

Barrierefreie Audio- und Videoinhalte

Audio- und Videoinhalte (AV-Inhalte) haben sich als wichtiger Bestandteil der heutigen Bildungslandschaft durchgesetzt. Sie eröffnen innovative Wege für das Lehren und Lernen, wodurch gängige pädagogische Ansätze ergänzt und umgestaltet werden. Auch Bibliotheken nutzen verstärkt multimediale Inhalte wie Aufzeichnungen von Schulungen, Podcasts oder instruktionale Videos. Neue digitale Plattformen und leicht zu bedienende Technologien haben die Verbreitung von Audio- und Videoinhalten demokratisiert und deutlich einfacher gemacht. Darüber hinaus erleichtert die Flexibilität dieser Medien das asynchrone Lernen, was den Bedürfnissen einer zunehmend heterogenen Nutzungsgruppe entgegenkommt und die Zugänglichkeit und Inklusion fördert.

Einleitung

Damit aber auch Menschen mit Behinderung oder gesundheitlichen Beeinträchtigungen AV-Inhalte nutzen können, müssen diese barrierefrei sein. Aber wer braucht welche Alternativen bei AV-Inhalten und warum? Welche Alternativen müssen bei welchen Inhalten bereitgestellt werden? Und wie können die Anforderungen an barrierefreie AV-Inhalte umgesetzt werden? Dieser Beitrag gibt Antworten auf diese und weitere Fragen. Weiterführende Hinweise geben die Links zu zusätzlichen Angeboten.

Alternative AV-Inhalte

Wer braucht welche Alternativen und warum?

Es hilft weiter, wenn man die Bedarfe der Nutzer*innen kennt und versteht. So kann man besser nachvollziehen, wer welche Alternativen zu AV-Inhalten benötigt und warum. Die Web Accessibility Initiative (WAI) erläutert detailliert, welche Bedarfe und Anforderungen Nutzer*innen mit verschiedenen Beeinträchtigungen haben (W3C Web Accessibility Initiative 2021):

- **Gehörlosigkeit:** Menschen, die gehörlos sind, können Audioinhalte nicht wahrnehmen und benötigen dafür eine Alternative. Häufig können gehörlose Menschen aber Text gut lesen. Die Informationen der Audioinhalte können sie als Transkript oder Untertitel lesen. Viele gehörlose Menschen bevorzugen auch die Gebärdensprache.

- Schwerhörigkeit: Menschen, die schwerhörig sind und über ein Resthörvermögen verfügen, bevorzugen oft die Audioinhalte, soweit sie diese hören können. Untertitel helfen ihnen dabei, nicht Gehörtes zu verstehen.
- Hörbeeinträchtigung: Ursache von Hörbeeinträchtigungen kann auch eine zentrale auditive Verarbeitungs- und Wahrnehmungsstörung sein. Menschen mit einer solchen Beeinträchtigung nutzen Untertitel, aber auch Transkripte, die sie in ihrem eigenen Tempo lesen können.
- Blindheit oder Sehbeeinträchtigung: Blinde oder stark sehbeeinträchtigte Menschen können die visuellen Inhalte eines Videos nicht oder nur stark eingeschränkt wahrnehmen. Eine Audiodeskription hilft ihnen, die visuellen Informationen zu verstehen.
- Taubblindheit: Taubblinde Menschen können weder die visuelle noch die auditive Komponente eines Videos wahrnehmen. Sie können aber mit einem deskriptiven Transkript die Inhalte des Videos nachlesen.
- Konzentrationsschwierigkeiten: Manche Menschen können sich nicht auf auditive oder visuelle Informationen konzentrieren und diese verstehen, wenn sich das Bildmaterial ändert. Für die meisten Videos benötigen sie auch beschreibende Transkripte.
- Neurodiversität: Auch Menschen mit Autismus-Spektrum-Störung oder Neurodiversität können von Audiodeskriptionen und Untertiteln profitieren (Fellowes 2011/2024). Erläuterungen in Form von Audiodeskriptionen können helfen, menschliche Emotionen in einem Video besser zu verstehen. Untertitel können als zusätzlicher Informationskanal dabei unterstützen, Videoinhalte besser und detaillierter zu erfassen.

Letztendlich profitieren alle Menschen von der barrierefreien Gestaltung von AV-Inhalten. Nutzer*innen können Transkripte überfliegen, anstatt einem langen Podcast zuzuhören. Untertitel helfen in geräuschsensiblen Umgebungen oder wenn die Nutzer*innen die Sprache des Videos (noch) nicht gut beherrschen.

Welche Alternativen für AV-Inhalte gibt es?

Im Zusammenhang mit der barrierefreien Gestaltung von AV-Inhalten tauchen Begriffe wie Untertitel, Audiodeskription, Caption oder Transkript auf (siehe auch WebAIM 2020):

- Untertitel: Während im Englischen zwischen *caption* und *subtitle* unterschieden wird und die Begriffe unterschiedliche Bedeutungen haben, gibt es in der deutschen Sprache nur das Wort „Untertitel“.
- Subtitles: Das englische Wort *subtitles* bezeichnet die verschriftlichte Übersetzung der gesprochenen Sprache eines Videoinhalts. Subtitle sind also für Menschen gedacht, die zwar hören, aber nicht die Sprache der Tonspur eines Videos verste-

hen können, weil sie eine andere Sprache sprechen. Für Subtitles werden auch Schilder und andere Schriftgrafiken übersetzt. Der Text wird in jedem Fall zeitsynchron mit dem Videoinhalt angezeigt.

- Captions: Die Zielgruppe von Captions sind Menschen, die gehörlos oder hörbbeeinträchtigt sind. Es werden nicht nur die gesprochenen Dialoge transkribiert, sondern auch schriftlich auf Musik, Geräusche und andere wichtige Audioinhalte hingewiesen. Zusätzlich enthalten die Captions Informationen zur Identifikation der sprechenden Personen, damit deutlich wird, wer etwas sagt. Captions liegen immer in der Originalsprache des Videos vor, es können aber auch zusätzlich Übersetzungen angeboten werden.
- Offene und geschlossene Untertitel: Offene Untertitel (engl. *open captions* – OC) sind permanent zu sehen und können nicht abgeschaltet werden, weil sie in das Videobild integriert sind. Geschlossene Untertitel (engl. *closed captions* – CC) können nach Bedarf ein- und ausgeschaltet werden.
- Transkript: Eine Textversion der gesamten gesprochenen Sprache und nicht-sprachlichen Audioinhalte eines Videos wird als Transkript bezeichnet. Die Zielgruppe solcher Transkripte sind Menschen, die gehörlos oder hörbeeinträchtigt sind. Sogenannte deskriptive oder beschreibende Transkripte enthalten zusätzlich noch Beschreibungen der visuellen Informationen eines Videos, die für das Verständnis notwendig sind. Bei einem interaktiven Transkript läuft dieses zeitsynchron mit dem Video mit und die gesprochenen Wörter werden im Transkript farbig hervorgehoben.
- Audiodeskription: Für Menschen, die den Videoinhalt nicht wahrnehmen können, werden die visuellen Inhalte in Form einer Audiodeskription bereitgestellt. Die Audiodeskription beschreibt visuelle Informationen, die für das Verstehen eines Videos wichtig sind. Im Idealfall fügt sich Audiodeskription in Sprechpausen des Videos ein.

Welche gesetzlichen Anforderungen müssen erfüllt werden?

Öffentliche Stellen sind gemäß der Europäischen Richtlinie 2016/2102 (Europäische Union 2016) dazu verpflichtet, ihre Webangebote, digitale Dokumente und mobile Anwendungen barrierefrei zu gestalten. Die dafür geltenden technischen Anforderungen legt der Durchführungsbeschluss (EU) 2021/1339 (Europäische Union 2021) fest und verweist auf die europäische Norm EN 301 549 (ETSI 2022) in der Version 3.2.1. Diese verweist in weiten Teilen auf die „Web Content Accessibility Guidelines“ (WCAG, W3C 2024), weicht aber auch in einzelnen Aspekten davon ab. Für die Barrierefreiheit von AV-Inhalten legt die EN 301 549 folgende Anforderungen fest:

- 7: Informations- und Kommunikationstechnologie mit Videofunktion
 - 7.1.1: Wenn Videos mit synchronisiertem Audioinhalt angezeigt werden, muss ein Modus zur Anzeige der verfügbaren Untertitel vorhanden sein. Wenn die Untertitel zuschaltbar sind (Closed Captions), muss es möglich sein, diese ein- und auszuschalten.
 - 7.1.2: Wenn Untertitel angezeigt werden, muss gewährleistet sein, dass die Anzeige der Untertitel mit dem Audioinhalt synchron ist.
 - 7.1.3: Wenn die IKT¹ Videos mit synchronisiertem Audioinhalt überträgt, umwandelt oder aufzeichnet, sollen die Untertitel erhalten bleiben, sich anzeigen lassen und mit dem Audioinhalt synchronisiert wiedergegeben werden. Auch zusätzliche Anzeigeparameter wie Position in der Anzeige, Schriftfarbe, -stil und Schriftart sollen möglichst erhalten bleiben.
 - 7.1.4: Wenn Untertitel zuschaltbar sind (Closed Captions), sollen diese durch die Nutzenden anpassbar sein. Dazu können die Vorder- und Hintergrundfarbe, die Schrift, Transparenz des Untertitel-Hintergrunds und Schriftkontur anpassbar sein. Aber es müssen nicht alle Parameter gleichermaßen geändert werden können, um die Anforderung zu erfüllen.
 - 7.1.5: Wenn Untertitel zuschaltbar sind (Closed Captions), müssen diese über eine Sprachausgabe wiedergegeben werden können.
 - 7.2.1: Verfügbare Audiodeskriptionen können ausgewählt und wiedergegeben werden. Entweder gibt es dafür ein eigenes Bedienelement oder es kann eine der Tonspuren des Videos als Audiodeskription ausgewählt werden.
 - 7.2.2: Wenn Audiodeskriptionen wiedergegeben werden, muss gewährleistet sein, dass die Anzeige der Untertitel mit dem Audioinhalt synchron ist.
- 9.1.2.1: Zeitbasierte Medien
 - 9.1.2.1.1: Für aufgezeichnete Nur-Audio-Inhalte und Nur-Video-Inhalte steht eine inhaltlich gleichwertige Alternative zur Verfügung. Das ist in der Regel ein Transkript oder ein deskriptives Transkript. Bei Nur-Video-Inhalten kann auch ein zusätzlicher Audio-Inhalt als Deskription wichtige visuelle Inhalte des Videos wiedergeben.
 - 9.1.2.1.2: Für aufgezeichnete Audioinhalte in synchronen Medien (Video mit Audio, Multimedia) stehen Untertitel zur Verfügung.
 - 9.1.2.1.3: Für aufgezeichnete Videoinhalte in synchronen Medien steht eine Audiodeskription oder eine Medienalternative zur Verfügung.
 - 9.1.2.1.4: Für live übertragene Audioinhalte in synchronen Medien stehen Untertitel zur Verfügung.
 - 9.1.2.1.5: Für aufgezeichnete Videoinhalte in synchronen Medien steht eine Audiodeskription zur Verfügung. (ETSI 2022)

¹ IKT = Informations- und Kommunikationstechnologie.

Das Kapitel 9 und alle darin enthaltenen Anforderungen der EN 301 549 beziehen sich auf Webinhalte, inhaltlich identische Anforderungen müssen aber auch Nicht-Webinhalte (Kapitel 10) und Software (Kapitel 11) erfüllen.

Vielleicht ist Ihnen aufgefallen, dass sich die Anforderungen 9.1.2.3 und 9.1.2.5 sehr ähneln. Das liegt daran, dass Anforderung 9.1.2.3 der WCAG-Stufe A entspricht, die Anforderung 9.1.2.5 der Stufe AA. Für die niedrigste Stufe A genügt eine Audiodeskription oder eine Medienalternative für Videoinhalte. Für die Stufe AA muss es eine Audiodeskription sein. Die WCAG-Stufen A und AA entsprechen den gesetzlichen Mindest-Anforderungen.

Die darüberhinausgehenden Anforderungen 1.2.6 (Gebärdensprache), 1.2.7 (erweiterte Audiodeskription), 1.2.8 (Medienalternative) und 1.2.9 (Live-Nur-Audio) der WCAG werden in der EN 301 549 lediglich für eine erweiterte Barrierefreiheit empfohlen. Das bedeutet, dass eine gemäß den gesetzlichen Regelungen barrierefreie Gestaltung von AV-Inhalten keine Gebärdensprachevideos erfordert, obwohl viele gehörlose Menschen bevorzugt in Gebärdensprache kommunizieren.

Was bedeutet das denn jetzt für das Lehren und Lernen?

Müssen jetzt alle Videoinhalte mit Untertiteln, Audiodeskriptionen und Transkripten versehen werden? Und was ist mit reinen Audioinhalten wie Podcasts? Ein Punkt ist dabei zu beachten: AV-Inhalte müssen dann zugänglich gestaltet werden, wenn die Audio- bzw. Videoinhalte nicht bereits eine Alternative zu Textinhalten darstellen und eindeutig als solche gekennzeichnet sind. Wenn beispielsweise mit einem schriftlichen Tutorial erläutert wird, wie ein Bibliothekskatalog genutzt werden kann, dann benötigt ein ergänzendes Video dazu keine Untertitelung, Audiodeskription oder Transkript. Denn in diesem Fall ist der ausführliche Text der Originalinhalt und das Video eine Alternative zu diesem Textinhalt.

In einer Checkliste erläutert die WAI, welche Alternativen für welche Form von Inhalten erforderlich sind. Die Regelungen der Bundes-BITV bzw. der EN 301 549 werden im Folgenden anhand typischer Szenarien illustriert.

Podcasts: Podcasts und andere Formen von aufgezeichneten Audioinhalte benötigen ein Transkript. Wenn Sie den Podcast aber lediglich als Annehmlichkeit für die Studierenden erstellen und die Inhalte auch als Text auf der Webseite zu finden sind, ist ein Transkript nicht erforderlich. Der Podcast wäre dann nur eine Alternative zu Textinhalten. Meistens sind Podcasts aber eigenständige Inhalte, beispielsweise Gespräche zwischen zwei oder mehr Personen. In diesem Fall müssen Sie für den Podcast ein Transkript erstellen. Denken Sie daran, dass es darum geht, Menschen, die den Audioinhalt nicht hören können, mit den Informationen zu versorgen, die im Podcast zu hören sind.

Videoaufzeichnungen: Häufig werden Vorlesungen, Videoanleitungen und ähnliche Videos als aufgezeichnete Inhalte bereitgestellt. Diese können mittlerweile mit geringem Aufwand am eigenen Rechner aufgenommen oder (semi-)professionell in einem Videostudio produziert werden. Wenn ein Video eine Tonspur mit gesprochener Sprache oder anderen Audioinhalten enthält, die wichtig sind, um das Video zu verstehen, werden Untertitel benötigt. Enthält das Video außerdem visuelle Informationen, die zum Verständnis erforderlich sind, dann müssen zusätzlich ein (deskriptives) Transkript und eine Audiodeskription bereitgestellt werden. Wenn eine sprechende Person im Video die relevanten visuellen Inhalte wie beispielsweise Bedienabläufe oder Folien ausführlich erläutert, kann es sein, dass auf ein deskriptives Transkript und eine Audiodeskription verzichtet werden kann.

Videokonferenzen: Seit der Corona-Pandemie werden an vielen Hochschulen und auch im beruflichen Umfeld synchrone Videokonferenzen als Ersatz für Präsenzveranstaltungen genutzt. Diese müssen untertitelt und relevante visuelle Informationen in zugänglicher Form verschriftlicht werden. Manche Videokonferenzsysteme stellen dafür Schnittstellen bereit, die Dienstleister nutzen, um Live-Untertitel in die laufende Videokonferenz einzuspielen, oder es werden automatische Live-Untertitel bereitgestellt. Gerade die Qualität solcher automatischen Untertitel ist von vielen Faktoren abhängig und nicht immer optimal. Eine pragmatische Möglichkeit ist, Schriftdolmetscher*innen zur Videokonferenz hinzuzuschalten. Diese verschriftlichen die gesprochenen Beiträge live. Die Nutzer*innen können die Schriftdolmetschung dann in einem separaten Fenster mitlesen. Aber was ist mit Folien oder mit Abbildungen die gezeigt werden? Auch für diese wichtigen visuellen Inhalte müssen zugängliche Alternativen gefunden werden. Präsentationen und andere visuelle Inhalte können auch vorab als barrierefreie Version zur Verfügung gestellt werden. Abbildungen kann die vortragende Person im notwendigen Umfang erläutern.

Wie können diese Anforderungen umgesetzt werden?

Es ist deutlich geworden, dass die barrierefreie Gestaltung von AV-Inhalte nicht trivial ist. Ein Video untertitelt man nicht mal eben nebenbei. Aber der Aufwand lohnt sich für alle Nutzenden, vor allem aber natürlich für Menschen mit Behinderung. Wie ein Video mit Audiodeskription, Untertitelung und Transkript dann aussehen kann, zeigt Abb. 1.

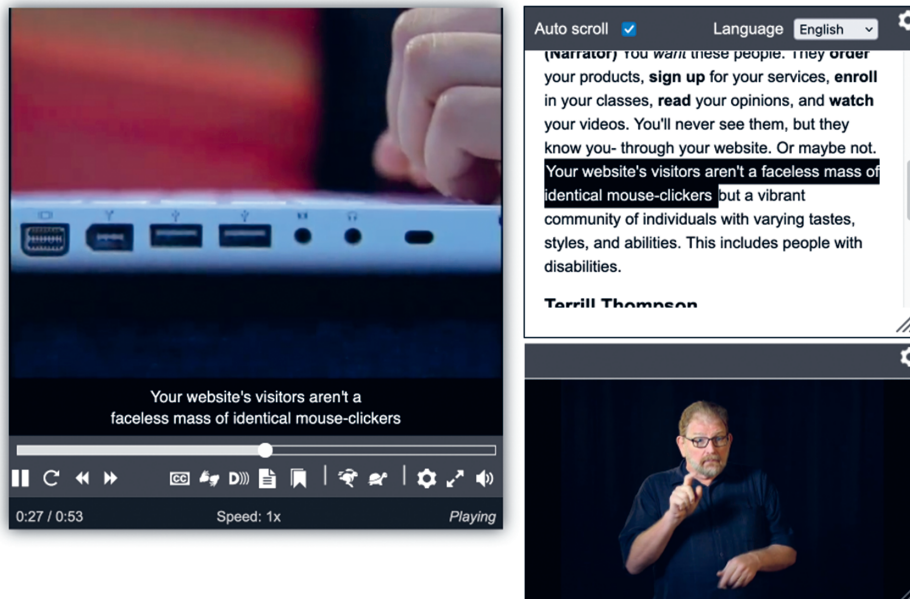


Abb. 1: AblePlayer spielt ein Video mit Untertiteln, deskriptivem Transkript und Gebärdensprachvariante ab.

Das Video in Abb. 1 enthält Untertitel in mehreren Sprachen, Audiodeskription sowie eine Gebärdensprachübersetzung des gesprochenen Textes. Zusätzlich kann ein interaktives Transkript angezeigt werden, das synchron mit dem Video läuft. Wird ein Wort im Transkript ausgewählt und angeklickt, springt das Video zur entsprechenden Zeitmarke.

Rahmenbedingungen schaffen

Die Erstellung von Untertiteln, Transkripten und Audiodeskriptionen erfordert personelle und finanzielle Ressourcen. Die Bereitstellung dieser Ressourcen ist Aufgabe der Anbieter der AV-Inhalte. Weil die barrierefreie Gestaltung so ressourcenintensiv ist, sollte überlegt werden, für welche Audio- und Videoinhalte Untertitel, Transkripte und Audiodeskriptionen erstellt werden. Hier wird man Prioritäten setzen müssen, bspw. bezogen auf die Dringlichkeit, die Nutzungsdauer oder die inhaltliche Relevanz der AV-Inhalte. Ebenso sollte entschieden werden, ob ein eigener Service aufgebaut oder externe Dienstleister beauftragt werden, denn beides hat Vor- und Nachteile.

Vorbereitung und Aufnahme

Die barrierefreie Gestaltung von AV-Inhalten beginnt schon vor der Aufnahme (W3C Web Accessibility Initiative 2021). Zunächst muss klar sein, welche Aspekte der Barrierefreiheit berücksichtigt und erfüllt werden müssen. Nicht jeder AV-Inhalt erfordert immer Untertitel, deskriptives Transkript sowie Audiodeskription. Wird beispielsweise der visuelle Inhalt ausführlich im AV-Inhalt beschrieben, kann es sein, dass eine Audiodeskription nicht mehr benötigt wird.

- Überlegen Sie, ob Sie das Projekt intern bearbeiten können oder teilweise oder vollständig an externe Firmen vergeben.
- Berücksichtigen Sie bei der Planung, dass Untertitel einen Teil des Videobilds verdecken. Audiodeskriptionen benötigen ebenso Pausen im gesprochenen Wort, damit sie in diesen Pausen wiedergegeben werden können.
- Ein detailliertes Skript vereinfacht die Videoerstellung. Es kann ebenso als Grundlage für Untertitel und ein Transkript genutzt werden.
- Sprechen Sie bei der Aufnahme langsam und deutlich, vermeiden Sie unnötige Hintergrundgeräusche und sorgen Sie für eine gute Bild- und Ton-Qualität. Investieren Sie daher Geld in qualitativ gute Geräte, wenn Sie AV-Inhalte selbst aufzeichnen. Die im Rechner eingebaute Kamera und das Mikrofon genügen in der Regel qualitativ nicht.
- Die Qualität eines AV-Inhalts steht und fällt mit der Vorbereitung. Erläutern bzw. verbalisieren Sie wichtige visuelle Inhalte wie Folien, Diagramme oder Schaubilder. Das integriert die Beschreibung in den AV-Inhalt. Eine Audiodeskription wird dadurch möglicherweise überflüssig oder fällt knapper aus.

Weitere Tipps und Hinweise stellt die WAI auf ihren Webseiten bereit (W3C Web Accessibility Initiative 2021). Die Universität Göttingen hat einen ausführlichen Leitfaden speziell zur Erstellung barrierefreier Lehrvideos im Hochschulkontext erstellt, der weitere Aspekte umfassend beleuchtet (Berginski u. a. 2023). Zwar hat die Anzahl der Webkonferenzen seit dem Ende der Covid-19-Pandemie abgenommen, aber dennoch hilft es, sich mit Informationen zur Vorbereitung und Durchführung barrierefreier Webkonferenzen vertraut zu machen. Die Bundesfachstelle Barrierefreiheit hält dazu weiterführende Links bereit (Bundesfachstelle Barrierefreiheit o. J.).

Untertitel, Transkripte und Audiodeskriptionen erstellen

Untertitel, Transkripte und Audiodeskriptionen selbst zu erstellen, ist zeit- und arbeitsintensiv. Die Basis für Untertitel und Transkripte ist immer eine Transkription des Audioinhalts eines AV-Inhalts. Entweder wird dieses manuell transkribiert oder zumindest als Grundlage per Spracherkennung erstellt. Ein so erstelltes Transkript muss dann weiterverarbeitet werden.

Wenn mehrere Personen sprechen, müssen die Sprechenden im Transkript identifiziert werden. Auch alle nicht-sprachlichen Inhalte, die für das Verständnis relevant sind, müssen transkribiert werden. Transkripte sollten die sprachlichen Inhalte so detailliert wie nötig und so ausführlich wie möglich wiedergeben. Es ist daher in der Regel nicht in Ordnung, Text zu paraphrasieren. Aber Verzögerungslaute wie „äh, mhh“ oder Wiederholungen müssen nicht zwingend transkribiert werden, sofern es nicht für das Verständnis notwendig ist. Wie genau und detailliert ein Audioinhalt transkribiert wird, hängt auch vom Kontext ab, in dem der AV-Inhalt eingesetzt wird.

Aus der Transkription können anschließend das Transkript und die Untertitel erstellt werden. Für Untertitel werden Zeitmarken eingefügt und die Untertitelabschnitte festgelegt. Zusätzlich können Untertitel mit besonderen Zeichen markiert oder farbig gestaltet werden. Für diese Arbeiten gibt es spezielle Software oder Online-services. Beispielsweise können Videos mit YouTube² automatisch untitled und die Untertitel danach manuell nachbearbeitet werden. Amara³ ist eine Onlineplattform zur kollaborativen Untertitelung von Videos. Wer seine Videos lokal am eigenen Rechner untiteln möchte, kann das mit kostenlosen Editoren wie CADET,⁴ Subtitle Edit,⁵ Tero Subtitler⁶ oder Jubler⁷ tun.

Aber wie sollen Untertitel denn jetzt genau aussehen? In vielen deutschsprachigen Leitfäden und Hinweistexten wird auf Untertitelrichtlinien verwiesen, die im Kontext von Videoproduktionen für Film und Fernsehen entstanden sind, wie zum Beispiel die UT-Richtlinien des NDR (NDR 2015). Dort wird bspw. empfohlen, maximal 37 Zeichen pro Zeile und ein bis zwei Zeilen pro UT-Einheit zu nutzen. Aber wo kommen diese 37 Zeichen überhaupt her? Offenbar sind es technische Einschränkungen, die zu dieser Zeichenanzahl geführt haben. Bei der Übertragung von Untertiteln per Tele-text stehen 40 Zeichen zur Verfügung, wobei davon mindestens 3 für Steuerzeichen vorgesehen sind. An anderer Stelle schlagen Díaz Cintas / Remael (2019) einen Sicherheitsbereich von 10 Prozent der Videobreite vor, ausgehend von einer Videogröße von 720 x 576 Pixeln. Innerhalb dieses Sicherheitsbereichs sollten keine Untertitel platziert werden. Anhand verschiedener Beispiele kommen sie so auf die typische Länge einer UT-Zeile von 37 Zeichen, die aber auch zwischen 33 bis 43 Zeichen schwanken kann (Díaz Cintas / Remael 2019, 84). Es gibt offenbar wenig empirisch abgesicherte Zahlen. Die meisten Angaben, auch zu Standzeiten und Lesegeschwindigkeiten, basieren scheinbar eher auf Erfahrungen und Traditionen. Man sollte schon hinterfragen, ob es sinnvoll ist, Untertitelrichtlinien für das Fernsehen auch für Online-Videos zu übernehmen.

2 YouTube: <https://www.youtube.com> (21.06.2024).

3 Amara: <https://amara.org/> (21.06.2024).

4 CADET: <https://www.wgbh.org/foundation/services/ncam/cadet> (21.06.2024).

5 Subtitle Edit: <https://www.nikse.dk/subtitleedit> (21.06.2024).

6 Tero Subtitler: <https://github.com/URUWorks/TeroSubtitler> (21.06.2024).

7 Jubler: <https://www.jubler.org> (21.06.2024).

Es fällt auf, dass die BBC in ihren Empfehlungen zu Standzeiten und Zeilenlängen deutlich flexibler agiert (BBC 2024). Als maximale Zeilenbreite werden 68 Prozent der Videobreite bei einem Video im Seitenverhältnis von 16 : 9 und sogar 90 Prozent bei einem Seitenverhältnis von 4 : 3 empfohlen. Und für die Berechnung von Standzeiten geht die BBC von einer Lesegeschwindigkeit von 160 bis 180 Wörtern pro Minute aus und empfiehlt eine minimale Standzeit von 0,3 Sekunden pro Wort. Die endgültige Entscheidung trifft aber immer die Person, die die Untertitel erstellt. Denn die Standzeiten hängen von weiteren Faktoren wie der Sprechgeschwindigkeit, der Textgestaltung und der Schnittsynchronisation ab.

In den vergangenen Jahren ist die Qualität der automatischen Spracherkennung (ASR) durch die Entwicklungen im Bereich der künstlichen Intelligenz (KI) deutlich besser geworden. Viele große IT-Firmen bieten mittlerweile entsprechende Services an, um Videos automatisch zu transkribieren und zu untertiteln. Mit der Software Whisper hat die Firma OpenAI die ASR mit großen Sprachmodellen (engl. *large language models* – LLM) kombiniert, um die Qualität noch weiter zu verbessern. Untersuchungen zeigen, dass Whisper qualitativ zuverlässiger arbeitet als andere ASR-Systeme (Kuhn u. a. 2024). Andere Forschende nutzen KI, um automatisiert Audiodeskriptionen zu erstellen (Stangl u. a. 2023; Bodi u. a. 2021) oder Feedback zur Qualität von manuell erstellten Audiodeskriptionen zu geben (Natalie u. a. 2023). Wer selbst ausprobieren möchte, wie gut Whisper zur automatischen Untertitelung taugt, findet im Internet eine kuratierte Liste mit Implementierungen von Whisper.⁸ Whisper kann außerdem zusammen mit o. g. Software Subtitle Edit sowie Tero Subtitler genutzt werden.

In der Regel reicht aber die Qualität solcher Services bei weitem noch nicht aus (W3C Web Accessibility Initiative 2021). Auch eine zuverlässige automatische Live-Untertitelung ist damit noch nicht möglich. Whisper hat – ähnlich wie ChatGPT – das Problem, Texte zu „halluzinieren“. Die damit erstellten Transkripte und Untertitel zeichnen sich durch relativ lange Satzkonstruktionen aus, was die Nutzung für die Untertitelung erschwert. Ohne menschliche Kontrolle und Nachbearbeitung sollten automatisch erstellte Untertitel und Transkripte nicht genutzt werden. Auf absehbare Zeit bleibt daher ein *human in the loop* (HITL) bei der Nutzung von KI-basierten Werkzeugen unverzichtbar (Trizna u. a. 2024).

Technische Umsetzung

Der HTML5-Standard bietet mit dem <video>-Element⁹ eine einfache Möglichkeit, Videoinhalte in Webseiten zu integrieren. Damit ist es auch möglich, Untertitel im

⁸ Implementierungen von Whisper: <https://github.com/sindresorhus/awesome-whisper> (21.06.2024).

⁹ Mozilla Developer Network – MDN (2024): <video> : The Video embed element. <https://developer.mozilla.org/en-US/docs/Web/HTML/Element/video> (28.05.2024).

Format WebVTT einzubinden.¹⁰ Für Audioinhalte gibt es analog dazu das <audio>-Element. Damit ist aber die Einbindung von Textalternativen nicht möglich. Hier wird empfohlen, Audioinhalte entweder auch mittels <video>-Element einzubinden oder auf technische Alternativen auszuweichen. Eine solche technische Alternative ist der MediaElementPlayer (Dyer o. J.). Noch vielfältigere Möglichkeiten bietet der AblePlayer (Thompson o. J.). Dieser ist vollständig barrierefrei bedienbar, unterstützt neben Untertiteln auch Deskriptionen, Kapitelnavigation und Metadaten. Nutzende können mit interaktiven Transkripten im Video navigieren, Gebärdensprachvarianten eines Videos aufrufen oder sich eine textbasierte Audiodeskription vorlesen lassen. Sowohl der MediaElementPlayer als auch der AblePlayer erfüllen damit die Anforderung, die Barrierefreiheit von AV-Inhalten zu unterstützen, als auch selbst barrierefrei bedienbar zu sein.

Viele Lern- (LMS) oder Contentmanagement-Systeme (CMS) bringen mittlerweile eigene Videoplayer mit, die in der Regel auch die Einbindung von Transkripten und Untertiteln unterstützen. So können auch Videos von Videoplattformen eingebunden und bei Bedarf mit Untertiteln, Textdeskription und Transkript versehen werden. Wie Untertitel zu vielen verschiedenen Videodiensten hinzugefügt werden, erläutert die University of Washington.¹¹

Bei einer Videokonferenz muss natürlich auch die eingesetzte Software barrierefrei sein. Die Anbieter entsprechender Software stellen im Internet ausführliche Informationen bereit, beispielsweise für Adobe Connect,¹² Zoom,¹³ Microsoft Teams,¹⁴ Google Meet,¹⁵ BigBlueButton¹⁶ oder Cisco WebEx.¹⁷ Die Bundesfachstelle Barrierefreiheit informiert ebenfalls über die Zugänglichkeit von populären Videokonferenzsystemen (Bundesfachstelle Barrierefreiheit o. J.).

10 Mozilla Developer Network – MDN (2023): Adding captions and subtitles to HTML video – Web media technologies MDN. https://developer.mozilla.org/en-US/docs/Web/Media/Audio_and_video_delivery/Adding_captions_and_subtitles_to_HTML5_video (28.05.2024).

11 University of Washington: Accessible Technology – Videos: <https://www.washington.edu/accesstech/videos/> (28.05.2024).

12 Adobe – Funktionen zur Barrierefreiheit in Adobe Connect: <https://helpx.adobe.com/content/help/de/de/adobe-connect/using/accessibility-features.html> (28.05.2024).

13 Zoom – Barrierefreiheit: <https://explore.zoom.us/de/accessibility/> (28.05.2024).

14 Barrierefreiheitstools für Microsoft Teams – Microsoft-Support: <https://support.microsoft.com/de-de/office/barrierefreiheitstools-für-microsoft-teams-2d4009e7-1300-4766-87e8-7a217496c3d5> (28.05.2024).

15 Google – Nutzerhandbuch zur Barrierefreiheit: <https://support.google.com/a/users/answer/1631886?hl=de> (28.05.2024).

16 BigBlueButton – Accessibility: <https://docs.bigbluebutton.org/accessiblity/> (28.05.2024).

17 Cisco Systems (2023): Funktionen für Barrierefreiheit für Meetings und Webinare. In: Webex-Hilfecenter. <https://help.webex.com/de-de/article/84har3/Funktionen-f%C3%BCr-Barrierefreiheit-f%C3%BCr-Meetings-und-Webinare> (28.05.2024).

Fazit

Die Gestaltung barrierefreier AV-Inhalte ist anspruchsvoll. Deshalb ist es sinnvoll, mit einer barrierefreien Umsetzung schon bei der Planung und Vorbereitung zu beginnen. Detaillierte Skripte zur Videoerstellung können im Anschluss genutzt werden, um Transkript, Untertitel und ggf. Audiodeskription zu erstellen. Oft ist es einfacher, mit der Hilfe von Expert*innen AV-Inhalte mit Untertiteln oder einer Audiodeskription zu versehen. Die hier aufgeführten Tipps und Hinweise geben Ihnen einen guten Einstieg in das Thema. Nutzen Sie die weiterführenden Links und Literaturhinweise, um mehr zu diesem spannenden Thema zu erfahren.

Autor

Dr. Björn Fisseler hat Sonderpädagogik studiert, im Schuldienst gearbeitet und in Rehabilitationswissenschaften promoviert. Er arbeitet als Experte für digitale Bildung an der FernUniversität in Hagen. Dabei unterstützt er Lehrende in Fragen rund um die Digitalisierung und Innovation in der Lehre. Seit mehr als 15 Jahren forscht er zu Fragen der barrierefreien und inklusiven Gestaltung digitaler Bildung. Seine Expertise bringt er in zahlreiche nationale und internationale Projekte ein (u. a. EADTU Taskforce Diversity & Inclusion, Hochschulforum Digitalisierung, Project Ed-ICT, ELoQ-Projekt).

Literatur und Quellen

- BBC (2024): Subtitle guidelines. <https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles> (28.05.2024)
- Berginski, Julia u. a. (2023): Leitfaden zur Erstellung barrierefreier Lehrvideos im universitären Kontext: Projekt DaLeLe4All. Göttingen: Universitätsverlag Göttingen. <https://doi.org/10.17875/gup2023-2355>
- Bodi, Aditya u. a. (2021): Automated video description for blind and low vision users. In: Extended abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3411763.3451810>
- Bundesfachstelle Barrierefreiheit (o. J.): Barrierefreie Webkonferenzen. https://www.bundesfachstelle-barrierefreiheit.de/DE/Fachwissen/Informationstechnik/Barrierefreie-Webkonferenzen/barrierefreie-webkonferenzen_node.html (28.05.2024)
- Díaz Cintas, Jorge / Remael, Aline (2019): Audiovisual translation: Subtitling. London: Routledge. <https://doi.org/10.4324/9781315759678>
- Dyer, John (o. J.): MediaElement.js – HTML5 video and audio unification framework. <https://www.mediaelementjs.com/> (28.05.2024)
- ETSI (2022): EN 301 549. Accessibility requirements for ICT products and services. <https://www.etsi.org/human-factors-accessibility/en-301-549-v3-the-harmonized-european-standard-for-ict-accessibility> (28.05.2024)

- Europäische Union (2016): Richtlinie (EU) 2016/2102 über den barrierefreien Zugang zu den Websites und mobilen Anwendungen öffentlicher Stellen. <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016L2102> (28.05.2024)
- Europäische Union (2021): Durchführungsbeschluss (EU) 2021/1339 der Kommission vom 11. August 2021 zur Änderung des Durchführungsbeschlusses (EU) 2018/2048 über die harmonisierte Norm für Websites und mobile Anwendungen. <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32021D1339> (28.05.2024)
- Fellowes, Judith (2011/2024): Autistic spectrum, captions and audio description. <https://sunsurfer.co.uk/autistic-spectrum-captions-and-audio-description/> (28.05.2024)
- Kuhn, Korbinian u. a. (2024): Measuring the accuracy of automatic speech recognition solutions. In: ACM Transactions on Accessible Computing 16 (4), S. 1–23. <https://doi.org/10.1145/3636513>
- Natalie, Rosiana u. a. (2023): Supporting novices author audio descriptions via automatic feedback. In: CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, April 2023, Article No. 77, S. 1–18. <https://doi.org/10.1145/3544548.3581023>
- NDR (2015): Untertitel-Standards. https://www.ndr.de/fernsehen/barrierefreie_angebote/untertitel/Untertitel-Standards,untertitelstandards102.html (28.05.2024)
- Stangl, Abigale u. a. (2023): The potential of a visual dialogue agent in a tandem automated audio description system for videos. In: ASSETS '23: Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility, October 2023, Article No. 32, S. 1–17. <https://doi.org/10.1145/3597638.3608402>
- Thompson, Terrill (o. J.): Able Player. <https://ableplayer.github.io/ableplayer/> (28.05.2024)
- Trizna, Michael u. a. (2024): Evaluating OpenAI Whisper and its variants for increasing the accessibility of audio collections. <https://doi.org/10.25573/data.24938859.v1>
- W3C (2024): Web Content Accessibility Guidelines (WCAG) 2.2. <https://www.w3.org/TR/WCAG22/> (28.05.2024)
- W3C Web Accessibility Initiative (2021): Making Audio and Video Media Accessible. In: Web Accessibility Initiative (WAI). <https://www.w3.org/WAI/media/av/> (28.04.2024)
- WebAIM (2020): Captions, Transcripts, and Audio Descriptions. <https://webaim.org/techniques/captions/> (28.05.2024)

