

## Bildunterschriften zu Containerformaten

- **Folie 3:** Unkomprimiertes Ausgangsmaterial (z.B. aus einer Aufzeichnung) besteht meist aus getrennten Audio- und Videoteilen, die jeweils kodiert und anschließend zusammen mit Synchronisations- und anderen Informationen in einem Container abgelegt werden. Zum Abspielen werden die Audio- und Videodaten wieder aus dem Container gelesen und separat an die jeweiligen Decoder weitergereicht, um wieder unkomprimiertes Audio- bzw. Videomaterial für die Wiedergabe zu erhalten.
- **Folie 6:** Das ISO Base Media File Format (ISO 14496-12) ist stark an Apples QuickTime-Format angelehnt. Viele andere Formate sind wiederum vom ISO Base Media File Format abgeleitet, z.B. 3GPP und MP4, das in MPEG-4 Part 14 und ISO 14496-14 definiert ist.
- **Folie 8:** Eine MP4-Datei besteht hauptsächlich aus einer *moov*- und einer *mdat*-Box. Während letztere die eigentlichen Audio-, Video- oder sonstigen Multimediadaten speichert, enthält die *moov*-Box weitere Boxes mit der Beschreibung dieser Daten, z.B. Länge und Abtastrate. Diese Beschreibungen sind in mehreren *trak*-Boxes gespeichert, wobei eine Box pro Spur (z.B. Videodaten, deutsche Audiodaten, englische Audiodaten) verwendet wird.
- **Folie 9:** Audio-, Video- und sonstige Multimediaspuren müssen nicht in jener MP4-Datei gespeichert sein, die die Beschreibung (*trak*-Boxes) dieser Spuren enthalten. Stattdessen können die Daten in anderen MP4-Dateien liegen, auf die aus der *trak*-Box heraus verwiesen wird. In jedem Fall liegen die Multimediadaten selbst aber in einer *mdat*-Box einer MP4-Datei.
- **Folie 10:** Die *moov*-Box ist hierarchisch aus anderen Boxes aufgebaut, nämlich meist aus einer *mvhd*-Box, die wiederum aus mehreren *trak*-Boxes – einer pro Spur – besteht. Jede *trak*-Box verfügt über Informationen zur jeweiligen Spur in Form einer *tkhd*-Box, die wiederum eine *mdia*-Box mit Details zu den Daten der Spur enthält. Neben weiteren Boxes (und Verschachtelungsebenen) sind hierin Informationen über den Speicherort der eigentlich Spurdaten (*dinf*- bzw. *dref*-Box) und deren Eigenschaften, z.B. Offset (*stco*-Box) und Größe (*stsz*-Box) der Einzeldaten, Position von I-Bildern (*stss*-Box) sowie Dekodier- (*stts*-Box) und Darstellungszeitpunkte (*ctts*-Box) gespeichert.
- **Folie 13:** Die Differenz zwischen dem Darstellungszeitpunkt (CT) und dem Dekodierzeitpunkt (DT) eines Frames wird als Composition Offset bezeichnet. Die Differenz zwischen den Dekodierzeitpunkten aufeinanderfolgender Frames (in Dekodierreihenfolge) wird als Decode Delta bezeichnet. Alle Größen sind in `timescale`-Einheiten angegeben.
- **Folie 15:** Mehrere Samples eines Datenstroms werden zu einem Chunk zusammengefasst. Die Anzahl der Samples pro Chunk ist meist variabel

und kann bei großen Samples eins betragen. Samples sind im Fall von Audiodaten meist die kodierten (Stereo-)Abtastwerte eines Blockes oder im Fall von Videodaten die kodierten Pixeldaten eines Einzelbildes.

- **Folie 17:** Audio- und Videodaten sollten idealerweise perfekt synchron sein, d.h. der Darstellungsbeginn eines Videobildes passt exakt zum Wiedergabezeitpunkt der dazugehörigen Audiodaten (oben). Sind Audio- und Videospur konstant zueinander versetzt (Mitte), fällt die Asynchronität nur dann nicht auf, wenn sie hinreichend klein ist. In diesem Fall ist eine Korrektur durch den konstanten Versatz einer der beiden Spuren möglich. Laufen Audio- und Videospur allerdings mit der Zeit immer weiter auseinander (unten), ist eine vollständige Resynchronisierung erforderlich, bei der die tatsächlichen Geschwindigkeiten der beiden Spuren mit jenen der Spurzeitstempel verglichen und entsprechend angepasst bzw. neu berechnet werden.