

Zoom-X Analysis - Audio Analyse

Urs-Benedict Braun
Otto-von-Guericke Universität
Magdeburg, Deutschland
urs-benedict.braun@st.ovgu.de

ACM Reference Format:

Urs-Benedict Braun. 2024. Zoom-X Analysis - Audio Analyse. In *Magdeburg '24: IT-Forensik: Zoom-X Report*, 25.09.2024, Magdeburg, DE. ACM, Magdeburg, MD, DE, 37 pages. <https://doi.org/0000000.0000000>

1 Zusammenfassung

Es handelt sich um einen Erweiterungsbericht des Papers "Meta and Media Data Stream Forensics in the Encrypted Domain of Video Conferences" [2] und untersucht im weiteren die Audiodaten-Analyse von Zoom-X. Dies erfolgt im Rahmen des Moduls "IT-Forensik" [7] der Otto-von-Guericke Universität im Sommersemester 2024.

2 Aufgabenstellung

Die Analyse in diesem Paper betrifft die (Audio)-Infrastruktur. Hier wird spezieller untersucht, ob im Audiodatenstrom (MCS_{AU}), welcher am Empfänger Client CL_2 empfangen wird, versteckte Information enthalten sind. Wenn möglich, sollten diese Aufgespalten werden.

3 Stand der Technik

Zoom-X ist die am meiste genutzte Lösung an der Otto-von-Guericke Universität, um Video-Konferenzen durchzuführen. Zoom-X wird von der Deutschen Telekom betrieben und deren Hardware ist (nach eigenen Bekunden) in Deutschland stationiert [1].

Audacity [12] ist ein freier und Open-Source Audioeditor, der auch Analysefunktionen bereitstellt. Diese Funktionen werden zur Analyse von den Audiodaten zu benötigt.

SOX [6] ist eine Linuxkommandozeilenanwendung zur Analyse und Bearbeitung von Audiodaten. SOX erstellt ein Spektrogramm anhand der Zeitachsen – dieses Programm wird ebenfalls zur Analyse der Audiodaten benötigt.

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

IFOR, September 25, 2024, Deutschland, MD

© 2024 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/0000000.0000000>

4 Konzept

4.1 Allgemeines Konzept

Ein Original liegt als WAV-Datei vor und dient als Referenz. Diese Datei wird über einen Media-Player abgespielt und an Zoom-X [15] übertragen. Der Sound wird sowohl am Quell-PC (bevor es an das Programm Zoom-X geht) als auch am Ziel-PC (nach der Bearbeitung durch Zoom-X) aufgezeichnet.

Zoom-X verarbeitet das Audio und sendet es über den Medienaudio-Datenstrom MCS_{AU} übers Internet zum Zoom-X Server (MMS: 64.224.37.50). Von dort aus wird es an den Empfänger-Client (CL_2) gesendet.

Im Anschluss daran werden die Sounds von Quelle und Ziel miteinander verglichen, um diese auf etwaige Unterschiede hinsichtlich Wasserzeichen zu kontrollieren. Danach wird der aufgezeichnete Sound mit dem Sound der Originaldatei verglichen, um Rückschlüsse auf Qualität und Veränderungen zu erhalten.

Zur Analyse der hierdurch erhaltenen Audiodaten wird mithilfe von SOX [6] je ein Spektrogramm erstellt. Die Spektrogramme beider Aufnahmen werden (Sender, Empfänger) verglichen.

Anschließend werden die Analysedaten auch in Audacity [12] betrachtet, um Auffälligkeiten in der Frequenz, der Sprachkadenz und den Korrelationen zu identifizieren.

4.2 Differenzaudio

Weiterhin werden die beiden Audiodaten synchronisiert. Hiervon wird eine Spur invertiert, um eine Differenzaudiodatei zu erzeugen. Diese Differenzaudiodatei wird dann sowohl durch Spektrogramme als auch durch Hörproben genauer untersucht.

Auf Seiten des Empfängers tritt der Jitter-Effekt auf. Jitter-Effekt bedeutet, dass die gesendeten und empfangenen Audiodaten nicht synchronisiert waren. Um diesen Effekt zu entfernen, wurde die empfangene Audioaufnahme (manuell) bearbeitet und mit der Senderdatei synchronisiert. Beide Tonspuren wurden in Audacity auf -1 dB Spitze normalisiert. Anschließend wurden die Spuren zu einer neuen Spur gemischt, wobei die Spur vom Empfänger invertiert wurde, um die Differenzaudiospur zu erhalten.

Dieses Verfahren basiert auf der Wellentheorie, bei der sich eine positive und eine negative Welle überlagern und gegenseitig auslöschen. Durch diese Methode wird festgestellt, ob Daten hinzugefügt oder entfernt wurden.

4.3 Verschiedene Sprachen

Es werden zwei Tests durchgeführt:

- ein Test mit einem deutschen Text, der synthetische Klänge enthält und
- ein weiterer Test mit einem englischen Text in zwei verschiedenen Dialekten (US-amerikanisches Englisch und britisches Englisch) ohne synthetische Klänge.

Der englische Test wird separat durchgeführt, um ausschließlich die sprachlichen Unterschiede zu untersuchen.

4.4 Kerbfilter Reproduktion (Experiment)

Als ersten Schritt wird der Effekt "Effekte-Rauschverminderung" auf die Originaldatei mit den folgenden Werten angewendet:

- Rauschverminderung: 12 dB
- Empfindlichkeit: 6
- Frequenz-Glättung (Bänder): 3

Als zweites wird der Kerbfilter mit den Parametern angewendet:

- Frequenz: 60 Hz
- Q-Faktor: 1

Als drittes wird der Kompressor mit den folgenden Einstellungen angewendet:

- Schwellenwert: 12 dB
- Grundrauschen: -40 dB
- Verhältnis: 2:1
- Ansprechzeit: 0,1 Sekunden
- Abklingzeit: 1 Sekunde
- Normalisieren auf -1 dB und Gleichspannungsversatz entfernen

Diese reproduzierte Datei wird mit der Original-Audio verglichen, eine Spur invertiert, analysiert und diese verglichen mit der Differenzaudio von Zoom-X.

5 Implementierung und Versuchsaufbau

5.1 VM Vorbereitung

In VirtualBox (Version 7.0.8 [11]) wird eine Umgebung mit den folgenden Einstellungen vorbereitet:

- 4 Kerne
- 4096 MB Arbeitsspeicher
- Netzwerkbrücke

Zunächst wird das Betriebssystem Windows 10 (Version 22H2, Oktober 2022 [13]) installiert. Danach folgen die Installationen der VirtualBox Guest Tools (Version 7.0.14 [11]) und der obligatorische Windows Update-Prozess.

Zur Erfassung des Datenverkehrs werden Wireshark (Version 4.2.4 x64 [8]) und ZAP Proxy (Version 2.14.0 [14]) installiert. ZAP Proxy benötigt zudem das Adoptium Java JRE [9], welches die Programme OpenJDK (Version 21.0.3, LTS, veröffentlicht am 16. April 2024) und OpenJRE (Build 21.0.3+9-LTS) umfasst.

Für die Übertragung des Audiosignals zwischen Abspielgerät, Aufnahmegerät und Zoom wird Virtual Audio Cable (Pack 43 [5]) verwendet. Zur Audioaufnahme wird Audacity (Version 3.5.1 [12]) eingesetzt. Abschließend erfolgt die Installation von Zoom-X (Version 6.0.4 (38135) [15]).

5.2 PC Aufbau

Unser Aufbau unterscheidet sich zum oben genannten Paper [2] in dem Sinne, das Sender- und Empfänger- PC sich im selben Netzwerk befinden und die Infrastruktur von Zoom-X benutzt wird.

5.3 Audio Wasserzeichen Detektion

Der Audioinhalt besteht aus folgenden Komponenten:

- Klang zur Synchronisation
- 10-minütige Gutenberg-"Vorlesung" (die ersten 10 Minuten der erzeugten Datei)
- DTMP-Abfolge (DTMF-Abfolge: audacity; Amplitude: 0,8 Dauer: 5 Sekunden; Aufgabenablauf: 55%)
- Sinusklang (Wellenform: Sinus; Frequenz: 44Hz; Amplitude: 0,8; Dauer: 30 Sekunden)
- Rechteckklang (Wellenform: rechteck; Frequenz: 44Hz; Amplitude: 0,8; Dauer: 30 Sekunden)
- Weißes-Rauschen (Rauschart: weiß; Amplitude: 0,6; Dauer: 30s)
- Stille
- Zirpen (Wellenform: Sinus; Start: 440Hz bei 0,8 Amplitude; Ende: 1320Hz bei 0,1 Amplitude; Dauer: 30 Sekunden)

Zeitstempel	Länge	Audioklang
1	1	Klang
2	600	Vorlesung / Gutenberg
601	5	DTMP
606	30	Sinus Klang
636	30	Rechteck Klang
666	30	Whitenoise
696	30	Stille
726	30	Zirpen

Table 1. Übersicht der Audioinhalte; Zeiteinheiten in Sekunden (de.wav)

Der Test besteht aus einer 10-minütigen synthetischen Sprachaufnahme (Raw-Data DT_1) und etwa 3 Minuten synthetischen Klängen, die über Virtual Audio Cable [5] übertragen werden. Diese Inhalte werden über Zoom-X übertragen, am Sender und Empfänger aufgezeichnet und anschließend miteinander verglichen. Ein Text-to-Speech-Programm "pico2wav" von SOX (v14.4.2) [6] erstellt diese Inhalte basierend auf einer Vorlage des Gutenberg-Projekts [16].

Ein zweiter Test beinhaltet die Übertragung einer 10-minütigen synthetische Sprachaufnahme in englischer

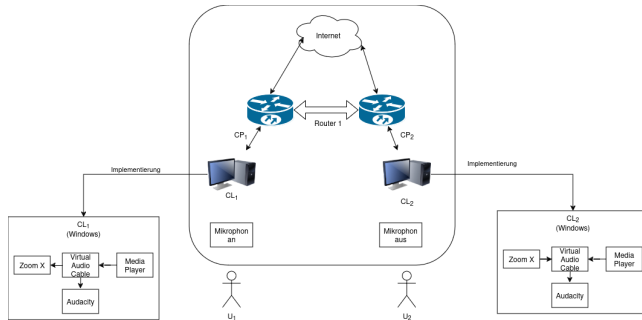


Figure 1. Aufbau des Audio-Versuches

Sprache, jeweils einmal mit einem US-amerikanischen und einem britischen Sprecher, um Unterschiede zwischen den Sprachvarianten zu untersuchen. Auch dieser Text stammt aus dem Gutenberg-Projekt [4].

5.4 Vorgehen der Audio Analyse

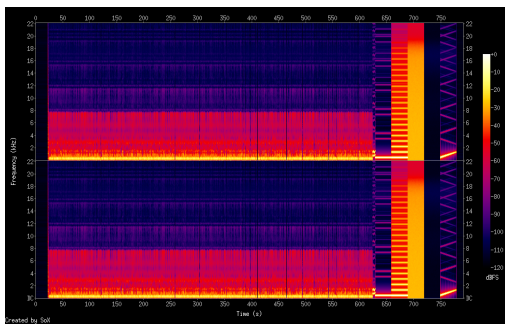
Auf der Senderseite (CL_1) wird Zoom-X so konfiguriert, dass das Mikrofon des Programms auf das Virtual Audio Cable (VBC) geroutet wird und der Media-Player auf dieselbe Quelle zugreift, um Musik abzuspielen. Audacity zeichnet die Audiodaten an dieser Stelle auf, bevor sie gesendet werden. Zoom-X überträgt das Audio anschließend über seine Server zum Empfänger.

Auf der Empfängerseite (CL_2) wird der Lautsprecher von Zoom-X auf das VBC eingestellt und von dort wird das Audio über Audacity aufgenommen. Anschließend wird die Datei im WAV-Format speichern und unterziehen sie einer näheren Untersuchung.

6 Evaluierung – Versuchsdurchführung – Diskussion der Ergebnisse

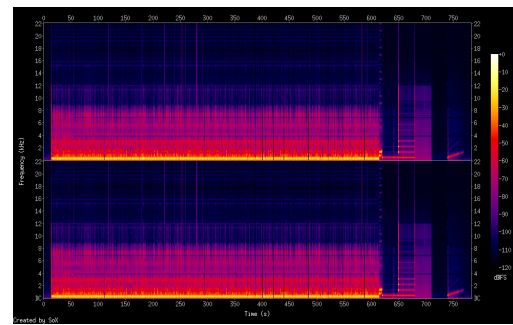
Alle Versuchsdurchführung erfolgen am Zoom-X Client (Patch-Notes von Zoom-X [10]). In diesen Abschnitt erfolgt die reine Audio Untersuchung.

6.1 Initialer Test

Figure 2. SOX Spektrogramm am Sender (CL_1 , Analyse 1, R_1)

R_n	Datum	Uhrzeit	Event	Datei
	26.04.24		Client Version 6.0.4 (38135)	
	24.04.24		DR1	
R1	03.05.24	16:38	Analyse 1	de.wav
R2	10.05.24	08:19	Analyse 2 – PC 2 Sender	de.wav
R3	10.05.24	08:29	Analyse 2 – PC 2 Empfänger	de.wav
R4	14.05.24	11:39	Analyse 3 – 1 PC (sendet)	de.wav
R5	14.05.24	12:01	Analyse 3 – 2 PC (sendet)	de.wav
	20.05.24		Client Version 6.0.10 (39647)	
	22.05.24		Hinweis verschiedene Sprecher	
R6	23.05.24	12:22	Analyse 4 – enUS (1. PC sendet)	enUS.wav
R7	23.05.24	12:33	Analyse 4 – enUS (2. PC sendet)	enUS.wav
	23.05.24	12:45	Analyse 4 – enGB (1. PC sendet)	enGB.wav
R9	23.05.24	13:00	Analyse 4 – enGB (2. PC sendet)	enGB.wav
	19.06.24		DR2	
	30.07.24		Update abgebrochen	
R10	30.07.24	10:10	Kontrolle verschlüsselter Kanal	de.wav
	07.08.24		DR3	
R11	07.08.24	13:10	Kontrolle verschlüsselter Kanal	all.wav
	25.09.24		DR4	

Table 2. Analysen und Events

Figure 3. SOX Spektrogramm am Empfänger (CL_2 , Analyse 1, R_1)

SOX. Beide Diagramme (DT_8 , 2, 3) haben gleiche x- und y-Achsen und stammen aus dem Datensatz R_1 . Für die Versuche $R_{1,4,5}$ spiegeln sich die Diagramme 2 im den Ergebnissen wieder. Für $R_{2,3}$ werden keine Audiodaten am Sender-Client (CL_1) aufgenommen, da dies ein Kontrolltest ist.

In der Abbildung 3 (Empfängerseitig) sowie im Anhang 8. 11 sind im Bereich der Stimmen zwei Lücken erkennbar, die im Original 2 bzw. 8. 7 im Anhang nicht vorhanden sind:

- Zwischen 3,8 kHz und 4 kHz
- Zwischen 6,8 kHz und 7 kHz

Diese Lücken könnten auf ein Wasserzeichen oder auf einen fehlerhaften Kerbfilter/Notch-Filter hinweisen.

In einem **Kontrolltest** ($R_{2,3}$) (siehe SOX Diagramm 8. 16) wird auf diese gefundene Eigenschaft erneut getestet, jedoch sind diese Lücken nicht mehr vorhanden. Somit ist dieses Ergebnis auf einen Fehler in der Einstellung für den Rausch- bzw. Kerbfilter seitens Zoom-X zurückzuführen.

Bei den synthetischen Klängen (ab Sekunde 620) gibt es keinen bis kaum einen Ausschlag, da diese durch Zoom-X eliminiert werden. Bei den Sinus- und Rechteckklängen sind nur noch vereinzelte leise Bruchstücke hörbar. In der Hörprobe erscheint dies wie ein dumpfes Geräusch und im SOX-Diagramm ist bei 500 Hz eine rote Linie mit -50 dBFS sichtbar.

Hintergrundrauschen (VBC-Fehler) sind in Form von "Schlieren" im SOX-Spektrogramm oberhalb der 8 kHz-Marke erkennbar, jedoch auch in der Senderdatei (Fig. 2). Somit ist dies kein Anzeichen auf ein Wasserzeichen seitens Zoom-X.

Ab Sekunde 657 zeigt Audacity einen leichten Ausschlag, obwohl nichts hörbar ist. Im SOX-Diagramm sind an dieser Stelle mehrere (7+1 im Kontrolltest) rote Linien verschiedener Intensitäten mit circa -40 dBFS erkennbar. Die erste Linie ist am intensivsten und die Intensität nimmt von Linie zu Linie ab. Eine dieser Linien befindet sich wahrscheinlich in der oben genannten Lücke.

Das weiße Rauschen (ab 678 s) wird vollständig herausgefiltert. Beim Klang an der Sekunde 739 gibt es einen leichten Ausschlag, der ebenfalls hörbar ist.

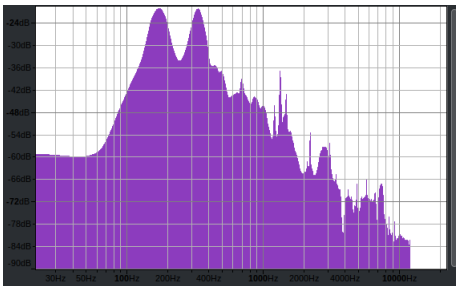


Figure 4. Audacity Spektrogramm am Empfänger (Cl_1 , Analyse 1, R_1) - x-Achse: Hz; y-Achse bei: -24db, relative Lautstärke

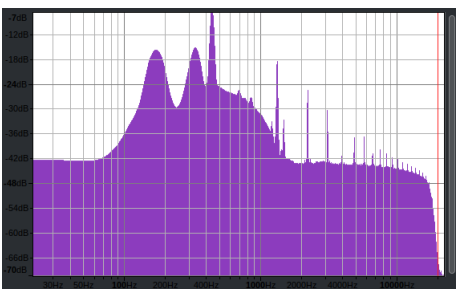


Figure 5. Audacity Spektrogramm am Sender (Cl_2 , Analyse 1, R_1) - x-Achse: Hz; y-Achse bei: -7db, relative Lautstärke

Audacity Spektrogramm. Im Audacity-Spektrogramm ist ab circa 10 kHz (X-Achse) ein "plötzliches Ende" erkennbar 8. 12 bzw. 4 im Anhang, welcher beim Sender 5 bzw. 8. 8 im Anhang nicht vorhanden sind. Dies deutet darauf hin,

dass Zoom-X das übertragene Audio bearbeitet und diese Frequenzen rausgefiltert hat.

Die y-Achsen beider Diagramme sind unterschiedlich, beim Sender (Cl_1) liegt die Lautstärke bei -7db, wobei beim Empfänger (Cl_2) -24db ist, währenddessen dessen Spitze bei -19,9db ist. Dies entspricht eine Reduktion und Dämpfung der Geräusche auf Empfängerseite.

Die Abbildung 6 zeigt im Vergleich zur Sender-Abbildung 7, dass die Audiodaten am Empfänger weicher ankommen und die "harten Kanten" geglättet werden, somit harmonisiert wurden.

In der Abbildung 8 wird deutlich, dass nicht mehr Störgeräusche (bzw. synthetische Klänge) am Empfänger-PC vorhanden sind als beim Sender 9.

Das gleichen Diagramm 4 zeigt auch das (gleiche) Ergebnis für die Audiospuren $R_{1,3,4,5}$ für die Empfängerseite an (gilt auch für die Paragraphen: Standard Korrelation und Erweiterte Autokorrelation).

Das gleichen Diagramm 5 zeigt das gleiche Ergebnis für die Audiospuren $R_{1,4,5}$ für die Senderseite an (gilt auch für die Paragraphen: Standard Korrelation und Erweiterte Autokorrelation).

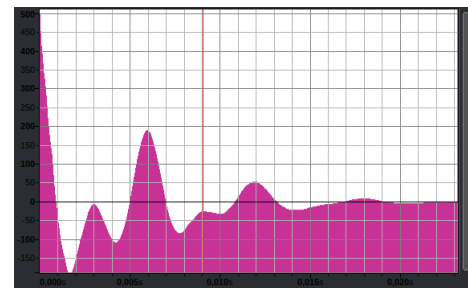


Figure 6. Audacity Spektrogramm (Standard-Autokorrelation) am Empfänger (Cl_2 , Analyse 1) - x-Achse: Sekunden; y-Achse bei: 500

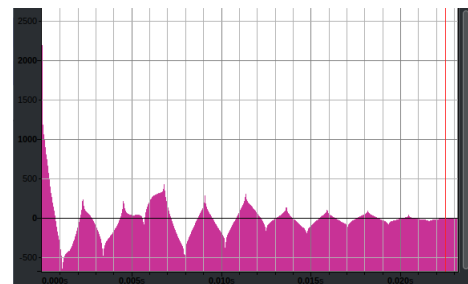


Figure 7. Audacity Spektrogramm (Standard-Autokorrelation) am Sender (Cl_1 , Analyse 1) - x-Achse: Sekunden; y-Achse bei: 2500

Standard Korrelation. Unterschiede

Auf der Empfängerseite werden höhere dominante Töne im Bereich von 100 bis 110 Hz gemessen, begleitet von weniger harmonischen Peaks, die durch abrupte Änderungen in den Werten sichtbar sind. Dies deutet darauf hin, dass die Komplexität des Audiosignals durch den Einsatz von Rausch- und Kerbfiltern reduziert wurde, was auf einen Informationsverlust während der Übertragung hinweist.

Im Gegensatz dazu zeigt die Senderseite eine niedrigere dominante Frequenz von 44 Hz. Die Kurven sind glatt und harmonisch, was auf eine höhere Komplexität des Audiosignals und einen höheren Informationsgehalt im Vergleich zur Empfängerseite hinweist.

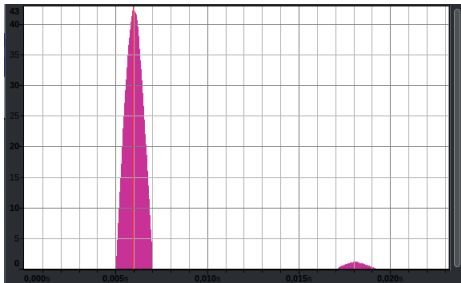


Figure 8. Audacity (Erweiterte-Autokorrelation) Spektrogramm am Empfänger (Cl_2 , Analyse 1) - x-Achse: Sekunden; y-Achse bei: 43

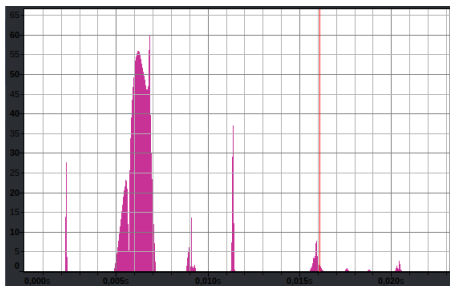


Figure 9. Audacity (Erweiterte-Autokorrelation) Spektrogramm am Sender (Cl_1 , Analyse 1) - x-Achse: Sekunden; y-Achse bei: 65

Erweiterte Autokorrelation. Auf der Empfängerseite ist im Vergleich zur Senderseite nur ein signifikanter Ausschlag (x-Achse: 0,005s) im selben Frequenzbereich zu erkennen. Kleinere Ausschläge außerhalb des Hauptausschlags wurden herausgefiltert, und der Hauptausschlag wurde geglättet.

Das Diagramm 9 (Sender) zeigt eine höhere Dichte an Frequenzen im niedrigeren Bereich, während das Diagramm 8 (Empfänger) einen sehr starken Peak bei 134 Hz aufweist und ansonsten eine relativ geringe Frequenzdichte zeigt.

Dieses Ergebnisse gelten ebenfalls für den erneut ausgeführten Kontrolltest.

Differenzaudio. Dieses Differenzaudio wird aus dem Versuch R_4 hergestellt. Im 2. Versuch R_5 wird dieses Ergebnis auch bestätigt.

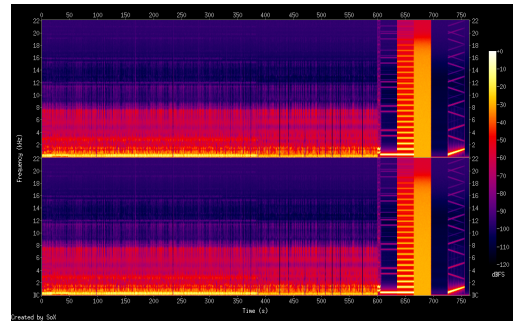


Figure 10. SOX Spektrogramm des Differenzaudios (1, R_4)

Das vorliegende Diagramm 10 repräsentiert eine inverse Telefonie-Aufnahme, bei der alle im Telefongespräch üblichen Frequenzen herausgefiltert wurden. Daher wirkt die Audioaufnahme unnatürlich und verzerrt.

Einige Frequenzen (in der Sprache) werden entfernt, da die "Vorlesung" im Differenzaudio hörbar und im Diagramm 2 erkennbar ist.

Die synthetischen Klänge werden von der Software Zoom-X nahezu vollständig eliminiert und sind daher in der Differenzaudio vorhanden.

Im Teil der Sprachaufnahme zeigen sich reduzierte Ausschläge in den Frequenzen bis 8 kHz.

Im Spektrogramm des Differenzaudio sind keine wiederkehrenden Muster oder ähnliche Artefakte zu erkennen. Somit lässt sich drauf schließen, dass (bisland) kein Wasserzeichen in der Datei vorhanden ist bzw. gefunden wurden ist.

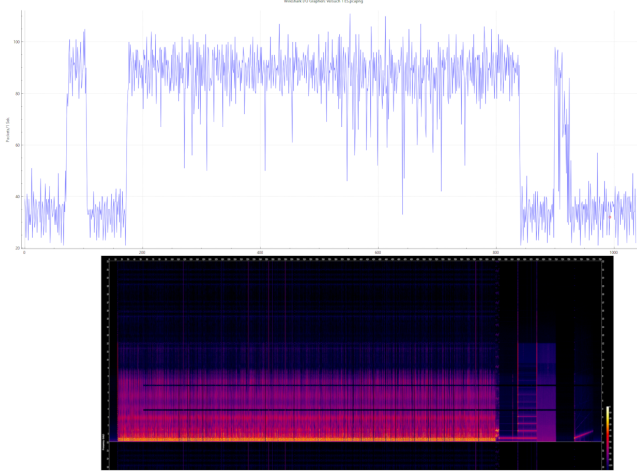


Figure 11. Übertragung von Audio über die Internetschnittstelle am Empfänger, R_1

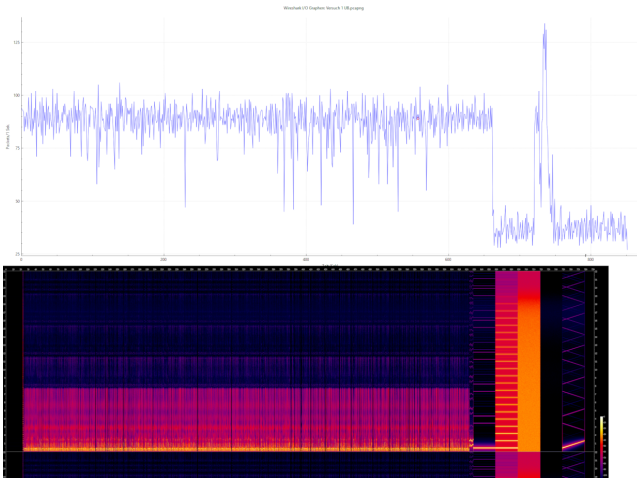


Figure 12. Übertragung von Audio über die Internetschnittstelle am Sender, R_1

Betrachtung verschlüsselter Kanal. Die R_1 Diagramme 11 und 12 stellen die Audiodatei mit dem Sendekanal von Zoom-X dar. Dabei wurden die Diagramme so über einander gelegt, sodass die Zeit-Achse mit dem:

- Audioinhalt der Testdatei (für die deutsche Sprache) mit dem
- Sendeverhalten an dem Zoom-X-Audio Server (MCS_{AU}) übereinstimmt.

Senderseitig (CL_1) wird festgestellt, dass die synthetischen Klänge herausgefiltert werden. Der Kerbfilter wird daher bereits am CL_1 angewendet und nicht erst auf den Zoom-X-Servern.

Empfängerseitig (CL_2) stimmt dies mit dem Sender überein, sodass davon ausgegangen werden kann, dass die Datei, die

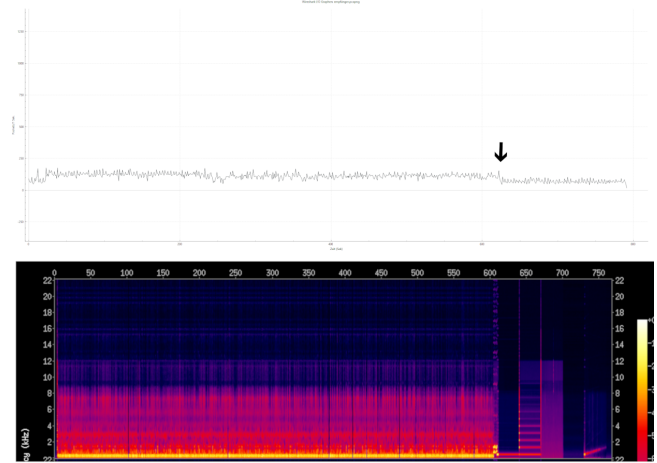


Figure 13. Übertragung von Audio über die Internetschnittstelle am Empfänger, R_{10} , Update Zyklus von Windows enthalten.

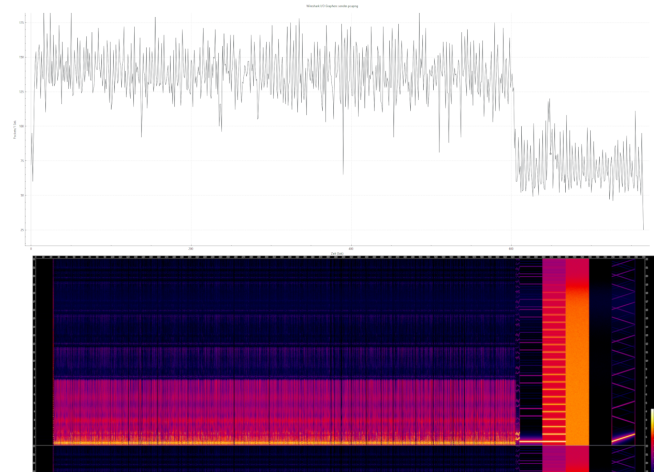


Figure 14. Übertragung von Audio über die Internetschnittstelle am Sender, R_{10}

vom Sender gesendet wurde, beim Empfänger unverändert ankommt.

Im Kontrolltest (R_{10}) wurde die erste Feststellung von R_1 bestätigt. Die Diagramme (13 und 14) belegen das gleiche Verhalten.

6.2 Test der anderen Sprecher und anderen Sprachen

Es wird festgestellt, dass in den SOX-, sowie Audacity-Spektrogrammen (R_6 bis R_9 , sowie R_1) keine nennenswerten Unterschiede erkennbar sind. Zur Veranschaulichung sind die entsprechenden Spektrogramme (enUS, enGB) hier abgebildet. Diese Ergebnisse gelten auch für den deutschen Text.

Beide Diagramme 17, 18 haben gleiche x- und y-Achsen.

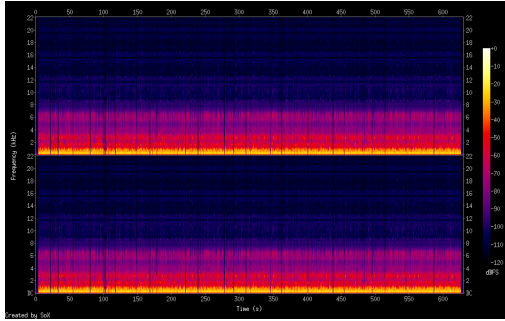


Figure 15. SOX Spektrogramm des Empfängers (Cl_2) in en-US, R_6)

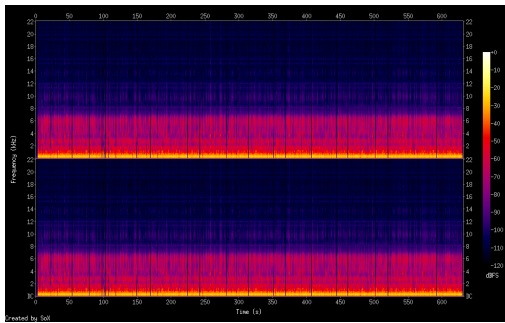


Figure 16. SOX Spektrogramm des Empfängers (Cl_2) in en-GB, R_8

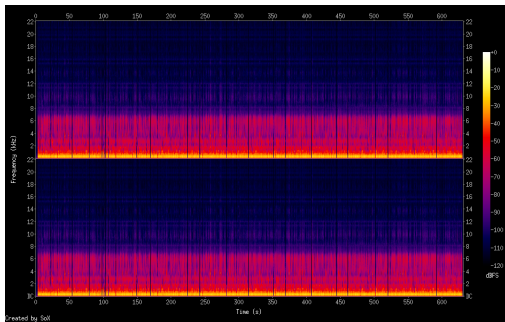


Figure 17. SOX Spektrogramm des Senders (Cl_1) in en-GB, R_8

Es wird festgestellt, dass in den SOX-, sowie Audacity-Spektrogrammen (R_6 bis R_9 , sowie R_1) keine nennenswerten Unterschiede (zu diesen Spektrogrammen in diesen Kapitel) erkennbar sind. Diese Ergebnisse gelten auch für den deutschen Text.

6.3 Experiment: Kerbfilter Reproduktion

Das SOX-Spektrogramm 19 zeigt die Reproduktion des Kerbfilters (als Differenzaudio) in Audacity. Dabei ist erkennbar, dass Informationen entfernt werden. Die Frequenzen der Sprachaufnahmen reichen bis etwa 7,8 kHz, wobei die intensivsten sprachlichen Klänge bei 3 kHz enden.

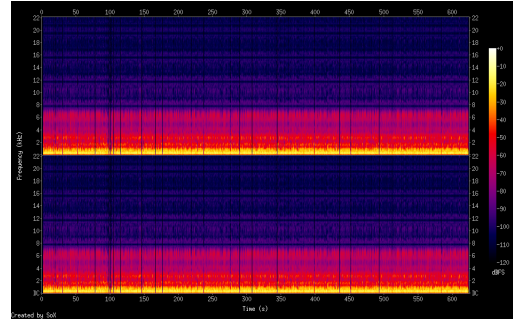


Figure 18. SOX Spektrogramm des Senders (Cl_1) in en-US, R_6

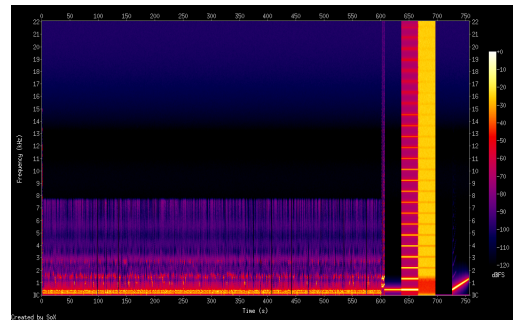


Figure 19. SOX Spektrogramm der Reproduktion des Rausch und Kerbfilters in Audacity

Das Differenzaudio wird aus der Original-Audio Datei, welche zum übertragen genutzt wird und der manuell bearbeiteten Datei (mit Rauschfilter) gebildet.

In der Hörprobe ähnelt das Ergebnis weitgehend dem Audio (aus Kapitel 6.1), wie es beim Empfänger ankommt. Dabei wurden mehr sprachliche Klänge herausgefiltert als bei der Verwendung von Zoom-X, was auf einen besseren Rauschunterdrückungsalgorithmus bei Zoom-X im Vergleich zu Audacity hinweist. Die synthetischen Klänge wurden bei Audacity nicht vollständig vom Filter entfernt, sondern nur abgeschwächt. Dies bestätigt die These, dass in Zoom-X kein Wasserzeichen vorhanden ist, da dieses Experiment die Beobachtungen der vorhergehenden Analysen unterstützt auf die Anwendung eines Kerbfilters/Rauschfilters.

Die im Kapitel 6.1 gesehenen "Schlieren" im Diagramm sind nicht zu erkennen, da das Programm Virtual Audio Cable, welches zum Transport der Audio Daten dient, nicht benutzt wurde. Somit ist diese Fehlerquelle nicht existent.

Vergleich

Die Reproduktion und die empfangenen Daten (mit Ausnahme der fehlerhaften ersten Übertragung R_1) ähneln sich. Beide weisen die "Artefakte" des Kerbfilters auf und bei der Sprache die gleichen Eliminierungen auf.

Das Audacity-Spektrogramm zeigt ein ähnliches Bild, jedoch wurden Geräusche und Störgeräusche nicht vollständig

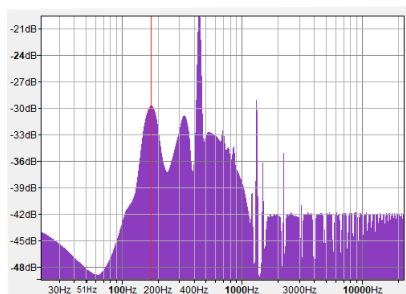


Figure 20. Audacity Spektrogramm - x-Achse: Herz; y-Achse bei: -21db

herausgefiltert. Die Störgeräusche bei 445 Hz, 1324 Hz und 2199 Hz sind weiterhin vorhanden, im Gegensatz zu den Ergebnissen bei der Verwendung von Zoom-X (vgl. Diagramm 5).

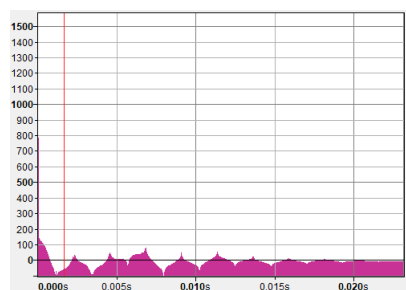


Figure 21. Audacity Spektrogramm (Standard-Autokorrelation) - x-Achse: Sekunden; y-Achse bei: 1500

Das Audacity-Spektrogramm zeigt am Empfänger niedrige, gezackte Ausschläge, die auf Störgeräusche hinweisen. Diese Ausschläge sind ein charakteristisches Merkmal und deuten darauf hin, dass bestimmte Frequenzen nicht vollständig herausgefiltert wurden.

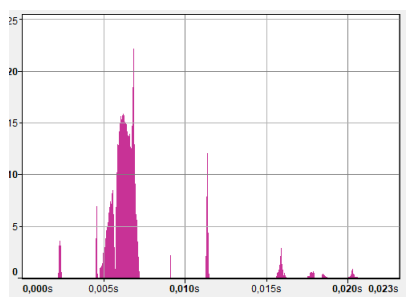


Figure 22. Audacity Spektrogramm (Erweiterte-Autokorrelation) - x-Achse: Sekunden; y-Achse bei: 25

In der erweiterten Autokorrelation zeigen sich Spitzen bei 447 Hz, 147 Hz, 88 Hz, 63 Hz, 56 Hz, 54 Hz und 49 Hz, die von

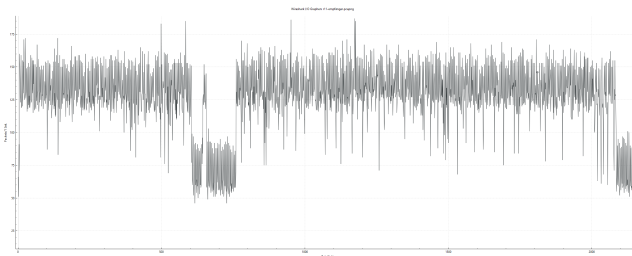


Figure 23. R_{11} I/O Graph am Empfänger

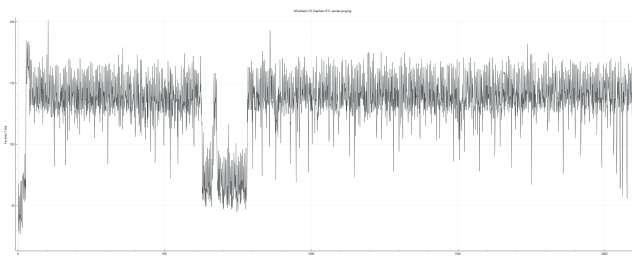


Figure 24. R_{11} I/O Graph am Sender

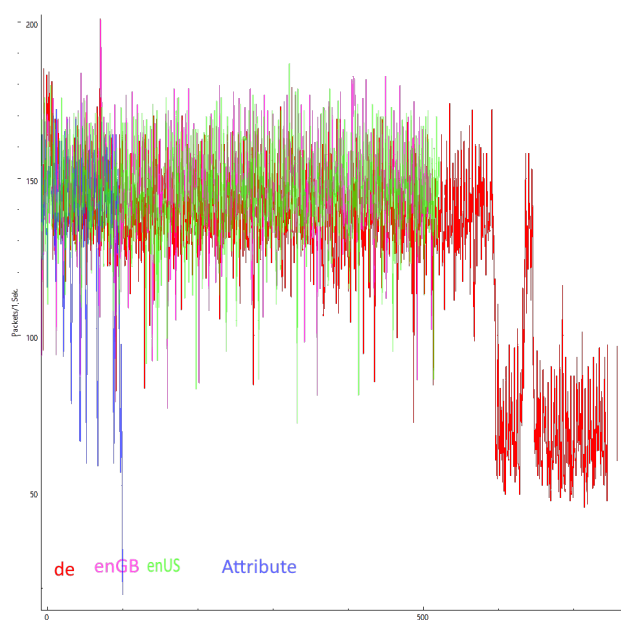


Figure 25. R_{11} Übereinander gelegter I/O Graph am Sender

Zoom-X herausgefiltert wurden. Zudem ist die Hauptspitze bei 1432 Hz kantig und nicht abgerundet wie im initialen Test.

6.4 Verschlüsselter Kanal bei verschiedenen Sprachen

Das Diagramm 26 zeigt, dass die Ausschläge im amerikanischen (grüner Graph) und britischen (pinker Graph) Englisch

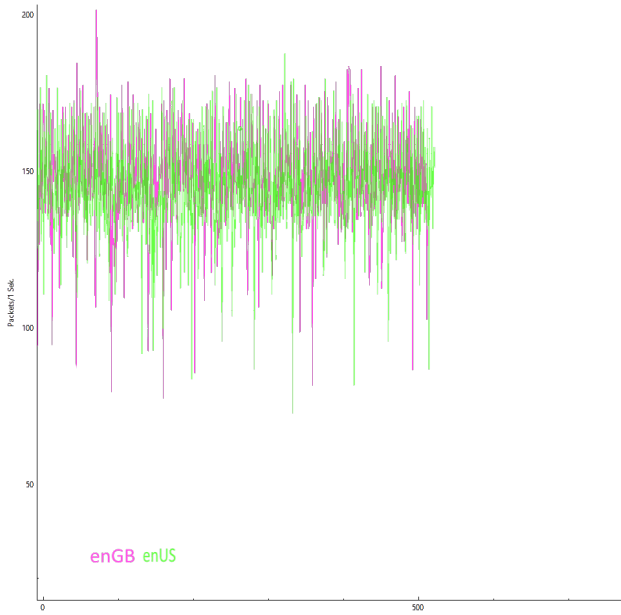


Figure 26. R_{11} enGB, enUS übereinander gelegter I/O Graph

weitgehend ähnlich verlaufen. Die zeitlichen Abstände zwischen den Ausschlägen beider Sprachvarianten betragen dabei maximal nur wenige Sekunden. Somit zeigt sich ein ähnliches Sendeverhalten.

Wird der deutsche Graph (roter Graph in Figur 25) hinzugezogen, zeigt sich, dass Zoom-X in den üblichen Sendepausen der beiden englischen Sprachvarianten auch im Deutschen Sendepausen einlegt. Allerdings treten im Deutschen insgesamt mehr Sendepausen (seitens Zoom-X) auf als in den englischen Varianten. Abgesehen davon ähneln sich das Sendeverhalten von Zoom-X.

Im Attribute-Projekt [3] (blauer Graph im 25) zeigen sich mehr Pausen bei den natürlichen Sprechern im Vergleich zu den synthetischen Sprechern der vorherigen Audiofiles. Die bietet potenziell eine Möglichkeit zur Identifizierung eines Sprecher.

Zwischen dem I/O-Graphen am Sender (24) und Empfänger (23) sind keine Unterschiede erkennbar, abgesehen von der Vorlaufzeit (der Wireshark Aufnahme).

Daraus lässt sich folgern:

- Bei synthetischen Sprechern sind kaum Unterscheidungen möglich.
- Bei natürlichen Sprechern könnten die Sprechpausen der Sprecher Hinweise auf Sprecher liefern.

6.5 Fazit bzw. neue Funde zum alten Paper

Zusammenfassend ist festzustellen:

- der Kerbfilter wird auf Senderseite angewendet
- der Zoom-X-Kerbfilter ist sehr gut
- Wasserzeichen sind bislang nicht gefunden wurden
- die Unterscheidungen (bzw. Identifizierung) von Sprachen im Spektrogramm ist **nicht** möglich
- die Unterscheidungen (bzw. Identifizierung) von Sprechern in Spektrogramm ist **nicht** möglich
- die Unterscheidungen (bzw. Identifizierung) von synthetischen Sprechern und Sprachen über Schnittstelle **kaum** möglich
- die Unterscheidungen (bzw. Identifizierung) von natürlichen Sprechern ist möglich, aufgrund individuellen Sprechweisen
- es wird derselbe Kerbfilter/Rauschunterdrückungsalgorithmus für alle Sprachen angewendet

7 Literaturverzeichnis

References

- [1] Telekom AG. 2024. *Zoom X für Ihr Business | Telekom Geschäftskunden*. Retrieved 17.07.2024 from <https://geschaeftskunden.telekom.de/magenta-business-collaboration/video-konferenzen/zoom-x>
- [2] Robert Altschaffel, Jonas Hielscher, Stefan Kiltz, and Jana Dittmann. 2021. *Meta and Media Data Stream Forensics in the Encrypted Domain of Video Conferences*. Retrieved 26.06.2024 from <https://dl.acm.org/doi/10.1145/3437880.3460412>
- [3] ATTRIBUT-Projekt. 2024. *ATTRIBUT-Projekt: „Was macht ihr da eigentlich genau? ist eine Fra....“*. Retrieved 07.08.2024 from <https://sparrow.cs.uni-magdeburg.de/@ATTRIBUT/112761785371442924>
- [4] Harold Baldwin. 1918. *Holding the line*. Retrieved 16.05.2024 from <https://www.gutenberg.org/cache/epub/73624/pg73624.txt>
- [5] VINCENT BUREL. 2024. *VB-Audio Virtual Apps*. Retrieved 16.05.2024 from <https://vb-audio.com/Cable/>
- [6] cbagwell, mansr, robs, and uklauer. 2024. *SoX - Sound eXchange*. Retrieved 16.05.2024 from <https://sourceforge.net/projects/sox/>
- [7] Jana Dittmann and Christian Krätzer. 2024. *Vorlesungsfolien 'IT-Forensik'*. FIN/OVGU Magdeburg, Sommersemester 2024.
- [8] Wireshark Foundation. 2024. *Download Wireshark*. Retrieved 16.05.2024 from <https://www.wireshark.org/download.html>
- [9] Eclipse Foundation. 2024. *Home | Adoptium*. Retrieved 16.05.2024 from <https://adoptium.net/de/>
- [10] Zoom Inc. 2024. *Release notes for Windows*. Retrieved 04.07.2024 from https://support.zoom.com/hc/en/article?id=zm_kb&sysparm_article=KB0068823
- [11] Oracle. 2023. *Downloads – Oracle VM VirtualBox*. Retrieved 16.05.2024 from <https://www.virtualbox.org/wiki/Downloads>
- [12] Audacity Team. 2024. *Audacity® Download*. Retrieved 16.05.2024 from <https://www.audacity.de/>
- [13] Chip.de Team. 2024. *Windows 10 ISO 64 Bit - Download - CHIP*. Retrieved 16.05.2024 from https://www.chip.de/downloads/Windows-10-ISO-64-Bit_81515537.html
- [14] ZAP Dev Team. 2024. *ZAP Download*. Retrieved 16.05.2024 from <https://www.zaproxy.org/download/>
- [15] Inc Telekom AG und Zoom Video Communications. 2024. *Zoom X für Ihr Business | Telekom Geschäftskunden*. Retrieved 16.05.2024 from <https://geschaeftskunden.telekom.de/magenta-business-collaboration/video-konferenzen/zoom-x>
- [16] Eufemia von Adlersfeld-Ballestrem. um 1900. *Amönenhof*. Retrieved 16.05.2024 from <https://www.projekt-gutenberg.org/adlersfe/amoenen/chap002.html>

8 Anhang

Definition: Kerbfilter / Notch-Filter. Ein Kerbfilter, auch als Notch-Filter bekannt, ist ein elektronischer Filter, der eine sehr schmale Frequenzbandbreite aus einem Signalspektrum entfernt oder stark dämpft, während die restlichen Frequenzen weitgehend unbeeinflusst bleiben.

(3) VCForensics: Meta and Media Data Stream Forensics - ZoomX

• Einleitung/Motivation

Videokonferenzsysteme (VC) nutzen Interaktionsmöglichkeiten (Handheben, Umfragen usw.), Text, Audio, Bilder und Video und andere Medien, um Kommunikation zu ermöglichen. Sie werden in Organisationen und Privatpersonen in zunehmendem Maße genutzt. Mit der zunehmenden Nutzung steigt auch der Bedarf Sicherheit und Datenschutz in diesen Videokonferenzsystemen zu bewerten und Videokonferenzsystemen.

• Aufgaben: Idee, Ansatz und Ausgangspunkt:

Ziel ist es, die Methodik und Untersuchungen nach <https://dl.acm.org/doi/10.1145/3437880.3460412> auf ZoomX anzuwenden, auf Spurensuche bezüglich Privacy zu gehen, Spurenlagen zu dokumentieren mittels des forensischen Prozessmodells nach Kiltz et al. und mit dem Stand im Paper zu vergleichen. Zur Wahrung des Datenschutzes sind möglichst computergenerierte Medien einzuspielen. Abstimmung mit JD und dem Betreuer ist wichtig!

Mögliche teaminterne Aufgabenaufteilung: Aufnahmen gemeinsam, Untersuchung aufgeteilt in unterschiedl. Mediendatenströme (Datenströme und biometrische Erkennungsmöglichkeiten: Interaktion-Text, Video, Audio).

Erwartetes Ergebnis:

- Untersuchungen der Gesamtkonfiguration und der einzelnen Datenströme wie Text, Tippverhlafen, Audio und Video sowie Interaktionen wie Handheben usw. Ggf erweiterte VC Möglichkeiten im Vergleich zu 2021 einbeziehen. Anzahl an Datenströmen und Aufzeichnungen mit den Betreuer und JD absprechen um einen ausreichenden Umfang und Signifikanz zu erzielen.
- Einordnung der Spurenlagen in das forensische Datenmodell und Spurenlagen tabellarisch und als Grafik.
- Dokumentation der Möglichkeiten und Grenzen der Untersuchung.
- Demonstratorinfrastruktur bestehend aus der Untersuchungsumgebung nach dem Papier mit ggf. Anpassungen für ZoomX sowie Bericht

Betreuer (Name und Kontakt):

Robert Altschaffel, Vertretung durch Stefan Kiltz

Team: 3-4 Studierende

Refs: Altschaffel, R.; Hielscher, J.; Kiltz, S.; Dittmann, J.: Meta and Media Data Stream Forensics in the Encrypted Domain of Video Conferences. In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, Virtual Event, Belgium, 22–25 June 2021; pp. 23–33. <https://dl.acm.org/doi/pdf/10.1145/3437880.3460412>



Figure 8.1 . Aufgabenstellung

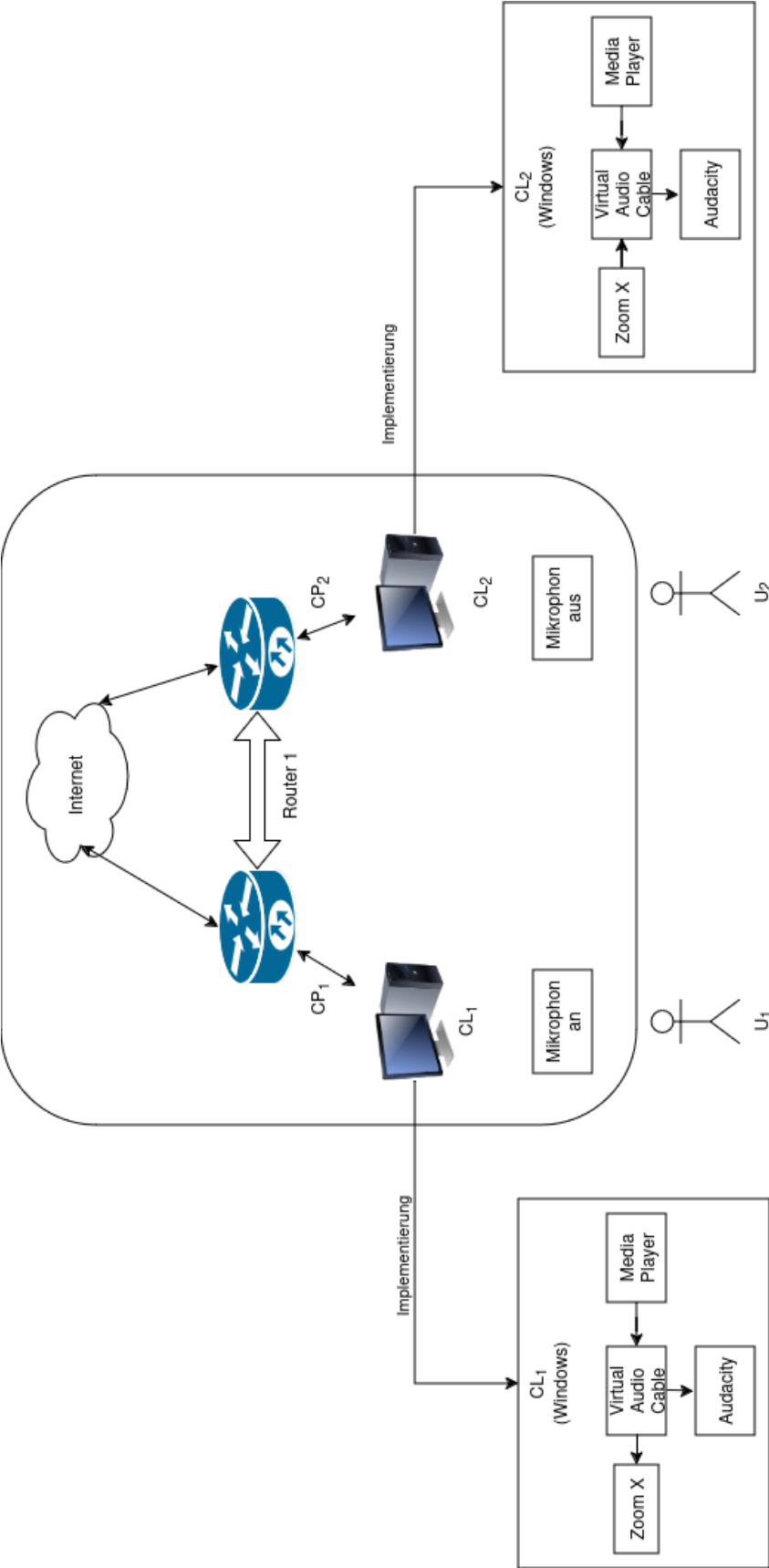


Figure 8.2 . Aufbau des (Audio)-Versuches

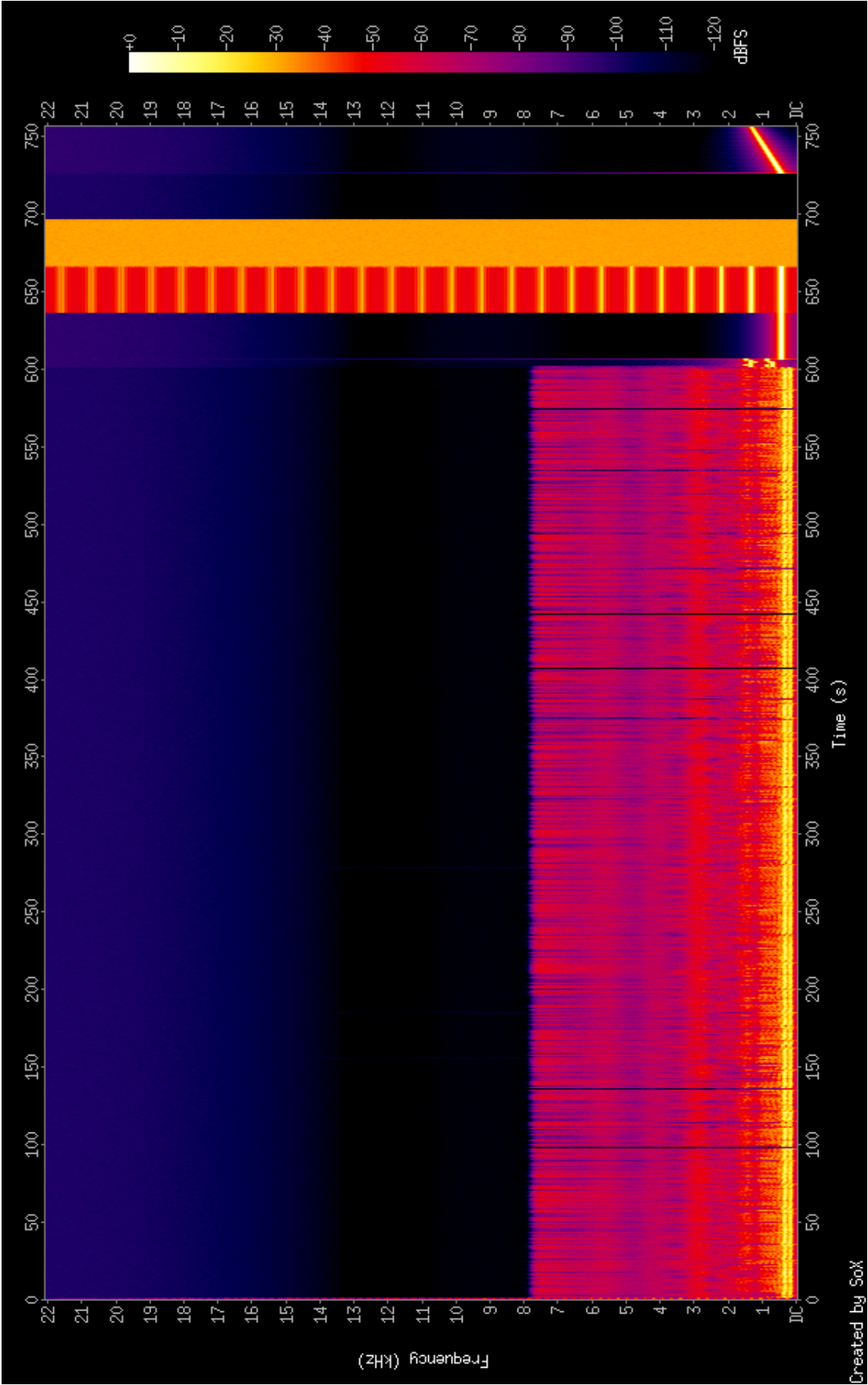


Figure 8.3 . SOX Spektrogramm der Raw-Data DT_1

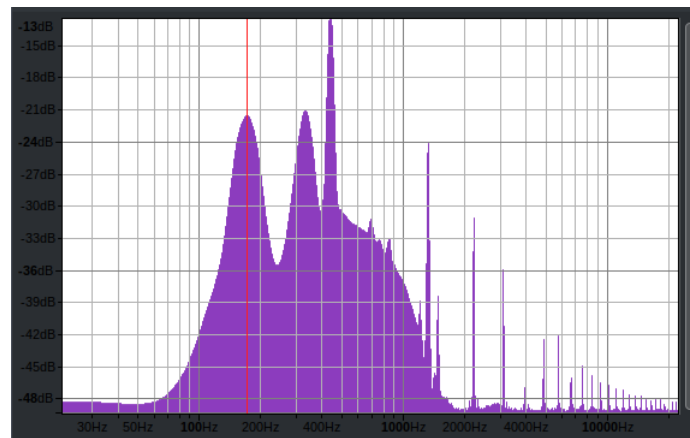


Figure 8.4 . Audacity Spektrogramm der Raw-Data DT_1

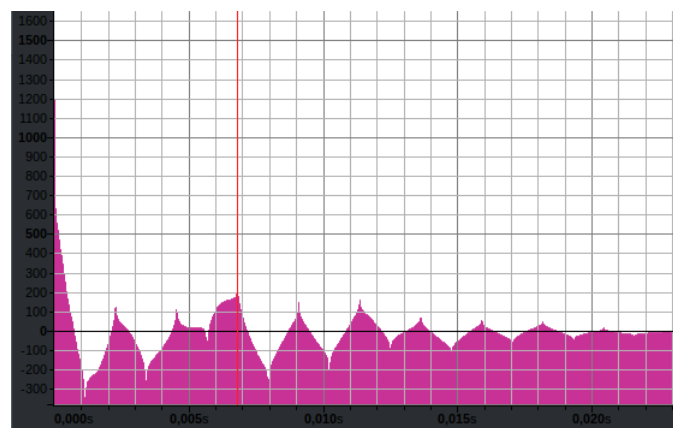


Figure 8.5 . Audacity Spektrogramm (Standard-Autokorrelation) Raw-Data DT_1

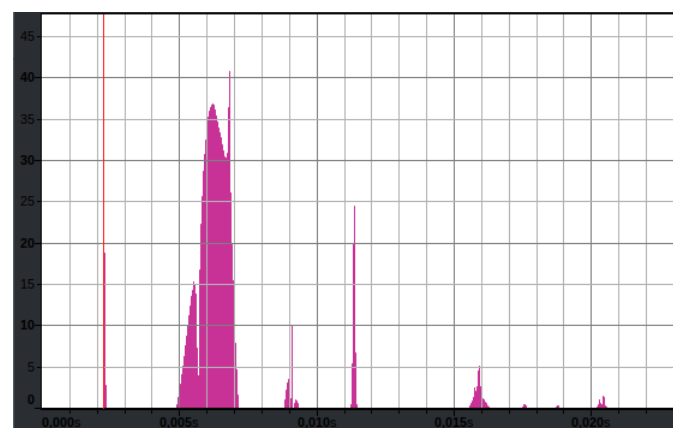


Figure 8.6 . Audacity (Erweiterte-Autokorrelation) Raw-Data DT_1

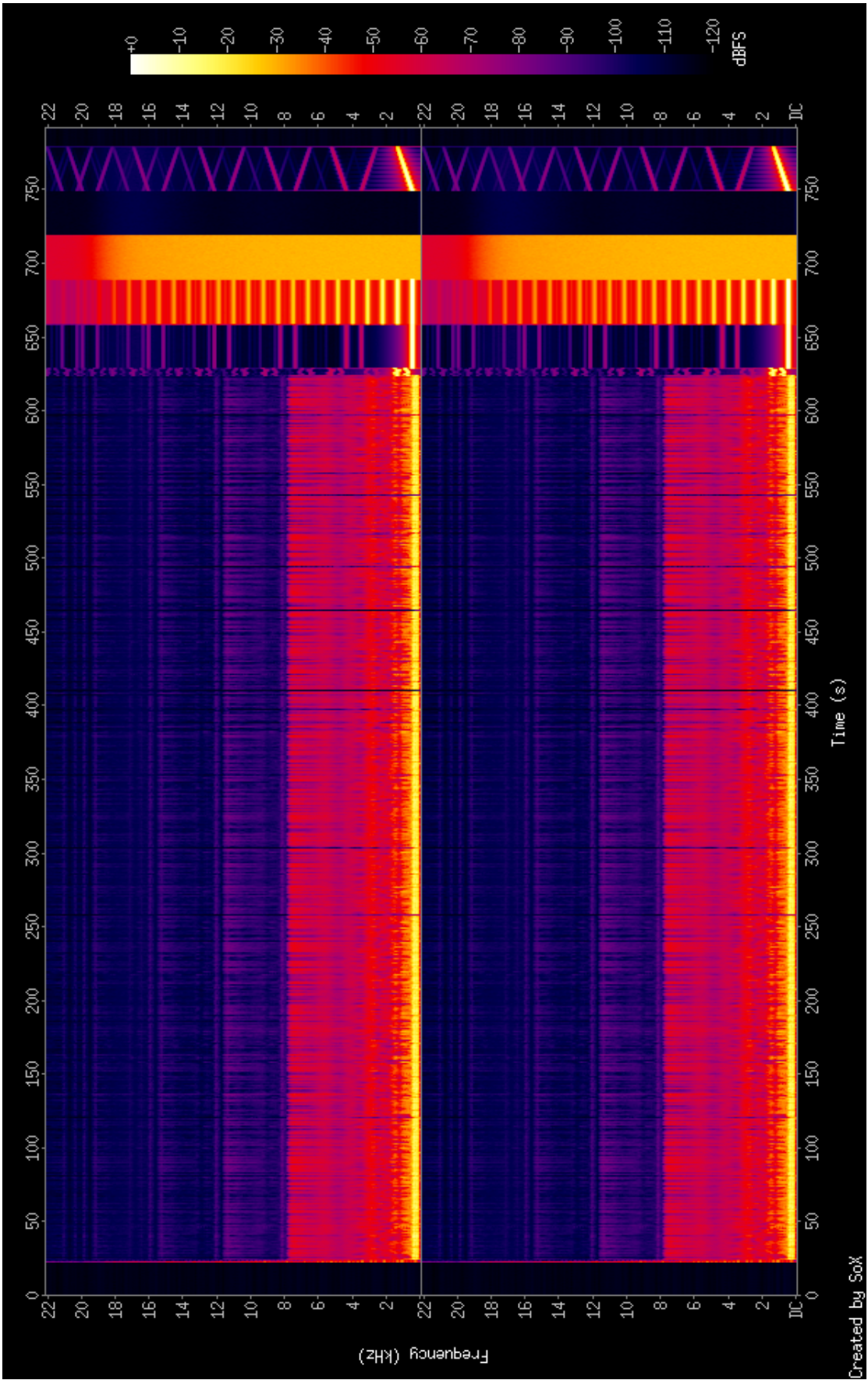


Figure 8.7 . SOX Spektrogramm am Sender (Analyse 1), R_1

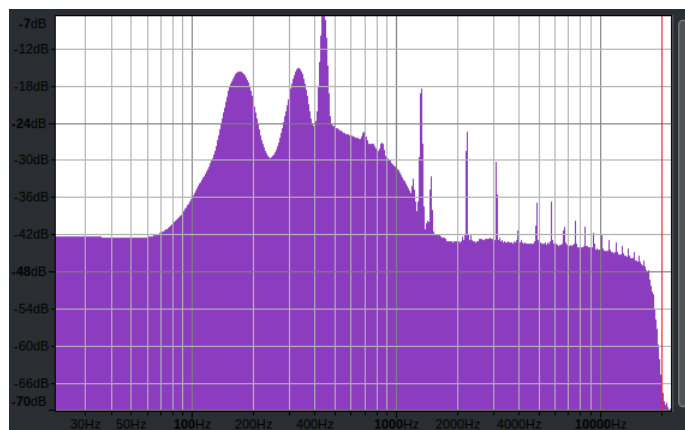


Figure 8.8 . Audacity Spektrogramm am Sender (Analyse 1), R_1

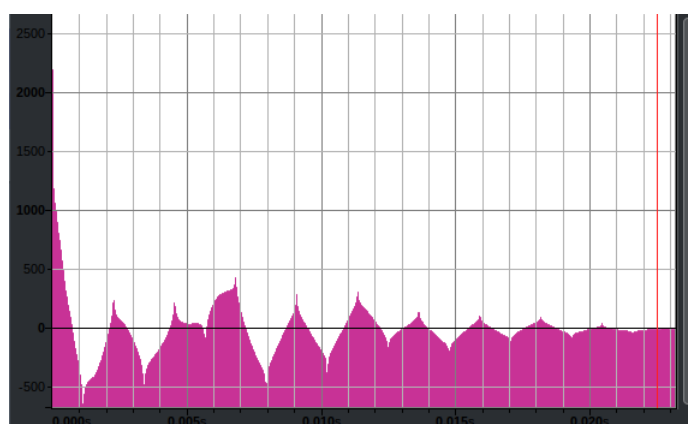


Figure 8.9 . Audacity Spektrogramm (Standard-Autokorrelation) am Sender (Analyse 1), R_1

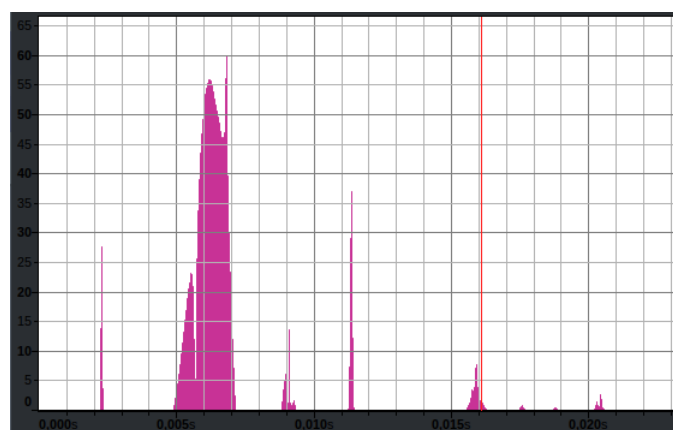


Figure 8.10 . Audacity (Erweiterte-Autokorrelation) Spektrogramm am Sender (Analyse 1), R_1

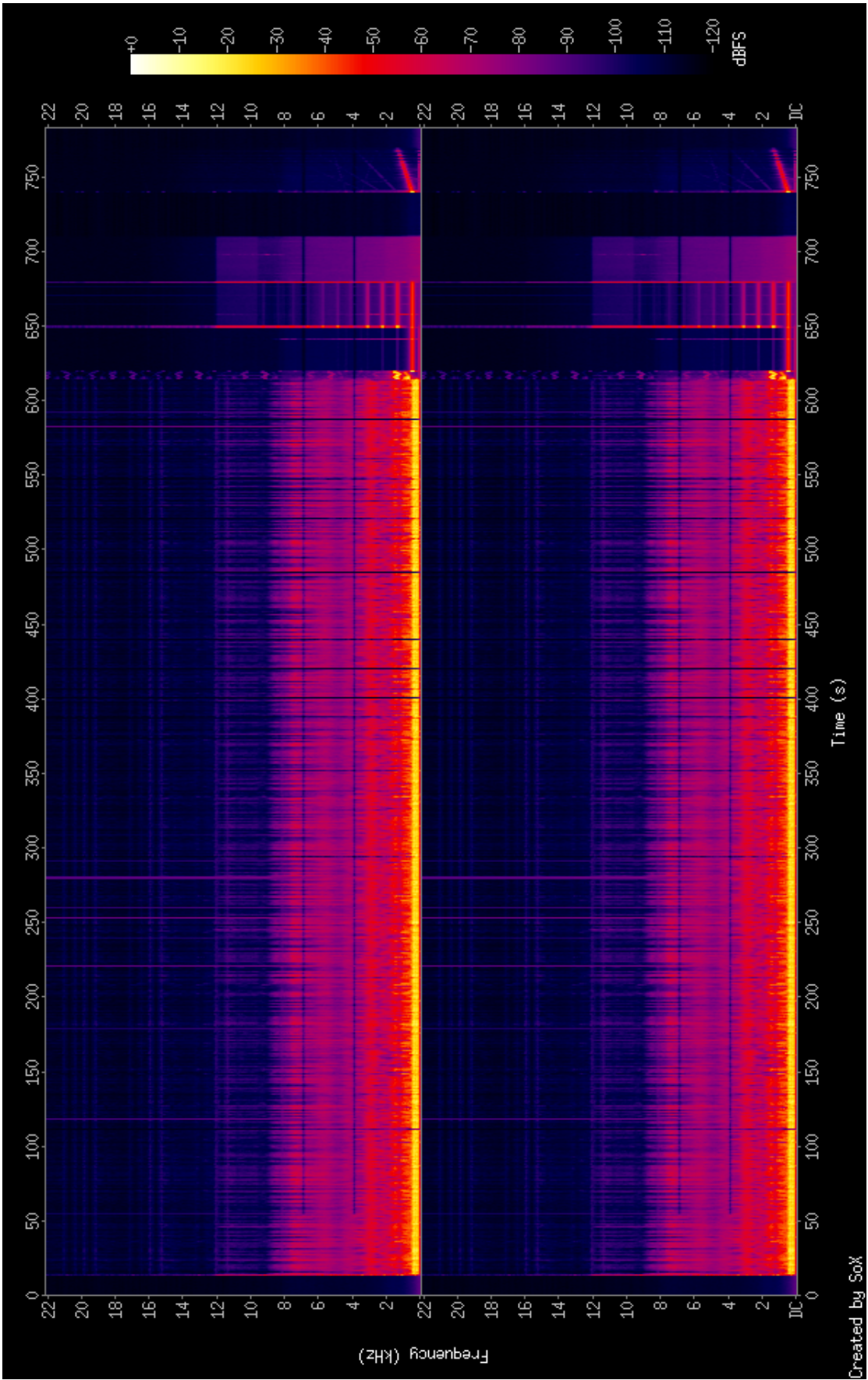


Figure 8. 11 . SOX Spektrogramm am Empfänger (Analyse 1), R_1

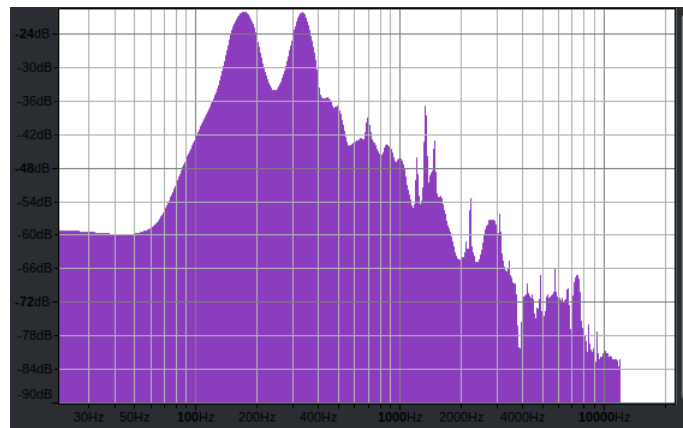


Figure 8. 12 . Audacity Spektrogramm am Empfänger (Analyse 1), R_1

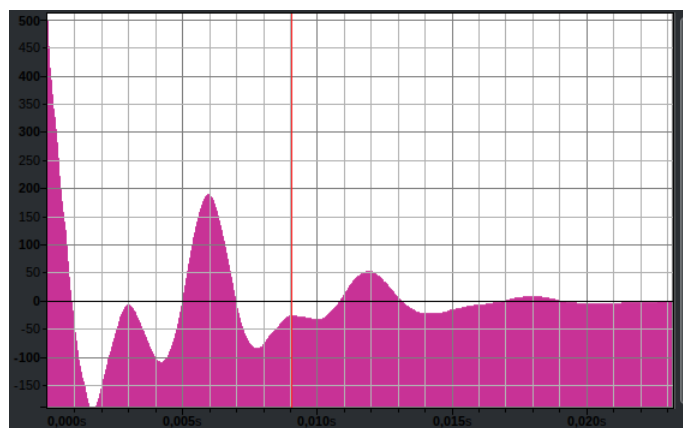


Figure 8. 13 . Audacity Spektrogramm (Standard-Autokorrelation) am Empfänger (Analyse 1), R_1

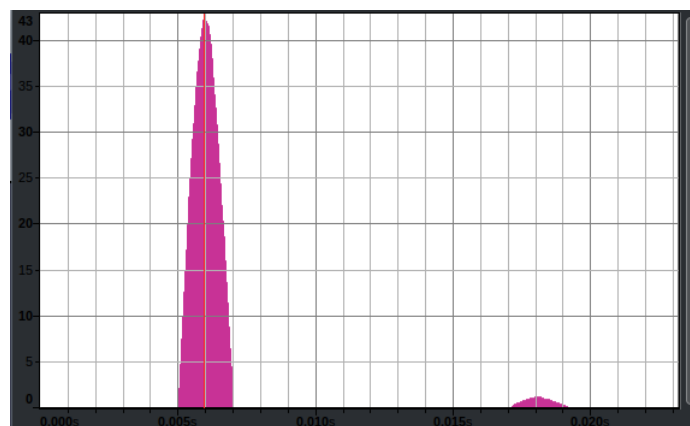


Figure 8. 14 . Audacity (Erweiterte-Autokorrelation) Spektrogramm am Empfänger (Analyse 1), R_1

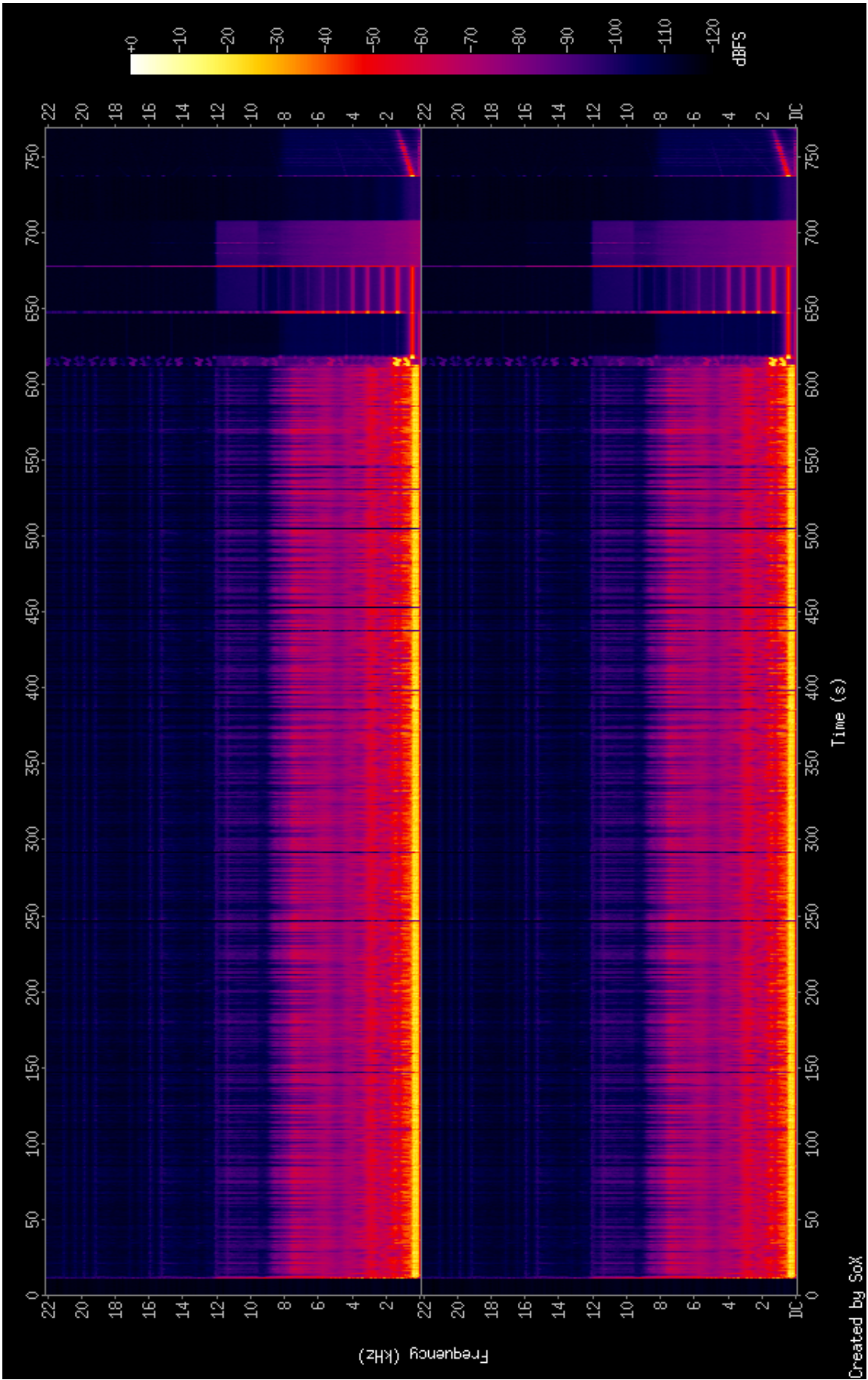


Figure 8.15 . SOX Spektrogramm PC 1 Empfängt (Empfangsspektrogramm), R3

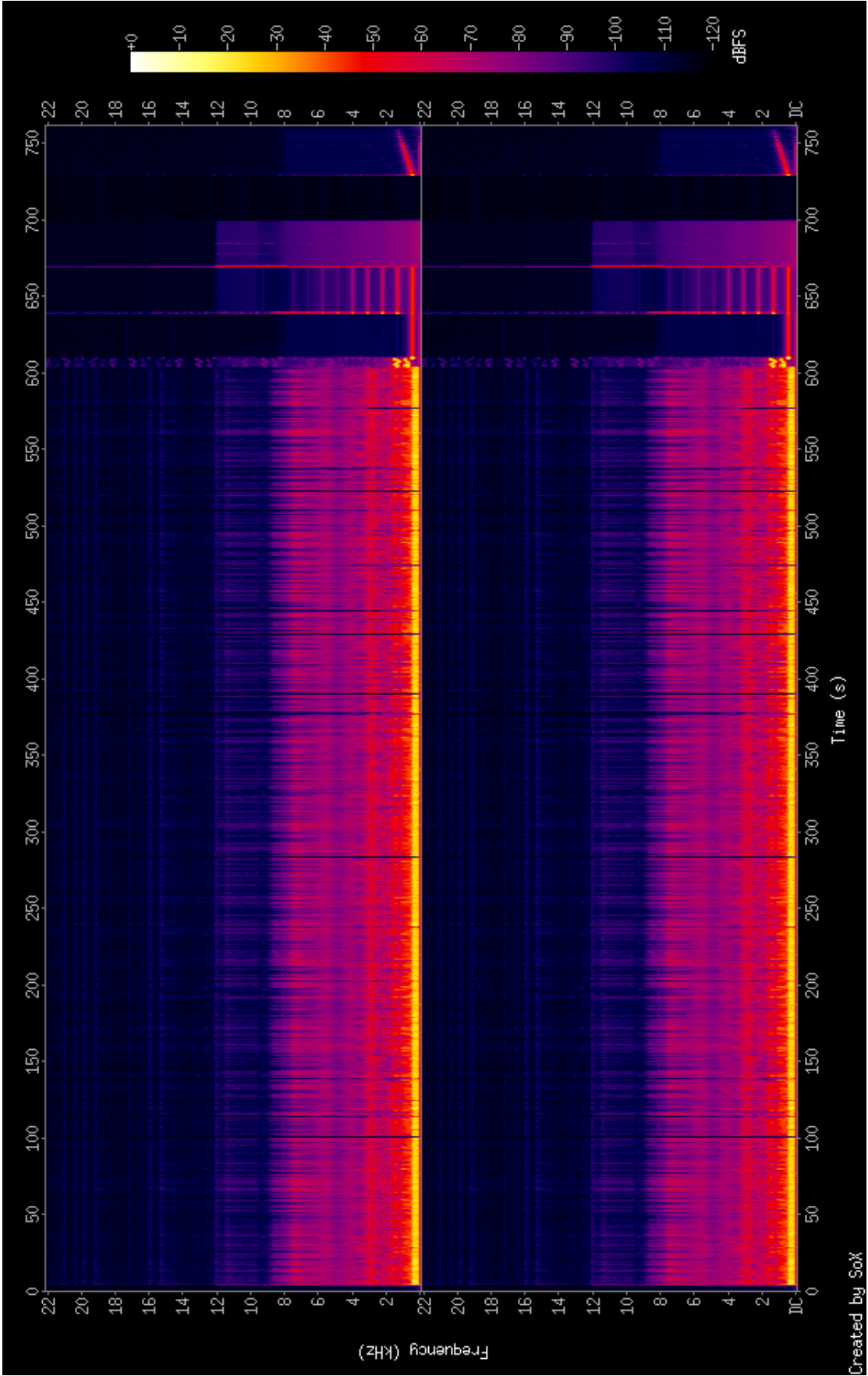


Figure 8. 16 . SOX Spektrogramm PC 2 Empfängt (Empfangsspektrogram), R_2

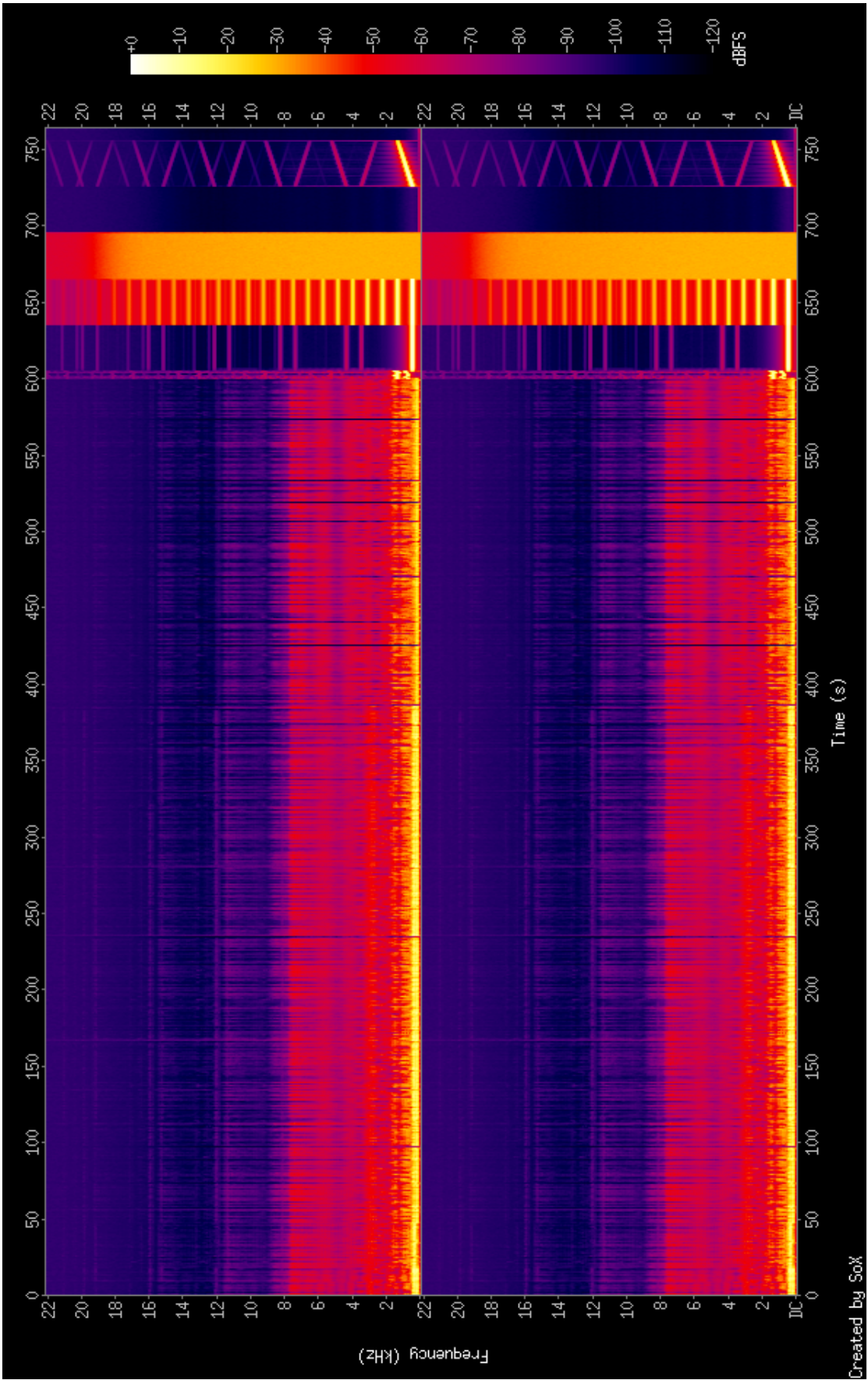


Figure 8. 17 . SOX Spektrogramm des Differenzdiagramm (Test 1, R_4)

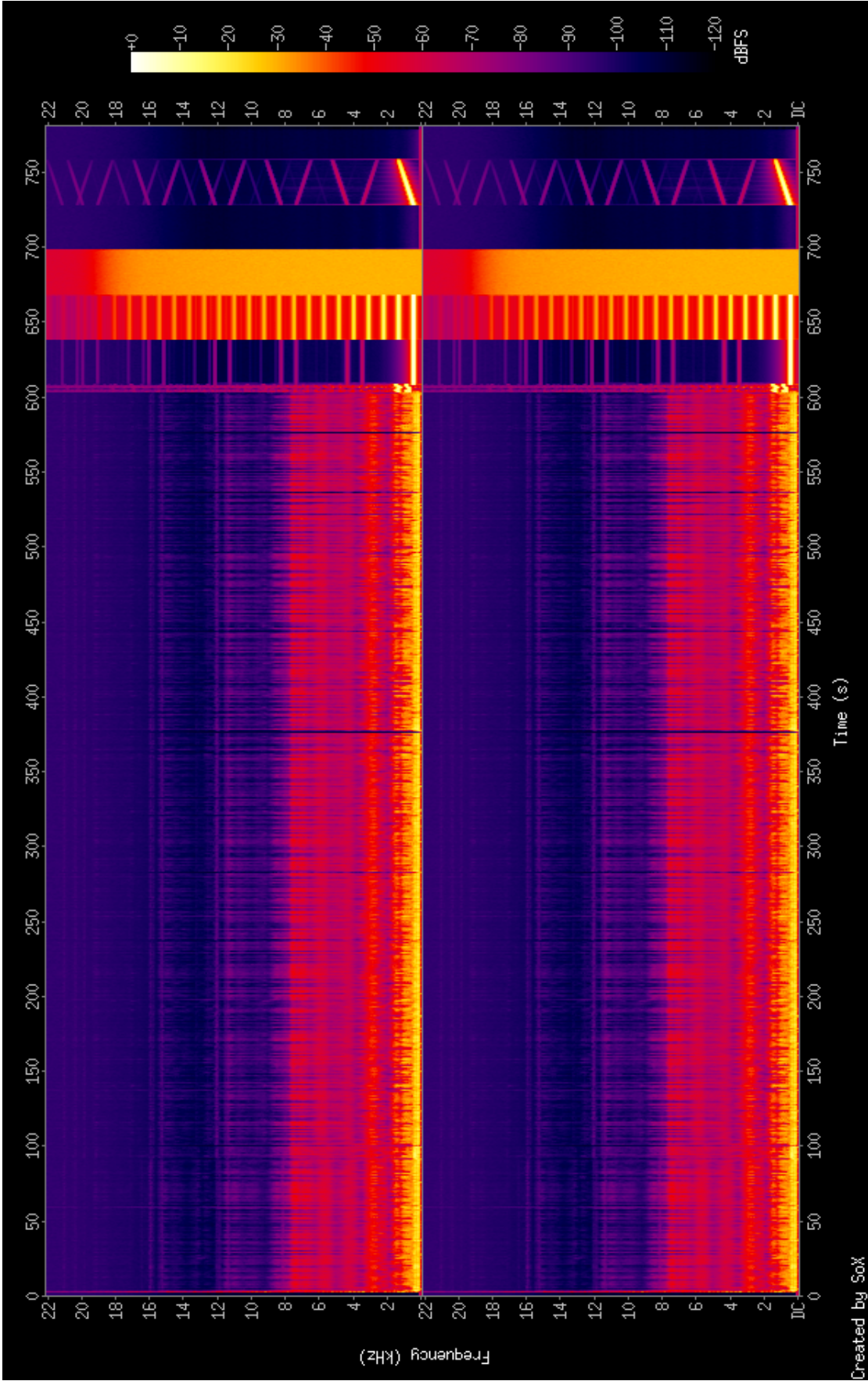


Figure 8. 18 . SOX Spektrogramm des Differenzdiagramm (Test 2, R_5)

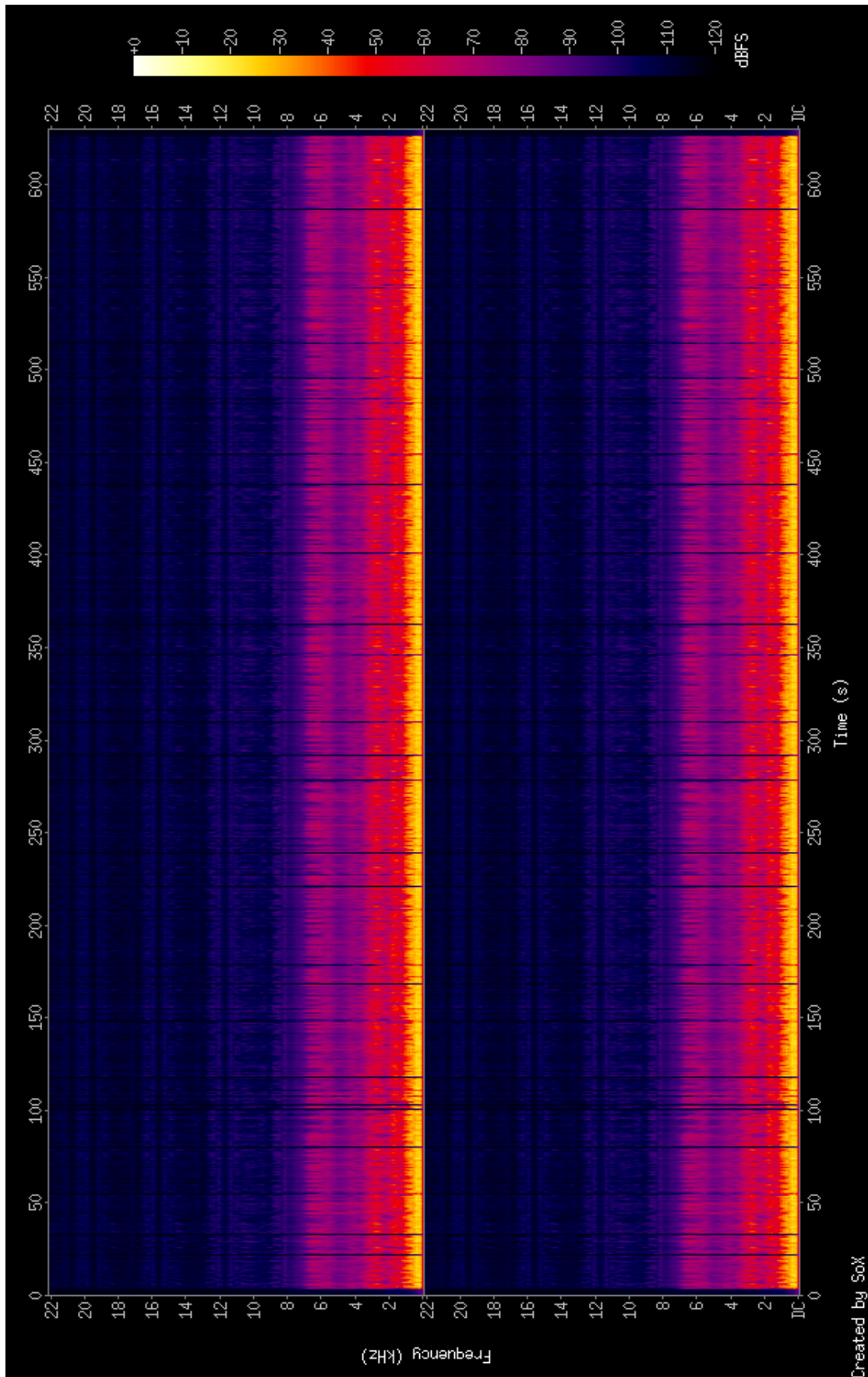


Figure 8. 19 . SOX Spektrogramm der englischen Sprache (enUS) des Empfängers, R_6)

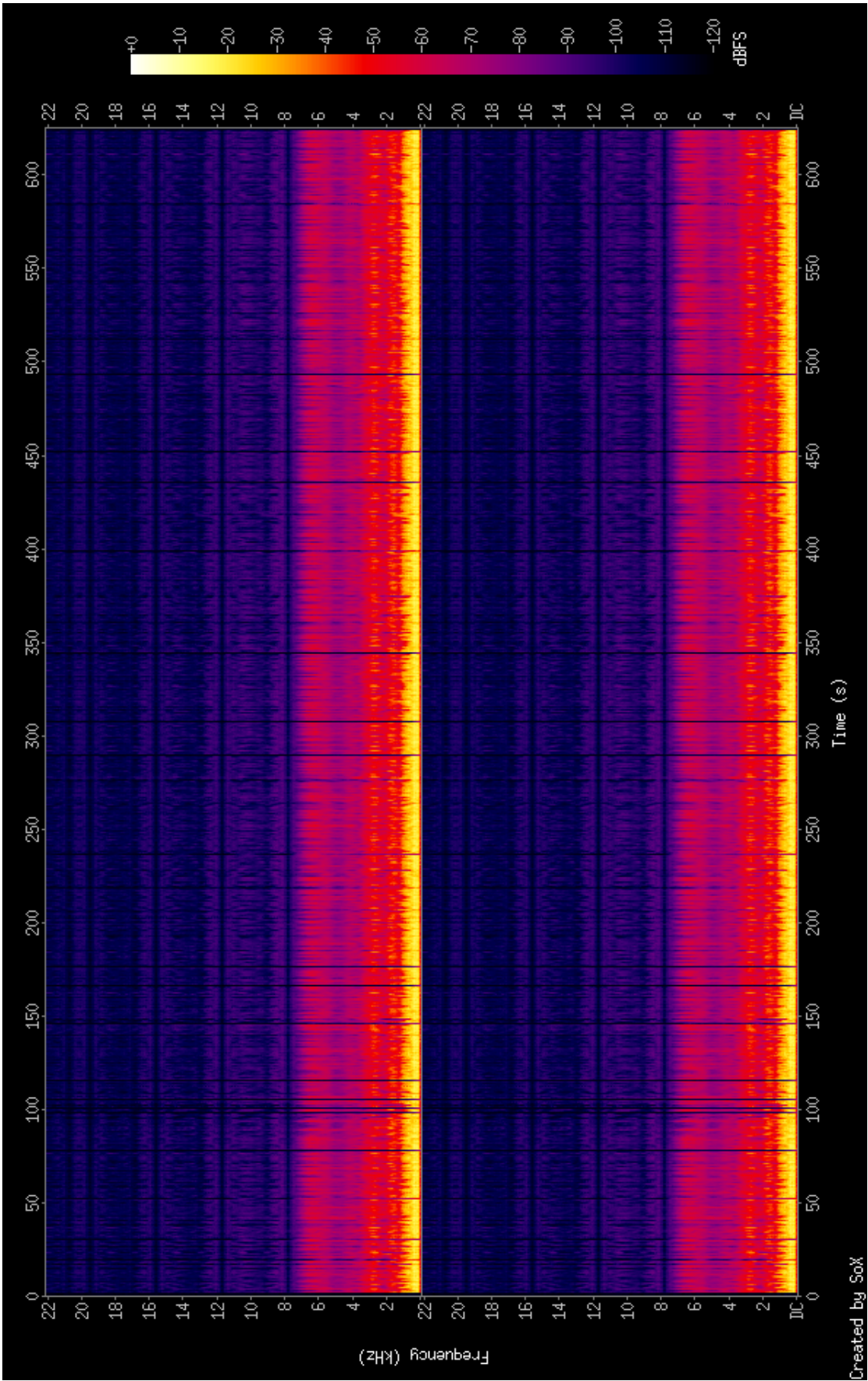


Figure 8. 20 . SOX Spektrogramm der englischen Sprache (enUS) des Senders, R_6

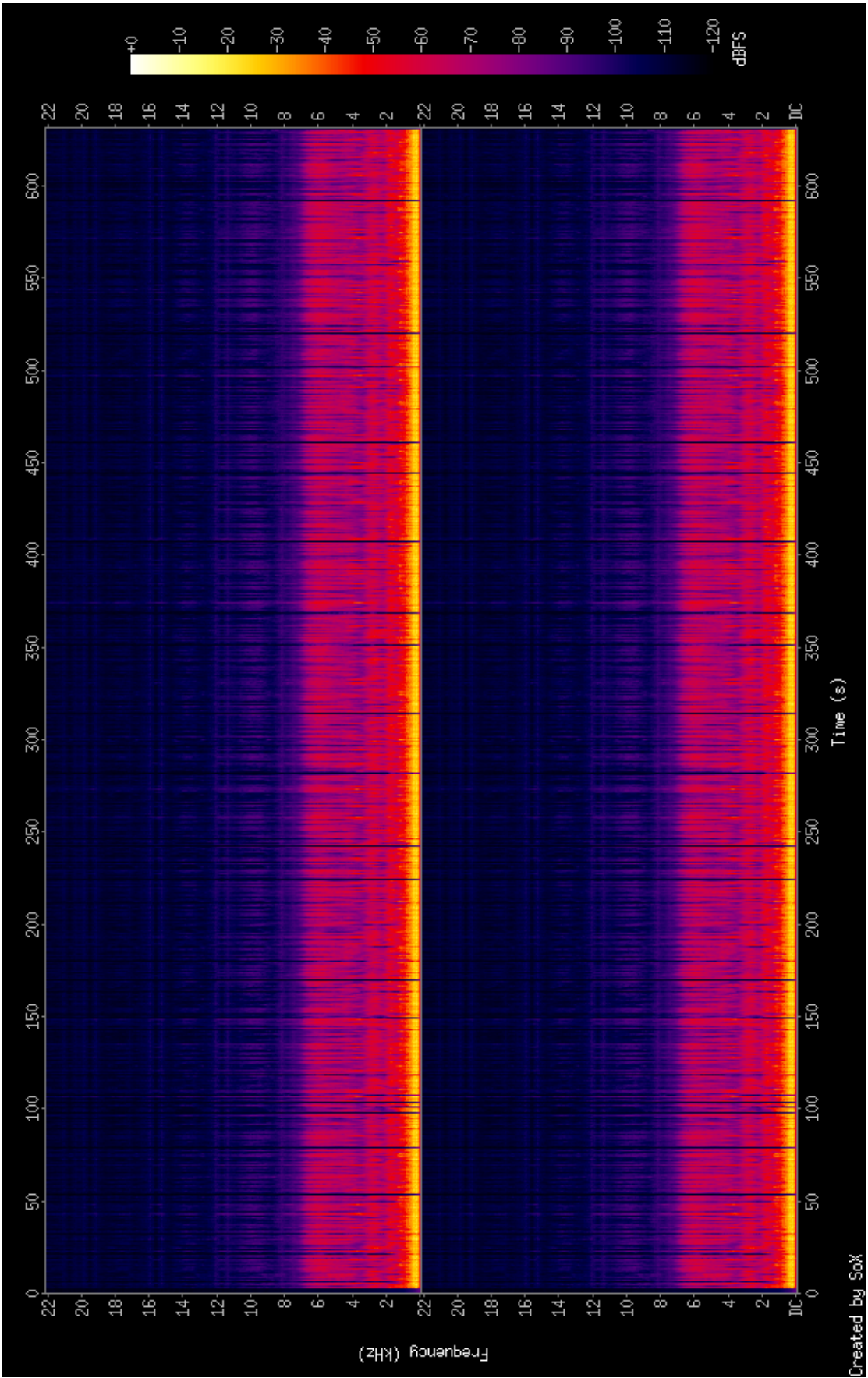


Figure 8. 21 . SOX Spektrogramm der englischen Sprache (enGB) am Empfängers, R₈

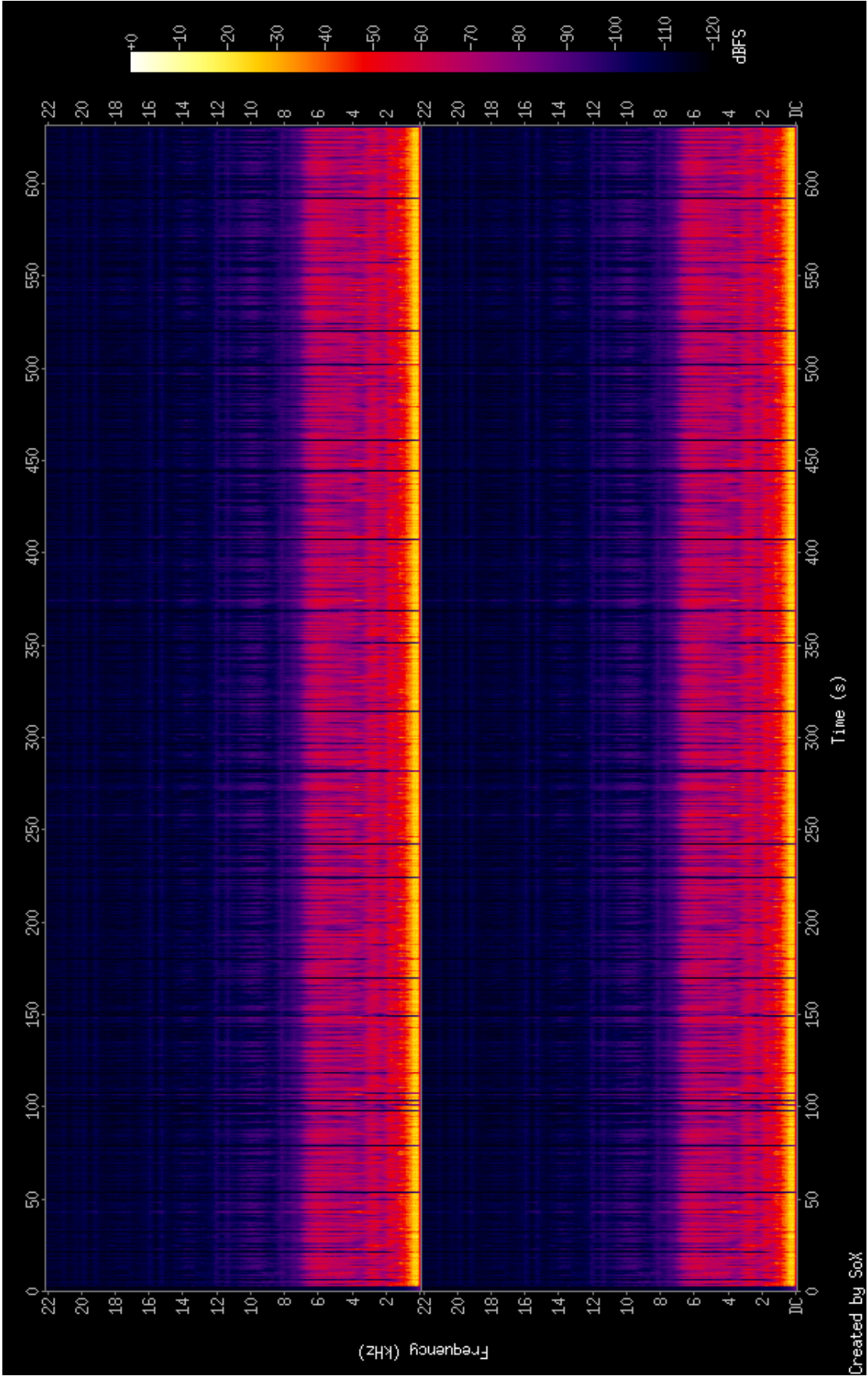


Figure 8. 22 . SOX Spektrogramm der englischen Sprache (enGB) des Senders, R_8

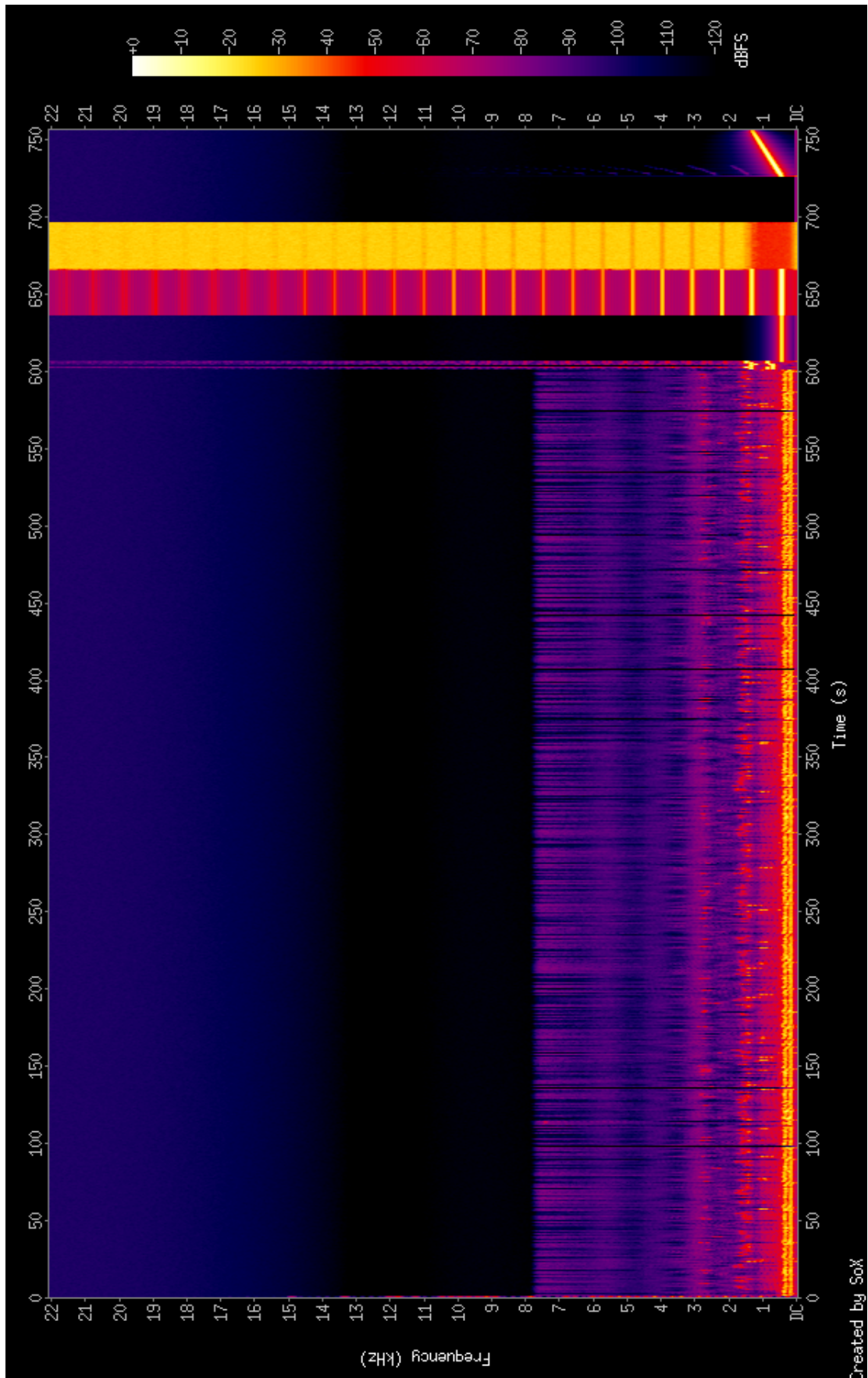


Figure 8. 23 . SOX Spektrogramm der Reproduktion des Rausch und Kerbfilters in Audacity

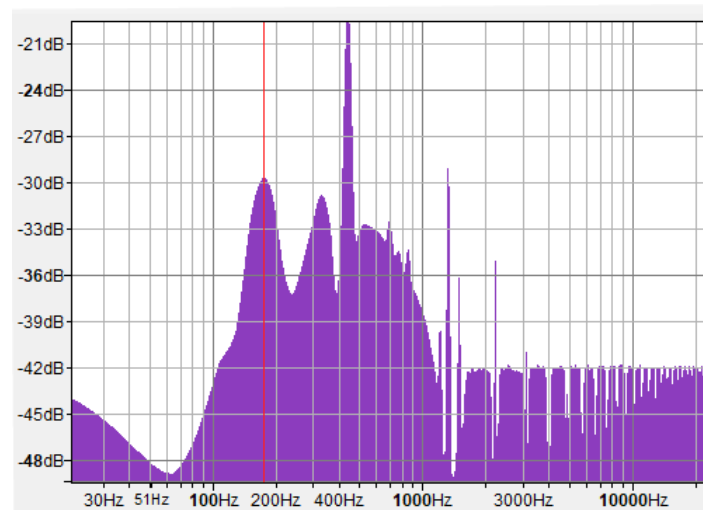


Figure 8.24 . Audacity Spektrogramm: Experiment Kerbfilter

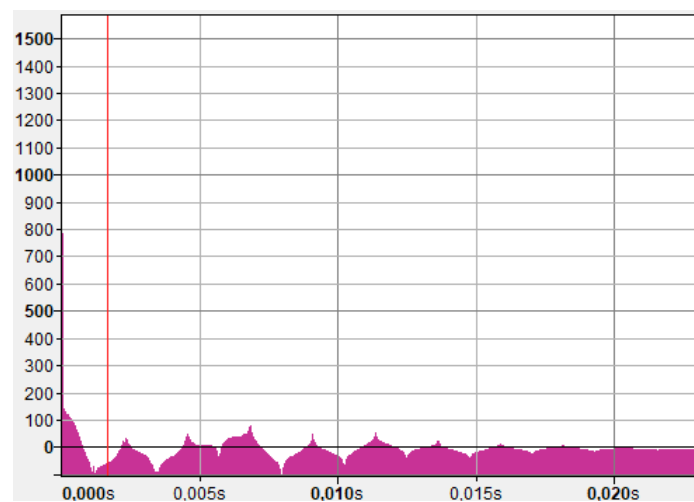


Figure 8.25 . Audacity Spektrogramm (Standard-Autokorrelation): Experiment Kerbfilter

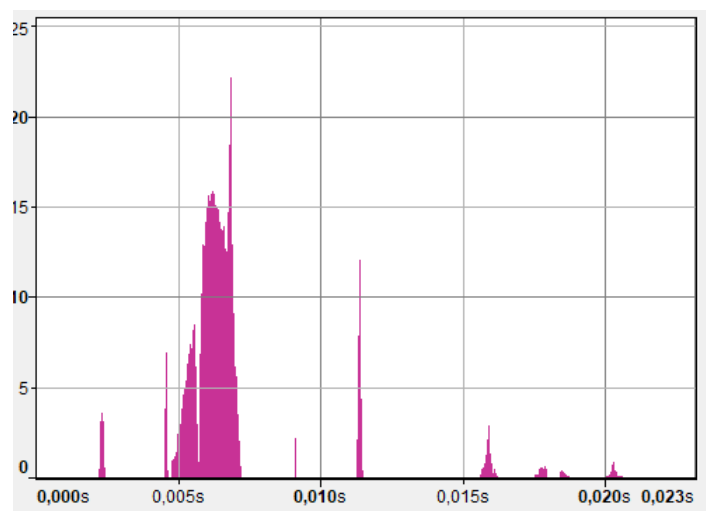


Figure 8. 26 . Audacity Spektrogramm (Erweiterte-Autokorrelation): Experiment Kerbfilter

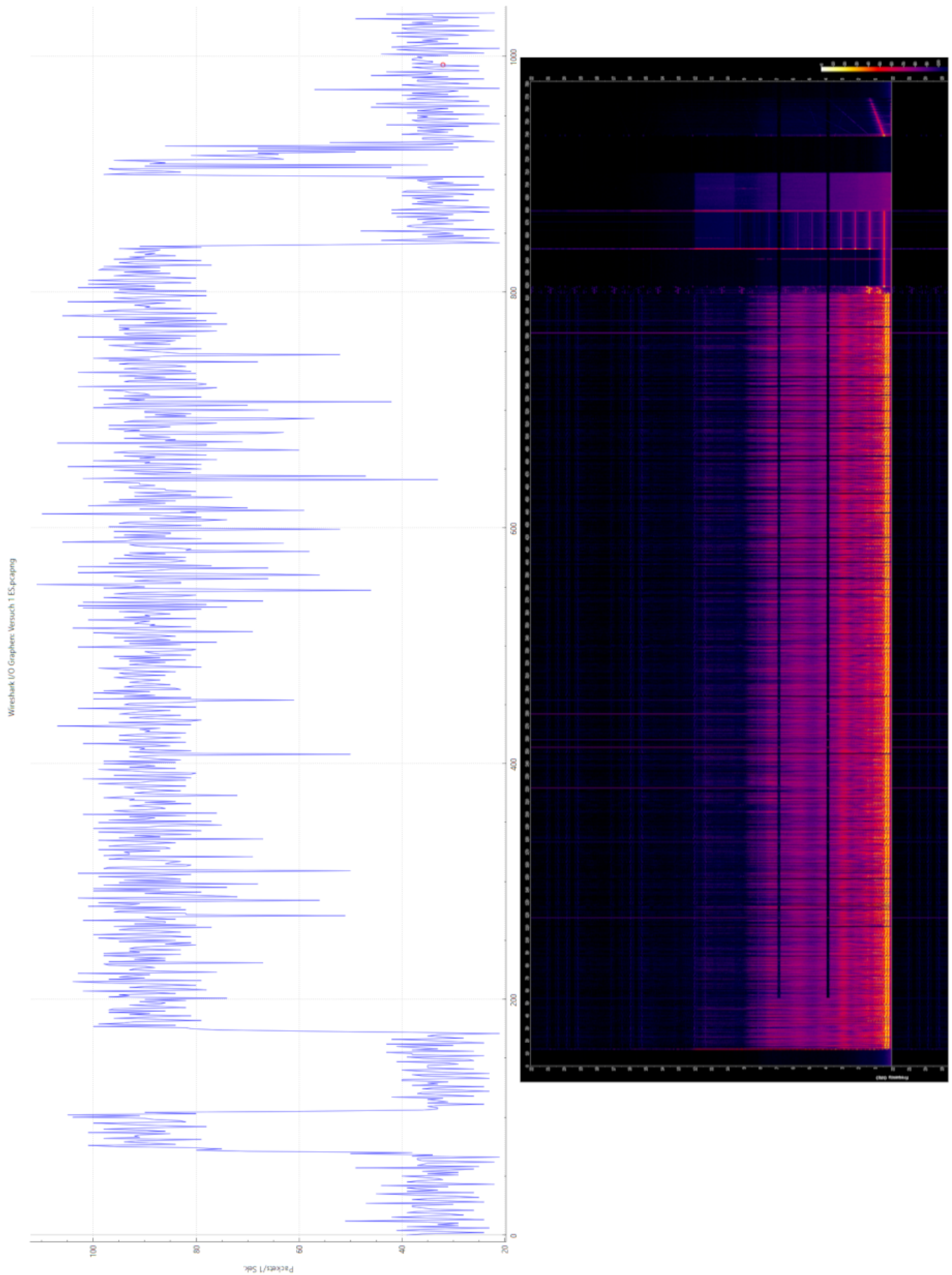


Figure 8.27. Übertragung von Audio über die Internetschnittstelle am Empfänger (R_1)

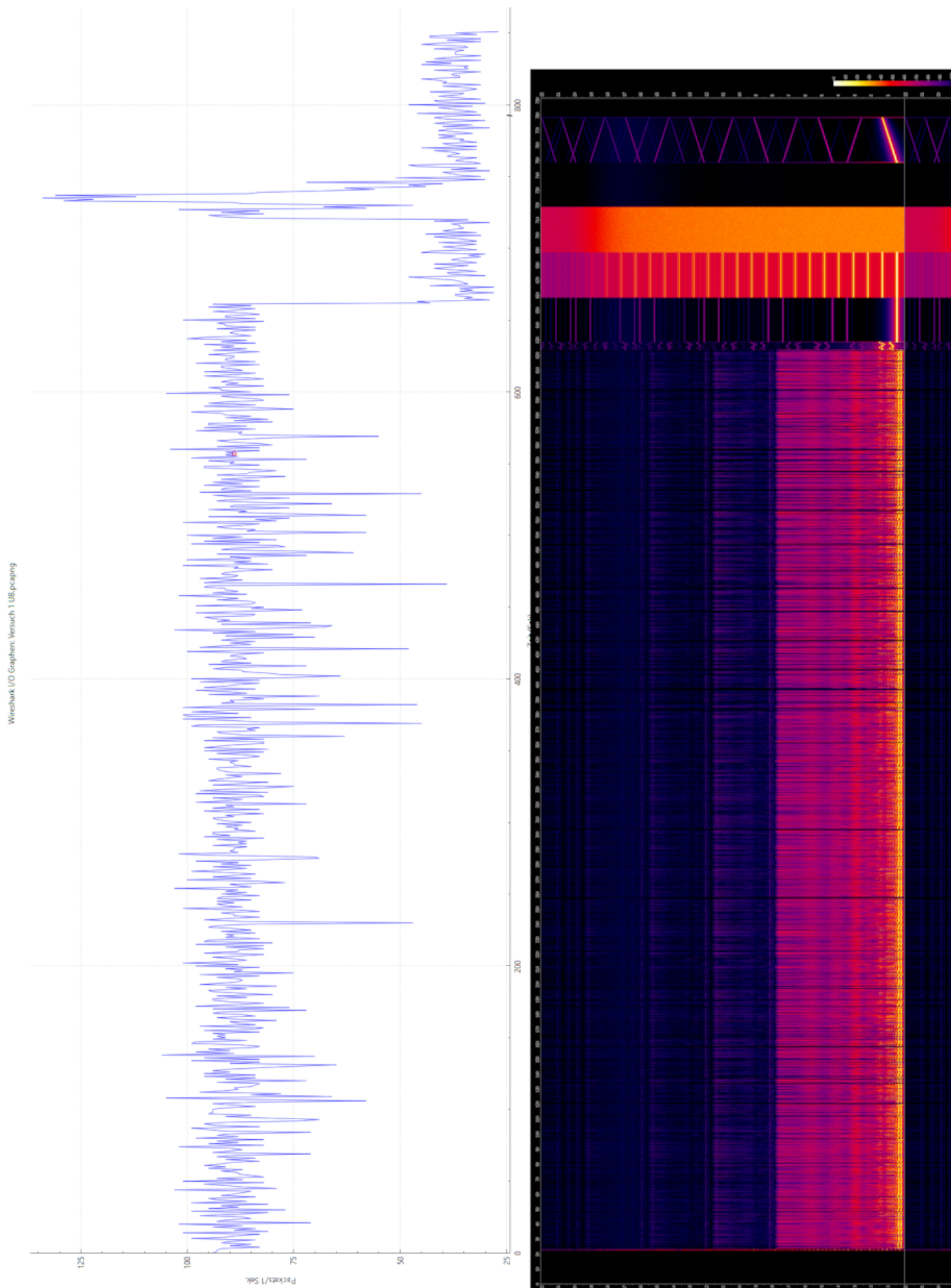


Figure 8.28. Übertragung von Audio über die Internetschnittstelle am Sender (R_1)

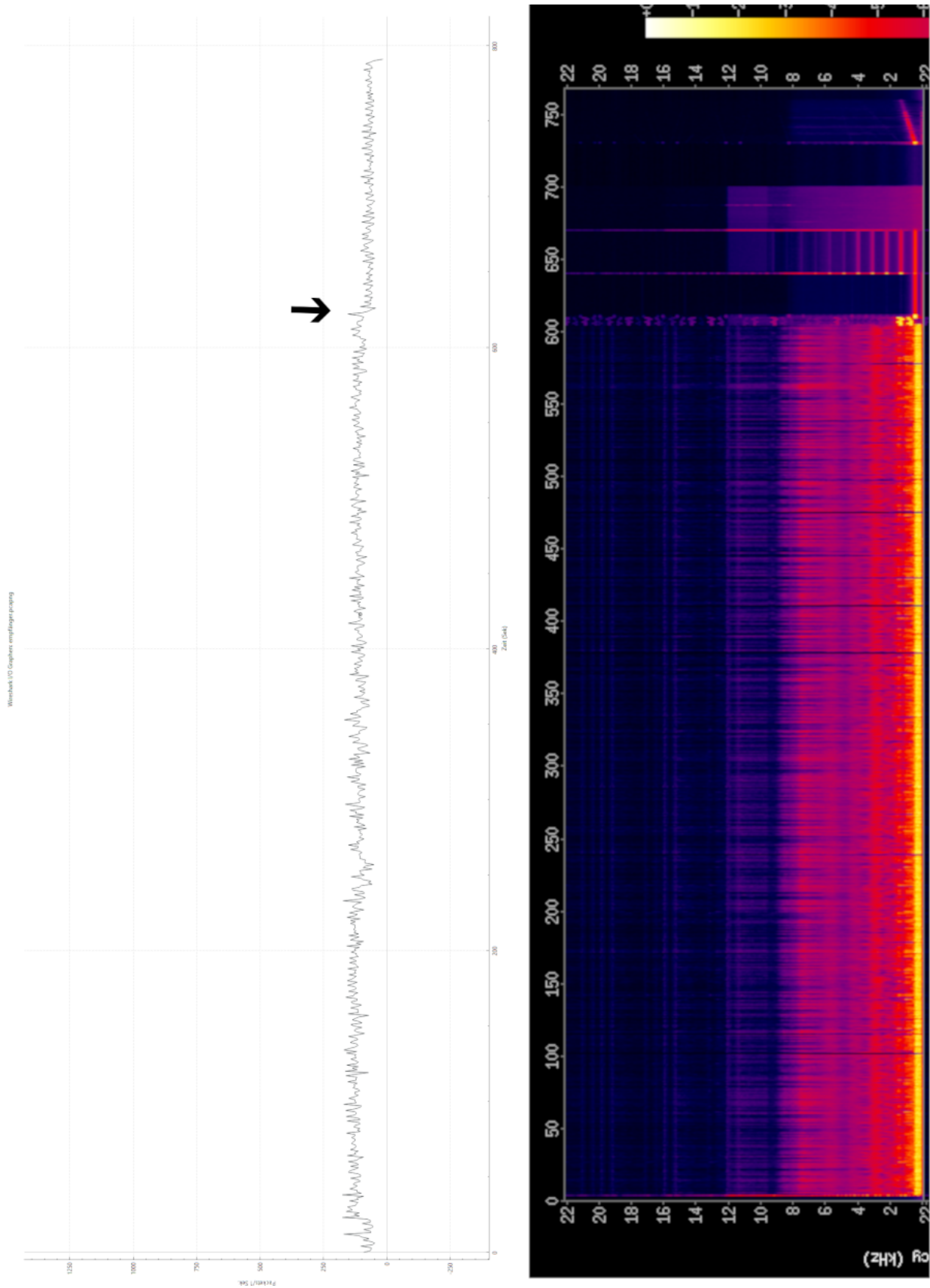


Figure 8. 29 . Übertragung von Audio über die Internetschnittstelle am Empfänger, R_{10}

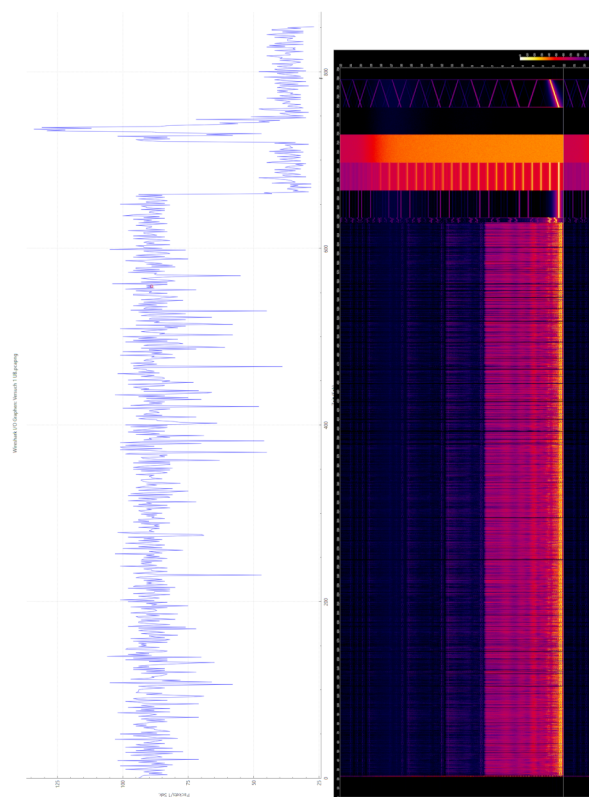


Figure 8.30 . Übertragung von Audio über die Internetschnittstelle am Sender, R_{10}

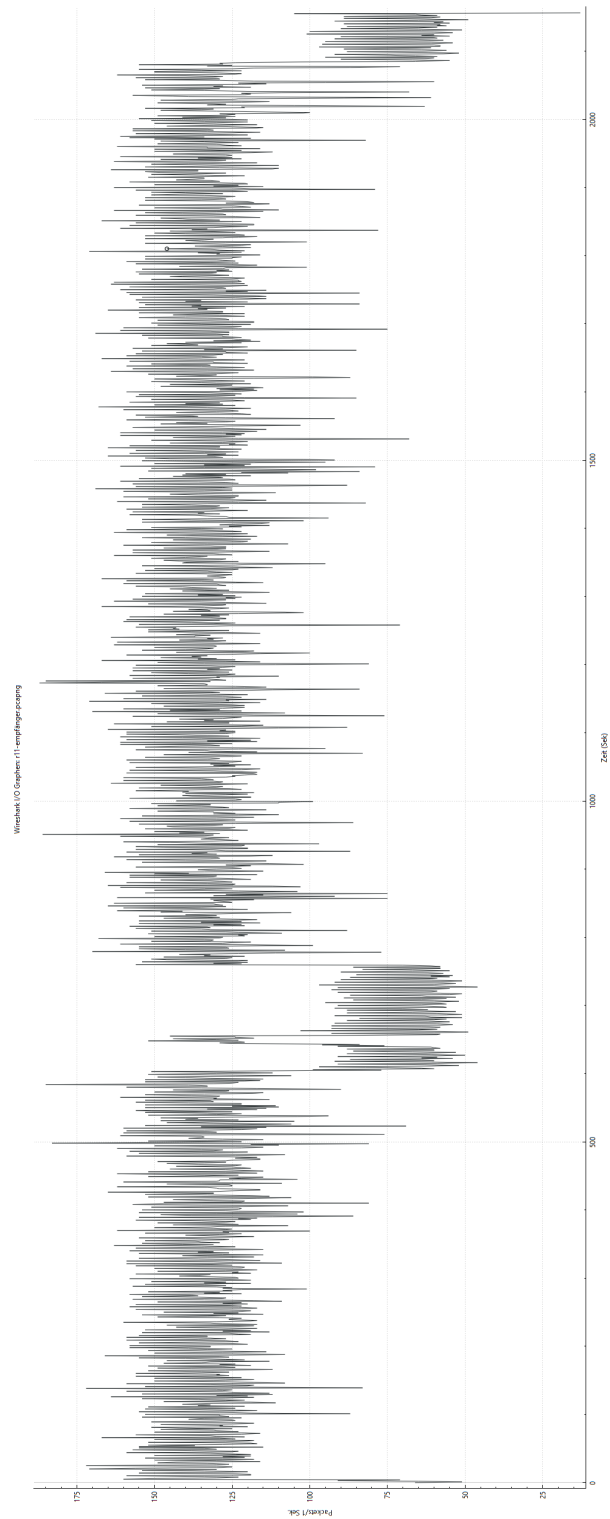


Figure 8. 31 . R_{11} I/O Graph am Empfänger

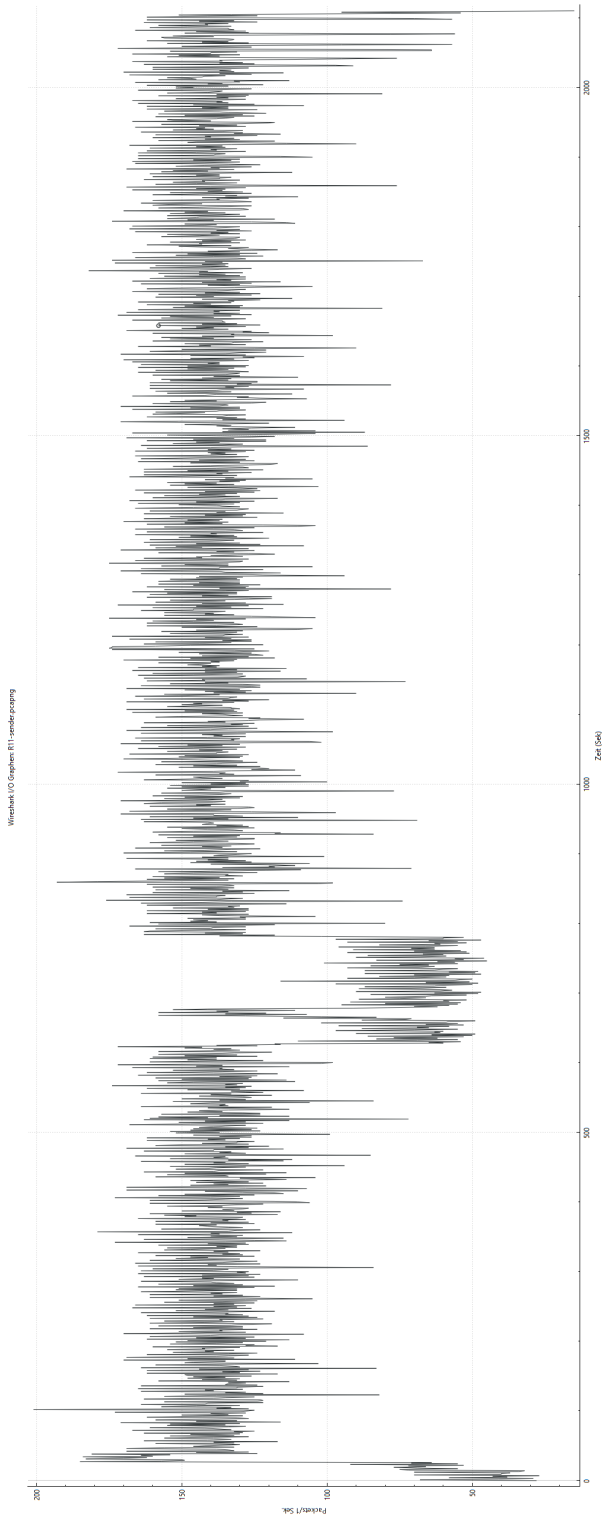


Figure 8.32 . R_{11} I/O Graph am Sender

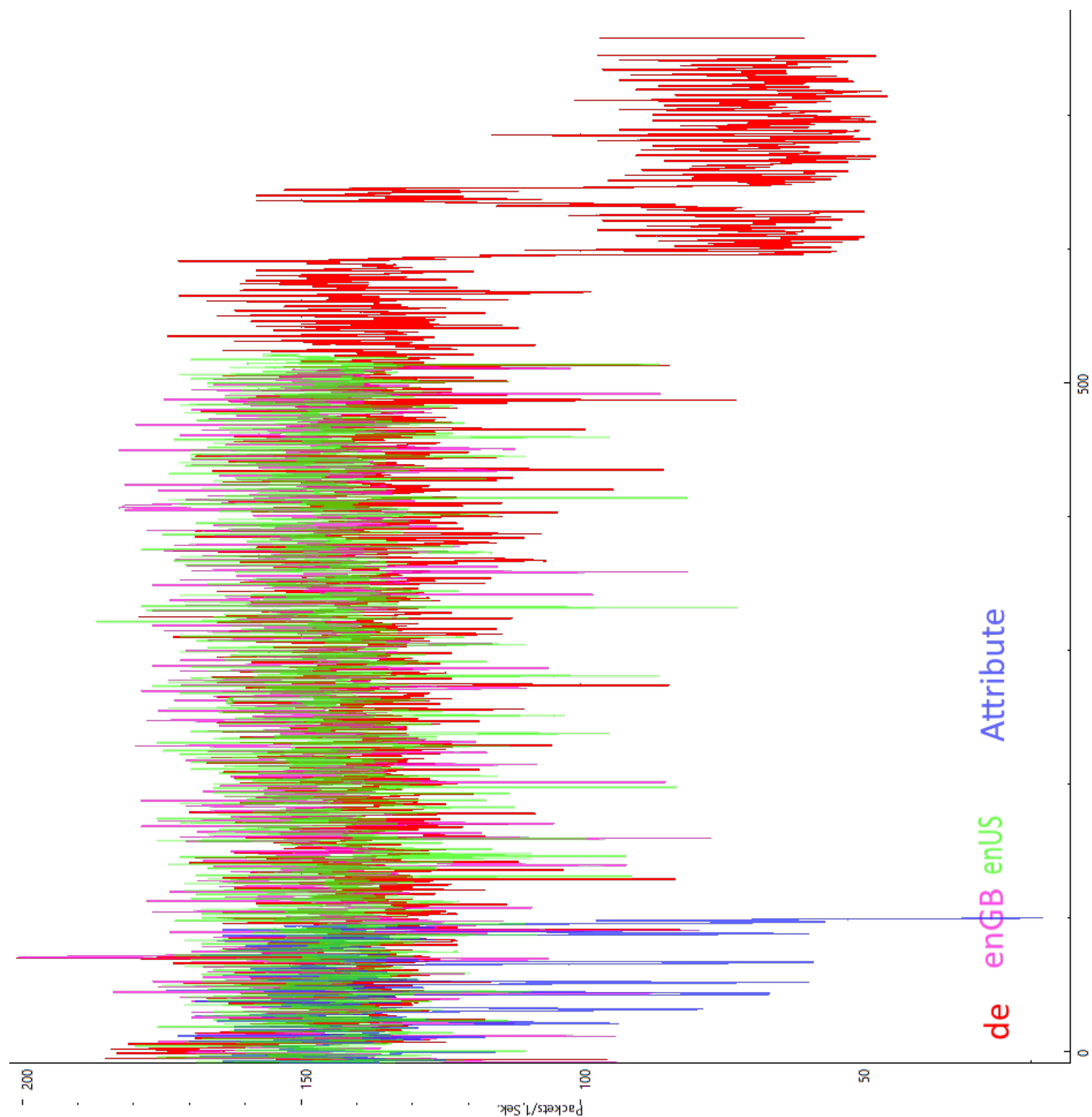


Figure 8.33 . R_{11} Übereinander gelegter I/O Graph am Sender

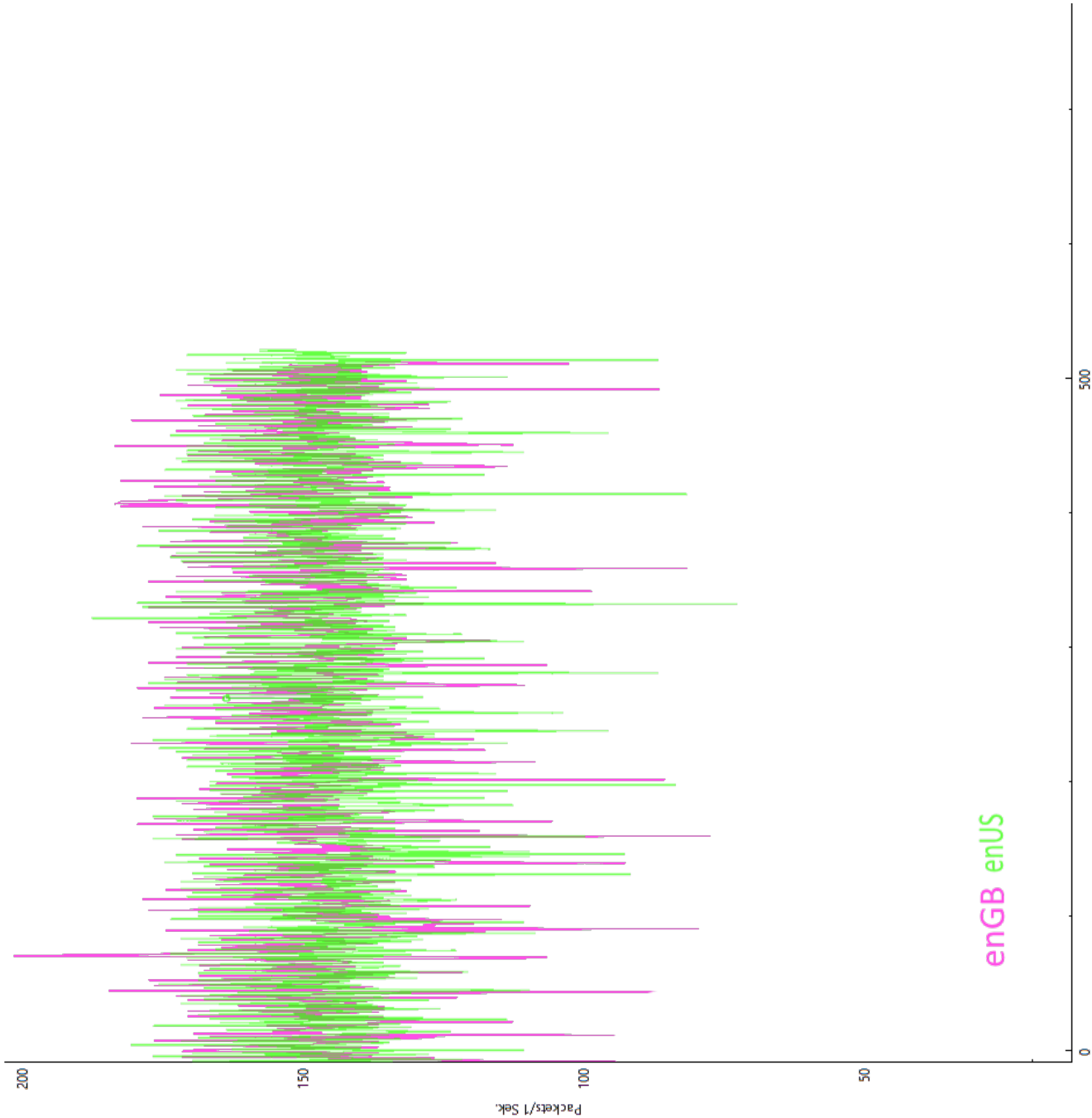


Figure 8. 34 . R_{11} enGB, enUS übereinander