

Flächendeckende Realisierbarkeit von barrierefreien Videos mithilfe automatisierter Untertitel

Sanja Grimminger, Dr. Sarah Voß-Nakkour, Linda Rustemeier,
Saba Mateen (Goethe-Universität Frankfurt am Main)

DOI: <https://doi.org/10.21248/gups.69153>



aus dem Sammelband

Digitale Barrierefreiheit in der Bildung weiter denken
Innovative Impulse aus Praxis, Technik und Didaktik

Herausgeber*innen

Dr. Sarah Voß-Nakkour, Linda Rustemeier, Prof. Dr. Monika M. Möhring,
Andreas Deitmer, Sanja Grimminger

Verlag

Universitätsbibliothek Johann Christian Senckenberg

1. Auflage 2023

DOI: <https://doi.org/10.21248/gups.62773>

ISBN 978-3-88131-102-1



Dieses Werk wurde unter der Lizenz „Creative Commons Namensnennung“
in Version 4.0 (abgekürzt „CC BY 4.0“) veröffentlicht.

Flächendeckende Realisierbarkeit von barrierefreien Videos mithilfe automatisierter Untertitel

Sanja Grimminger, Dr. Sarah Voß-Nakkour, Linda Rustemeier, Saba Mateen
(Goethe-Universität Frankfurt am Main)

Abstract:

Untertitel manuell zu erstellen, erfordert zum jetzigen Zeitpunkt viel Zeit und personelle Ressourcen. Für Einzelne ist es mühsam ein geeignetes Tool zu finden und einen guten Workflow zu erarbeiten, geschweige denn die schnell steigende Anzahl an Videos zu bearbeiten. Bedingt durch die Corona-Pandemie und das damit einhergehende Arbeiten im digitalen Lehrbetrieb ist die Anzahl der Videos noch einmal immens gestiegen. Zu Untertiteln sind nicht nur aufgenommene Vorlesungen, sondern beispielsweise auch Informationsvideos auf einer Homepage, Erklärfilme, aber auch digitale Live-Veranstaltungen. Dies sprengt die begrenzten Ressourcen, die den Hochschulen zur Verfügung stehen. Im Beitrag wird untersucht, inwieweit automatische Untertitel die Lösung dieses Problems sein können. Dafür werden vier Tools auf ihre funktionellen Unterschiede und die Qualität ihrer Ergebnisse untersucht. Anschließend wird erarbeitet, welche dieser Kriterien für Hochschulen relevant sind, um eine flächendeckende Untertitelung in allen Videos gewährleisten zu können. Die Erkenntnisse basieren auf den praktischen Erfahrungen von **studiumdigitale**, der zentralen eLearning-Einrichtung der Goethe-Universität Frankfurt am Main, mit dem Einsatz verschiedener Tools in Projekten sowie aus dem Erfahrungsaustausch mit anderen Hochschulmitarbeiter*innen im Rahmen des Verbundprojekts HessenHub. Das Ziel dabei ist, einen schlanken Workflow im Sinne einer Lean Media Production, d.h. eine innovative Herangehensweise, die möglichst ressourcenschonend und perspektivisch nachhaltig ist, zu erarbeiten.

Schlüsselbegriffe: Barrierefreie Videos, automatische Untertitelung



1. Untertitel als wesentlicher Faktor von barrierefreien digitalen Medien

Bei der Suche nach Antworten auf die Frage, wie digitale Barrierefreiheit an Hochschulen bestmöglich umgesetzt werden kann, spielen Videos eine wesentliche Rolle. Videos umfassen nicht nur Vorlesungsaufzeichnungen und Lehr-/Lernvideos wie Erklärfilme, sondern beispielsweise auch Informationsvideos auf einer Homepage oder digitale Live-Veranstaltungen. Bedingt durch die Pandemie und das damit einhergehende Arbeiten im digitalen Lehrbetrieb ist die Anzahl der Live- und On-demand-Abrufe von Videos sowie die Nutzungszeit immens gestiegen. Dies belegen interne Zahlen des Videoportals „MediaSite“ der Goethe-Universität Frankfurt am Main: Waren es im Jahr 2019 insgesamt noch 571.969 Aufrufe, stieg die Anzahl 2020 auf über 3.157.655 und im Jahr 2021 auf über 8.906.374 Zugriffe.

Damit Videos barrierefrei und damit allen zugänglich sind, gilt es, bestimmte Voraussetzungen zu erfüllen. Die Barrierefreie-Informationstechnik-Verordnung (BITV 2.0, 2019) schreibt im §3 vor: „Die in § 2 genannten Angebote, Anwendungen und Dienste der Informationstechnik sind barrierefrei zu gestalten. Dies erfordert, dass sie wahrnehmbar, bedienbar, verständlich und robust sind.“ In dem Projekt „BIK für Alle“ (BIK, 2018) werden drei Merkmale genannt, die für barrierefreie Online-Videos gelten:

- „[...] sie sind barrierefrei erreichbar- und bedienbar (barrierefreie Einbindung),
- für höreingeschränkte Menschen wird eine Untertitelung zur Verfügung gestellt und
- wichtige visuelle Informationen werden für blinde Menschen über eine Audiodeskription vermittelt.“

Voraussetzung dafür ist ein barrierefreier Videoplayer, wie der [AblePlayer](#). Dieser stellt sicher, dass ein Video auch für Personen zugänglich und nutzbar ist, die die Computermaus nicht bedienen können und das Video über die Tastatur steuern bzw. für Personen, denen die Bedienelemente ansonsten zu klein wären. Des Weiteren müssen gleichwertige Informationen über verschiedene Sinne wahrnehmbar gemacht werden (Zwei-Sinne-Prinzip). Nach der BITV 2.0-Bedingung 1.2.1 ist das Sichtbare über Audiodeskription wiederzugeben und das Hörbare mit Untertiteln zu verschriftlichen (Liquid Impressions KG). Untertitel und Audiodeskriptionen müssen



über den Videoplayer ein- und ausgeblendet werden können.

In diesem Beitrag stehen Untertitel im Fokus und damit die Frage, wie sich Tools zur Untertitelung hinsichtlich ihrer Funktionen sowie der Qualität ihrer Ergebnisse unterscheiden. Anschließend wird erarbeitet, was aus den gewonnenen Erkenntnissen auf die Umsetzung an Hochschulen zu berücksichtigen ist bzw. welche Punkte relevant sind, um eine flächendeckende Untertitelung in allen Videos gewährleisten zu können. Es soll im Sinne einer Lean Media Production eine innovative und schlanke Herangehensweise erarbeitet werden, die möglichst ressourcenschonend und perspektivisch nachhaltig ist (Voß-Nakkour, 2013).

1.1 Wer benötigt Untertitel und wie werden sie erstellt?

Untertitel fördern die Verständlichkeit von Videos im Allgemeinen. Sie sind vor allem für Personen mit einer Hörbeeinträchtigung oder gehörlosen Menschen von enormer Bedeutung, unterstützen aber auch Menschen, wie z.B. Nicht-Muttersprachler*innen oder Personen, die sich in einer lauten Geräuschkulisse befinden. Ziel von Untertiteln ist es, allen denselben Kenntnisstand zu vermitteln (NDR, 2015). Für die Vorgehensweise, wie Untertitel zu erstellen sind, gibt es Standards von den öffentlich-rechtlichen Fernsehsendern in Deutschland, Österreich und der Schweiz (ebd.) oder auch ISO-Standards (ISO, 2021). An dieser Stelle wird lediglich ein kurzer Überblick über die Arbeitsschritte gegeben: Bei der manuellen Erstellung von Untertiteln ist zunächst eine Transkription nötig, wobei der Text in kleine Häppchen aufgeteilt wird. Manche Tools bieten eine automatische Erstellung von Untertiteln an. Diese werden mittels Spracherkennung (Speech-to-Text) erstellt: „Speech recognition or speech to text includes capturing and digitizing the sound waves, transformation of basic linguistic units or phonemes, constructing words from phonemes and contextually analyzing the words to ensure the correct spelling of words that sounds the same“ (Al Smadi, Al Issa, Trad & Al Smadi, 2015). Neben dem gesprochenen Wort werden je nach Kontext auch Hintergrundgeräusche transkribiert. Auf die Transkription folgt die Anpassung der Untertitel an die Timeline des Videos. Dabei ist darauf zu achten, dass der Text synchron zu Bild und Ton eingeblendet wird. Anschließend steht die manuelle Überprüfung bzw. Korrektur der Untertitel an. Am Ende kann eine Untertiteldatei mit dem gewünschten Format exportiert werden. Je nachdem, wie das Video verwendet werden soll und welcher Videoplayer zur Darstellung genutzt wird, entscheidet man sich beispielsweise für



SRT (Subrip), WebVTT (Web Video Text Track) oder STL (EBU Subtitle Data Exchange format). Ausführlichere Anleitungen zur Untertitelerstellung findet man z.B. bei der TU Dortmund: „[Leitfaden zur Erstellung von Untertiteln](#)“.

1.2 Probleme bei der Erstellung von Untertiteln

Die manuelle Untertitelung erfordert einen hohen personellen Einsatz und ist somit sehr zeit- und auch kostenintensiv. Im Gegensatz dazu schont die automatische Erstellung zwar diese personellen Ressourcen, ist dafür jedoch kostenintensiv in der Anschaffung eines Dienstes. Im Verbundprojekt „HessenHub“ wird das Thema „barrierefreie Videos“ immer wieder diskutiert. Die Erfahrungen der zentralen Einrichtungen und Lehrenden ist, dass es für einzelne Personen sehr mühsam und zeitaufwendig ist manuell Untertitel zu erstellen, einen effizienten Workflow zu erarbeiten, geschweige denn die schnell steigende Anzahl an Videos zu bearbeiten. Auch das Finden sowie die kostenpflichtige Anschaffung eines geeigneten Tools, welches automatisierte Untertitel erstellen kann, werden als Herausforderung genannt. Der Arbeitskreis „[Barrierefreie Videos in der Hochschullehre](#)“ hebt außerdem den Aspekt hervor, dass Videos zeitnah bereitgestellt werden sollten, da ansonsten die Akzeptanz der Nutzer*innen verringert werden könnte – z.B. Studierende, die mit Videos wie Vorlesungsaufzeichnungen lernen (Puhl & Lerche, 2019: 88). Diese Faktoren sprengen den Rahmen der begrenzten Ressourcen, die den Hochschulen zur Verfügung stehen. Auch die Erfahrungen von **studiumdigitale** mit automatischer Untertitelung durch das Tool „Amberscript“ hat gezeigt, dass nicht nur das Transkribieren, sondern auch die Nachbearbeitung, d.h. die synchrone Anpassung der Timeline von Untertitel an Bild und Ton, viel Zeit erfordert, um bei den Betrachtenden ein optimales Verständnis zu fördern. Dazu kommt die manuelle Verbesserung von undeutlich gesprochenem Text sowie Verständnisfragen bezüglich der Richtigkeit von fachspezifischen Begriffen. Ein Beispiel: Trotz automatischer Untertitelung von 15 Vorlesungen (im Schnitt ca. 60 Min. lang) wurden durchschnittlich ca. 4h 20 Min. in einem ersten Durchgang und 2h im zweiten für die Optimierung der Untertitel benötigt. Die Nachbearbeitung ist jedoch nötig, wie Puhl und Lerche (2019: 99 f.) hervorheben: „Den Hochschulen muss bewusst sein, dass die Richtigkeit der Untertitelung von hoher Bedeutung ist, da sie für die Prüfungsvorbereitung genutzt werden und Inhalte demnach Prüfungsrelevanz haben könnten. Die manuelle Nachbearbeitung der Untertitel ist folglich geboten und sollte von Personen durchgeführt werden, die Kompetenzen in der fachkultur-spezifischen Sprache haben.“



2. Kategorien von Tools

Es gibt eine Reihe von Tools, die das Erstellen von Untertiteln ermöglichen.

Sie unterscheiden sich in folgenden Punkten:

1. Kosten: Ist das Tool kostenlos oder kostenpflichtig?
2. Art der Anwendung: Handelt es sich um ein browserbasiertes Tool oder eine Desktopanwendung?
3. Datenspeicherung: Wo werden die Videos und Textdateien gespeichert bzw. wird die Datenschutz-Grundverordnung ([DSGVO](#)) eingehalten?
4. Manuelle oder automatische Erstellung: Werden die akustischen Informationen von einer Person transkribiert oder automatisch erstellt und ist ggf. künstliche Intelligenz (KI) involviert? Um den Prozess der automatischen Untertitelung zu optimieren, simulieren dabei Maschinen menschliche Intelligenz, sind lernfähig und optimieren sich selbst (Amberscript, 2021).
5. Art der Untertitelbearbeitung: Werden Position von Start- (Einblendung des Untertitels) und Endpunkt (Ausblendung des Untertitels), um die Untertitel synchron an die Sprache im Video anzupassen, per Time Code numerisch verändert (z.B. 00:02:15 - 00:02:24) oder in einer Timeline per Drag-and-drop?

2.1 Erste Datenerhebung zum Vergleich von automatischen Untertiteln

Bevor im weiteren Verlauf im Rahmen des Netzwerkes „digitale Barrierefreiheit in hessischen Hochschulen“ ([DigiBar](#)) des Projekts „HessenHub“ eine größer angelegte Studie zum Thema „automatische Erstellung von Untertiteln“ folgt, werden in diesem Beitrag die Ergebnisse einer Vorstudie vorgestellt. Mit diesem Vorgehen werden vorab mögliche Kriterien auf ihre Aussagekraft überprüft, um bei der Hauptstudie einen möglichst hohen Informationswert zu bekommen. Die Erhebung zielt zum einen darauf ab, die sich unterscheidenden Eigenschaften der Tools zu identifizieren und deren Ausmaß zu verdeutlichen. Zum anderen werden Kriterien in Bezug auf ihre



Relevanz in der Hochschullehre eingeordnet und anschließend herausgearbeitet, welche Tools das Potenzial haben, zur Etablierung der standardisierten Untertitelung barrierefreier Videos beizutragen.

In der Erhebung werden die Ergebnisse von automatisch erstellten Untertiteln von vier Tools ausgewertet. Die Wahl fiel auf diese Tools, da **studiumdigitale** und ihre Projektpartnerin, die Technische Hochschule Mittelhessen (THM), bereits Erfahrungen mit ihnen gesammelt haben.

[Premiere Pro](#) von Adobe ist ein kostenpflichtiges Videobearbeitungsprogramm, das einzeln oder in der Creative Cloud mit anderen Anwendungen abonniert werden kann. Man kann es als Monats- oder Jahresabo erwerben. Es handelt sich um eine Desktopanwendung. Mit der Version Premiere Pro 15.4 gibt es die Möglichkeit, automatisch Untertitel zu erstellen, die der DSGVO entsprechen. Sie können sowohl numerisch als auch per Drag-and-drop bearbeitet werden.

[Amberscript](#) ist ebenfalls kostenpflichtig. Die Preise sind hier nach Stunden gestaffelt, was bedeutet, dass die Laufzeit der eigenen Videos die Abokosten bestimmt. Es gibt die Möglichkeit, zusätzliche Pakete zu abonnieren, wie beispielsweise professionelle Transkripte und Untertitel oder Übersetzungen. Es ist ein browserbasiertes Tool, das DSGVO-konform ist und somit vollständigen Datenschutz bietet. Die Bearbeitung der Untertitel erfolgt durch numerische Zeitangaben, wobei eine Möglichkeit zur Bearbeitung per Drag-and-drop demnächst freigeschaltet werden soll. Dieses Tool wird seit 2019 von **studiumdigitale** verwendet.

Mit [Panopto](#) können Videos aufgezeichnet, geteilt und transkribiert oder untertitelt werden. Es wird an der THM verwendet. Die Kosten für ein Jahresabo variieren je nach Anzahl der Benutzer*innen des Unternehmens bzw. der eingeschriebenen Vollzeitstudierenden einer Hochschule. Es handelt sich um ein browserbasiertes Tool, das zur Einhaltung der DSGVO verpflichtet ist. Der Startpunkt des Untertitels kann numerisch vorgegeben werden.

[YouTube](#) ist wohl das bekannteste Videoportal. Es gehört zum Großkonzern Google. Die Nutzung der Videos sowie die Erstellung der Untertitel sind kostenlos. Es handelt sich um ein browserbasiertes Tool, das allerdings nicht der DSGVO entspricht. Die Untertitelbearbeitung ist sowohl numerisch als auch per Drag-and-drop möglich.



Für die Vergleichsstudie wurde das Video mit dem Titel „[Hörbeeinträchtigungen erfahren](#)“ ausgewählt, das von **studiumdigitale** konzipiert und produziert wurde. Folgende Kriterien waren maßgeblich für die Entscheidung: Es kommen drei verschiedene Sprecher*innen vor. Es sind sowohl männliche Stimmen als auch eine weibliche Stimme zu hören. Außerdem ist das Video mit einer Länge von 03:15 Minuten relativ kurz und ermöglicht eine schnelle Auswertung. Die Auswahl liegt des Weiteren darin begründet, dass die eigentliche Tonspur durch Veränderungen so verzerrt bzw. überlagert wird, dass der Ton nicht klar verständlich ist. Neben Tief- und Hochtonschwerhörigkeit werden Tinnitus, Schallleitungsstörung und Hintergrundgeräusche simuliert sowie eine Fremdsprache (kroatisch) verwendet. Dies scheint im Hinblick auf die Auswertung der Tools interessant, da es in nicht professionell gedrehten Videos oft zu suboptimalen Verhältnissen bei der Tonaufnahme kommen kann und so der Versuch unternommen wird, diese mit Hilfe der veränderten Tonspur im Video zu simulieren. Auch dieser Aspekt soll in der größer angelegten Studie überprüft werden.

Die Vorgehensweise beruht auf folgenden Schritten: Zunächst wurde die Videodatei in das jeweilige Tool geladen und die automatischen Untertitel erstellt. Bei YouTube erforderte dies einen zusätzlichen Schritt in den Einstellungen, da die ursprünglich erstellten Untertitel Wort für Wort und nicht in Textblöcken eingeblendet werden, wenn das Video abgespielt wird. Auch bei Amberscript ist vor der Auswertung ein zusätzlicher Schritt nötig, um das Transkript in Untertitel umzuwandeln. Die Auswertung wurde von zwei Personen mit Erfahrung in der Untertitelerstellung durchgeführt. Die Ergebnisse wie Zeitmarken und auftretende Auffälligkeiten sowie Fehler wurden in einer Tabelle protokolliert. Zum Vergleich wurde das final veröffentlichte Video hinzugezogen. Um die Daten möglichst objektiv und wiederholbar zu machen, wurde weitestgehend auf die qualitative Auswertung verzichtet.



2.2 Auswertung

Quantitativ wurden folgende Kategorien herausgearbeitet und untersucht:

- I Rechtschreibfehler, durch die der Lesefluss zwar gestört, das Wort aber immer noch verständlich ist, z.B. „Hör Beeinträchtigung“ statt „Hörbeeinträchtigung“
- II Übersetzungsfehler, durch die dem eigentlichen Wort ein anderer Sinn gegeben wird, z.B. „hier helfen“ statt „Hörhilfen“
- III Worte, in denen ein Teil oder das ganze Wort gefehlt hat, z.B. „Leistung“ statt „Hörleistung“
- IV Satzzeichenfehler (d.h. Punkte und Kommata), die nicht gesetzt wurden
- V Untertiteldarstellung (Zeilenanzahl, Zeichen pro Zeile, Standzeit der Zeichen pro Sekunde)

Die Auswertung der sprachlichen Ergebnisse (I-IV) stellen die Abschnitte ohne Simulation der Hörbeeinträchtigung dar. Auf diese wird im Anschluss ein qualitativer Blick geworfen. Die Auswertung der Untertiteldarstellung (V) erfolgt bezogen auf das gesamte Video. Auswertung I-IV:

- Premiere Pro: 12 Rechtschreibfehler, 3 Übersetzungsfehler, 5 fehlende (Teil-)Worte und 0 Satzzeichenfehler
- Amberscript: 11 Rechtschreibfehler, 5 Übersetzungsfehler, 5 fehlende (Teil-)Worte und 3 Satzzeichenfehler
- Panopto: 11 Rechtschreibfehler, 4 Übersetzungsfehler, 5 fehlende (Teil-)Worte und 4 Satzzeichenfehler
- YouTube: 82 Rechtschreibfehler (8 Rechtschreibfehler und 74 Fehler bezüglich fehlender Großschreibung), 6 Übersetzungsfehler, ein fehlendes (Teil-)Wort und 46 Satzzeichenfehler (15 Satzzeichenfehler und 31 fehlende Punkte nach Satzende mit darauffolgender Großschreibung)



Premiere Pro, Amberscript und Panopto weisen nur geringfügige Unterschiede bei der Spracherkennung auf. In den meisten Fällen wurden sogar dieselben Fehler identifiziert. YouTube erstellt Untertitel ohne Beachtung von Groß- und Kleinschreibung sowie ohne Interpunktion. Dadurch ist die Fehlerrate um ein Vielfaches höher. Um die Ergebnisse mit den anderen Tools trotzdem vergleichbar zu machen, wurden bei einer zweiten Auswertung fehlende Punkte und die darauffolgende Großschreibung als Satzzeichenfehler gezählt.

Hinsichtlich der Abschnitte mit Störungen haben die Tools, bis auf YouTube, eine leicht höhere Fehlerrate aber dennoch gute Ergebnisse geliefert. Bei YouTube fehlen hingegen ganze Textpassagen. Der fremdsprachige Teil konnte bei allen Tools nicht sinnvoll transkribiert werden.

Zu beachten ist, dass dem analysierten Video ein Skript zugrunde lag und natürliche Sprache, wie sie beispielsweise in einer Vorlesung verwendet wird, andere Merkmale aufweist, wie z.B. undeutliches Sprechen wie Nuscheln, Akzente oder Füllwörter und Laute wie „äh“. Daher ist es spannend zu sehen, inwieweit sich die Ergebnisse in der Folgestudie bei anderen Videos unterscheiden. Eine weitere Frage, die sich stellt ist, ob diese „Unvollkommenheiten“ transkribiert werden sollen.

Entscheidend für einen ressourcenschonenden Arbeitsablauf mit einem Tool ist nicht nur die Qualität der Spracherkennung, sondern auch die Untertiteldarstellung. Laut Untertitel-Standard (NDR, 2015) sollen Untertitel aus maximal zwei Zeilen bestehen, je Zeile max. 37 Zeichen genutzt werden sowie die Standzeit 13 bis 15 Zeichen pro Sekunde betragen. Die untersuchten Untertitel wurden dafür als Untertiteldatei exportiert und mit dem Tool „Subtitle Edit“ überprüft. Dazu wurden die Textblöcke insgesamt gezählt sowie die Anzahl der Blöcke, die im Nachgang manuell verbessert werden müssen. Um eine Vergleichbarkeit zu ermöglichen, werden die Ergebnisse auch in Prozentzahlen ausgedrückt. Auswertung V:

- Premiere Pro: 55 Untertitel-Einblendungen, davon bei 25 zu viele Zeichen pro Zeile und bei 13 zu viele Zeichen pro Sekunde (45% zu viele Zeichen pro Zeile; 24% zu viele Zeichen pro Sekunde)
- Amberscript: 63 Untertitel-Einblendungen, davon bei 7 zu viele Zeichen pro Zeile und bei 50 zu viele Zeichen pro Sekunde (11% zu viele Zeichen pro Zeile; 79% zu viele Zeichen pro Sekunde)



- Panopto: 29 Untertitel-Einblendungen, davon bei 29 zu viele Zeichen pro Zeile und bei 27 zu viele Zeichen pro Sekunde + fließender Text, keine Absätze (100% zu viele Zeichen pro Zeile; 93% zu viele Zeichen pro Sekunde)
- YouTube: 31 Untertitel-Einblendungen, davon bei 28 zu viele Zeichen pro Zeile und 21 zu viele Zeichen pro Sekunde (90% zu viele Zeichen pro Zeile; 68% zu viele Zeichen pro Sekunde)

Es zeigen sich deutliche Unterschiede bei der Darstellung der Zeichen pro Zeile. Hier schneidet Amberscript am besten ab, bei Premiere Pro muss gut die Hälfte der Untertitel nachbearbeitet werden. Panopto und YouTube haben im Vergleich zu den Erstgenannten nahezu die Hälfte der Gesamtanzahl der Untertitелеinblendungen. Dies hat auch eine sehr viel höheren Nachbearbeitungsaufwand zur Folge. Bei der Zeichenanzahl pro Sekunde schneidet Premiere Pro am besten ab. Hier muss ca. ein Viertel der Untertitel nachbearbeitet werden. Bei den anderen sind es drei Viertel und mehr.

3. Fazit und Ausblick

Die erste Datenerhebung macht deutlich, dass bei allen vier Tools eine Nachbearbeitung der Untertitel dringend nötig ist und entsprechende Ressourcen eingeplant werden müssen. Ist die Fehlerrate hinsichtlich der Spracherkennung bei drei von vier Tools noch überschaubar, erfordert die Untertiteldarstellung bei mindestens einem Viertel der Untertitel eine Nachbearbeitung. Hier ist der Frage nachzugehen, ob die Untertitel-Standards, wie die des NDR, ihre Gültigkeit behalten. Ferner stellen Hochschulen besondere Anforderungen an ein Tool. Vor allem der Aspekt der Datenspeicherung und somit die Anwendung der DSGVO stellt das erste Ausschlusskriterium dar. Ist diese gewährleistet, sollte auf eine effiziente Untertitelbearbeitung per Drag-and-drop geachtet werden, um ressourcenschonendes Arbeiten zu ermöglichen. Ob es sich um ein browserbasiertes Tool oder eine Desktopanwendung handelt, kann je nach Kontext anders bewertet werden. Letztendlich bleibt die Kostenfrage zu klären. Welches Modell sich anbietet, liegt in der Entscheidung der Hochschule.



Bei der automatischen Untertitelung wird zukünftig lernfähige KI eine noch größere Rolle spielen, denn wenn Tools die manuellen Verbesserungen (z.B. von Fachbegriffen) für zukünftige Untertitelungen mit einbeziehen können oder Sprecher*innenerkennung eingesetzt wird, wird der Arbeitsaufwand mit der Zeit geringer und die Arbeit perspektivisch nachhaltiger. Durch große Mengen an Trainingsdaten kann durch maschinelles Lernen ein immer besser werdendes Ergebnis erwartet werden (Amberscript, 2021). Dazu wird zurzeit beispielsweise auch am Karlsruher Institut für Technologie (KIT) geforscht, wo Prof. Dr. Waibel und sein Team das erste Echtzeit-Sprachübersetzungssystem entwickelt haben.

Ein anderes Beispiel ist das Forschungs- und Entwicklungsprojekt „Speaker“ der beiden Fraunhofer-Institute IIS und IAIS, wo mittels einer Sprachassistentenplattform eine „offene, transparente und sichere Sprachassistentenanwendung“ bereitgestellt wird, die ein besonderes Augenmerk auf Datensouveränität legt (Speaker).

Für die Zukunft ist zu wünschen, dass die KI-basierten Untertitelungen so ausgereift sind, dass nicht nur die Untertitelung von Videos ohne großen Aufwand realisiert werden kann, sondern auch Live-Veranstaltungen möglichst optimal untertitelt werden und damit alle Menschen teilhaben können.



Quellen

Ableplayer. Fully accessible cross-browser HTML5 media player. Online unter: <https://ableplayer.github.io/ableplayer/> (zuletzt aufgerufen am 1.2.2022)

Adobe. Online unter: <https://www.adobe.com/products/premiere.html> (zuletzt aufgerufen am 25.1.2022)

Amberscript. Online unter: <https://www.amberscript.com/de/> (zuletzt aufgerufen am 25.1.2022)

Amberscript (10.8.2021). Wie funktioniert eine Spracherkennungssoftware? Online unter: <https://www.amberscript.com/de/blog/spracherkennungssoftware-erklart/> (zuletzt aufgerufen am 13.2.2022)

BIK für Alle (2018). Leitfaden barrierefreie Online-Videos. Online unter: <https://bik-fuer-alle.de/leitfaden-barrierefreie-online-videos.html> (zuletzt aufgerufen am 2.2.2022)

BITV 2.0 (2019). Barrierefreie-Informationstechnik-Verordnung. Online unter: https://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html (zuletzt aufgerufen am 7.2.2022)

HessenHub. Netzwerk digitale Hochschullehre Hessen. DigiBar. Netzwerk Digitale Barrierefreiheit an Hochschulen. Online unter: <https://www.hessenhub.de/netzwerk-und-news/digitale-barrierefreiheit/> (zuletzt aufgerufen am 25.1.2022)

Intersoft Consulting. Datenschutz Grundverordnung, DSGVO. Online unter: <https://dsgvo-gesetz.de/> (zuletzt aufgerufen am 2.2.2022)

ISO (2021). ISO/IEC 20071-23:2018. Online unter: <https://www.iso.org/standard/70722.html> (zuletzt aufgerufen am 15.2.2022)

Liquid Impressions KG. BITV 2.0 (2019). Online unter: <https://www.barrierefreie-webseite.de/barrierefreies-webdesign/bitv-20-2019/anforderung-12> (zuletzt aufgerufen am 10.2.2022)

NDR (2015). Untertitel Standards. Online unter: <https://www.ndr.de/fernsehen/barrierefreie-angebote/untertitel/Untertitel-Standards,untertitelstandards102.html> (zuletzt aufgerufen am 25.1.2022)



Panopto. Online unter: <https://www.panopto.com/de/> (zuletzt aufgerufen am 25.1.2022)

Puhl, S. & Lerche S. (2019). Barrierefreie Videos in der Hochschullehre In: Tolle, P.; Plümmer, A.; Horbach, A. (Hrsg.), Hochschule als interdisziplinäres barrierefreies System, kassel university press (zuletzt aufgerufen am 13.2.2022)

Speaker. Eine Sprachassistentenplattform „Made in Germany“. Online unter: <https://www.speaker.fraunhofer.de/> (zuletzt aufgerufen am 13.2.2022)

Al Smadi, T., Al Issa, H.A., Trad, E. & Al Smadi, K.A. (2015). Artificial Intelligence for Speech Recognition Based on Neural Networks. Online unter: https://www.scirp.org/html/2-3400378_55265.htm (zuletzt aufgerufen am 10.2.2022)

Technische Hochschule Mittelhessen (THM). Online unter: <https://www.thm.de/site/> (zuletzt aufgerufen am 25.01.2022)

TU Dortmund (2017). Leitfaden zur Erstellung von Untertiteln. Online unter: <https://digitale-lehre.tu-dortmund.de/storages/digitale-lehre/r/Dokumente/Anleitung-Untertitelung.pdf> (zuletzt aufgerufen am 15.1.2022)

Voß-Nakkour, S. (2013). Lean Media Production. Konzept und Unterstützung durch das Autorensystem LernBar. Dissertation. Online unter: <http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/29053> (zuletzt aufgerufen am 05.2.2022)

YouTube. Online unter: <https://www.youtube.com/> (zuletzt aufgerufen am 25.1.2022)

