

# Echtzeitsystem zur mehrkanaligen Breitbandtelefonie

Marco Gimm, Katharina Rebbe, Gerhard Schmidt

Christian-Albrechts-Universität zu Kiel, E-mail: (mgi,stu108730,gus)@tf.uni-kiel.de

## Einleitung

Auch in Zeiten von immer größeren Übertragungsraten werden häufig lediglich geringe Teile des Sprachspektrums bei der mobilen Telefonie übertragen. Dies wirkt sich vor allem negativ auf die Sprachqualität aus. Neue Standards wie HD-Voice sollen Abhilfe schaffen und ermöglichen eine höhere Bandbreite. Eine weitere Entwicklung der letzten Jahre ist, dass mobiles Telefonieren wesentlich günstiger geworden ist. Bei vielen Anbietern kann für einen festen monatlichen Betrag unbegrenzt telefoniert werden. Auch besitzen heute viel Menschen mindestens ein Mobiltelefon. Ein anderer Ansatz zur Verbesserung der Sprachqualität - neben HD-Voice - ist es, mehrere Mobilfunkverbindungen parallel für ein einzelnes Gespräch zu nutzen, um so die Bandbreite zu erhöhen. Der Vorteil hierbei ist die Modularität, sodass auch noch höhere Bandbreiten sowie eine Stereotelefonie möglich sind. Dazu wird zunächst ein Modulationsverfahren vorgestellt, welches erlaubt die Frequenzbänder von Sprachsignalen auf verschiedenste Weise aufzuteilen und außerdem eine Referenz einzuprägen. Anschließend können diese über eine oder mehrere Mobilfunkverbindungen übertragen werden. Auf der Gegenseite werden die Laufzeiten der einzelnen Verbindungen mittels einer effektiven Implementierung von Kreuzkorrelationen geschätzt, sodass die einzelnen Signale zeitlich angeglichen werden können. Anschließend werden diese wieder zu ihren ursprünglichen Signalen zusammengefügt.

## Aufbau des Systems

Die prinzipielle Aufgabe des Systems ist es, Signale in verschiedene Frequenzbereiche aufzuteilen. Diese werden anschließend benutzerdefiniert zusammengefasst bzw. aufgeteilt und wenn notwendig in das Basisband verschoben, um sie anschließend zu übertragen. So ist es denkbar, ein breitbandiges Signal in mehrere schmalbandige Anteile aufzuteilen und anschließend wieder zusammenzufügen. Eine weitere Möglichkeit wäre es eine breitbandige Verbindung zu nutzen, um mehrere schmalbandige Signale darüber zu versenden, die vorher zu einem kombiniert wurden. Anschließend können diese beim Empfänger wieder separiert werden. Um dies möglichst generisch implementieren zu können, bietet sich eine Verarbeitung im Teilbandbereich an. Hierfür wird eine STFT Filterbank-Struktur nach [1] verwendet. Die Filterbank, welche für die Signalaufteilung zuständig ist, wird als Analysefilterbank bezeichnet. Diese bilden den ersten Signalverarbeitungsblock. Jeweils am Ende der Signalverarbeitung befinden sich die Synthesefilterbänke, welche dazu dienen die Frequenzstützstellen wieder zu einem Gesamtsignal zusammenzufügen. Da dieses nicht zwingend dieselbe

Bandbreite, wie das Eingangssignal, hat, muss die Filterbanklänge nicht der der Analysefilterbank entsprechen. Grundsätzlich ist es hier aber sinnvoll, Längen von Zweier- bzw. Vierer-Potenzen zu wählen, da so auf Radix-2 bzw. Radix-4 FFT-Implementierungen zurückgegriffen werden kann. Die Signalverarbeitung wurde in Sender- und Empfängersignalverarbeitungsblöcke untergliedert. Der Senderblock gestaltet sich, wie in Abb. 1 dargestellt.

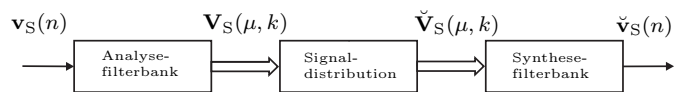


Abbildung 1: Struktur der Senderseite

Darin ist  $\mathbf{v}_S(n) \in \mathbb{R}^{A_{S,E} \times R_{S,E}}$  eine Matrix bestehend aus  $A_{S,E}$  Eingangssignalen, die mit einem Rahmenversatz von  $R_{S,E}$  blockweise abgetastet werden. Der Sender ist für die Signalverteilung zuständig. Zunächst werden  $A_{S,E}$  Analysefilterbänke gerechnet. An deren Ausgang liegt das Signal  $\mathbf{V}_S(\mu, k) \in \mathbb{C}^{A_{S,E} \times N_{\text{FFT},S,Ana}}$  an. Wobei  $N_{\text{FFT},S,Ana}$  der FFT-Länge der Analysefilterbank des Senders entspricht.  $\mu$  beschreibt die Frequenzstützstelle und  $k$  den Rahmenindex. Anschließend werden die Signale zu den  $A_{S,A}$  Signalen zusammengefügt bzw. aufgeteilt. Der Vektor  $\check{\mathbf{V}}_S(\mu, k) \in \mathbb{C}^{A_{S,A} \times N_{\text{FFT},S,Syn}}$  wird mit Hilfe der Synthesefilterbänke der Länge  $N_{\text{FFT},S,Syn}$  zum Ausgangssignalvektor  $\check{\mathbf{v}}(n) \in \mathbb{R}^{A_{S,A} \times R_{S,A}}$ , wobei es sich bei  $R_{S,A}$  um den Rahmenversatz am Ausgang und bei  $A_{S,A}$  um die Anzahl der Ausgangssignale handelt.

Eine einfache Neuordnung der Frequenzstützstellen würde allerdings für Verzerrungen sorgen, sodass eine der Verschiebung entsprechende Phasenkorrektur nach [2] stattfinden muss:

$$\check{V}_S^{(e,a)}\left(\mu + \Delta^{(e,a)}(\mu, k), k\right) = V_S^{(e,a)}(\mu, k) e^{j\phi^{(e,a)}(\mu, k)} \quad (1)$$

mit  $\phi^{(e,a)}(\mu, k) = \frac{2\pi\Delta^{(e,a)}(\mu, k)kR_{S,A}}{N_{\text{FFT},E,Syn}}$   $\check{V}_S^{(e,a)}(\mu + \Delta^{(e,a)}(\mu, k), k)$  ist der jeweilige Teil des Ausgangssignals (Index  $a$ ), welcher aus einem Teil des Eingangssignals (Index  $e$ ) besteht. Dieses kann ganzzahlig um  $\Delta^{(e,a)}(\mu, k)$  bezüglich der ursprünglichen Frequenzstützstelle verschoben sein. Würden alle Signale über die selbe Verbindung versendet werden, könnte der Empfänger so gestaltet werden, dass dieser identisch zum Sender ist. Dieser müsste lediglich so konfiguriert werden, dass die Signalverteilung genau invers zu der des Senders vorgenommen wird. Das Gleiche gilt für die Modulationsfilter. Davon kann allerdings nicht ausgegangen werden. Bei der Übertragung der Signale muss von mehreren Verbindungs-Unterschieden ausgegangen werden. Einige Beispiele hierfür sind:

- Unterschiedliche Signallaufzeit durch Verwendung verschiedener Komponenten,
- Unterschiedliche Signallaufzeit durch verschiedenes Routing,
- Unterschiedliche Signallaufzeit durch Verwendung verschiedener Netzbetreiber,
- Unterschiedliche Codierungsverfahren durch Verwendung unterschiedlicher Komponenten.

Laut [3] lässt sich ein Laufzeitunterschied von Sprache, die zum einen per Direktschall und zum anderen über einen zweiten Kanal wiedergegeben wird, ab einer Laufzeit von etwa 35 ms wahrnehmen und wird als störend empfunden. Für einen korrekten räumlichen Effekt bei mehrkanaligen Signalen ist die Richtung der ersten Wellenfront entscheidend. Eine Mobilfunkverbindung hat durch das *Routing* Laufzeiten von 70 ms bis 100 ms [4], sodass alleine hierdurch eine Differenz von bis zu 30 ms vorhanden sein kann. Dazu kommen noch weitere mögliche Laufzeitunterschiede, bspw. durch die Verwendung unterschiedlicher Geräte. Folglich muss mit einer Überschreitung der kritischen Laufzeit bei dieser Anwendung gerechnet werden. Um diese auszugleichen und somit für eine erhöhte Sprachqualität zu sorgen, wird der Empfänger im Gegensatz zum Sender um eine Laufzeitschätzung sowie einen Laufzeitausgleich ergänzt. Eine Schätzung der Laufzeitunterschiede wurde mittels

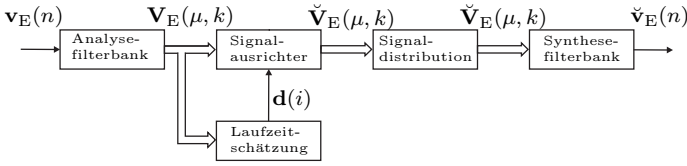


Abbildung 2: Struktur der Empfängerseite

Kreuzkorrelationen realisiert. Diese können nur verwertbare Ergebnisse liefern, wenn die Signale, die miteinander korreliert werden auch gleiche Anteile aufweisen, welche lediglich durch die Verbindungen unterschiedlich verzögert werden. Daher ist es notwendig jedem Ausgangssignal auf der Senderseite ein Referenzsignal hinzuzufügen. Im Empfänger muss bekannt sein, welche spektralen Anteile zu dem Referenzsignal gehören, sodass diese zur Schätzung der Laufzeitdifferenz genutzt und anschließend aus dem Signal entfernt werden können. Abb. 3 zeigt eine exemplarische Signalaufteilung im Teilbandbereich für eine Konfiguration mit einem Eingangssignal und drei Ausgangssignalen. Die in Graustufen gekennzeichneten Signalanteile sind jeweils das Nutzsignal bzw. dessen Anteile. Der Signalanteil im Karomuster beschreibt das Referenzsignal, welches bei allen drei Ausgangssignalen identisch ist.

Ein Rahmen am Eingang des Empfängers besitzt die Länge  $R_{E,E}$ . Die Analysefilterbank verfügt über eine gewisse Überlappung. Üblicherweise wird in diesem System eine Überlappung von 75% genutzt, sodass die FFT-Ordnung dem vierfachen Rahmenversatz entspricht. Somit beträgt der Zeitbereich, der pro Rahmen betrachtet werden kann:

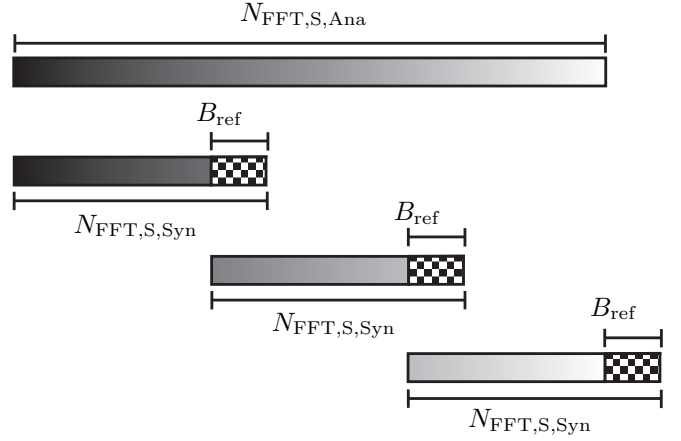


Abbildung 3: Exemplarische Signalaufteilung im Teilbandbereich mit einem Eingangssignal und drei Ausgangssignalen

$$t_{CC} = \frac{4R_{E,E}}{f_s}, \quad (2)$$

wobei es sich bei  $f_s$  um die Abtastrate handelt. Bei einem Rahmenversatz von  $R_{E,E} = 64$  und einer Abtastrate von  $f_s = 44100$  Hz bedeutet das einen Zeitraum von 5,8 ms. Um auch einen größeren Laufzeitunterschied feststellen zu können, bedarf es demnach eine Betrachtung von mehr als einem Rahmen. Daher werden die Rahmen in einem Ringspeicher hinterlegt. Die maximale Länge hierfür ist wählbar und wird auf die maximal erwartete Länge der Laufzeitdifferenz festgelegt. Weiteres Potential zur Einsparung von Rechenlast bietet die Tatsache, dass nur Teilbänder betrachtet werden müssen, welche das Referenzsignal beinhalten. Das hätte zur Folge, dass alle übrigen Teilbänder zu Null gesetzt, trotzdem aber die volle FFT-Länge genutzt würde. Eine Möglichkeit, dies zu vermeiden, ist es das Referenzsignal in das Basisband zu verschieben und die FFT-Länge  $N_{FFT,CC}$  entsprechend zu der Länge des Referenzsignals  $B_{Ref}$  zu setzen. Zudem muss berücksichtigt werden, dass der Rahmenversatz  $R_{CC}$  ebenfalls angepasst werden muss, sodass die Beziehung

$$N_{FFT,CC} = 4R_{CC} \quad (3)$$

für eine Überlappung von 75% gegeben ist. Dadurch ändert sich ebenfalls die Abtastrate der Kreuzkorrelierten, da das Vorgehen einer Unterabtastung mit dem Faktor

$$M_{CC} = \frac{N_{FFT,E,Ana}}{2B_{ref}}, \quad (4)$$

entspricht. Es ist nicht zu erwarten, dass eine große Variation in der Laufzeit herrscht. Daher wurde ein Verfahren entwickelt, mit welchem die Laufzeitschätzung verteilt über mehrere Rahmen berechnet wird. Das hat den Vorteil, dass nicht für jeden neuen Rahmen Kreuzkorrelationen über die gesamte Länge berechnet werden müssen. Auch eine Berechnung, die in einem konstanten Zeitabstand vorgenommen wird, wäre ineffizient, da so Spitzen in der Rechenlast geschaffen würden.

Damit ein System echtzeitfähig ist, muss es eine definierte Aufgabe in einer vorgegebenen Zeit erledigen

[5]. In diesem Fall bedeutet dies, dass ein Rahmen verarbeitet werden muss, bevor ein neuer vorhanden ist. Demnach ist es sinnvoll, die Verarbeitungszeit für einen Rahmen möglichst konstant zu halten. Die Berechnung der Kreuzkorrelation lässt sich effizient im Frequenzbereich durchführen. Wie bereits erwähnt, soll hier ein Laufzeitunterschied zwischen den Signalen  $v_{CC}^{(j)}(n)$  und  $v_{CC}^{(h)}(n)$  festgestellt werden können, der die Dauer eines Rahmens übersteigt. Dafür wurde eine effiziente Umsetzung der Schätzung der Kreuzkorrelation unter Verwendung der einzelnen Kurzzeitkreuzleistungsdichtespektren implementiert. Soll ein Laufzeitunterschied von bis zu  $L$  Rahmen festgestellt werden, ist es somit notwendig von jedem der zu korrelierenden Signale  $L$  Spektren zu speichern. Nun lassen sich  $2L - 1$  einzelne Kurzzeitkreuzleistungsdichtespektren, die als Matrix  $\mathbf{S}_{v_{CC}^{(j)}v_{CC}^{(h)}} \in \mathbb{C}^{(2L-1) \times N_{\text{FFT,CC}}}$  zusammengefasst werden, wie folgt berechnen:

$$S_{v_{CC}^{(j)}v_{CC}^{(h)}}^{(u)}(\mu, k) = \left( V_{CC}^{(j)}(\mu, k-p) \right)^* V_{CC}^{(h)}(\mu, k-o) \quad (5)$$

mit  $u = (L-1) + (p-o)$ .

Dabei gilt  $o, p \in [0, 1 \dots, L-1]$  und  $o \cap p \in [0]$ . Die einzelnen Kurzzeitkreuzleistungsdichtespektren werden zusätzlich mit einem IIR-Filter erster Ordnung mit dem Glättungsparameter  $\alpha$  geglättet.

$$\overline{S_{v_{CC}^{(j)}v_{CC}^{(h)}}^{(u)}(\mu, k)} = \alpha \overline{S_{v_{CC}^{(j)}v_{CC}^{(h)}}^{(u)}(\mu, k-1)} + (1-\alpha) \overline{S_{v_{CC}^{(j)}v_{CC}^{(h)}}^{(u)}(\mu, k)} \quad (6)$$

Jedes zuvor bestimmte Kreuzleistungsdichtespektrum beinhaltet nun die Fouriertransformierte der Kreuzkorrelierten des aktuellen Rahmens eines der zur korrelierenden Signale und entweder des aktuellen oder eines verzögerten Rahmens des jeweils anderen zu korrelierenden Signals. Diese sollen im nächsten Schritt in den Zeitbereich transformiert werden, wobei anschließend eine Fensterung mit der Fensterfunktion  $w(\kappa)$  vorgenommen wird. Die einzelnen Kreuzkorrelationsfunktionen werden zu  $\hat{\mathbf{s}}_{v_{CC}^{(j)}v_{CC}^{(h)}}^{(u)}(\kappa) \in \mathbb{R}^{(2L-1) \times 2L R_{CC}-1}$  zusammengefasst:

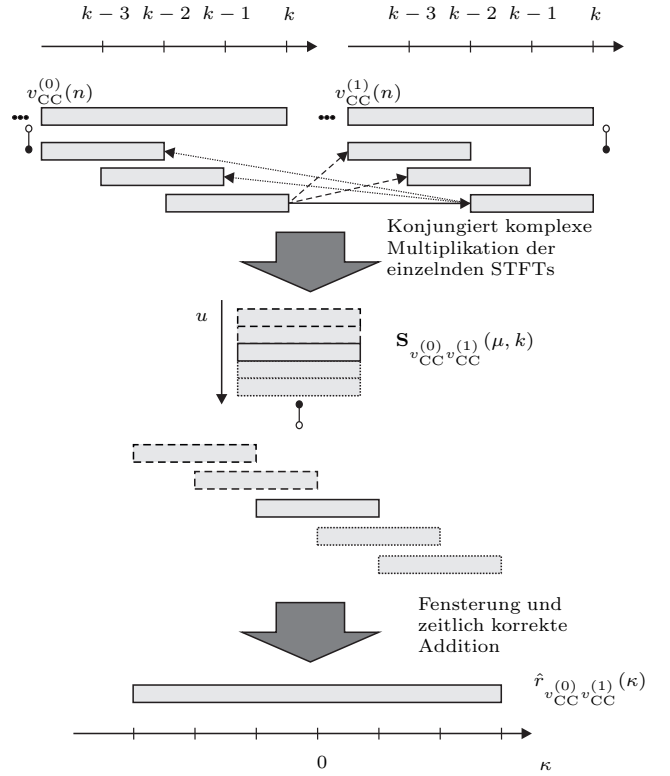
$$\hat{\mathbf{s}}_{v_{CC}^{(j)}v_{CC}^{(h)}}^{(u)}(\kappa) = \begin{cases} w(\kappa) \mathcal{F}^{-1} \left\{ \overline{S_{v_{CC}^{(j)}v_{CC}^{(h)}}^{(u)}(\mu, k)} \right\}, \\ \text{für } 0 \leq \kappa < N_{\text{FFT,CC}} \\ 0, \text{ sonst.} \end{cases} \quad (7)$$

Da die Korrelationsschätzung innerhalb einer Filterbankstruktur mit Überlappung stattfindet, muss diese auch bei der Zusammensetzung der gesamten Kreuzkorrelierten aus den einzelnen Intervallen berücksichtigt werden. Die gesamte Schätzung der Kreuzkorrelierten ergibt sich somit als Summe aller gefensterten Schätzungen, die entsprechend der zugrundeliegenden zeitlichen Differenz der Rahmen, aus denen sie gebildet wurden, zu verschieben sind:

$$\hat{r}_{v_{CC}^{(j)}v_{CC}^{(h)}}^{(u)}(\kappa + LR_{CC}) = \sum_{u=0}^{2L-1} \hat{\mathbf{s}}_{v_{CC}^{(j)}v_{CC}^{(h)}}^{(u)}(\kappa - uR_{CC}). \quad (8)$$

Die hier vorgestellte Variante der Kreuzkorrelation

macht es möglich die Berechnungen beliebig aufzuteilen, sodass die Rechenlast gleichmäßig verteilt werden kann. Die geschätzte Laufzeit wird anschließend dazu verwen-



**Abbildung 4:** Schematische Darstellung der rahmenbasierten Schätzung der Kreuzkorrelierten mit  $L = 3$

det die Signale wieder auszurichten. Dazu ist es notwendig mindestens  $L$  Rahmen von jedem Signal zu speichern. Dies geschieht ebenfalls in einem Ringspeicher. Anschließend wird bestimmt, welches Signal die größte Laufzeit aufweist. Alle anderen Signale müssen entsprechend der Laufzeitsdifferenz zu diesem Signal verzögert werden. Die Verzögerung, sowie die Speicherung der Rahmen findet im Teilbandbereich statt. Daher geschieht zunächst eine Unterteilung der notwendigen Verzögerung in ganze Rahmen  $\tau_A^{(i)}$  und den entsprechenden Rest  $\tau_S^{(i)}$ . Der Index  $i \in [0, 1, \dots, A_{S,E} - 1]$  beschreibt hier, um welche Laufzeitdifferenz es sich handelt. Bei einer Anzahl von  $A_{S,E}$  Eingangssignalen am Empfänger gibt es genau  $A_{S,E} - 1$  Laufzeitdifferenzen bezüglich des ersten Signals:

$$\tau_A^{(i)} = \lfloor d(i) f_S \rfloor, \quad (9)$$

$$\tau_R^{(i)} = \left\lfloor \frac{\tau_A^{(i)}}{R_{E,E}} \right\rfloor, \quad (10)$$

$$\tau_S^{(i)} = \frac{\tau_A^{(i)}}{R_{E,E}} - \left\lfloor \tau_R^{(i)} \right\rfloor. \quad (11)$$

Die Laufzeitunterschiede in ganzen Rahmen werden nun durch Verzögern selbiger ausgeglichen. Der Rest wird wiederum durch die Anwendung des Verschiebungssatzes der Fouriertransformation angeglichen:

$$\check{V}_E^{(i)}(\mu, k) = V_E^{(i)}\left(\mu, k - \tau_R^{(i)}\right) \cdot e^{-j \frac{2\pi \tau_S^{(i)}}{N_{\text{FFT,E,Ana}}}}. \quad (12)$$

Die soeben beschriebene Laufzeitangleichung ist für nahezu konstante Laufzeiten mit nur geringen Änderungen praktikabel. Sollte nun eine plötzliche Veränderung eintreten, sodass die Laufzeiten sowie die einzelnen Laufzeitrelationen sich ändern, würde dies dazu führen, dass eine abrupte Veränderung stattfindet. So würden ggf. so viele Rahmen verloren gehen, dass dies deutlich durch Knackgeräusche zu hören wäre. Um dem vorzubeugen, wurde die Signalangleichung so gestaltet, dass neue Werte der Laufzeitschätzung zunächst einmal als Vorgabe gesetzt werden. Allerdings wird jedes Mal nur ein maximaler Wert  $\Delta\tau$  an Veränderung zugelassen, sodass die tatsächliche Verzögerung dem aktuellen Laufzeitunterschied langsam folgt.

## Implementation und Verifikation

Die beschriebenen Komponenten wurden innerhalb des *Kiel Real-Time Application Toolkit* (kurz KiRAT) implementiert. Dieses bietet zum einen die Möglichkeit, die implementierten Komponenten durch Verwendung externer Signale in Echtzeit zu testen. Andererseits ist es möglich zuvor aufgenommene Signale ebenfalls in Echtzeit zu verwenden. Dies wurde hier zunächst getan, um die Funktionalität des Systems nachzuweisen. Dabei wurde eine Konfiguration gewählt, welche zwei Signale mit einer Bandbreite von 8 kHz in je zwei Kanäle mit 3,5 kHz aufteilt. Die entstandenen Signale wurden abgespeichert und wiederum als Eingangssignale des Empfängers genutzt. Durch diese Art Test mit verschiedenen Konfigurationen konnte die Funktionalität der Signalverteilung nachgewiesen werden. Die Spektrogramme eines aufgeteilten und anschließend wieder zusammengeführten Eingangssignals sind in Abb. 5 zu sehen.

Weiterhin wurde der zuvor beschriebene Test ebenfalls mit der Verwendung eines Referenzsignals durchgeführt. Als Referenzsignal wurde ein Teil eines der Eingangssignale genutzt, welcher eine Bandbreite von 500 Hz aufweist. Die Ausgangssignale wurden wieder aufgenommen, GSM kodiert und dekodiert und künstlich verzögert. Auch hier wurden verschiedene Verzögerungen zwischen 0 und 200 ms getestet, wobei der maximale festgestellte Fehler bei  $<1$  ms lag. Die Funktionalität der Laufzeitschätzung und des Laufzeitausgleichs konnte so ebenfalls nachgewiesen werden. Außerdem wurde gezeigt, dass die Funktionalität der korrelationsbasierten Laufzeitschätzung trotz Kodierung gegeben ist.

## Fazit und Ausblick

In dieser Arbeit wurde ein Überblick über ein generisches System zur mehrkanaligen Breitbandtelefonie gegeben, welches ebenfalls in der Lage ist viele weitere Anwendungsfälle abzudecken. Hierzu wurden zunächst verschiedene Anwendungsfälle und mögliche Einsatzgebiete aufgezählt, woraus anschließend verschiedene Anforderungen abgeleitet wurden. Daraufhin wurde ein System, welches diese Anforderungen erfüllt, theoretisch beschrieben. Dieses wurde zuletzt implementiert und getestet, sodass die grundsätzliche Funktionalität nachgewiesen werden konnte.

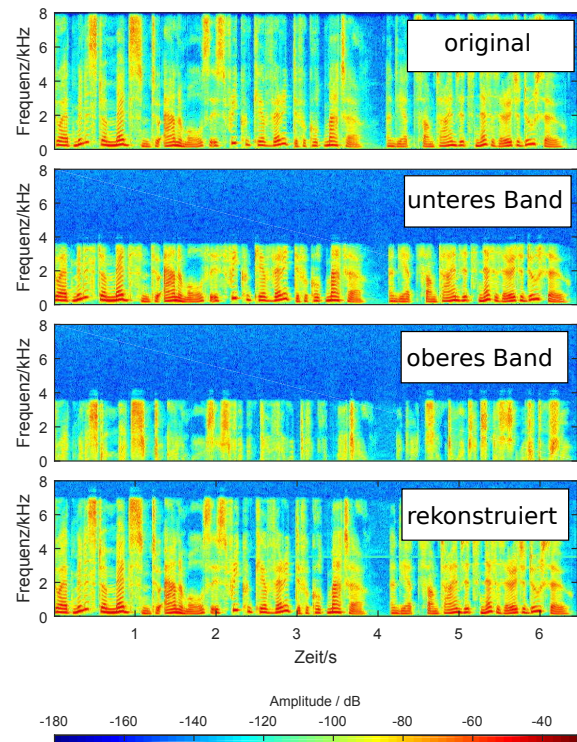


Abbildung 5: Spektrogramme der beschriebenen Verifikation.

Im nächsten Schritt sollen weitere Tests in Echtzeit vorgenommen werden. Dafür wird ein Versuchsaufbau herangezogen, worin das System eine realistische Anwendung unter Verwendung verschiedener Mobilfunkkanäle zu erfüllen hat. Die resultierenden Signale sollen aufgenommen und anhand verschiedener objektiver als auch subjektiver Methoden beurteilt werden.

## Literatur

- [1] J. Allen, „Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 25, Nr. 3, S. 235–238, 1977.
- [2] J. Withopf, S. Rohde, und G. Schmidt, „Application of Frequency Shifting in In-Car Communication Systems,” in *11. ITG Fachtagung Sprachkommunikation*, 2014.
- [3] S. Gelfand, *Hearing: An Introduction to Psychological and Physiological Acoustics, Third Edition*. Taylor & Francis, 1998.
- [4] P. Stuckmann, *The GSM Evolution: Mobile Packet Data Services*. WILEY, 2002.
- [5] *Informationstechnik - Begriffe*, Deutsches Institut für Normung Std. DIN ISO/IEC 2382, 2015.