



**BUNDESRECHTSANWALTSKAMMER**

## **Stellungnahme Nr. 63 November 2023**

**Entwurf eines Gesetzes zur digitalen Dokumentation der strafgerichtlichen Hauptverhandlung (Hauptverhandlungsdokumentationsgesetz – DokHVG)  
(in der Fassung als Regierungsentwurf vom 10. Mai 2023)**

**Hier: technische Aspekte zur Aufzeichnung und Transkription von Audioaufzeichnungen**

### **Mitglieder des Ausschusses Strafprozessrecht:**

Rechtsanwalt Dr. Matthias Dann  
Rechtsanwalt Prof. Dr. Michael Gubitza  
Rechtsanwältin Dr. Vera Hofmann  
Rechtsanwalt Prof. Dr. Christoph Knauer, Vorsitzender  
Rechtsanwalt Dr. jur. Andreas Minkoff  
Rechtsanwalt Maximilian Müller  
Rechtsanwalt Jürgen Pauly  
Rechtsanwältin Anette Scharfenberg  
Rechtsanwältin Dr. Alexandra Schmitz  
Rechtsanwältin Stefanie Schott  
Rechtsanwalt Prof. Dr. Gerson Trüg

Prof. Dr. René Krenz-Baath (Berichtersteller)

Rechtsanwältin Eva Melina Buchmann, Bundesrechtsanwaltskammer

#### **Bundesrechtsanwaltskammer**

The German Federal Bar  
Barreau Fédéral Allemand  
[www.brak.de](http://www.brak.de)

#### **Büro Berlin – Hans Litten Haus**

Littenstraße 9    Tel. +49.30.28 49 39 - 0  
10179 Berlin    Fax +49.30.28 49 39 -11  
Deutschland    Mail [zentrale@brak.de](mailto:zentrale@brak.de)

#### **Büro Brüssel**

Avenue des Nerviens 85/9    Tel. +32.2.743 86 46  
1040 Brüssel    Fax +32.2.743 86 56  
Belgien    Mail [brak.bxl@brak.eu](mailto:brak.bxl@brak.eu)

**Verteiler:** Bundesministerium der Justiz und für Verbraucherschutz  
Bundesministeriums des Innern, für Bau und Heimat  
Ausschuss für Recht- und Verbraucherschutz des Deutschen Bundestages  
Ausschuss für Inneres und Heimat des Deutschen Bundestag  
Fraktionsvorsitzende der CDU/CSU, SPD, BÜNDNIS 90/DIE GRÜNEN, DIE LINKE  
Rechtspolitischen Sprecher der Fraktionen CDU/CSU, SPD, BÜNDNIS 90/DIE GRÜNEN,  
DIE LINKE  
Arbeitskreise Recht der Bundestagsfraktionen  
Landesjustizminister/Justizsenatoren der Länder  
Innenministerien und Senatsverwaltungen für Inneres der Länder  
Bundesgerichtshof  
Rechtsanwaltskammern  
Bundesverband der Freien Berufe  
Bundesnotarkammer  
Bundessteuerberaterkammer  
Patentanwaltskammer  
Deutscher Anwaltverein  
Deutscher Juristinnenbund  
Deutscher Notarverein  
Deutscher Richterbund  
Neue Richtervereinigung e.V.  
Deutscher Steuerberaterverband e. V.  
Wirtschaftsprüferkammer  
Institut der Wirtschaftsprüfer in Deutschland e.V.  
Bund Deutscher Kriminalbeamter  
Verbraucherzentrale Bundesverband e.V.  
Deutscher Juristentag e.V.  
Redaktionen der NJW, NStZ, NZWiSt, Beck Verlag, ZAP, AnwBI, DRiZ, FamRZ, FAZ,  
Süddeutsche Zeitung, Die Welt, taz, Handelsblatt, dpa, Spiegel, Focus, Deubner Verlag  
Online Recht, LTO, Beck aktuell, Jurion, Juris Nachrichten, Juve, LexisNexis Rechtsnews,  
Otto Schmidt Verlag, Kriminalpolitische Zeitschrift, Strafverteidiger Forum, Zeitschrift  
HRR Strafrecht

Die Bundesrechtsanwaltskammer ist die Dachorganisation der anwaltlichen Selbstverwaltung. Sie vertritt die Interessen der 28 Rechtsanwaltskammern und damit der gesamten Anwaltschaft der Bundesrepublik Deutschland mit rund 166.000 Rechtsanwältinnen und Rechtsanwälten<sup>1</sup> gegenüber Behörden, Gerichten und Organisationen – auf nationaler, europäischer und internationaler Ebene.

## Stellungnahme

### Vorbemerkung

Die technische Umsetzbarkeit und Verlässlichkeit von Transkriptionssystemen wurde zum Teil sehr kritisch bewertet. Ziel dieser Stellungnahme ist es daher, die technischen Aspekte zur Aufzeichnung und Transkription von Audioaufzeichnungen in Strafgerichtsprozessen zu erläutern und aktuelle sowie zukünftige Entwicklungen zu beleuchten. Final werden konkrete Empfehlungen formuliert, um eine erfolgreiche Durchführung der im Gesetzentwurf vorgesehenen Pilotphase zu ermöglichen/abzusichern. Diese Stellungnahme knüpft an die [Stellungnahme Nr. 23/2023](#) zum Regierungsentwurf an.

### Transkription von Audioaufzeichnungen in Strafgerichtsprozessen

Die Entwicklung automatisierter Sprache-zu-Text Transkription hat in den vergangenen Jahren signifikante Fortschritte erfahren. Hierbei ist einerseits die Zuverlässigkeit der Spracherkennung, die Verringerung der notwendigen Rechenressourcen sowie die Verfügbarkeit auf verschiedensten Software- und Hardware-Plattformen verbessert worden. Neben den etablierten Software-Lösungen hat sich der Umfang und die Qualität frei zugänglicher Sprache-zu-Text Transkription rasant entwickelt. In verschiedenen Benchmarking-Verfahren wurden lediglich minimale Qualitätsunterschiede dokumentiert.

Insbesondere die Etablierung von generischen (KI-basierten) Sprache-zu-Text Transkriptionsmethoden bietet eine Vielzahl von Vorteilen, welche nunmehr beleuchtet werden.

Die Qualität und Robustheit einer automatisierten Transkription leitet sich vornehmlich aus der Fähigkeit ab verschiedenste Gesprächssituationen (Einzelvortrag, Diskussion, spontane Interviews), Sprachvarianten (Dialekt, Lautstärke, Sprachgeschwindigkeit usw.) und ggf. Fachtermini zu erkennen und entsprechend erfolgreich zu transkribieren.

Herkömmliche Transkriptionstechniken verfügten i.A. über spezialisierte Implementationen und Techniken, um in sehr spezifischen Situationen eine hohe Transkriptionsqualität zu erreichen. Nachteilig war hierbei, dass sich diese Produkte beispielsweise nur sehr eingeschränkt auf variierende Gesprächssituationen oder Fachtermini einstellen können. Der Vorteil von modernen KI-basierten Sprache-zu-Text-Modellen ist, dass sich diese durch automatisierte wiederkehrende Lernphasen individuellen und neuen Situationen anpassen können, ohne die darunterliegende Hardware und Software ändern zu müssen. Ein weiterer Vorteil ist, dass die verbesserten Sprache-zu-Text-Modelle in Cloud-basierten Verfahren sofort und in lokal betriebenen Lösungen durch ein einfaches Software-Update zur Verfügung stehen.

Auch ist denkbar, dass auf Basis identischer Hardware- und Software-Komponenten abgewandelte Versionen der Sprache-zu-Text-Modelle in verschiedenen Teilen der Bundesrepublik (Oberbayern,

---

<sup>1</sup> Im Interesse einer besseren Lesbarkeit wird nicht ausdrücklich in geschlechtsspezifischen Personenbezeichnungen differenziert. Die im Folgenden gewählte männliche Form schließt alle Geschlechter gleichberechtigt ein.

Nordfriesland, Sachsen) zum Einsatz kommen, um den unterschiedlichen sprachlichen Gegebenheiten gerecht zu werden.

### **Technische Anforderungen in den Gerichten**

Einer der wiederholt formulierten Kritikpunkte ist der Aufwand zur Installation und zum Betrieb der Transkriptionstechnik. Hierbei ist zu beachten, dass der digitale Wandel auf verschiedenen Ebenen in den Gerichten Einzug gehalten hat bzw. in naher Zukunft halten wird. Ein wesentlicher Faktor hierbei ist die verpflichtende Einführung der E-Akte und die flächendeckende Internetanbindung der Gerichte. Durch die Einführung der E-Akte ist von einer vorhandenen IT-Infrastruktur bestehend aus mindestens einem PC und einer kabelgebundenen Datenverbindung auszugehen. Eine Verwendung von drahtlosen Übertragungstechniken ist nicht zu empfehlen, da der fehlerfreie Betrieb mit einfachen Mitteln, z.B. Störsender, unterbunden werden könnte. Somit ist ein Großteil der Hardware-Voraussetzungen zur Transkription vorhanden. Zusätzlich sind eine von den jeweiligen Gegebenheiten abhängige Anzahl von Mikrofonen notwendig inklusive der Anschlusskomponente an den jeweiligen PC.

Durch die flächendeckende Internetanbindung der Gerichte ist die Anwendung von Cloud-basierten Transkriptionssystemen möglich. Hierbei steht aus technischer Sicht vor allem die Zuverlässigkeit dieser Anbindung im Vordergrund. Cloud-basierte Lösungen setzen eine stabile Internetverbindung während der gesamten Aufzeichnungszeit voraus. Hierbei ist zu bemerken, dass die notwendige Datenübertragungsrate als minimal zu betrachten ist. Ein zu transkribierender Audio-Stream benötigt nach unseren Experimenten bei hoher Qualität 100kBit/s<sup>2</sup>. Dies entspricht ca. 0,3% der Übertragungskapazität (Upload) einer durchschnittlichen Internetanbindung in Deutschland im September 2023 (31Mbit/s)<sup>3</sup>.

Weiterhin stehen freie Sprache-zu-Text Transkriptionswerkzeuge zur Verfügung, welche keine Internetanbindung voraussetzen und lokal beispielsweise auf dem PC im Gerichtssaal betrieben werden können. Hier ist das „Common Voice“-Projekt der Mozilla Foundation zu nennen<sup>4</sup>. Dieses Projekt entwickelt Sprachmodelle in verschiedenen Sprachen, welche übertragen auf lokale Systeme ohne Internetanbindung kostenlos zur real-time Sprache-zu-Text Transkription eingesetzt werden können. Das Sprachmodell wird ständig weiterentwickelt und in regelmäßigen Intervallen zum Download angeboten. Aktuell (14.9.2023) beinhaltet das deutsche Modell 1301 Stunden Sprachmaterial, welches von 18.352 SprecherInnen zur Verfügung gestellt wurde.

### **Nutzbarkeit / Usability von Transkriptionslösungen**

Wiederholt sind Kosten für zusätzliches Personal bzw. Personalaufwand zur Bedienung der Transkriptionstechnik als ein signifikanter Hinderungsgrund für die umfängliche Einführung der Transkription diskutiert worden. Diesem Aspekt ist mit einem besonderen Fokus auf die Usability der zu beschaffenden Software-Lösung zu begegnen. Hierbei sollte die Bedienung und das Monitoring während der Verhandlung auf einfache Art und Weise, sowie ohne erweiterte IT-Kenntnisse möglich sein. Erfahrungen an schwedischen Gerichten ergaben, dass selbst Richter fortgeschrittenen Alters nach angemessener Einführung keine Probleme mit der Bedienung und dem Monitoring der audiovisuellen Aufnahmen hatten.

---

<sup>2</sup> Court 2025 – Demonstrator, S. 391-395, Bericht der Expertinnen- und Expertengruppe zur Dokumentation der strafgerichtlichen Hauptverhandlung – Anlagenband -

<sup>3</sup> Statista: <https://de.statista.com/statistik/daten/studie/416534/umfrage/durchschnittliche-internetgeschwindigkeit-in-deutschland>

<sup>4</sup> <https://commonvoice.mozilla.org/de>

## **Empfehlungen zur erfolgreichen Umsetzung**

Zusätzlich zu den o.g. Ausführungen sind folgende Empfehlungen hilfreich zur erfolgreichen Umsetzung des Vorhabens.

1. Eine sofortige (real-time) Transkription während der Gerichtsverhandlung ist vor dem Hintergrund verschiedener Aspekte zu empfehlen:
  - a. Eine Zuordnung der jeweiligen Ausführungen zu einer spezifischen Person kann bei einer nachträglichen Transkription unmöglich sein.
  - b. Eine Prüfung der Transkription und somit folglich des Transkriptionssystems kann unmittelbar während der Verhandlung erfolgen und von den Prozessbeteiligten (Richter, Anwälte, Staatsanwaltschaft) verfolgt werden.
  - c. Ein Ausfall bzw. eine signifikante Einschränkung der Aufzeichnung wird unmittelbar erkannt und folglich kann auf herkömmliche Protokollierungsmethoden zurückgegriffen werden.
2. Die Robustheit eines Transkriptionssystems kann durch die Anwendung redundanter Transkriptionsmethoden erhöht werden. D.h. im Falle eines Ausfalles der Internetanbindung oder eines Cloud-Dienstes, könnte eine lokale (wenn auch weniger leistungsfähige) Lösung die Transkription weiterführen bis der reguläre Cloud-basierte Service wieder zur Verfügung steht.
3. Die Qualität des resultierenden Transkripts hängt signifikant von der Audioqualität der Aufzeichnung ab. Im derzeitigen Gesetzentwurf sind jedoch keinerlei Vorgaben zur technischen Umsetzung enthalten. Entsprechend ist zu empfehlen hier entsprechende Vorgaben hinzuzufügen. Hierfür existieren entsprechende DIN- bzw. ISO-Normen.

## **Fazit**

Der erfolgreichen Anwendung von automatisierter Sprache-zu-Text Transkription von Audioaufzeichnungen in Strafgerichtsprozessen ist aus aktueller technischer Perspektive möglich.

Im Hinblick auf die derzeitigen und zukünftigen rasanten Entwicklungen in allen Bereichen der Informationstechnologie und insbesondere der Künstlichen Intelligenz, welche hier besonders zur Anwendung kommt, ist eine weitere kontinuierliche und signifikante Verbesserung von Transkriptionslösungen in den kommenden Monaten klar erwartbar.

- - -