



DROCON –DROpout CONcealment for wireless digital audio transmission

Projektleiter: Robert Höldrich (hoeldrich@iem.at)

Mitarbeiterin: Cornelia Falch (falch@iem.at)

Gefördert vom BMVIT im Rahmen der Forschungsinitiative FIT-IT Embedded Systems.

Kooperationspartner: AKG – Acoustics GmbH

1 Zusammenfassung

DROCON ist eine Projektkooperation von AKG Acoustics Wien und dem Institut für Elektronische Musik und Akustik der Kunstuniversität Graz. Im Bereich der drahtlosen Übertragung von Musik- und Sprachsignalen zielt DROCON darauf ab, kurze Signalausfälle zu verschleiern. Dabei verfolgt das Projekt das konkrete Ziel, die Ausfallsicherheit von Übertragungstrecken von Audiosignalen zu gewährleisten. Es soll als eigenständige Signalverarbeitungseinheit ("embedded system") in ein System eingegliedert werden, das hauptsächlich aus folgenden Teilen besteht, siehe auch Abb. 1:

Signalaufnahme – Kodiereinheit (Sender) – (digitale) Übertragungstrecke – Dekodiereinheit (Empfänger) – **Ausfallsicherheitseinheit** – *"Verfügbarkeit des Signals"*. Abb. 2 zeigt einen kurzen Ausschnitt eines fehlerbehafteten Signals, wie es auf Empfängerseite vor (blau durchgezogen) bzw. nach (rot strichliert) der Ausfallsicherheitseinheit verfügbar ist.

Da das System für den Echtzeitbetrieb gedacht ist, liegt die große Herausforderung darin, die gesamte Verzögerungszeit möglichst gering zu halten. Die Ausfallsicherheitseinheit sollte daher keine negativen Auswirkungen auf die Systemlatenz haben, weshalb für die Signalverarbeitung Techniken mit sehr kurzen Latenzzeiten gefordert werden.

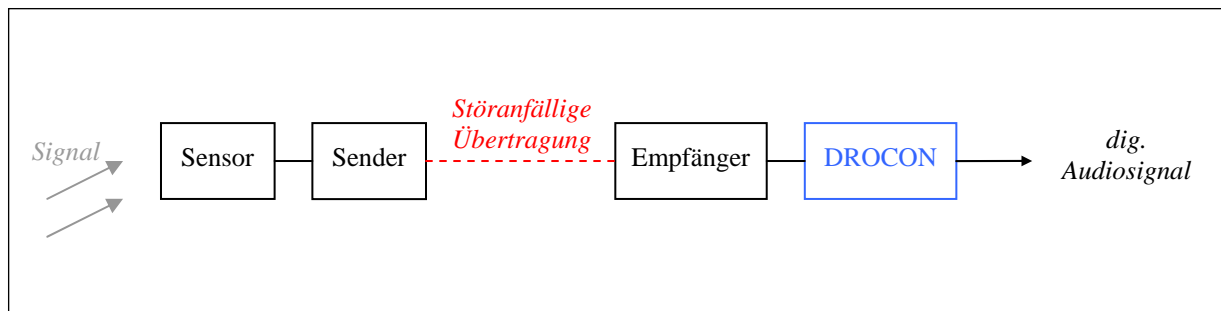


Abb.1: Gesamtes Übertragungssystem

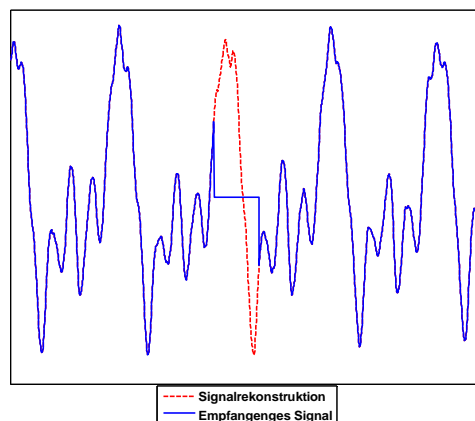


Abb.2: Zeitverlauf eines mit einem Ausfall behafteten Signalausschnitts und Rekonstruktion der fehlenden Samples

DROCON verfolgt die innovative Strategie, eine Reihe unterschiedlicher Signalverarbeitungs-Verfahren parallel zu implementieren. Mithilfe einer geeigneten Heuristik soll zur Laufzeit die optimale Methode der Signalfortführung gewählt werden. Dabei spielen die Art der Signale, die Länge der Ausfälle sowie deren statistische Häufigkeitsverteilung eine wesentliche Rolle.

2 Methoden im Überblick

Die Verminderung der Signalqualität durch Störungen oder Ausfälle ist in der Audiotechnik inakzeptabel und muss durch geeignete Methoden verhindert werden. Zu diesem Zweck werden verschiedenartige Modelle zur Fortsetzung zeitdiskreter Signale aus unterschiedlichen Bereichen der Signalverarbeitung untersucht und entsprechend der konkreten Anforderungen spezifiziert. Wesentliche Merkmale der sogenannten Fehlerverschleierung in der digitalen Audioübertragung stellen neben der Erhaltung der Signalqualität unter anderem die Echtzeitfähigkeit der

Verarbeitungseinheit, die Schnittstelle zu einer vorangehenden Stufe (z.B. Signal-Dekodiereinheit), die Länge der Signalausfälle sowie deren statistische Häufigkeitsverteilung dar.

Ein derzeit aktuelles Themengebiet beschäftigt sich mit der Restauration fehlerhafter Signale. Es soll eruiert werden, inwieweit die darin angewandten Methoden entsprechend der im vorliegenden Projekt geforderten Kriterien adaptiert werden können. Beispielsweise spielen bei der Digitalisierung von altem, analogen Audiomaterial Verfahren zur Datenwiederherstellung eine wesentliche Rolle, weshalb deren Entwicklung bzw. Verbesserung nach wie vor ein aktuelles Forschungsgebiet darstellt. Im Allgemeinen geschieht Signalkonvertierung und -aufbereitung nicht in Echtzeit, somit besteht nicht die Notwendigkeit, strenge Restriktionen bezüglich der Verarbeitungsgeschwindigkeit zu setzen.

Ähnliches gilt für einen weiteren, interessanten und ebenfalls aktuellen Forschungsschwerpunkt – der Audioübertragung in Netzwerken (z.B. "Internet-Streaming"). Die Echtzeitfähigkeit solcher Systeme wird durch diverse Übertragungsprotokolle, Buffer, etc. beeinflusst, weshalb die Grenzen dabei in jedem Fall weitaus weniger kritisch sind als in der vorliegenden Arbeit.

Aufgrund einer ausführlichen Literaturrecherche und unter Berücksichtigung oben genannter Einschränkungen stellen Extrapolations- und Prädiktionsalgorithmen potentielle Verfahren zur Überbrückung von Signalausfällen dar. Die Herausforderung bei diesen Methoden liegt darin, neue Signalsamples ausschließlich mithilfe von Information aus der Signalvergangenheit zu präzisieren. Ein adäquates Verfahren muss in der Lage sein, am Beginn einer Signalunterbrechung ohne zeitliche Verzögerung dieses fortzuführen, das heißt, die korrekte Gewichtung und Kombination der letzten Abtastpunkte muss jederzeit sofort verfügbar sein. Die Anzahl der erforderlichen Samples, welche als Beobachtungsfenster bzw. Beobachtungsintervall bezeichnet werden, ist von essentieller Bedeutung. Einerseits muss das Beobachtungsfenster groß genug gewählt werden, um ausreichend Information über die Zusammensetzung des Signals aufzuzeigen. Da Musik und Sprache allerdings hochgradig nichtstationär sind und innerhalb eines Fensters keine weitere zeitliche Auflösung der spektralen Anteile möglich ist, ergibt sich automatisch eine maximale Länge der Intervalle. Eine grobe Abschätzung des Parameters findet in der ersten Projektphase statt.

Das Prinzip der Signalextrapolation basiert auf den Grundlagen der linearen Algebra und kann mathematisch mittels einer Übertragungsmatrix A beschrieben werden. In allgemeiner Form formuliert die Matrix eine mögliche Variante der Linearkombination, mit der aus einem nicht-bandbegrenzten Signal unbestimmter Länge (theoretisch $\rightarrow \infty$) ein Ausschnitt definierter Länge und spektralen Inhalts ermittelt wird. Die Extrapolation erfordert eine Matrixinversion, da hier die Formulierung wie folgt lautet: Ein bandbegrenztes Signal endlicher Länge (entspricht jenem des Beobachtungsintervalls) soll auf unbestimmte Länge extrapoliert werden. Da generell die Länge des Beobachtungsfensters nicht mit der Länge der Extrapolation übereinstimmt, ist A eine rechteckige Matrix, deren Inversion durch ein numerisches Näherungsverfahren gelöst werden muss. Eine gängige Möglichkeit stellt die Berechnung der Pseudoinversen mithilfe von "least-squares"-Lösungen (LS), wie z.B. der "minimum norm LS" (MNLS) oder der "weighted LS" (WLS) Methode, dar. Zur Verbesserung der Stabilität dieser Verfahren können zusätzliche Regularisierungsmatrizen eingesetzt werden. Andererseits verspricht die Verwendung der quadratischen Norm (Energieabschätzung) – z.B. als "mean square extrapolation filter" – gegenüber MNLS und WLS verbesserte Eigenschaften. Ein alternativer Ansatz bedient sich der Singulärwertenerweiterung der nun quadratischen Matrizen $A^T A$ bzw. AA^T und führt zur Bestimmung der sogenannten "discrete prolate spheroidal sequences" – einer Menge von orthogonalen Lösungsfunktionen. Ein Teilbereich des ersten Projektabschnitts widmet sich der praktischen Evaluierung dieser theoretischen Ansätze, welche in Matlab implementiert werden.

In der zeitkontinuierlichen Domäne können beliebige Signale ohne notwendige Beschränkungen exakt und analytisch über die Taylorreihenentwicklung berechnet werden. Dies gilt nicht für den diskreten Fall, weshalb A als Beschränkung des Zeit- und Frequenzbereichs interpretiert werden kann. Die Kriterien für die zeitliche Limitierung sind prinzipiell durch das Beobachtungsfenster gegeben, können jedoch in ihrer Form (Rechteck, Hamming oder ähnlicher Fensterfunktionen) variiert werden. Den einfachsten Fall der Eingrenzung des Frequenzbereichs stellt äquivalent dazu eine Fensterung in der Frequenzdomäne – Bandpassfilterung – dar. Um eine ausreichend genaue Annäherung des extrapolierten Signals zum Original zu erreichen, müssen die Bandpässe sehr schmalbandig ausgeführt werden. Die Implementation einer geeigneten Filterbank hat eine zeitliche Verzögerung des

Signals zur Folge, die sich dementsprechend nachteilig auf die Performanz des Gesamtsystems der Audioübertragung auswirkt.

Eine andere Art, das Signal im Frequenzbereich zu beschneiden, bietet die lineare Prädiktion. Im Gegensatz zur Bandpass-Filterbank wird hier direkt die spektrale Information des Signals verwertet. Diese Methode verspricht einen wesentlichen Vorteil bei der Extrapolation. Wiederum besteht die Herausforderung darin, die Verarbeitungszeit so gering wie möglich zu halten. Da für die lineare Prädiktion die Koeffizienten des Prädiktors laufend neu abgeschätzt werden müssen, ist ein rekursiver Algorithmus notwendig. Es ist jedoch möglich, die Rekursion mit den Signalblöcken der Übertragung zu synchronisieren und damit eine kontinuierliche Aktualisierung des Systems zu erreichen, ohne die Latenzzeit negativ zu beeinflussen.

3 Subjektive Testreihen

Eine ausgedehnte Evaluierung der Qualität der Verschleierungstechniken wird mit Hilfe von zwei Testreihen durchgeführt. Diese sind zeitlich nacheinander angeordnet, sodass Erkenntnisse, die aus dem ersten Durchlauf gewonnen werden, in der zweiten Untersuchung berücksichtigt und damit gezielt weitere Details extrahiert werden können. Dementsprechend sind die Ziele vorerst sehr allgemein gewählt, wodurch sich ein relativ umfassendes Testdesign ergibt.

- Vergleich der beiden Algorithmen "Extrapolation" vs. "Interpolation": Es soll untersucht werden, ob und in welchem Ausmaß sich die Qualität dieser Techniken auf unterschiedliche Signale (Instrumente, Sprache) bzw. Musikstile auswirkt. Obwohl insgesamt drei Algorithmen inklusive einer zusätzlichen leichten Modifikationsstufe entwickelt wurden, werden letztere aufgrund der ohnedies komplexen Teststruktur in Reihe 1 nicht berücksichtigt.
- Die gewählten Audiobeispiele beinhalten Vertreter der Genres Pop, Klassik und Jazz, die Instrumente Violine und Klavier, sowie ein Sprachbeispiel.
- In die originalen Musikdateien werden künstlich Ausfälle eingefügt, um ein reales Einsatzszenario der digitalen Übertragungsstrecke zu simulieren. Die Statistik der Ausfälle basiert auf Erfahrungswerten ähnlicher Datenübertragungen.

- Zur Evaluierung der Auswirkung des Ausfallszeitpunkts innerhalb eines Testsignals werden verschiedene Ausfallszenarien zufällig erzeugt (die jedoch alle dieselbe Ausfallstatistik aufweisen).
- Die den Algorithmen inhärenten Parameter werden im Vorhinein grob adjustiert und "eingefroren", um eine Beeinflussung dieser auf das Testergebnis zu vermeiden.
- Der Testablauf beruht auf dem Standard ITU-R BS.1116-1, der speziell zur subjektiven Beurteilung von geringfügigen Signalverschlechterungen eingeführt wurde.
- Da die Versuchspersonen allgemein als nicht geschult in dieser Methode der psychoakustischen Messverfahren einzustufen sind, ist vor dem eigentlichen Test eine entsprechende Einführungsphase notwendig. Die erste Testreihe umfasst eine Gruppe von 29 Probanden, die zweite wird mit 15 durchgeführt.
- Die Reliabilität der Antworten wird durch zwei Testdurchläufe mit jeweils exakt denselben Testsamples geprüft, allerdings findet vor jedem Zyklus eine vollständige Randomisierung dieser statt.

Die Auswertung der Ergebnisse anhand statistischer Methoden zeigt, dass das gewählte Testdesign keine allgemeingültigen, signifikanten Aussagen über den qualitativen Unterschied der beiden Algorithmen zulässt. Hierfür sind erweiterte Tests mit einer größeren Anzahl an Musikbeispielen sowie Ausfallszenarien notwendig. Diese Untersuchungen werden jedoch derzeit nicht weiter verfolgt, vielmehr wird die zweite Testreihe in Bezug auf Umfang bzw. Testdauer reduziert. Aus diesem Grund spezialisiert sich das neue Design auf die Extrapolationstechnik.

Ein ausgeprägter Zusammenhang besteht einerseits zwischen Art des Audiobeispiels und Verschleierungsgüte, zum anderen lässt sich relativ deutlich erkennen, dass der Ausfallszeitpunkt einen nicht vernachlässigbaren Einfluss auf das Leistungsverhalten der Algorithmen hat. Komplexe Audiosignale, die aus mehreren überlagerten Einzelsignalen bestehen (z.B. Instrumentengruppen, Orchester), sind für die betrachteten Verschleierungstechniken besser geeignet als solierende Instrumente, weshalb letztere im zweiten Testdesign genauer untersucht werden. Zusätzlich wird die Ausfallstatistik mit Hilfe aktueller Daten aus verbesserten Testübertragungstrecken adaptiert, welche die vorhin festgelegten, harten

Anforderungen etwas entschärft. Als weitere Modifikation für die zweite Testreihe erfolgt eine gezielte Optimierung der dem Algorithmus inhärenten Parameter.

Ein Vergleich der Auswertungen von Testreihe 1 und 2 zeigt, dass die Veränderungen im Testdesign positive Auswirkungen auf die Funktionalität der Verschleierungseinheit hat. Alternativ dazu erfolgt eine Gegenüberstellung mit einer anderen, derzeit aktuellen Verschleierungstechnik, wobei ein Leistungsvorsprung der vorgeschlagenen Methode(n) festzustellen ist, siehe Abb. 3.

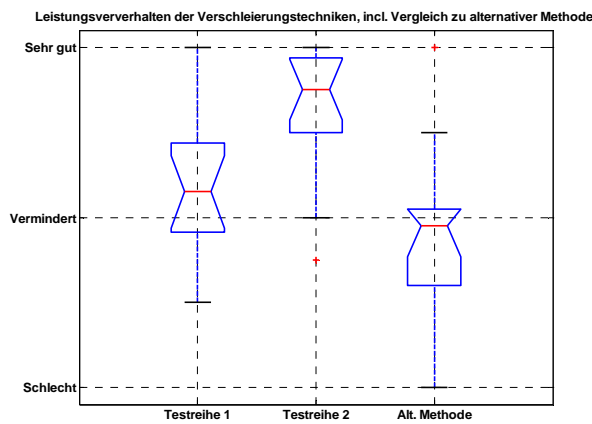


Abb. 3: Statistische Auswertung der Testergebnisse