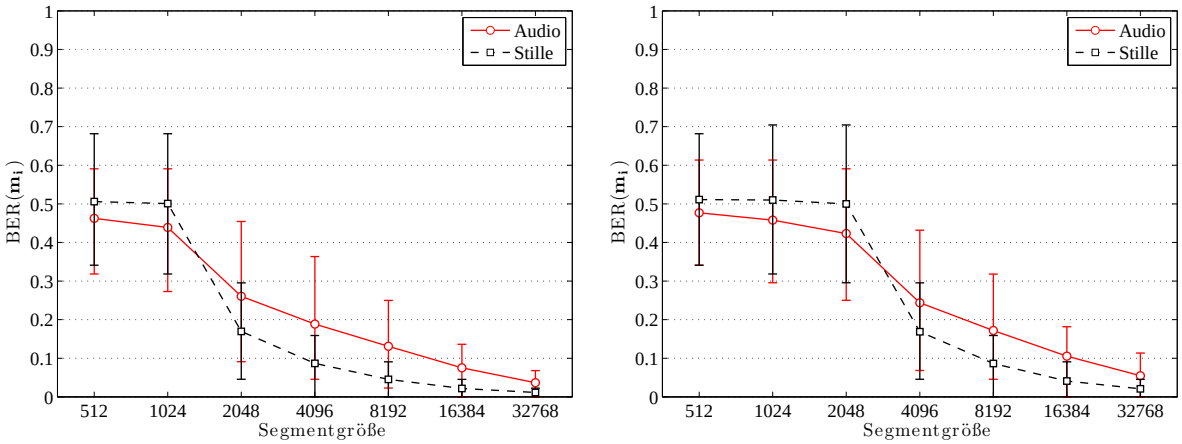


**Abbildung A.11.:** Bitfehlerrate des Inhaltmerkmals bei einer Störungslänge von 1 024 Samples (links) und 2 048 Samples (rechts).

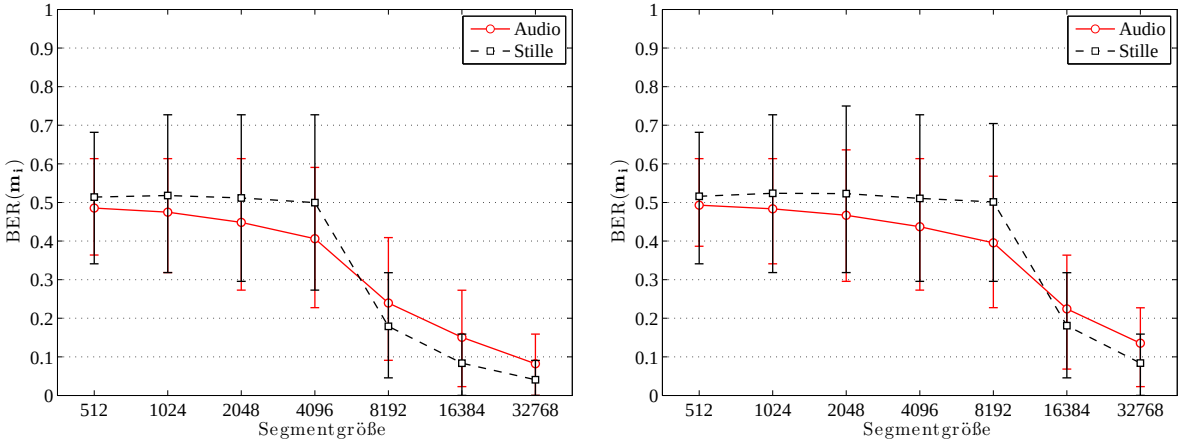


**Abbildung A.12.:** Bitfehlerrate des Inhaltmerkmals bei einer Störungslänge von 4 096 Samples (links) und 8 192 Samples (rechts).

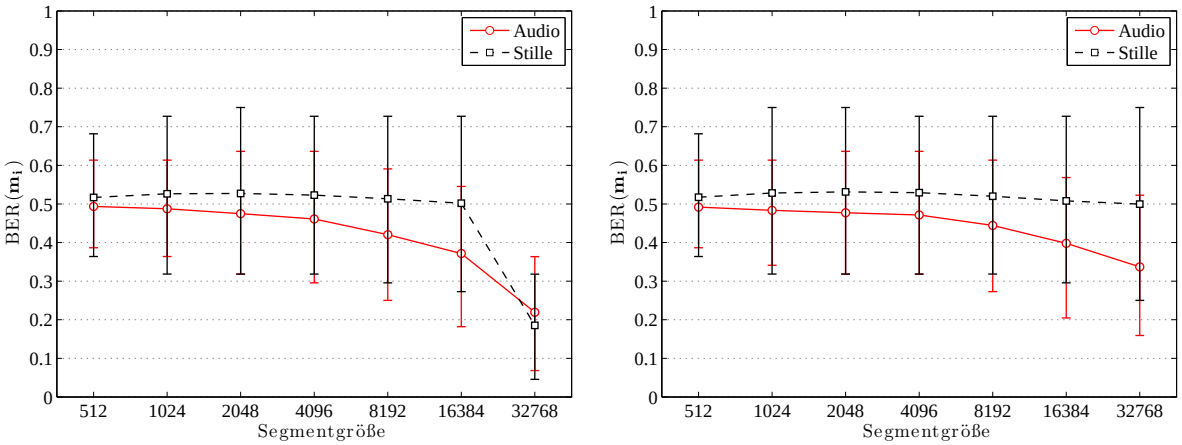


**Abbildung A.13.:** Bitfehlerrate des Inhaltmerkmals bei einer Störungslänge von 16 384 Samples (links) und 32 768 Samples (rechts).

# A.3. Leistungsanalyse der Wasserzeichentechnik

## A.3.1. Eignung der Frequenzgruppen

Die Abbildungen A.14 bis A.18 zeigen die Simulationsergebnisse hinsichtlich der Robustheit des Wasserzeicheninfomation in Abhängigkeit der Frequenzgruppen $F1$ bis $F25$ gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen. Für die Analyse wurden Segmentlängen von $l_W = \{512, 1\,024, 2\,048, 4\,096, 8\,192, 16\,384\}$ sowie eine Einbettungsstärke von $f = 0.05$ verwendet. Der Einbettungsbereich der Wasserzeicheninformation ist jeweils auf die Koeffizienten der Frequenzgruppe $F\,j$ beschränkt. Die Abbildungen stellen in Bezug auf die Frequenzgruppe $F\,j$ den Mittelwert sowie den Wertbereich der Bitfehlerrate der Wasserzeicheninfomation $\mathbf{w}$ bei Nutzung aller Segmentlänge $l_W$ für die Simulationen dar.

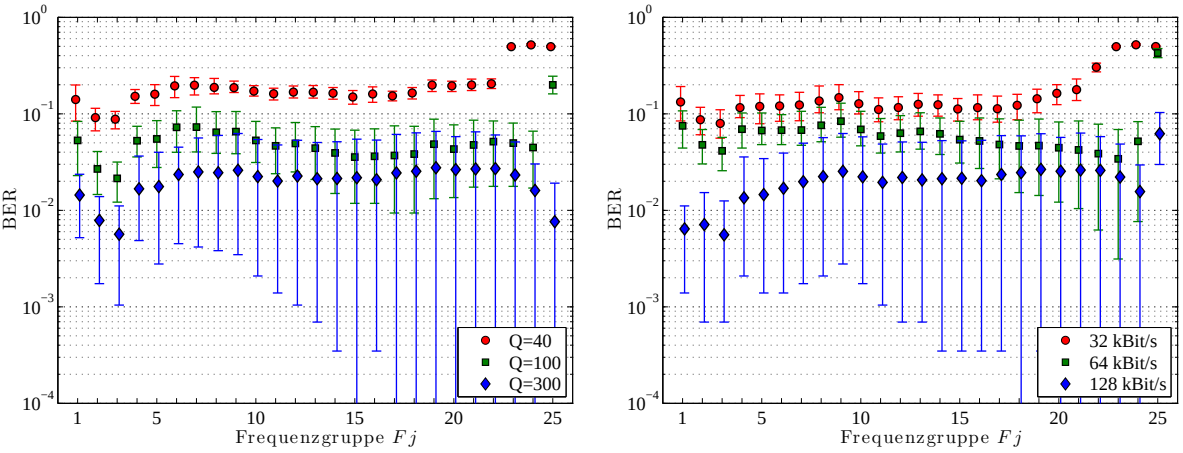


**Abbildung A.14.:** Robustheit der Wasserzeicheninformation bei einer Einbettungsstärke $f = 0,05$ gegenüber AAC-Kompression (links) und MP3-Kompression (rechts) in Abhängigkeit der Segmentlänge $l_W$.

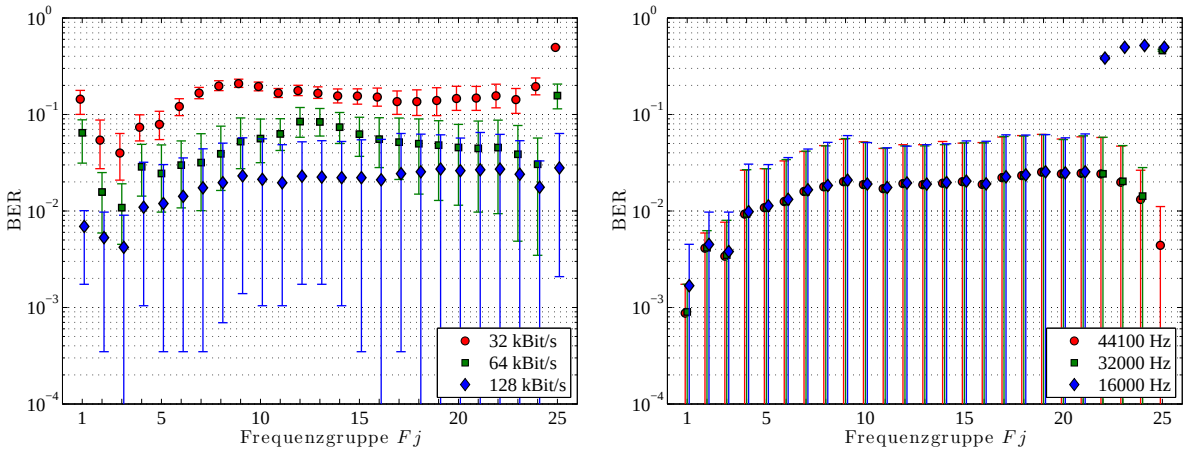


**Abbildung A.15.:** Robustheit der Wasserzeicheninformation bei einer Einbettungsstärke $k = 0,05$ gegenüber Vorbis-Kompression (links) und Unterabtastung (rechts) in Abhängigkeit der Segmentlänge $l_W$.

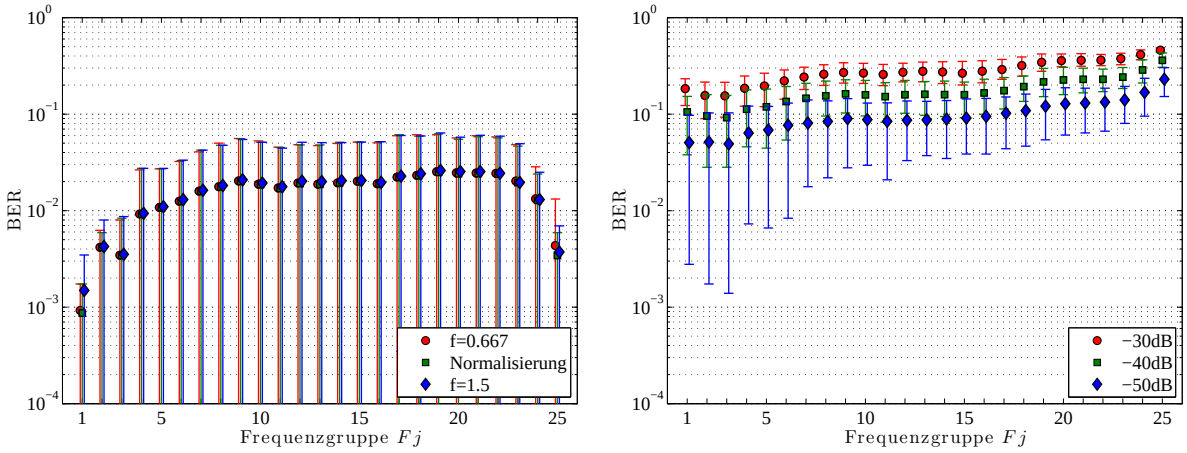


**Abbildung A.16.:** Robustheit der Wasserzeicheninformation bei einer Einbettungsstärke $f = 0,05$ gegenüber Laustärkeveränderung (links) und Weißem Rauschen (rechts) in Abhängigkeit der Segmentlänge $l_W$.

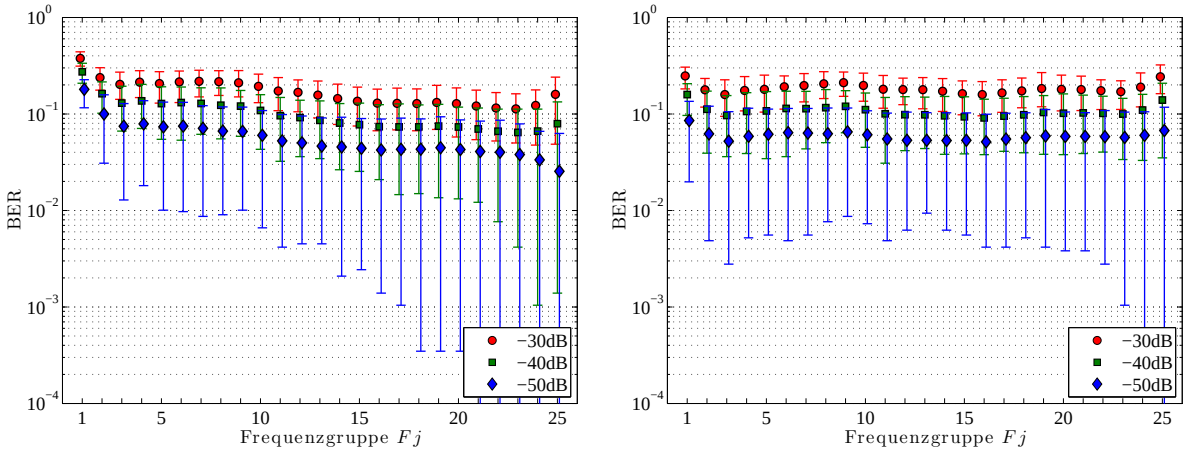


**Abbildung A.17.:** Robustheit der Wasserzeicheninformation bei einer Einbettungsstärke $f = 0,05$ gegenüber $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts) in Abhängigkeit der Segmentlänge $l_W$.

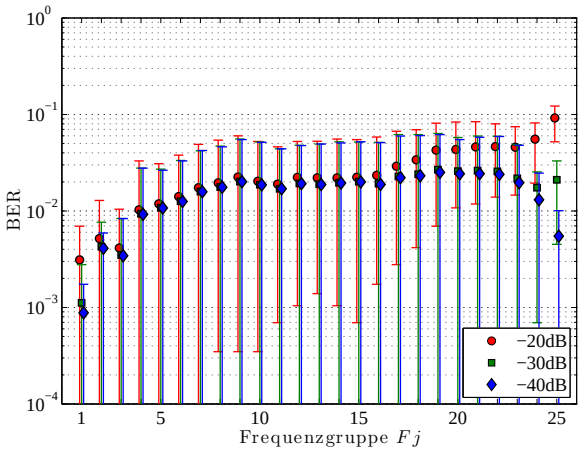


**Abbildung A.18.:** Robustheit der Wasserzeicheninformation bei einer Einbettungsstärke $f = 0,05$ gegenüber dynamischen Rauschen (links) in Abhängigkeit der Segmentlänge $l_W$.

## A.3.2. Robustheit der Wasserzeicheninfomation

Die Abbildungen A.19 bis A.23 zeigen die Simulationsergebnisse hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen. Für die Analyse wurden Segmentlängen von $l_W = \{512,\ 768,\ 1\,024,\ 1\,536,\ 2\,048,\ 3\,072,\ 4\,096,\ 6\,144,\ 8\,192,\ 12\,288,\ 16\,384\}$ verwendet sowie eine Wasserzeichentransparenz mit einem ODG von -1 eingestellt. Der Einbettungsbereich der Wasserzeicheninformation wird durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Abbildungen stellen in Bezug auf die Segmentlänge $l_W$ den Mittelwert sowie den Wertbereich der Bitfehlerrate der Wasserzeicheninformation $\mathbf{w}$ dar.

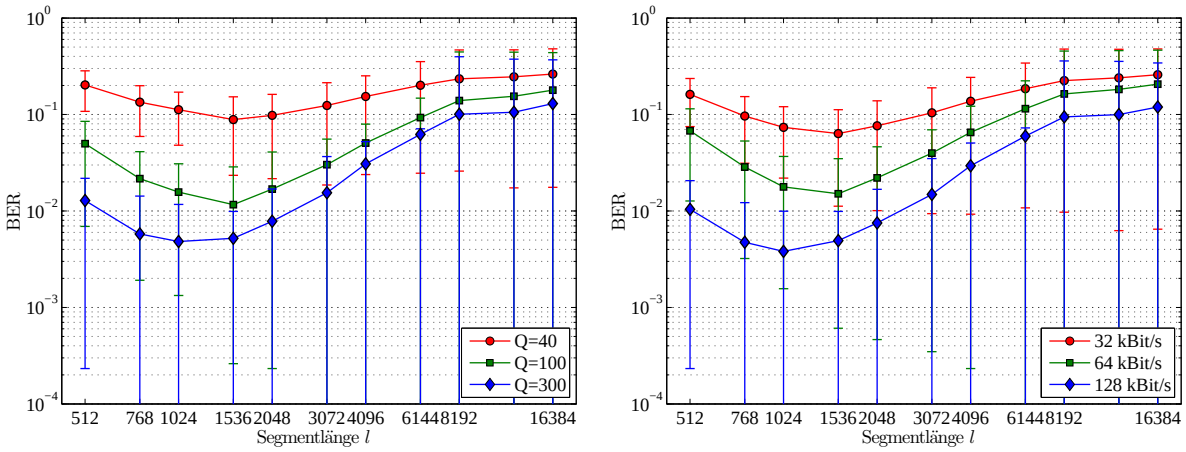


**Abbildung A.19.:** Robustheit der Wasserzeicheninformation bei einem ODG von -1 gegenüber AAC-Kompression (links) und MP3-Kompression (rechts) in Abhängigkeit der Segmentlänge $l_W$.

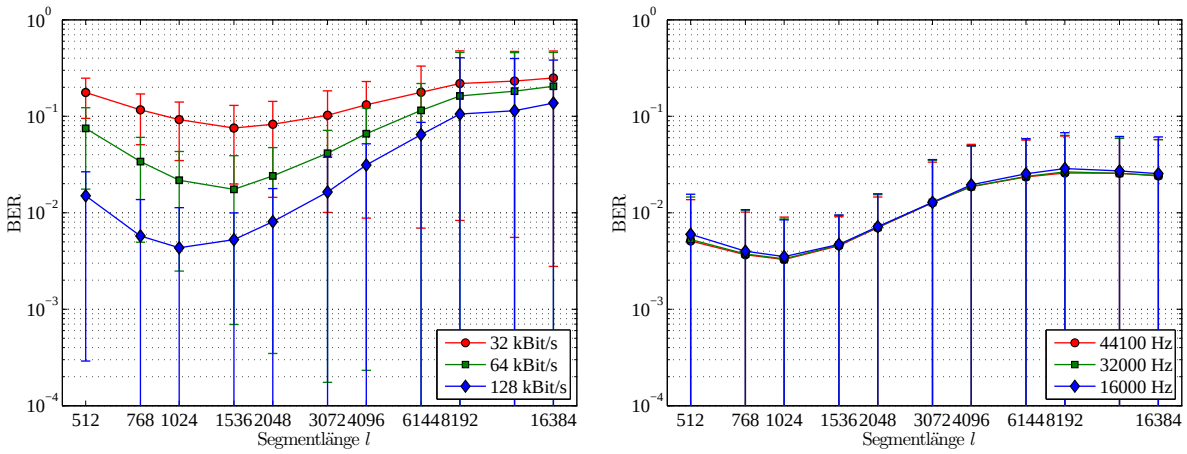


**Abbildung A.20.:** Robustheit der Wasserzeicheninformation bei einem ODG von -1 gegenüber Vorbis-Kompression (links) und Unterabtastung (rechts) in Abhängigkeit der Segmentlänge.

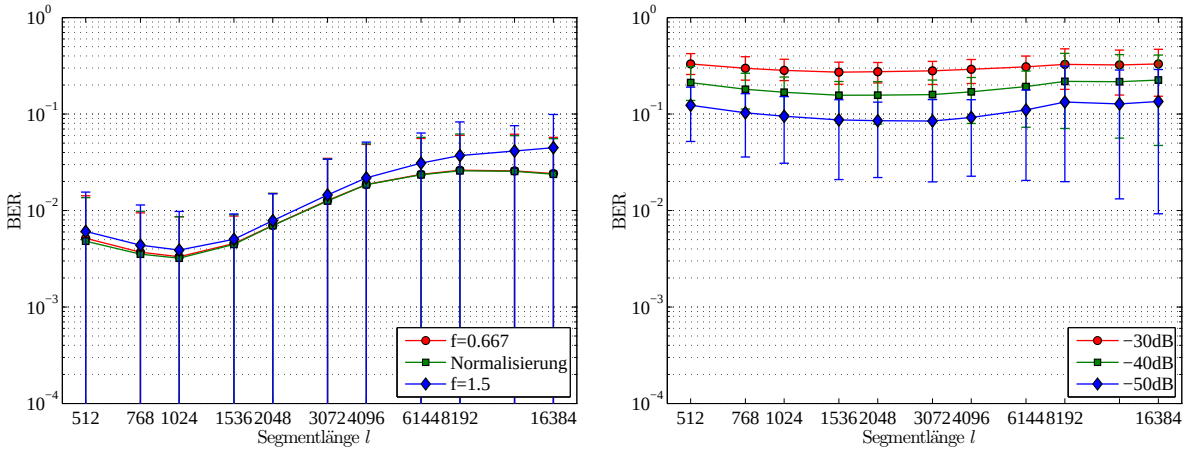


**Abbildung A.21.:** Robustheit der Wasserzeicheninformation bei einem ODG von -1 gegenüber Lautstärkeveränderung (links) und Weißem Rauschen (rechts) in Abhängigkeit der Segmentlänge.

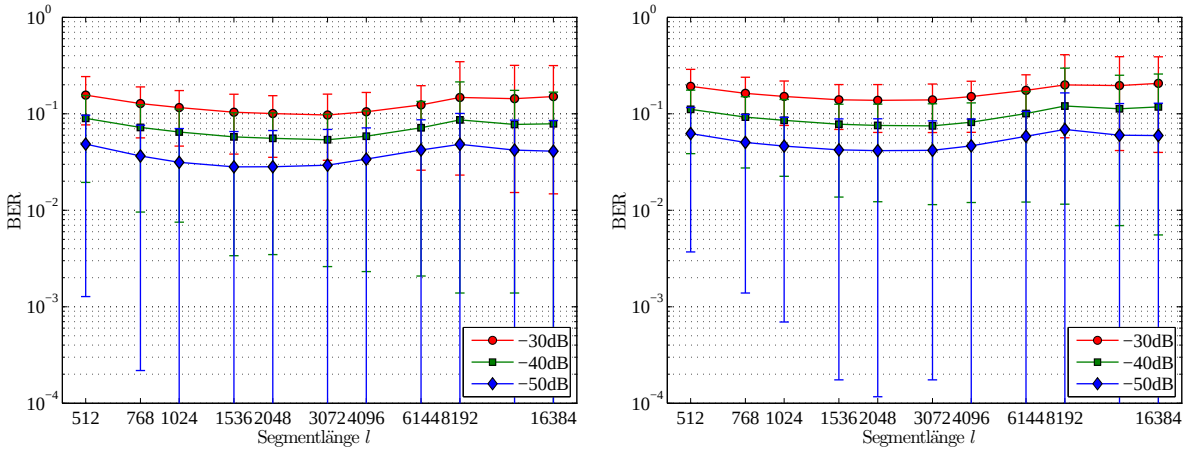


**Abbildung A.22.:** Robustheit der Wasserzeicheninformation bei einem ODG von -1 gegenüber $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts) in Abhängigkeit der Segmentlänge.

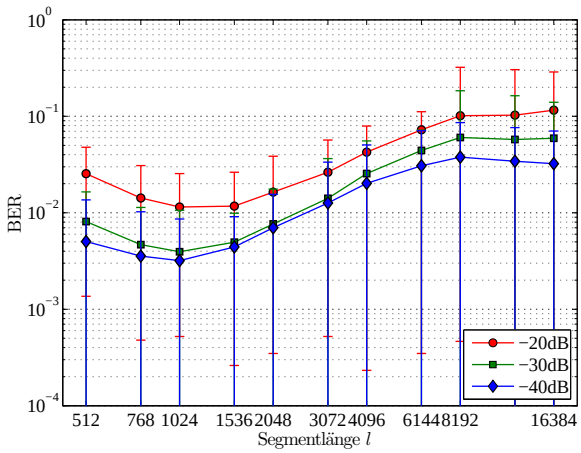


**Abbildung A.23.:** Robustheit der Wasserzeicheninformation bei einem ODG von -1 gegenüber dynamischen Rauschen (links) in Abhängigkeit der Segmentlänge.

# A.4. Leistungsanalyse des Grundkonzepts

## A.4.1. Robustheit des Inhaltsmerkmals

Die Abbildungen A.24 bis A.28 zeigen die Simulationsergebnisse für die Kombination von Merkmalsextraktion und Wasserzeichentechnik hinsichtlich der Robustheit des Inhaltsmerkmals gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Abbildungen stellen in Bezug auf die Intensität der Störungen den Wertebereich I, den Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■ und den Mittelwert □ der Bitfehlerrate der Merkmalsvektoren $\mathbf{m_i}$ dar.

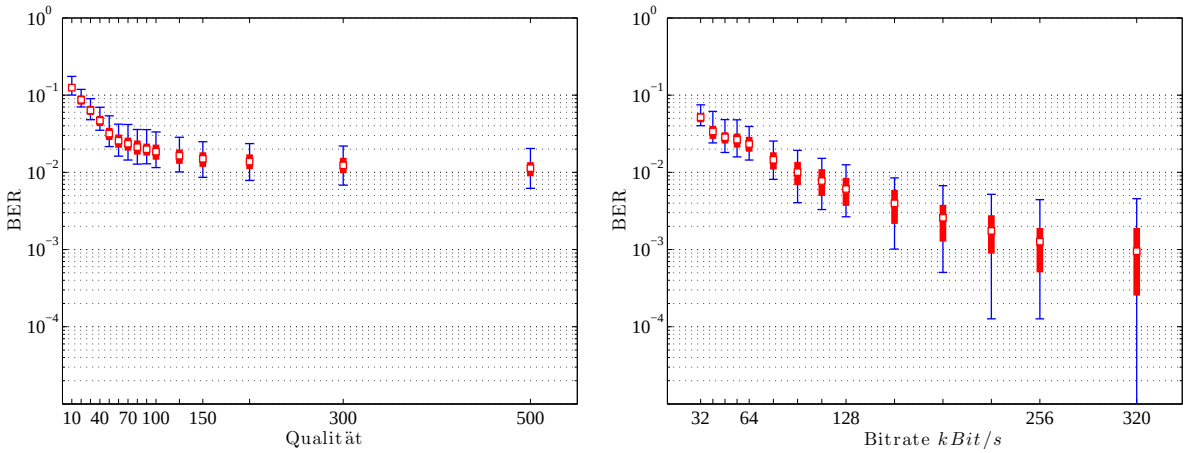


**Abbildung A.24.:** Robustheit des Inhaltsmerkmals gegenüber AAC-Kompression (links) und MP3-Kompression (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

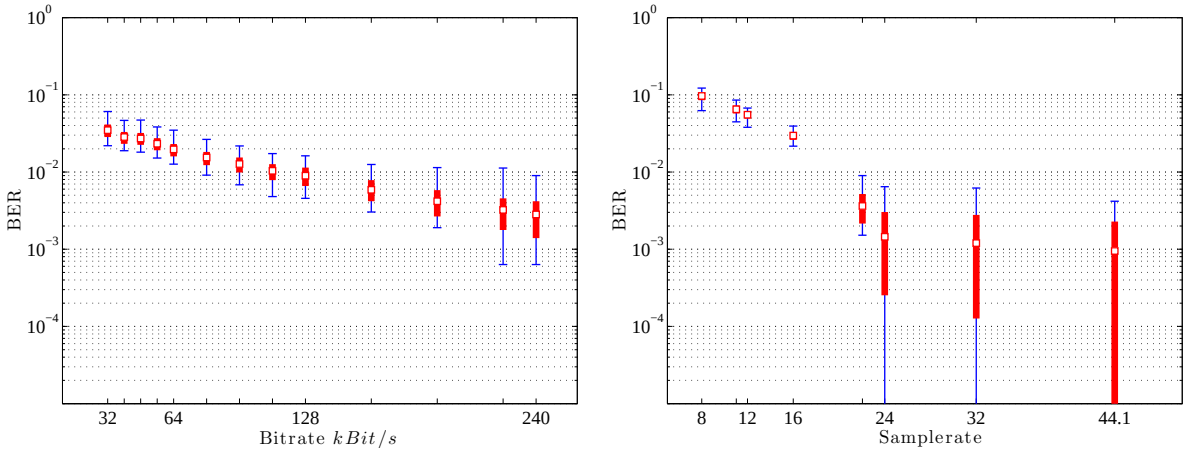


**Abbildung A.25.:** Robustheit des Inhaltsmerkmals gegenüber Vorbis-Kompression (links) und Unterabtastung (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

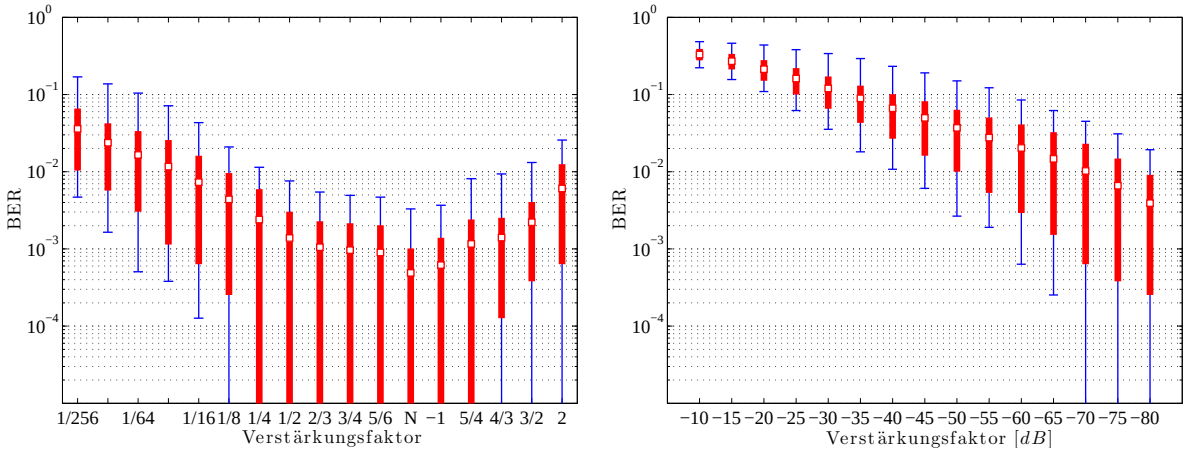


**Abbildung A.26.:** Robustheit des Inhaltsmerkmals gegenüber Laustärkeveränderung (links) und Weißem Rauschen (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

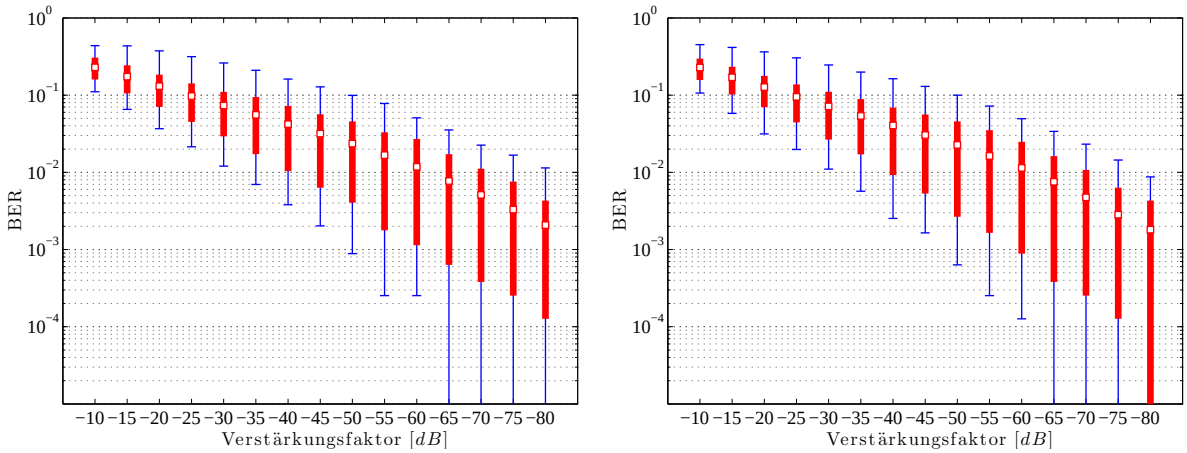


**Abbildung A.27.:** Robustheit des Inhaltsmerkmals gegenüber $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

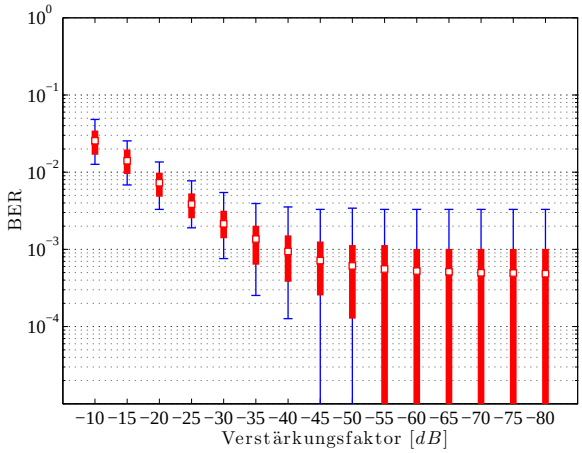


**Abbildung A.28.:** Robustheit des Inhaltsmerkmals gegenüber dynamischen Rauschen (links). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

## A.4.2. Robustheit der Wasserzeicheninformation

Die Abbildungen A.29 bis A.33 zeigen die Simulationsergebnisse für die Kombination von Merkmalsextraktion und Wasserzeichentechnik hinsichtlich der Robustheit der Wasserzeicheninformation **w** gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von

-1 eingestellt. Die Abbildungen stellen in Bezug auf die Intensität der Störungen den Wertebereich I, den Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■ und den Mittelwert □ der Bitfehlerrate der Wasserzeicheninformation **w** dar.

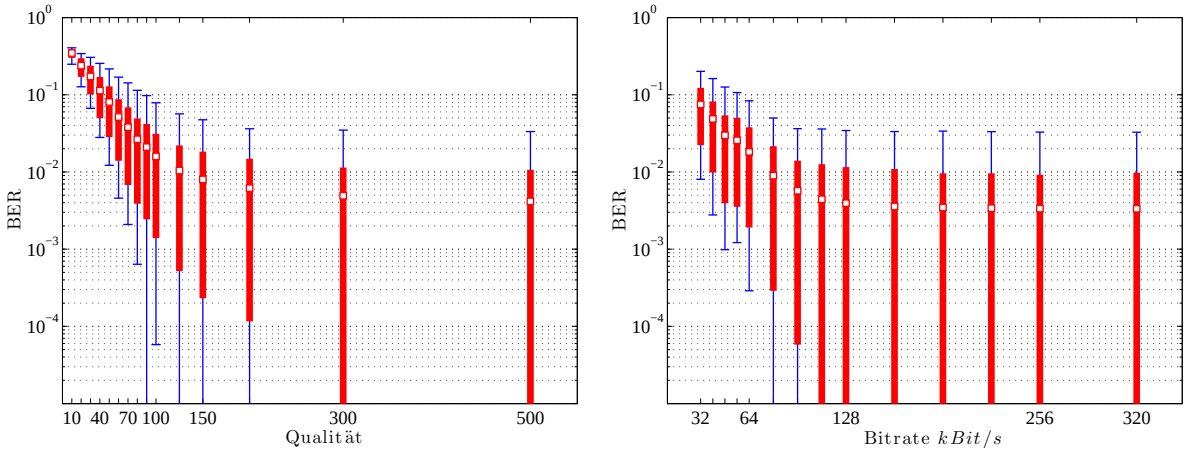


**Abbildung A.29.:** Robustheit der Wasserzeicheninformation gegenüber AAC-Kompression (links) und MP3-Kompression (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

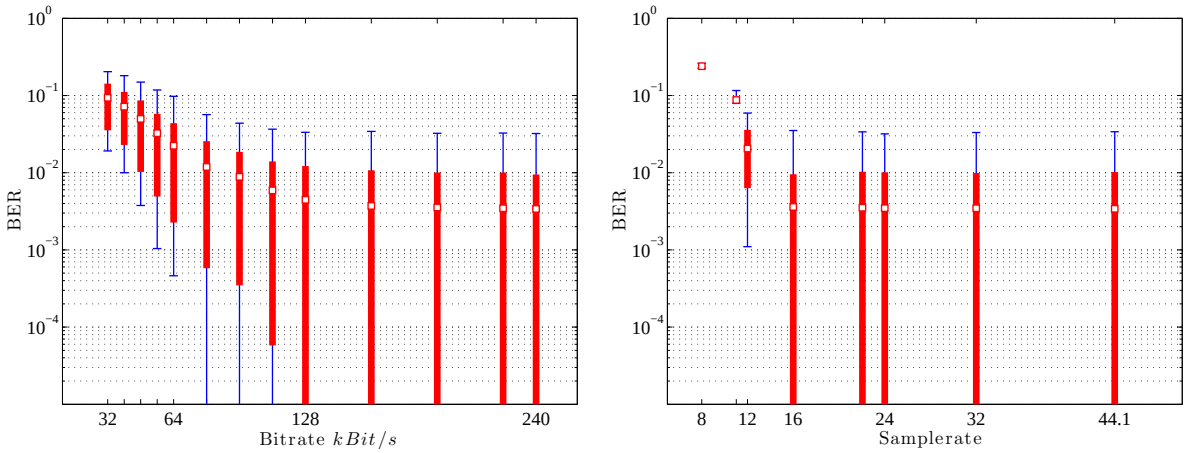


**Abbildung A.30.:** Robustheit der Wasserzeicheninformation gegenüber Vorbis-Kompression (links) und Unterabtastung (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

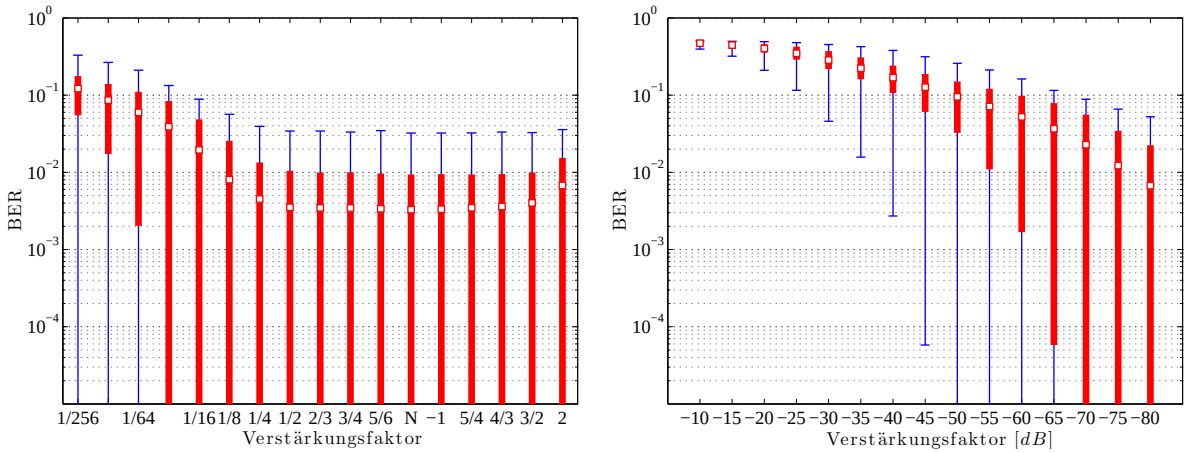


**Abbildung A.31.:** Robustheit der Wasserzeicheninformation gegenüber Laustärkeveränderung (links) und Weißem Rauschen (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

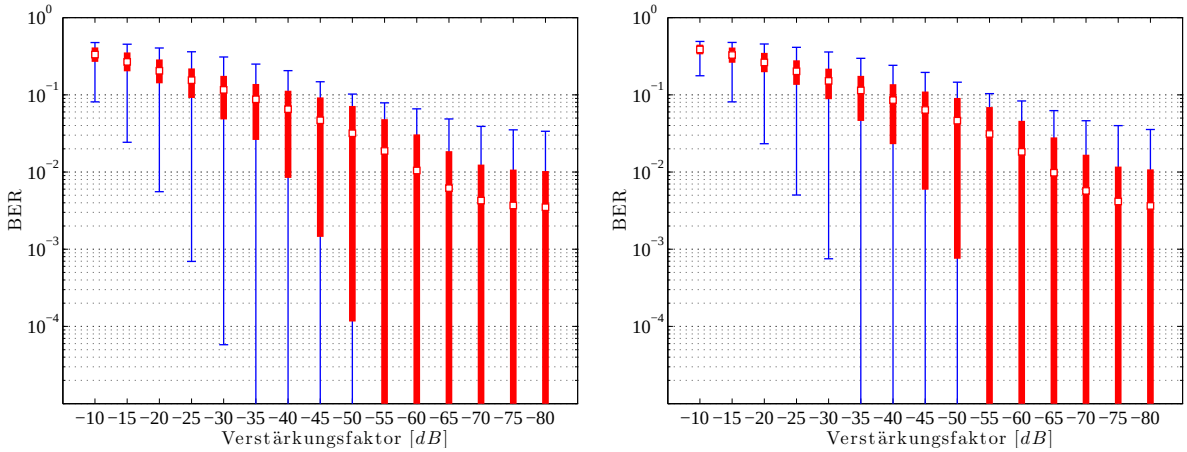


**Abbildung A.32.:** Robustheit der Wasserzeicheninformation gegenüber $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

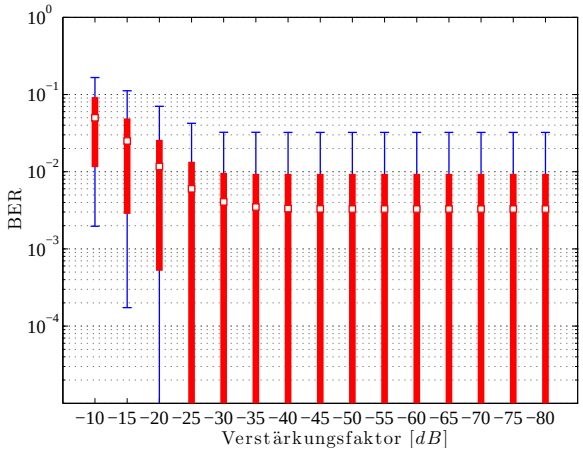


**Abbildung A.33.:** Robustheit der Wasserzeicheninformation gegenüber dynamischen Rauschen (links). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

# A.5. Leistungsanalyse der Merkmalsverstärkung mittels Totzone

## A.5.1. Robustheit des Inhaltsmerkmals

Die Abbildungen A.34 bis A.38 zeigen die Simulationsergebnisse für die Erweiterung des Grundsystems um die Merkmalsverstärkung hinsichtlich der Robustheit des Inhaltsmerkmals gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Merkmalsverstärkung erfolgt mit dem Schwellwert $Th_{tot} = 0,0001$. Die Abbildungen stellen in Bezug auf die Intensität der Störungen den Wertebereich I, den Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■ und den Mittelwert □ der Bitfehlerrate der Merkmalsvektoren $\mathbf{m_i}$ dar. Als Vergleichsgrößen sind der Wertebereich I, der Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■ und der Mittelwert □ der Bitfehlerrate der Merkmalsvektoren $\mathbf{m_i}$ für das Grundsystem ohne Merkmalsverstärkung dargestellt.

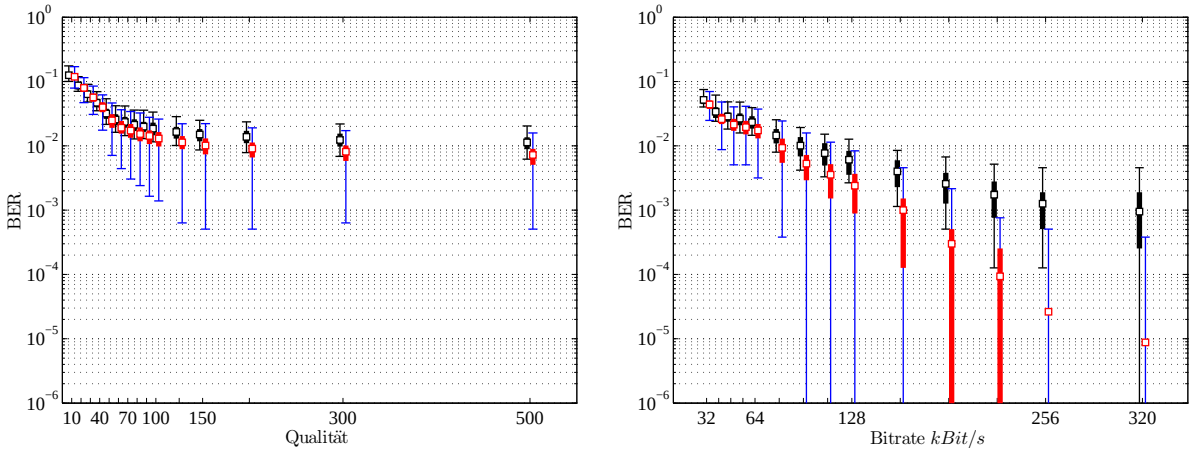


**Abbildung A.34.:** Robustheit des Inhaltsmerkmals gegenüber AAC-Kompression (links) und MP3-Kompression (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

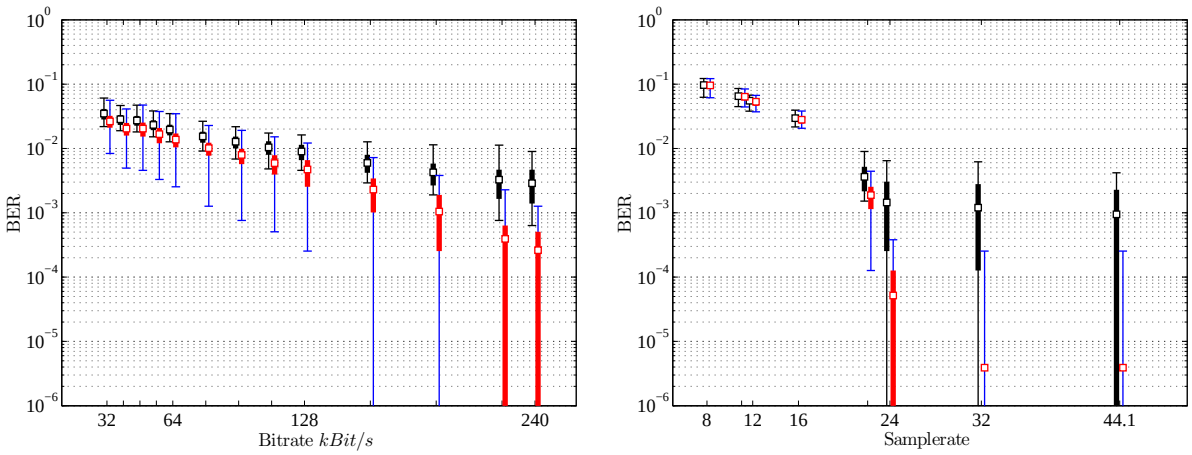


**Abbildung A.35.:** Robustheit des Inhaltsmerkmals gegenüber Vorbis-Kompression (links) und Unterabtastung (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

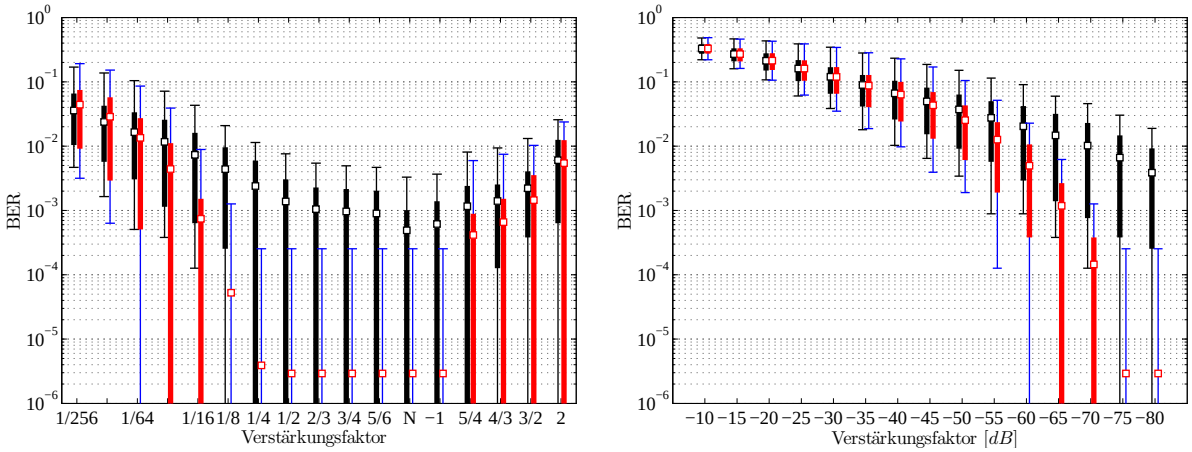


**Abbildung A.36.:** Robustheit des Inhaltsmerkmals gegenüber Laustärkeveränderung (links) und Weißem Rauschen (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

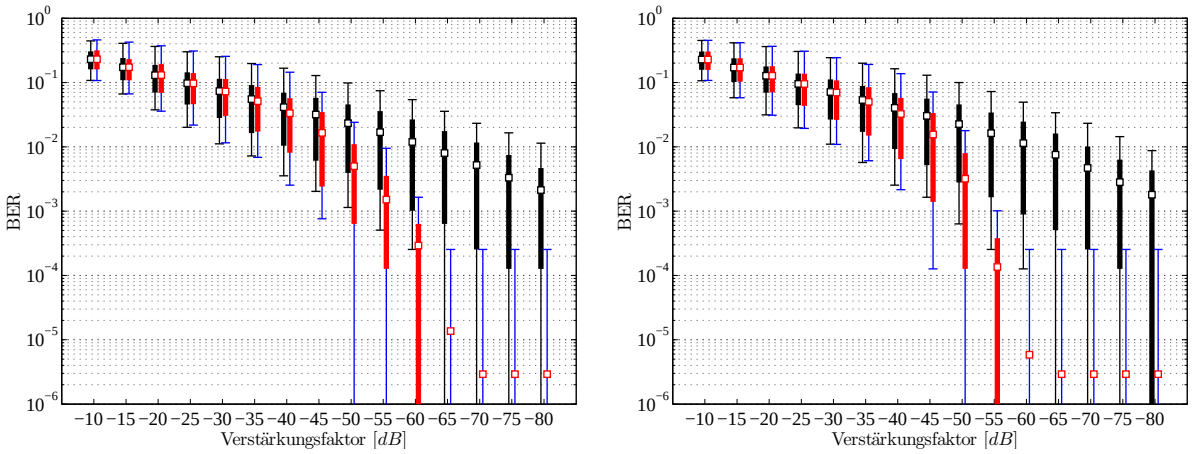


**Abbildung A.37.:** Robustheit des Inhaltsmerkmals gegenüber $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

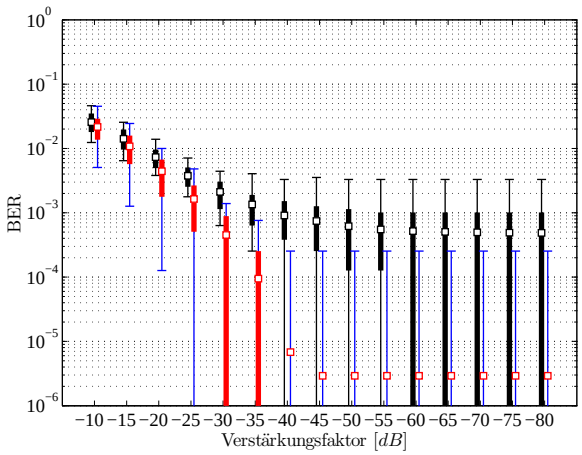


**Abbildung A.38.:** Robustheit des Inhaltsmerkmals gegenüber dynamischen Rauschen (links). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

## A.5.2. Robustheit der Wasserzeicheninformation

Die Abbildungen A.39 bis A.43 zeigen die Simulationsergebnisse für die Erweiterung des Grundsystems um die Merkmalsverstärkung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Merkmalsverstärkung erfolgt mit dem Schwellwert $Th_{tot} = 0,0001$. Die Abbildungen stellen in Bezug auf die Intensität der Störungen den Wertebereich I, den Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■ und den Mittelwert □ der Bitfehlerrate der Wasserzeicheninformation **w** dar. Als Vergleichsgrößen sind der Wertebereich I, der Bereich zwischen dem 15,85%- und 84,15%-Fraktil ■ und der Mittelwert □ der Bitfehlerrate der Wasserzeicheninformation **w** für das Grundsystem ohne Merkmalsverstärkung dargestellt.

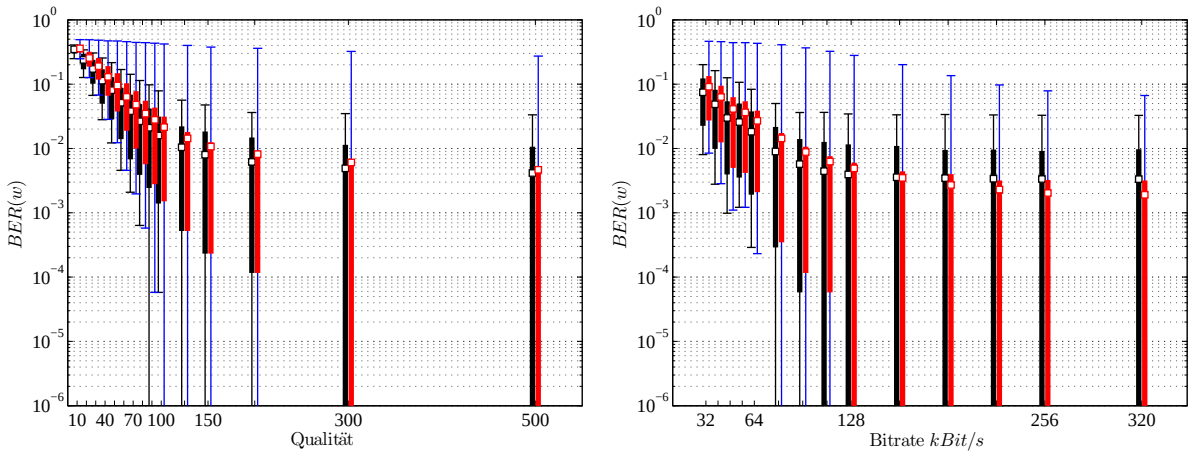


**Abbildung A.39.:** Robustheit der Wasserzeicheninformation gegenüber AAC-Kompression (links) und MP3-Kompression (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

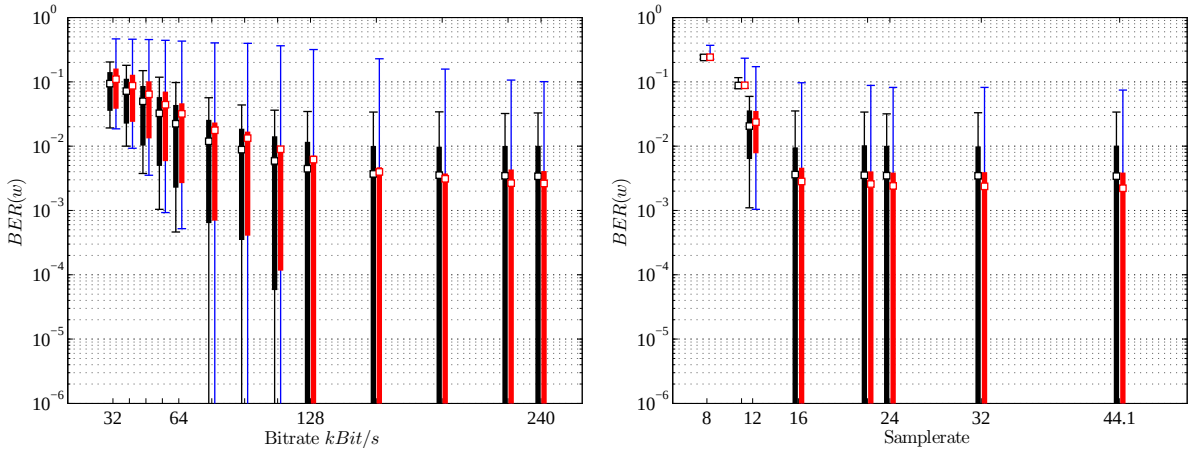


**Abbildung A.40.:** Robustheit der Wasserzeicheninformation gegenüber Vorbis-Kompression (links) und Unterabtastung (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

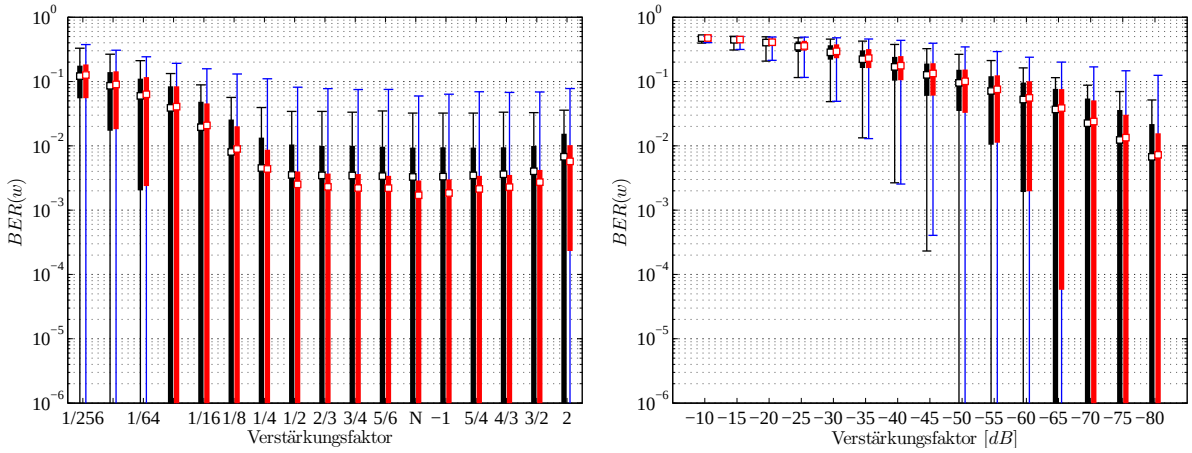


**Abbildung A.41.:** Robustheit der Wasserzeicheninformation gegenüber Laustärkeveränderung (links) und Weißem Rauschen (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

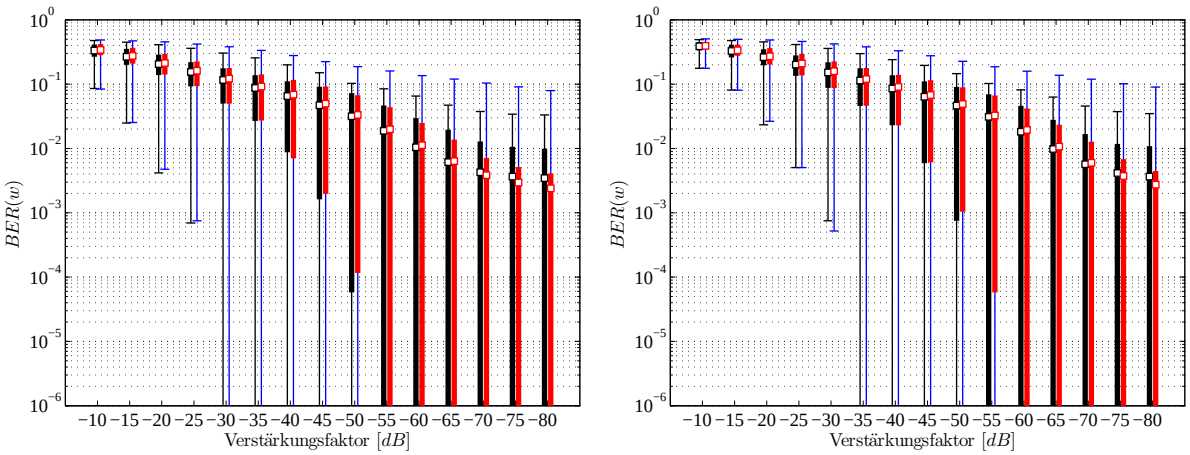


**Abbildung A.42.:** Robustheit der Wasserzeicheninformation gegenüber $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

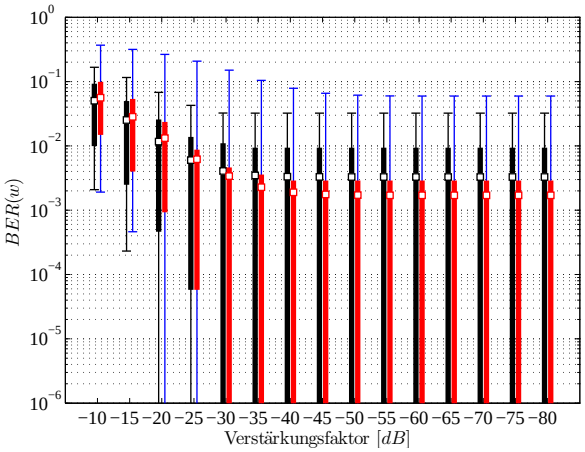


**Abbildung A.43.:** Robustheit der Wasserzeicheninformation gegenüber dynamischen Rauschen (links). Segmentlängen: $l_M = 8192$, $l_W = 1024$; Merkmalsdomain: $F2$-$F23$; Einbettungsdomain: $F15$-$F20$; Transparenz: ODG $= -1$

# A.6. Leistungsanalyse der *Soft-Input*-Decodierung

## A.6.1. normierte Distanz $\varepsilon'_{norm}$

Die Abbildungen A.44 bis A.44 zeigen die Simulationsergebnisse für die Schätzung der Dichten $\hat{p}(\varepsilon'_{norm})$ der geschätzten normierten Distanzen $\varepsilon'_{norm}$ für zulässige Störungen bzw. Störungen des Graubereichs. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen

$F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Bestimmung erfolgte unter Kenntnis der gesendeten Wasserzeicheninformation $w$.

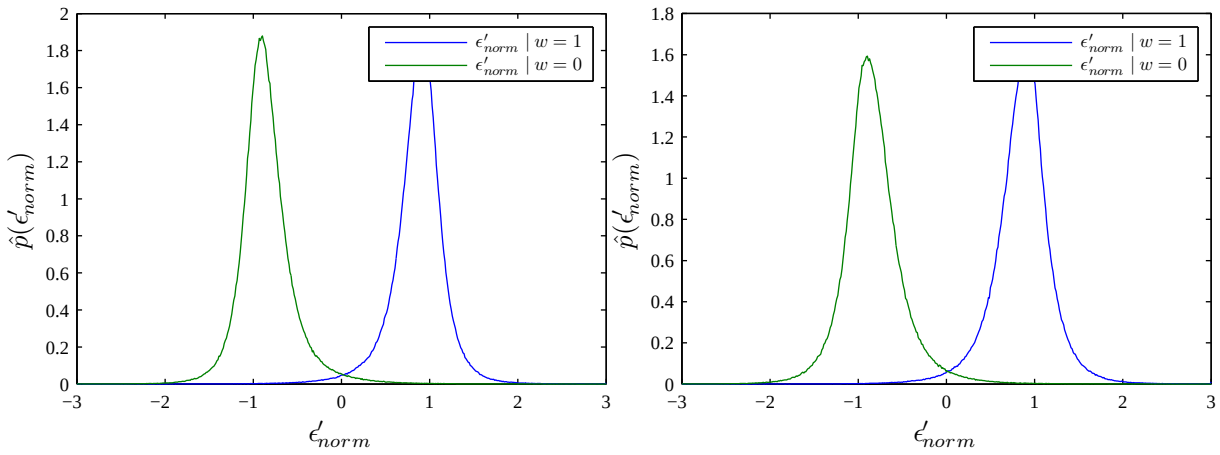


**Abbildung A.44.:** Geschätzte Dichte $\hat{p}(\varepsilon'_{norm})$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch verlustbehaftete AAC-Kompression: Q = 100 (links) und MP3-Kompression: 64 kBit/s (rechts) und unter der Bedingung, dass die gesendete Wasserzeicheninformation $w$ mit 1 bzw. 0 vorlag.

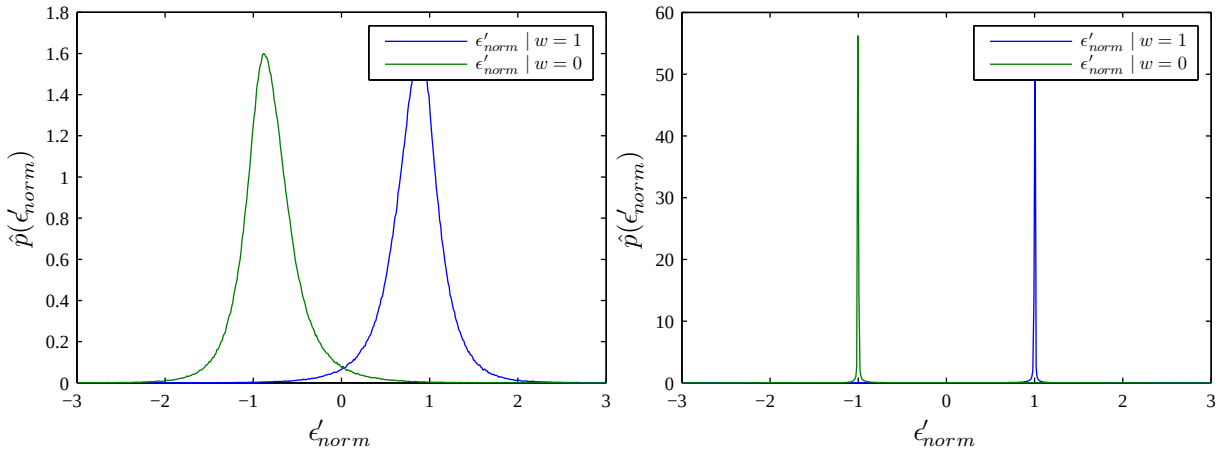


**Abbildung A.45.:** Geschätzte Dichte $\hat{p}(\varepsilon'_{norm})$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch Vorbis-Kompression: 64 kBit/s (links) und Unterabtastung: 16 kHz (rechts) und unter der Bedingung, dass die gesendete Wasserzeicheninformation $w$ mit 1 bzw. 0 vorlag.

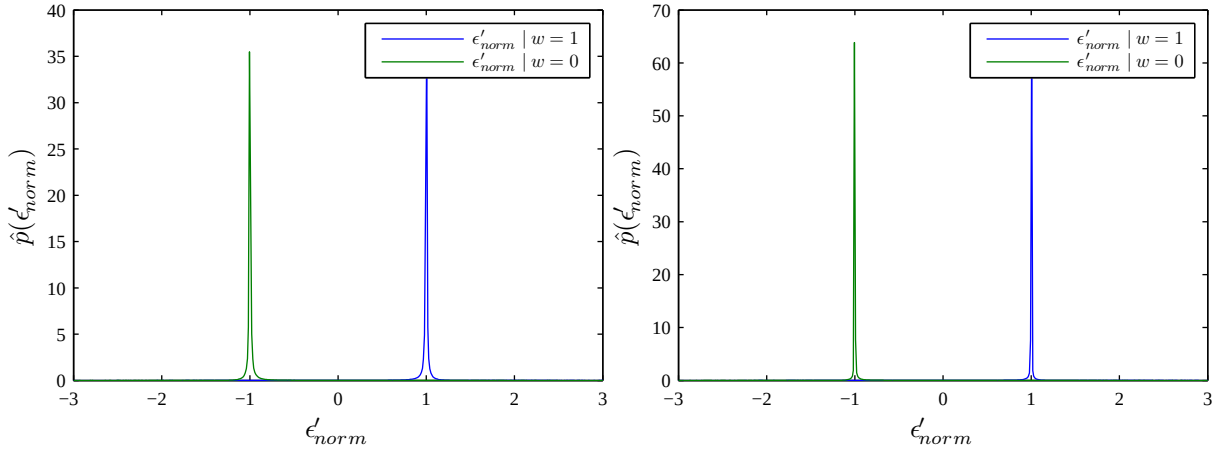


**Abbildung A.46.:** Geschätzte Dichte $\hat{p}(\varepsilon'_{norm})$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch Unterabtastung: 32 kHz (links) und Lautstärkeveränderung: k=1,34 (rechts) und unter der Bedingung, dass die gesendete Wasserzeicheninformation $w$ mit 1 bzw. 0 vorlag.

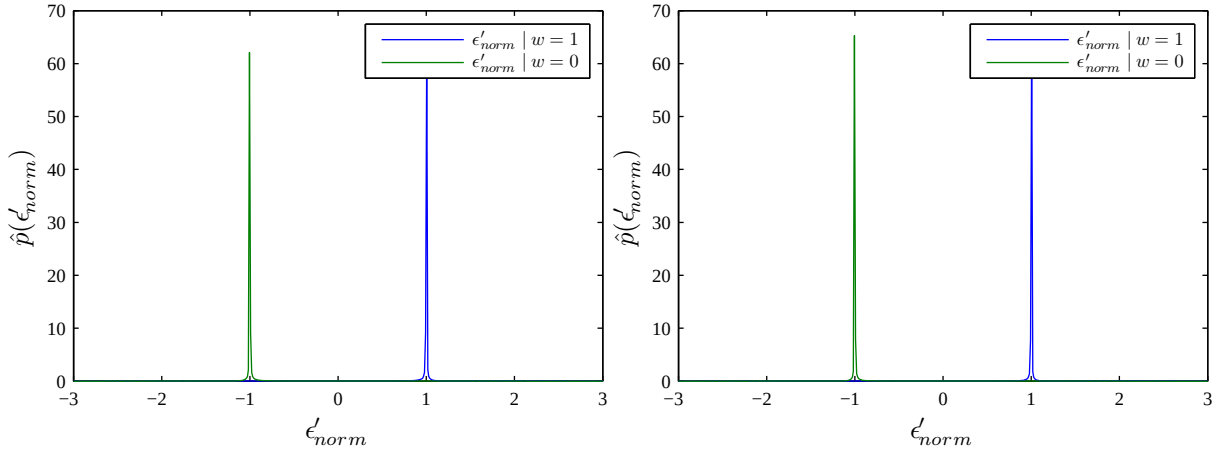


**Abbildung A.47.:** Geschätzte Dichte $\hat{p}(\varepsilon'_{norm})$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch Normalisierung (links) und Lautstärkeveränderung: k=0,667 (rechts) und unter der Bedingung, dass die gesendete Wasserzeicheninformation $w$ mit 1 bzw. 0 vorlag.

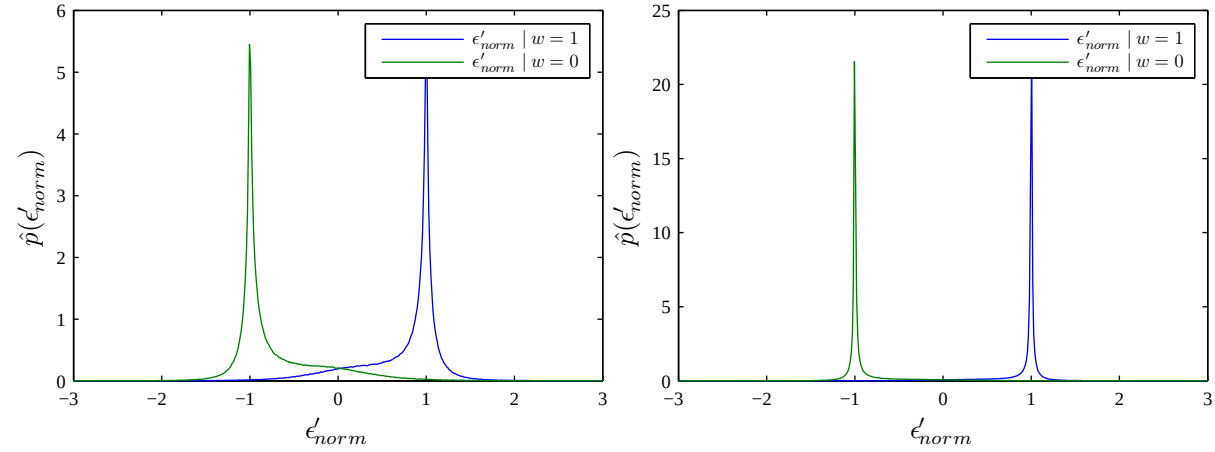


**Abbildung A.48.:** Geschätzte Dichte $\hat{p}(\varepsilon'_{norm})$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch Weißes Rauschen: k=-50 dB (links) und $1/f^2$-Rauschen: k= -50 dB (rechts) und unter der Bedingung, dass die gesendete Wasserzeicheninformation $w$ mit 1 bzw. 0 vorlag.

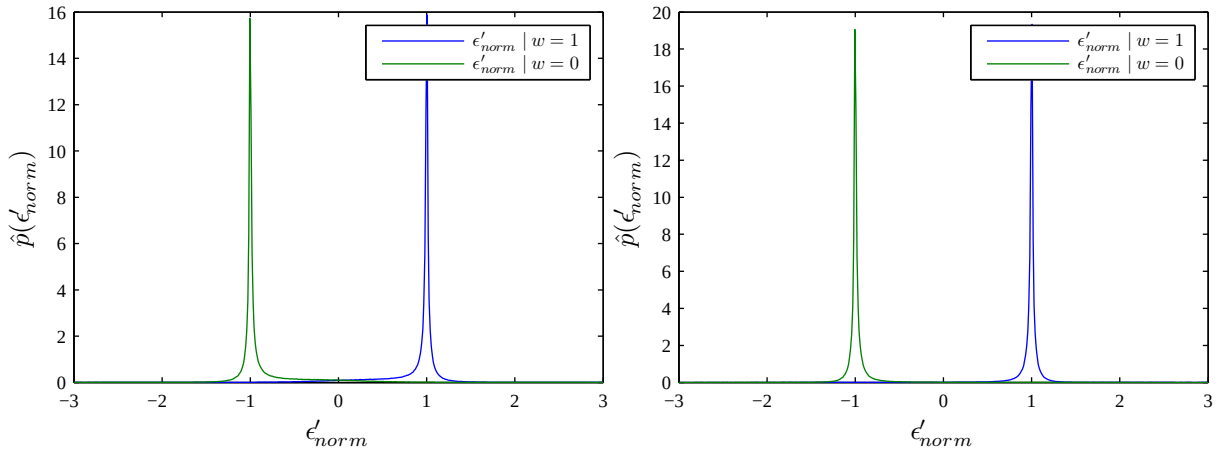


**Abbildung A.49.:** Geschätzte Dichte $\hat{p}(\varepsilon'_{norm})$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch $1/f$-Rauschen: k=-50 dB (links) und dynamisches Rauschen: k=-30 dB (rechts) und unter der Bedingung, dass die gesendete Wasserzeicheninformation $w$ mit 1 bzw. 0 vorlag.

## A.6.2. Kanalinformation

Die Abbildungen A.50 bis A.55 zeigen die Simulationsergebnisse für die aus den geschätzten Dichten $\hat{p}(\varepsilon'_{norm})$ der normierten Distanzen $\varepsilon'_{norm}$ resultierende Kanalinformation $L_c(\hat{p}(\varepsilon'_{norm}))$ bei Störung der Trägerdaten durch zulässige Störungen bzw. Störungen des Graubereichs. Die Bestimmung erfolgte unter Kenntnis der gesendeten Wasserzeicheninformation $w$.

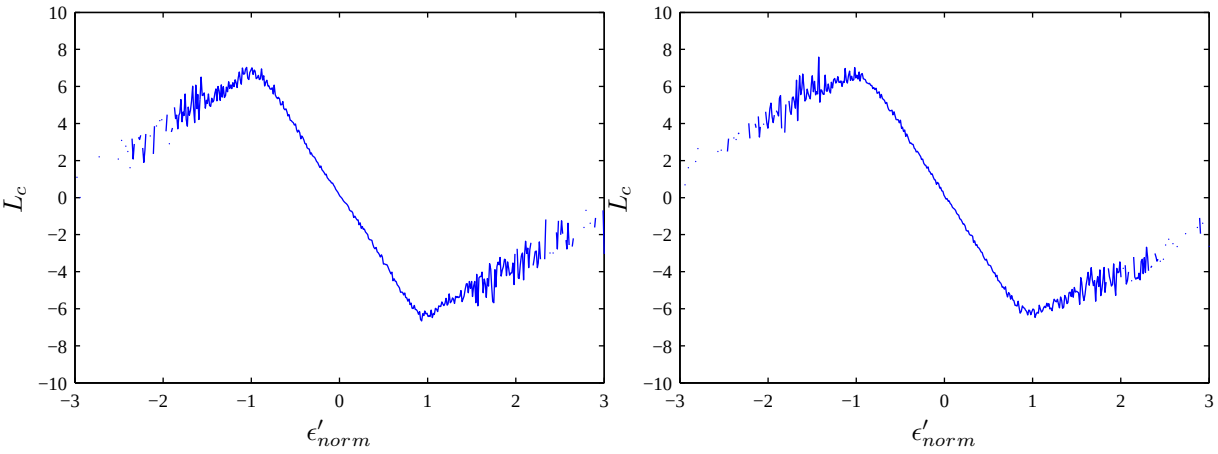


**Abbildung A.50.:** Kanalinformation $L_c(\hat{p}(\varepsilon'_{norm}))$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch verlustbehaftete AAC-Kompression: Q = 100 (links) und MP3-Kompression: 64 kBit/s (rechts).

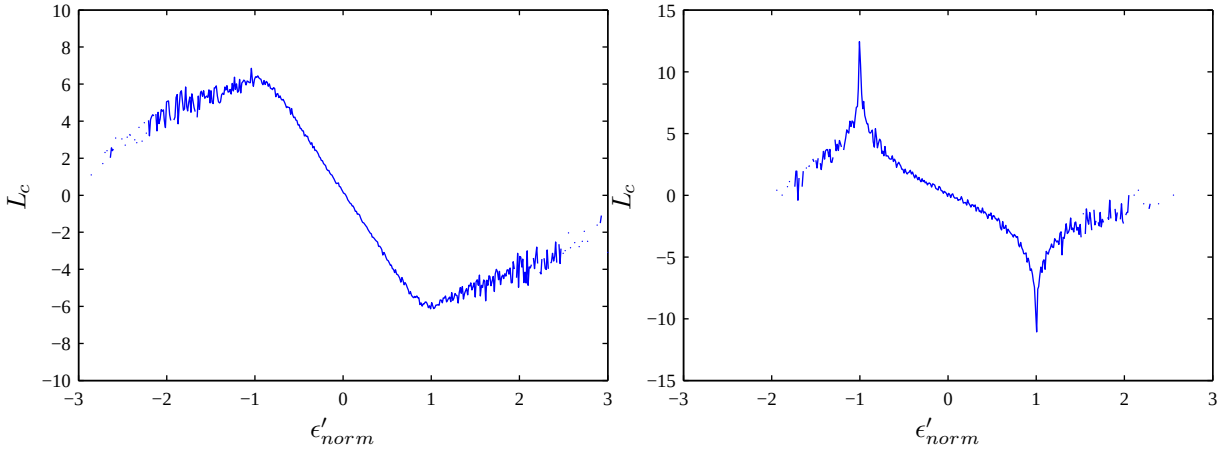


**Abbildung A.51.:** Kanalinformation $L_c(\hat{p}(\varepsilon'_{norm}))$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch Vorbis-Kompression: 64 kBit/s (links) und Unterabtastung: 16 kHz (rechts).

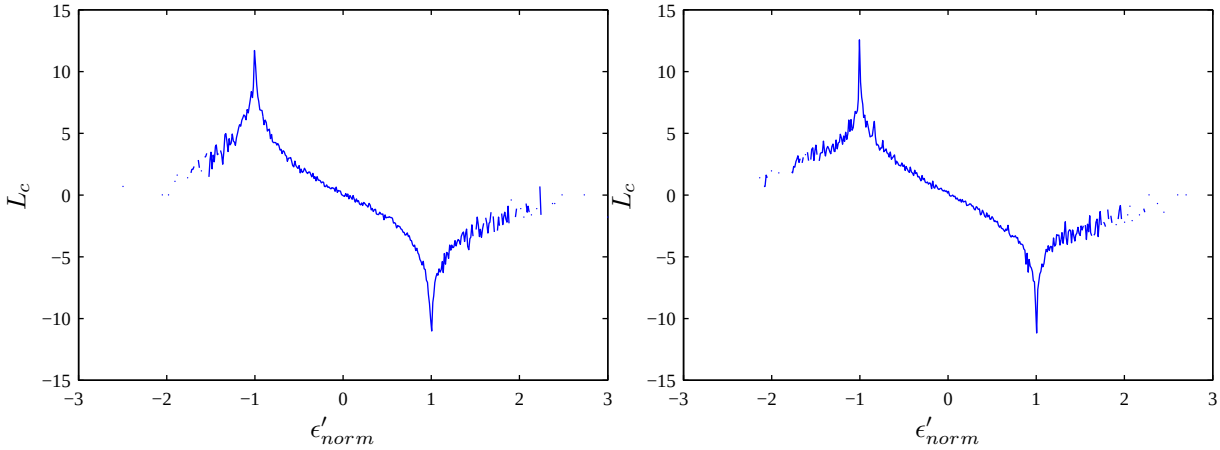


**Abbildung A.52.:** Kanalinformation $L_c(\hat{p}(\varepsilon'_{norm}))$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch Unterabtastung: 32 kHz (links) und Lautstärkeveränderung: k=1,34 (rechts).

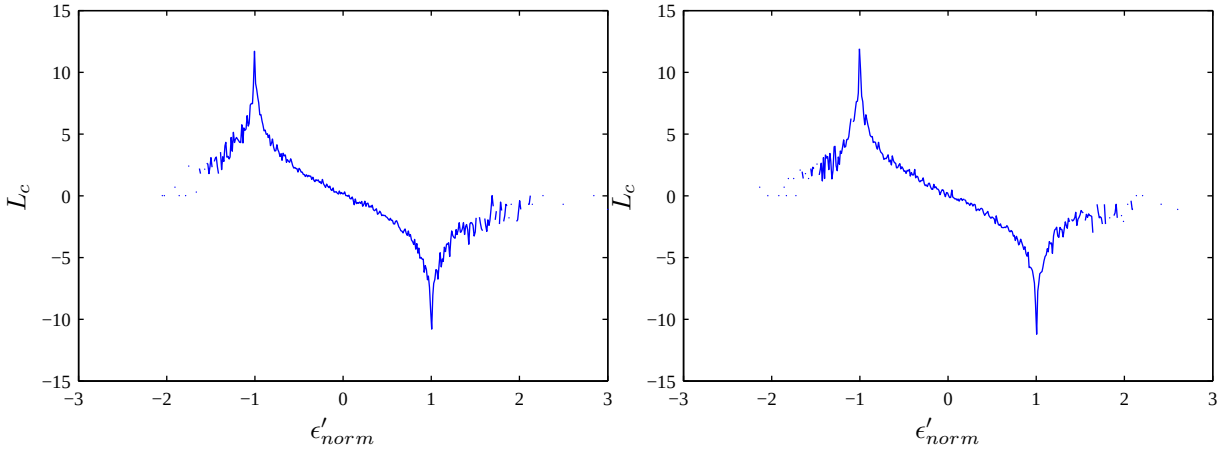


**Abbildung A.53.:** Kanalinformation $L_c(\hat{p}(\varepsilon'_{norm}))$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch Normalisierung (links) und Lautstärkeveränderung: k=0,667 (rechts).

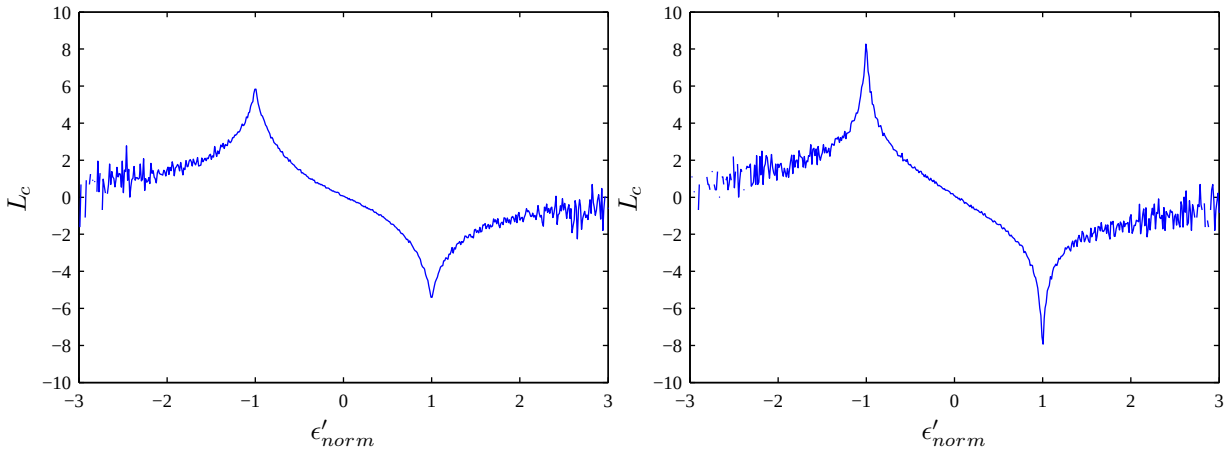


**Abbildung A.54.:** Kanalinformation $L_c(\hat{p}(\varepsilon'_{norm}))$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch Weißes Rauschen: k=-50 dB (links) und $1/f^2$-Rauschen: k= -50 dB (rechts).

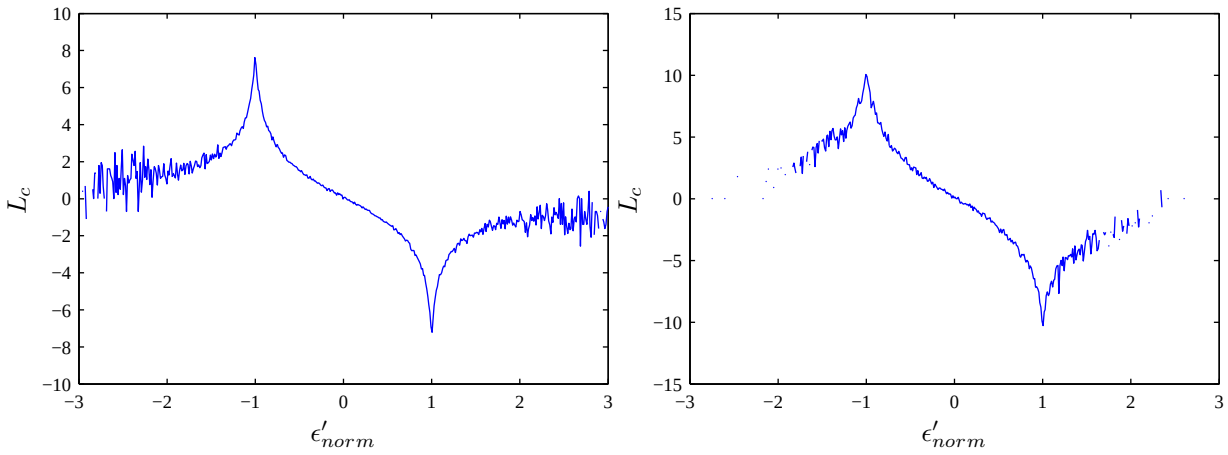


**Abbildung A.55.:** Kanalinformation $L_c(\hat{p}(\varepsilon'_{norm}))$ der normierten Distanz $\varepsilon'_{norm}$ bei Störung der Trägerdaten durch $1/f$-Rauschen: k=-50 dB (links) und dynamisches Rauschen: k=-30 dB (rechts).

## A.6.3. Fehlerkorrektur für das Grundsystem

### A.6.3.1. Fehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.56 bis A.60 zeigen die Simulationsergebnisse für die Erweiterung des Grundsystems um die *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$

bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die Fehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar, also den Anteil der Audiorahmen für den die verschlüsselte Nutzinformation $\mathbf{v}$ wenigsten einen Bitfehler enthält. Als Vergleichsgrößen werden die Fehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt.

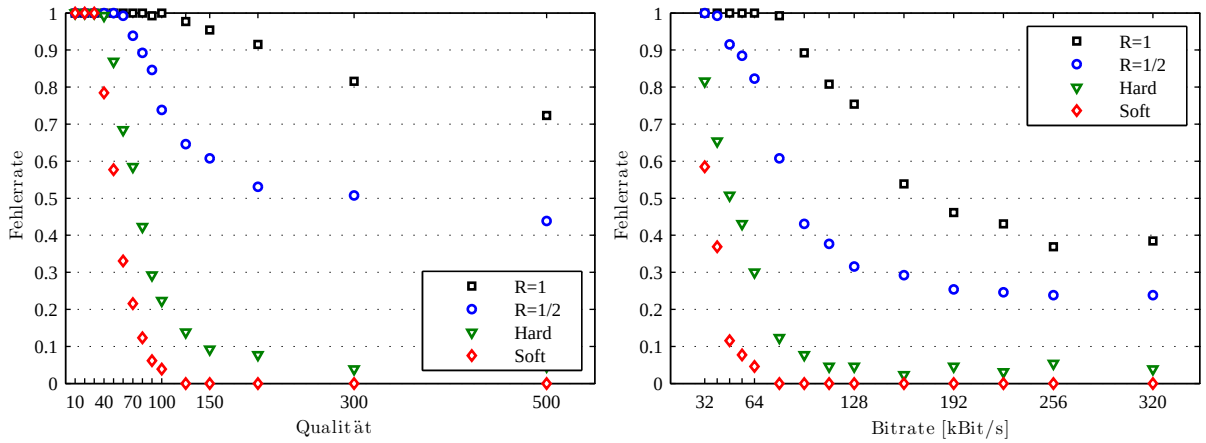


**Abbildung A.56.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

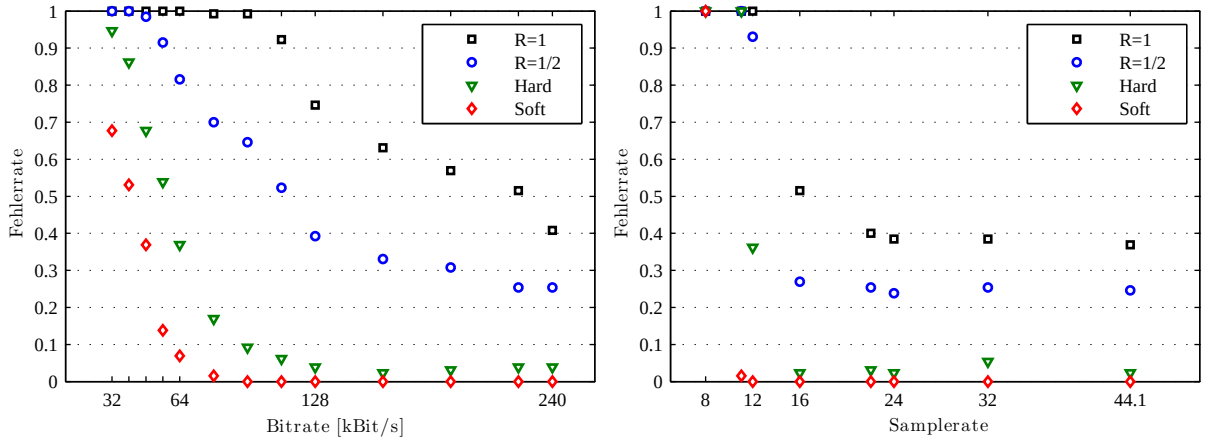


**Abbildung A.57.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

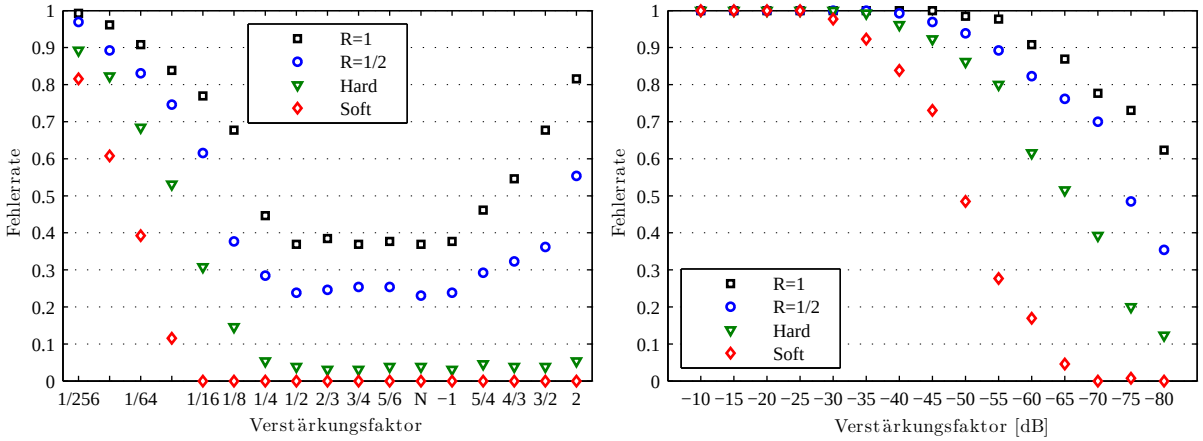


**Abbildung A.58.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

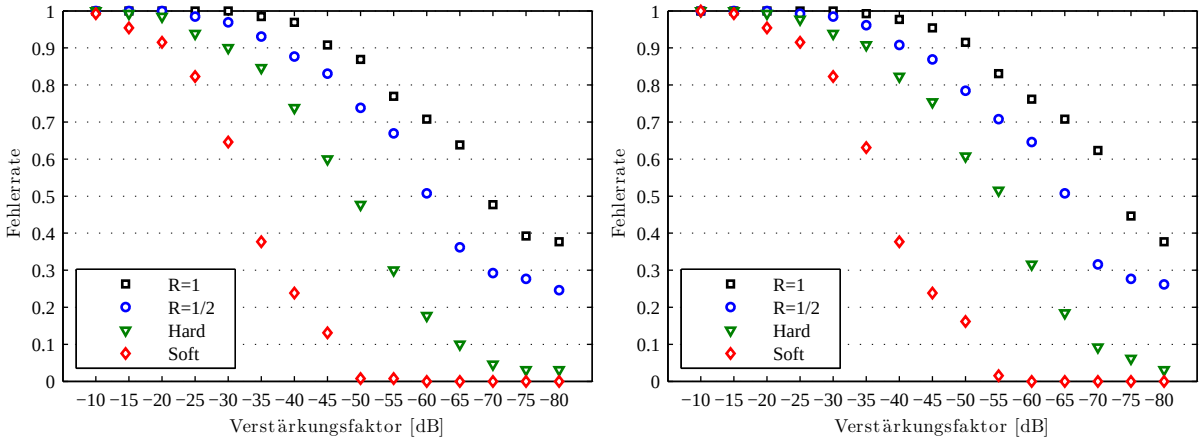


**Abbildung A.59.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

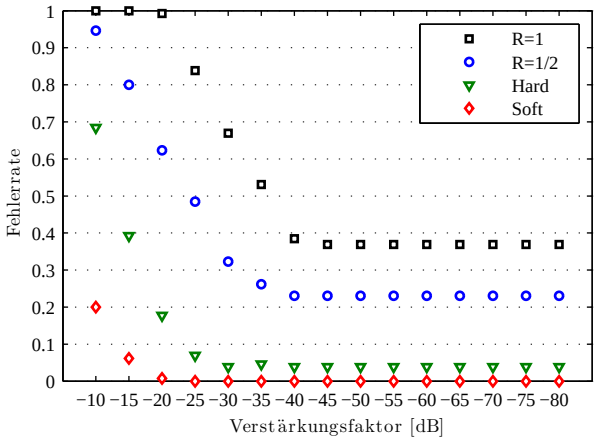


**Abbildung A.60.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch dynamisches Rauschen (links).

### A.6.3.2. Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.61 bis A.65 zeigen die Simulationsergebnisse für die Erweiterung des Grundsystems um die *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar. Als Vergleichsgrößen werden die mittleren Bitfehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt.

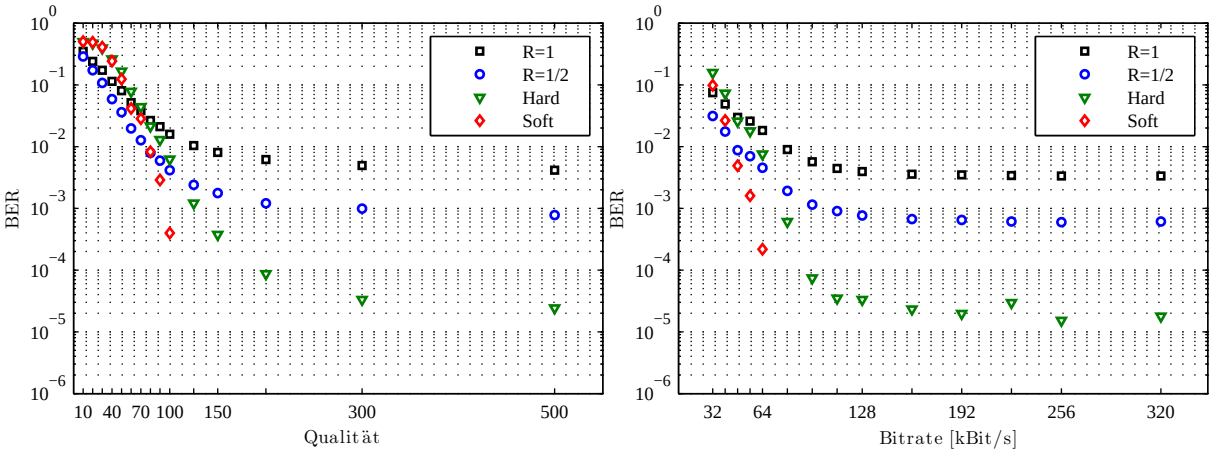


**Abbildung A.61.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

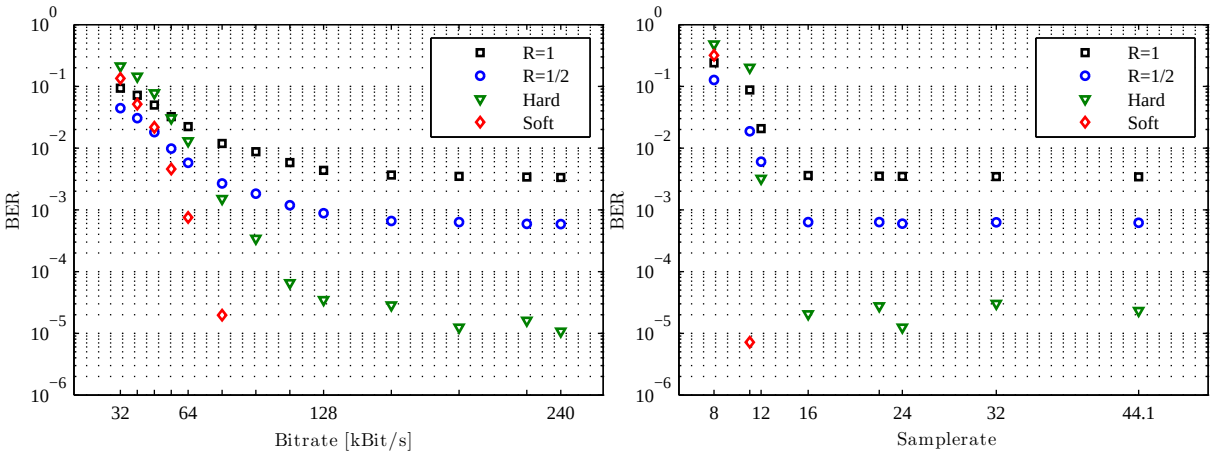


**Abbildung A.62.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

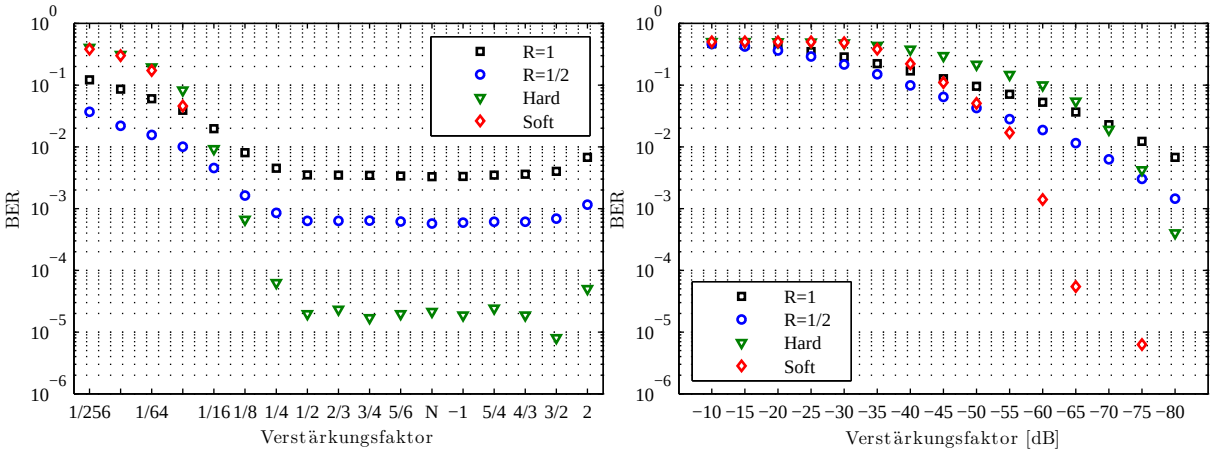


**Abbildung A.63.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

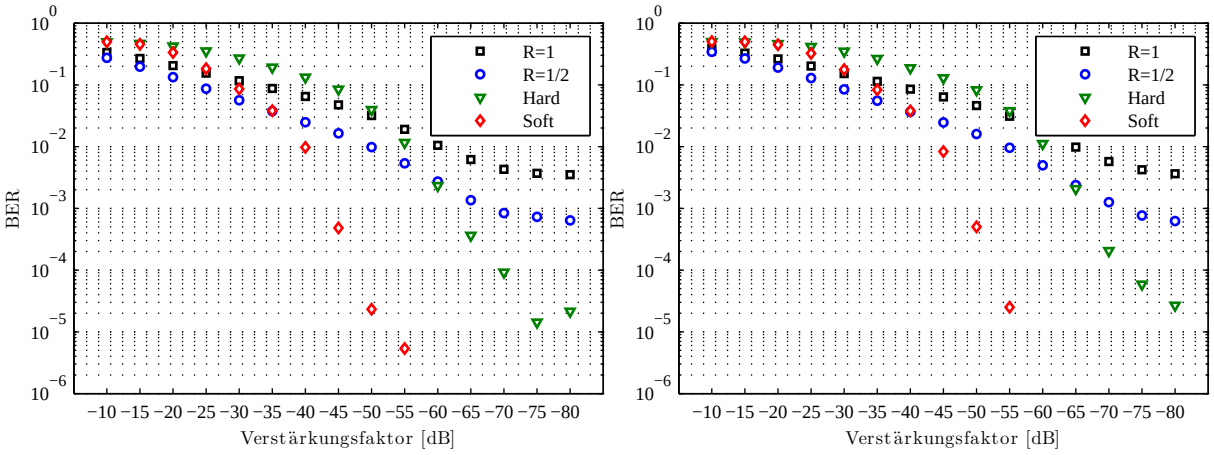


**Abbildung A.64.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

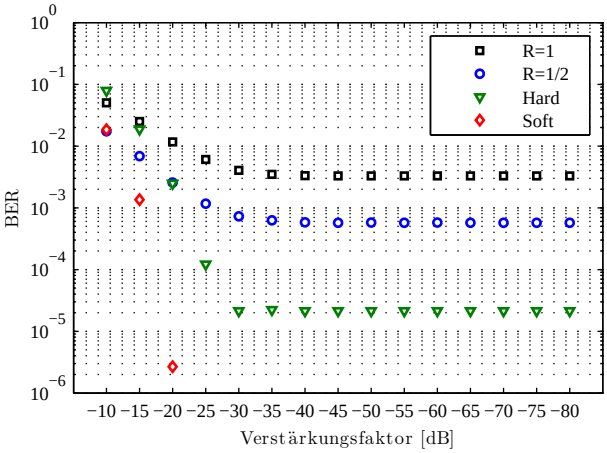


**Abbildung A.65.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch dynamisches Rauschen (links).

### A.6.3.3. Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.66 bis A.70 zeigen die Simulationsergebnisse für die Erweiterung des Grundsystems um die *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die maximale Bitfehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar, also den schlechtesten Wert für die Robustheit in Abhängigkeit der Testdaten. Als Vergleichsgrößen werden die maximalen Bitfehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt.

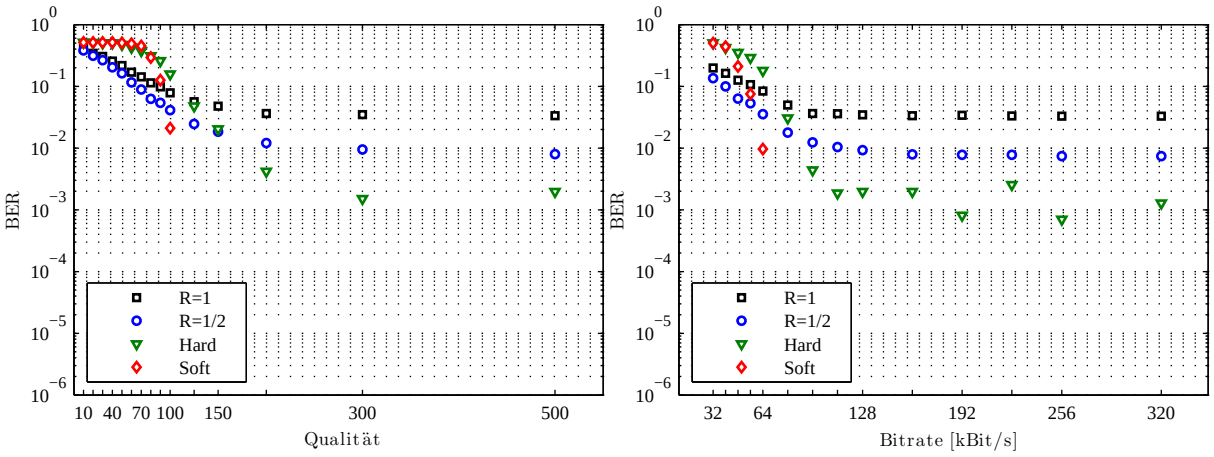


**Abbildung A.66.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

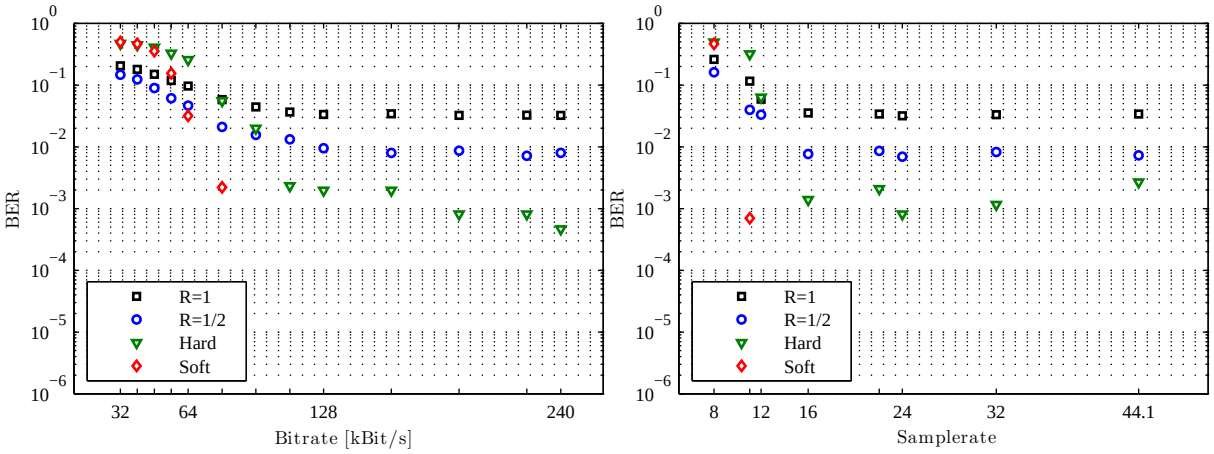


**Abbildung A.67.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

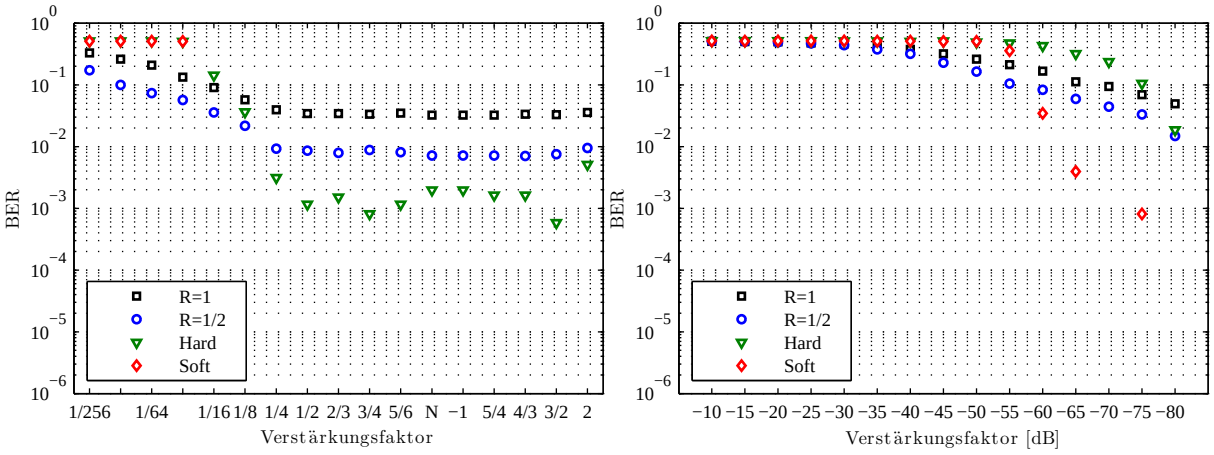


**Abbildung A.68.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

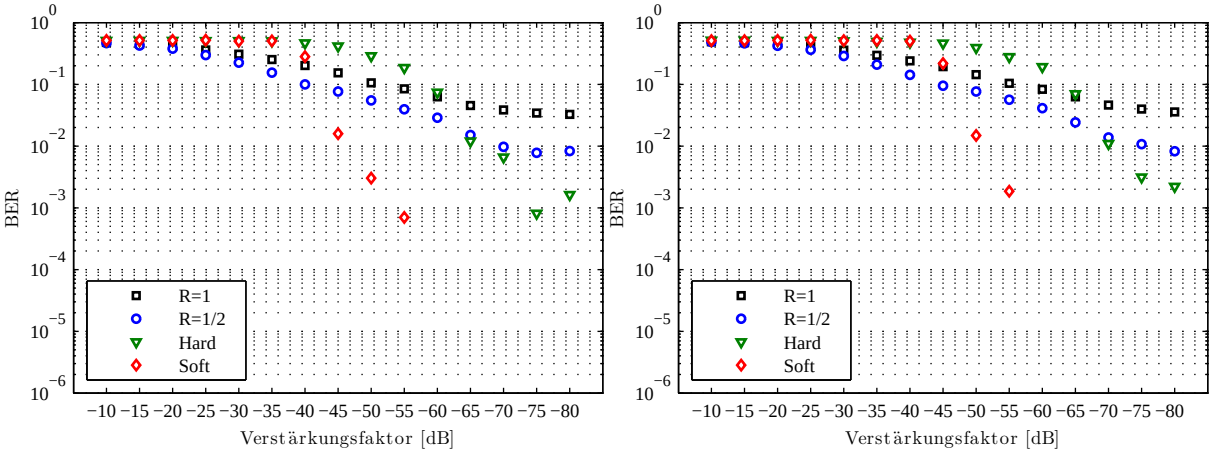


**Abbildung A.69.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

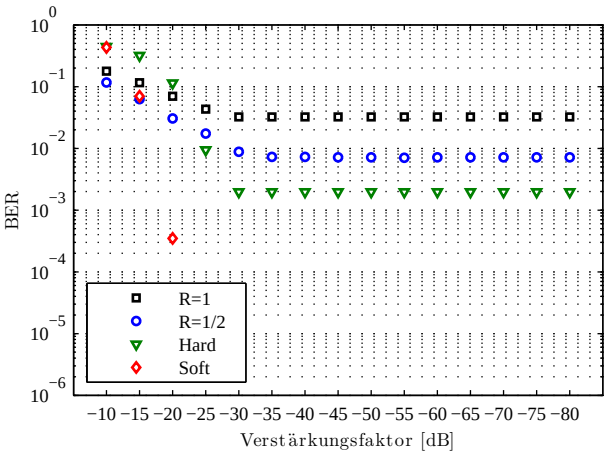


**Abbildung A.70.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem bei Störung der Trägerdaten durch dynamisches Rauschen (links).

## A.6.4. Fehlerkorrektur für das Grundsystem mit Totzone

### A.6.4.1. Fehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.71 bis A.75 zeigen die Simulationsergebnisse für die Erweiterung des Grundsystems mit Merkmalsverstärkung um die *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch

die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Merkmalsverstärkung erfolgt mit dem Schwellwert $Th_{tot}$=0,0001. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die Fehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar, also den Anteil der Audiorahmen für den die verschlüsselte Nutzinformation **v** wenigsten einen Bitfehler enthält. Als Vergleichsgrößen werden die Fehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt.

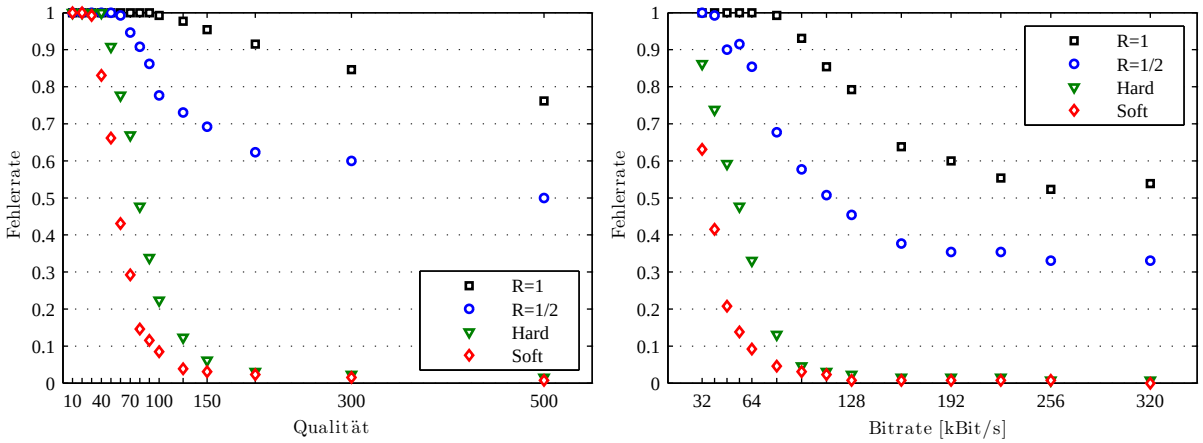


**Abbildung A.71.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

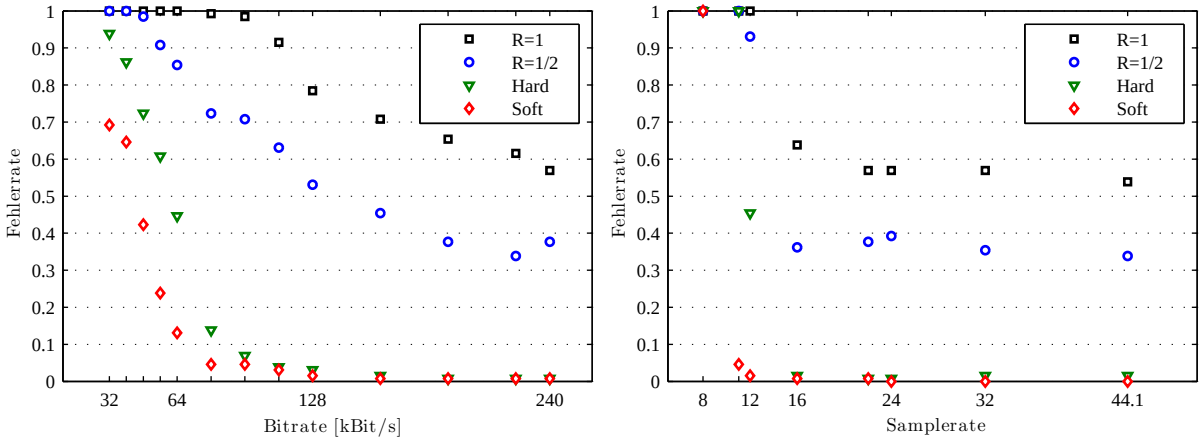


**Abbildung A.72.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

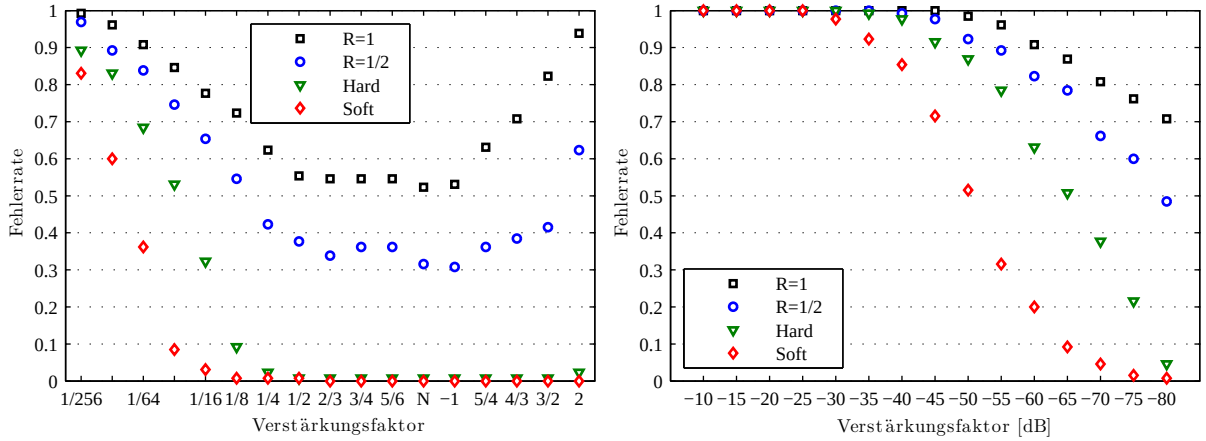


**Abbildung A.73.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

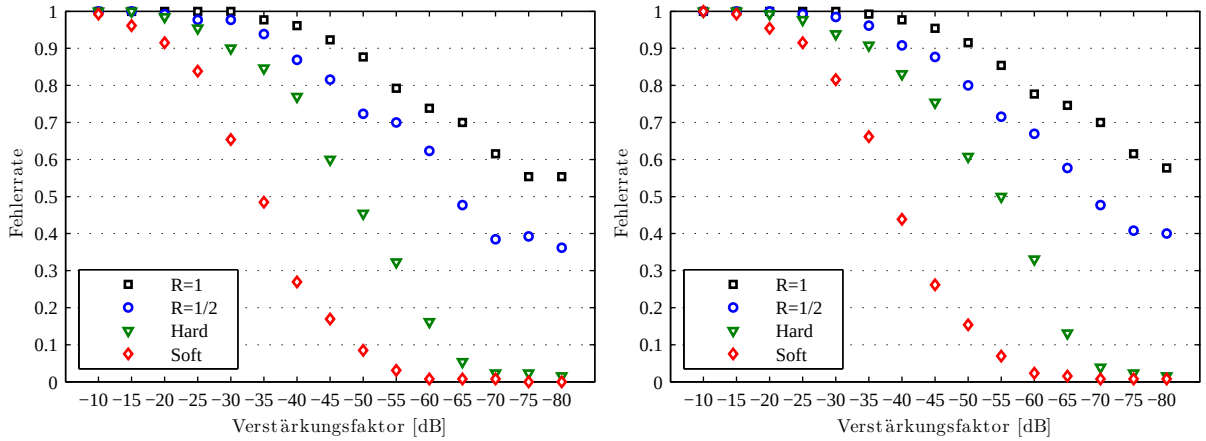


**Abbildung A.74.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

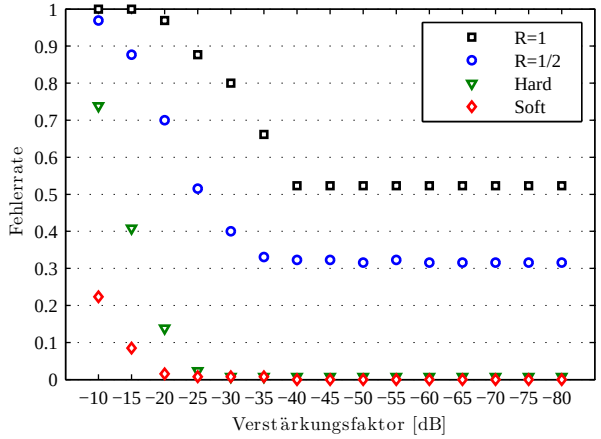


**Abbildung A.75.:** Fehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch dynamisches Rauschen (links).

### A.6.4.2. Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.76 bis A.80 zeigen die Simulationsergebnisse für die Erweiterung des Grundsystems mit Merkmalsverstärkung um die *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Merkmalsverstärkung erfolgt mit dem Schwellwert $Th_{tot}$=0,0001. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar. Als Vergleichsgrößen werden die mittleren Bitfehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt.

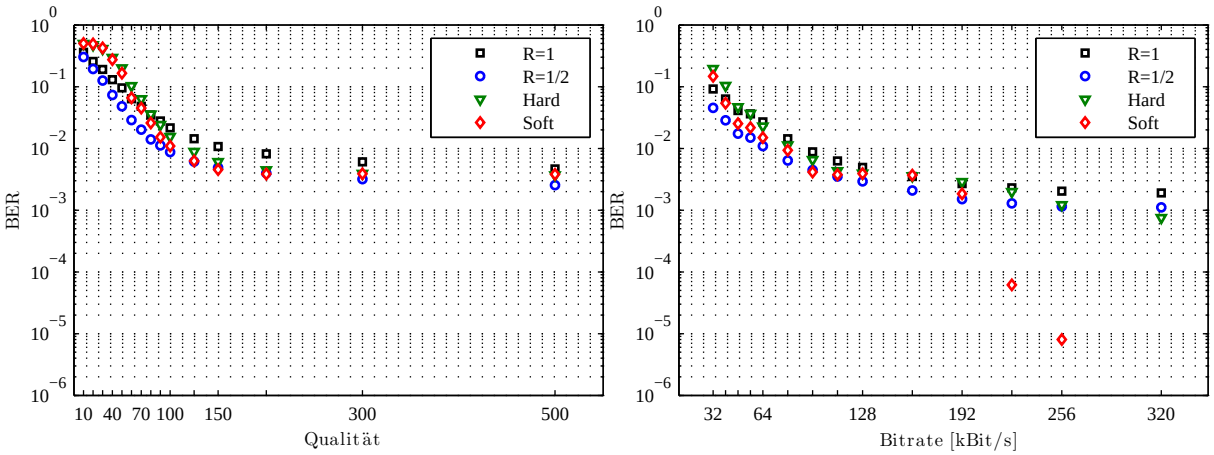


**Abbildung A.76.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmals bei Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

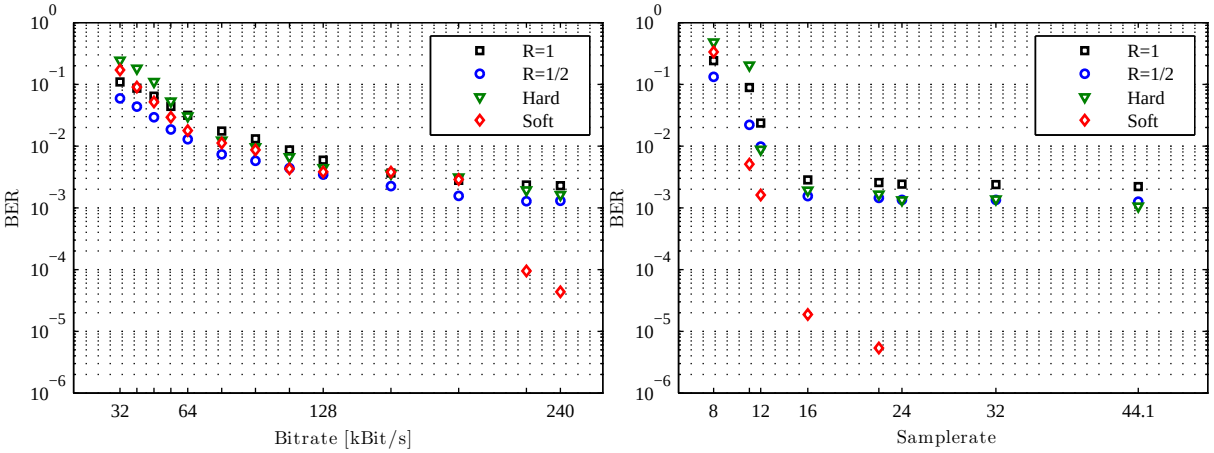


**Abbildung A.77.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmals bei Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

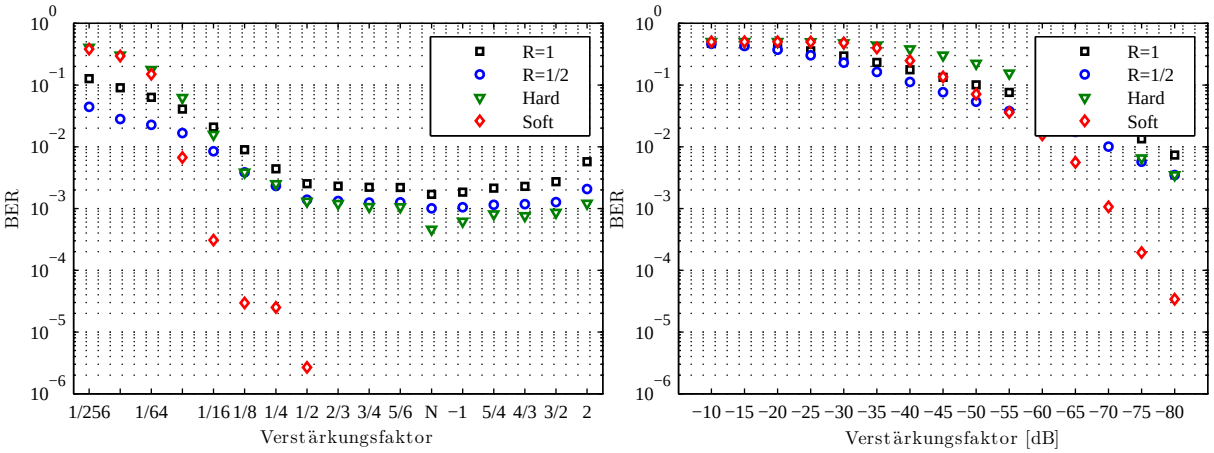


**Abbildung A.78.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmals bei Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

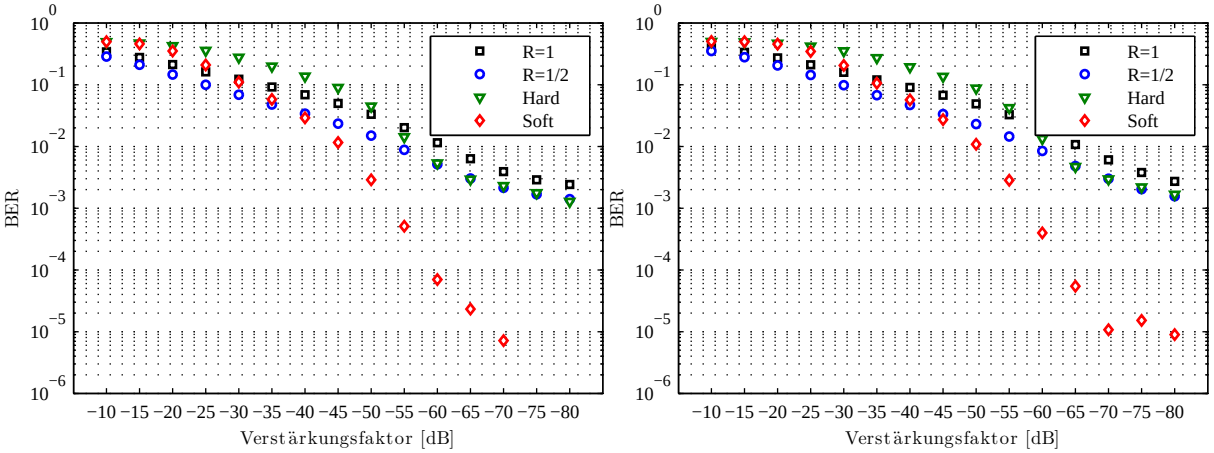


**Abbildung A.79.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmals bei Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

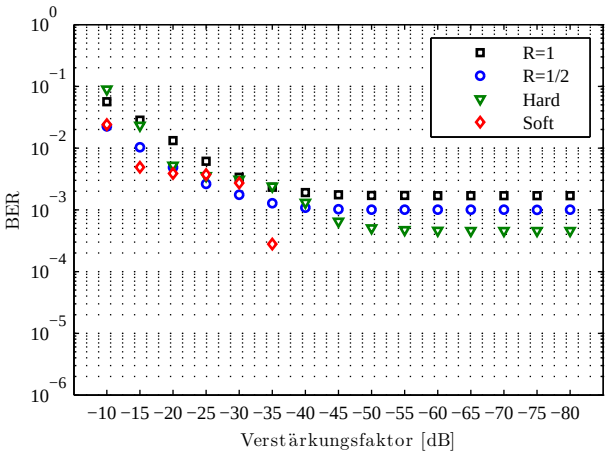


**Abbildung A.80.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmals bei Störung der Trägerdaten durch dynamisches Rauschen (links).

### A.6.4.3. Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.81 bis A.85 zeigen die Simulationsergebnisse für die Erweiterung des Grundsystems mit Merkmalsverstärkung um die *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F15$ bis $F20$ gebildet. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Merkmalsverstärkung erfolgt mit dem Schwellwert $Th_{tot}$=0,0001. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die maximale Bitfehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar, also den schlechtesten Wert für die Robustheit in Abhängigkeit der Testdaten. Als Vergleichsgrößen werden die maximalen Bitfehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt.

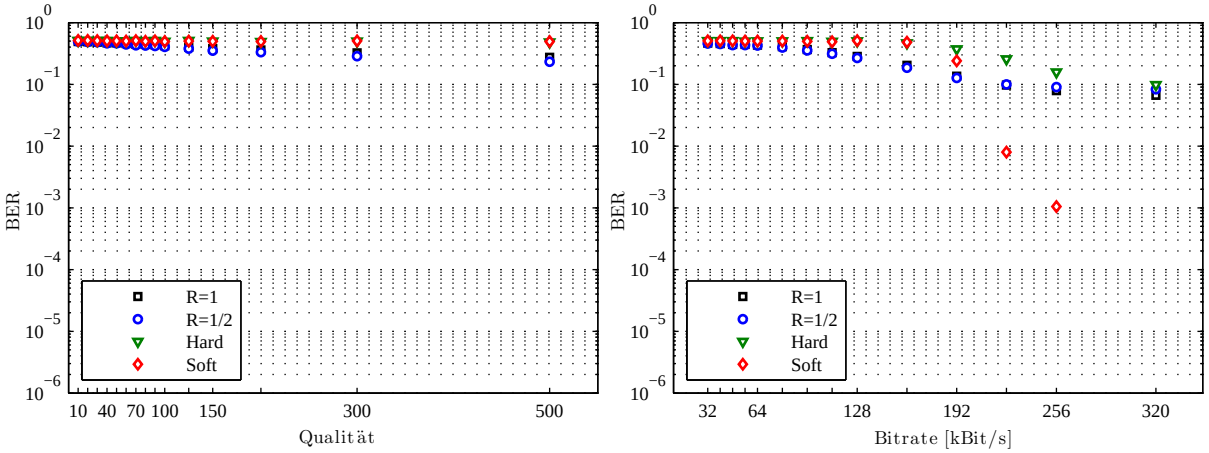


**Abbildung A.81.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

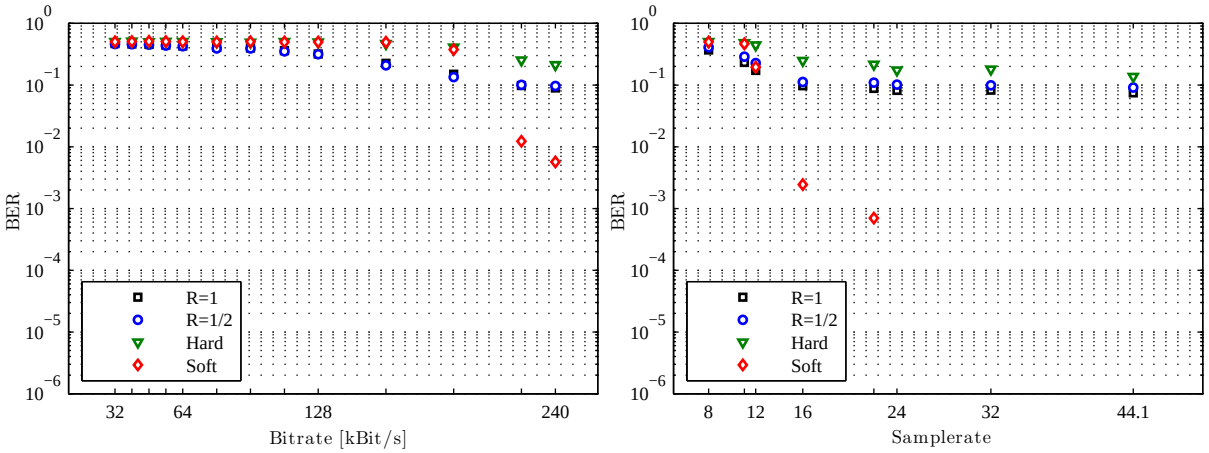


**Abbildung A.82.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

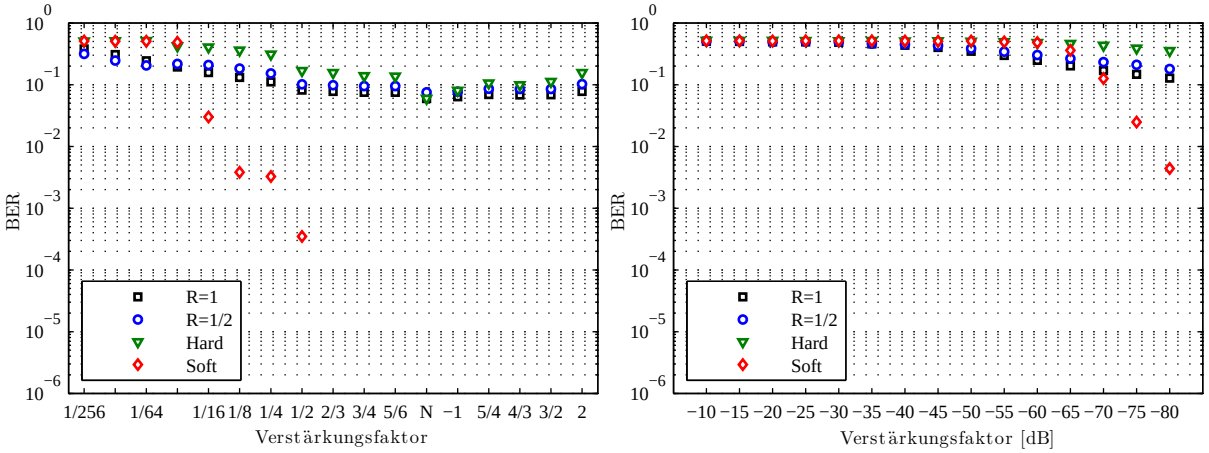


**Abbildung A.83.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

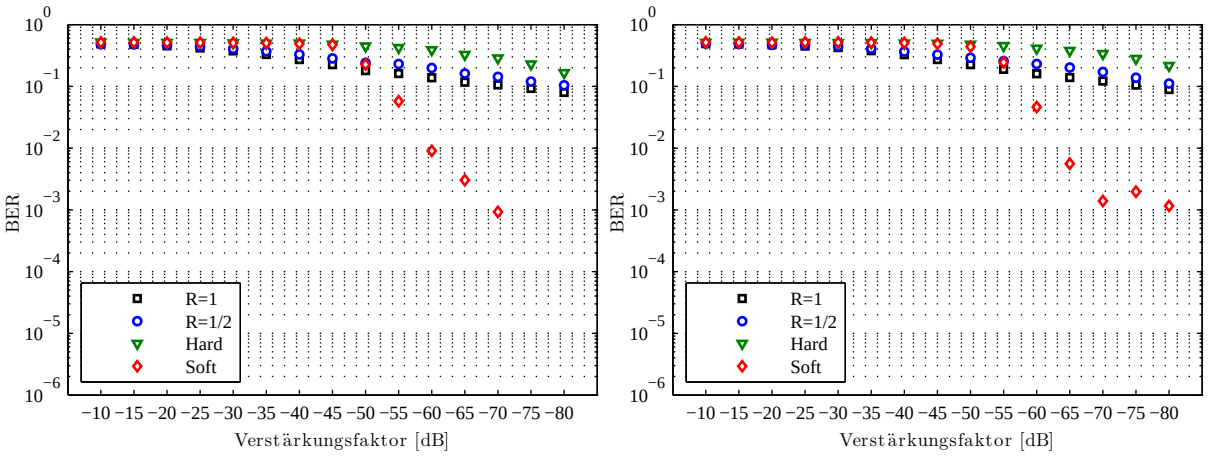


**Abbildung A.84.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

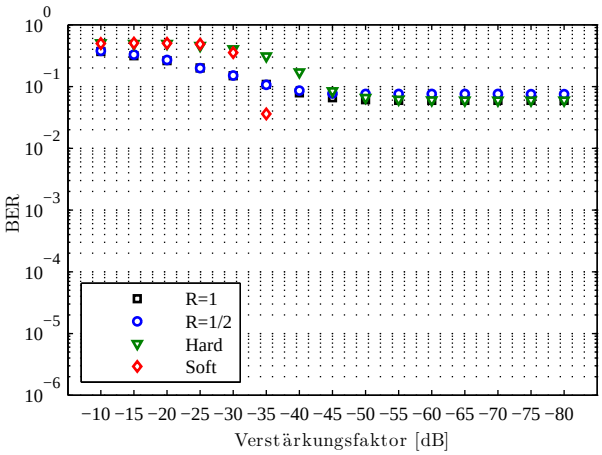


**Abbildung A.85.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens für das Grundsystem mit Totzone des Inhaltsmerkmales bei Störung der Trägerdaten durch dynamisches Rauschen (links).

# A.7. Leistungsanalyse der hierarchischen Wasserzeicheneinbettung

## A.7.1. Fehlerkorrektur für die hierarchische Einbettung

### A.7.1.1. Fehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.86 bis A.90 zeigen die Simulationsergebnisse für die Erweiterung der hierarchischen Einbettung um die *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem

Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F19$ bis $F20$ gebildet. Die hierarchische Einbettung erfolgt über zwei Ebenen. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die Fehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar, also den Anteil der Audiorahmen für den die verschlüsselte Nutzinformation **v** wenigsten einen Bitfehler enthält. Als Vergleichsgrößen werden die Fehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt. Weiterhin sind die Ergebnisse für das Grundsystem aus Anhang A.6.3.1 mit grauen Symbolen in die Abbildungen eingefügt.

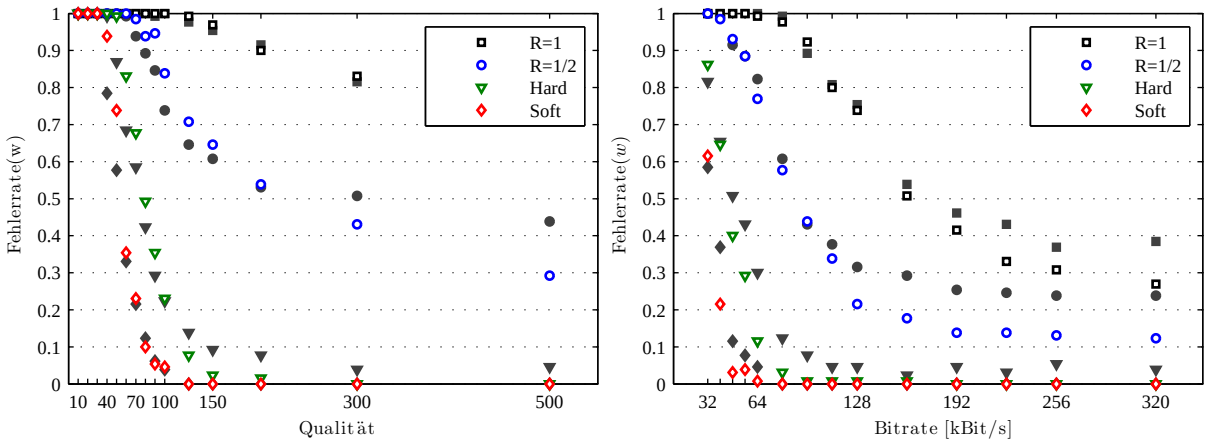


**Abbildung A.86.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

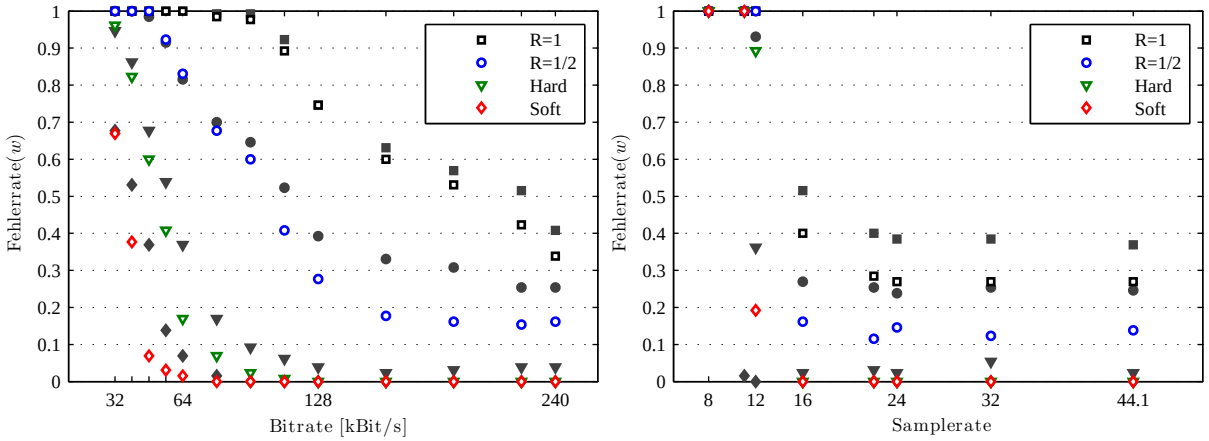


**Abbildung A.87.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

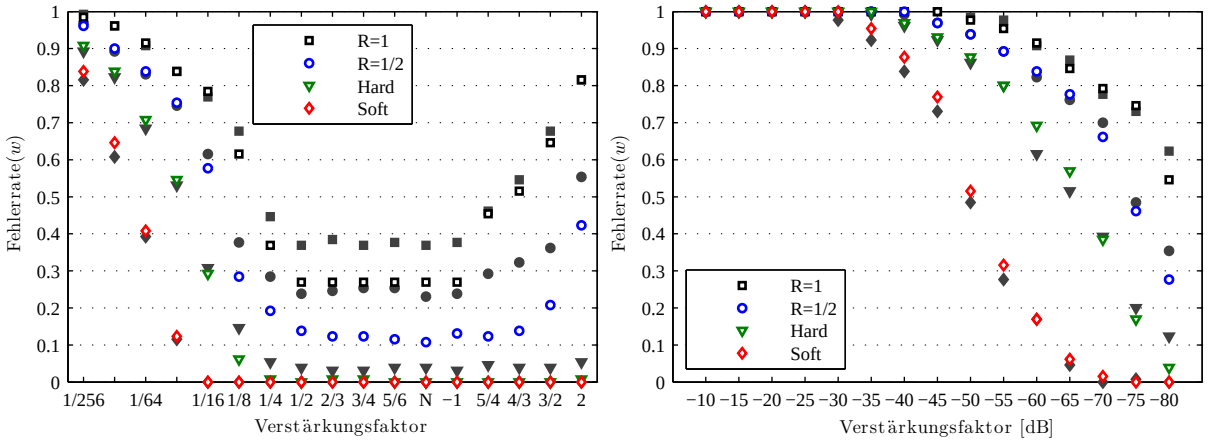


**Abbildung A.88.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

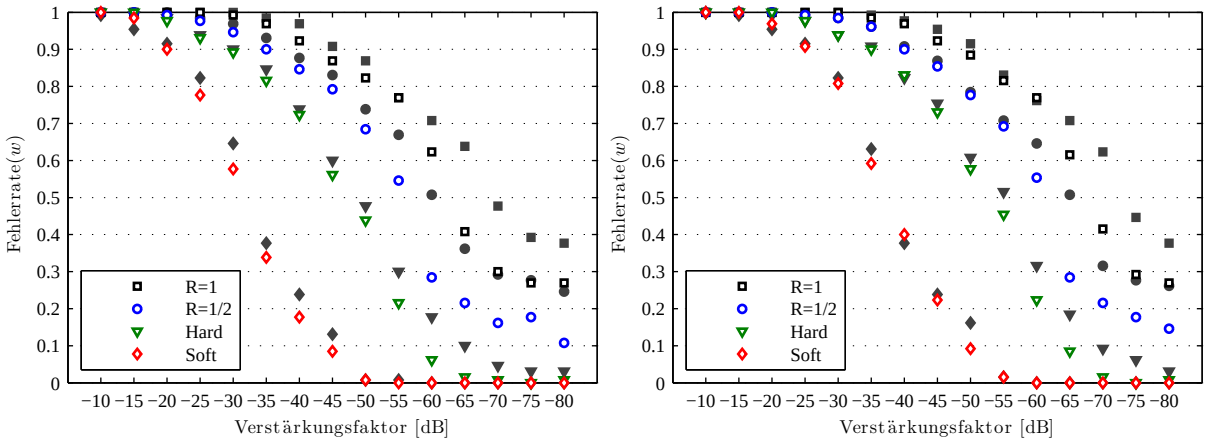


**Abbildung A.89.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

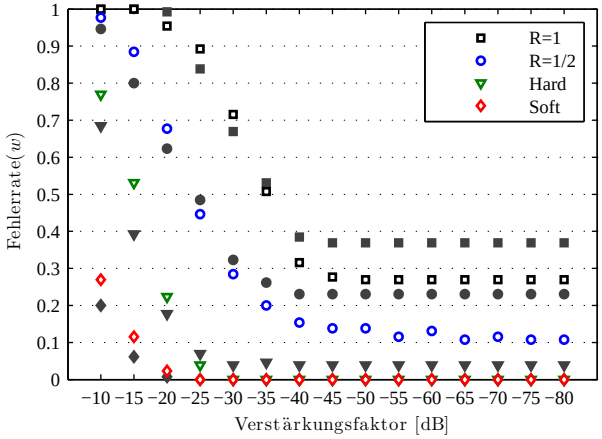


**Abbildung A.90.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch dynamisches Rauschen (links).

### A.7.1.2. Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.91 bis A.95 zeigen die Simulationsergebnisse für die Erweiterung der hierarchischen Einbettung um die *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F19$ bis $F20$ gebildet. Die hierarchische Einbettung erfolgt über zwei Ebenen. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar. Als Vergleichsgrößen werden die mittleren Bitfehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt. Weiterhin sind die Ergebnisse für das Grundsystem aus Anhang A.6.3.2 mit grauen Symbolen in die Abbildungen eingefügt.

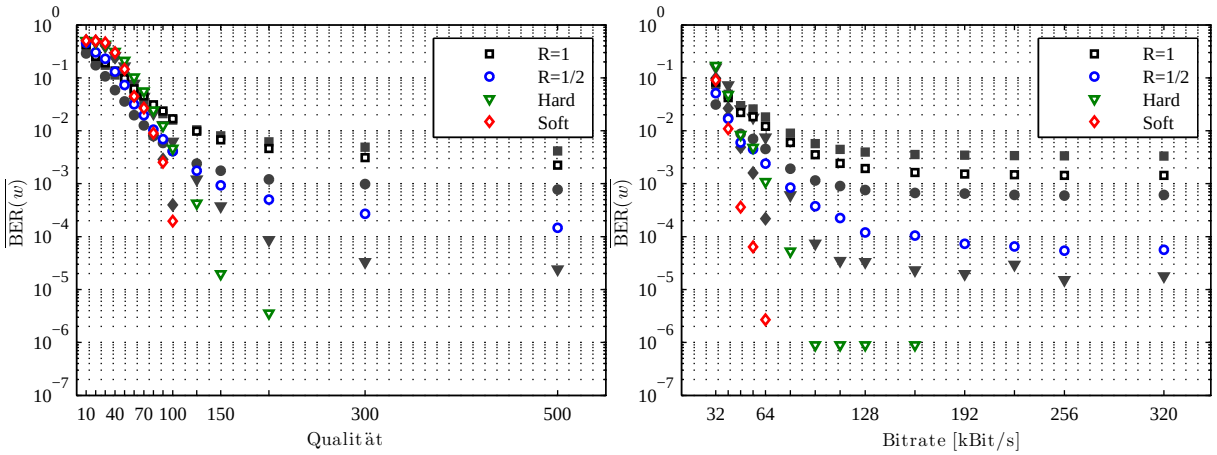


**Abbildung A.91.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

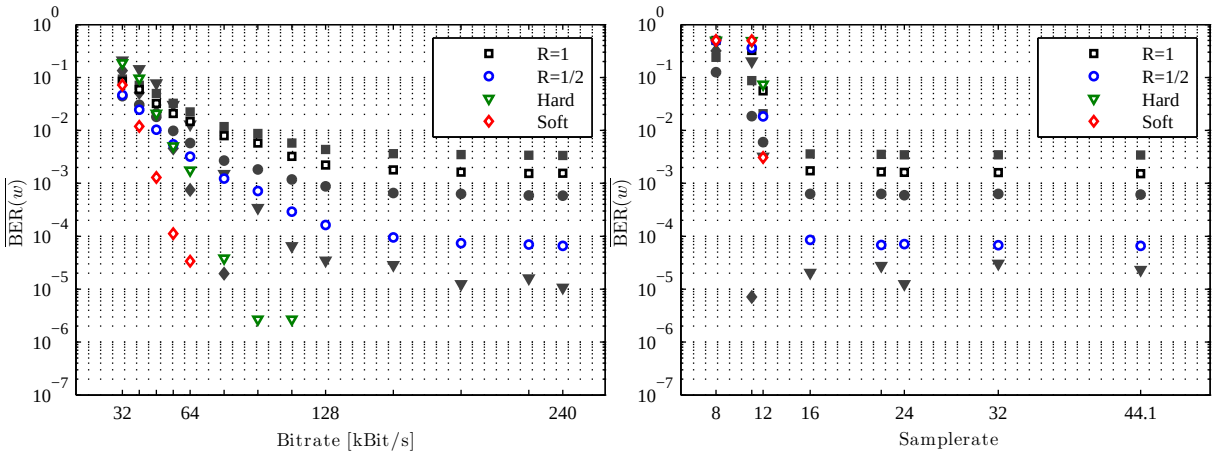


**Abbildung A.92.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

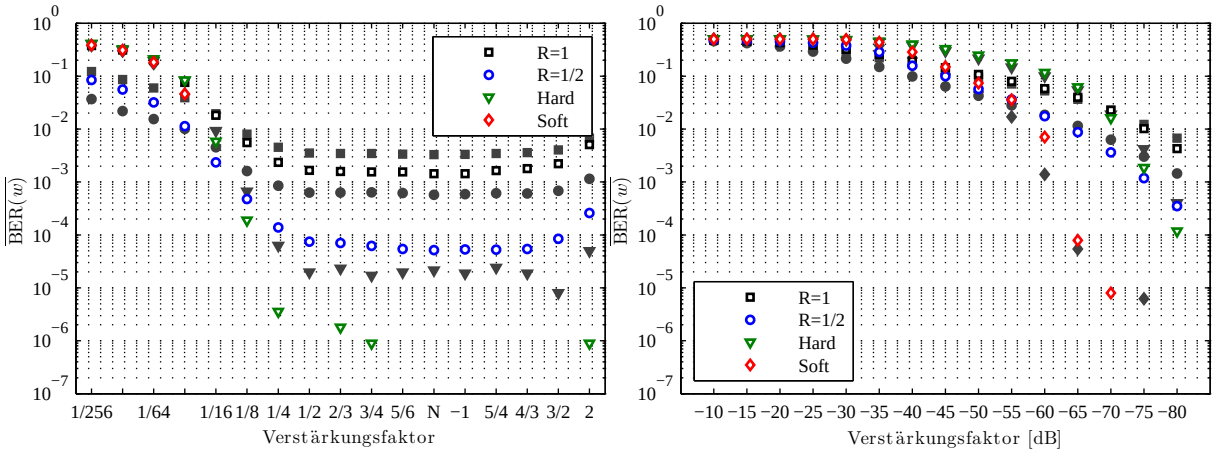


**Abbildung A.93.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

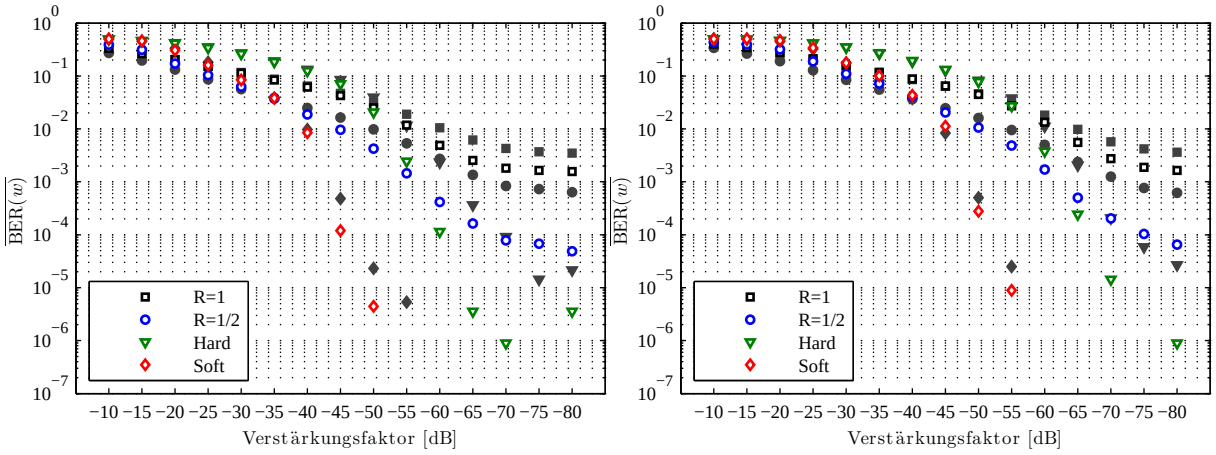


**Abbildung A.94.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

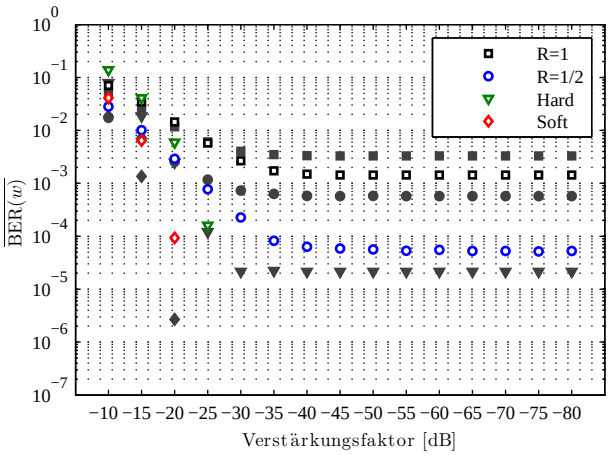


**Abbildung A.95.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch dynamisches Rauschen (links).

### A.7.1.3. Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.96 bis A.100 zeigen die Simulationsergebnisse für die Erweiterung der hierarchischen Einbettung um die *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F19$ bis $F20$ gebildet. Die hierarchische Einbettung erfolgt über zwei Ebenen. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die maximale Bitfehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar, also den schlechtesten Wert für die Robustheit in Abhängigkeit der Testdaten. Als Vergleichsgrößen werden die maximalen Bitfehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt. Weiterhin sind die Ergebnisse für das Grundsystem aus Anhang A.6.3.3 mit grauen Symbolen in die Abbildungen eingefügt.

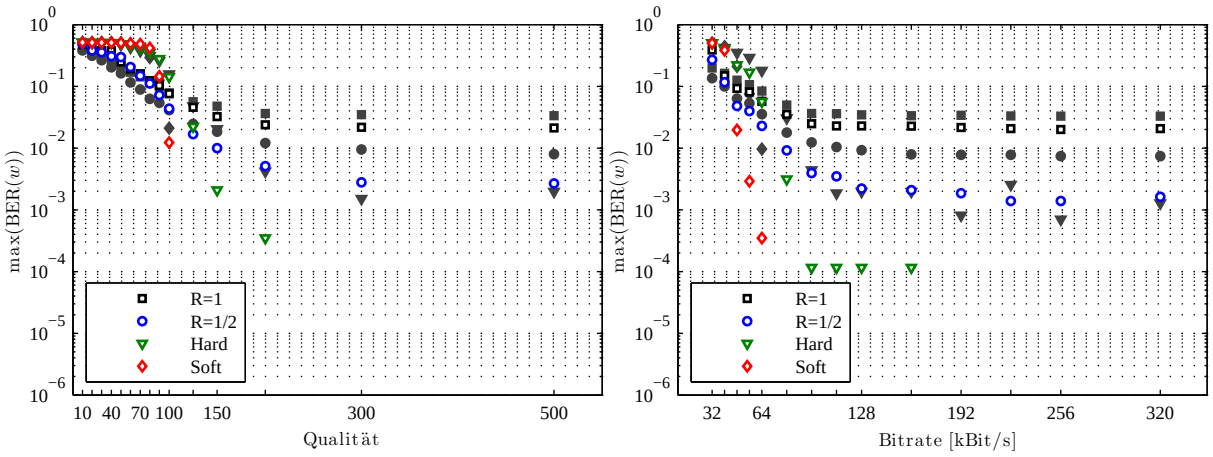


**Abbildung A.96.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

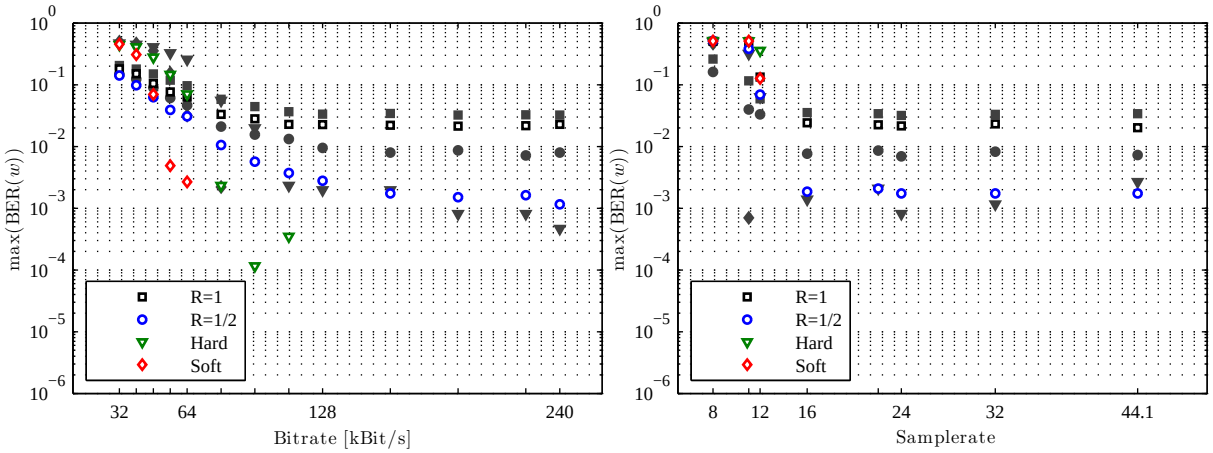


**Abbildung A.97.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

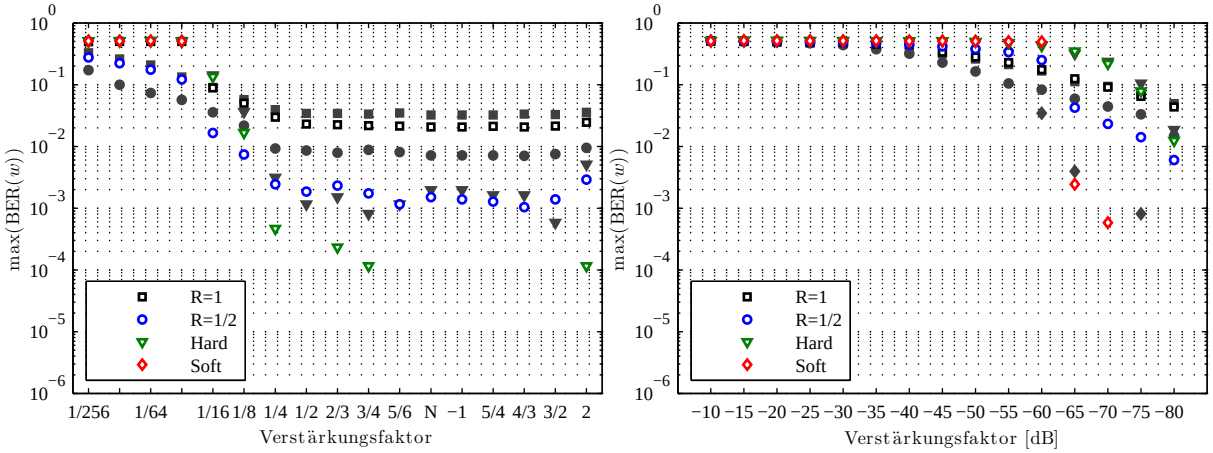


**Abbildung A.98.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

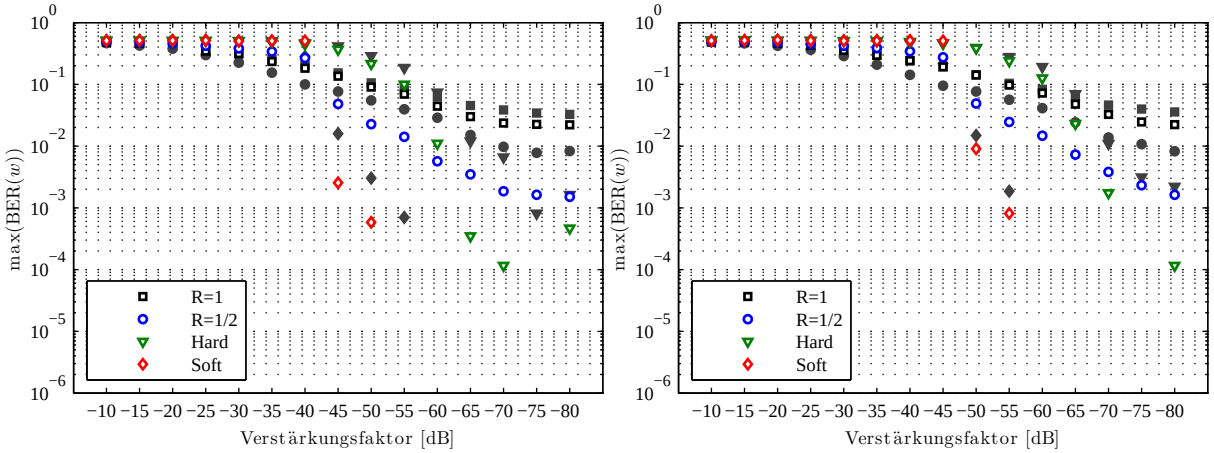


**Abbildung A.99.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

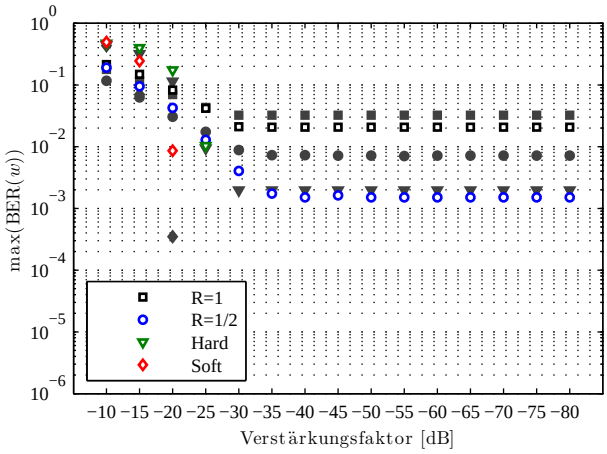


**Abbildung A.100.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung und Störung der Trägerdaten durch dynamisches Rauschen (links).

## A.7.2. Fehlerkorrektur für die hierarchische Einbettung mit Totzone

### A.7.2.1. Fehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.101 bis A.105 zeigen die Simulationsergebnisse für die hierarchische Einbettung mit Merkmalsverstärkung und *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch

die Frequenzgruppen $F19$ bis $F20$ gebildet. Die hierarchische Einbettung erfolgt über zwei Ebenen. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Merkmalsverstärkung erfolgt mit dem Schwellwert $Th_{tot}$=0,0001. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die Fehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar, also den Anteil der Audiorahmen für den die verschlüsselte Nutzinformation **v** wenigsten einen Bitfehler enthält. Als Vergleichsgrößen werden die Fehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt. Weiterhin sind die Ergebnisse für das Grundsystem mit Merkmalsverstärkung aus Anhang A.6.4.1 mit grauen Symbolen in die Abbildungen eingefügt.

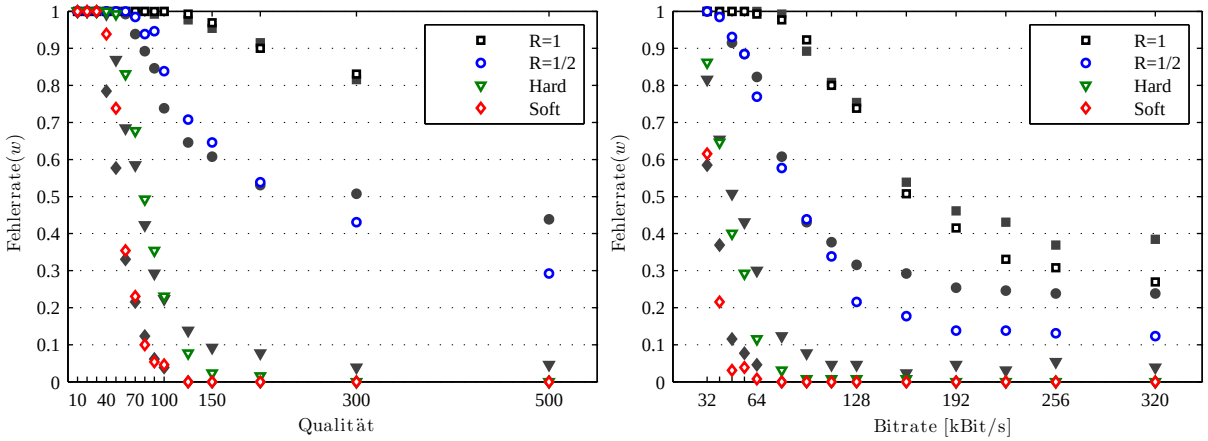


**Abbildung A.101.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

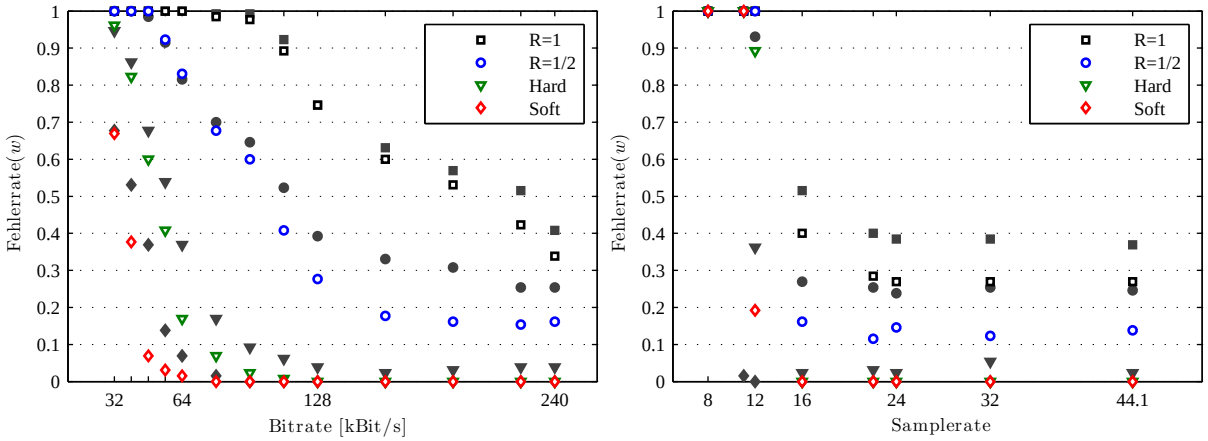


**Abbildung A.102.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

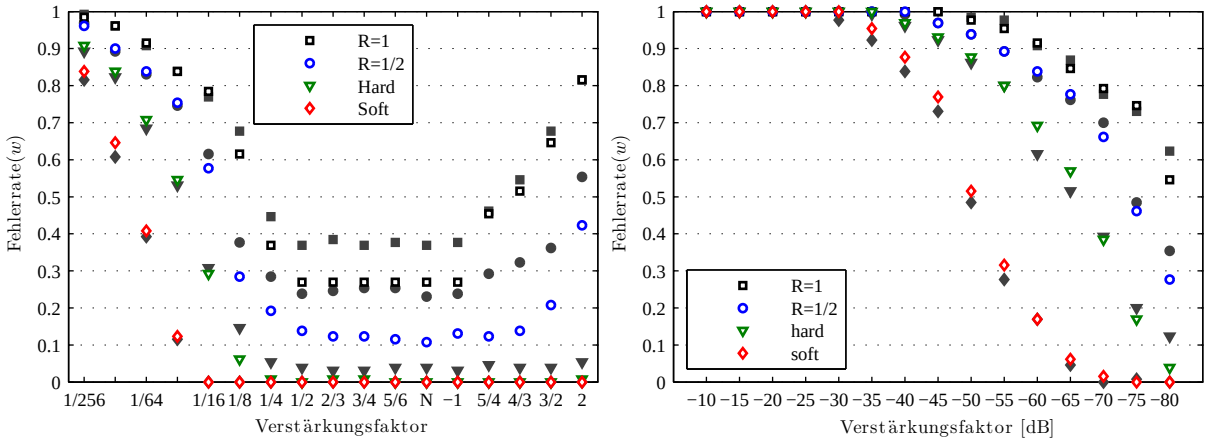


**Abbildung A.103.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

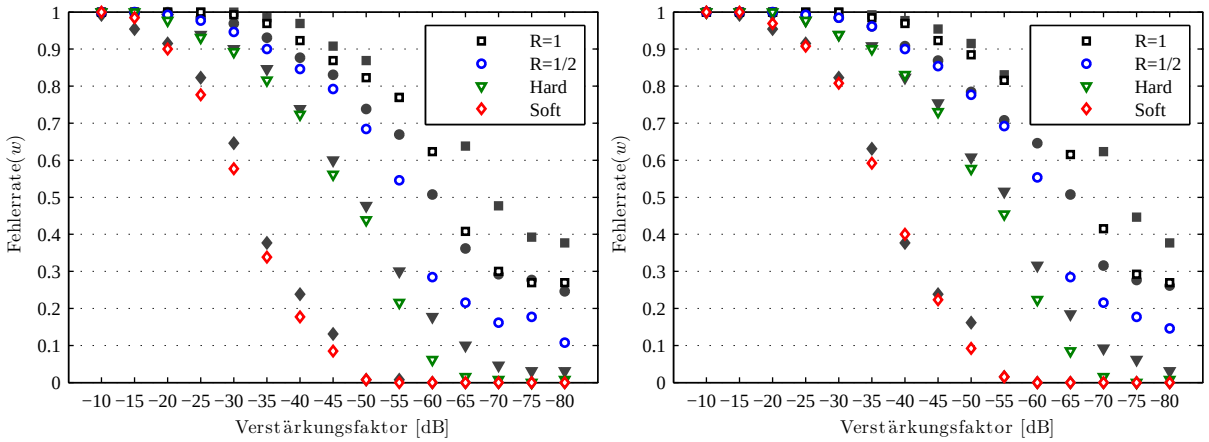


**Abbildung A.104.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

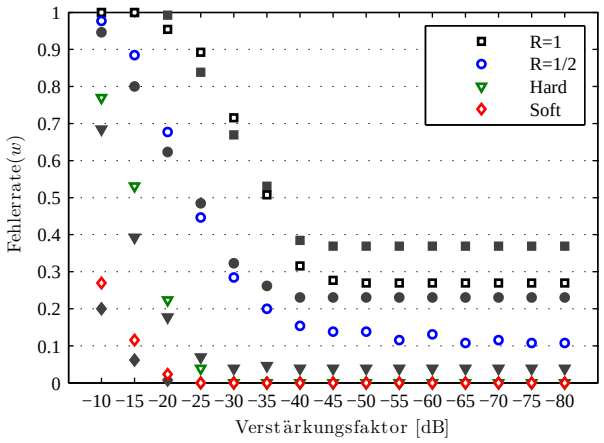


**Abbildung A.105.:** Fehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch dynamisches Rauschen (links).

### A.7.2.2. Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.106 bis A.110 zeigen die Simulationsergebnisse für die hierarchische Einbettung mit Merkmalsverstärkung und *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F19$ bis $F20$ gebildet. Die hierarchische Einbettung erfolgt über zwei Ebenen. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Merkmalsverstärkung erfolgt mit dem Schwellwert $Th_{tot}$=0,0001. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar. Als Vergleichsgrößen werden die mittleren Bitfehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt. Weiterhin sind die Ergebnisse für das Grundsystem mit Merkmalsverstärkung aus Anhang A.6.4.2 mit grauen Symbolen in die Abbildungen eingefügt.

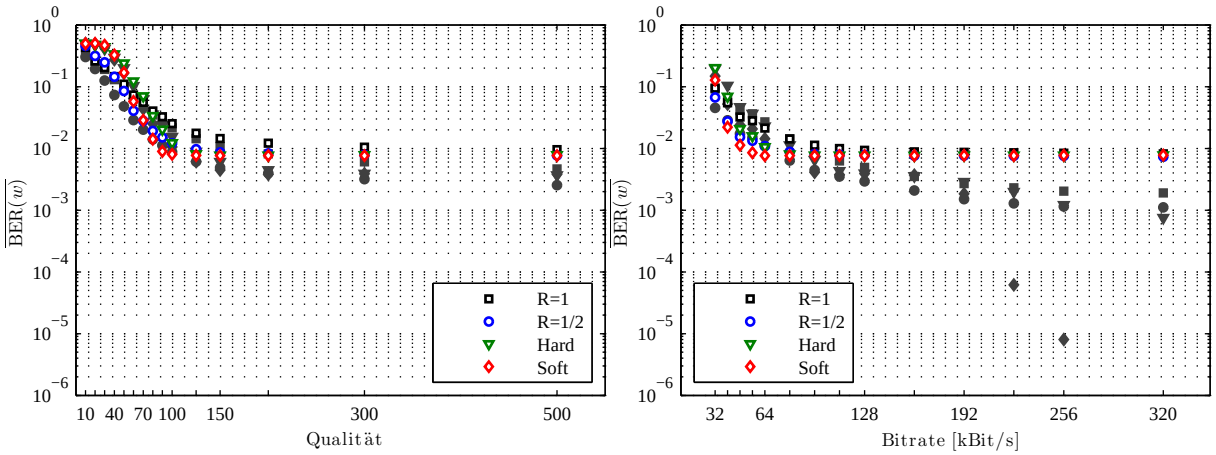


**Abbildung A.106.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

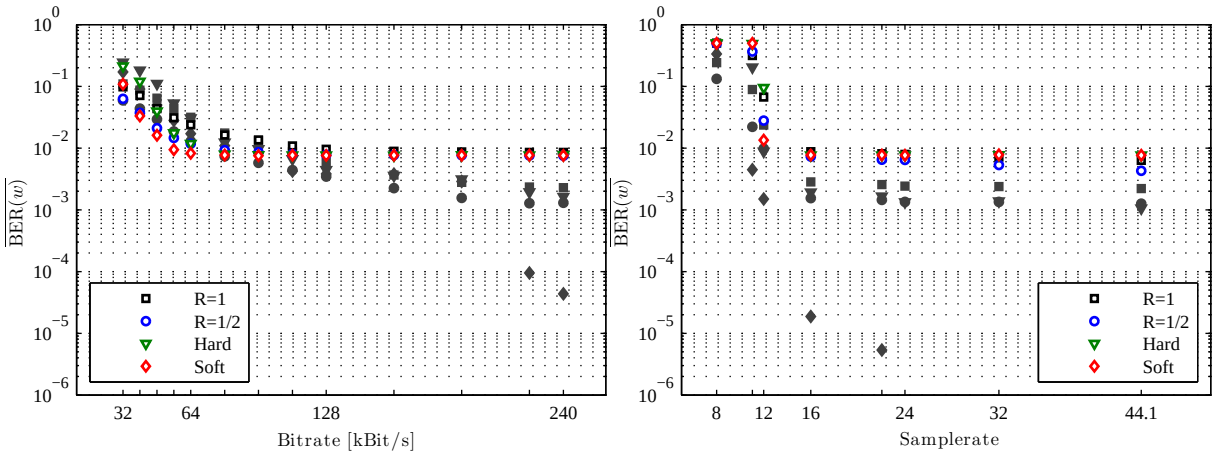


**Abbildung A.107.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

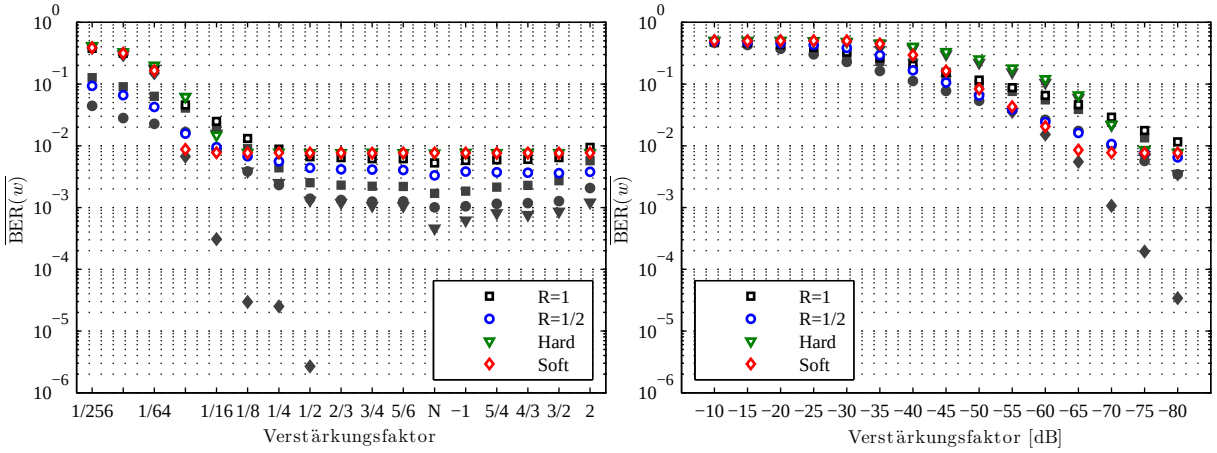


**Abbildung A.108.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

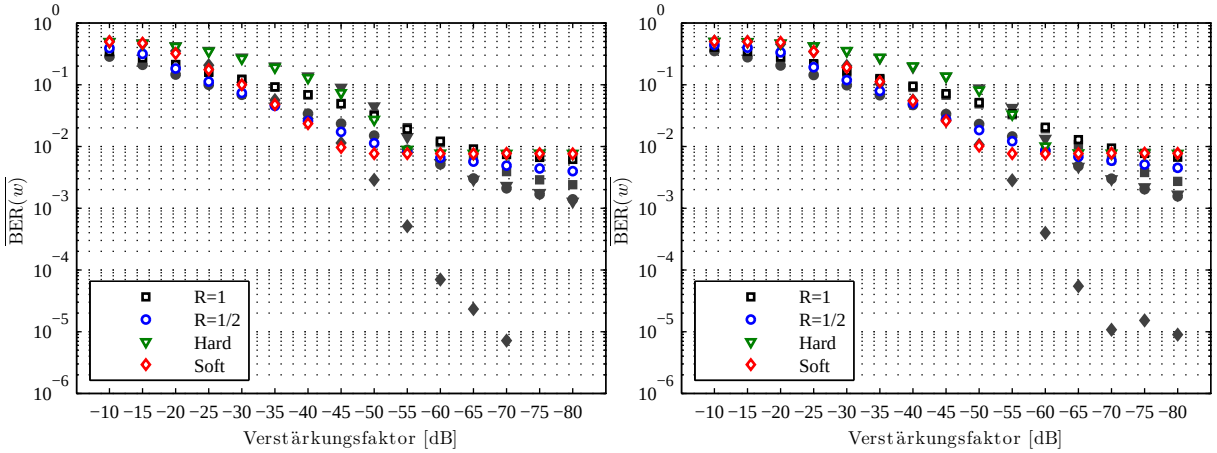


**Abbildung A.109.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

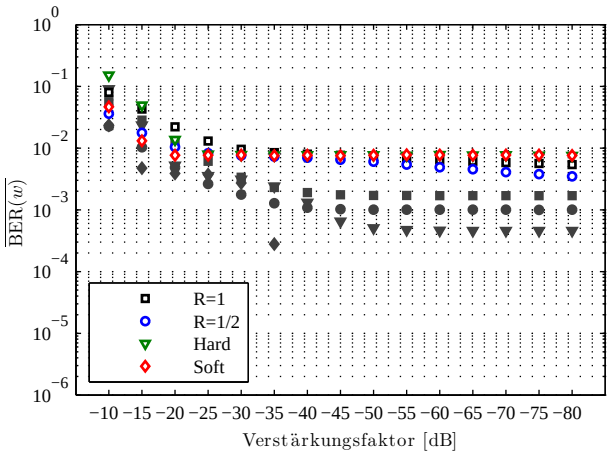


**Abbildung A.110.:** Mittlere Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch dynamisches Rauschen (links).

### A.7.2.3. Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens

Die Abbildungen A.111 bis A.115 zeigen die Simulationsergebnisse für die hierarchische Einbettung mit Merkmalsverstärkung und *Soft-Input*-Decodierung hinsichtlich der Robustheit der Wasserzeicheninformation gegenüber den in Abschnitt 4.4.1.3 festgelegten Störungen mit erweitertem Parametersatz. Für die Analyse wurde eine Segmentlänge von $l_M = 8\,192$ für die Extraktion der Inhaltsmerkmale und eine Segmentlänge von $l_W = 1\,024$ für die Wasserzeicheneinbettung verwendet. Der Extraktionsbereich für die Inhaltsmerkmale wird durch die Frequenzgruppen $F2$ bis $F23$ und der Einbettungsbereich für die Wasserzeicheninformation durch die Frequenzgruppen $F19$ bis $F20$ gebildet. Die hierarchische Einbettung erfolgt über zwei Ebenen. Die Transparenz der Wasserzeicheneinbettung wird auf ein ODG von -1 eingestellt. Die Merkmalsverstärkung erfolgt mit dem Schwellwert $Th_{tot}$=0,0001. Die Abbildungen stellen in Bezug auf die Intensität der Störungen die maximale Bitfehlerrate der Nutzinformation eines Audiorahmens („Soft“) dar, also den schlechtesten Wert für die Robustheit in Abhängigkeit der Testdaten. Als Vergleichsgrößen werden die maximalen Bitfehlerraten für die Wasserzeicheninformation („R=1“), die Wasserzeicheninformation mit halber Kapazität („R=1/2“) und der Nutzinformation bei einer *Hard-Input*-Decodierung („Hard“) dargestellt. Weiterhin sind die Ergebnisse für das Grundsystem mit Merkmalsverstärkung aus Anhang A.6.4.3 mit grauen Symbolen in die Abbildungen eingefügt.

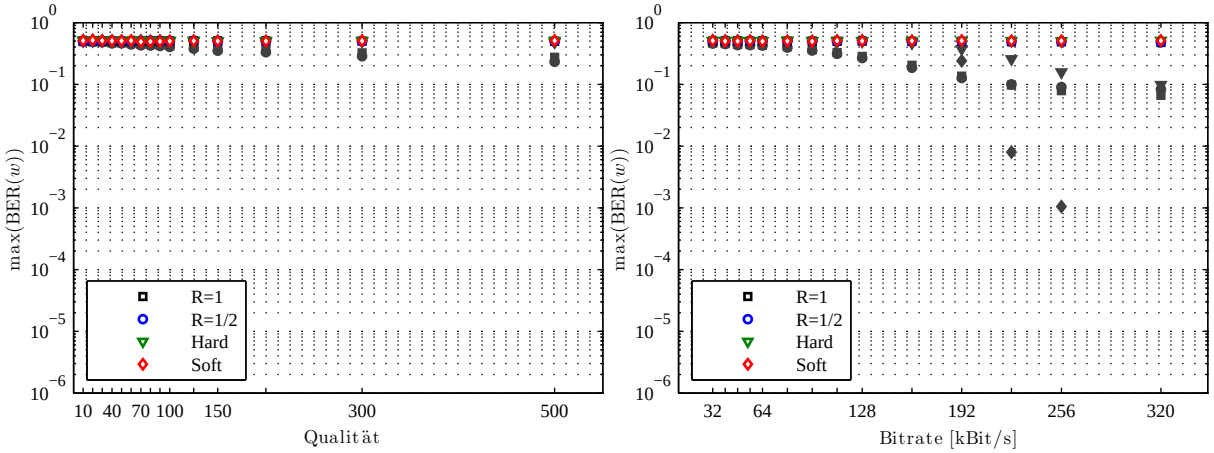


**Abbildung A.111.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch AAC-Kompression (links) und MP3-Kompression (rechts).

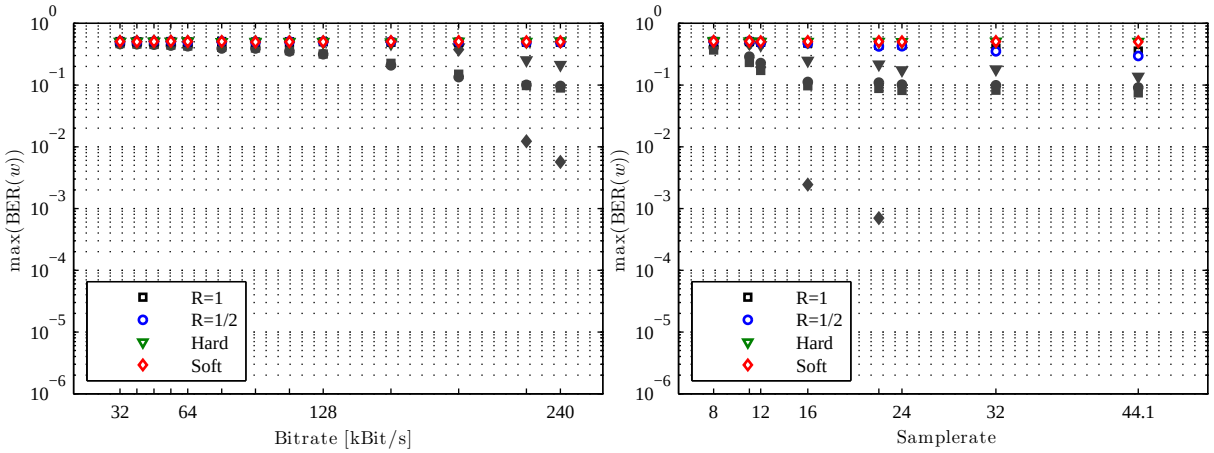


**Abbildung A.112.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch Vorbis-Kompression (links) und Unterabtastung (rechts).

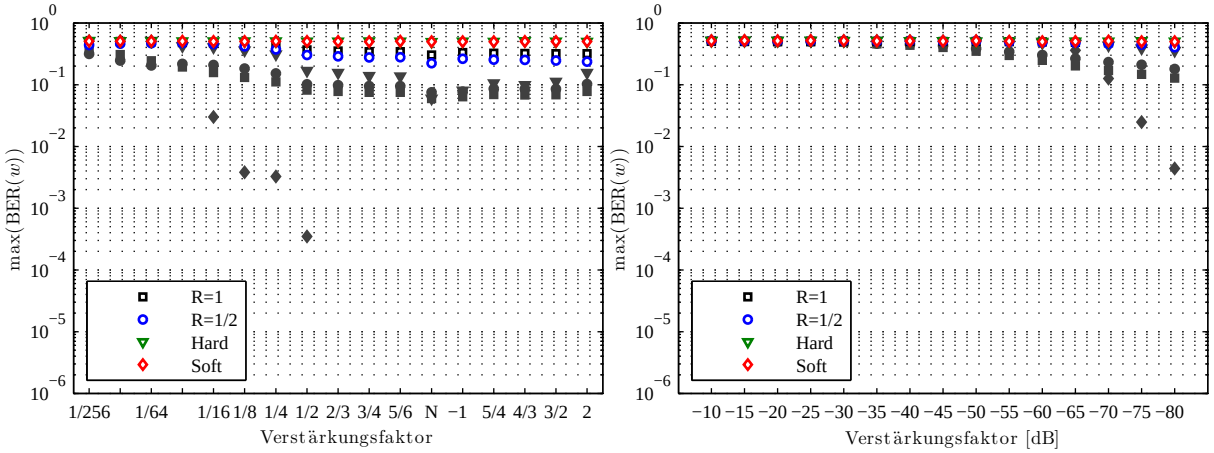


**Abbildung A.113.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch Laustärkeveränderung (links) und Weißes Rauschen (rechts).

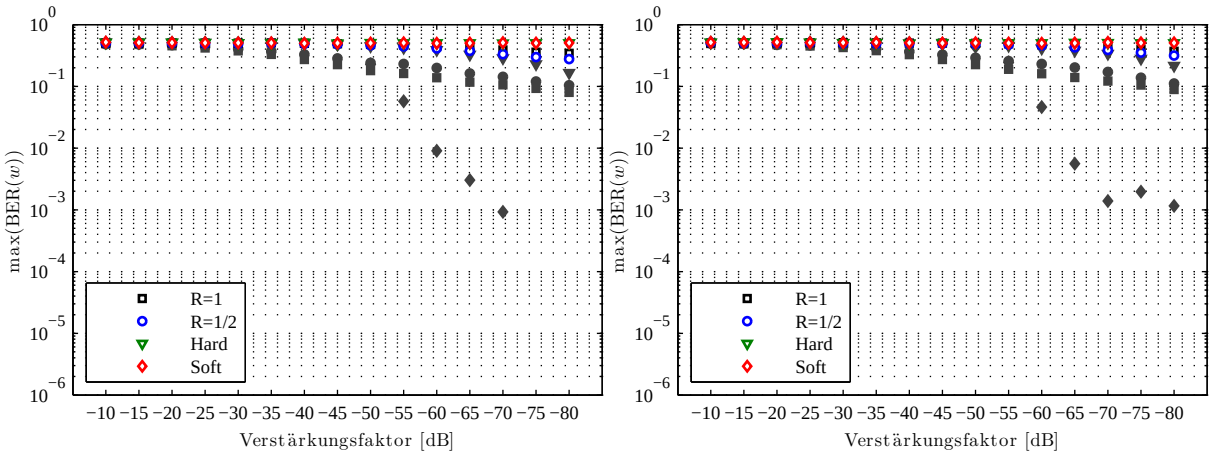


**Abbildung A.114.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch $1/f^2$-Rauschen (links) und $1/f$-Rauschen (rechts).

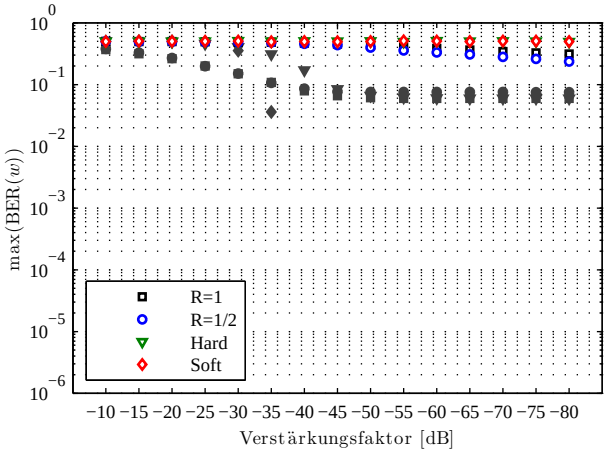


**Abbildung A.115.:** Maximale Bitfehlerrate der Nutzinformation eines Audiorahmens bei hierarchischer Einbettung mit Totzone und Störung der Trägerdaten durch dynamisches Rauschen (links).

# Lebenslauf

**Persönliche Daten**

Dipl.-Ing. Michael Gulbis
Geb. am 30. September 1980 in Rostock

**Beruflicher Werdegang**

| | |
|---|---|
| seit Sep. 2012 | Modulverantwortlicher „Einspeisung“<br>von Strom nach EEG und KWK-Gesetz |
| Aug. 2011 - Aug. 2012 | Softwareentwickler<br>Bereich Softwareentwicklung / Billing / Verbrauchsabrechnung, SIV.AG, Roggentin<br>○ ERP-System kVASy® |
| Okt. 2008 - Jul. 2011 | wissenschaftlicher Projektmitarbeiter, Institut für Nachrichtentechnik, Universität Rostock<br>○ Mobile Assistenzsysteme für Routeninformationen und Krankenakten<br>○ Softwareoptimierung des HHi. H.264 Baseline Videodecoders |
| Dez. 2005 - Sep. 2008 | Stipendiat im Graduiertenkolleg der Deutschen Forschungsgemeinschaft (DFG): „Verarbeitung, Verwaltung, Darstellung und Transfer multimedialer Daten - technische Grundlagen und gesellschaftliche Implikationen“<br>○ Entwicklung effizienter Watermarking-Verfahren zum inhaltsbasierten Integritätsschutz von Sprachdaten |

**Hochschulstudium**

| | |
|---|---|
| 28. Nov. 2005 | Diplom-Ingenieur, Informationstechnik, Note: „sehr gut“<br>○ Konzeptentwicklung und Implementierung eines Audiowasserzeichens zum Integritätsschutz von Audiodaten, Note: „gut“ |
| Okt. 2004 - Mär. 2005 | Praktikum am Kompetenzzentrum für Mediensicherheit (MERIT), Fraunhofer-Institut für Integrierte Publikations- und Informationssysteme (IPSI), Darmstadt<br>○ Forschung und Entwicklung von Integritäts-Wasserzeichen für Videodaten |
| 12. Jan. 2004 | Bachelor of Science, Informationstechnik, Note: „gut“<br>○ Effiziente Algorithmen zur digitalen Spektralanalyse, Note: „sehr gut“ |
| Okt. 2000 - Nov. 2005 | Studium der Informationstechnik, Universität Rostock |

**Wehrdienst**

| | |
|---|---|
| Nov. 1999 - Aug. 2000 | Grundwehrdienst in Eutin und Rendsburg, Fernmelder (HG) |

**Schulbildung**

| | |
|---|---|
| 10. Jul. 1999 | Abitur am Fritz-Reuter-Gymnasium, Kühlungsborn, Note: 1,6 |

# Erklärung

Ich erkläre, dass ich die eingereichte Dissertation selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Rostock, 8. März 2013 Michael Gulbis