FORSCHUNG

AUDIOFORENSIK, PARTIELLES AUDIO-MATCHING UND AUDIO-PHYLOGENIE-ANALYSE:

TECHNOLOGIEN FÜR CONTENT-VERIFIZIERUNG UND MEDIENMANAGEMENT

PATRICK AICHROTH, HANNA LUKASHEVICH

Audio-Forensik, Partielles Audio-Matching und Audio-Phylogenie-Analyse sind drei sehr spezialisierte Technologien, die jedoch in einer Vielzahl von Anwendungen wie Manipulationserkennung, Duplikaterkennung, Rechtemanagement, Qualitätskontrolle und Programmanalyse eingesetzt werden können. Der Artikel beschreibt die entsprechenden Ansätze und erklärt, wie sie – insbesondere in Kombination – zur Lösung sehr unterschiedlicher Probleme eingesetzt werden können.

Audio forensics, partial audio matching and audio phylogeny analysis are three fairly specialised technologies, but they can be used in a wide range of applications such as manipulation detection, duplicate detection, rights tracking, quality control and program analysis. The article describes the corresponding approaches and explains how they can be used to address very different problems.

Einführung

Bei automatischer Metadatenextraktion und KI denkt man zunächst an besonders bekannte Werkzeuge wie Gesichts- und Objektdetektion und -erkennung oder Sprach- und Sprechererkennung. Es gibt jedoch auch weniger bekannte Technologien, die bei erstaunlich vielen Fragestellungen hilfreich sein können.

Dazu gehören Audioforensik, Audio-Partial-Matching und Audio-Phylogenie-Analyse. In diesem Artikel zeigen wir, wie sie funktionieren und für welche Anwendungsgebiete sie eingesetzt werden können, z. B. für Audio-Verifikation, aber auch für die Erkennung von Wiederverwendung und von Duplikaten, für Rechtemanagement und Programmanalyse.

Audio-Verifikation: Probleme, Ansätze und Herausforderungen

Audio-Fakes: Man kann seinen Ohren nicht immer trauen

Fakes lassen sich heute relativ kostengünstig und einfach erstellen, denn Content ist in großen Mengen verfügbar, Bearbeitungstools kosten nicht viel und es gibt nahezu unbegrenzte Verbreitungswege. Besonders einfach lässt sich ein Fake erstellen, indem man vorhandenes Material nutzt und bearbeitet und so einen neuen, verfälschten Kontext oder veränderte Botschaften erzeugt.

Dies gilt insbesondere für Audio-Material. Sprache ist unser wichtigstes Kommunikations- und Informationsmittel und stellt eine perfekte Angriffsfläche für Manipulationen dar. Die Bearbeitung von Audiomaterial ist vermutlich der effizienteste Weg, um Fälschungen zu erstellen. Das Kopieren, Einfügen oder Entfernen einiger weniger Worte oder Sätze aus Sprachaufnahmen kann eine völlig verzerrte oder neue Bedeutung erzeugen und so erheblichen Schaden anrichten. Ist die Bearbeitung handwerklich gut gemacht, ist sie nur sehr schwer zu erkennen. Es überrascht deshalb nicht, dass es in den letzten Jahren zu einem spürbaren Anstieg von entsprechenden „Verdachtsfällen" gekommen ist – vor allem im Journalismus, aber auch bei Betrugsfällen und bei polizeilichen Ermittlungen.

In all diesen Fällen benötigt man eine Einschätzung darüber, ob die Aufnahmen als authentisch im obigen Sinne einzuschätzen sind, d. h. ob sie mit dem angenommenen Aufnahmekontext (z. B. Zeit und Ort) übereinstimmen und nicht nachträglich bearbeitet wurden.

Fälschungen erkennen: Authentifizierung vs. Manipulationserkennung

Es gibt generell zwei sehr unterschiedliche Ansätze zur Verifikation von Inhalten: Die Inhaltsauthentifizierung mit kryptographischen Mitteln, d. h. digitalen Signaturen, und die Manipulationserkennung und -lokalisierung, d. h. Detektion spezifischer Manipulationen.

Digitale Signaturen können verwendet werden, um Content in einem fest definierten „Referenzzustand" im Content-Lebenszyklus zu signieren, z. B. während oder direkt nach der Aufzeichnung. So kann nachgewiesen werden, dass der Content nicht nachträglich bearbeitet wurde: Jede Änderung des Inhalts führt zu einer fehlgeschlagenen Validierung/Prüfung. Digitale Signaturen scheinen damit die perfekte Lösung für Content-Verifikation zu sein. Doch gibt es dabei auch eine Reihe von Nachteilen:

Erstens ist es je nach Anwendungsfall häufig erforderlich, digitale Signaturen robust gegen Änderungen bei Produktion und Verbreitung von Inhalten zu machen, die in der Praxis üblich sind, den Content aber inhaltlich nicht verändern, wie z. B. Formatkonvertierungen oder Transcodierungen. Zweitens erfordern digitale Signaturen besondere Modifikationen an der für Aufnahme und Verarbeitung von A/V-Daten verwendeten Hard- und Software, was mit einem entsprechenden Aufwand und oft mit Einschränkungen der Usability verbunden ist – weshalb sich digitale Signaturen in diesem Bereich auch nie in der Breite durchgesetzt haben. Drittens liefert die Authentifizierung nur eine Antwort auf die Frage, ob etwas manipuliert wurde, aber nicht, *wie genau* die Inhalte verändert wurden, was für eine detaillierte

Analyse aber von entscheidender Bedeutung ist. Aufgrund dieser Einschränkungen ist einleuchtend, dass die Verwendung digitaler Signaturen (ggfs. mit medienspezifischen Modifikationen) grundsätzlich sinnvoll ist und besonders bei „kontrollierten" Prozessen wie z. B. autorisierten Aufnahmen durch Behörden das Mittel der Wahl, aber für die Anforderungen einer Verifizierung von Inhalten beliebiger Herkunft nicht ausreichend.

Verfahren zur **Manipulationserkennung und -lokalisierung** bei Bild-, Video und Audiodaten unterliegen diesen Einschränkungen nicht: Sie basieren auf der Überlegung, dass Aufnahme und Verarbeitungsschritte jeweils spezifische Spuren oder „Footprints" im Material hinterlassen, die sich grundsätzlich detektieren und für eine Bestätigung bzw. Falsifizierung von Annahmen und Aussagen über den Inhalt verwenden lassen, z. B. darüber, ob mit einem bestimmten Gerät aufgenommen und nicht nachträglich bearbeitet wurde. Sie setzen keine Änderungen bei Aufnahme- und Verarbeitungswerkzeugen voraus und erlauben auch detailliertere Analysen dazu, wie und wo genau Inhalte verändert wurden, was wiederum auch Rückschlüsse auf Vorgehensweise und Motivation liefern kann. Diese Flexibilität und Leistungsfähigkeit hat allerdings einen Preis: Es werden vielerlei Detektoren benötigt, um die Bandbreite möglicher Manipulationen abzudecken, und deren Realisierung ist oft anspruchsvoll und erfordert spezifisches Knowhow bzgl. Signalanalyse, Statistik oder maschinellem Lernen, den typischen Werkzeugen einer medienforensischen Analyse.

Audioforensische Detektoren

Aus der Vielzahl audioforensischer Detektoren (die selbstverständlich alle auch auf Audioströme in Videos anwendbar sind) sollen im Folgenden drei beschrieben werden, die für den Anwendungsbereich und das Vorhaben besonders wichtig sind.

Alle drei beruhen auf der Erkennung von sog. „acquisition footprints" und „editing footprints": **„Acquisition footprints"** (siehe Bild 1) sind Aufnahmespuren, die u. a. durch Aufnahmegerät, Aufnahmecodierung oder ähnlich entstehen. Sie sind bei Originalaufnahmen *immer* vorhanden und können *sowohl detektiert als auch*, z. B. bzgl. Aufnahmegerät, Aufnahmezeit etc., *verifiziert werden*. „Editing footprints" dagegen sind Spuren, die durch eine Bearbeitung des Materials, z. B. das Entfernen von Abschnitten, Einbringen von Material aus anderer Quelle, nachträgliche Codierung etc., entstehen, oft auch in Form von Inkonsistenzen der vorgenannten „acquisition footprints". Sie sollten bei Originalaufnahmen (normalerweise) nicht vorhanden sein und können infolgedessen *nur detektiert*, aber nicht verifiziert werden.

Als erstes Beispiel soll hier die **ENF-Analyse** dienen. Dieses Verfahren basiert auf der Analyse der elektrischen Netzfrequenz (ENF) [1], die in vielen Aufnahmen als mehr oder weniger hörbares Brummen existiert (siehe Bild 2). Sie schwankt je nach Aufenthaltsort innerhalb der vorgegebenen Toleranzen für das Stromnetz leicht um die 50 bzw. 60 Hz im Zeitverlauf, wobei diese Schwankungen aufgrund der engen Verflechtung von Stromnetzen recht einheitlich sind.

Die ENF ist damit in zweierlei Hinsicht für audioforensische Analysen interessant: Erstens ist es möglich, ENF-Referenzdatenbanken aufzubauen oder ENF-Daten von Energieprovidern einzuholen und durch Extraktion der ENF als „acquisition footprints" und entsprechende Abgleichverfahren (sog. „Pattern Matching") Aufnahmezeitpunkt und -ort einzugrenzen oder gar zu bestimmen [2–7]. Dazu sind aber relativ lange Abschnitte mit extrahierbarer ENF erforderlich, oder

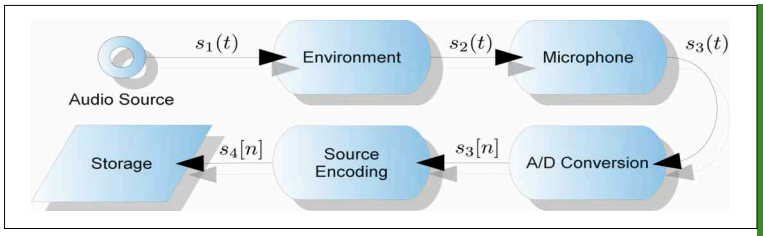


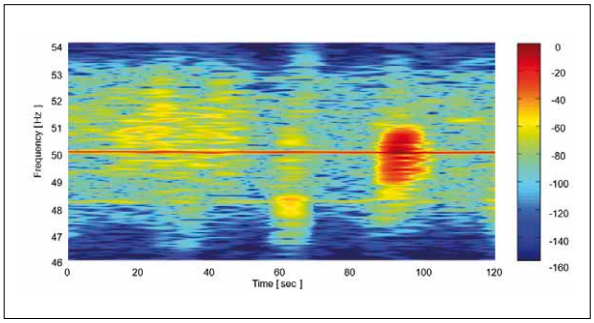
Bild 1. Schritte im Aufnahmeprozess Quelle: Fraunhofer IDMT

um die Plausibilität einer zeitlichen Abfolge von Abschnitten zu prüfen (siehe Bild 4) [10]. Zweitens schlagen sich Schnitte und Manipulationen in Form von Inkonsistenzen in der ENF v.a. in Form von Phasensprüngen als „editing footprints" im Material nieder, die detektiert werden können [9–11], wobei eine Analyse der Obertöne noch zu einer deutlichen Verbesserung führen kann [12].

Eine ENF-Analyse kann natürlich nur dann vorgenommen werden, wenn auch ENF im Signal vorhanden ist. Da diese auch akustisch oder via Induktion in die Aufnahme gelangen kann, z. B. wenn ein Stromkabel in der Nähe verläuft, kommt dies auch bei Mobilgeräten und Smartphones vor, die ENF tritt aber bei älteren Aufnahmen besonders häufig und ausgeprägt auf.

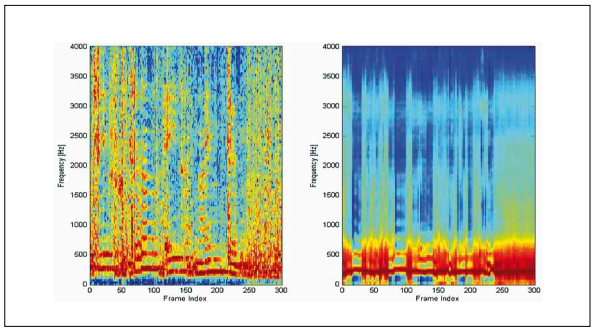
Ein zweites Verfahren ist die **Mikrofonklassifizierung** [13, 14]. Das Mikrofon hat den mit Abstand stärksten Einfluss auf den Aufnahmeprozess, und man kann eine Kombination von Signalanalyse und lernbasierten Verfahren einsetzen, um zu einer gegebenen Aufnahme eine originäre Klangquelle und den Einfluss des Aufnahmegeräts bzw. Mikrofons zu ermitteln (siehe Bild 5), d. h. man kann die Frequenzantwort des entsprechenden Mikrofons/Aufnahmegeräts berechnen (siehe Bild 6). Diese Technik kann einerseits verwendet werden, um den Typ des Aufnahmegeräts einer Aufnahme zu bestimmen bzw. zu verifizieren [15]. Andererseits lassen sich damit auch Manipulationen entlarven, bei denen Material aus unterschiedlichen Typen von Aufnahmegeräten verwendet wurde [16].

Als drittes Beispiel schließlich sei die **Coding-Analyse** einschließlich des sog. „Inverse Decoding" [17] genannt. Bei letzterem geht es darum, durch eine Rekonstruktion der Arbeitsschritte eines Codecs im Signal nicht nur zu erkennen, ob entsprechende verlustbehaftete Codierschritte im Audiomaterial zu finden sind, sondern auch welche Codierparameter wie z. B. die Bitrate dabei verwendet wurden. Initiale einfache Ansätze für MP3 [18] wurden zu einem kombinierten Detektor für zahlreiche Codecs und Codierparameter für MP3, AAC, MP3PRO [19] sowie später für GSM Sprachcodecs und A-law und mu-law PCM [20] erweitert. Diese Verfahren können nicht nur verwendet werden, um Aussagen bzgl. Aufnahmegeräten und Aufnahmeprozess zu prüfen, sondern



Quelle: Fraunhofer IDMT

Bild 2. Audiomaterial mit ENF



Quelle: Fraunhofer IDMT

Bild 3. Aufnahme und originäre Klangquelle

auch um festzustellen, ob Material mit unterschiedlichen Codierungsspuren (Codec, Bitrate) zusammengeschnitten wurde. Durch die Erkennung von Inkonsistenzen bei den Spuren bzgl. des Framing Grid können für viele Codecs sogar Manipulationen innerhalb von Material erkannt werden, das der gleichen Aufnahme entstammt. Eine wichtige Ergänzung zum „Inversen Decoding" stellen Verfahren auf Basis von maschinellem Lernen bzw. Convolutional Neural Networks dar, mit denen sich die Existenz und Parameter von mehreren Codierungsschritten detektieren lassen [21].

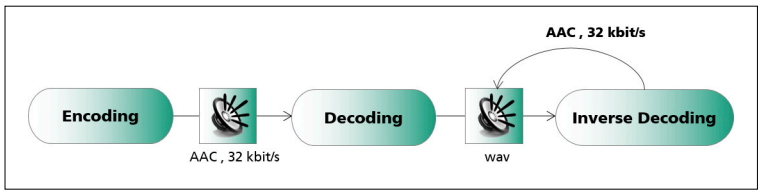
Analyse von „Kopien": Partielles Audio-Matching und Phylogenie-Analyse

Eine umfassende audioforensische Analyse setzt in vielen Fällen voraus, auch mit Kopien bzw. Teilkopien umzugehen, und zwar aus zwei Gründen: Erstens, weil bei Fakes oft Material wiederverwendet wird, was vorher schon einmal da war. Zweitens, weil bei der Verbreitung von Material z. B. über Videoplattformen bewusst oder unbewusst zahlreiche Kopien und Teilkopien entstehen und es oft sehr schwer oder gar unmöglich ist zu sagen, in welcher Reihenfolge diese entstanden sind und welches das Original war. Für solche Problemstellungen sind zwei Analyseverfahren besonders hilfreich, die im Folgenden beschrieben sind: Partielles Audio-Matching und Audio-Phylogenie-Analyse.

Beim „klassischen" Audio-Matching [22] werden sog. "Fingerprints", eine Art Zusammenfassung charakteristischer Audio-Merkmale zu einem Inhalt, verwendet, um den Ausschnitt eines Stücks (typischerweise ein Musikstück) in einer Referenz-Datenbank wiederzufinden bzw. zu identifizieren. **Partielles Audio-Matching** nutzt ebenfalls „Fingerprints" (siehe Bild 7), verwendet aber ein anderes Matching-Verfahren mit einer anderen Zielstellung: Es zielt auf die Erkennung und Lokalisierung von Teilduplikaten bzw. Wiederverwendung von ggfs. sehr kurzen Abschnitten (typischerweise bis zu ca. 3–4 sec) in einem Datenset ab, ohne dass vorher bekannt wäre, ob und welche Teilüberlappungen existieren [23].

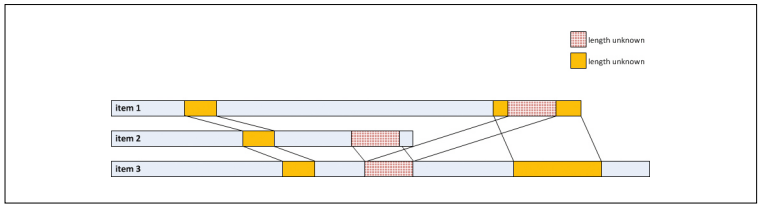
Partielles Audio-Matching kann im Kontext der Content-Verifizierung eingesetzt werden, um bei Ermittlungsarbeiten eine teilweise Wiederverwendung von Material effizient zu erkennen und zu lokalisieren, was manuell sehr zeitaufwändig und ab einer bestimmten Datenmenge quasi unmöglich ist. Außerdem kann eine modifizierte Version des Verfahrens eingesetzt werden, um die missbräuchliche Wiederverwendung von sehr kurzen Abschnitten (bis ca. 150 ms) zwecks Manipulation, sogenannte "copy-move-forgeries", zu erkennen [24].

Die **Phylogenie-Analyse** hingegen widmet sich wiederum dem Problem der „Entwicklungsgeschichte" von Items, d.h. der Frage, in welcher Reihenfolge „Near-Duplicates", also Kopien durch Transformationen wie z. B. Codierung entstanden sind (vgl. Bild 5). Damit lassen sich Fragen beantworten wie: „Welche Audiodatei wurde zuerst produziert bzw. veröffentlicht?" und „Welche Audiodateien sind Vorgänger/Nachfolger von einer anderen Audiodatei"? Bei entsprechenden Verfahren werden Ähnlichkeitsanalysen und logische Annahmen (z. B. bzgl. Qualitätsverlusten durch Codierung) verwendet, um die (wahrscheinliche) Entwicklungshistorie des Materials zu rekonstruieren. Die aktuellsten Verfahren erlauben eine entsprechende Analyse und Erkennung von Transformationen für MP3 und AAC Codierung sowie Fading und Trimming [25], benötigen für eine entsprechende Analyse aber sehr viel Rechenzeit. Im Zuge des Projektvorhabens soll das Verfahren über eine Verwendung neuester Ansätze des maschinellen Lernens optimiert und auf weitere Codecs erweitert werden.



Quelle: Fraunhofer IDMT

Bild 4. Funktionsprinzip des Inversen Decoder



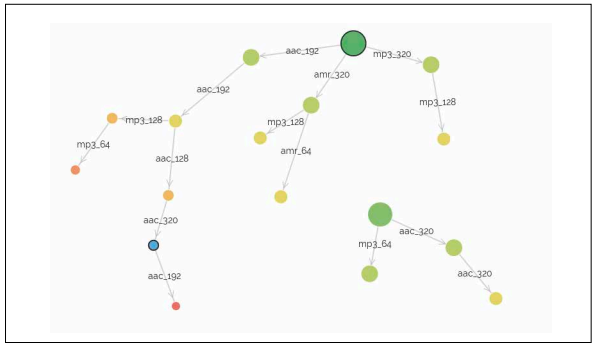
Quelle: Fraunhofer IDMT

Bild 5. Partielles Matching – Detektion von Teilkopien

Content-Verifizierung: Herausforderungen

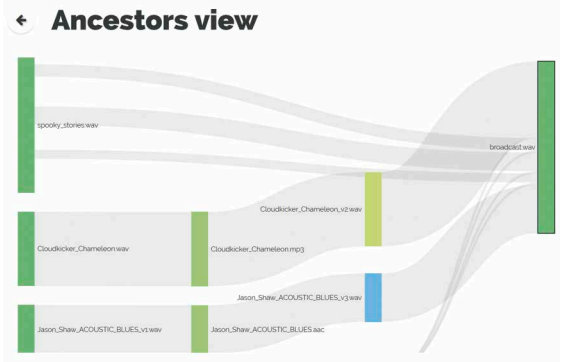
Die oben genannten Werkzeuge bilden eine nützliche Toolbox zur Unterstützung der Content-Verifizierung, aber bei deren Einsatz stellen sich immer auch diverse Herausforderungen:

- Es gibt zahllose „Angriffsvarianten" und auf neue oder verbesserte Detektoren kommen früher oder später wiederum verbesserte Angriffe, die die Entwicklung neuer Detektoren notwendig machen: Forensik ist ein Katz-und-Maus-Spiel.
- Eine „Fälschungskultur" für Mediendaten ist erforderlich: Je mehr Informationen über den Erfassungs- und Bereitstellungskontext bereitgestellt werden, desto mehr Ziele für eine Plausibilitätsprüfung gibt es, was wiederum umso besser für die Verifizierung ist. Derzeit wird dies aber noch viel zu wenig praktiziert.
- Für eine umfassende Content-Verifizierung muss ein breites Spektrum von Manipulationsdetektoren abgedeckt werden. Um dieses Ziel zu erreichen, ist erstens noch viel Forschung und Entwicklung in diesem Bereich erforderlich, zweitens sollten multimodale und interdisziplinäre Ansätze der Standard sein. Eine Inhaltsüberprüfung sollte nicht nur alle Datentypen (Video, Bild, Audio, Text, Metadaten) abdecken, sondern neben Medienforensik und IT-Forensik auch andere Disziplinen wie Linguistik, Psychologie, Datenanalyse usw. einschließen.
- Verifizierungs-Werkzeuge müssen auf Arbeitsabläufe und praktische Anforderungen angepasst und geschickt in die praktischen Arbeitsabläufe integriert werden, um



Quelle: Fraunhofer IDMT

Bild 6. Phylogenie-Analyse: Automatische Erkennung von Eltern-Kind-Beziehungen



Quelle: Fraunhofer IDMT

Bild 7. Visualisierung der „Vorfahren" einer Produktion

insbesondere im journalistischen Bereich Nutzen zu bringen. Nutzer benötigen außerdem entsprechende Schulungen, um zu verstehen, wann welche Tools eingesetzt und wie ihre Ergebnisse zu interpretieren sind. Zum Beispiel müssen Benutzer mit Wahrscheinlichkeiten statt Gewissheiten Anwender können. Solche Herausforderungen werden aktuell im von Google DNI finanzierten Projekt „Digger" angegangen, das die Integration von Audio-Forensik-Tools in eine vorhandene Plattform zur Inhaltsüberprüfung, TrulyMedia, vorsieht.

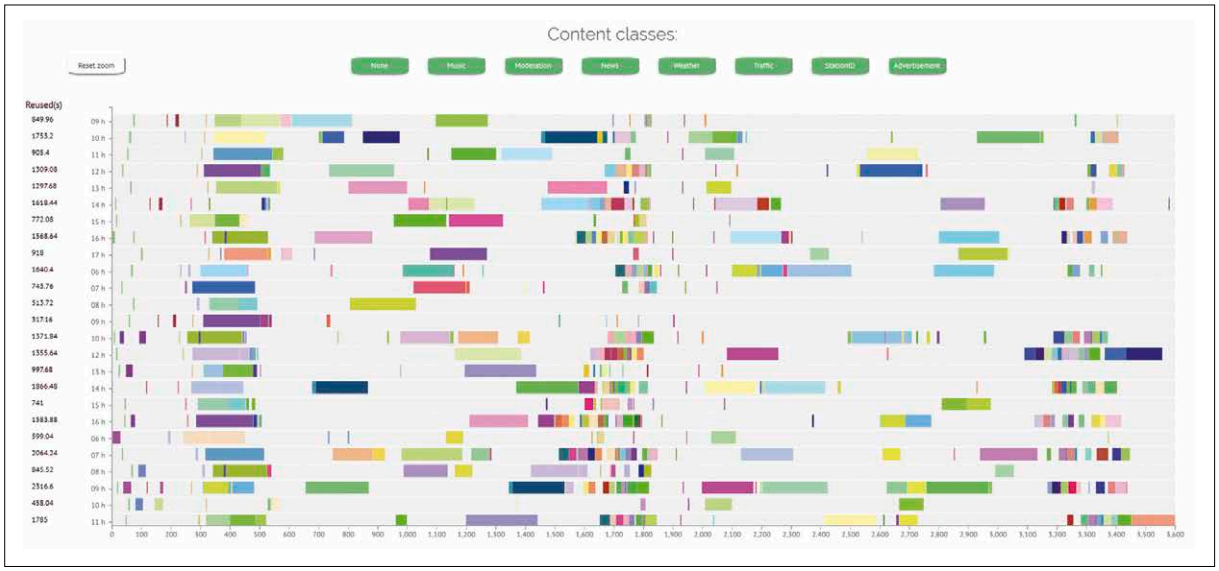
Über diese bestehenden Herausforderungen hinaus gibt es aber noch eine weitere: Eine neue Generation von Sprachsynthesetechnologien auf der Basis von Deep Learning/GANs ermöglicht die Erstellung von sehr realistischer synthetischer Sprache, die insbesondere für den journalistischen Bereich eine große Bedrohung darstellt, aber auch für Audio-Kommunikation im Allgemeinen. Es wird daher wichtig sein, rechtzeitig Detektoren für diese Art von Audio-Fälschungen zu entwickeln.

Weitere Anwendungsgebiete

Die beschriebenen Technologien können noch für mehrere andere Anwendungen genutzt werden:

Wenn Inhalte produziert werden, besonders wenn Legacy-Systeme benutzt werden oder externe Akteure ins Spiel kommen, dann liegen für eine Produktion oft nur unzureichende Informationen darüber vor, welches Material letztlich genau wo in die Produktion eingegangen ist. Dies führt häufig zu fehlerhaften, unvollständigen und inkonsistenten Metadaten und zusätzlichen Kosten. Partielles Audio-Matching kann dann für **automatisches Metadaten- und Rechtetracking** benutzt werden und somit die Produktion gegen das (potenziell) verwendete Rohmaterial vergleichen werden. So wird automatisch erkannt und lokalisiert, welches Material letztlich zum Einsatz kam. Das Ergebnis ist eine detaillierte Auflistung, an welcher Stelle im Endmaterials welches Rohmaterial verwendet wurde. Bild 7 zeigt eine Visualisierung einer entsprechenden Analyse für eine gegebene Produktion „broadcast.wav". Zur **De-Duplizierung** eignet sich das Tool ebenfalls: So lassen sich Teilkopien aufspüren, und diese können anschließend mit der Phylogenie-Analyse untersucht werden, um Originale zu identifizieren, und anschließend alle Teilkopien zur Löschung vormerken – denn in vielen Fällen reicht es aus, das Original aufzubewahren.

Man kann partielles Audio-Matching aber auch als Basis verwenden, um eine **Programmanalyse** zu realisieren: Durch eine Analyse des Programms werden sich wiederholende Segmente automatisch erkannt und lokalisiert. Diese Information kann als Ausgangspunkt dienen, um die Wiederholungsrate in einem Programm/Sender zu ermitteln, aber inhaltliche Überlappungen zwischen Sendern und Programmen bzw. deren „Einzigartigkeit" zu analysieren. Aufgrund der zeitlichen Anordnung und Länge kann man aber auch gut auf den Inhalt schließen und Vergleiche mit anderen Programmen uns Sendern vornehmen. Bild 8 zeigt ein Beispiel für eine einfache Visualisierung eines analysierten



Quelle: Fraunhofer IDMT

Bild 8. Visualisierung von Wiederholungen im Radioprogramm

Bild: Fraunhofer IDMT

PATRICK AICHROTH

ist seit 2003 wissenschaftlicher Mitarbeiter und seit 2006 Leiter der Gruppe „Mediendistribution und Sicherheit" am Fraunhofer IDMT.

www.idmt.fraunhofer.de

Bild: Fraunhofer IDMT

HANNA LUKASHEVICH

ist seit 2006 wissenschaftliche Mitarbeiterin und seit 2014 Leiterin der Gruppe „Semantische Musiktechnologien" am Fraunhofer IDMT.

www.idmt.fraunhofer.de

Radioprogramms (Segmente mit demselben Inhalt haben dieselbe Farbe).

Kombiniert man diese Analyse noch mit automatischer Musik-Analyse und Sprach-Musik-Unterscheidung sowie Spracherkennung und Textanalyse, so lassen sich umfangreiche statistische Informationen über Musik- und Nachrichtenprogramme daraus ableiten.

Ein weiteres Anwendungsgebiet ist die **Qualitätskontrolle**: Für die Aggregation und den Ingest kann es sehr wichtig sein sicherzustellen, dass Inhalte nicht mit undokumentierten vorherigen Codierungsschritten angeliefert werden, was zu unbeabsichtigten Qualitätsproblemen führt, z. B. durch Transkodierung von MP3 zu AAC. Dies kann durch den oben beschriebenen inversen Decoder realisiert werden.

Schließlich lassen sich die genannten Technologien auch zur **Synchronisierung** verwenden: Wenn mehrere A/V-Aufzeichnungen desselben Ereignisses (z. B. aus verschiedenen Kameraperspektiven aufgenommen) gefunden und synchronisiert werden müssen, kann partielles Matching dafür eingesetzt werden.

Zu guter Letzt können die o. g. Verfahren zur Audio-Manipulationsdetektion auch für die **automatische Schnitterkennung** verwendet werden, um undokumentierte Edits während des Produktionsprozesses im Audiomaterial nachträglich zu annotieren.

Referenzen

[1] S. Gupta, S. Cho und C.-C. J. Kuo, „Current developments and future trends in audio authentication," Multimedia in Forensics, Security and Intelligence, 2012, pp. 50–59.
[2] M. Kajstura, A. Trawinska und J. Hebenstreit, „Application of the electrical network frequency (ENF) criterion: A case of a digital recording," Forensic Science International, Bd. 155, Nr. 2–3, pp. 165–717, 2005.
[3] C. Grigoras, „Applications of ENF analysis in forensic authentication of digital audio and video recordings," Journal of the Audio Engineering Society, Bd. 57, Nr. 9, pp. 643–661, 2009.
[4] R. Sanders, „Digital audio authenticity using the electric network frequency," in Proc. of Audio Engineering Society Conference, 2008.
[5] C. Grigoras, „Digital audio recording analysis: The electric network frequency (ENF) criterion," International Journal of Speech Language and the Law, Bd. 12, Nr. 1, pp. 1350–1771, 2005.
[6] A. Cooper, „The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings an automated approach," in Proc. of Audio Engineering Society Conference on Audio Forensics – Theory and Practice, 2008.
[7] M. Huijbregtse und Z. Geradts, „Using the ENF criterion for determining the time of recording of short digital audio recordings," in Proc. of 3rd International Workshop Computational Forensics, 2009.
[8] S. Mann, L. Cuccovillo, P. Aichroth und C. Dittmar, „Combining ENF Phase Discontinuity Checking and Temporal Pattern Matching for Audio Tampering Detection," in Proc. of Workshop on Audiosignal and Speech processing (WASP), Koblenz, Germany, 2013.
[9] D. Nicolalde und J. Apolinario, „Evaluating digital audio authenticity with spectral distances and ENF phase change," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), 2009.
[10] D. Rodriguez, J. Apolinario und L. Biscainho, „Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," IEEE Trans. Information Forensics and Security, Bd. 5, Nr. 3, pp. 534–543, 2010.
[11] L. Cuccovillo, S. Mann, P. Aichroth, M. Tagliasacchi und C. Dittmar, „Blind Microphone Analysis and Stable Tone Phase Analysis for Audio Tampering Detection," in Proc. of 135th AES Convention, New York, USA, 2013.
[12] L. Cuccovillo und P. Aichroth, „Increasing the Temporal Resolution of ENF Analysis via Harmonic Distortion," in Proc. of AES International Conference of Audio Forensics, 2017.
[13] C. Kraetzer, M. Qian, M. Schott und J. Dittmann, „Context model for microphone forensics and its application in evaluations," in Proc. of SPIE Conference on Media Watermarking, Security, and Forensics, 2011.
[14] C. Kraetzer, A. Oermann, J. Dittmann und A. Lang, „Digital audio forensics: A first practical evaluation on microphone and environment classification," in Proc. 9th of Workshop Multimedia and Security, 2007.
[15] L. Cuccovillo, S. Mann, M. Tagliasacchi und P. Aichroth, „Audio Tampering Detection via Microphone Classification," in Proc. of IEEE Multimedia Signal Processing Workshop (MMSP), Pula, Italy, 2013.
[16] L. Cuccovillo und P. Aichroth, „Open-set microphone classification via blind channel analysis," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016.
[17] S. Möhrs, J. Herre und R. Geiger, „Analyzing decompressed audio with the „Inverse Decoder" – towards an operative algorithm," in Proc. of 112th convention Audio Engineering Society, Munich, Germany, 2002.
[18] J. Herre und M. Schug, „Analysis of Decompressed Audio - The ‚Inverse Decoder," in Proc. of 109th AES Convention, New York, USA, 2000.
[19] P. Bießmann, D. Gärtner, C. Dittmar, P. Aichroth, M. Schnabel, G. Schuller und R. Geiger, „Estimating MP3PRO Encoder Parameters From Decoded Audio," in Proc. of Workshop on Audiosignal and Speech processing (WASP), Koblenz, Germany, 2013.
[20] L. Cuccovillo und P. Aichroth, „Inverse decoding of PCM A-law and µ-law," in Proc. of AES International Conference of Audio Forensics, 2019.
[21] D. Seichter, L. Cuccovillo und P. Aichroth, „AAC encoding detection and bitrate estimation using a convolutional neural network," in Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016.
[22] A. Wang, „An Industrial Strength Audio Search Algorithm," in Proc. of Internation Conference on Music Information Retrieval (ISMIR), 2003.
[23] M. Maksimovic, P. Aichroth und L. Cuccovillo, „Detection and Localization of Partial Audio Matches," in Proc. of IEEE International Conference on Content-Based Multimedia Indexing (CBMI), 2018.
[24] M. Maksimovic, L. Cuccovillo und P. Aichroth, „Copy-Move Forgery Detection and Localization via Partial Audio Matching," in Proc. of AES International Conference of Audio Forensics, 2019.
[25] M. Maksimovic, L. Cuccovillo und P. Aichroth, „Phylogeny analysis for MP3 and AAC coding transformations," in Proc. of IEEE International Conference on Multimedia and Expo (ICME), 2017.