

# FORSCHUNGSPROJEKT ZU UNTERTITELN IN 360°-VIDEOS

PETER THO PESCH

Virtual Reality ist seit einiger Zeit ein allgegenwärtiges Thema. Gleichzeitig stellt das Aufsetzen einer VR-Brille, auch auf Messen, für viele eine Hürde dar. VR-Erlebnisse werden mal positiver, mal negativer wahrgenommen. Die Zahl der verfügbaren Endgeräte, sowie der Umfang der angebotenen Inhalte nimmt dabei weiter stetig zu. Auch Rundfunkanstalten setzen auf VR-Technologie, um neue Formate zu testen. Neben „echtem VR“ geschieht dies in erster Linie in Form von 360°-Videos, die idealerweise mit einer VR-Brille, aber auch über Tablets, Smartphones und PCs betrachtet werden können. Barrierefreie Dienste wie Untertitel, Gebärdensprache und Audiodeskrip-

tion werden dabei oft stiefmütterlich behandelt. Wo es im Fernsehen längst etablierte Standards gibt (technische als auch redaktionelle), fehlt es bei den immersiven Angeboten an Erfahrung dazu, wie hör- und sehgeschädigten Nutzern die Inhalte am besten vermittelt werden können. Genau hier setzt das EU-Projekt Immersive Accessibility (ImAc) an. Die insgesamt neun Projektpartner haben es sich zum Ziel gesetzt, barrierefreie Angebote für immersive Inhalte zu verbessern. Dieser Artikel stellt die Ansätze und Ergebnisse für Untertitel in 360°-Videos vor.

Virtual Reality has been an omnipresent topic for some time now. At the same time, putting on VR glasses, even at trade fairs, is a hurdle for many people. VR experiences are sometimes perceived more positively, sometimes more negatively. The number of available devices, as well as the amount of offered content is constantly increasing. Broadcasters also offer content, primarily 360°-videos, which can ideally be viewed with VR glasses, but also via tablets and PCs. Barrier-free services such as subtitles, sign language and audio description are often neglected. Where television has established standards (both technical and editorial), there is a lack of experience with the immersive offerings as to how hearing and visually impaired users can best communicate the content. This is precisely where the EU project Immersive Accessibility (ImAc) comes in. The nine project partners have set themselves the goal of improving barrier-free offerings for immersive content. This article presents the approaches and results for subtitles in 360° videos.



Abbildung 1. Aktiver Untertitel mit Pfeil-Indikator zum Anzeigen der Sprecherposition in der Szene.



Abbildung 2. Aktiver Untertitel mit Radar-Indikator zum Anzeigen der Sprecherposition in der Szene.

## Indikatoren für Sprecher – Am Bildschirm fixierte Untertitel

Untertitel werden bereits auf verschiedenen Plattformen auch für 360°-Videos angeboten und dann meist im unteren Sichtfeld zentriert dargestellt. Das heißt dreht man den Kopf, wandern die Untertitel mit und bleiben somit lesbar.

Für diese Darstellungsform wurde im Projekt untersucht, ob zusätzliche Indikatoren für Sprecher als hilfreich empfunden werden. Dafür werden neben den Untertiteln Pfeile eingeblendet, die den Zuschauer zum Sprecher führen. In einer anderen Option wird eine Art Radarsystem eingeblendet, welches die Sprecher anzeigt. Beide Varianten kamen bei den Nutzern gut an, wobei die Pfeil-Indikatoren etwas beliebter sind. Im Folgenden sind zwei Beispiele beschrieben.

**Pfeil-Indikator:** Ein Pfeil wird links oder rechts neben dem Untertitel eingeblendet und zeigt in die Richtung des jeweiligen Sprechers. Befindet sich der Sprecher im sichtbaren Bildausschnitt, wird der Pfeil ausgeblendet. Beispiel (Abbildung 1) zeigt einen Ausschnitt eines 360°-Videos. Ein Reporter ist im Bild sichtbar, aber der eingeblendete Untertitel gehört zu dem Produzenten. Dieser ist nicht im Bildausschnitt zu sehen.

**Radar-Indikator:** Ein Radarsystem wird im Bild durchgehend eingeblendet. Es zeigt die 360°-Szene von oben und deutet die aktuelle Blickrichtung des Zuschauers an. Die zu einem Untertitel gehörige Sprecherposition wird durch ein kleines Dreieck (in der Farbe des Untertitels) am Rand des Radars signalisiert. Im Beispiel (Abbildung 2) zeigt dies an, dass sich der Sprecher (in diesem Fall der Reporter) rechts des aktuellen Bildausschnitts befindet.

### Alternative Darstellungsform – Am Video fixierte Untertitel

Eine zweite Darstellungsform für Untertitel im 360°-Raum unterscheidet sich grundlegend von oben gezeigten Beispielen. Dabei wird der Untertitel nicht am aktuellen Bildausschnitt, sondern am Video fixiert. Er verhält sich damit zunächst einmal wie ein im Video eingebrannter Untertitel. Ein großer Nachteil bei diesem Ansatz ist, dass der Untertitel für den Zuschauer nur sichtbar ist, wenn dieser in die entsprechende Richtung guckt. Eben genau wie Videoinhalte in einer 360°-Szene. Trotzdem lohnt sich ein näherer Blick auf diese Darstellungsform, denn sie führen bei der Nutzung von VR-Brillen möglicherweise zu einer höheren Qualität der immersiven Erfahrung. Das lassen zumindest frühere Untersuchungen der Ludwig-Maximilians-Universität München vermuten [Rothe].

In der Praxis wird diese Darstellungsform zum Beispiel so genutzt, dass der Untertitel an mehreren Positionen im 360°-Video eingebrannt wird (das heißt, er ist in diesem Fall auch nicht mehr abschaltbar). So kann gewährleistet werden, dass der Zuschauer immer einen der Untertitel sieht, egal in welche Richtung er gerade guckt.

Diese Variante wurde im ImAc Projekt mit 40 Testpersonen gegenüber der Variante mit am Bildschirm fixierten Untertiteln getestet. Die Untertitel wurden dabei an drei jeweils um 120° versetzten Positionen am Video fixiert. Gut 82,5 Prozent der Testpersonen bevorzugten jedoch insgesamt die am Bildschirm fixierten Untertitel. Als Hauptgrund wird hierfür genannt, dass man sich besser umsehen kann. Nutzer, welche die am Video fixierten Untertitel bevorzugten nannten als Grund, dass diese besser lesbar sind. Im EU-Projekt ImAc wurde weiter eine Variante getestet, bei welcher die Untertitel am Sprecher fixiert werden. Zusätzlich sollte eine Hilfestellung gegeben werden, wann immer der Zuschauer den Sprecher und damit auch den Untertitel nicht im Blickfeld hat. Dabei wurden verschiedene Möglichkeiten diskutiert. Die langjährige Erfahrung des IRT im Bereich der Untertitelung ist hierbei sowohl in die Konzepte zur Darstellung der Untertitel als auch in Lösungen zum geeigneten Transport der entsprechenden Metadaten eingeflossen. Getestet wurde am Ende die Hilfestellung durch die Einblendung eines Pfeils, der am Bildschirm fixiert ist und den Zuschauer zum Sprecher lenkt. In Diskussion war auch die Möglichkeit zusätzlich zu genanntem Pfeil den Untertiteltext mit einzublenden. Eine entsprechende Demo findet sich im IRT Lab [IRTLab]. In Hinblick auf die untenstehenden Testergebnisse hätte diese Option möglicherweise zu einer größeren Akzeptanz für diese Darstellungsoption geführt.



Abbildung 3. Am Sprecher fixierte Untertitel.

Die Ergebnisse der durchgeführten Nutzertests mit Schwerhörigen und Tauben zeigen, dass hier jedoch noch Luft für Verbesserungen ist. Der Großteil der Nutzer bevorzugt die Fixierung am Bildschirm (knapp 80 Prozent bei insgesamt 24 Testpersonen aus zwei Testgruppen). Als Hauptgrund wird genannt, dass sie in diesem Modus die Untertitel nicht verpassen können. Er gewähre mehr Freiheit sich umzuschauen.

Allerdings wurden auch mehrere Vorteile der am Video fixierten Untertitel genannt. Ein Nutzer berichtet, dass die Darstellung intuitiver und besser integriert ist, ein anderer, dass es mehr wie im echten Leben ist. Außerdem sei die Identifizierung des zum Untertitel gehörenden Sprechers einfacher und ein Nutzer sagte, jeder Sprecher bekomme so seine eigene Stimme. Die am Video fixierten Untertitel sind laut der Testpersonen auch besser zu lesen, da die Darstellung qualitativ besser ist. Letzteres ist zum Teil aber der Performance auf den Endgeräten geschuldet, die bei den am Bildschirm fixierten Untertiteln eher zu Flimmern neigt. Grund sind kleine Verzögerungen bei der Bereitstellung der Sensordaten der VR-Brille, wenn das Gerät an seine Leistungsgrenze stößt.

Insgesamt wurde auch festgestellt, dass die Art des Contents und insbesondere die Gesprächsgeschwindigkeit und die Verteilung der Personen in der Szene eine Rolle dabei spielt, welche Variante bevorzugt wird. Eine Testperson wünschte sich explizit einen gemischten Modus, der je nach Art des Inhalts die Untertitel am Bildschirm oder am Video fixiert.

Detaillierte Ergebnisse der durchgeführten Tests sind auf der Projektwebseite [ImAc] dokumentiert.

### Implementierungen und neue Metadaten für die 360°-Szene

Ein Web-Player, der die vorgeschlagenen Untertiteldarstellungen unterstützt, wurde vom Projektpartner i2CAT entwickelt und auf GitHub veröffentlicht [i2catPI]. Auch Editor-Tools für die Erstellung zusätzlicher Metadaten wurden im Projekt entwickelt.

Alle ImAc Entwicklungen basieren auf Web-Standards (DASH, WEB-GL, HTML5, IMSC), die für die Zusatzfunktionen



#### PETER THO PESCH

Peter tho Pesch ist aktuell für die IRT-Aktivitäten im Bereich Barrierefreiheit in immersiven Medien zuständig.

[www.irt.de](http://www.irt.de)

Quelle: IRT

erweitert wurden. Eine Integration in bestehende Web-Auspielwege ist gut möglich.

Für die im Projekt entwickelten Darstellungsoptionen müssen zusätzliche Metadaten zur Player-Applikation übertragen werden. Im Projekt wurde dafür das Untertitelformat „Internet Media Subtitles and Captions“ (IMSC) [IMSC] genutzt. IMSC ist ein (Unter-) Profil der Timed Text Markup Language (TTML) und lässt nutzerdefinierte Erweiterungen der Metadaten in gewissem Rahmen zu. So sind die im Projekt erstellten Untertiteldateien weiterhin standardkonform und lassen sich auch von anderen Playern wie bspw. dem VLC abspielen – dort jedoch bisher ohne die in ImAc entwickelten Features.

Das IRT bringt die Projektergebnisse in Standardisierungsgruppen des W3C oder der MPEG ein, so dass kompatible, einheitliche Lösungen entwickelt werden können. ➤

#### Referenzen und Links

[D5.4] Zusammenfassung der Nutzerergebnisse im ImAc-Projekt (Deliverable D5.4); <http://www.imac-project.eu/documentation/deliverables/>

[i2catPI] Github Projekt des von i2CAT entwickelten Web Players; <https://github.com/ua-i2cat/ImAc>

[ImAc] Webseite des ImAc Projektes; <http://www.imac-project.eu/>

[IMSC] W3C Recommendation IMSC 1.0.1; <https://www.w3.org/TR/ttml-imsc1.0.1/>

[IRTLab] Artikel und Demo zu am Sprecher fixierten Untertiteln; <https://lab.irt.de/subtitles-in-360-media/>

[Rothe] Sylvia Rothe, Kim Tran, Heinrich Hußmann; Dynamic Subtitles in Cinematic Virtual Reality, in International Conference in Interactive Experiences for Television and Online Video (TVX) 2018, Hanyang University, Seoul, Korea, June 2018

#### Danksagung

Weitere Infos, sowie einen Zugang zum Portal mit Videoinhalten, die in ImAc um barrierefreie Angebote ergänzt wurden, gibt es auf der Projektseite [www.imac-project.eu](http://www.imac-project.eu). Neuigkeiten aus dem Projekt erhalten Sie auch über Twitter (@ImAcProject).

Für dieses Projekt wurde im Rahmen der Finanzhilfvereinbarung Nr. 761974 Fördermittel aus dem Programm der Europäischen Union für Forschung und Innovation „Horizon 2020“ bereitgestellt.



## "DIGITAL MEDIA ONTOLOGY" IN ERSTER VERSION

Um relevante Inhalte im linearen Fernsehen und auf Video-on-Demand-Portalen zu finden, bedarf es qualitativ und quantitativ hochwertiger Beschreibungen der Inhalte, sogenannter Metadaten. Das Forschungsprojekt dwerft hat hierfür nun eine erste Version der „digital media ontology“ (dmo) fertiggestellt und veröffentlicht. Klassische Metadatenmodelle sind nicht dafür geeignet, den Komplexitätsgrad dessen, was inhaltlich innerhalb einer Szene geschieht, abzubilden – die Wissensmodelle aus dem Bereich der semantischen Technologien (Ontologien) können dies aber leisten. Die erste dmo-Version stellt das Herzstück der Technologieplattform dar und ist das erste Wissensmodell, das die gesamte mediale Wertschöpfungskette – von der Idee über die Produktion bis hin zum Archiv – abbildet. Die feingranulare Struktur dieses Wissensmodells erlaubt nicht nur präzise Suchen bis hin auf Szenenebene, sie bildet gleichzeitig auch maximale Interoperabilität zu anderen, existierenden Ontologien auf Element-Ebene ab.

Auf Basis der digital media ontology kann eine übersetzende Vernetzung stattfinden, die eine Vielzahl von in der Filmbranche eingesetzten Produktions- und Distributionstechnologieherstellern verstehen und einbinden kann. Mit diesen Herstellern möchte das Forschungsbündnis dwerft in den Austausch treten und in der Testphase einbinden, um die Verwertung von Filmen in der Multi-Plattform-Distribution deutlich zu verbessern und den Sendern und Produzenten die Interoperabilität und Schnittstellenunabhängigkeit zur Verfügung zu stellen.

Die dwerft forscht an der „Linked Media Data Cloud“ (LMDC), in der sämtliche Daten semantisch miteinander

in Beziehung gesetzt und dann für alle Beteiligten individuell und intuitiv nutzbar gemacht werden. Die Plattform soll allen Software-Lösungen in Film und TV eine transparente Möglichkeit von Geben und Nehmen bieten, statt einzelnen Anbindungen unterschiedlicher Software-Lösungen. Mit der Technologieplattform LMDC können Anwendungstools aus allen Bereichen der Wertschöpfungskette dynamisch vernetzt werden. Durch die verlustfreie Speicherung und intelligente Verbindung von Metadaten in einer semantisch gestützten Datenbank soll eine modulare Anpassungsfähigkeit an die Marktdynamik erreicht werden.

Neben den Redakteuren, die damit viel einfacher und umfassender passende Inhalte finden können, profitieren auch die Zuschauer. Mediatheken können auf der Basis der entwickelten Lösung mit neuen Suchmöglichkeiten ausgestattet werden – aufbauend auf den automatisiert gewonnenen Informationen, wie beispielsweise zu Drehorten oder zur Hintergrundmusik. Aber auch innerhalb der Wertschöpfungskette werden Arbeiten erleichtert, da die Weitergabe von Metadaten nicht mehr manuell, sondern automatisiert erfolgen kann. „Wir sind sehr froh gemeinsam mit den assoziierten Partnern den Erfahrungsaustausch bezüglich unserer Forschungsarbeit voranzutreiben. Außerdem bietet sich für unser Projekt eine sehr gute Möglichkeit, Lösungsansätze ganzheitlich und mit dem vernetzten Ansatz zu betrachten und in der Praxis zu testen“, sagt Mark Gülbahar, technisch-wissenschaftlicher Leiter der dwerft und Projektleiter beim Institut für Rundfunktechnik.

➤ [www.digitalmediaontology.org](http://www.digitalmediaontology.org)