HOCHSCHULE
MITTWEIDA
UNIVERSITY OF
APPLIED SCIENCES

BACHELORARBEIT

Frau
Carmen Wührl

Erstellung eines Audiodatensatzes zur sequentiellen Lokalisierung von Manipulationen

Mittweida, 2023

Fakultät: Angewandte Computer- und Biowissenschaften

BACHELORARBEIT

Erstellung eines Audiodatensatzes zur sequentiellen Lokalisierung von Manipulationen

Autor:
Frau

Carmen Wührl

Studiengang:
Allgemeine und digitale Forensik

Seminargruppe:
FO19w3-B

Erstprüfer:
Prof. Dr. rer. nat. Dirk Labudde

Zweitprüfer:
B. Sc. Svenja Preuß

Einreichung:
Mittweida, 30.05.2023

Faculty Angewandte Computer- und Biowissen-
schaften

BACHELORTHESIS

Creation of an audio data set for the sequential localization of manipulations

author:
Ms.

Carmen Wührl

seminar group:
FO19w3-B

submission:
Mittweida, 30.05.2023

Bibliografische Beschreibung:

Wührl, Carmen Maria:
Erstellung eines Audiodatensatzes zur sequentiellen Lokalisierung von Manipulationen. - 2023. – 63 Seiten, 26 Abbildungen, 12 Tabellen,
Mittweida, Hochschule Mittweida, University of Applied Sciences, Fakultät Angewandte Computer- und Biowissenschaften, Bachelorarbeit, 2023

Referat:

Die vorliegende Bachelorarbeit beschäftigt sich mit der Erstellung eines Audiodatensatzes zur sequentiellen Lokalisierung von Manipulationen. Die Motivation sich mit diesem Thema zu beschäftigen, resultiert aus der geringen Menge an öffentlichen Datensätzen im Hinblick der Multimediamanipulation und der Wichtigkeit von Audio in der Forensik (Khan et al., 2018; Luge, 2017). Dabei werden zunächst die Grundlagen aus den Themenbereichen Audio, Datensatz sowie Manipulation dargestellt. Für die Erstellung des Datensatzes, wurde zunächst eine Vielzahl an Daten bereitgestellt, indem mittels einem Pythonskript, Videos, von YouTube heruntergeladen sowie die Audiospur getrennt und im mp4-Format gespeichert wurden. Weiterhin erfolgte auf der Datenmenge, der Prozess der Datenbereinigung sowie das Umbenennen der Audiodateien. Anschließend ereignet sich die Darlegung des Konzeptes und die theoretische Beschreibung der Manipulierung sowie die exemplarische Durchführung der Manipulation. Daraufhin erfolgt die theoretische Darlegung der Aufteilung des Datensatzes in Test- und Trainingsdaten. Die Ergebnisse spiegeln wider, dass das geschriebene Pythonskript funktioniert und nahezu keine Fehler während des Downloads entsteht. Weiterhin zeigen sie auf, dass die exemplarische Durchführung funktioniert. Allerdings benötigt es zum einen noch die Umsetzung des in der Theorie dargelegten Manipulationsschrittes und zum anderen, darauf aufbauend, etwaige Evaluierungsschritte.

Inhalt

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

MP3	MPEG-2 Audio Layer 3
MP4	MPEG-4
MPEG	Moving Pictures Experts Group
Wav	wave

1 Einleitung

Audio spielt im Alltag eine wichtige Rolle, denn das Hören zählt zu den essenziellsten Sinnen (Fischer, 2016). Der Fachbegriff Audio leitet sich vom lateinischen Begriff „audire“, zu Deutsch „hören“ ab (Meroth & Tolg, 2008). Dabei stellt ein Audio eine „über das Internet abrufbare Tonaufnahme“ dar (*Audio* | *Duden*, 2023). Somit ist es nicht verwunderlich, dass das Audio, zu dem am häufigsten verwendeten Kommunikationsmittel gehört (Ajmi et al., 2022). Weiterhin sind digitale Audiodateien durch die Digitalisierung omnipräsent (Pan et al., 2012).

1.1 Motivation

Hinsichtlich der Forensik stellt das Themengebiet Audio im Vergleich mit anderen Fachbereichen ein neues Teilgebiet dar (Luge, 2017). Jedoch kann die Audioforensik bereits bei sehr vielen forensischen Fragen herangezogen werden. Als Beispiel anzuführen sind Manipulations- sowie Echtheitsanalyse oder Nebengeräusch- und Geräuschidentifizierung (Luge, 2017). Dabei wird unter Audioforensik „die Anwendung von Wissenschaft und wissenschaftlichen Methoden im Umgang mit digitalen Beweisen in Form von Audio verstanden“[1] (Dzulfikar et al., 2021, S. 145). Dadurch kann das Beweismittel Audio aufgrund der Fähigkeit, relevante Informationen für das Gerichtsverfahren zu liefern, dazu beitragen, Kriminalfälle aufzuklären (Dzulfikar et al., 2021).

In puncto Strafverfolgung beinhaltet die menschliche Stimme unter gewissen Voraussetzungen, einen erheblichen Pluspunkt im Gegensatz zu weiteren Beweismitteln, wie beispielsweise Fingerabdrücken sowie DNA-Spuren. Zumal ein Sprecher ohne jeglichen Zweifel im Gegensatz zu einem zurückgelassenen Fingerabdruck, mit der Straftat in Verbindung gesetzt werden kann. Da sich aus rechtlicher Sicht ein Täter zum Beispiel mittels einem Erpresseranruf oder einer telefonischen Lösegeldforderung mindestens der Mitwisserschaft strafbar macht (Kuenzel, 2023).

Weiterhin spiegelt sich die Wichtigkeit der Beschäftigung mit dem Bereich der Audioforensik durch die nun folgenden Beispiele zweier Kriminalfälle wider. Häufig stellt sich die Frage in der Audioforensik, wer als erster geschossen hat. Zum Beispiel, wenn eine Seite darauf besteht, aus Notwehr gehandelt zu haben, wobei die andere Seite exakt auf dem Gegenteil beharrt (R. Maher, 2016). Notwehr lässt sich als „diejenige Verteidigung, welche erforderlich

[1] Übersetzt durch Autor

ist, um einen gegenwärtigen rechtswidrigen Angriff von sich oder einem anderen abzuwenden“ definieren (Rechtskunde — Leicht Verständlich, 1973, S. 35). Ein weiteres Beispiel ist die folgende kurz dargelegte Situation. Es wurden hintereinander zwei sich anreihende Schüsse abgegeben, obwohl zwei Polizeiautos in der Nähe parkten, war keine Kamera auf die Schützen gerichtet. Die Autos nahmen lediglich den Ton beider Schüsse auf. Daher stellen sich die Fragen nach dem Ursprung der Schüsse sowie der Anzahl der Schützen oder ob es sich um zwei Schüsse aus derselben Waffe handelt (R. Maher, 2018).

1.2 Problemstellung

Durch den Anstieg der Digitalisierung ist die Menge an Multimedia-Daten, die durch intelligente Geräte produziert werden, rasant angestiegen. Dadurch sind etliche Herausforderungen entstanden, um wertvolle Hinweise aus den Multimedia-Daten zu gewinnen (Abbasi et al., 2022). Bei Multimedia handelt es sich um „das Zusammenwirken [sowie] die Anwendung von verschiedenen Medien“ (Multimedia | Duden, 2023).

Jedoch werden etwaige Untersuchungen oftmals erschwert, da heutzutage nahezu jedermann in der Lage ist, Audiosignale zu manipulieren (Xiang et al., 2022). Der Begriff Manipulation bedeutet eine Person mittels absichtlicher Einflussnahme in eine gewisse Richtung zu führen (*Manipulieren | Duden*, 2023). Hinzu kommt, dass durch neue Entwicklungen Täter nicht mehr komplette Audiodateien verfälschen. Vielmehr werden nun einzelne Abschnitte manipuliert (Rahman et al., 2022). Die Gefahren sind hierbei, dass beispielsweise Negationswörter wie „nicht“ in Aussagen eingebaut werden und somit der Inhalt abgewandelt wird (L. Zhang et al., 2022). Weiterhin werden entgegen der Gefahr, wie das bereits aufgeführte Beispiel der teilweisen Manipulation zeigt, nur eine geringe Anzahl an Studien hinsichtlich dieser Manipulationsart ausgeführt (Rahman et al., 2022).

Auch bezüglich des Vorteiles gegenüber anderen Beweismitteln ist ein bekanntes Hindernis der Stimme, mit, wie viel Gewissheit davon ausgegangen werden kann, dass die sprachliche Evidenz richtig zugeordnet wird (Kuenzel, 2023).

Weiterhin stellt die fehlende Verfügbarkeit von umfassenden, öffentlich zugänglichen Datensätzen zur Bewertung existierender und neuer Algorithmen für Multimedia-Forensiker eine signifikante Herausforderung dar (Khan et al., 2018).

1.3 Zielsetzung

Daher gilt als zielführende Aufgabe in dieser Arbeit, das Erstellen eines Datensatzes zur sequenziellen Lokalisierung von Manipulationen, um somit künftige Forschungen hinsichtlich Gegenmaßnahmen gegenüber Audiomanipulationen zu unterstützen. Der Begriff sequenziell bezeichnet dabei „fortlaufend, nacheinander erfolgend“ (Sequenziell | Duden, 2023). Hierfür werden die einzelnen vorbereitenden Schritte getroffen und das Konzept der

Manipulierungsidee dargelegt und an einem Beispiel exemplarisch ausgeführt. Woraufhin der Schritt der Datensatzaufteilung theoretisch beleuchtet wird.

1.4 Kapitelübersicht

Insgesamt besteht diese Bachelorarbeit aus 6 Kapiteln, die Einleitung miteingeschlossen. Zu Beginn erfolgt die Darlegung der wichtigsten Grundlagen, die für das weitere Verständnis vonnöten sind. Dabei wird der Grundlagenteil in weitere vier Abschnitte untergliedert. In Kapitel 3, den Materialien und Methoden, werden zum einen die verwendeten Materialien näher erläutert und zum anderen das Vorgehen im Methodenteil detailliert beschrieben. Daraufhin erfolgt die Präsentation der Ergebnisse in Kapitel 4. Darauf folgend beschäftigt sich Kapitel 5 mit der Diskussion der Ergebnisse. Abschließend findet die Darlegung des Ausblicks und des Fazits im letzten Kapitel statt.

2 Grundlagen

Zum Verständnis der vorliegenden Bachelorarbeit müssen grundlegende Konzepte bzw. elementare Termini näher erläutert werden, um die darauffolgenden Erörterungen der Themenbereiche Audio, Manipulation und Datensatz nachvollziehen zu können. Mathematische Funktionen und Umrechnungen sowie biologische Erklärungen werden nur dann berücksichtigt, wenn sie zur weiteren Erfassung der Inhalte unerlässlich sind.

2.1 akustische Grundbegriffe

Zunächst ist es wichtig den Begriff Signal eindeutig darzulegen, denn es handelt sich hierbei um das bedeutungsvollste Objekt der Sprachverarbeitung (Pfister & Kaufmann, 2017). Dabei stellt diese ein „durch Digitaltechnik ermöglichtes maschinelles Aufnehmen, Erkennen, Interpretieren und Erzeugen von Sprachlauten, sprachlichen Signalen“ dar (*Sprachverarbeitung* | *Duden*, 2023). Der Begriff Signal kann als mathematische Funktion oder Zahlenfolge definiert werden, die dazu dient, sich ändernde Größen zu charakterisieren und Informationen zu kodieren (Weinzierl, 2008). Es bildet sich, da das Gesprochene einer Person und die dadurch entstehenden Schallwellen mittels eines Mikrofons in ein elektrisches Signal umgewandelt werden (Pfister & Kaufmann, 2017). Aus Abbildung 1 geht hervor, dass bei Audiosignalen die horizontale unabhängige Variable in der Regel einen zeitlichen Verlauf repräsentiert, während die vertikale abhängige Variable den Schalldruck oder die elektrische Spannung darstellt (Weinzierl, 2008). Sie stellen eine Vielfältigkeit an akustischen Signalen, einschließlich Hintergrundgeräusche sowie gesprochene Kommunikation dar (De Benito-Gorrón et al., 2019). Da die elektrische Spannung, in dieser Arbeit keinen weiteren Einfluss nimmt, wird auf eine ausführliche Darlegung verzichtet. Jedoch wird eine kurze Definition aufgezeigt, um den Begriff einordnen zu können. Spannung gibt die „Differenz der elektrischen Potenziale zweier Punkte, aufgrund deren zwischen diesen beiden Punkten ein elektrischer Strom fließen kann“ wieder (*Spannung* | *Duden*, n.d.).

Peterson (2012) definiert Schall, als „die Wirkung, die die Schwingungen der Luft oder anderer Medien auf das Hörorgan und seine zentralen Verbindungen ausüben. Schall ist mechanische Strahlungsenergie, wobei die Bewegung der Teilchen des materiellen Mediums, durch das er sich ausbreitet (Gas, Flüssigkeit oder Festkörper), längs der Übertragungslinie erfolgt.“[2] (S. 329). Grafisch werden die Schallwellen, also Druckschwankungen an einem Punkt mittels Sinusfunktionen, wie in Abbildung 1 abgebildet, dargestellt. (Werner, 2019).

[2] Übersetzt vom Autor

Dabei handelt es sich bei der Sinusform um einen „zeitliche[n] Verlauf welcher der mathematischen Sinusfunktion entspricht“ (Winzker, 2023, S.34). Weiterhin beschreibt die Schwingung „ein[en] Vorgang, dessen Merkmale sich mehr oder weniger regelmäßig zeitlich wiederholen und dessen Richtung mit ähnlicher Regelmäßigkeit wechselt“ (Guicking, 2016, S.2).

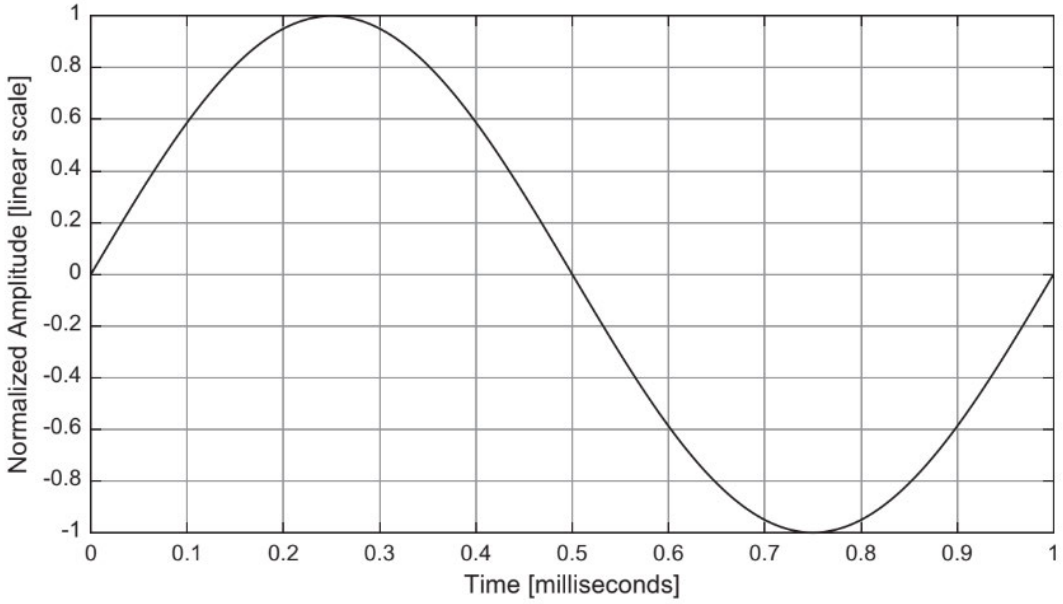

Abbildung 1: Darstellung eines Signals (R. Maher, 2018, S.7) Die X-Achse bildet die Zeit, in Millisekunden, und die Y-Achse die Amplitude ab (Weinzierl, 2008)

Zudem ist der Begriff Audio, für den Zusammenhang dieser Bachelorarbeit wichtig. Gemäß Ajmi et al. (2022) handelt es sich bei Audio „um eine Wellenform … bei der sich die Amplitude mit der Zeit ändert“[3] (S. 2).

Der Ausdruck Amplitude definiert „die maximale Höhe einer Schwingung.... Sie repräsentiert die Tonstärke, das heißt, je größer die Amplitude eines Tones ist, desto lauter wird er gehört. Je geringer die Amplitude ist, umso leiser hören wir den Ton“ (Bühler et al., 2018, S.3). Allerdings existiert keinerlei linearer Zusammenhang bei Amplitude und Hörempfinden. Daraus lässt sich folgern, dass eine doppelte Amplitude nicht gleichbedeutend ist mit einer doppelten Lautstärke (Bühler et al., 2018). Zum Verständnis, der Definition Amplitude, kann die folgende Abbildung verwendet werden.

[3] Übersetzt vom Autor

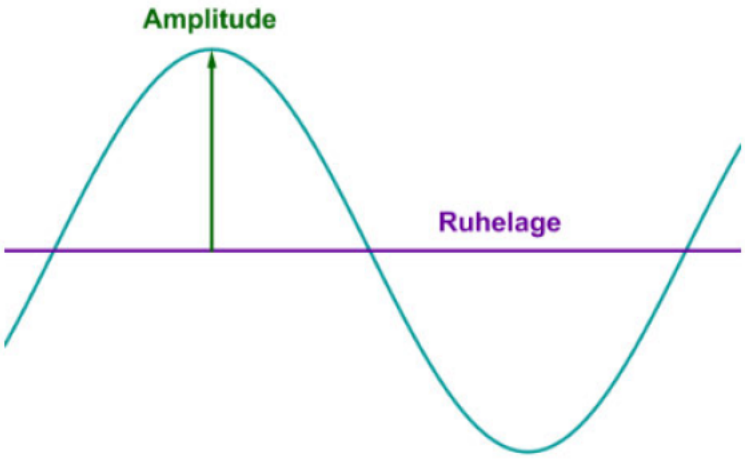

Abbildung 2: Darstellung des Begriffes Amplitude (Amplitude Und Ruhelage Der Trigonometrischen Funktionen - Lernen Mit Serlo!, 2023)

Des Weiteren wird unter der Frequenz „die Anzahl der Zyklen, die innerhalb eines bestimmten Zeitraums auftreten [verstanden]. Sie wird normalerweise in Hertz (Hz) gemessen“ (Bühler et al., 2018, S. 3). Im Gegensatz dazu spiegelt der Begriff Periodendauer, laut Bühler et al. (2018), „die Dauer einer vollständigen Schwingung“ (S. 3) wider. Der Terminus Periodendauer ist wichtig, da die Frequenz den Kehrwert der Periodendauer darstellt (Bühler et al., 2018). Zur besseren Erfassung der beiden Definitionen kann Abbildung 3 herangezogen werden.

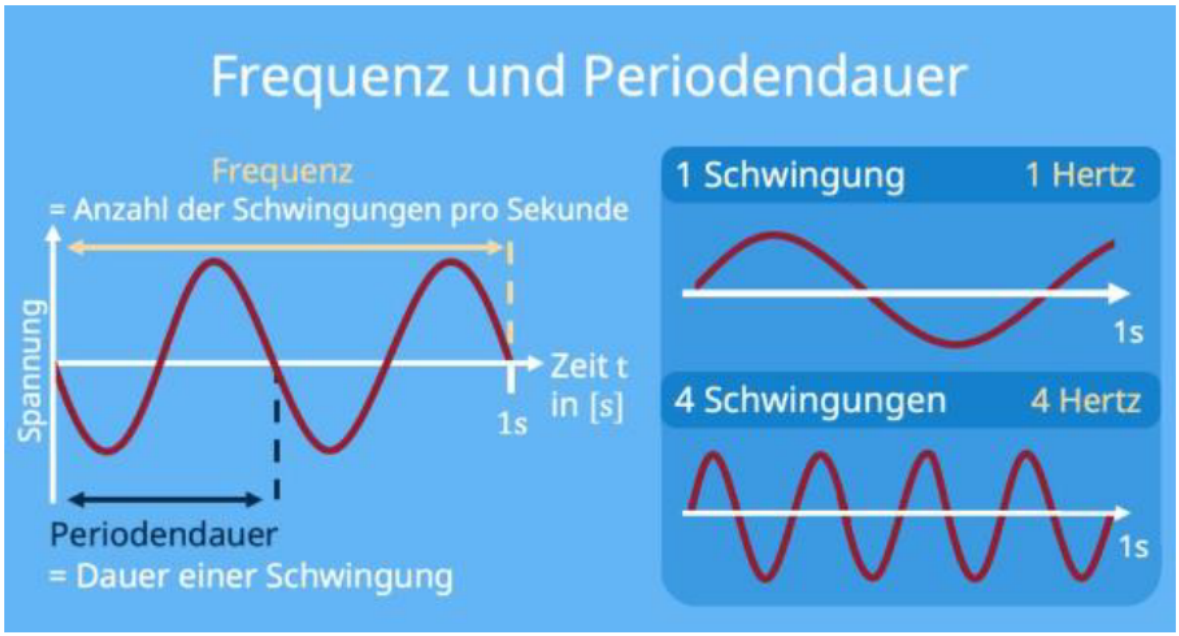

Abbildung 3: Erklärung von Frequenz & Periodendauer (Frequenz • Definition, Einheit Und Formel, 2023)

2.2 grafische Darstellungsformen

Die Deutung akustischer Informationen setzt die Ohren voraus, demgegenüber vermögen zusätzlich die Augen während einer forensischen Audioanalyse zu helfen (R. Maher, 2018).

Vor allem für die Messung von exakten Zeitpunkten sowie Amplituden sind Augen klar im Vorteil. Diesbezüglich bietet ein Wellenformanzeigeprogramm eine grafische Darstellung, welche unterstützenden herangezogen werden kann „um hörbare Ereignisse, Zeitintervalle, Signaländerungen und andere Signalattribute zu identifizieren“[4] (R. Maher, 2018, S. 48).

Bei Wellenformen handelt es sich um eine Form der Veranschaulichung „welche die Amplitude des Audiosignals [wie aus Abbildung 4 zu entnehmen ist] über der Zeit anzeigt“ (R. C. Maher, 2020).

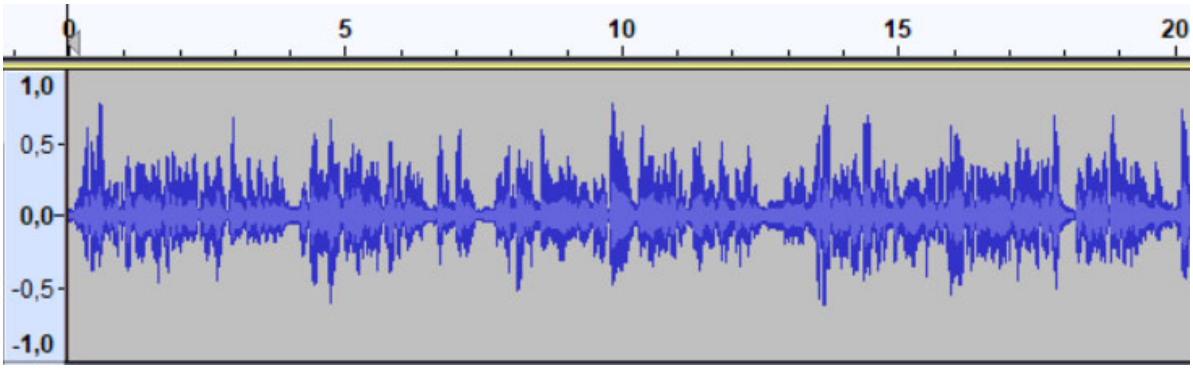

Abbildung 4: Darstellung der Wellenform, von einem 20-sekündigen Ausschnitt der Audioaufnahme, 24_originaleaudiodatei_00-50, des erstellten Datensatzes. Dabei zeigt die x-Achse die Zeit und die y-Achse die Amplitude an (R. C. Maher, 2020)

Darüber hinaus kann neben der Untersuchung der Wellenform im Zeitbereich ein Spektrogramm helfen, bedeutungsvolle Signalmerkmale zu erkennen (R. Maher, 2018).

Es handelt sich hierbei um ein besonderes Diagramm, das mittels Berechnung der Kurzzeit-Fourier-Transformation (des Spektrums) erstellt wird (R. Maher, 2018). Die Kurzzeit-Fourier-Transformation spiegelt die gängigste Technik der Zeit-Frequenz-Technologie wider, welche benutzt wird, um bei einer Untersuchung der im Signal innewohnenden Spektralanteile mit Zeitpunkten bzw. Zeitintervallen zu verknüpfen (Mertins, 2013). Weiterhin ist der Begriff Spektrum ein vielfältig definierbarer Begriff. Für die Bachelorarbeit wird die Definition von M. Werner (2010) „als die Signalbeschreibung im Frequenzbereich“ herangezogen.

Weiterhin definiert der Begriff Welle eine „Schwingung, die sich fortpflanzt“ (*Welle* | *Duden*, 2023).

[4] Übersetzt vom Autor

Beide Darstellungsformen bilden die Energie eines Audiosignals ab, wobei die waagerechte Achse (x-Achse) die Zeitskala darstellt. Jedoch handelt es sich beim Spektrogramm bei der senkrechten Achse (y-Achse), wie Abbildung 5 zeigt, um die Darstellung der Signalfrequenzskala in Hertz (R. Maher, 2018).

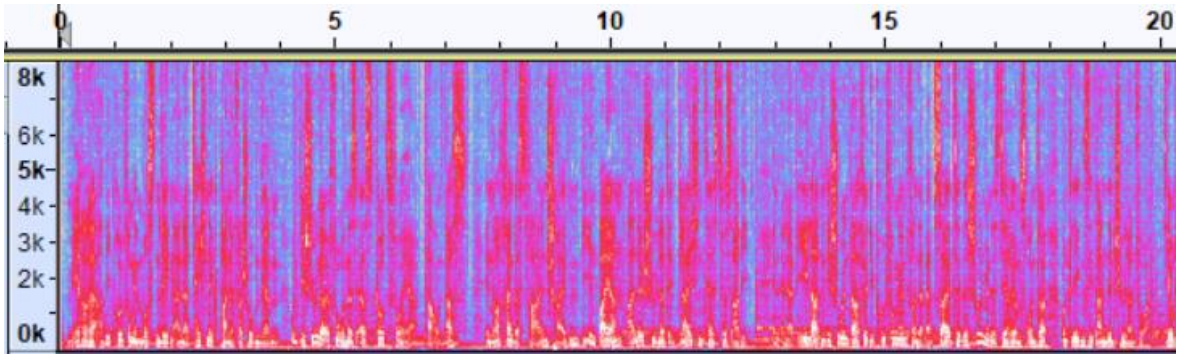

Abbildung 5: Veranschaulichung eines Spektrogramms, eines 20-sekündigen Ausschnittes der Audioaufnahme, 24_originaleaudiodatei_00-50, des erstellten Datensatzes. Hierbei zeigt die x-Achse die Zeitskale und die y-Achse die Signalfrequenzskala in Hertz (R. Maher, 2018).

Die Fourier-Transformation ermöglicht die Analyse eines Signals auf Grundlage seiner Grundfrequenzkomponenten. Mithilfe der Fourier-Transformation wird die Amplitude jeder Grundfrequenz eines Signals, das in seine Grundfrequenzen zerlegt wurde, bestimmt. In Abbildung 6 wird der Prozess der Fourier-Transformation bildlich dargestellt. Zunächst untergliedert das Spektrogramm die Länge einer Schallquelle in kleinformatige Abschnitte. Diese unterlaufen im nächsten Schritt die Fourier-Transformation, um die jeweilige Frequenz zu ermitteln. Im letzten Schritt werden die Ergebnisse aller Fourier-Transformationen der einzelnen Segmente in einem Diagramm gesammelt. Weiterhin werden die einzelnen Frequenzamplituden durch verschiedene Farben abgebildet. Die Intensität der Farbe eines Audiosignals ist abhängig von dessen Energie (Ghadekar et al., 2023).

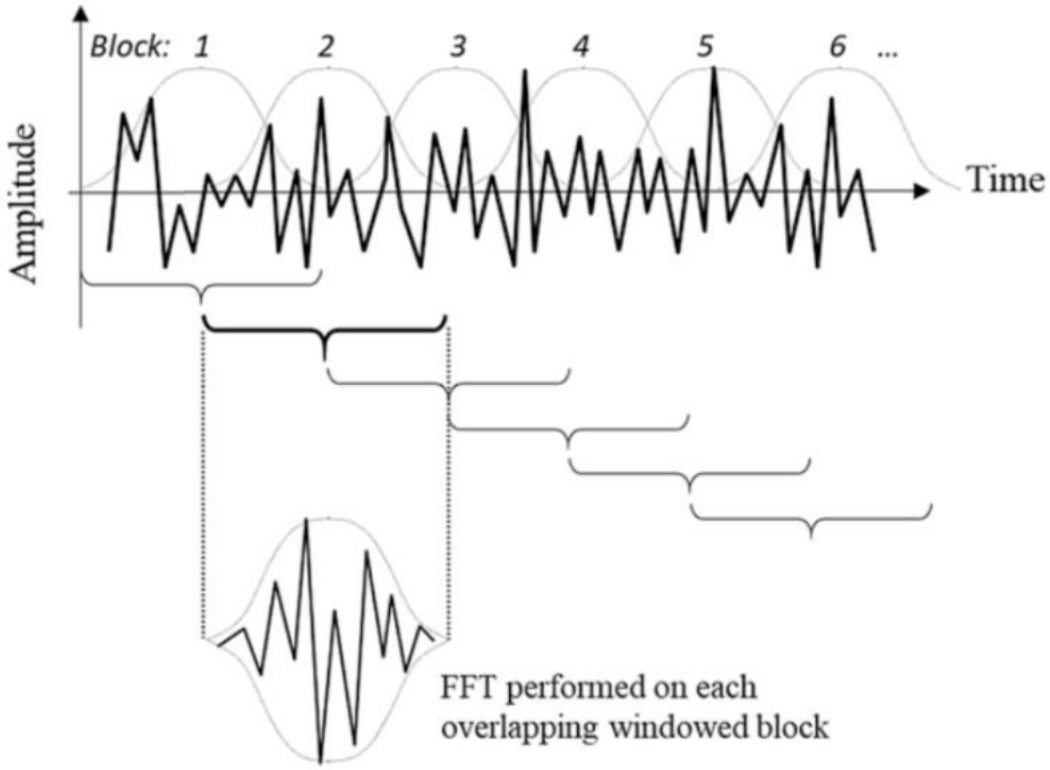

Abbildung 6:Darstellung der Fourier-Transformation während der Erstellung eines Spektrogramms (Maher, 2018, S.42) (Übersetzung durch Autor: Time = Zeit und FFT performed on each overlapping windowed block = Durchführung der Fouriertransformation an jedem überlappenden Block

Eine weitere Darstellungsform ist das Mel-Spektrogramm bzw. Melgram, welche eine Zeit-Frequenzmatrix darstellt. Dabei spiegelt die Frequenzachse die Mel-Frequenz-Skala wider, welche eine logarithmische Wahrnehmungsdarstellung des Spektrums ist. Die Mel-Spektrogramm Umwandlung beruht auf der Ausrechnung der Kurzzeit-Fourier Transformation. Die Frequenzbänke werden mittels einer Mel-Filterbank in die Mel-Skala umgewandelt (De Benito-Gorrón et al., 2019

Filterbanken werden oft als vorbereitende Schritte in vielen Anwendungen verwendet und können zur Merkmalsextraktion eingesetzt werden (Gautama & Van Hulle, 1999). Die Intention einer Filterbank ist die „Aufteilung oder Zerlegung des Signals in Teilbänder“ (Schuller, 2023, S. 1). Dieser Prozess, resultiert in Teilbandsignalen, wobei ein Frequenzteilband jeweils einem spezifischen Filter in der Bank zugeordnet ist (Penedo et al., 2019). Im Detail kommt die Mel-Filterbank zum Einsatz, wenn die Charakteristiken der menschlichen Hörverarbeitung dargestellt werden sollen (Tak et al., 2017).

Unter der Mel-Skala lässt sich eine Skala der Tonhöhe verstehen, welche von Menschen mit konstanter Entfernung vernommen werden (Ghadekar et al., 2023). Dabei wird die Hertz Skala, wie aus Abbildung 7 zu entnehmen ist, auf der Mel-Skala neu zugeordnet. Dabei gilt

der Ton von 1000 Hz als Referenzpunkt, er wird mit der Tonhöhe von 1000 mel gleichgestellt (Ghadekar et al., 2023).

Diese Umwandlung wird getroffen, da Begutachtungen aufzeigten, dass eine individuelle Wahrnehmung eines Tones, also die sogenannte Tonheit, nicht linear zu der jeweiligen Frequenz auftritt. Angesichts dessen wurde empfohlen, das Spektrum eher über die Mel-Skala, anstatt über die lineare Frequenzskala zu betrachten (Pfister & Kaufmann, 2017).

Die Tonheit besitzt die Einheit Mel, die Bezeichnung stammt von dem englischen Begriff melody (Möser, 2009).

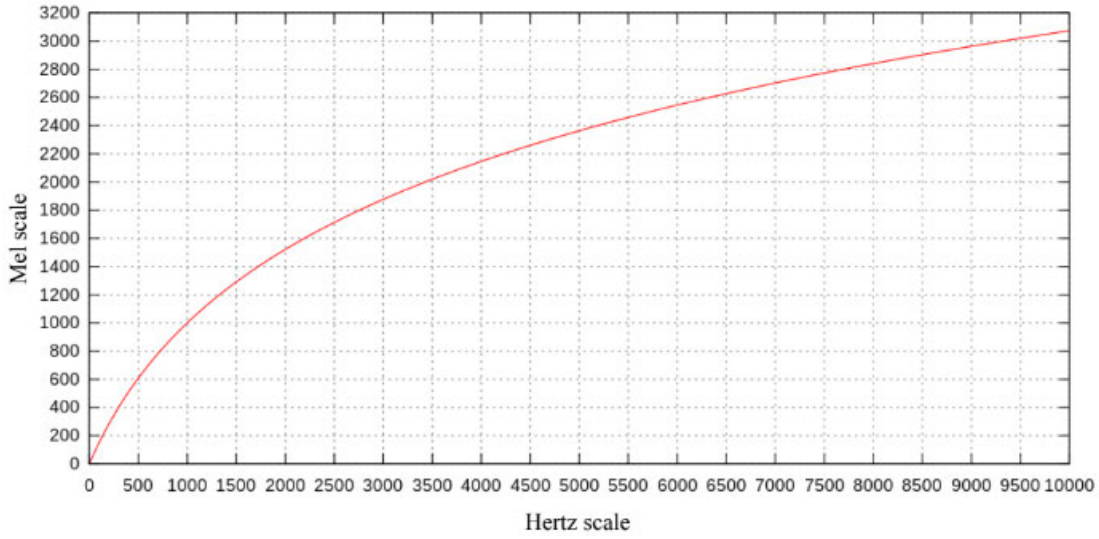

Abbildung 7: Gegenüberstellung der Hertz und Mel-Skala (Ghadekar et al., 2023) Dabei spiegelt die X-Achse die Hertz Skala und die Y-Achse stellt die Mel-Skala dar.

Zum Verbildlichen des Unterschiedes zwischen dem Spektrogramm und Melgram kann die Abbildung 8 herangezogen werden. Dabei spiegelt das in a) dargestellte Spektrogramm die Audiodatei, 24-originaleaudiodatei_00-50, wider und b) stellt das Mel-Spektrogramm zu dieser Datei dar.

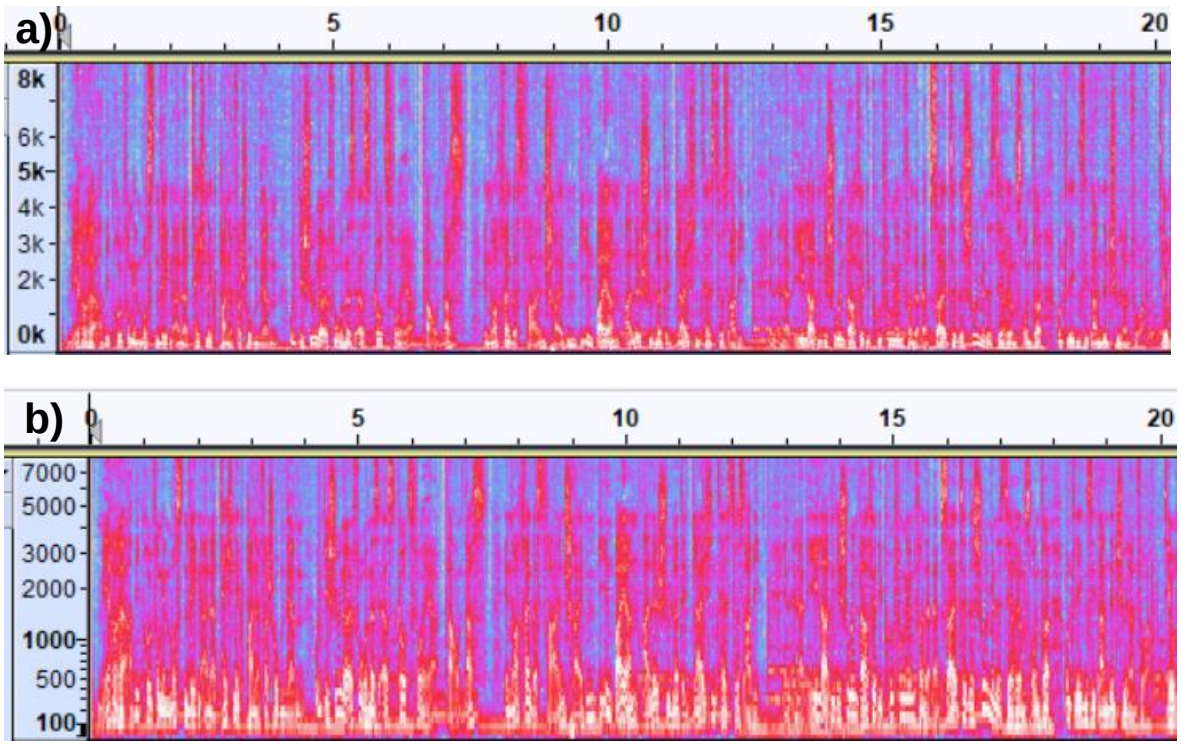

Abbildung 8: Gegenüberstellung des in a) abgebildeten Spektrogramms und dem in b) dargestellten Melgram. Dabei stellen beide Teilabbildungen, einen Ausschnitt aus der Audiodatei 24-originaleaudiodatei_00-50, aus dem erstellten Datensatz dar

Eine weitere Variante ist das Cochleagram, dabei stellt dies eine ähnliche Form des Spektrogramms dar. Jedoch versucht diese Form das menschliche Gehör nachzuahmen (Buermann & Van Meer, 2020). Aus Abbildung 9, geht hervor, dass sich die Cochlea im Innenohr befindet und das Hörorgan darstellt (Wang & Brown, 2006). Weiterhin lässt sich das Erscheinungsbild, laut R. Maher (2018), als einen „knöcherne[n], spieralförmige[n] Hohlraum, der das weiche biologische Gewebe umhüllt und schützt, das besonders empfindlich auf schallinduzierte Vibrationen reagiert“[5] beschreiben (S.20). Somit ist die Cochlea „ein integraler Bestandteil der menschlichen Hörperipherie, und seine Funktionen umfassen sowohl die Frequenzanalyse der vom Mittelohr empfangenen Schwingungen und der Transduktion dieser Schwingungen in ein neurales Signal.“ (Mill, 2008, S. 20).

[5] Übersetzt vom Autor

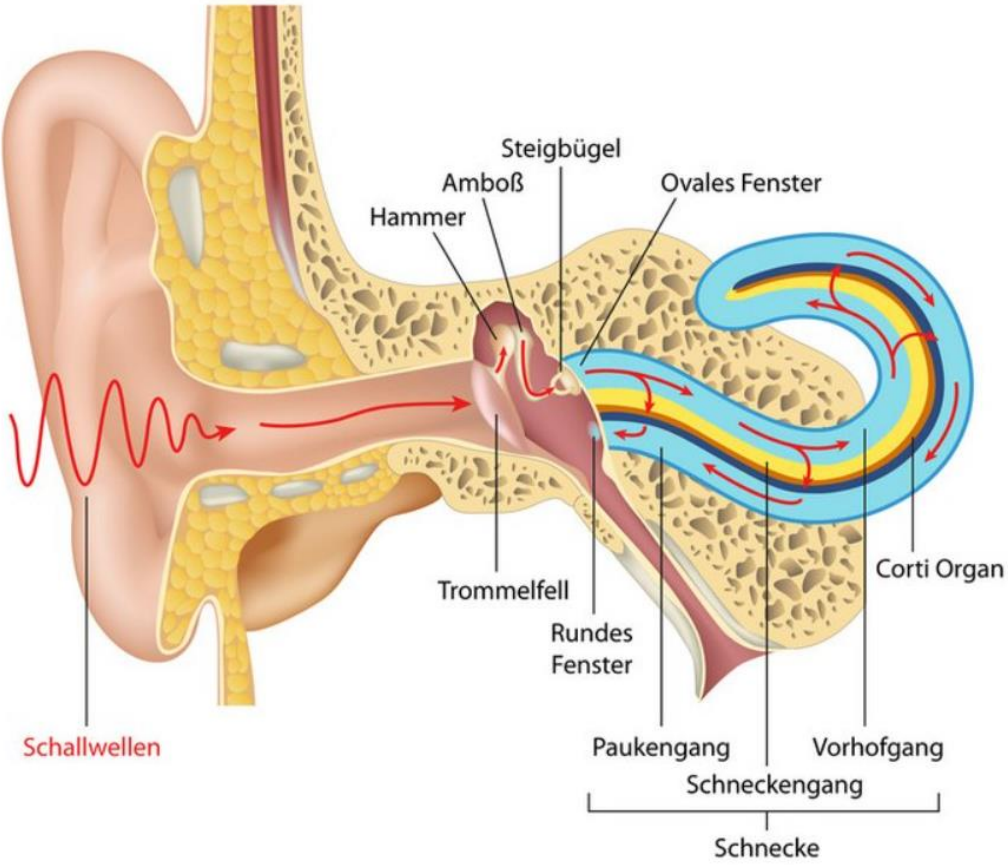

Abbildung 9: Darstellung des Ohres und der Cochlea (Hörschnecke) (CI Und Hörimplantate | Cochlea Implantat-Zentrum | Klinikum Stuttgart, n.d.) Die Cochlea übermittelt zum einen den Schall an das eigentliche Hörorgan und zum anderen führt es eine Filterung durch, wodurch verschiedene Frequenzen an den einzelnen Orten die Härchen des Cortischen Organs reizen (Meroth & Tolg, 2008)

Das Spektrogramm und das Cochleagramm gleichen sich nahezu hinsichtlich der grafischen Darstellung, wie aus Abbildung 10 zu entnehmen ist. Die Abweichung zwischen beiden Arten ist, dass für das Imitieren des Hörorgans, bei der Erstellung des Cochleagramms einer Audiodatei, ein Gamma-Filter zur Anwendung kommt. Dies geschieht, um die Verhaltensweise genauer nachzubilden (Buermann & Van Meer, 2020). Nach Sharan und Moir (2015a) „wird der Gammatonfilter häufig für den Zweck der linearen Filtermodellierung der Frequenzselektivität der menschlichen Cochlea herangezogen“ (S. 441). Dies ist wichtig, da die große Anzahl an Haarzellen, innerhalb der Hörschnecke auf einer spezifischen Frequenz schwingen (Sharan & Moir, 2015b).

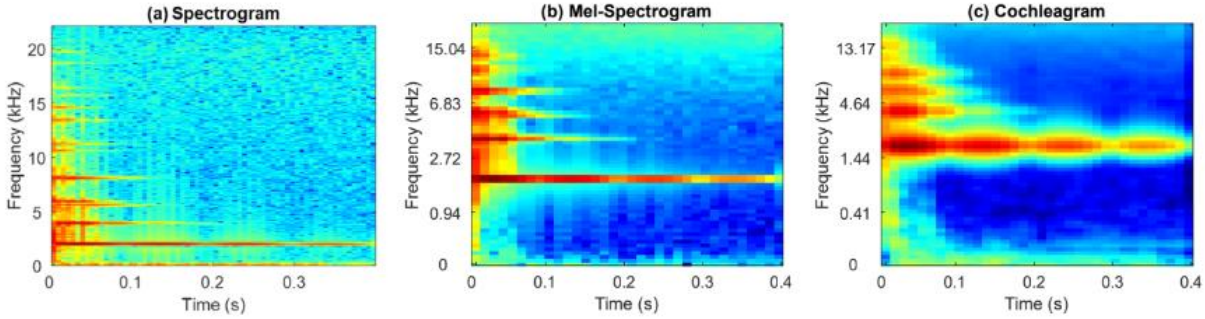

Abbildung 10: Gegenüberstellung Spektrogramm, Mel-Spektrogramm und Cochleagram (Sharan & Moir, 2019, S.4)

Der Vorteil gegenüber dem Spektrogramm charakterisiert sich durch die unterschiedliche Verteilung der Frequenzinformationen. Bei einer Darstellung eines Spektrogramms werden, wie bereits dargelegt, die beherrschenden Frequenzinformationen entlang der Zeit dargestellt. Dabei werden die Frequenzkomponenten gleichbleibend längs der orthogonalen bei unveränderter Bandbreite grafisch dargestellt. Allerdings beinhalten viele Schallsignale größere Frequenzkomponenten innerhalb des niedrigeren Frequenzbereiches, infolgedessen diese Informationen in einer solchen Zeit-/ Frequenzdarstellung nicht komplett abgebildet werden (Sharan & Moir, 2015b). M. Werner (2008) definiert den Begriff Bandbreite als „eine wichtige Kenngröße zur Beschreibung von Signalen im Frequenzbereich. Sie gibt die Breite des Intervalls im Spektrum an, in dem die ‚wesentlichen' Frequenzkomponenten des Signals liegen" (S. 161).

Dementgegen beinhaltet das Cochleagram schmale Frequenzkomponenten im niedrigeren Bereich und gegenteilig breite Frequenzkomponenten im oberen Bereich. Daher sollte bei der Analyse von einem akustischen Ereignis, bei dem der größte Teil an spektraler Energie im niedrigeren Frequenzbereich liegt, ein Cochleagream herangezogen werden (Sharan & Moir, 2019). Da es eindeutig eine höhere Anzahl an spektraler Information beinhaltet (Sharan & Moir, 2015b).

2.3 Manipulation

Ein wichtiger Teil der Bachelorarbeit stellt der Themenbereich Manipulation dar. Für das Grundlagenverständnis werden hierfür verschiedene Manipulationsarten und deren Gefahren sowie Audio-Tools im Hinblick der Manipulation dargelegt. Des Weiteren liegt der Fokus in der Bachelorarbeit auf das Beschreiben von Manipulationsarten des Audiosignals, da es sich im Allgemeinen bei forensischen Audiobeweisen um Tonaufnahmen handelt (R. Maher, 2018). Etwaige weitere Manipulation werden dabei nicht beleuchtet.

2.3.1 Manipulationsarten

Es ist ein relevanter Aspekt, dass forensische Audioaufnahmen häufig nicht unter optimalen Voraussetzungen aufgenommen werden, deshalb beinhalten die Aufnahmen oftmals Rauschen, Ausschnitte, Verzerrungen, Störgeräusche sowie andere Mängel. Dadurch wird der Zustand und die Klarheit der Sprache beschädigt. Dies gilt vor allem für geheim aufgenommene Aufnahmen, denn dieser Faktor unterbindet oftmals eine gute Platzierung des Mikrofons, dabei stellen sehr hallende Aufnahmen sowie Störgeräusche, beispielsweise das Reiben des Mikrofons an der Kleidung, die Resultate dar (R. Maher, 2018).

Für die vorliegende Bachelorarbeit sind normale Störgeräusche, die zum einen bereits während der Aufnahme entstehen und zum anderen ohne böswilligen Hintergrund vorgenommen werden, von den Manipulationsarten, die nun beschrieben werden, zu unterscheiden, welche absichtlich herangezogen werden, mit arglistigen Hintergedanken.

Zu nennen ist zum einen das Verändern der sogenannten akustischen Szene (Yi et al., 2022). Der Ausdruck beschreibt die Umgebung des auditiven Aufnahmeortes (Bhagtani et al., 2022). Weiterhin ist eine Teilaufgabe der Audioforensik die sogenannte Wiederherstellung von Verbrechens- oder Unfallszenen. Auch hier können erhebliche Folgen auftreten, wenn die akustische Szene manipuliert wird (Yi et al., 2022).

Weiterhin kann zur Verdeutlichung der Bedrohung die folgende Situation herangezogen werden. Zum Beispiel kann bei einem Notruf eines verfolgten Opfers, der wirkliche Standort nicht passend festgestellt werden, fall das Lokalisierungssystem von einer akustischen Szenemanipulationstechnologie attackiert wird (Yi et al., 2022). In Abbildung 11 ist ein Beispiel für eine Umgebungsmanipulation dargestellt. Dabei spiegelt (a) eine reale Äußerung in einer Szene wider und (b) zeigt die Unterschiede in der Wellenform nach einer Umgebungsmanipulation[6] (Yi et al., 2022)

[6] Übersetzung von (a) und (b) durch Autor

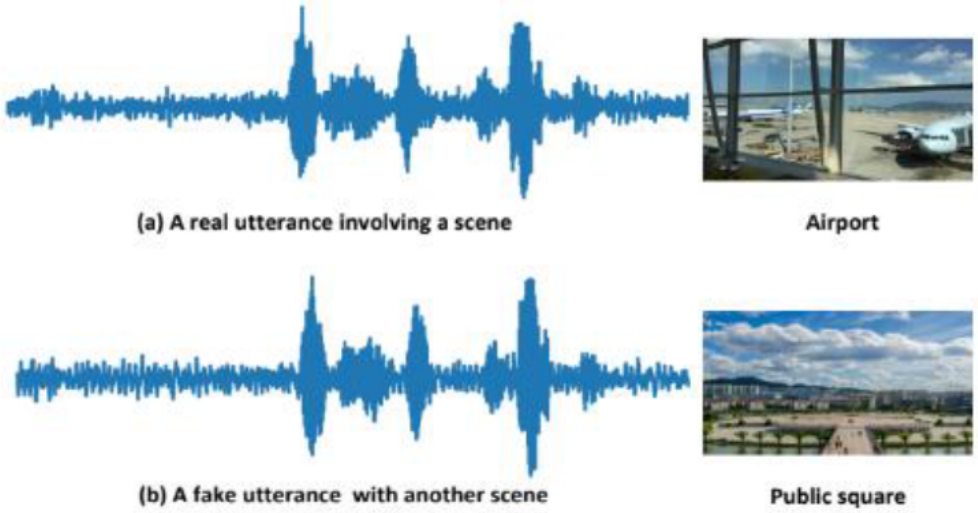

Abbildung 11: Verbildlichung der akustischen Umgebungsmanipulation (a) spiegelt eine reale Äußerung in einer Szene (b) zeigt die Unterschiede der Wellenform nach einer Umgebungsmanipulation (Yi et al., 2022, S.5)

Zum anderen ist ein weiteres oftmals verwendetes Mittel das Splicing, es bedeutet, dass Ausschnitte eines Audiosignals in ein anderes eingeflochten werden (Zakariah et al., 2018).

Solche verfälschten Audios beschweren die Vertrauenswürdigkeit von Beweisen. Weiterhin besitzen diese Manipulierungen das Potenzial für negative Konsequenzen auf die Allgemeinheit durch das Ausweiten von Fake News (Z. Zhang et al., 2022). Abbildung 12 zeigt ein Beispiel für das Splicing auf. Dabei wird der Ausschnitt „How are you?“ zu Deutsch „Wie geht es dir?“ aus einer eigenen Aufnahme in eine andere eingefügt. Daraufhin ändert sich die Originalaussage „Can you speak one more time?“ – „Können Sie noch einmal sprechen?“ in die nun neue Aussage „Can you speak ‚I am good‘ one more Time?“- „Kannst du noch einmal ‚Mir geht es gut‘ sagen [7] (L. Zhang et al., 2022).

[7] Die englischen Segmente wurden vom Autor übersetzt

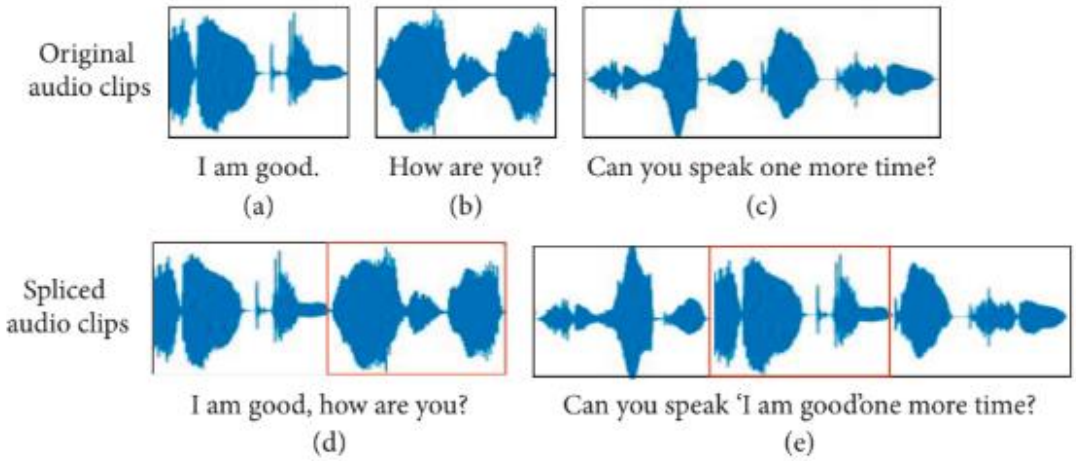

Abbildung 12: Ablauf des Splicing. In der ersten Reihe sind die originalen Audioaufnahmen und in der unteren Reihe ist das Ergebnis dargestellt (Z. Zhang et al., 2022, S.5)

Weiterhin können Bereiche eines Signals entfernt oder an einen anderen Punkt innerhalb des Audiosignals versetzt werden (Bhagtani et al., 2022). Diese Variante wird auch als „Copy-Moves“ bezeichnet (L. Zhang et al., 2022).

Eine weitere Methode ist das Teilspoofing. Dabei werden, „synthetisierte oder transformierte Audiosegmente in eine echte Sprachäußerung eingebettet“[8] (L. Zhang et al., 2022, S.1). Wie in der Einleitung dargelegt, können selbst schon kleine Änderungen innerhalb einer Tonaufnahme zur Veränderung der Intention führen. Zudem steigern sie das Potenzial für Missbrauch im Hinblick von Nachahmung oder Betrug. Beispielsweise können Abschnitte künstlicher Sprache, in eine echte Aussage eingliedert sowie Segmente durch falsche Sprache ausgetauscht werden (L. Zhang et al., 2022). In Abbildung 13 sind für das Erkennen von Unterschieden nach der Manipulation mittels Teilspoofing ein Spektrogramm sowohl vor als auch nach der Manipulation aufgezeigt.

[8] Übersetzt vom Autor

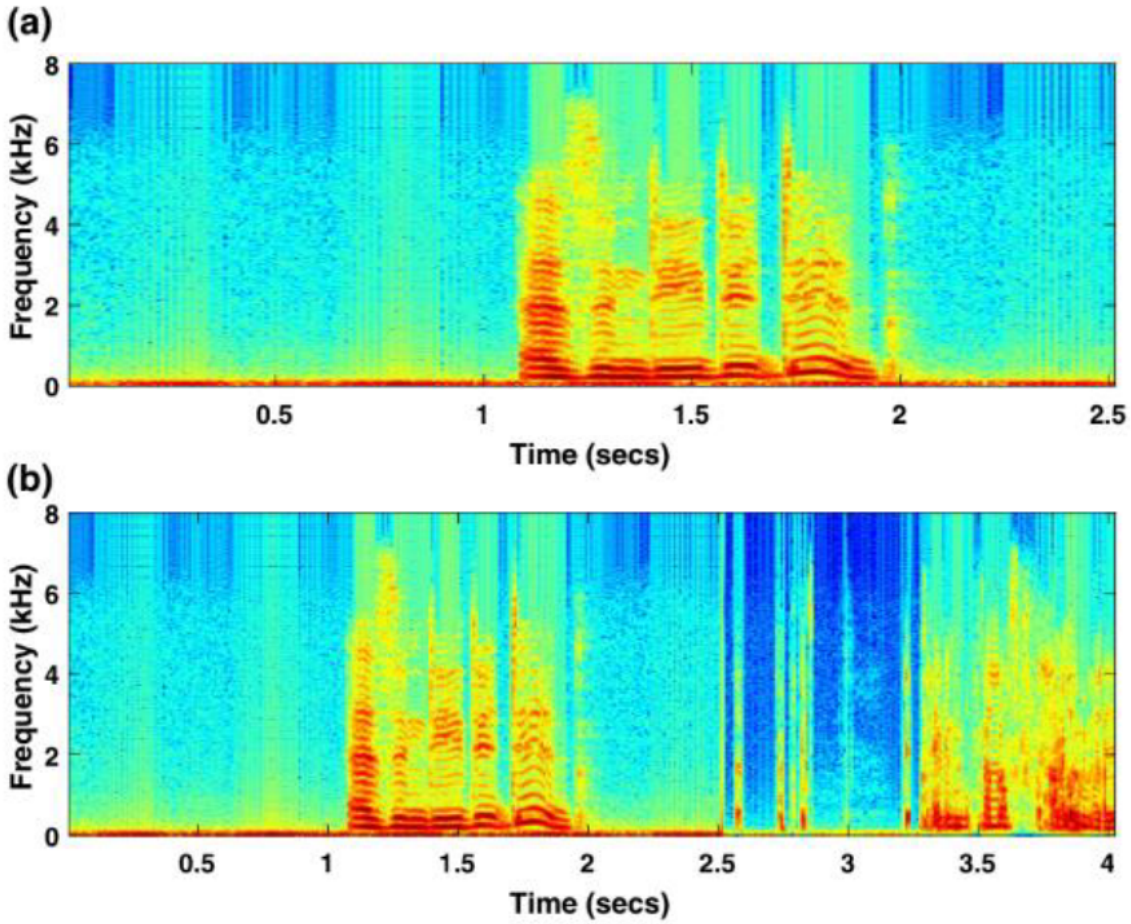

Abbildung 13: Vergleich eines in a) nicht manipulierten Spektrogramms und in b) das Ergebnis nach der Manipulationsart Teilspoofing (Kishore Kumar et al., 2021, S. 198)

In Tabelle 1 sind für das Verständnis und dem Verinnerlichen der Unterschiede die verschiedenen Manipulationsarten außer dem Manipulieren der akustischen Szenen nochmals aufgelistet.

Tabelle 1: Unterschiede der Manipulationsarten

Art	Definition
Copy-Move	Kopieren/ Einfügen innerhalb der gleichen Audioaufnahme (Bhagtani et al., 2022)
Spleißen	Die Ausschnitte stammen aus unterschiedlichen Aufnahmen (Zakariah et al., 2018)
Teilspoofing	Die Audioaufnahmen kommen nicht aus andersartigen echten Audioaufnahmen → künstlich erstellt (L. Zhang et al., 2022)

2.3.2 Manipulationstools

Im Hinblick der Manipulierung haben sich verschiedene Anwendungen etabliert, deswegen ist eine Auswahl dieser in der folgenden Tabelle aufgeführt. Allerdings muss angemerkt werden, dass die Liste nicht alle verfügbaren Tools auflistet.

Tabelle 2: Darstellung verschiedener Manipulationstools

Name	Beschreibung	Quelle
Adobe Audition	• leistungsfähiges digitales Audio-Toolkit • emulieren ein Audiostudio mit Mehrspurproduktionen, Sounddateibearbeitung und digitale Signalverarbeitungseffekte	(Li et al., 2014)
Sound Forge	• ein PC-basiertes Programm zur Bearbeitung von WAV-Dateien • erlaubt das Hinzufügen komplexer Spezialeffekte	(Li et al., 2014)
Pro Tools	• eine integrierte High-End-Audioproduktions- & -bearbeitungsumgebung	(Li et al., 2014)
Wavosaur	• Sound-Editor, Audio-Editor, Wav-Editor-Software für die Bearbeitung, Verarbeitung & Aufnahme von Sounds, wav & mp3-Dateien	(Wavosaur Free Audio Editor With VST and ASIO Support, n.d.)
GoldWave	• Soundeditor • extrahieren von Audio-Dateien aus anderen Multimedia-Dateien	(GoldWave Für Windows 7/8/10/11, 2023)
Audacity	• Programm zur Audio-Bearbeitung und kommt mit den bekanntesten Audio-Formaten zurecht	(Audacity - Gratis Musik Mischen, 2023)
WavePad	• bearbeitet und erstellt eine Vielzahl von Audioaufnahmen	(Tracktion 4 - Audio-Editor, 2023)
Mp3DirectCut	• Audioeditor für kodierte MP3 & AAC • Ohne erneute Kodierung können Tracks geschnitten, zugeschnitten, geteilt, zusammengefügt & überblendet werden	(mp3DirectCut - Cut MP3 and AAC in Windows - Download, 2023)
Reapear	• digitale Audioproduktionsanwendung für Computer • unterstützt eine breite Palette von Hardware, digitalen Formaten & Plugins	(REAPER | Audio Production Without Limits, n.d.)
Nero WaveEditor 2019	• Audiobearbeitungssoftware • zahlreiche Effekte & Einstellungen • unterstützt eine Reihe von Ausgangsaudioformaten	(Nero WaveEditor 2019, 2023)

2.4 Datensatz

Wie bereits in der Einleitung dargelegt, ist es das Ziel einen Datensatz zu erstellen. Hierbei ist es vonnöten zunächst den Begriff Daten als auch Datensatz klar zu definieren. Das Wort Daten kann „als eine beliebige Sammlung von Symbolen …, die eine Reihe von Messungen oder Beobachtungen über ein Ereignis oder einen Vorfall darstellen“[9] angesehen werden (Foxwell, 2020, S.4). Sie „bestehen … aus Zahlen, Zeichen, Wörtern, Bildern und anderen Symbolen, die bestimmte Arten und Merkmale aufweisen und Eigenschaften haben, die direkt darauf schließen lassen, wie man ihre Bedeutungen und Beziehungen

[9] Übersetzt vom Autor

zusammenfasst und visualisiert“[10] (Foxwell, 2020, S. 4). Bei einem Datensatz handelt es sich um eine „Gruppe in bestimmter Hinsicht zusammenhängender Daten einer Datei“ (*Datensatz | Duden*, 2023). Weiterhin wird aufgezeigt, welche Schritte zur Erstellung von einem Datensatz wichtig und unabdingbar sind. Auch werden allgemeine Anforderungen an eben solche dargelegt und die Folgen von Voreingenommenheit sowie Überanpassung abgebildet.

2.4.1 Datensatzanforderungen

Auf der einen Seite kann der Begriff Anforderungen hinsichtlich Qualität sowie Verwendbarkeit betrachtet werden. Auf der anderen Seite können Ansprüche an einen Datensatz aus ethischer, seriöser oder praktikabler Sicht diskutiert werden (Foxwell, 2020). Weiterhin werden in diesem Unterpunkt die Anforderungen an Metadaten sowie an die Datensatzgröße der einzelnen Daten beleuchtet.

2.4.1.1 Anforderungen an die Qualität und Verwendbarkeit

Es ist zu beachten, dass nicht alle Nuancen, die ethisch ausschlaggebend sind, aus rechtmäßiger Sicht gesteuert werden können bzw. gesteuert werden sollten. Im Gegenzug dazu, sind aus rechtlicher Sicht Blickwinkel vorhanden, die aus sachbezogenen Gründen existieren, obschon sie moralisch nicht notwendig sind. Grundsätzlich sollte die Rechtssetzung durchgehend eventuelle ethische Rückschlüsse durchdenken sowie moralische Forderungen abdecken (Gutachten der Datenethikkommission, 2019). Jedoch wird in dieser Bachelorarbeit der Bereich Datensatz nur aus technischer Sicht betrachtet und nicht im Hinblick auf Ethik diskutiert.

Verschiedene Anforderungen zur ersten Begriffsauslegung, die die Qualität und Verwendbarkeit betreffen, sind in Tabelle 3 aufgelistet.

[10] Übersetzt vom Autor

Tabelle 3: Anforderungen an einen Datensatz in Anlehnung an (Foxwell, 2020, S. 5-6)

Bezeichnung	Definition
Genauigkeit	Messungen und Merkmale müssen das Beobachtete korrekt wiedergeben
Relevanz	die für die Analyse ausgewählten Elemente müssen sich direkt auf das Phänomen beziehen
repräsentativ	die Datentypen müssen so ausgewählt werden, dass sie den untersuchten Sachverhalt angemessen widerspiegeln
wohldefiniert	die Bedeutung der Datenelemente muss in einem Schema, Metadaten oder Datenwörterbuch eindeutig definiert sein
vollständig	die ausgewählten Datenelemente müssen alle potenziell relevanten Messungen und Merkmale aufweisen
granular	ausgewählte Datentypen sollten eine ausreichende Bandbreite und Detailgenauigkeit aufweisen, um die gesamte Variabilität der Datenelemente zu erfassen
Fehlervermeidung	• menschliche Fehler bei der Datenerfassung und -aufzeichnung • Umgang mit fehlenden Daten
Quellenverständnis	Verstehen der Quelle und Bedeutung von Ausreißern

2.4.1.2 Metadaten

Zusätzlich zu diesen Anforderungen sind für einen Datensatz Metadaten vonnöten. Diese legen den Inhalt dar und liefern zusätzlich Angaben zum Eigentümer bzw. Ersteller sowie dem Standort. Erstklassige Metadaten sind hilfreich für Leser sowie Analytiker und ermöglichen eine eindeutige Identifizierung der Quellen und erleichtern die Nachvollziehbarkeit und Verifizierbarkeit der Ergebnisse (Foxwell, 2020). In der Theorie sollten die Metadaten die Möglichkeit geben, die Methoden und Ergebnisse zu replizieren (Foxwell, 2020). Aus Tabelle 4 kann entnommen werden, welche Informationen zu einwandfreien Metadaten gehören.

Tabelle 4: Informationen in Metadaten in Anlehnung an (Foxwell, 2020, S.49)

Art	Inhalt
Datenelemente (auch bekannt als Variablen)	• Elementnamen und Beschreibungen • Datentypen • Maßeinheiten • zulässiger Wertebereich • spezielle Kodierung,
Datensatzdatei	• Ersteller/ Eigentümer • Datum der Erstellung • Dateityp und -größe • Nutzung/ Datenschutz/ Freigabebeschränkung • Referenz/ Standort

2.4.1.3 Datensatzgröße

Trotz grundlegender Recherche hinsichtlich des betreffenden Punktes der Datensatzgröße, konnte keine exakte Definition gefunden werden, wie groß ein Datensatz, im Hinblick des Themenbereiches Audiomanipulation, mindestens sein sollte. Aus diesem Grund wurde eine Auswahl von Datensätzen bezüglich der Größe inspiziert. Dabei wurde darauf geachtet, dass die zum Vergleich herangezogenen Datensätzen in irgendeiner Weise einen

Bezug zur Audioforensik oder dem Themengebiet der Manipulation innehaben. Es sei darauf hingewiesen, dass eine einzelne Betrachtung mit Datensätzen, die die Manipulation innerhalb der Audioforensik widerspiegeln, durch die in der Einleitung bereits dargelegten Knappheit an solchen Datensätzen nicht möglich ist. Die zum Vergleich herangezogenen Datensätzen sind in der folgenden Tabelle aufgelistet. Bei der Anzahl der Daten ergibt sich ein Mittelwert von 1837 Daten (vgl. Tabelle 5).

Tabelle 5: Zum Vergleich herangezogene Datensätze

Name	Beschreibung	Anzahl der Daten	Quelle
Forensischer Datensatz für die digitale Multimedia Forensik	Leistungsbewertung von Methoden der Audioforensik, einschließlich Mikrofon-Identifikationsmethoden, akustische Umgebungsidentifikation, Codec-Identifizierung, Erkennung doppelter Kompression	660 Audiodateien, die in vier verschiedenen Sprachen aufgenommen wurden	(Khurram Khan et al., 2017)
Akustische Klangdatenbank in realen Umgebungen für das Verstehen von Klangszenen und die Freihand-Spracherkennung	• Sammlung von Klangszenendaten • Projekt, das für Studien zum Verstehen von Geräuschen unabdingbar ist, einschließlich der Lokalisierung von Geräuschquellen, der Suche nach Geräuschen, der Erkennung und Spracherkennung in realen akustischen Umgebungen	4000 manuell segmentierte Sound-Event-Dateien, die zu 50 Klassen gehören, 80 Dateien pro Klasse.	(Nakamura et al., 2000)
FaceForensics+++	ein forensischer Datensatz	besteht aus 1000 Original-Videosequenzen, die mit vier automatischen Gesichtsmanipulationsmethoden manipuliert wurden: Deepfakes, Face2Face, FaceSwap und NeuralTextures	(FaceForensics++, 2020)
Schusswaffen Audiodatensatz	Die Audios der Waffenmodelle wurden auf YouTube mit Videos gesammelt, die für jedermann zugänglich sind	insgesamt 851 Dateitypen mit 8 Waffenmodellen erhalten	(Gunshot Audio Dataset, 2021)

2.4.2 Fehler bei der Datensatzerstellung

Neben den Anforderungen an einen Datensatz, gibt es auch verschiedene Fehler, die während der Datensatzerstellung auftreten können. Diese sind oftmals das Resultat von Inkorrektheiten auf menschlicher Seite und Planungsfehler (Foxwell, 2020). Tabelle 6 zeigt Fehler auf, die während der verschiedenen Phasen der Datensammlung auftreten können.

Tabelle 6: Ursachen für schlechte Daten in Anlehnung an (Foxwell, 2020, S. 7-8)

Phase	Bezeichnung	Erklärung
Fehler bei der Erstellung	methodische Fehler	schlecht konzipierte Experimente, Erhebungen oder Instrumentierung
	schlechte Dokumentation	unklare Definitionen von Begriffen und fehlende oder verwirrende Schemata, Metadaten oder Datenverzeichnisse
	falsche Spezifizierung von Datentypen und -formaten	Missverständnisse den Zweck und die Auswahl von Datentypen und das Vergessen oder Vermeiden von Standard-Datentypen
Fehler bei der Erhebung	unzugängliche Erhebungsanweisungen und -methoden	Mangel an klaren Verfahren für die Datenerfassung
	unwirksame Durchsetzung der Datenerfassungsregeln	Mangel an Überwachung und Beaufsichtigung
	Fehlinterpretation von Datenelementen	Mangelnde Klarheit über die Bedeutungen
	Transkription/ Typos	keine Überprüfung oder Validierung der aufgezeichneten Daten
	Betrügerische Antworten/ Beobachtungen	Absichtlich missverständliche oder unsinnige Antworten
	fehlende Daten	Versäumnis, Gründe für Nichtantworten zu verstehen und zu korrigieren
	unmögliche, außerhalb des Bereichs liegende Daten	keine Überprüfung der Grenzen von Datenelementen
Fehler bei der Nacherhebung und Analyse	verschieben/ kopieren	Fehler bei der Aufzeichnung oder Lagerung
	Fehlinterpretation	Missverstehen der Bedeutung von Datenelementen oder Antworten
	Aktualität, Daten "verrotten"	Verfall von Zeiten, Orten oder anderen Merkmalen
Fehler bei der Aufzeichnung	Aufzeichnung von Zahlen mit führenden Nullen	0013 statt 13
	Verwendung des Großbuchstabens O für die Zahl Null (0); schwer zu erkennen	
	Vertauschen von Buchstaben oder Zahlen	LA für AL, 32 für 23
	Inkonsistente Verwendung von Namenskonventionen	Italien/ Italia, US/ USA, Deutschland/ Allemagne

2.4.3 Voreingenommenheit

Neben den in 2.3.1 aufgeführten Anforderungen an einen Datensatz und dem Abschnitt 2.3.2 Fehler bei der Datensatzerstellung, ist der Gegenstand dieses Kapitels die enorme Gefahr der Verzerrung, die der Prozess der Datenauswahl mit sich bringt. Diese Auseinandersetzung ist wichtig, da Voreingenommenheit der Qualität der Daten in einem Maße schadet wie keine anderen Ursachen (Foxwell, 2020). Dabei handelt es sich hinsichtlich der Datenerfassung sowie -erhebung um jede absichtliche und unabsichtliche Bevorzugung für einen analytischen Prozess, die zu unkorrekten und irreleitenden Folgerungen führt. Dies stellt einen

Fehler innerhalb der Forschung dar, denn die Absicht ist die Ausführung von Messungen und Beobachtungen hinsichtlich einer Besonderheit mit der Intention, das wahre Verhalten und die innehabenden Eigenschaften zu definieren (Foxwell, 2020).

Aus Tabelle 7 sind Formen von Voreingenommenheit und weiteren Problemen aufgelistet, die diese verursachen können.

Tabelle 7: Formen von Voreingenommenheit in Anlehnung an (Foxwell, 2020, S. 62)

Name	Erklärung
kognitive Voreingenommenheit	• Unfähigkeit, etwas wahrzunehmen bzw. falsche Wahrnehmung • vorurteilsbehaftete Gedanken und Verhaltensweisen
Stichprobenverzerrung	• um gute Daten zu erhalten, brauchen wir eine gute Stichprobe • die gesamte Teilmenge, der Stichprobe, sollte repräsentativ für die Zielpopulation sein • sicherstellen, dass keine wichtigen Datenkategorien übersehen, werden • eine bevorzugte Auswahl eines Forschers kann eine Abweichung von der Zufälligkeit verursachen
Finanzierung/ Interessenskonflikt	• Wer zahlt für die Forschung? • Haben sie Einfluss auf die Ressourcen, Methoden, Schlussfolgerungen oder Veröffentlichungen?
kleine Probe	zuverlässige statistische Hypothesentests erfordern ausreichenden Umfang
Verfügbarkeit/ Convenience-Bias	Auswahl von Methoden bzw. Messungen, da sie leicht zu beschaffen und nicht spezifisch und relevant sind
erwartungsabhängige Verzerrung	Interpretation oder Anpassung von Messungen durch vorherige Überzeugungen
Messungsverzerrung	• Verwendung fehlerhafter Instrumente zur Datenerfassung • testen & validieren der Instrumente vor der Messung

In diesem Absatz wird beleuchtet, inwiefern eine Verminderung von Voreingenommenheit erreicht werden kann. Diese komplett zu eliminieren, ist angesichts der unbewussten Vorurteile, nahezu ausgeschlossen. Ein jeder sollte sich jedoch bewusst sein, dass dieses Problem existiert und in Leidenschaft gezogen werden kann, ohne eigene Beabsichtigung. Das Bemerken von Voreingenommenheit, innerhalb der Forschung, wird durch die Ausbildung, der Erfahrung sowie der Tiefe des Wissens im Hinblick des Untersuchungsgebiets gesteigert. Als weitere Möglichkeit für die Verminderung von Voreingenommenheit, ist die Aufsicht von Experten oder Peer-Review, wobei Gruppen grundsätzlich als fähiger gelten, Fehler sowie Verzerrungen zu erkennen (Foxwell, 2020). Nach Weingart (2012) bezeichnet der Ausdruck Peer-Review „die Prüfung von Publikationen vor ihrer Veröffentlichung und von Forschungsanträgen vor ihrer Bewilligung durch kompetente Kollegen (peers) ausschließlich nach universalistischen Kriterien“ (S.145). „Peer Review hat eine doppelte Funktion. Nach ‚innen‘ soll sie das Vertrauen in die Verlässlichkeit und Wechselseitigkeit der wissenschaftlichen Kommunikation zur Sicherung ihrer Offenheit schaffen. Nach ‚außen‘ gegenüber der Öffentlichkeit soll sie Vertrauen in die Verlässlichkeit des produzierten Wissens herstellen, u. a. um die Ressourcen für die Forschung zu legitimieren“ (Weingart, 2015, S.25). Weiterhin kann es, gegenüber einigen Formen von Voreingenommenheit, helfen aufgeschlossen zu sein, die eigene Hypothese zu falsifizieren (Foxwell, 2020).

Zudem ist zu beachten, dass nicht jeder Ausreißer, extreme oder absurde dokumentierte Messwerte, als Fehler zu deuten ist. Es sollten dennoch alle im Allgemeinen beleuchtet werden, um festzustellen, ob sie als neutral zu klassifizieren sind oder eine nähere Betrachtung benötigen. Zusätzlich stellt der Zweifel als auch die Ablehnung bezüglich von Ausreißern eine bekannte Form der Voreingenommenheit dar (Foxwell, 2020).

2.4.4 Datenbereinigung

Das nun folgende Kapitel enthält Empfehlungen, wie mit Fehlern in der Datensammlung umzugehen ist. Fehler können ungeachtet eines ausführlichen Studiendesigns, kompetent erledigter Vorarbeit sowie Pläne hinsichtlich Fehlervermeidung nicht gänzlich ausgeschlossen werden. Hilfestellung kann dabei das Verfahren der Datenbereinigung geben. Dadurch können sowohl Fehler erkannt als auch berichtigt bzw. zumindest die Konsequenzen auf die Studienergebnisse vermindert werden (Van Den Broeck & Brestoff, 2013).

Hinzuzufügen ist, dass nicht immer ein klarer Schnitt zwischen Fehlern sowie richtigen Werten zu erkennen ist. Ein als verdächtig eingestufter Datenpunkt respektive Muster sowie fehlende Werte benötigen eine ausführliche Begutachtung. Solche Punkte vermögen sowohl auf Unterbrechungen im Datenfluss als auch auf die nicht Disponibilität von Informationen hinzuweisen (Van Den Broeck & Brestoff, 2013).

Vor diesen Hintergründen gelten vorher festgelegte Regeln als wichtiger Teil in der Forschung. Daher bietet es sich an, Datenbereinigung als planmäßigen Prozess bestehend aus Inspektion, Diagnose sowie Handhabung von Datenunregelmäßigkeiten anzusehen (Van Den Broeck & Brestoff, 2013).

Es empfiehlt sich zu versuchen Fehler von sich aus zu entdecken, anstatt durch Zufall während weiteren anderen Studienaktivitäten außerhalb der Datenbereinigung darauf zu stoßen. Dadurch wird Zeit innerhalb der Analyse- und Schreibphase eingespart, da eine wiederholte Untersuchung nach jeder Berichtigung der sämtlichen Datenfehler einen enormen Zeitaufwand darstellt (Van Den Broeck & Brestoff, 2013).

Außerdem sollte dieser Ablauf nicht auf der Originaldatei vollzogen werden, sondern auf einer Kopie. Weiterhin empfiehlt es sich, Aufzeichnungen darüber zuführen, welche Korrekturen vorgenommen werden. Zunächst ist es von Vorteil eine völlige Datenuntersuchung zu vollziehen, um einen Eindruck zu erlangen, wie einschneidend das Ausmaß der Fehler ist. Zusätzlich sollte überprüft werden, ob die einzelnen Datenelemente richtig benannt wurden und über eine Nummerierung verfügen (Foxwell, 2020).

Aus Abbildung 14 ist der vereinzelte Prozess der Datenbereinigung, unterteilt in drei Schritte, zu entnehmen. Dabei erfolgt im ersten Schritt das Screenen der Daten. Dieser Vorgang beinhaltet das Entdecken von Mängeln bzw. Überschuss von Daten, Ausreißer und Unstimmigkeiten, seltsame Muster sowie einen Verdachtsanalysebericht. Daraufhin erfolgt im nächsten Schritt, die sogenannte Diagnose. Dies können Fehler und fehlende

Daten, wahre Extremwerte oder echte normale Werte sein. Es kann auch zu keiner Diagnose kommen, wobei der Datenpunkt aber dennoch als verdächtig klassifiziert wird. Abschließend geschieht durch das Korrigieren, Löschen oder unverändert lassen die Behandlung (Van Den Broeck & Brestoff, 2013, S. 391).

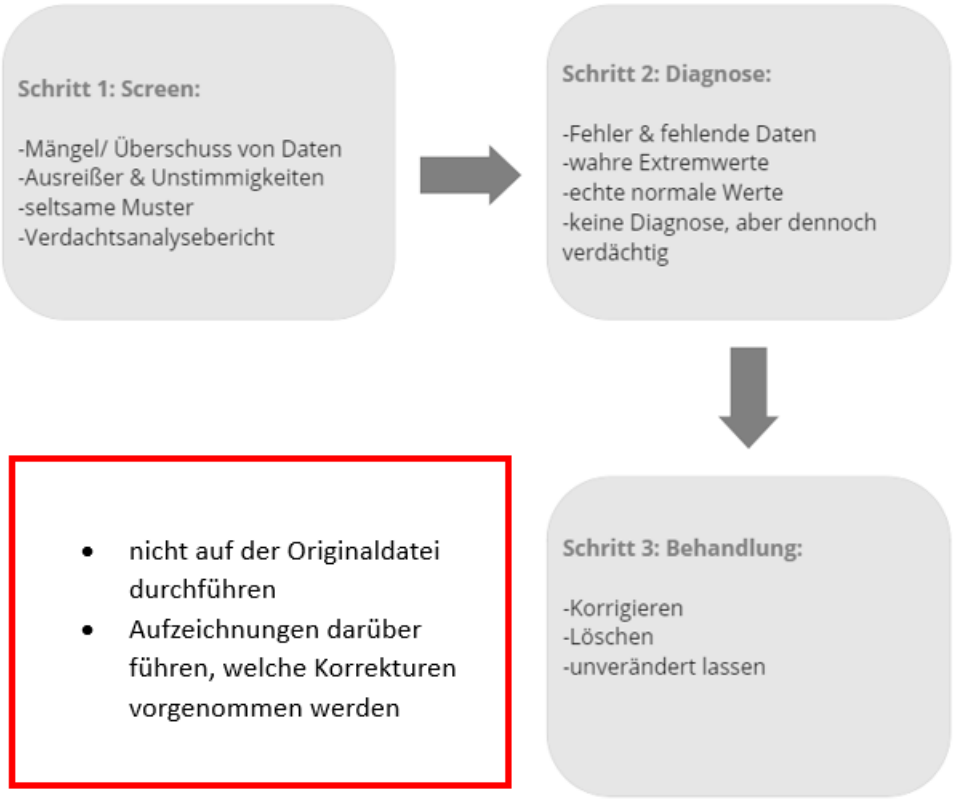

Abbildung 14: Veranschaulichung des Prozesses der Datenbereinigung. Hierbei stammen die Informationen der grauen Kästen von Van Den Broeck & Brestoff (2013, S. 391) und die Informationen des roten Kästchens sind in Anlehnung an Foxwell (2020, S. 77)

2.4.5 Datenanalysetools

Für das Untersuchen, Zusammenfassen und Analysieren von Daten haben sich eine Vielzahl von verschiedenen Werkzeugen etabliert. Ebenso wie bei der Vorstellung unterschiedlicher Manipulationstools, im Abschnitt 2.2.2, kann hier eine Vollständigkeit der Liste nicht gewährleistet werden. Daher ist in Tabelle 8 lediglich eine Auswahl von Applikationen dargestellt, jedoch handelt es sich bei diesen um die verbreitetsten Tools, wobei sich jene Erkenntnis auf verschiedene Erhebungen innerhalb des Arbeitsgebietes stützt (Foxwell, 2020).

Tabelle 8: Tabellarisierung unterschiedlicher Datenanalysetools in Anlehnung an (Foxwell,2020)

Name	Beschreibung
Python	Universalprogrammiersprache, die auf viele verschiedene Problemklassen angewandt werden kann *(General Python FAQ, n.d.)*
Excel	Ein Tabellenkalkulationsprogramm aus dem Haus Microsoft, mit dem Zahlen und Texte in tabellarischer Form dargestellt und ausgewertet werden können (Manuela.Lenz, 2020)
R	Softwareumgebung für statistische Berechnungen und Grafiken *(R: The R Project for Statistical Computing, n.d.)*
SQL	Standardsprache zum Speichern, Bearbeiten und Abrufen von Daten in Datenbanken *(SQL Tutorial, n.d.)*

Die Auswahl eines bzw. einer Kombination von Tools ist nicht nur abhängig von den einzelnen Anforderungen der Datenerstellung sowie –transformation, statistischen Zusammenfassungen und Untersuchungen oder der Art der Datenvisualisierung sowie deren Qualität und Verfügbarkeit von Bibliotheken, sondern es spielen auch individuelle Vorlieben und die Benutzerfreundlichkeit eine Rolle (Foxwell, 2020).

2.4.6 Dateiformate

Die Speicherung von Daten kann in verschiedenen Formaten erfolgen. Bei einem Dateiformat handelt es sich um „die innere logische Struktur einer Datei“ (Böhringer et al., 2014, S. 156). Hierbei gibt es eine Vielzahl von unterschiedlichen Formaten, wobei beispielsweise auf die Auflistung aus dem Buch Kompendium der Mediengestaltung, von Böhringer et al. (2014), verwiesen wird. Diese Aufführung reiht allein bei den Audio- und Videoformaten 13 unterschiedliche Dateiformate auf (Böhringer et al., 2014). Daher sind lediglich, die Audioformate, aus Tabelle 9 zu entnehmen, die für die im Kapitel 3.2 dargelegten Methoden herangezogen werden.

Tabelle 9: Auflistung der für die Bachelorarbeit wichtige Audio- und Videoformate in Anlehnung an (Böhringer et al., 2014, S. 232-233).

Format	Name	Erklärung
MP3	Kurzform von MPEG-2 Audio Layer 3	• verlustbehaftetes Audioformat, das zum weltweiten Standard wurde
MP4	Kurzform von MPEG-4	• Containerformat für multimediale Daten • MP4 kann neben Audio und Video auch Grafiken und Text (Untertitel) enthalten.
WAV	Wave	• Verlustfreies Audioformat (Microsoft) • für den Einsatz in Multimedia-Produktionen ist eine Konvertierung in ein anderes Format (z. B. MP3) anzuraten.

Bei MPEG handelt es sich um die Kurzform für „Moving Pictures Experts Group“ (Böhringer et al., 2014, S. 233). Dieser Begriff bezeichnet „eine Expertengruppe, die sich mit der Standardisierung von Videos und Computeranimationen beschäftigt“ (MP3 | Duden, 2023). Weiterhin stellt auch MPEG ein „Containerformat für multimediale Daten“ dar (Böhringer et al., 2014, S. 233).

Hansch and Rentschler (2012) definieren Containerformate, als „Dateiformate, die unterschiedliche Datenformate enthalten können. Dabei legen Containerformate nur die Struktur und die Art der Aufbewahrung von Inhalten fest und können hinsichtlich ihrer Möglichkeiten stark variieren.“ (S. 25).

2.4.7 Datensatzaufteilung

Ein wichtiger Schritt bei der Datensatzerstellung ist die Datensatzaufteilung. Hierbei gibt es verschiedene Möglichkeiten, den Datensatz aufzuteilen. Die Aufteilung in Trainings-, Validierungs- und Testdaten stellt eine konventionelle Methode dar (Huang et al., 2008). Eine Erklärung für die einzelnen Aufteilungskategorien ist aus Tabelle 10 zu entnehmen.

Tabelle 10: Erklärung der einzelnen Aufteilungskategorien, mittels der Darlegung der Bezeichnung und des Einsatzortes sowie dem Zweck in Anlehnung an Von Der Hude (2020, S.145)

Bezeichnung	Einsatzort und Zweck
Testdaten	Herangezogen bei der Erstellung des Modells
Validierungsdaten	Überprüfung der Modellgüte und gegebenenfalls Veränderung des Modells (diese Daten sind also auch an der Modellbildung beteiligt)
Trainingsdaten	Überprüfung der Güte des endgültigen Modells (diese Daten sind nicht an der Modellbildung beteiligt)

Der Begriff Güte stellt den „Grad der guten Beschaffenheit eines Erzeugnisses, einer Leistung o. Ä“ dar (Güte | Duden, 2023).

Die folgenden Ausführungen stützen sich überdies auf empirische Studien, laut denen die besten Resultate erlangt werden, wenn 20 - 30% als Testdaten und die verbleibenden 70 - 80% als Trainingsdaten gebraucht werden (Gholamy et al., 2018). Aus diesem Grund wird in der vorliegenden Bachelorarbeit eben jener Ansatz detaillierter erläutert und weitere Möglichkeiten nicht näher betrachtet.

Die Unterteilung in einen Trainings- sowie einen Testdatensatz, ist notwendig beim Lernen einer Dependenz aus Daten, um eine Beeinträchtigung durch Überanpassung zu verhindern (Gholamy et al., 2018).

Die Problematik Überanpassung wird in Anlehnung an Aggarwal (2018) wie folgt definiert: „Die Tatsache, dass die Anpassung eines Modells an einen bestimmten Trainingsdatensatz nicht garantiert, dass er eine gute Vorhersageleistung auf unsichtbare Testdaten liefert. Auch wenn das Modell die Ziele auf den Trainingsdaten perfekt vorhersagt. Mit anderen Worten: Es besteht immer eine Lücke zwischen der Trainings- und Testdatenleistung, die insbesondere groß sind, wenn die Modelle komplex sind und der Datensatz klein ist“ (S. 25).

Zunächst wird ein Modell mittels des Trainingssets trainierts und durch die Testdaten wird die Akkuratheit des entstandenen Modells überprüft (Gholamy et al., 2018). Der Ausdruck „Training eines Modells“, kann durch die Darstellung der folgenden Situation verstanden werden. Häufig hat ein Modell hinsichtlich eines physikalischen Phänomens viele unbekannte Parameter, welche mittels der bekannten Daten dezidiert werden sollen. Statistisch gesehen gilt, umso mehr Datenpunkte herangezogen werden, desto akkurater sind die dadurch entstehenden Schätzungen. Unter dieser Perspektive könnte die falsche Schlussfolgerung gezogen werden, dass der beste Weg zur Erörterung der Parameter des Modells, darin liegt, jegliche gegenwärtige Daten für die Eruierung heranzuziehen. Solch eine Konklusion funktioniert nur, falls sichergestellt werden kann, dass das verwendete Modell das dazugehörige Ereignis angemessen definiert. Es bleibt jedoch unbeachtet, dass in der Realität kaum sichergestellt werden kann, dass das derzeitige verwendete Modell faktisch akkurat ist. Sollten jegliche erhältliche Daten genutzt werden, für die Bestimmung der Modellparameter, kann ein Resultat die bereits erwähnte Überanpassung sein (Gholamy et al., 2018).

Für die dargelegte Einteilung, von 20 bis 30% für Trainings- und 70 bis 80% für Testdaten, werden Schätzung der Genauigkeit erhalten, die zum einen vor dem Hintergrund, dass sie den Approximationsfehler keinesfalls unterschätzen und dadurch die Genauigkeit keineswegs überschätzen, gültig sind. Zum anderen sind sie innerhalb der validen Schätzungen die exaktesten, also deren Überschätzung des Näherungsfehlers gilt als die kleinstmögliche (Gholamy et al., 2018). Laut Lubbe (2023) ist der sogenannte Approximationsfehler bzw. Näherungsfehler als „eine feste, nicht stochastische Größe, die die systematische Abweichung zwischen den wahren Parametern und ihrer begrenzten Annäherung durch das Modell charakterisiert“ (S. 1) zu verstehen.

2.4.8 Datenschutzrecht

Bei der Arbeit mit Daten, darf der Aspekt des Datenschutzes nicht vernachlässigt werden, denn dies ist einer der Eckpfeiler innerhalb der augenblicklichen Informations- & Mediengesellschaft. Während des Gebrauches moderner Kommunikationstechnologien gelangen absichtlich oder unabsichtlich zusehends mehr persönliche Daten in den Umlauf. Aus diesen Gründen bietet das Datenschutzrecht einen gesetzlichen Rahmen, um einem Missbrauch vorzubeugen (Kaesler, 2013).

Mittels der Abbildung 15 wird eine, von der Datenethikkommission vorgeschlagene, Darlegung der Regelungen bezüglich des Umgangs mit Daten vorgestellt. Zunächst ist eine vorausschauende Verantwortung notwendig. Dies beinhaltet zum einen die Einschätzung von Konsequenzen, inklusive der Option der Verletzung der Rechte einer anderen Person. Der Punkt Achtung von Rechten beteiligter Personen knüpft daran nahtlos an, denn bei jeglicher Nutzung von Daten ist es notwendig durchgehend eben jene zu achten. Bei der Wohlfahrt durch Nutzen und Teilen von Daten handelt es sich um die Besonderheit, dass es sich bei ihnen um nicht –rivale Güter handelt und sich diese daher nicht aufgrund eines parallelen Gebrauchs durch unterschiedliche Anwender sowie Zwecke abnutzen (Gutachten Der Datenethikkomission, 2018).

Hinter dem Begriff Zweckadäquate Datenqualität verbirgt sich der gewissenhafte Umgang um eine tunlichst genaue Wiedergabe der aktuellen Wirklichkeit oder eine ziemlich passgenaue Vorhersage einer künftigen Realität. Weiterhin ist die risikoadäquate Informationssicherheit wichtig, da verloren gegangene Daten nur schwer wiedererlangt werden können. Außerdem gibt es eine Vielzahl von oftmals übersehenen Angriffsmöglichkeiten durch außerhalb, wodurch sich eine gesonderte Angreifbarkeit hinsichtlich Manipulierung und Zerstörung ergibt. Abschließend ist es wichtig eine interessenadäquate Transparenz zu schaffen, um die Wahrnehmung bzw. eine Verletzung der Datenrechte überhaupt überprüfen zu können (Gutachten Der Datenethikkomission, 2018).

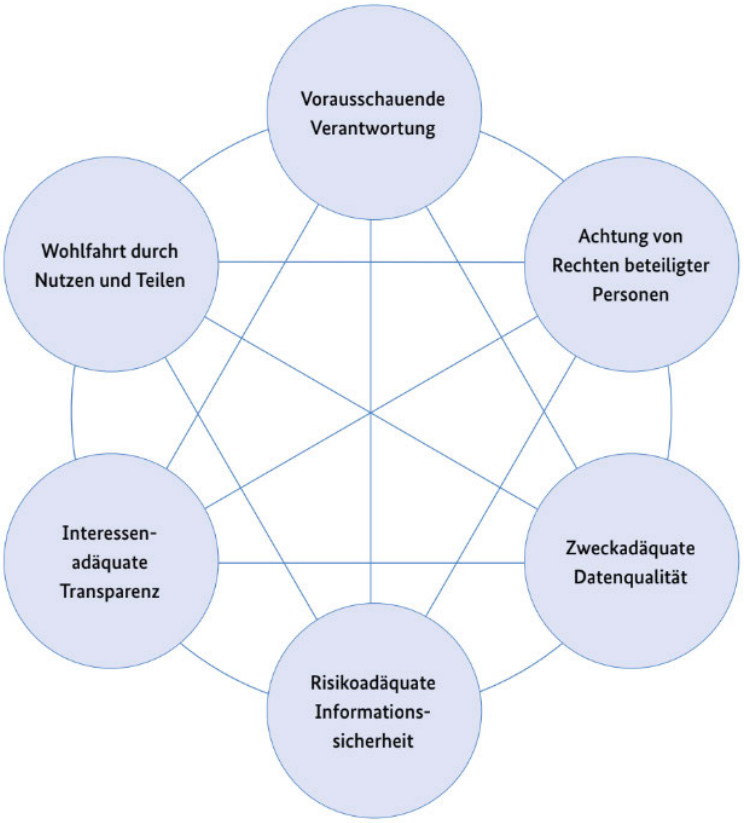

Abbildung 15: Anforderungen an den Umgang mit Daten (Gutachten Der Datenethikkomission, 2018, S. 84)

3 Materialien und Methoden

Hinsichtlich der Erstellung des Datensatzes werden im Abschnitt 3.1 die verwendeten Materialien aufgelistet und erklärt, um die darauffolgende Darlegung der Methoden im Abschnitt 3.2 nachvollziehen zu können. Dadurch soll sichergestellt werden, dass zum einen das Vorgehen und zum anderen die erzielten Ergebnisse, in Kapitel 4, nachvollzogen werden können.

3.1 Materialien

Für das Bearbeiten der Zielaufgabe wurde eine Remotedesktopverbindung zu einem Hochschulrechner der Hochschule Mittweida verwendet.

Darüber hinaus wurden noch weitere Hilfsmittel bzw. Materialien herangezogen. Eine alphabetische Auflistung sowie eine Beschreibung dieser sind der Tabelle 10 zu entnehmen. Weiterhin ist die verwendete Python Versionen mitangegeben. Dies ist wichtig, da Python 3 gegenüber der vorherigen Fassung, Python 2, über bedeutende Veränderungen verfügt (Hetland, 2017). Daher sollte bei einer weiteren Verarbeitung auf die richtige Version geachtet werden, da bei anderen Fassungen nicht garantiert werden kann, dass die gleichen Ergebnisse erreicht werden können.

Tabelle 11: Alphabetische Auflistung und Beschreibung der verwendeten Materialien

Name	Beschreibung	Version
Hochschulrechner	Zugriff mittels Remotedesktopverbindung	
Python	Programmiersprache (*Welcome to Python.org*, 2023)	3.11.2
YouTube	Videoplattform (Miletic, 2021)	

3.2 Methoden

In diesem Kapitel wird nun das Vorgehen und der Ablauf der einzelnen Schritte für das Erstellen eines Audiodatensatzes detailliert erläutert. Zunächst wird im Abschnitt 3.2.1 das Vorgehen der Datenbereitstellung dargelegt. Woraufhin im Folgenden, Punkt 3.2.2 die Verfahrensweise der Datenbereinigung aufgezeigt wird. Der Abschnitt 3.2.3 schließt mit der theoretischen Aufbereitung der Manipulationsidee und die exemplarische Durchführung des Manipulationsschrittes an. Nachfolgend wird die Vorgehensweise bei der Datensatzaufteilung ausgebreitet. Worauf abschließend der Punkt der Aufbereitung der Datensatzstruktur erfolgt.

3.2.1 Datenbereitstellung

Um einen Datensatz zu erstellen, werden, wie in Abschnitt 2.3.1.3 beschrieben, zunächst eine Vielzahl von Daten benötigt. Zu diesem Zweck wurde die Videoplattform YouTube verwendet, da diese eine Vielzahl an öffentlich zugänglichen Videos bereitstellt (Miletic, 2021). Für die Schritte Herunterladen und Trennen der Audiospur wurde ein Python Skript verwendet, welches automatisch die Videos der angegebenen Playlist herunterlädt und zeitgleich die Audiospur trennt, sodass nur diese im Format mp4 gespeichert und in einem eigens für die heruntergeladenen Audioaufnahmen erstellten Ordner abgelegt wird.

Dabei kam die Bibliothek pytube zum Einsatz. Hierbei handelt es sich um eine „leichtgewichtige, abhängigkeitsfreie Pythonic-Bibliothek (und ein Kommandozeilenprogramm) zum Herunterladen von YouTube-Videos“[11] (*Pytube — Pytube 15.0.0 Documentation*, n.d.). Weiterhin benötigt pytube zum einen eine Pythonversion von 3.6 oder höher und zum anderen pip (Pytube, n.d.). Dabei handelt es sich um den Packetmanager von Python (Hetland, 2017). Zu beachten ist, dass die PyPi-Version nicht selten veraltet ist (Pytube, n.d.). PyPi, Kurzform für Python Package Index, stellt ein Software-Verzeichnis von Python dar. Sie ist essenziell, um die durch die Python-Community „entwickelte und geteilte Software zu finden und zu installieren.“ (*PyPI · the Python Package Index*, 2023). Aufgrund der Unstabilität von PyPi-Verison wurde mittels pip pytube mit dem Befehl 'pip install git+HTTPS://GitHub.com/nficano/pytube' aus dem Quellcode installiert.

3.2.2 Datenaufbereitung

Im nächsten Schritt folgte das manuelle Umbenennen der einzelnen mp4 Dateien nach dem in Abbildung 16 dargelegten Schema. Das Muster der Benennung setzt sich aus den folgenden Bausteinen zusammen: eine eindeutige von 00 beginnende Identifizierungsnummer, der Begriff originaleaudiodatei, um aufzuzeigen, dass die Datei nicht manipuliert wurde und die Angabe der Audiolänge in Minuten und Sekunden. Die einzelnen Elemente werden mit einem Unterstrich getrennt, lediglich die Audiolänge untergliedert sich nochmals in Minuten und Sekunden auf, welche mittels eines Bindestriches getrennt werden.

[11] Übersetzt vom Autor

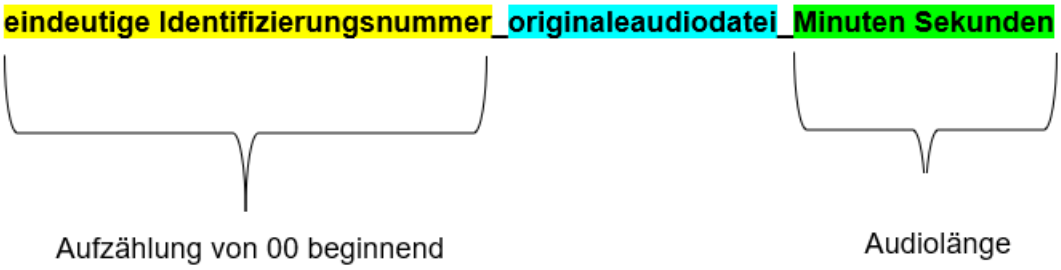

Abbildung 16: Benennungsschema der Audiodateien. Der Aufbau lautet hierfür wie folgt: eine eindeutige von 00 beginnende Identifizierungsnummer, der Begriff originaleaudiodatei, um aufzuzeigen, dass die Datei nicht manipuliert wurde und Audiolängenangabe in Minuten und Sekunden

Ein weiterer Punkt der Datenaufbereitung stellt der in den Grundlagen beschriebener Prozess der Datenbereinigung dar. Dabei wurden die Audios auf Fehler beim Download überprüft. Dies hat, zum Grund, dass sich eine nicht voll funktionsfähige Audiodatei für eine weitere Verarbeitung nicht eignet.

3.2.3 Grundkonzept

In diesem Kapitel erfolgt zunächst die Darlegung des Konzeptes der Manipulierungsidee. Die Idee des Plans basiert zum einen auf dem Paper von Jia et al. (2019), mit dem Titel: „Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis“ und zum anderen sowohl auf dem Projekt, „Deep Fake Audio VIdeo with colab“ von Ajithvallabai (n.d.) als auch auf dem Projekt, „Real-Time Voice Cloning“ von CorentinJ (n.d.), sowie auf der Masterarbeit von Jemine (2019), welche den gleichnamigen Titel „Real-Time Voice Cloning“ trägt. Eine detaillierte Aufschlüsselung, welche Elemente der einzelnen Quellen innerhalb des Konzeptes zu tragen kommen wird durch die ausführliche Aufschlüsselung im Punkt 3.2.4 und dessen Unterpunkte deutlich.

Grundlegend soll ein Text erstellt werden und dieser unter Heranziehen, einer Audiodatei des angefertigten Datensatzes wiedergegeben werden (Jia et al., 2019; CorentinJ, (n.d.), Ajithvallabai, (n.d.), Jemine (2019)). Daraufhin soll, dass erstellte künstliche Audio, in die Referenzdatei des Datensatzes eingefügt werden. Dabei ist es das Ziel eine bestimmte Menge an Audiodaten des Datensatzes zu manipulieren, indem für jede Datei dieser Quantität dieses Konzept durchgeführt wird. Diese Beschreibung stellt per im Punkt 2.3.1 und in Tabelle 1 dargelegter Definition die Manipulationsart Teilspoofing dar.

3.2.4 Datenmanipulierung

In diesem Abschnitt wird der Schritt der Erstellung der artifiziellen Audiodateien in der Theorie beschrieben. Dabei fliesen in die Erklärungen, das in 3.2.3 genannte Paper von Jia et al. (2019) und die Projekte, (Ajithvallabai, n.d.) und (CorentinJ, n.d.), sowie die Masterarbeit von Jemine (2019) mit ein. Da es sich bei den beiden angegebenen Projekten nicht um eine

in schriftlicher Form und somit als unveränderliche anzusehende Abhandlung handelt, stüzen sich die Erläuterungen auf die Versionen vom Zugriffsdatum vom 17.05.2023. Das Projekt von Ajithvallabai (n.d.). enthält neben der Audiogeneration, auch Anleitungen zur Videogeneration sowie die Verknüpfung von Audio und Video. Jedoch wird aufgrund des Bezuges der Bachelorarbeit zum Themenbereich Audio, lediglich dieser Abschnitt näher beleuchtet.

Bei der folgenden Beschreibung der Manipulation steht hinsichtlich des Verständnisses der einzelnen Schritte, die Nachvollziehbarkeit einerseits bezüglich einer möglichen Umsetzung und andererseits das grundsätzliche Verstehen des Vorgangs im Vordergrund, daher wird nicht jede Fachterminologie umfassend definiert, insofern sie für die Erfassung nicht unabdingbar ist.

3.2.4.1 Ablaufbeschreibung

Die grundlegende Idee stammt aus dem Paper von Jia et al. (2019). Das darin beschriebene Modell, besteht aus drei eigenständig trainierten Komponenten und ist in der Lage Audio, von unterschiedlichen Sprechern, zu erzeugen. Ein weiterer wichtiger Punkt ist hierbei, dass die Sprachgenerierung auch bei Stimmen funktioniert, mit denen das Modell nicht trainiert wurde (Jia et al., 2019). Die einzelnen Schritte sind in Abbildung 17 abgebildet.

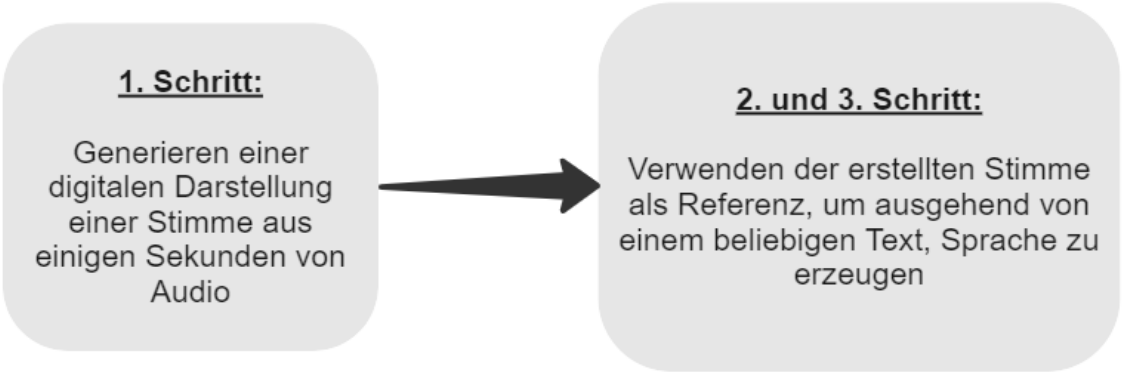

Abbildung 17: Darstellung der einzelnen Schritte in Anlehnung an (CorentinJ, n.d.)

Da das Paper von Jia et al. (2019) keine Umsetzung anbietet, wird die Implementierung von Jemine (2019) mit einem neueren Vocoder, wodurch das Modell in Echtzeit durchführbar ist, dargelegt. Laut Sinha (2010) versucht ein Voice Coder, auch Vocoder genannt, „das Verhalten des menschlichen Sprachproduktionssystems während jedem Sprachsegment nachzubilden“ (S. 113). Weiterhin führt dies „zu einer erheblichen Verringerung der für die Darstellung des Sprachrahmens erforderlichen Datenmenge, da nicht mehr die ursprüngliche Wellenform kodiert wird, sondern einfach eine Reihe von Modelldaten.“ (Sinha, 2010, S113f). Ein Vorteil des Projektes ist, dass es nicht erforderlich ist etwas lokal zu installieren. Weiterhin ist ein Pluspunkt, dass es sowohl für Männer-, als auch für Frauenstimmen funktioniert (Ajithvallabai, n.d.). Nun folgen die einzelnen Instruktionen für eine erfolgreiche Durchführung des Projektes.

Da die Umsetzung von Jemine (2019) als Open Source veröffentlicht ist, wurde diesbezüglich eine grafische Schnittstelle, dargestellt in Abbildung 18, bereitgestellt, damit ein Nutzer ohne weitere Vorbereitung, das Framework verwenden kann. Die grafische Schnittstelle trägt den Namen „SV2TTS toolbox“ (Jemine, 2019).

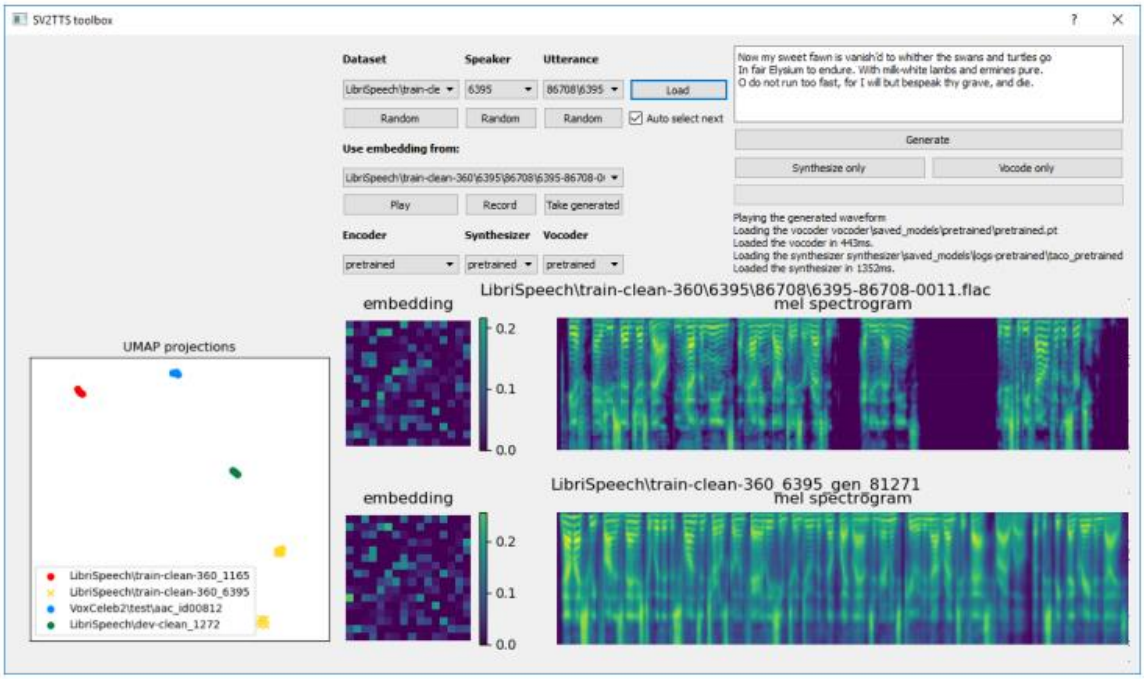

Abbildung 18: Veranschaulichung der grafischen Schnittstelle mit dem Namen SV2TTS toolbox (Jemine, 2019, S. 31)

Als Erstes entscheidet sich der Anwender für eine Audiodatei aus einem Datensatz. Zu diesem Zweck kann die Toolbox sowohl mit einer Vielzahl an bewährten Datensätzen umgehen als auch entsprechend verändert werden, um weitere zu ergänzen (Jemine, 2019).

Nach der Ladung des Ausdrucks wird die Einbettung ausgerechnet. Zusätzlich wird ein Mel-Spektrogramm, ersichtlich in der Abbildung 18, generiert, welches aber nicht für eine weitere Berechnung notwendig ist und somit als Referenzpunkt angesehen werden kann (Jemine, 2019).

Daraufhin wird auf einer erfolgreichen Einbettung folgend, ein Spektrogramm erstellt. Darüber hinaus kann nun mittels des Textfeldes, der grafischen Schnittstelle, hervorgehoben in Abbildung 19, ein willkürlicher Text eingegeben und dadurch ein künstliches Audio erzeugt werden (Jemine, 2019).

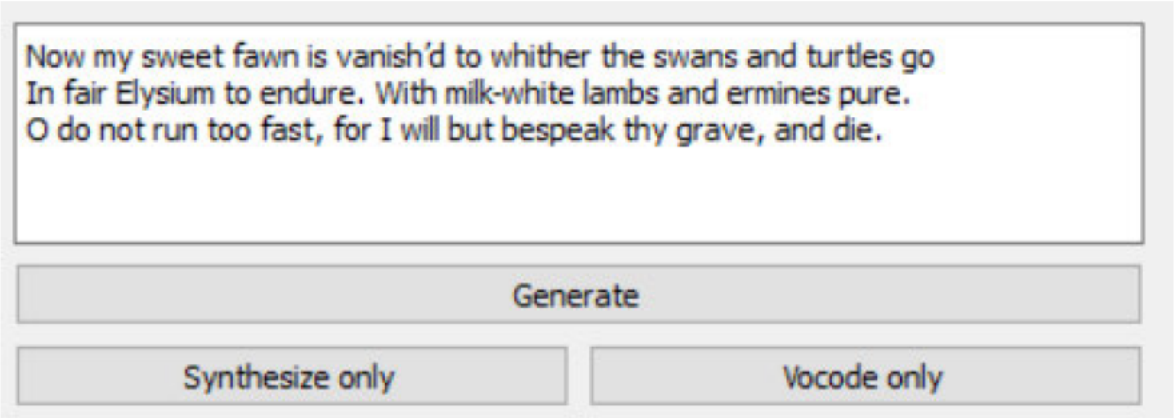

Abbildung 19: Hervorheben des Textfeldes der Toolbox, zur Generierung eines Audios mit der erlernten Stimme (Jemine, 2019)

Für den Text gelten einige Anforderungen, welche jetzt aufgeführt werden. Es können bis zu 12 Wörter dazugegeben werden. Zudem dürfen keine Satzzeichen verwendet werden, da diese nur Geräusche bewirken. Für die Trennung der einzelnen Wörter soll jeweils nur ein Leerzeichen zum Einsatz kommen. Wenn es gewünscht ist, mehrere Textabsätze zu erstellen, dann sollte das generierte Audio unter einer andersartigen Bezeichnung gespeichert werden. Woraufhin weitere Audios angefertigt werden können. Weiterhin wird beim Speichern in der Beschreibung des Projektes das Containerformat wav verwendet, welches für die folgende Schritte als Voraussetzung gilt. Außerdem muss eine Audiodatei der Zielperson zur Verfügung stehen. Bei diesem Audio, sollte lediglich die Person sprechen und etwaige Hintergrundgeräusche sollten nicht vorhanden sein (Ajithvallabai, n.d.).

Das YouTube-Video mit dem Titel Real-Time Voice Cloning Toolbox zeigt den Ablauf der Erstellung unter der Verwendung der Toolbox auf (Corentin Jemine, 2019).

3.2.4.2 Installationsanforderungen

Der nun folgende Abschnitt bietet keine vollständige Installationsanleitung. Jedoch sind für die Installation einige Vorkehrungen zu treffen, welche nun dargelegt werden. Hinsichtlich der Betriebssysteme werden Windows und Linux unterstützt. Zudem wird die Verwendung eines Grafikprozessors befürwortet, um das Training und die Geschwindigkeit verbessern zu können. Eine weitere Empfehlung stellt der Gebrauch von Python 3.7 dar. Zwar sollte auch Python 3.5 oder höher in Ordnung sein, allerdings müssen hierfür Anpassungen getroffen werden. Weiterhin muss ffmpeg, um Audiodateien lesen zu können, installiert werden. Zudem muss auch PyTorch eingearbeitet werden. Abschließend sollten die letzten Voraussetzungen mittels des folgenden Befehls: pip install -r requirements.txt, installiert werden. Zum Schluss kann die Toolbox gestartet werden (CorentinJ, n.d.). PyTorch stellt

ein „schnelles, flexibles Experimentieren und [eine] effiziente Produktion“[12] sicher (*PyTorch*, n.d.).

CorentinJ (n.d.) bietet weitere optionale Implementierungen, die heruntergeladen werden können. Da diese nicht unbedingt notwendig sind und eher für Demo-Zwecke genutzt werden, wird auf eine ausführliche Beschreibung dieser verzichtet.

3.2.5 exemplarische Durchführung

Zum Verständnis des Manipulationskonzeptes wird dieses anhand eines Beispiels exemplarisch durchgeführt. Als Referenzaudiodatei für die Erstellung der künstlichen Stimme, wurde von Jia et al. (2019), das Audio, 1320_00000 (*Audio Samples From “Transfer Learning From Speaker Verification to Multispeaker Text-To-Speech Synthesis,”* n.d.), des Sprechers, LibriSpeech 1320, benutzt. Abbildung 20 zeigt die verschriftliche Form des Audios 1320_000000 dar.

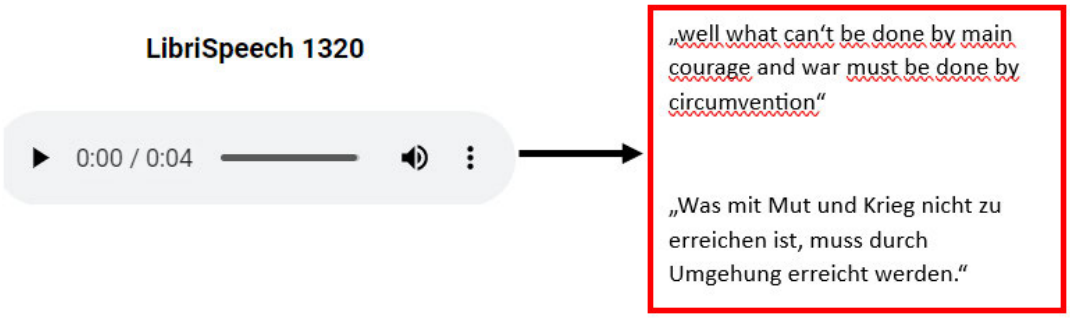

Abbildung 20: Verschriftlichung des Audios 1320_00000 vom Sprecher LibriSpeech 1320. Das Audio 1320_00000 stammt von *Audio Samples From “Transfer Learning From Speaker Verification to Multispeaker Text-To-Speech Synthesis“* (n.d.). Das Audio wurde vom Autor übersetzt

Daraufhin wurde die folgende Äußerung „This work reflects a quest for lost identity, a recuperation of an unknown past“, zu Deutsch „Dieses Werk spiegelt die Suche nach einer verlorenen Identität, die Wiedererlangung einer unbekannten Vergangenheit wider“[13] mit der Stimme des Sprechers erzeugt (*Audio Samples From “Transfer Learning From Speaker Verification to Multispeaker Text-To-Speech Synthesis,”* n.d.). Das Ergebnis stellt das von, Jia et al. (2019), veröffentlichte synthetisiertes Audio mit dem Titel, 1320_00073, dar. Auch dieses ist unter der folgenden Quelle, *Audio Samples From “Transfer Learning From Speaker Verification to Multispeaker Text-To-Speech Synthesis”* (n.d.) abrufbar.

[12] Übersetzt vom Autor

[13] Übersetzt vom Autor

Anschließend wurde die, für das Beispiel, herangezogene Datei, 1320_00073, in die Datei 24_originaleaudiodatei_00-50 eingefügt. Um dies zu erreichen, wurde das Programm Audacity herangezogen.

Als erstes wurde die Tonspur, der Audiodatei 1320_00073, über den Reiter Datei und der Auswahl Öffnen, geladen. In Abbildung 21 ist die Wellenform, dieser Audiodatei abgebildet. Mittels dem Befehl Auswählen und der Wahl Alles aus der erscheinenden Liste wurde die vollständige Audiospur markiert und durch das Kommando Bearbeiten und Kopieren kopiert.

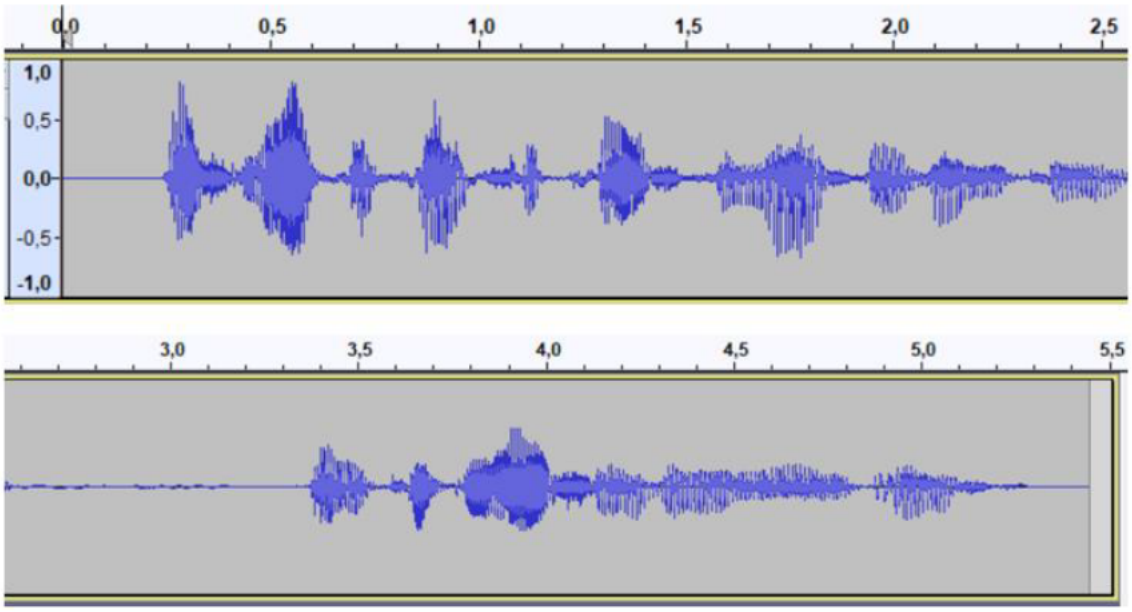

Abbildung 21: Darstellung der Wellenform, der Audiodatei 1320_00073 (*Audio Samples From "Transfer Learning From Speaker Verification to Multispeaker Text-To-Speech Synthesis,"* n.d.). Für die bessere Darstellung wurde das Spektrogramm in zwei Teile unterteilt. Es sei angemerkt, dass die Wellenform normalerweise fortlaufend dargestellt wird.

Weiterhin wurde in einem neuen Fenster die Zielaudio, in unserem Fall 24_originaleaudiodatei_00-50, des eigenen erstellten Datensatzes, mittels der gleichen Vorgehensweise geladen. Anschließend wurde, aus den beiden Fenstern, das Fenster der Zieldatei gewählt und der Cursor willkürlich innerhalb der Audiodatei gesetzt. Dadurch wurde in der Audioaufnahme eine Markierung gesetzt, welche in Abbildung 22 zu sehen ist. Darauf aufbauend, wurde durch das Auswählen von Bearbeiten und Einfügen, die Audiodatei an dem entsprechenden Ort eingesetzt.

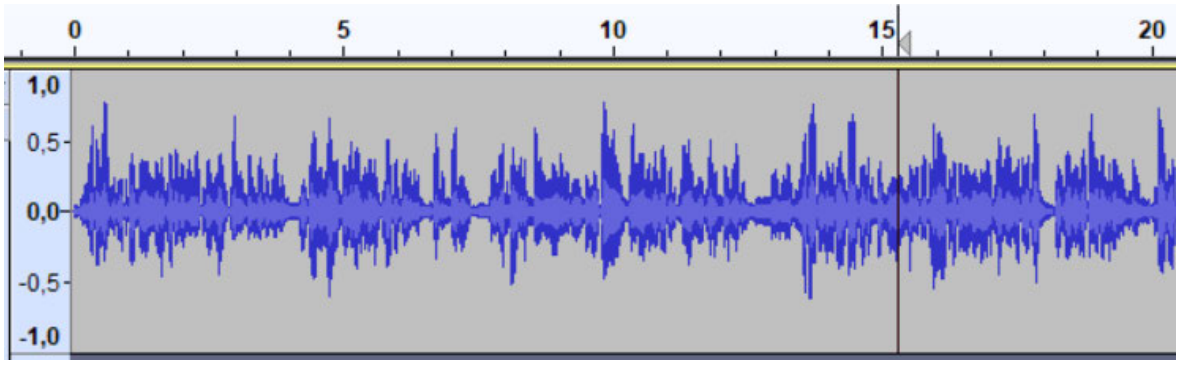

Abbildung 22: Setzen einer Markierung innerhalb der Audioaufnahme, sichtbar durch den grauen Strich in der Wellenform der Audioaufnahme 24_originaleaudiodatei_00-50

3.2.6 Datensatzaufteilung

In diesem Abschnitt wird nun das Vorgehen der Datensatzaufteilung beschrieben. Die Datensatzaufteilung erfolgte nach dem Train-Test Konzept. Auch hierfür sollte ein Pythonskript verwendet werden.

Zunächst wird die Programmierbibliothek NumPy herangezogen. Dabei handelt es sich um „ein Open-Source-Projekt, das numerische Berechnungen mit Python ermöglicht“ (*About Us*, 2023). Weiterhin wird, für die Aufgabe der Datensatzaufteilung, scikit-learn benötigt (Python, 2023). Dabei stellt diese ein Python-Bibliothek dar (Python, 2023). Zudem ist es ein „einfache[s] und effiziente[s] Tool[ ] für die prädiktive Datenanalyse“ (*Scikit-learn: Machine Learning in Python — Scikit-learn 1.2.2 Documentation*, n.d.). Abschließend braucht es noch die scikit-learn Funktion train_test_split(), unter Zuhilfenahme dieser kann der Datensatz aufgespaltet werden (Python, 2023).

3.2.7 Datensatzstruktur

Weiterhin wird die Datensatzspeicherstruktur verbessert, um eine übersichtliche Aufbereitung der Datenstruktur wiedergeben zu können. Dabei spielt nicht nur der Datensatz eine Rolle, sondern auch das verwendete Pythonskript und die für den Datensatz wichtigen Excel-Tabellen.

Weiterhin ist der in 3.2.5 exemplarisch dargelegte Ablauf der Manipulierung zum Verständnis der Konzeptidee mit darzulegen.

4 Ergebnisse

Nachdem im letzten Kapitel die Materialien und Methoden vorgestellt wurden, sollen nun in diesem Kapitel die erzielten Ergebnisse aufgezeigt werden. Dies geschieht, indem, in Abschnitt 4.1, aufgeführt wird, wie viele Daten schlussendlich vorbereitet werden konnten. Daraufhin legt der Punkt 4.2, das Resultat der Datenbereinigung dar. Daraufhin, wird in 4.3 das Ergebnis der exemplarischen Durchführung der Manipulation beschrieben. Da die Datensatzaufteilung in der Theorie vollzogen wurde, wird dieser Schritt, bei der Ergebnisvorstellung übersprungen. Abschließend erfolgt das Aufzeigen der Datensatzstruktur.

4.1 Datenbereitstellung

Insgesamt wurden von der Plattform YouTube 1840 Videos heruntergeladen und die Audiospur getrennt. Aus der folgenden Tabelle können die dafür verwendeten Playlists entnommen werden. Da diese unter ständiger Veränderung liegen, sind zudem das Zugriffsdatum aus Tabelle 12 abzulesen.

Tabelle 12: Angabe der verwendeten Playlists, von YouTube, zur Erstellung des Datensatzes unter Angabe des Kanalnamens, des Zugriffdatums und der Quelle

Name der Playlist	Kanalname	Zugriffsdatum	Quelle
Inside JVA	ZDFheute Nachrichten	30.04.2023	(ZDFheute Nachrichten, n.d.a)
frontal	ZDFheute Nachrichten	30.04.2023	(ZDFheute Nachrichten, n.d.b)
ARD-Morgenmagazin	tagesschau	30.04.2023	(tagesschau, n.d.a)
#mittendrin	tagesschau	30.04.2023	(tagesschau, n.d.b)
Nachrichten & Aktuelles	BR24	30.04.2023	(BR24, n.d.a)
Bayern: News & Hintergründe aus der Region	BR24	10.05.2023	(BR24, n.d.b)
TechTalk	tagesschau	10.05.2023	(tagesschau, n.d.c)

Abbildung 23 zeigt einen Ausschnitt der heruntergeladenen Videos sowie die Angabe des Dateiformates und den Speicherort.

	Name	Extension	Folder Path
2	10 JAHRE AFD Die Gründer packen aus I Spurensuche I frontal.mp4	.mp4	E:\Audiostreams\
3	15 Windräder Planungen für Windpark im Frankenwald BR24.mp4	.mp4	E:\Audiostreams\
4	18000 Euro im Monat Wie Handwerker mit dem Strompreis kämpfen I Abendschau I BR24.mp4	.mp4	E:\Audiostreams\
5	1983 in Deutschland Als wir kurz vor dem III Weltkrieg standen Die Story Kontrovers BR24.mp4	.mp4	E:\Audiostreams\
6	2022 – das Jahr in Bayern Abendschau BR24.mp4	.mp4	E:\Audiostreams\
7	25 Jahre Crashtests ADAC Test- und Technikzentrum Abendschau BR24.mp4	.mp4	E:\Audiostreams\
8	2G Wie Corona-Regeln die Gesellschaft zerteilen I frontal.mp4	.mp4	E:\Audiostreams\
9	49-Euro-Ticket Keine einheitlichen Regelungen BR24.mp4	.mp4	E:\Audiostreams\
10	49-Euro-Ticket Verkauf heute gestartet BR24.mp4	.mp4	E:\Audiostreams\
11	50 Jahre Olympia-Attentat offene Fragen offene Wunden report München BR24.mp4	.mp4	E:\Audiostreams\
12	6-Seen-Platte in Duisburg Kritik an Bauprojekt tagesthemen mittendrin.mp4	.mp4	E:\Audiostreams\
13	67-mal exorziert mit Segen des Bischofs Exorzismus-Fall Anneliese Michel belastet bis heute BR24.n	.mp4	E:\Audiostreams\
14	70 Thronjubiläum von Elisabeth II Queen-Fans feiern in München Abendschau BR24.mp4	.mp4	E:\Audiostreams\
15	75 Jahre Israel Blick auf eine wechselvolle Geschichte BR24.mp4	.mp4	E:\Audiostreams\
16	9-Euro-Ticket Pro und contra BR24 1600.mp4	.mp4	E:\Audiostreams\
17	9-Euro-Ticket So lief das Pfingst-Wochenende BR24.mp4	.mp4	E:\Audiostreams\
18	90 Prozent schaffen Abschluss Clermont-Ferrand-Mittelschule hat besonderes Konzept BR24 Shorts	.mp4	E:\Audiostreams\
19	A93 gen Österreich Viele Lkw wenig Parkplätze – Polizei kämpft gegen Verstöße Abendschau BR24.	.mp4	E:\Audiostreams\
20	Ab wann ist ein Erdbeben gefährlich I BR24 Shorts.mp4	.mp4	E:\Audiostreams\
21	Abfall im Tank Bamberger Busse fahren befüllt mit Biomüll BR24.mp4	.mp4	E:\Audiostreams\
22	Abgehängte Schüler Die Scherben der Corona-Pandemie mehrwert BR24.mp4	.mp4	E:\Audiostreams\
23	Abitur-Prüfungen in Bayern Ist eine Technik-Panne wie in NRW möglich BR24 Shorts.mp4	.mp4	E:\Audiostreams\
24	Ablaufende Corona-Impfzertifikate Was muss ich tun BR24.mp4	.mp4	E:\Audiostreams\
25	ABS fürs Radl – Jugend forscht zur Fahrradsicherheit BR24.mp4	.mp4	E:\Audiostreams\
26	Absage Deutsche Bahn baut kein neues ICE-Werk im Großraum Nürnberg BR24.mp4	.mp4	E:\Audiostreams\

Abbildung 23: Ausschnitt der heruntergeladenen Videos, unter Angabe des Dateiformates und des Speicherortes

Die geringste Dauer hat das Video, BR24 (2022), mit einer Dauer von 0 Minuten und 19 Sekunden. Dahingegen besitzt das längste Video, BR24 (2023) eine Länge von 58 Minuten und 15 Sekunden.

4.2 Datenaufbereitung

Das Ergebnis der Umbenennung ist in Abbildung 24 abgebildet. Da der Datensatz aus 1840 Audioaufnahmen besteht, stellt die Auflistung nur einen Auszug des umbenannten Datensatzes dar.

	Name	Extension	Folder Path
2	01_originaleaudiodatei_12-14.mp4	.mp4	E:\Audiostreams_umbenennung\
3	02_originaleaudiodatei_05-03.mp4	.mp4	E:\Audiostreams_umbenennung\
4	03_originaleaudiodatei_07-00.mp4	.mp4	E:\Audiostreams_umbenennung\
5	04_originaleaudiodatei_05-22.mp4	.mp4	E:\Audiostreams_umbenennung\
6	05_originaleaudiodatei_01-34.mp4	.mp4	E:\Audiostreams_umbenennung\
7	06_originaleaudiodatei_01-31.mp4	.mp4	E:\Audiostreams_umbenennung\
8	07_originaleaudiodatei_13-37.mp4	.mp4	E:\Audiostreams_umbenennung\
9	08_originaleaudiodatei_02-50.mp4	.mp4	E:\Audiostreams_umbenennung\
10	09_originaleaudiodatei_03-26.mp4	.mp4	E:\Audiostreams_umbenennung\

Abbildung 24: Ausschnitt der heruntergeladenen Audios, nach erfolgreicher Umbenennung unter Angabe des Dateiformates und des Speicherortes

Wie bereits aufgeführt, erfolgte während der Datensatzaufbereitung, der Prozess der Datenbereinigung. Durch die manuelle Durchsicht konnte festgestellt werden, dass sich alle heruntergeladenen Audioaufnahmen öffnen lassen und sie sich somit ohne Komplikationen weiterverarbeiten lassen.

4.3 Datenmanipulierung

Nach der exemplarischen Durchführung des Manipulationsschrittes ergibt sich die nun eine neue Audioaufnahme. Die Wellenform dieser ist der Abbildung 25 zu entnehmen. Abschließend wird die nun entstandene Audiodatei unter dem Namen 24_manipulierteaudiodatei_00-55 abgespeichert.

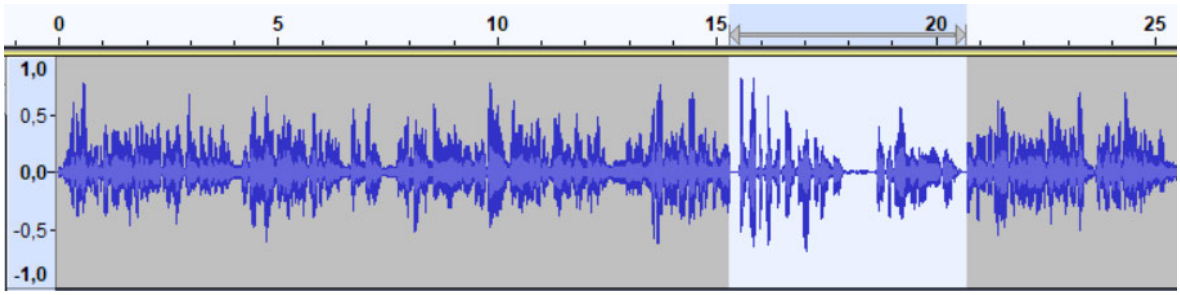

Abbildung 25: Darstellung des Ergebnisses der Manipulation. Der hellblau markierte Bereich stellt die Audiodatei 1320_00073 (*Audio Samples From "Transfer Learning From Speaker Verification to Multispeaker Text-To-Speech Synthesis,"* n.d.) dar.

4.4 Datensatzspeicherstruktur

Die Speicherstruktur des Datensatzes, nach dem Öffnen des Ordners Datensatz, ist in der Abbildung 26 abgebildet.

Name	Änderungsdatum	Typ
Excel Tabellen	28.05.2023 16:24	Dateiordner
Exemplarische Durchführung	29.05.2023 19:37	Dateiordner
Heruntergeladenen_Audioaufnahmen	10.05.2023 15:31	Dateiordner
Python Skript	28.05.2023 14:19	Dateiordner
Umbenannte Audioaufnahmen	18.05.2023 13:08	Dateiordner

Abbildung 26: Veranschaulichung der Speicherstruktur des Datensatzes, nach dem Öffnen des Ordners Datensatz

5 Diskussion

In dieser Bachelorarbeit galt es als Ziel einen Datensatz, zur sequentiellen Lokalisierung von Manipulationen zu erstellen. Die grundlegenden Konzepte sind hierfür zu erläutern. Woraufhin die vorbereitenden Schritte getroffen werden und die Konzeptbeschreibung der Manipulationsidee und die Erläuterung eben dieser sowie der exemplarischen Durchführung erfolgt. Darauf folgend schließt sich die theoretische Beschreibung der Datenaufteilung an. Durch dieses Kapitel erfolgt nun das kritische Auseinandersetzen mit der vorliegenden Bachelorarbeit.

Festzustellen ist, dass es sich allgemein sehr schwierig gestaltet, die drei Themenbereiche, Audio, Manipulierung, Datensatz im Grundlagenteil darzulegen, um auf der einen Seite den vorgegebenen Rahmen nicht um weiten zu überschreiten, aber auf der anderen Seite genug Informationen zu bieten, um sowohl einen grundlegenden Überblick zu erhalten als auch für das weitere Verständnis jegliche benötigten Mittel zu erklären. Allerdings können nun durch die breite Fächerung des Grundlagenteils, viele Informationen genauso für andere Fragestellung, die nur einen der aufgeführten Themenbereiche betreffen, herausgenommen werden.

Dieser Pluspunkt spiegelt sich zudem in der Angabe der verwendeten Python Bibliotheken wider, da dadurch diese auch für andere Datensätze herangezogen werden können, ohne weitere eigene Recherchen.

Bei der Betrachtung der Datensatzgröße in 2.4.1.3, wurden vier Datensätze zum Vergleich hinsichtlich der auszuwählenden Datenmenge, herangezogen. Dabei ergab sich die Durchschnittsmenge von 1837 Daten. Diese Anzahl wurde, wenn auch nur minimal, vom erstellten Datensatz übertroffen. Allerdings sei zu beachten, dass vier Datensätze eine geringe Menge für einen Vergleich darstellen. Weiterhin hatten diese keinen exhaustiven Bezug zur Audiomanipulation und Audioforensik.

Hinsichtlich der Auswahl der einzelnen Playlists wurde im Vorfeld nicht darauf geachtet, ob sie Videos ohne andere Stimmen oder etwaigen Hintergrundgeräuschen, beinhalten. Dies wird zwar bei dem im Schritt 3.2.3 beschriebenen Projekt der Manipulierung, als notwendig angesehen, um eine qualitative artifizielle Stimme erstellen zu können (Ajithvallabai, n.d.). Da jedoch, im Rahmen der vorliegenden Bachelorarbeit, keine Aufnahme von einzelnen Audios erfolgen konnte, wurde die Verwendung von YouTube, als umfassende Videoplattform als Lösung für dieses Problem herangezogen. Aus der in Tabelle 11 aufgeführten Kanalnamen geht hervor, dass es sich um verschiedene Nachrichtenkanäle handelt. Somit kann nicht gänzlich ausgeschlossen werden, dass andere Personen zu Wort kommen oder störende Hintergrundgeräusche vorhanden sind. Dies kann als ein Fehler bei der Erhebung (vgl. Tabelle 6), angesehen werden, und zwar der unzugänglichen Erhebungsanweisungen

und -methoden, welche ein Defizit an eindeutigen Verfahren hinsichtlich der Datenerfassung definiert (Foxwell, 2020). Hierbei hätte noch eine bessere Auswahl, hinsichtlich der unterschiedliche Playlists vollzogen werden können. Hierfür wäre eine mögliche Lösung von den einzelnen Sprechern der unterschiedlichen Aufnahmen in einem Rahmen, der die geforderten Bedingungen erfüllt, Referenzaudios zu erstellen, die dann für die Erstellung der künstlichen Stimme herangezogen werden. Jedoch stellt sich dieses Verfahren aufgrund der hohen Anzahl an Sprechern als utopisch dar.

Zudem gab es bei den einzelnen Playlists keine weiteren Informationen einerseits zu dem jeweiligen Aufnahmeprozess und andererseits zu vorherigen Manipulierungen. Dabei spielt es keine Rolle, ob diese mit einer böswilligen Intention vollzogen wurden oder nur aus Zwecken der Verschönerung erfolgten. Dadurch wird bei beiden Gesichtspunkten die Überprüfbarkeit der Schritte erschwert, da diese nicht lückenlos nachvollzogen werden können.

Im nächsten Punkt erfolgte der Prozess der Datenaufbereitung, wie in Abschnitt 3.2.2 dargelegt. Dabei wurden die einzelnen Schritte manuell vollzogen, dies stellt einen enormen Zeitaufwand dar, auch wenn der Schritt Umbenennung und die Datenbereinigung gleichzeitig stattfinden konnten. Weiterhin muss bei einer manuellen Umbenennung mehr auf Flüchtigkeitsfehler geachtet werden, da eine erneute Durchsicht der einzelnen umbenannten Audiodateien, nach dem in Abbildung 16 dargelegten Schema, sich aufwendiger gestaltet als, beispielsweise bei einem Python Skript einmal die gewünschte Namensgebung anzugeben und diesen Schritt zu automatisieren.

Der Vorteil der manuellen Umbenennung spiegelt sich vor allem in der Angabe der Dauer der einzelnen Audios wider. Da dadurch eine weitere Art der Überprüfbarkeit eingebaut wurde, ob eine Manipulierung der Audiodateien erfolgte. Weiterhin kann direkt aus der Bezeichnung, ohne dem Zwischenschritt Öffnen, die Länge, der etwaigen Datei herausgelesen werden. Ein weiterer Pluspunkt des Vorgehens stellt das Auffinden von Fehlern während des Downloads dar. Allgemein ist ein weiterer Vorteil der Umbenennung hinsichtlich des in Abbildung 16 gelb gefärbten eindeutigen Identifizierungsnummer, dass somit bei einer weiteren Verarbeitung die einzelnen Dateien einfacher angesprochen und zugeordnet werden können. Darüber hinaus spielte der Name der einzelnen Audios keine Rolle im weiteren Verlauf, da für den Datensatz lediglich die Audioaufnahme von Nutzen ist, wodurch sie obsolet sind. Die bessere Zuordnung, wird nach der exemplarischen Durchführung besonders deutlich, durch die Bezeichnung 24_manipulierteaudiodatei_00-55 wird aus der Zahl 24 bereits ersichtlich, auf welche Originaldatei sich die manipulierte Datei bezieht. Weiterhin lässt sich ablesen, dass die Datei manipuliert wurde. Zudem wird dies durch die Änderung in der Zeitangabe deutlich, wodurch auch hier die doppelte Überprüfbarkeit gegeben ist.

Bei der Auseinandersetzung mit den aus Tabelle 3 zu entnehmenden Anforderungen an einen Datensatz fällt auf, dass sich nicht alle aufgeführten Punkte auf diesen Datensatz anwenden lassen. Dem Punkt Fehlervermeidung, wurde, wie bereits erwähnt, im Bereich der Datensatzaufbereitung, Genüge getan. In puncto Vollständigkeit und Granularität

bedarf es weitere Untersuchungen, ob im Datensatz alle möglichen Szenarien einer Audioaufnahme abgebildet wurden.

In der vorliegenden Bachelorarbeit wurde, in Abschnitt 2.4.3, aufgeführt, dass die Voreingenommenheit, als eine der größten Gefahren im Laufe der Datensatzerstellung anzusehen ist (Foxwell, 2020). Diesbezüglich kann im Hinblick auf den erstellten Datensatz, die Aussage getroffen werden, dass jegliche, in der Tabelle 7, beschriebenen Erscheinungsarten nicht allumfassend zutreffen. Jedoch sei angeführt, dass die Auswahl der Beschaffungsmethode auf die Videoplattform YouTube fiel, da diese, wie bereits dargelegt, eine Vielzahl an Videos bereithält. In diesem Punkt könnte die Voreingenommenheit hinsichtlich der Verfügbarkeit in den Sinn kommen, welche die „Auswahl von Methoden bzw. Messungen, da sie leicht zu beschaffen und nicht spezifisch und relevant sind“[14] beschreibt (Foxwell, 2020, S. 62). Allerdings ist zu erwähnen, dass YouTube sehr wohl, den Anforderungen betreffend der Bereitstellung von Videos und demzufolge auch Audioaufnahmen entspricht. Inwiefern diese in der beschriebenen Methode funktionieren, muss noch evaluiert werden, aber in erster Linie wurden Audioaufnahmen benötigt und diese Aufgabe ist erfüllt. Auch für den Punkt, kleine Probe, welcher faktisch eine Voreingenommenheit darstellt (vgl. Tabelle 7), müssten erst weitere Evaluierungen erfolgen, um dies besser einschätzen zu können. Dennoch wurde sich mit diesem Problem schon, in der vorliegenden Bachelorarbeit, kritisch auseinandergesetzt, um aufzuzeigen, dass es ein Bewusstsein für diesen Fall gibt, was wie, von Foxwell (2020) aufgeführt, und in dieser Bachelorarbeit wiedergegeben, helfen kann, eben eine solche Voreingenommenheit zu minimieren.

Bezüglich der Methodik setzen die aufgeführten Projekte das Dateispeicherformat wav voraus, dies liegt noch nicht vor, da das erstellte Pythonskript, die Audioaufnahmen lediglich im Format mp4 abspeichern konnte. Bei einer Umsetzung der dargestellten Manipulationsart ergibt sich somit noch weitere Verarbeitungsschritte, die vorher noch durchgeführt werden und nicht außer Acht gelassen werden sollten, um eine optimale Weiterverarbeitung garantieren zu können. Erschwert wird dies, da das Projekt von Ajithvallabai (n.d.) keine automatische Umwandlung, für viele Dateien bietet.

Weiterhin stellt es sich als herausfordern dar, das in 3.2.3 dargelegte Grundkonzept, kritisch zu bewerten. Gleichwohl geht aus der Darlegung hervor, dass dieses noch aus sehr vielen einzelnen Teilen besteht. Zudem lässt sich nicht aus den herangezogenen Quellen herauslesen, ob die Erstellung der künstlichen Stimme nicht automatisch für eine Vielzahl von verschiedenen Daten von statten geht. Eine manuelle Manipulierung, ist bei einer solchen Datensatzgröße als unakzeptabel anzusehen.

Des Weiteren wurden im Grundlagenteil, der vorliegenden Bachelorarbeit insgesamt vier verschiedene Manipulationsarten vorgestellt. Inwiefern es für eine andere Manipulationsart

[14] Übersetzt durch Autor

eine automatische Implementierung der Manipulationsaufgabe gibt, ist keine wertige Aussage zu treffen. Demgegenüber ist jedoch festzuhalten, dass sich die von Jemine (2019) erstellte Toolbox, auf ein veröffentlichtes Paper stützt, und somit eine fundierte wissenschaftliche Grundlage hat. Eine solche Anwendung konnte hinsichtlich anderer Manipulationsarten, durch die Autorin, nicht gefunden werden. Zusätzlich besteht kein Zwang, exakt diese Projekte zu verwenden. Das Vorhandensein der Daten, in einem nicht manipulierten Zustand, bietet einen weiteren Vorteil, da diese nun auch für themenfremde Aufgaben herangezogen werden können, die eine hohe Anzahl an Audiodateien voraussetzt. Darüber hinaus wurde im ganzen Methodenteil darauf geachtet, alle verwendeten Python Bibliotheken mit anzugeben, um durchgehend eine Wiederholung der einzelnen Schritte bieten zu können. Zudem wurde durch die exemplarische Ausführung gezeigt, dass das Konzept durchaus funktionieren kann, insofern sich die einzelnen Schritte automatisieren lassen.

Allerdings ist zu erwähnen, dass dabei eine englische Audiodatei für das Einfügen in den erstellen Datensatz herangezogen wurde. Sowohl durch die Diskrepanz der Sprachen und den Unterschieden hinsichtlich des Sinnes beider Aufnahmen, wäre die Manipulierung bei einer Überprüfung durch einen Hörer leicht zu erkennen. Daher muss zum einen sichergestellt werden, ob die Stimmenerzeugung auch in der Sprache Deutsch funktioniert und zum anderen benötigt es eine Idee, inwiefern Texte erzeugt werden können, die nach einem Einfügen auch Sinn ergeben.

In der vorliegenden Bachelorarbeit wurde lediglich die Aufteilungsmethode in Trainings- und Testdaten ausführlich beschrieben. Es wurden keine weiteren andersartigen Methoden aufgeführt, einen Datensatz aufzuteilen. Hierdurch und durch das Fehlen weiterer Vergleichsdatensätze, kann nicht hundertprozentig gesagt werden, dass bessere Ergebnisse, durch das Verwenden einer anderen Methode, erzielt werden könnten. Zudem wurde zwar die Bedeutung von Validierungsdaten aufgeführt, aber nicht bei der Durchführung beachtet. Weiterhin wurde das Vorgehen lediglich in der Theorie dargelegt, somit kann keine Aussage getroffen werden, welche prozentuale Aufteilung vorgenommen werden sollte. Allerdings wurde in den Grundlagen im Abschnitt 2.4.7 der prozentuale Bereich umfassend definiert.

Abschließend dürfen die im Abschnitt 2.4.8 aufgeführten Punkte des Datenschutzrechtes in dieser kritischen Auseinandersetzung nicht fehlen. Bei der Erstellung des Datensatzes wurde auf einen gewissenhaften Umgang mit den Daten geachtet. Zudem handelt es sich um öffentlich zugängliche Videos. Inwiefern bereits vorher eine Missachtung etwaiger Datenschutzrechte stattfand, kann nicht festgestellt werden.

6 Fazit und Ausblick

Als zielführende Aufgabe wurde, für die vorliegende Bachelorarbeit, zum einen das Erstellen eines Audiodatensatzes zur sequentiellen Lokalisierung von Manipulation und zum anderen das theoretische Beschreiben der Manipulationsaufgabe, gesetzt. Hierfür erfolgte zunächst die umfassende Darlegung von verschiedenen Grundlagen aus den Bereichen Audio, Manipulation und Datensatz, um eine breitgefächerte Basis zu bieten. Nachfolgend wurden mittels Python, verschiedene YouTube-Videos heruntergeladen und zeitgleich die Audiospur getrennt. Anschließend stand sowohl die Umbenennung als auch der Prozess der Datenbereinigung an, welche manuell durchgeführt wurden. Darauf aufbauend erfolgte die Darlegung des Konzeptes der Manipulierung in der Theorie. Dabei wurden zwei Projekte vorgestellt, wobei eines die Open Source Implementierung, mit einigen Anpassungen, des Papers von Jia et al. (2019) darstellt. Zudem wurde das Konzept exemplarisch durchgeführt. Im Anschluss wurde die Aufteilung des Datensatzes unter Verwendung der Python Bibliothek scikit-learn und der daraus herausgezogenen Funktion train_test_split() in der Theorie dargelegt. Abschließend erfolgte das Aufbereiten der Datensatzstruktur.

Die vorliegende Bachelorarbeit schafft zunächst, durch den breitgefächerten Grundlagenteil eine gute Ausgangsbasis, sich mit den Bereichen Audio, Audiomanipulation sowie Datensatz aus dem Blickwinkel der Forensik zu beschäftigen. Die in dem Punkt Methoden dargelegten Python Bibliotheken bieten eine Repetition der einzelnen Schritte. Auch wenn der Schritt der Manipulierung lediglich theoretisch dargelegt wurde, sind die vorbereitenden Punkte für die Erstellung eines Datensatzes getroffen und dieser kann somit auch für andere Aufgaben der Manipulierung herangezogen werden.

Die Ergebnisse spiegeln wider, dass das Herunterladen der Audiostreams aus den einzelnen Playlists ohne Probleme funktioniert, womit dieses Skript auch für andere Arbeiten, die diese Aufgabe innehaben, herangezogen werden kann. Weiterhin zeigt der Methodenteil, dass das Umbenennen noch ein langwieriges Vorgehen darstellt, obgleich der Prozess der Datenbereinigung parallel abläuft. Die Diskussion zeigte noch etwaige Schwachstellen auf und bot zudem eine kritische Auseinandersetzung mit den gewonnenen Ergebnissen.

Die vorliegende Arbeit dient also in vielerlei Hinsicht als Grundlage für weitere Arbeiten, sei es die Evaluierung oder das Durchführen der in der Theorie dargelegten Manipulierung. Außerdem könnte eine Verbesserung des, als sehr zeitraubend dargestellten Punkt der Datenaufbereitung erfolgen oder noch weiter ausgedehnt werden. Idealerweise könnte die Implementierung des in der Theorie dargelegten Schrittes der Manipulierung erfolgen. Wodurch eine automatische Verfälschung durchgeführt werden kann.

Zusammenfassend, lässt sich festhalten, dass es noch eine Vielzahl an Evaluierungsverfahren sowie Weiterentwicklungen benötigt, um eine abschließende Aussage über die Eignung des Datensatzes zur sequentiellen Lokalisierung von Manipulation zu treffen.

Literaturverzeichnis

Abbasi, A., Javed, A. R., Yasin, A., Jalil, Z., Kryvinska, N., & Tariq, U. (2022). A Large-Scale Benchmark Dataset for Anomaly Detection and Rare Event Classification for Audio Forensics. *IEEE Access*, *10*, 38885–38894. https://doi.org/10.1109/access.2022.3166602

About Us. (2023). https://numpy.org/about/

Aggarwal, C. C. (2018). *Neural Networks and Deep Learning: A Textbook*. Springer.

Ajithvallabai. (n.d.). *GitHub - ajithvallabai/Deepfakes_audio_video: Contains colab files for making audio and video with deep fakes*. GitHub. https://github.com/ajithvallabai/Deepfakes_audio_video

Ajmi, S. A., Hayat, K., Obaidi, A. M. A., Kumar, N., Najmuldeen, M., & Magnier, B. (2022). Digital Audio Forensics: Blind Human Voice Mimicry Detection. *HAL (Le Centre Pour La Communication Scientifique Directe)*. https://doi.org/10.48550/arxiv.2209.12573

Amplitude und Ruhelage der trigonometrischen Funktionen - lernen mit Serlo! (2023, January 6). serlo.org. https://de.serlo.org/mathe/1569/amplitude-und-ruhelage-der-trigonometrischen-funktionen

Audacity - Gratis Musik mischen. (2023, June 5). [Video]. CHIP Online. https://www.chip.de/downloads/Audacity_13010690.html

Audio | *Duden*. (2023). Duden. https://www.duden.de/node/9651/revision/1393123

Audio samples from "Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis." (n.d.). https://google.github.io/tacotron/publications/speaker_adaptation/index.html

Bhagtani, K., Yadav, A. K., Bartusiak, E. R., Xiang, Z., Shao, R., Baireddy, S., & Delp, E. J. (2022). An Overview of Recent Work in Media Forensics: Methods

and Threats. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2204.12067

Böhringer, J., Bühler, P., Schlaich, P., & Sinner, D. (2014). *Kompendium der Mediengestaltung: III. Medienproduktion Print*. Springer-Verlag.

BR24. (2022, October 8). *Volles Fundbüro nach der Wiesn | BR24 #Shorts* [Video]. YouTube. https://www.youtube.com/watch?v=LVVOdTkYsFk

BR24. (2023, April 26). *Hightechland Bayern?! Mission: Zukunft dahoam! Regierungserklärung von Markus Blume | BR24live* [Video]. YouTube. https://www.youtube.com/watch?v=yGhQLBudxME

BR24. (n.d.a). *Nachrichten & Aktuelles* [Playlist] YouTube. https://www.youtube.com/watch?v=pFPmyvmMCtE&list=PLvAQXonIfGbNLbtYt-IE3ljaQJ1Vc-yS0

BR24. (n.d.b). *Bayern: News & Hintergründe aus der Region* [Playlist]. YouTube. https://www.youtube.com/watch?v=pFPmyvmMCtE&list=PLvAQXonIfGbPmrcijb0lBb_CyIQnDYKav

Buermann, M., & Van Meer, T. (2020). *Speech recognition using very deep neural networks: Spectrograms vs Cochleagrams*.

Bühler, P., Schlaich, P., & Sinner, D. (2018). *AV-Medien: Filmgestaltung – Audiotechnik – Videotechnik*. Springer-Verlag.

CI und Hörimplantate | Cochlea Implantat-Zentrum | Klinikum Stuttgart. (n.d.). https://www.klinikum-stuttgart.de/kliniken-institute-zentren/cochlea-implantat-zentrum/ci-und-hoerimplantate#prettyPhoto

Corentin Jemine. (2019, June 12). *Real-Time Voice Cloning Toolbox* [Video]. YouTube. https://www.youtube.com/watch?v=-O_hYhToKoA

Literaturverzeichnis

CorentinJ. (n.d.). *GitHub - CorentinJ/Real-Time-Voice-Cloning: Clone a voice in 5 seconds to generate arbitrary speech in real-time*. GitHub. https://github.com/CorentinJ/Real-Time-Voice-Cloning

Datensatz | *Duden*. (2023). Duden. https://www.duden.de/node/30536/revision/1219439

De Benito-Gorrón, D., Lozano-Diez, A., Toledano, D. T., & Gonzalez-Rodriguez, J. (2019). Exploring convolutional, recurrent, and hybrid deep neural networks for speech and music detection in a large audio dataset. *Eurasip Journal on Audio, Speech, and Music Processing*, *2019*(1). https://doi.org/10.1186/s13636-019-0152-1

Dzulfikar, H., Adinandra, S., & Ramadhani, E. (2021). The Comparison of Audio Analysis Using Audio Forensic Technique and Mel Frequency Cepstral Coefficient Method (MFCC) as the Requirement of Digital Evidence. *JOIN (Jurnal Online Informatika)*, *6*(2), 145. https://doi.org/10.15575/join.v6i2.702

FaceForensics++. (2020, April 10). Kaggle. https://www.kaggle.com/datasets/sorokin/faceforensics

Fischer, W. (2016). *Digitale Fernseh- und Hörfunktechnik in Theorie und Praxis: MPEG-Quellcodierung und Multiplexbildung, analoge und digitale Hörfunk- und Fernsehstandards, DVB, DAB/DAB+, ATSC, ISDB-T, DTMB, terrestrische, kabelgebundene und Satelliten-Übertragungstechnik, Messtechnik*. Springer-Verlag.

Foxwell, H. (2020). *Creating Good Data: A Guide to Dataset Structure and Data Representation*. Apress.

Frequenz • Definition, Einheit und Formel. (2023). Studyflix. https://studyflix.de/ingenieurwissenschaften/frequenz-5716

Literaturverzeichnis

Gautama, T., & Van Hulle, M. M. (1999). Self-Organized Feature Extraction Achieved with a Parameterized Filterbank. *Neural Processing Letters*.

General Python FAQ. (n.d.). Python Documentation. https://docs.python.org/3/faq/general.html#what-is-python-good-for

Ghadekar, N. P., Shetty, N. V., Maheshwari, N. P., Shah, N. R., Shaha, N. A., & Sonawane, N. V. (2023). Non-Facial Video Spatiotemporal Forensic Analysis Using Deep Learning Techniques. *Proceedings of Engineering and Technology Innovation*, *23*, 01–14. https://doi.org/10.46604/peti.2023.10290

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Departmental Technical Reports (CS)*.

GoldWave für Windows 7/8/10/11. (2023). CHIP Online. Retrieved April 19, 2023, from https://www.chip.de/downloads/GoldWave-fuer-Windows-7_8_10_11_12993799.html

Guicking, D. (2016). *Schwingungen: Theorie und Anwendungen in Mechanik, Akustik, Elektrik und Optik*. Springer-Verlag.

Gunshot audio dataset. (2021, June 22). Kaggle. https://www.kaggle.com/datasets/emrahaydemr/gunshot-audio-dataset

Gutachten der Datenethikkomission. (2018). *Datenethikkommission Der Bundesregierung*.

Güte | Duden. (2023). *Duden*. https://www.duden.de/node/61176/revision/1233494

Hansch, P., & Rentschler, C. (2012). *Emotion@Web: Emotionale Websites durch Bewegtbild und Sound-Design*. Springer-Verlag.

Hetland, M. L. (2017). *Beginning Python: From Novice to Professional*. Apress.

Literaturverzeichnis

Huang, G. B., Mattar, M., Berg, T. L., & Learned-Miller, E. (2008). Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments. *Workshop on Faces in "Real-Life" Images: Detection, Alignment, and Recognition*.

Jemine, C. (2019). *Real-Time Voice Cloning* [Masterarbeit]. Université de Liège, Liège, Belgique.

Jia, Y., Zhang, Y., Weiss, R., Wang, Q., Shen, J., Ren, F., Chen, Z., Nguyen, P., Pang, R., Moreno, I., & Wu, Y. (2019). Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis. *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*.

Kaesler, C. (2013). *Recht für Medienberufe: Kompaktes Wissen zu allen rechtstypischen Fragen*. Springer Vieweg.

Khurram Khan, M., Zakariah, M., Malik, H., & Raymond Choo, K.-K. (2017). A novel audio forensic data-set for digital multimedia forensics. *Australian Journal of Forensic Sciences*.

Kuenzel, H. (2023). Der gegenwärtige Stand der forensischen Sprachverarbeitung. *Kriminalistik*, *11*, 676–684.

Kumar, A., Paul, D., Pal, M., Sahidullah, & Saha, G. (2021). Speech frame selection for spoofing detection with an application to partially spoofed audio-data. *International Journal of Speech Technology*, *24*(1), 193–203. https://doi.org/10.1007/s10772-020-09785-w

Li, Z., Drew, M. S., & Liu, J. (2014). *Fundamentals of Multimedia*. Springer Science & Business Media.

Lubbe, D. (2023). ADVANTAGES OF USING UNWEIGHTED APPROXIMATION ERROR MEASURES FOR MODEL FIT ASSESSMENT. *Psychometrika*.

Literaturverzeichnis

Luge, H. (2017). Audioforensik. In *Forensik in der digitalen Welt* (pp. 215–238).

Maher, R. (2016). Gunshot recordings from a criminal incident: Who shot first? *Journal of the Acoustical Society of America*, *139*(4), 2024. https://doi.org/10.1121/1.4949969

Maher, R. (2018). Principles of Forensic Audio Analysis. In *Modern acoustics and signal processing*. Springer Nature. https://doi.org/10.1007/978-3-319-99453-6

Maher, R. C. (2020). Forensische Audioanalyse. In *Handbuch der Audiotechnik* (pp. 1–16).

manipulieren | *Duden*. (2023). Duden. https://www.duden.de/node/151899/revision/1226835

Manuela.Lenz. (2020, April 21). *Was ist Excel? Eine Einführung in das mächtige Tabellen-Tool*. WinTotal.de. https://www.wintotal.de/was-ist-excel/

Meroth, A., & Tolg, B. (2008). *Infotainmentsysteme im Kraftfahrzeug: Grundlagen, Komponenten, Systeme und Anwendungen*. Friedr. Vieweg & Sohn Verlag.

Mertins, A. (2013). Signaltheorie. In *Springer eBooks*. https://doi.org/10.1007/978-3-8348-8109-0

Miletic, B. (2021, August 3). *Was ist Youtube eigentlich? Eine Definition*. Futura, Erkunde Die Welt. https://www.futura-sciences.com/de/was-ist-youtube-eigentlich-definition_5477/#:~:text=Eine%20Definition,-Abgelegt%20unter%3A%20Internet&text=Definition%20von%20Youtube%3A%20YouTube%20ist,der%20meistbesuchten%20Websites%20der%20Welt.

Mill, R. R. (2008). *The application of auditory signal processing principles to the detection, tracking and association of tonal components in sonar* [Doktorarbeit].

Literaturverzeichnis

Möser, M. (2009). *Messtechnik der Akustik*. Springer.

MP3 | Duden. (2023). *Duden*. https://www.duden.de/node/92020/revision/1397208

mp3DirectCut - Cut MP3 and AAC in Windows - Download. (2023). Retrieved April 19, 2023, from https://mpesch3.de/

Multimedia | Duden. (2023). *Duden*. https://www.duden.de/node/99687/revision/1442091

Nakamura, S., Hiyane, K., Asano, F., Nishiura, T., & Yamada, T. (2000). Acoustical Sound Database in Real Environments for Sound Scene Understanding and Hands-Free Speech Recognition. *Language Resources and Evaluation*.

Nero WaveEditor 2019. (2023). CHIP Online. Retrieved April 19, 2023, from https://www.chip.de/downloads/Nero-WaveEditor-2019_57673298.html

Pan, X., Zhang, X., & Lyu, S. (2012). *Detecting splicing in digital audios using local noise level estimation*. https://doi.org/10.1109/icassp.2012.6288260

Penedo, S. R. M., Netto, M. L., & Justo, J. F. (2019). Designing digital filter banks using wavelets. *EURASIP Journal on Advances in Signal Processing*, *2019*(1). https://doi.org/10.1186/s13634-019-0632-6

Peterson, H. A. (2012). Physeal Injury Other Than Fracture. In *Springer eBooks*. https://doi.org/10.1007/978-3-642-22563-5

Pfister, B., & Kaufmann, T. (2017). *Sprachverarbeitung: Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. Springer-Verlag.

PyPI · The Python Package Index. (2023). PyPI. https://pypi.org/

Python, R. (2023). Split Your Dataset With scikit-learn’s train_test_split(). *realpython.com*. https://realpython.com/train-test-split-python-data/

PyTorch. (n.d.). https://pytorch.org/features/

Pytube. (n.d.). *GitHub - pytube/pytube: A lightweight, dependency-free Python library (and command-line utility) for downloading YouTube Videos.* GitHub. https://github.com/pytube/pytube

pytube — pytube 15.0.0 documentation. (n.d.). https://pytube.io/en/latest/

R: The R Project for Statistical Computing. (n.d.). https://www.r-project.org/

Rahman, H., Graciarena, M., Castan, D., Cobo-Kroenke, C., McLaren, M., & Lawson, A. (2022). Detecting Synthetic Speech Manipulation in Real Audio Recordings. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2209.07498

REAPER | *Audio Production Without Limits*. (n.d.). https://www.reaper.fm/

Rechtskunde — leicht verständlich. (1973). In *Gabler Verlag eBooks*. https://doi.org/10.1007/978-3-663-13087-1

Schuller, G. (2023). *Filterbänke und Audiocodierung: Komprimierung Von Audiosignalen Mit Python*. Springer.

scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation. (n.d.). https://scikit-learn.org/stable/

sequenziell | Duden. (2023). *Duden*. https://www.duden.de/node/165596/revision/1248654

Sharan, R. V., & Moir, T. (2015a). *Cochleagram image feature for improved robustness in sound recognition*. https://doi.org/10.1109/icdsp.2015.7251910

Sharan, R. V., & Moir, T. (2015b). Subband Time-Frequency Image Texture Features for Robust Audio Surveillance. *IEEE Transactions on Information Forensics and Security*. https://doi.org/10.1109/tifs.2015.2469254

Sharan, R. V., & Moir, T. J. (2019). Acoustic event recognition using cochleagram image and convolutional neural networks. *Applied Accoustics*, 62–66.

Sinha, P. (2010). Speech Processing in Embedded Systems. In *Springer eBooks*. https://doi.org/10.1007/978-0-387-75581-6

Spannung | *Duden*. (n.d.). Duden. https://www.duden.de/node/169375/revision/1244250

Sprachverarbeitung | *Duden*. (2023). Duden. https://www.duden.de/node/170932/revision/1243777

SQL Tutorial. (n.d.). https://www.w3schools.com/sql/

tagesschau. (n.d.a). *ARD-Morgenmagazin* [Playlist]. YouTube. https://www.youtube.com/watch?v=WUZPwgBF9CA&list=PLkKDSXRppVa4HQ0wcWr08A2zenoo69VtG

tagesschau. (n.d.b.). *#mittendrin* [Playlist]. YouTube. https://www.youtube.com/watch?v=9Za16ZQgRfc&list=PLkKDSXRppVa6NEqZBjcVriy85uKaDLfKj

tagesschau. (n.d.c). *TechTalk* [Playlist]. YouTube. https://www.youtube.com/watch?v=K2VAwUwtjGs&list=PLkKDSXRppVa6DgNTUVhPnZhZ9ZLrFLMgM

Tak, R. N., Agrawal, D. M., & Patil, H. A. (2017). Novel Phase Encoded Mel Filterbank Energies for Environmental Sound Classification. *Pattern Recognition and Machine Intelligence*, 317–325.

Tracktion 4 - Audio-Editor. (2023, October 1). [Video]. CHIP Online. https://www.chip.de/downloads/WavePad_28979806.html

Van Den Broeck, J., & Brestoff, J. R. (2013). *Epidemiology: Principles and Practical Guidelines*. Springer Science & Business Media.

Von Der Hude, M. (2020). Predictive Analytics und Data Mining. In *Springer eBooks*. https://doi.org/10.1007/978-3-658-30153-8

Wang, D., & Brown, G. C. (2006). Fundamentals of Computational Auditory Scene Analysis. In *IEEE eBooks*. https://doi.org/10.1109/9780470043387.ch1

Wavosaur free audio editor with VST and ASIO support. (n.d.). https://www.wavosaur.com/

Weingart, P. (2012). Wissenschaftssoziologie. In *Handbuch Wissenschaftspolitik* (pp. 141–155).

Weingart, P. (2015). *Wissenschaftssoziologie*.

Weinzierl, S. (2008). *Handbuch der Audiotechnik*. Springer.

Welcome to Python.org. (2023, May 22). Python.org. https://www.python.org/about/

Welle | Duden. (2023). Duden. https://www.duden.de/node/203795/revision/1370554

Werner, M. (2008). *Signale und Systeme: Lehr- und Arbeitsbuch mit MATLAB®-Übungen und Lösungen*. Springer Science & Business Media.

Werner, M. (2010). *Nachrichtentechnik: Eine Einführung für alle Studiengänge*. Springer Science & Business Media.

Werner, M. W. (2019). Digitale Signalverarbeitung mit MATLAB®. In *Springer eBooks*. https://doi.org/10.1007/978-3-658-18647-0

Winzker, M. (2023). *Elektronik für Entscheider: Grundwissen für Wirtschaft und Technik*. Springer-Verlag.

Xiang, Z., Bestagini, P., Tubaro, S., & Delp, E. J. (2022). Forensic Analysis and Localization of Multiply Compressed MP3 Audio Using Transformers. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp43922.2022.9747639

Literaturverzeichnis

Yi, J., Zhang, D., Tao, J., Tian, Z., Fan, C., Ma, H., & Fu, R. (2022). SceneFake: An Initial Dataset and Benchmarks for Scene Fake Audio Detection. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2211.06073

Zakariah, M., Khan, M. S., & Malik, H. (2018). Digital multimedia audio forensics: past, present and future. *Multimedia Tools and Applications*, *77*(1), 1009–1040. https://doi.org/10.1007/s11042-016-4277-2

ZDFheute Nachrichten. (n.d.a). *Inside JVA* [Playlist]. YouTube. https://www.youtube.com/watch?v=ehFMxoKKzhs&list=PLdPrKDvwrog4UuKY7nKjEeWzfDyv-6Aid

ZDFheute Nachrichten. (n.d.b). *frontal* [Playlist]. YouTube. https://www.youtube.com/watch?v=nCSowzAvo4Q&list=PLdPrKDvwrog6_sr6vetqlqb2g3KIfl9YW

Zhang, L., Wang, X., Cooper, E., & Evans, N. (2022). The PartialSpoof Database and Countermeasures for the Detection of Short Fake Speech Segments Embedded in an Utterance. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

Zhang, Z., Zhao, X., & Yi, X. (2022). ASLNet: An Encoder-Decoder Architecture for Audio Splicing Detection and Localization. *Security and Communication Networks*, *2022*, 1–9. https://doi.org/10.1155/2022/8241298

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Chemnitz, den 30.05.2023

Carmen Maria Wührl