Richtlinien

August 2009

Erläuterungen zum Bild/Ton-Versatz

Herausgeber: Institut für Rundfunktechnik

Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland
Ständiges ARD-Büro
Bertramstraße 8
60320 Frankfurt/Main

Telefon (069) 59 06 07
Telefax (069) 155 20 75
E-Mail: ard-buero@ard.de

Zweites Deutsches Fernsehen
ZDF-Straße 1
55127 Mainz-Lerchenberg

Telefon (06131) 70 0
Telefax (06131) 70 2157
E-Mail: info@zdf.de

Österreichischer Rundfunk
Würzburggasse 30
A - 1136 Wien

Telefon +43 1 87878-0
Telefax +43 1 87878 12738
E-Mail: online@orf.at

tv productioncenter zürich ag
Fernsehstrasse 1-4
CH - 8052 Zürich

Telefon +41 1 305 40 00
Telefax +41 1 305 40 10
E-Mail: info@tpcag.ch

ASF-Experten-Gruppe

Herr	Gierlinger	Friedrich	IRT	(Vorsitz)
Herr	Kurz	Walter	ZDF	
Herr	Lehmann	Hartmut	SWR	
Herr	Schiebener	Jörg	BR	
Herr	Metzger	Stefan	SWR	
Herr	Kaiser	Martin	RBT	
Frau	Wieland	Heike	SWR	(Zeitweise)

Das hier vorliegende Dokument wurde im Auftrag des Arbeitskreises Systemservice Fernsehen (ASF) von der Arbeitsgruppe „Audio/Videodelay“ erarbeitet. Es gibt den durch die ASF 2008 verabschiedeten Stand der Arbeiten wieder.

Schutzrechte-Hinweis

Herausgegeben im Auftrag der oben genannten Rundfunkanstalten vom:

Institut für Rundfunktechnik GmbH
Entwicklungsplanung/Öffentlichkeitsarbeit
Floriansmühlstrasse 60
80939 München

Telefon (089) 323 99 204
Telefax (089) 323 99 205
E-Mail: presse@irt.de
Homepage: www.irt.de

Inhaltsverzeichnis

1. Einleitung

Durch die Einführung der Digitaltechnik und neuer datenreduzierter Codierverfahren in der Fernsehtechnik entstehen durch die damit verbundenen Prozesslaufzeiten für Video und Audio wesentlich längere und vor allem unterschiedliche Laufzeiten. Werden diese Differenzen nicht innerhalb tolerierbarer Grenzen gehalten, kann die Qualität des Programmmaterials beim Zuschauer zum Teil stark darunter leiden.

Bei HDTV ist das Thema relevanter, da bei der Darstellung von Details auf großflächigen Displays aufgrund des Lupeneffektes die Laufzeitdifferenzen wahrnehmbarer werden.

Dieses Papier soll dazu dienen, die diesbezüglichen Ursachen, Zusammenhänge und Lösungsansätze aufzuzeigen. Es werden die einzelnen Prozesse und Phänomene in der Produktions- und Übertragungskette erläutert, und es wird auf Lösungsansätze verwiesen, auch wenn sie manchmal als trivial angesehen werden können. Wichtig ist, dass aufgezeigt wird, wie jeder Bearbeitungsschritt zum gesamten Bild/Ton-Versatz in der Kette beiträgt. Es wird gezeigt, dass Fehler an den Stellen korrigiert werden sollten, an denen sie auftreten. Die Ergebnisse dieser Studie sollten in den Produktionsbereichen, in der technischen Planung und auch in den Normungsgremien Berücksichtigung finden.

Bei Zeitangaben in Fields und Frames muss berücksichtigt werden, auf welches Fernsehsystem sich diese Angabe bezieht. Für progressive Systeme in Europa kann die Bildwiederholrate 50 Hz (720p/50 bzw. 1080p/50) oder 25 Hz (1080p/25) betragen. Systeme im Interlace-Verfahren arbeiten mit einer Bildwiederholrate von 25 Hz (576i/25 bzw. 1080i/25) und somit beträgt die Halbbildfrequenz 50 Hz.
(Bildwiederholrate 50 Hz entspricht 20 ms Framelänge
Bildwiederholrate 25 Hz entspricht 40 ms Framelänge)

Im Audiobereich angegebene Zeitwerte beziehen sich auf die Standard-Abtastfrequenz von 48 kHz.

2. Physikalische und physiologische Grundlagen

2.1 Schallgeschwindigkeit, Lichtgeschwindigkeit

Die Schallgeschwindigkeit beträgt in Luft 331.6 m/s (bei 0 ^{o}C und 1013 mbar Luftdruck). Für die weiteren Betrachtungen wird ein Wert von 330 m/s zugrunde gelegt. Somit der Schall in 40 ms (1 Vollbild) etwa 13 m zurücklegt.

Die durch die Lichtgeschwindigkeit (ca. 300 000 km/s im Vakuum und in Luft) bedingte Laufzeit kann in einem Gesichtsumfeld bzw. einer Szene (mehrere Meter bis einige Kilometer) gänzlich vernachlässigt werden.

2.2 „Empfinden" von Bild/Ton-Versatz

Nimmt ein Mensch ein optisch-akustisches Ereignis wahr, z. B. ein Blitz und der daraus resultierende Donner in 1000 m Entfernung, so tritt das akustische Ereignis etwa 3 s später ein. Hier ist schon zu erkennen, dass ein Versatz folgerichtig weder gesehen, noch gehört, sondern ***empfunden*** wird. Die Verbindung der optischen und akustischen Wahrnehmung findet erst bei der „Verarbeitung" beider Reize im Gehirn statt.

2.3 Empfindlichkeit des Menschen für Bild/Ton-Versatz

Das in 2.2 beschriebene Beispiel zeigt ein ganz natürliches, physikalisches Phänomen, bei dem das akustische dem optischen Ereignis nacheilt. Diese Tatsache ist aufgrund der Evolution des Menschen bzw. in unseren Erfahrungswerten als vollkommen natürlich zu betrachten. Wird aber das optische gegenüber dem akustischen Ereignis verzögert (hier: der

Donner kommt vor dem Blitz), so wirkt dies unnatürlich und wird als sehr störend empfunden.

Wie auch die EBU Technical Recommendation R37 „The relative timing of the sound and vision components of a television signal“ zeigt, ist die Sensibilität für “Ton vor Bild” größer. Laut dieser Recommendation nimmt die Hälfte der Zuschauer einen Versatz “Ton vor Bild” bereits ab 40 ms wahr, erkennt jedoch “Ton nach Bild” erst ab 60 ms.

2.4 Szenen- und interessenabhängige Sensibilität für Bild/Ton-Versatz

Wie störend ein Bild/Ton-Versatz empfunden wird, hängt auch vom Inhalt der Szene ab. Sprechende Menschen, gewisse Sportarten (z.B. Tennis) oder Tätigkeiten wie Holzhacken oder Schlagzeug Spielen werden sensibler wahrgenommen als fahrende Autos oder rauschende Bäche. Selbst bei sprechenden Menschen hängt es davon ab, ob es sich um die eigene Muttersprache oder um eine Fremdsprache handelt, beziehungsweise ob die Mund- und Lippenbewegungen der Sprache zugeordnet werden können.

Beeinflusst wird das Beurteilungsvermögen auch, wenn der Zuschauer Insider hinsichtlich des dargestellten Themas ist. So ist bei Musik ein Musiker, der z.B. selbst Geige oder Schlagzeug spielt, aufmerksamer als ein Laie, der zu der Musik bzw. dem Instrument keine oder kaum eine Beziehung hat.

Bei wachsendem Interesse und wachsender Begeisterung für den Inhalt eines Beitrages nimmt die Wahrnehmung für Qualitätsmängel wie Bild/Ton-Versatz allerdings ab.

Auch die Szenengestaltung spielt für die Sensibilität bezüglich Bild/Ton-Versatz eine Rolle. Eine „Totale“ wird unkritischer empfunden als eine Nahaufnahme mit Details.

Professionelle Mitarbeiter in der Produktion (Cutter, MAZ-Techniker, Toningenieure, ..), also Experten, sind laut einer Untersuchung [1] etwa doppelt so empfindlich beim Beurteilen des Bild/Ton-Versatzes als Laien.

Die Beispiele zeigen, wie schwierig es ist zu beurteilen, welche Toleranzen zulässig erscheinen. Weiterhin ist zu beachten, dass nicht die Grenze, ab wann ein Betrachter den Bild/Ton-Versatz bewusst erkennt, für die Definition einer Toleranzgrenze herangezogen werden kann, sondern dass schon viel früher ein gewisses “Unwohlsein” beim Betrachter entsteht, ohne dass ihm die Ursache bewusst wird; der Genuss des Programms ist jedoch dennoch eingeschränkt.

3. Beschreibung der Standards

3.1 Begriffsdefinition: Ton vor-/nacheilend bzw. +/-

Die Verzögerungszeit zwischen Bild und Ton (delay time) wird in der Literatur und bei den Richtlinien in Millisekunden angegeben. Dabei wird die Lage des Videosignals als Bezug genommen. Ein positiver Wert (+) bedeutet, dass das Audio-Signal dem Video-Signal voreilt, ein negativer Wert (-) steht für das Nacheilen des Audio-Signals. Im englischen Sprachraum wird der Bild/Ton-Versatz eindeutiger mit “sound advanced with reference to vision” und “sound delay with reference to vision” bezeichnet. Im deutschsprachigen Raum hat sich eine ähnliche Bezeichnung etabliert: “Bild vor Ton” bzw. “Ton vor Bild”.

3.2 Hinweis auf Richtlinien

In der EBU Recommendation R37(2007) „The relative timing of the sound and vision components of a television signal“ wird empfohlen, dass der Bild/Ton-Versatz am

Sendeübergabepunkt kleiner 40 ms (Ton vor Bild) bzw. kleiner 60 ms (Bild vor Ton) sein soll. Die empfohlenen Werte wurden aufgrund von subjektiven Tests ermittelt, bei denen 50 % der Teilnehmer die oben angegebenen Grenzen eines Bild/Ton-Versatzes erkennen konnten.

Außerdem wird seit der Ausgabe 2006 empfohlen, wenn immer es möglich ist, Vorkehrungen zu treffen, um den Bild/Tonversatz zu minimieren. Die Genauigkeit des Bild/Tonversatzes sollte an jedem Punkt innerhalb von 5 ms Ton vor Bild und 15 ms Ton nach Bild liegen.

In der ITU sind in den beiden Bereichen „Telecommunication“ und „Radiocommunication“ die unten aufgeführten Empfehlungen bezüglich des Bild/Ton- Versatzes festgelegt. Die ITU-T J.100 (Section Telecommunication) „Tolerances for transmission time differences between the vision and sound components of a television signal“ basiert auf der ehemaligen CMTT.717 (Study Programme 21A/CMTT). Diese bezieht sich auf den Report 1081 (Question 35/11). Auch der Report 412-4 (Study Programme 21A/CMTT) bezieht sich auf den Report 1081. Dieser Report sowie auch die derzeit gültige Recommendation ITU-T J.100 empfehlen einstimmig, dass bei jeder Verbindung, die für den internationalen Fernseh-signalaustausch verwendet wird, der Bild/Ton-Versatz maximal 20 ms (Ton vor Bild) bzw. 40 ms (Bild vor Ton) betragen darf. Die Recommendation ITU-R BT.1359 (Section Radiocommunication) beschreibt, bezugnehmend auf die Frage „Question ITU-R 35-4/11“, ein sehr umfangreiches Szenario, das im folgenden erläutert werden soll.

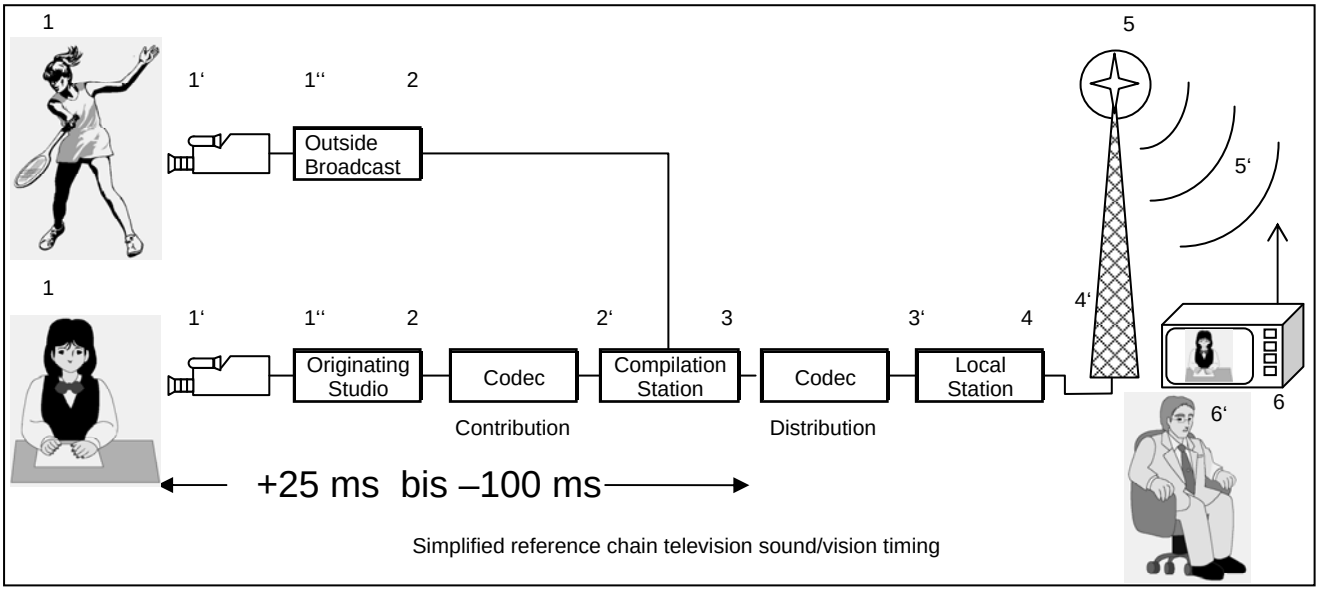

Bild 1: Figure1 der ITU-R BT.1359 für einen Referenzkanal

Die Empfehlung lautet wie folgt (aus dem Englischen übersetzt):

1. Die Referenz [(timing zero) Bild/Ton-Versatz 0 ms] für die nachfolgende Messung des relativen Bild/Ton-Versatzes ist mit dem Übergabepunkt Ausgang Compilation Station (Position 3 in Bild 1) definiert. In deutschen Rundfunkanstalten kann dies der Hauptschaltraumausgang sein.

2. Die gesamte Toleranz des Bild/Tonversatzes (zwischen Position 1‘ und 6‘) soll +90 ms bzw. –185 ms nicht überschreiten.

3. Die Zeittoleranz zwischen der Bildquelle (Position 1) und dem in Punkt 1 festgelegten Referenzpunkt darf die Grenzwerte +25 ms und –100 ms nicht überschreiten. (Anmerkung: Das ist der Bereich, in dem der Programmproduzent Kontrolle über den Bild/Ton-Versatz ausüben kann. Allgemein betrachtet ist es nicht möglich, eine korrekte oder beabsichtigte Bild/Ton-Beziehung festzustellen, erstens wegen des in Bild 2 gezeigten Undetectability plateau und zweitens wegen der vom Produzenten gewollten abweichenden Bild/Ton-Beziehung.)

4. Der Bild/Ton-Versatz-Unterschied zwischen dem in Punkt 1 definierten Referenzpunkt und dem Sendereingang sollte innerhalb von +22,5 ms und -30 ms liegen.

5. Wenn keine Korrektur des Bild/Tonversatzes durch den Programm-produzenten mehr möglich ist, soll jeder Abschnitt, der nicht der Kontrolle des Programmproduzenten unterliegt dem Signal nicht mehr als ±2 ms Bild/Ton-Versatz zufügen.

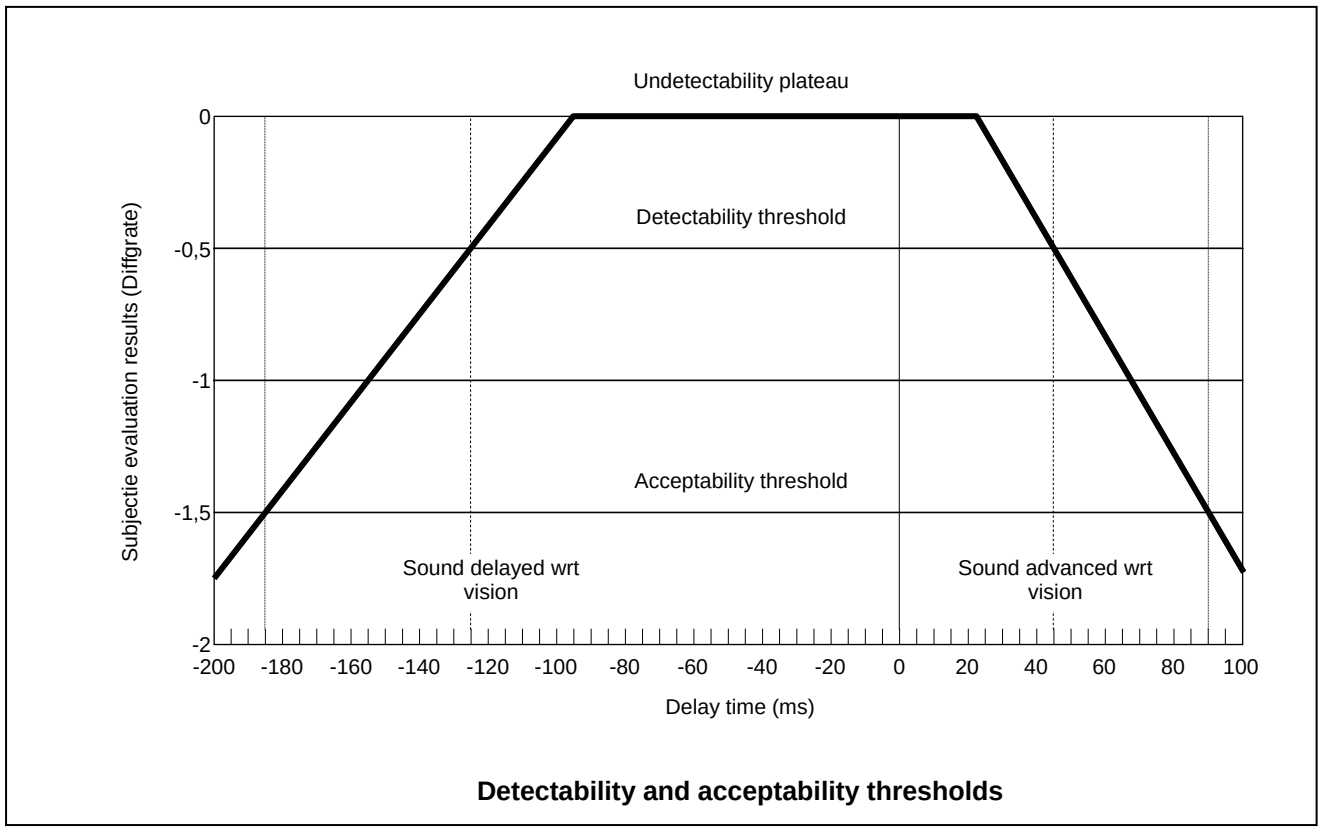

Bild 2: Figure 2 der ITU-R BT.1359 mit den Sichtbarkeits- und Akzeptierbarkeitsgrenzen

Im Anhang 1 der Recommendation ITU-R BT.1359 werden die empfohlenen Werte erklärt (aus dem Englischen übersetzt).

1. Es ist aus jahrelanger Erfahrung von der Filmprojektion her bekannt, dass der Bild/Ton-Versatz sehr wichtig ist und dass ein exakt bestimmbarer Punkt gefunden werden kann, ab dem der Versatz vom Betrachter als störend empfunden wird. Die ITU-R BR.265 zeigt, dass die Genauigkeit der Bild-/Ton-Beziehung innerhalb ± eines halben Bildes liegen soll. Beim Film bedeutet dies eine akzeptierbare Abweichung von ±22 ms.

2. Die verschiedenen Aufnahmetechniken erzeugen scheinbar unvermeidbare und unvorhersehbare Schwankungen des Bild/Tonversatzes von etwa einem halben Teilbild.

3. Subjektive Untersuchungen in Japan, in der Schweiz und in Australien zeigten eine hohe Übereinstimmung der Beurteilung des Bild/Ton-Versatzes von TV-Material für das NTSC- bzw. das PAL-System. Im Mittel wird ein Bild/Ton-Versatz ab +45 ms und ab –125 ms erkannt und bis +90 ms und –185 ms akzeptiert.

4. Der Bereich, in dem ein Bild/Ton-Versatz gerade noch wahrgenommen wird, kann zur Bestimmung einer Empfehlung nicht herangezogen werden, weil dieser Bereich in der Verantwortung der Programmmacher liegt. Außerdem ist kein empfohlener Weg bekannt, die präzise Bild/Ton-Beziehung zu ermitteln. Es kann die unbefriedigende Situation vorliegen, dass eine Bild/Ton-Beziehung gewählt wurde,

die nahe an einer Wahrnehmbarkeitsgrenze liegt. Damit steht dann nur noch ein sehr begrenzter Raum für zusätzliche Fehler zur Verfügung, bis der Bild/Ton-Versatz unakzeptabel wird.

5. Wegen des Bereiches, in dem Bild/Ton-Versätze nicht erkannt werden können, muss der erlaubte Bild/Tonversatz 0,5 Einstufungspunkte (bei einer 5-stufigen Skala) oberhalb der subjektiv ermittelten Wahrnehmbarkeitsgrenze liegen. In subjektiven Untersuchungen wurden 60 ms als eine Beeinträchtigungsstufe Bild vor Ton und 45 ms als eine Beeinträchtigungsstufe Ton vor Bild ermittelt. Dies führt zu den in Bild 2 gezeigten Flanken des Diagramms und davon abgeleitet auch zu den Werten der Nichterkennbarkeit des Bild/Ton-Versatzes. Demnach sind Fehler zulässig, die innerhalb der Grenzen von 95 ms Bild vor Ton und 22,5 ms Ton vor Bild liegen.

Die in der Recommendation ITU-R BT.1359 angegeben Werte wurden mit Nachrichtensprecherinnen in Japan, der Schweiz und in Australien ermittelt und betrachten ***nicht*** Szenen, die wesentlich kritischer sein können (z.B. Sport- oder Musikbeiträge).

Außerdem werden in diesem von uns vorgelegten Papier Wege gezeigt, wie die Bild/Ton-Beziehung ermittelt werden kann (siehe Kapitel 5). Aus diesem Grund ist für ein allgemein gültiges Umfeld die EBU-Festlegung mit ihren enger spezifizierten Werten zu Grunde zu legen.

4. Ursachen für Audio/Video-Delay

4.1 Allgemeines

In diesem Kapitel werden mögliche Ursachen aufgezeigt, die zu einem Audio/Video Delay führen können. Betrachtet wird dabei der gesamte Signalweg von der Aufnahme bis zum heimischen TV-Empfänger.

4.2 Abtastung - mechanische Klappe

Die zeitdiskrete Abtastung einer Szene führt bei elektronischen Kameras zu einem Abtastintervall von einem Halbbild gleich 20 ms und beim Film zu einem Abtastintervall von einem Bild gleich 40 ms. Dies bedeutet, dass ein abzubildendes Ereignis, das keinerlei Bezug zum Abtastsystem hat, einen zeitlichen Versatz bis zu 20 ms bzw. 40 ms aufweisen kann. Dieser Effekt führt bei Verwendung von „asynchronen“ Messeinrichtungen zur Erfassung des Bild/Ton-Versatzes (z.B. einer mechanischen Klappe), zu einer Messunsicherheit von 0 bis 20 ms (40 ms).

4.3 Produktionstechnik

4.3.1 Elektronische Kameras

Bei Messungen an Röhrenkameras bestätigt sich der in Kapitel 4.2 beschriebene Abtastvorgang.

Grundsätzlich tritt bei Kameras mit Halbleitersensoren eine Signalverarbeitungszeit von ca. 3 ms auf.

Bedingt durch den Bildaufbau und das Auslesen der Bildinformationen ergeben sich für den CCD-Sensor und den CMOS-Sensor unterschiedliche Laufzeiten durch das jeweilige Sensorprinzip.

Bei Kameras mit CCD-Sensoren kommt es durch das Auslesen des Bildspeichers zu einer zusätzlichen Verzögerung zwischen 0 und 20 ms (bei progressiven Verfahren mit einer Bildwiederholrate von 25 Hz bis zu 40 ms). Im Gegensatz dazu tritt diese bei Kameras mit CMOS-Sensor durch die direkte Adressierung der Bildpunkte nicht auf (Verhalten wie Röhrenkamera).

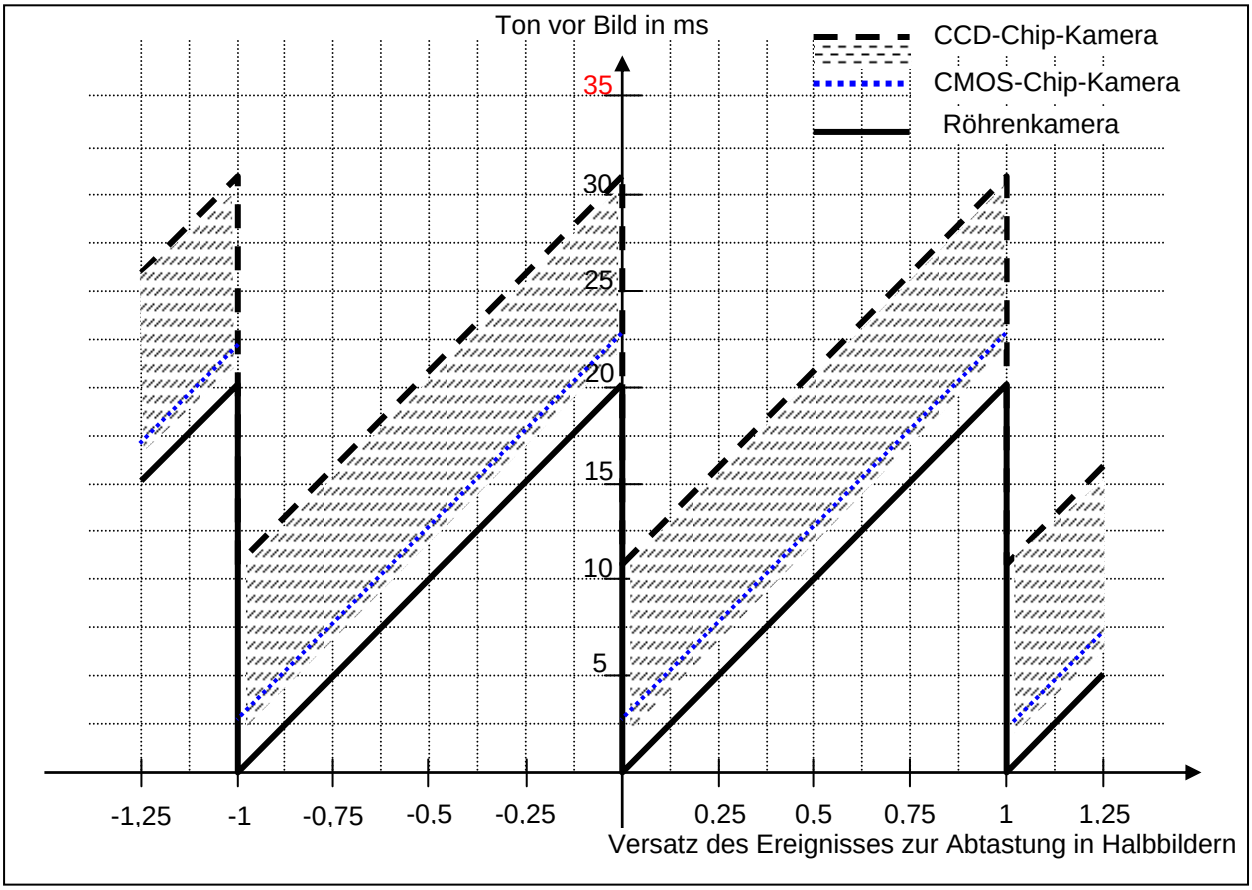

Bild 3: Systembedingter Bild/Ton-Versatz durch die elektronische Kameraabtastung

4.3.2 Drahtlose Kamerasysteme

Durch die notwendige Datenreduzierung des Videosignals zur digitalen drahtlosen Übertragung entsteht eine zusätzliche Signallaufzeit. Diese ist abhängig von dem benutzten Videocodierstandard und liegt meist zwischen 40 und 60 ms. Sollte zusätzlich eine Videosynchronizer-Funktion notwendig werden muss diese noch eingerechnet werden.

4.3.3 Position der Mikrofone zur Kamera

Durch die Schallgeschwindigkeit bedingt passt in den seltensten Fällen der Bildeindruck exakt zum Höreindruck. Wird zum Beispiel mit einer Kamera auf ein Objekt gezoomt, das sich in einer Entfernung von etwa 6,50 m befindet, hat der Ton gegenüber dem Bild schon eine Verzögerung von 20 ms, wenn das Mikrofon direkt neben der Kamera platziert ist. Will man dies vermeiden, müssten bei jedem Zoom und Schwenk die Mikrofone mitgeführt werden.

4.3.4 Playback-Problem

Der Akteur, der im Playback-Verfahren vor der Kamera steht, bekommt seine Information für die entsprechenden Lippenbewegungen von einem Monitorlautsprecher. Steht dieser Lautsprecher weit entfernt, verzögern sich die Bildbewegungen gegenüber dem Playback-Signal, das zum Zuschauer übertragen wird. Bei einer Entfernung von ca. 10 m eilt dann der Ton für den Fernsehzuschauer dem Bild um 30 ms voraus.

4.3.5 Videoprojektoren und Flachbildschirme

Die meisten Videoprojektoren (mit Techniken wie z. B. **DMD**, **D**igital **M**icromirror **D**evice oder **LCD**, **L**iquid **C**rystal **D**isplay, **ILA**, **I**mage **L**ight **A**mplifier) verfügen über einen Bildspeicher

für die Signalaufbereitung, der grundsätzlich eine konstante Laufzeitverzögerung des Bildes von mehr als 20 ms zur Folge hat.

Bei Displays, die für die Signalverarbeitung Vollbildspeicher benötigen, tritt üblicherweise eine Verzögerungszeit für die Videoinformation zwischen 20 ms und 100 ms auf. Die tatsächliche Signalverarbeitungszeit ist abhängig vom Format des Eingangssignals, den verwendeten Algorithmen zur Bilddarstellung, dem Darstellungsmodus (z.B. 100-Hz-Mode) und dem Displaytyp. Zur Zeit werden Displays angeboten, die zu Lasten der Bildqualität auf geringere Verarbeitungszeiten optimiert sind.

4.3.6 Multiplex-Darstellung

Für Monitorwände und Bearbeitungsplätze besteht die Möglichkeit, mit einem Monitor mehrere Videosignale im Multiplex darzustellen. Je nach Arbeitsweise und Auflösung kommen auch hier Bildspeicher zum Einsatz, die das Videosignal einschließlich der Verarbeitungszeit im Display zwischen 80 ms und 120 ms verzögern.

4.3.7 Virtuelles Studio

Die Produktionslandschaft bei den Rundfunkanstalten verändert sich zunehmend. Das verlangt eine höhere Auslastung der Studios, was wiederum Auswirkungen auf die Art der Herstellung von diversen Sendungen hat. So bietet sich bei Magazinsendungen eine neue Produktionsform an: die Herstellung im virtuellen Studio.

Das virtuelle Studio bietet den Vorteil, dass verschiedene Sendungen in kurzer Zeitfolge im Studio aufgezeichnet werden können bzw. dass aus diesem live gesendet werden kann. Dieser schnelle Wechsel der Produktionen wird dadurch erreicht, dass der jeweils aktuelle Hintergrund (Kulisse) im Computer in Echtzeit berechnet wird. Welcher Ausschnitt aus der Kulisse aktuell ist, wird dem Rechner durch die reale Studiokamera vorgegeben. Dazu werden die Bewegungsdaten der realen Kamera erfasst, digitalisiert und dem Computer zugeführt. In Abhängigkeit von diesen Daten wird dann das Bild der virtuellen Kamera erzeugt und für die momentane Szene zur Verfügung gestellt.

Für das Erfassen der Bewegungsdaten sind zur Zeit verschiedene Tracking-Verfahren bekannt. Allen Verfahren gemeinsam ist, dass sie zur Aufbereitung der Bewegungsdaten der Kamera Zeit benötigen. Je nach Verfahren treten Verzögerungszeiten zwischen 80 ms und 120 ms auf. Größere Prozessorleistungen ermöglichen geringere Verarbeitungszeiten.

Dem Moderator bereiten diese Bearbeitungszeiten große Schwierigkeiten, weil er seine Stimme in Echtzeit hört, sein Bildsignal aber erst um die Bearbeitungszeit verspätet sieht. Auch bei Einspielungen von Videobeiträgen in das virtuelle Studio oder bei Satelliten-Liveschaltungen etc. ist dieses Problem vorhanden und muss beachtet werden.

4.3.8 Browsing bei Redaktionssystemen

Mit der Einführung von Redaktionssystemen wird der PC auch zur Darstellung des Roh- bzw. Programmmaterials am Redaktionsplatz zur redaktionellen Beurteilung bzw. zur Erstellung einer Rohschnittliste genutzt. Aufgrund der eingeschränkten Übertragungskapazität der Netzwerke kommen hier datenreduzierte Formate wie z. B. Real-Player-Format, MPEG1, MPEG2 oder MPEG4 zum Einsatz. Es hängt nun vom Systemhersteller ab, ob Timestamps für die Verkoppelung von Audio und Video angewandt und in wie weit diese beim Decodiervorgang genutzt werden. Die Erfahrung zeigt, dass diesbezüglich mit einem Bild/Ton-Versatz zu rechnen ist und Kompromisse eingegangen werden müssen (vergl. auch Kapitel 4.6.3).

4.4 Signal-Processing

4.4.1 Videosynchronizer im Zusammenspiel mit Audiodelays

In der Videotechnik werden zur Anpassung von asynchronen Signalen Videosynchronizer verwendet. Synchronizer für digitale Komponentensignale benötigen aufgrund der 2-V-Periodizität 1 Vollbild (20 ms bzw. 40 ms) als Speichertiefe.

Da asynchrone Signale zueinander driften, ergibt sich keine konstante Verzögerungszeit, was zu Problemen bei der automatischen Nachsteuerung von Audiodelays (Audiosynchronizer ∗) führt. Kritisch ist dabei der Zeitpunkt, bei dem der Videosynchronizer seine Verzögerungszeit von der maximalen zur minimalen ändert, da sich dies beim Ton nicht störungsfrei nachbilden lässt (siehe auch 4.4.12).

Werden Videosynchronizer zur zeitlichen Anpassung von synchronen Signalen (zwei Standorte bzw. Ü-Wagen phasenstarr mit GPS-Synchronisation) eingesetzt, arbeiten sie immer mit einer konstanten Verzögerungszeit. Diese kann messtechnisch ermittelt und das Audiodelay dementsprechend mit dieser Verzögerungszeit fest eingestellt werden.

4.4.2 Digitale Video Effekt-Geräte (DVE) und Format-Converter 16:9 ⇔ 4:3

Digitale **V**ideo **E**ffekt-Geräte (DVE) basieren in der Regel auf einer Vollbildspeichertechnologie. Ein DVE wird innerhalb eines Studios verwendet, und somit liegen nur synchrone Videosignale am DVE-Eingang an. Die Ausgänge eines DVEs liegen dann wieder als Quelle am Bildmischer auf. Die Verzögerungszeit für das Videosignal beträgt unter diesen Voraussetzungen konstant 40 ms. Mischt, schaltet oder blendet man die Originalquelle und das im DVE manipulierte identische Bild ein, müsste auch hier der Ton entsprechend nachgesteuert werden. Eine solche automatisierte Audiodelay-Steuerung ist derzeit nur mit Einschränkungen realisierbar.

Formatkonverter (16:9 ⇔ 4:3) kann man als Spezialfall eines zweidimensionalen DVEs betrachten. Werden ganze Beiträge bei Sendung oder Postproduktion konvertiert, so muss der daraus resultierende Bild/Ton-Versatz korrigiert werden.

Für die Up/Down-Konvertierung, Framerate-Konvertierung und Scaling werden abhängig vom Fabrikat unterschiedliche Rechenzeiten benötigt, die im Audiosignalweg berücksichtigt werden müssen.

4.4.3 PAL-Encoder/-Decoder

Bei PAL-Decodern mit digitalem Processing sind für hochwertige Anforderungen Decodieralgoryhtmen mit zeitlicher Filterung (dreidimensional) implementiert. Auch eine Synchronizer-Funktion zur zeitlichen Anpassung der Videosignale bzw. zur Anpassung asynchroner Signale an eine Zeitebene je nach Systemanforderung kann optional zum Einsatz kommen (vergl. Kapitel 4.4.1). Es können geräteabhängig Videolaufzeiten üblicherweise bis zu 60 ms auftreten.

PAL-Encoder enthalten u. U. zur Laufzeitanpassung Line- bzw. Frame-Synchronizer. Hier können Laufzeiten bis zu 40 ms auftreten.

4.4.4 Video-File-Transfer

Beim Video-File-Transfer selbst kann systembedingt kein zusätzlicher Bild/Tonversatz entstehen.

Für die Erzeugung und Ausspielung von Video-Files ist ein entsprechender Time Stamp – Mechanismus zur Sicherstellung der Bild/Tonbeziehung vorgesehen. Bei der Ausspielung ist darf zu achten, dass die entsprechenden Time Stamps im Datenstrom ausgewertet werden.

∗Auf dem Markt werden Karten angeboten, die das Tondelay und einen Samplerate-Konverter (siehe Abschnitt 4.4.6) in sich vereinen und somit als Audiosynchronizer bezeichnet werden. Diese Karten interpolieren bei einer Änderung der Verzögerungszeit den Ton und vermeiden so das störende Knacken. Wie schnell die Nachregelung geschieht, ist von der Größe der aufzuholenden Verzögerung abhängig .

4.4.5 Videomischpult

In Bildmischern mit „zeilenorientierter“ Verarbeitung können optional Bildspeicher integriert werden, die sich dann auf die gesamte Verzögerungszeit des Gerätes auswirken. Außerdem sind auch digitale Bildmischer erhältlich, die über „bildbasierendes“ Processing verfügen.
Bildmischer selbst mit Autofaser verursachen eine vernachlässigbar kleine Verzögerung von wenigen Zeilen. Größere Verzögerungen können entstehen durch Up/Down-Konverter in den Ein- und Ausgängen sowie integrierten DVEs (siehe 4.4.2).

4.4.6 Audiomischpult

Digitale Mischpulte benötigen längere interne Verarbeitungszeiten als ihre analogen Vorgänger. Durch das Einschleifen von signalbeeinflussenden Elementen wie Filter oder dynamische Regeleinheiten wird jeweils ein Rechenprozess angestoßen, der bei älteren Systemen zu einer Erhöhung der Verarbeitungszeit von bis zu ca. 20 ms führt. Heutzutage wird eine konstante Durchlaufzeit, unabhängig vom Signalverarbeitungsprozess, von verschiedenen Herstellern realisiert. Deren typische Gesamtsystemdurchlaufzeit (d.h. analoger Eingang – Processing – analoger Ausgang) beträgt bis zu 3 ms.

4.4.7 Kreuzschiene

Die Latenzzeiten der Audiokreuzschiene ergeben sich aus der entsprechenden Konstellation d.h. analoger oder digitaler Ein- und Ausgang, Verwendung von Sample Rate Converter (SRC), MADI oder AES Verbindungen u.s.w. Die Laufzeitkonstellation bedingt durch die A/D- und D/A-Wandlung ergibt sich bei Nutzung des analogen Ein- und Ausgangs und liegt bei älteren Anlagen bei 3,3 ms, modernere Systeme benötigen zwischen 1 und 1,6 ms.

Moderne digitale Studiomischpulte sind häufig in direkter Einheit mit der Audiokreuzschiene aufgestellt. Im Gesamtsystem ergeben sich Gesamtzeiten von 1,5 bis zu 3,2 ms.

4.4.8 Abhöreinheiten

Neben den Mischpulten, die sowohl die Stereo- als auch die Mehrkanalabhörmöglichkeit bieten, werden für das Surround-Monitoring externe Controller angeboten. Die Laufzeit ohne Berücksichtigung von Dolby Decodierzeiten beträgt von 1,2 ms bis zu 3,4 ms.

4.4.9 Dolby – Mehrkanal

Die Toncodierungsformate Dolby Digital und Dolby E benötigen entsprechende Encodier – und Decodierzeiten. Diese Coder finden sich als “Standalone“ Geräte als auch in Mischpulten, Kreuzschienen und Abhörcontrollern wieder. Für diese Coderrechenzeiten ergeben sich folgende Werte:

Dolby E => ein Format zur Signalverteilung
Encodierung 40 ms
Decodierung 40 ms

Dolby Digital => ein Format zur Signalübertragung
Encodierung 187 ms
Decodierung 11 bis 15 ms

4.4.10 Abtastratenwandler (**S**ample **R**ate **C**onverter**, SRC)**

Der Sample Rate Converter wird als Synchronizer in der digitalen Audiowelt verwendet. Er wandelt zwischen Abtastfrequenzen von 32 kHz, 44,1 kHz, 48 kHz und 96 kHz. Außerdem

ist er in der Lage, zwei Signale grundsätzlich gleicher Abtastfrequenz, die aber asynchron sind, auf den „Mastertakt‘‘ hinzuziehen. Diese SRCs werden sowohl als Einzelgerät als auch in den verschiedensten Audiosystemen (z. B.: Kreuzschienen, Mischpulte) integriert bzw. verwendet.

Die typische Durchlaufzeit eines Sample Rate Converter beträgt bis zu 3,5 ms.

4.4.11 Durchlaufzeiten von digitalen Audio-Geräten

Nachfolgend sollen einige Geräte mit deren typischen Durchlaufzeiten aufgeführt werden. Leider fehlen diese Angaben häufig in den technischen Unterlagen. Es ist deshalb empfehlenswert, diese ggf. messtechnisch zu erfassen. Im Allgemeinen liegen diese Zeiten unter 2 ms und machen sich somit erst bei einer Kaskadierung von mehreren Geräten bemerkbar.

Analog/Digital-Wandler	0,4 ms – 1 ms
Digital/Analog-Wandler	0,7 ms – 1,5 ms
Grundlaufzeit von Delays (im D/D-Betrieb ∗)	bis zu 1 ms
Begrenzer, Transienten–Limiter (im D/D-Betrieb)	ca. 0,3 ms
Lautsprecher mit DSP	bis zu 80 ms
Digitale Mikrophone	0,5 ms – 1 ms

4.4.12 Embedded Audio

Die Übertragung des Audiosignals innerhalb des digitalen Videosignals (SDI/HDSDI bzw. SDTI/HDSDTI) wird wegen der damit verbundenen Vorteile (z. B.: gemeinsame Signalführung) immer häufiger eingesetzt. Das Argument gemeinsamer Weg und somit gleiche Laufzeiten gilt nicht in jedem Fall. Nachfolgend sollen vier Verarbeitungsprozesse beschrieben werden, die zusätzliche Laufzeiten für das Audiosignal zur Folge haben.

Zu beachten ist, dass bei Verarbeitung von Dolby E Signalen keine Bit beeinflussenden Geräte wie z.B.: SRC verwendet werden dürfen, die das Audiosignal definitiv unbrauchbar machen. Eine für diesen Fall mögliche Signalverarbeitung wird unter Punkt. 4 aufgeführt.

1. Embedden / Deembedden eines synchronen digitalen Audiosignals

   Die synchronen AES/EBU-Audiodaten müssen für die Einbindung als Ancillary-Daten in den seriellen Videodatenstrom aufbereitet werden. Je nach Hersteller benötigen die Embedder, auch Multiplexer oder Combiner genannt, für diese Funktion einige Millisekunden. Abhängig von der Lage der Audiodaten zur nächstmöglichen H-Lücke kommt noch eine weitere variable Laufzeit von bis zu 64 µs hinzu. Am Deembedder treten in Analogie ähnliche Laufzeiten für die Rückführung der embedded Audiosignale in ein AES/EBU-Signal auf.

2. Embedden / Deembedden eines asynchronen digitalen Audiosignals

   Um ein zu einem seriell digitalen Videosignal asynchrones AES/EBU-Signal zu embedden, muss das Audiosignal über einen Sample Rate Converter mit dem seriell digitalen Videosignal synchronisiert werden. Auf dem Markt sind Embedder erhältlich, welche die Sample Rate Converter–Funktion integriert haben (SMPTE 272M). Wird ein Embedder verwendet, der diese Funktion nicht enthält, muss ein externer Sample Rate Converter verwendet werden.

3. Synchronisation eines seriell-digitalen Videosignals mit einem embedded Audiosignal auf einen Studiotakt

   Soll ein asynchrones SDI-Signal direkt durch entsprechende Frame-Synchronizer dem Studiotakt angepasst werden, unterscheidet man zwischen drei Arbeitsweisen:

∗ A/A-Betrieb bedeutet analog Eingang/analog Ausgang; D/D-Betrieb bedeutet digital Eingang/digital Ausgang

Im ersten Fall werden Audio- und Videodaten ohne Demultiplexen nach dem „FIFO-Prinzip“ in einem gemeinsamen Speicher abgelegt und wieder ausgelesen. Wird dabei die Speicherkapazität überschritten bzw. unterschritten, muss ein Frame herausgenommen bzw. ein Frame aufgefüllt werden, was zu Tonstörungen führen kann. Einige Synchronizer bieten die Möglichkeit durch das automatische Herunterregeln und wieder Aufziehen des Audiosignals, solche Tonstörungen zu vermeiden. Die Audio/Video-Beziehung bleibt erhalten.

Im zweiten Fall werden die Audiodaten im Eingangs-Processing des Fram-Synchronizers deembedded, konstant um ca. 20 ms verzögert und am Ausgang wieder embedded. Je nach zeitlicher Lage des Eingangssignals zum Studiotakt ergeben sich unterschiedliche Audio/Video-Delays im Bereich von ± 20 ms.

Im dritten Fall wird das Audiosignal deembedded, über einen Sample Rate Converter einem Audiodelay zugeführt, das dynamisch die Laufzeit des Audiosignal an die Videoverzögerung anpasst. Damit bleibt auch in diesem Fall die Audio/Video-Beziehung erhalten.

4. Synchronisation eines seriell-digitalen Videosignals mit einem embedded DolbyE Audiosignal an einen Studiotakt

   Ein Dolby E Signal in einem seriell-digitalen Videosignal muss nach dem Deembedden aufgrund der unbedingt einzuhaltende Datentransparenz decodiert, synchronisiert, wieder Dolby E encodiert und erneut embedded werden. Das Videosignal muss um die dafür benötigte Zeitspanne von mindestens 80 ms verzögert werden.

   Frame Synchronizer mit dem sogenannten “Guardband cropping“-Verfahren ermöglichen eine Synchronisation des Dolby E Signals ohne zusätzliche Codierverfahren und damit ohne zusätzliche Latenzzeit. Geräte mit dieser Funktion gewährleisten automatisch eine korrekte Platzierung des Dolby E Rahmens im seriell-digitalen Videosignal.

4.4.13 Allgemeine Bemerkungen zu Überwachungsplätzen

Um das Programmmaterial, insbesondere bezüglich Audio/Video-Delay, richtig beurteilen zu können, müssen die Kontrollpunkte für das Audio- und das Videosignal jeweils an äquivalenten Stellen im Signalweg liegen. Beurteilungsplätze, die das Monitorsignal/Bild z.B. hinter einem Frame-Synchronizer und das Audio-Signal vor dem entsprechenden Ton-Delay abgreifen, können zu keiner zuverlässigen Aussage bezüglich des Bild/Ton-Versatzes kommen.

4.5 Übertragung

Bei Übertragungen ist nicht immer gewährleistet, dass das Video- und Audiosignal den gleichen Weg und somit die identische Laufzeit haben. So kann zum Beispiel der internationale Ton mit dem Videosignal über Satellit und der nationale Ton incl. Kommentar terrestrisch oder über See-Kabel geführt werden. Diese Laufzeitunterschiede müssen in jeden Fall ausgeglichen werden.

Die Kapitel 4.5.1 bis 4.5.6 sollen zeigen, mit welchen Laufzeiten bei den jeweiligen Übertragungswegen und Codieralgorythmen zu rechnen ist.

Selbst bei einer Zuführung der Audiosignale als embedded Audio werden die Bild- und Tonsignale im Coder mit Datenreduktion getrennt verarbeitet.

4.5.1 MPEG-2

Grundsätzlich stellt die Paketierung von MPEG2-Elementarströmen in den MPEG2-„Program Stream“ (Nutzung bei DVD) oder in den „Transport Stream“ (Nutzung bei DSNG und DVB) durch die Einfügung der „Timestamps“ selbst die Mechanismen für eine lippensynchrone Wiedergabe bereit. Korrekt konfigurierte Systeme weisen daher kein Audio/Video-Delay auf. Nach anfänglichen Schwierigkeiten beherrschen mittlerweile eigentlich alle Anbieter von Encodern oder Decodern diese Mechanismen.

Die Gesamtlaufzeit ist dabei durch die Video-Codierung und insbesondere die Pufferung im Encoder und im Decoder definiert. Sie errechnet sich wie folgt:

Laufzeit [s] = MPEG_Puffergröße [Bits] / Video_Netto_Bitrate [Bits/s]
+ (Anzahl_B-Frames + 1) * Bildperiode

mit MPEG_Puffergröße = 1.83 Mbit
Anzahl_B-Frames (zwischen I- oder P-Frames) meist 2 oder 3

Zusätzlich ist noch ein konstanter Offset von 1-2 Bildern für die eigentliche Verarbeitung im Encoder und im Decoder zu addieren. Typischerweise ergeben sich somit Laufzeiten von etwa 400 ms bei einer Übertragungsrate von 8 Mbit/s. Alle Parameter sind dem Encoder während des Codiervorgangs bekannt, sodass der Bild/Tonbezug erhalten bleibt.

Eine Verkürzung der Laufzeit (etwa für Reportage-Anwendungen) ist durch Verringerung des aktiv genutzten Puffers (wiederum ausschließlich durch Konfiguration des Encoders) - allerdings auf Kosten der Codier-Effizienz möglich. Diese Betriebsart wird heute von vielen Herstellern als sogenannter „Low-Delay-Modus“ (etwa 200 ms) unterstützt.

Nicht zu verwechseln ist dieser Modus mit einem von MPEG2 selbst definierten „Low-Delay-Modus“, der in der Regel zwar den gesamten Pufferbereich ausnutzt und damit sehr effizient arbeitet, jedoch empfängerseitig gelegentlich in Abhängigkeit vom Bildinhalt eine ruckweise Wiedergabe durch „skipped Frames“ aufweist. Diese Betriebsart wurde für den Einsatz im Computer-Bereich definiert und dürfte für Broadcast-Anwendungen wohl keine Rolle spielen.

4.5.2 MPEG-Audio Layer II

Die Zahl der Anwendungen von Datenreduktionsverfahren hat im professionellen Audiobereich in den letzten Jahren stark zugenommen. Auch im Konsumerbereich sind Systeme wie Minidisk, DAB, ADR und PC-Anwendungen etabliert. Man erreicht mit den Erkenntnissen aus der Psychoakustik, insbesondere durch Ausnutzung des Verdeckungseffektes, eine volle Audiobandbreite von 20 kHz bei einer im Vergleich zur linearen PCM-Codierung wesentlich geringeren Datenmenge. Benötigt man für ein nicht datenreduziertes 16-Bit-Stereosignal bei 48 kHz Abtastrate etwa 1,5 Mega Bit pro Sekunde (Mbps), so beträgt die Bitrate nach der MPEG-Audio-Codierung (MPEG-Layer-I bzw. MPEG-Layer-II) zwischen 8 kbps und 384 kbps. Die Codierung der Audiosignale erfolgt für den Broadcastbereich in der Regel nach dem ISO-MPEG-Layer-II-Format, ebenfalls bekannt unter dem Namen MUSICAM (**M**asking-pattern adapted **U**niversal **S**ubband **I**ntegrated **C**oding **A**nd **M**ultiplexing).

Im MPEG-Layer-II-Verfahren sind folgende Systemdurchlaufzeiten vorhanden: Für die Filterung werden ca. 11 ms, für die Framebildung1 aufgrund der Framelänge 24 ms und für die Übertragungstechnik 1 Granule gleich 2 ms benötigt. Außerdem muss, bevor der Decoder mit der Decodierung des MPEG-Signals beginnen kann, noch die komplette Kontrollinformation im Decoder vorhanden sein, was, abhängig von der Datenrate, zusätzlich ca. 5 ms Verarbeitungszeit zur Folge hat. Damit beträgt die vom MPEG-Layer-II-Codec verursachte theoretische Systemdurchlaufzeit mindestens 42 ms. Dabei weichen die in der Praxis zu erreichenden Zeiten von den theoretischen erheblich ab. Bei den zur Zeit vorhandenen Implementationen sind die Verzögerungszeiten durchschnittlich doppelt so hoch wie in der entsprechendenTheorie.

1 In diesem Fall wird unter „Frame“ nicht ein Videoframe (40 ms) verstanden, sondern eine Rahmenbildung für die MPEG-Layer II-Datenebene (24 ms).

Je nach benutztem MPEG-Audio Layer entstehen unterschiedliche Laufzeiten. Bei MPEG-Layer I[2] etwa 50 ms, beim Layer II[3] sind es knapp 100 ms und beim Layer III[4] etwa 200 ms. Wird ein nach MPEG codiertes Audiosignal im MPEG-System-Datenstrom übertragen, sorgen „Time Stamps“ (siehe Kapitel 4.5.4) für die korrekte Beziehung von Audio- und Videodaten. Da MPEG-Audio-Layer II auch eigenständig verwendet werden kann, müssen diese Zeiten berücksichtigt werden. Es ist offensichtlich, dass abhängig vom jeweils verwendeten Codierverfahren und dessen herstellerabhängiger Implementation nicht zu vernachlässigende Verzögerungen entstehen.

4.5.3 MEPG-4/AVC / H.264

Ein weiteres hocheffizientes Videokompressionsverfahren entstand aus dem Standard MPEG-4 Part 2 durch eine gemeinsame Weiterentwicklung der ITU und der ISO/IEC MPEG. Der Standard MPEG-4 Part 10 ist auch unter dem Namen AVC/H.264 bekannt. In der ITU findet man ihn unter ITU-T H.264. Später erweiterte das DVB Konsortium den ETSI Standard TS 101154 um den Part AVC/H.264. Er wird wegen seiner effektiven Komprimierung vorrangig für HDTV eingesetzt und ist eines der gängigen Videokompressionsverfahren für den Blu-ray Disc-Standard. Auch zur hochauflösenden Fernsehübertragung über DVB-S2 kommt H.264 zur Anwendung.

Die Bitratenreduktion gegenüber MPEG-2 kann bei vergleichbarer Qualität durchaus über 50 % betragen.

Je nach Applikation, Datenrate und Prozessrechenleistung können Codec-Laufzeiten von 200ms bis 850ms auftreten.

Die Verarbeitung des Audiosignals erfolgt wie bei MPEG 2 (Kapitel 4.5.1)

4.5.4 Rundfunkservice-Multiplexer

Bei der Übertragung von Video-/Audiodaten über ATM durch den Rundfunkservice-Multiplexer wird das Videosignal MPEG-codiert. Das Audiosignal wird entweder auch MPEG-codiert oder gemäß SMPTE 302M transparent direkt in den MPEG-Transportstrom eingefügt. MPEG stellt auch hier die Werkzeuge zur Verfügung, um die Bild/Tonbeziehung aufrecht zu erhalten. Die PAL-Ein- und Ausgänge werden durch vorgeschaltete PAL-Coder und –Decoder realisiert und sind bezüglich der Bild/Tonbeziehung gesondert zu behandeln (siehe Kapitel 4.4.3).

4.5.5 Video over IP / IPTV / Internet-TV

Bei der Übertragung von Inhalten über das Internetprotokoll unterscheidet man zurzeit drei verschiedene Varianten.

Video over IP ermöglicht den Austausch von Live Videodaten (Streaming, relativ niedrige Datenrate) z.B. im Rahmen von Chats. In der Regel wird dabei gleichzeitig Voice over IP zum Einsatz gebracht. Die Audio- und Videokommunikation erfolgt an einem PC-Arbeitsplatz.

Mit IPTV (Internet Protocol Television; deutsch: Internet-Protokoll-Fernsehen) wird die digitale Übertragung von breitbandigen Anwendungen in Realzeit, wie Fernsehprogrammen und Filmen, über ein digitales Datennetz bezeichnet. Hierzu wird das auch dem Internet zugrunde liegende Internet Protocol (IP) verwendet. Beim IPTV wird von einem Telekommunikations-Anbieter einem bestimmten Nutzerkreis – den Abonnenten – ein festes

2 MPEG-Layer I ist ein vereinfachtes MUSICAM mit niedriger Komplexität
3 MPEG-Layer II ist identisch mit MUSICAM, der Name MUSICAM wurde vor der Normierung in ISO/IEC verwendet
4 MPEG-Layer III ist eine Kombination von ASPEC – ein vom Fraunhofer Institut entwickeltes Verfahren – und MUSICAM

Programmbouquet mit definierter Qualität in seinem Breitbandnetz zur Verfügung gestellt. Die Ausgabe kann sowohl über PC als auch über eine Settop-box erfolgen.

Beim Internet-Fernsehen („TV over Internet“ auch WEB-TV) können beliebige Inhalte und Programme, die frei im Netz zugänglich sind, zu jeder Zeit und überall von Jedermann heruntergeladen werden (kostenpflichtig oder kostenfrei). Im Gegensatz zum IPTV hat das Internet-Fernsehen keine Qualitätsgarantie (Quality of service).

Diese Begriffe werden allerdings nicht einheitlich verwendet:

Sollten diese Übertragungsverfahren in irgendeiner Weise für Sendungen genutzt werden, so sind auch hier erhebliche Latenzzeiten zu erwarten (vergl. Kapitel 4.5.8).

4.5.6 Satellitenübertragung

Kommen geostationäre Satelliten in 37 000 km Höhe über dem Äquator für die Übertragung zum Einsatz, so ist ein einzelner Weg von überschlagsmäßig 40 000 km anzunehmen (der Satellit steht nicht senkrecht über den Linkstationen). Bei Lichtgeschwindigkeit (300 000 km pro Sekunde) und einem Gesamtweg von 80 000 km braucht das Signal 267 ms Streckenlaufzeit.

4.5.7 Tonübertragung via ISDN (Integrated Services Digital Network)

Das Reduktionsverfahren nach ISO MPEG ermöglicht eine qualitativ gute Übertragung des Audiosignals über ISDN-Verbindungen. In Europa sind das meist digitale Strecken mit n x 64 kbit/s, in Nordamerika teilweise mit n x 56 kbit/s. Für den Verbindungsaufbau muss sichergestellt sein, dass der Datenübertragungsdienst frei geschaltet ist und dass das verwendete Datenformat und -protokoll (1TR6 oder DSS1) übereinstimmt.

Zur Anbindung der Studiosignale werden sogenannte Codecs eingesetzt. Hier kommt der Datendienst (Serviceindikator 7) zum Einsatz. Diese Datendienste werden teilweise schon jetzt nicht mehr von den Providern unterstützt bzw. ist die Abkündigung absehbar. Ersetzt werden diese Dienste durch „Audio over IP“ (vergl. Kapitel 4.5.9). Wie schon in Kapitel 4.5.4 beschrieben, kommt es hier zu teilweise erheblichen Verzögerungen des Audiosignals. Benutzt man zum Beispiel als Kommentarleitung den Weg über ISDN-Verbindung, müssen die Signale zeitlich angepasst werden. Unter Umständen muss das Videosignal sogar verzögert werden.

Die Laufzeiten des ISDN-Netzes selbst sind im Vergleich zu denen der Codecs vernachlässigbar, solange keine Satellitenstrecken mit beteiligt sind. Die vom Codec selbst verursachten Signallaufzeiten liegen, wie in Kapitel 4.5.2 beschrieben, zwischen 50 ms und 200 ms. Für den Multiplexvorgang, der hochbitratige Signale auf mehrere 64-kBit/s-Kanäle (sogenannte B-Kanäle) verteilt und nach der Übertragung wieder zusammenbindet, werden abhängig von der Implementation und der Anzahl der B-Kanäle eventuell zusätzlich geringfügige Zeiten benötigt (herstellerabhängig ist dieser Zuwachs vernachlässigbar oder kann bis zu 450 ms Gesamtlaufzeit führen). Werden während der Übertragung vom Provider einzelne B-Kanäle umgeroutet, entstehen zwar keine nennenswerte Laufzeitänderungen, jedoch kann nicht gewährleistet werden, dass die Übertragung störungsfrei weiterläuft.[5]

Die ISDN-Technik wird zunehmend bis zur endgültigen Abkündung durch IP-Übertragung (vergl. 4.5.8 und 4.5.9) durch die Provider als Übergangsphase ersetzt. Obwohl die „Frontends“ immer noch ISDN-konform sind (2 B-Kanäle) werden oft schon im Backbone-Bereich der Provider die Verbindungen als IP-Verbindung behandelt. In diesem Fall wird durch die Gateways in der Nutzungsart Telefonie ein geringes Delay dazu kommen. Die Übertragung über einen Audiocodec ist dann aber wegen des fehlenden Datenkanals nicht

[5] In der ITU-T J.52 wird ein Verfahren zur Übertragung von hochqualitativen digitalen Tonsignalen über 1 bis 6 B-Kanäle beschrieben. Mit dem dort beschriebenen Synchronisationsverfahren können die unvermeidbaren Störungen bei den oben erwähnten Umroutvorgängen minimiert werden.

mehr möglich. Die Laufzeiten über IP-Netze können variieren, je nachdem welche Qualitätsstufe (QoS, Jitter), Provider und welche Entfernung zu überbrücken ist. Es ist mit Laufzeiten in Größenordnungen von 5ms bis 100ms zu rechnen.

4.5.8 Voice over IP (VoIP), IP-Telefonie

Unter der IP-Telefonie, eine Kurzform für die Internet-Protokoll-Telefonie, auch Internet-Telefonie oder Voice over IP (kurz VoIP) genannt, versteht man das Telefonieren über Datennetze, welche nach Internet-Standards aufgebaut sind. Dabei werden für Telefonie typische Informationen, d. h. Sprache und Steuerinformationen beispielsweise für den Verbindungsaufbau, über ein auch für Datenübertragung nutzbares Netz übertragen. Bei den Gesprächsteilnehmern können sowohl Computer, auf IP-Telefonie spezialisierte Telefonendgeräte, als auch über spezielle Adapter angeschlossene klassische Telefone die Verbindung ins Telefonnetz herstellen.

Neben der erforderlichen Übertragungskapazität haben Qualitätsmerkmale wie Verzögerung, Schwankungen in der Übertragung (Jitter) und Paketverlustrate erheblichen Einfluss auf die resultierende Sprachqualität und Laufzeit. Durch Priorisierung und geeignete Netzplanung ist es möglich, eine mit der herkömmlichen Telefonie vergleichbare Sprachqualität und Zuverlässigkeit zu realisieren.

Der Transport von Daten benötigt eine gewisse Laufzeit. Diese ist bei herkömmlicher Telefonie im wesentlichen die Summe der Signallaufzeiten auf den Übertragungskanälen. Bei Telefonie über IP-Netze kommen weitere Verzögerungen durch die Paketierung und Zwischenspeicherung sowie gegebenenfalls Kompression und Dekompression der Daten hinzu. Bei der Telefonie stellen, unabhängig von der verwendeten Technik, gemäß ITU-T Empfehlung G.114 bis 400 ms Laufzeit in einer Richtung die Grenze dar, bis zu der die Qualität von Kommunikation in Echtzeit noch als akzeptabel gilt. Ab etwa 125 ms kann die Laufzeit vom Menschen jedoch schon als störend wahrgenommen werden. Daher empfiehlt die ITU-T bei hoch-interaktiven Kommunikationsformen generell eine Laufzeit von 150 ms nicht zu überschreiten.

Um Datenjitter (unterschiedliche Laufzeiten von Datenpaketen) zu kompensieren, werden so genannte „Pufferspeicher“ eingesetzt, die eine zusätzliche Verzögerung der empfangenen Daten bewirken, um die Daten isochron auszugeben. Pakete deren zeitlicher Abstand größer ist als der gewählte Pufferspeicher, können nicht mehr in den Ausgabedatenstrom eingearbeitet werden. Die Größe des Pufferspeichers (in ms) addiert sich zur Gesamtlaufzeit und erlaubt die Wahl zwischen größerer Verzögerung oder höherer Paketverlustrate.

Werden solche VoiP-Verbindungen z.B. für die Kommentare oder n-1-Anwendung (siehe auch 4.5.10) eingesetzt, so sind sehr lange und variable Laufzeiten in Größenordnungen von bis zu 400 ms zu berücksichtigen.

4.5.9 Audio over IP (AoIP)

Im Gegensatz zu VoIP gibt es bei AoIP, das meist in geschlossenen Netzen eingesetzt wird, unter anderem garantierte Eigenschaften (QoS) des Dienstes. In der EBU Tech.-Doc. 3326 „Audio contribution over IP“ (siehe Bild 4) sind u.a. unterschiedliche Codierverfahren nach den Kriterien Reduzierung der Datenrate, unterschiedliche Übertragungsqualitäten oder geringe Laufzeiten zur professionellen Übertragung von analogen bzw. digitalen (AES/EBU) Audiosignalen über IP-Netzen beschrieben.

Für eine hohe Qualität bei niedriger Laufzeit ist eine PCM-linear-Codierung (Low–Delay/High–Quality Codierung) vorgesehen. Die Codec-Laufzeiten können zwischen 2 ms bis 80 ms betragen. Die zusätzlichen Laufzeiten über IP-Netze können extrem variieren, je nachdem welche Qualitätsstufe (QoS), Übertragungsart (UMTS, DSL, BGAN, ...), Provider und welche Entfernung zu überbrücken ist. Es ist mit Gesamtlaufzeiten von 10 ms bis 200 ms, in Extremfällen bis in den Sekundenbereich zu rechnen.

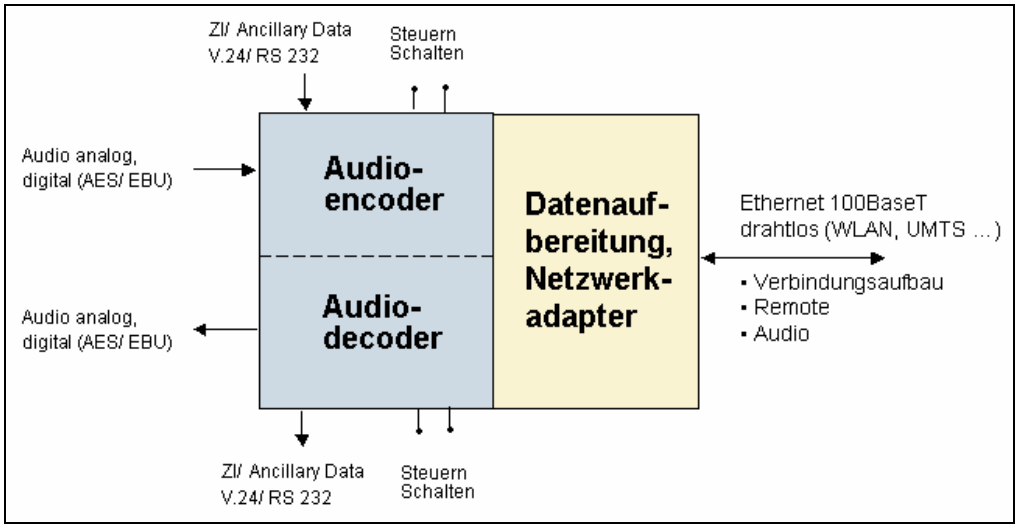

Bild 4: Aufbau von Audio over IP-Codecs

4.5.10 "Konferenzschaltung" / n-1-Signal

Der Signalweg, bei dem externe Signale über Satellit zugeführt werden, findet heute immer häufiger Anwendung, muss aber aufgrund der nachfolgenden Signallaufzeitproblematik besonders berücksichtigt werden.

Die einfache Signallaufzeit über einen Satelliten von der Außenstelle zum Studio beträgt je nach Orbitposition ca. 270 ms (s. Beispiel Kapitel 4.5.6). Zusätzlich müssen beim Einsatz von datenreduzierten Audiocodecs deren Laufzeit berücksichtigt werden. Der Zuschauer einer Konferenzsituation hört zum Zeitpunkt 0 die Frage des Moderators an die Außenstelle. Das Übermitteln dieser Frage über die Satellitenstrecke dauert, wie erwähnt, ca. ¼ Sekunde. Das Audio/Video-Signal der Antwort benötigt ebenfalls ¼ Sekunde für den Rückweg. Für den Zuschauer entsteht damit der Eindruck, dass die Außenstelle ½ Sekunde später reagiert. Beim Einspielen der Frage an der Außenstelle durch einen Lautsprecher gelangt auch dieses Signal an das Mikrofon der Außenstelle und ist somit für den Zuschauer als Echo der Frage hörbar. Diese Problematik kann man an der Außenstelle durch Verwendung von Ohrhörern umgehen.

Neben diesen Effekten, die der Zuschauer wahrnimmt, sind Signallaufzeiten auch für den Kommentator problematisch. Nimmt z. B. ein Kommentator seine eigene Stimme um mehr als 10 ms verzögert auf dem Kopfhörer wahr, kann er dadurch irritiert werden.

4.5.11 Simulcast-Übertragung Rundfunk/Fernsehen

Im Gegensatz zu früher, als es nur die analoge terrestrische Übertragung gab, sind die Übertragungsmöglichkeiten wesentlich vielfältiger geworden. Wird z. B. bei Konzerten der Ton im Hörfunk, Fernsehen oder Online gleichzeitig (simulcast) übertragen ist ein synchrones Eintreffen von Bild- und Tonsignal nicht möglich. So beträgt z. B. die Laufzeit bei einer Livesendung über DVB-T bis zu 6 Sekunden.
Eine Simulcast-Übertragung sollte von den Rundfunkveranstaltern nicht mehr beworben werden, da sie keinerlei Möglichkeiten haben, dem Zuschauer eine Synchronität bei paralleler Ausstrahlung zu gewährleisten.

4.6 Endgeräte

4.6.1 100-Hz-Fernsehempfänger

Kommt bei Heimfernsehgeräten die 100-Hz-Technik zum Einsatz, so ist mit den gleichen Verzögerungszeiten wie in Kapitel 4.3.4 beschrieben, also mit bis zu 20 ms zu rechnen.
Mittlerweile bieten die Geräte vermehrt die Möglichkeit das Audiosignal zeitlich an das Videosignal anzupassen.

4.6.2 Projektoren und Flachbildschirme

Für Projektoren und Flachbildschirme im Consumer-Bereich gelten die gleichen Bedingungen wie für die Broadcast-Geräte.
(siehe Kapitel 4.3.5)

4.6.3 Set-Top-Box (STB)

Die EBU Tech Doc 3311 mit dem Titel „Guidelines for Multichannel Audio in DVB“ beschreibt die Audio/Video Synchronisation. Das DVB-System nützt den Time-Stamp-Mechanismus für die Synchronisation zwischen Audio und Video des gesendeten Signales des MPEG-Transportstromes. Das Timing zwischen Encoder und Decoder soll innerhalb 1ms gehalten werden.

Weiterhin sollen die Audiosignale (z.B. unterschiedliche Sprachen, DolbyDigital) synchron mit dem Videosignal (für das gleiche Fernsehprogramm) an allen Kontrollpunkten des Übertragungskanals und die Audio/Video Synchronisation in der Set-Top-Box innerhalb einer Toleranz von -5ms (Audio früher) und +15ms (Audio später) zwischen den decodierten Video- und Audioausgängen liegen und sollen nicht driften.

Zu beachten ist hier, dass in dieser Norm voreilendes Audio mit minus, nacheilendes Audio mit plus gekennzeichnet wurden. Dies ist unterschiedlich zu den anderen relevanten Normen (vergl. Kapitel 3.1 und 3.2).

In der Praxis implementieren die verschiedenen Hersteller den Time-Stamp-Meachnismus auf unterschiedlichste Weise. Es gibt Set-Top-Boxen die den Zeitabgleich nicht kontinuierlich, sondern nur in größeren Zeitabständen durchführen. Dadurch ist ein Driften des Bild-/Tonversatzes außerhalb der Toleranzgrenzen möglich.

4.6.4 HDMI-Schnittstelle bei Endgeräten

Der HDMI 1.3a Standard enthält ein Verfahren, das Lippensynchronität auf Protokollebene ermöglicht. Damit kann der Bild-Tonbezug für getrennte Geräte für Ton- und Bildwiedergabe korrigiert werden. Dies gilt nur für die internen Signallaufzeiten der Geräte und beinnhaltet nicht den im Signal ggf. enthaltenen Bild-Tonversatz.

4.6.5 Fernsehempfang mittels PC

Werden PCs für multimediale Darstellung von Programmmaterial genutzt, besteht eine Vielfalt von Wiedergabemöglichkeiten. Die Prozesslaufzeiten der TV-Karten können sehr unterschiedlich sein. Die Audioinformation wird i.d.R. über den analogen Line-Eingang der Soundkarte ohne Verzögerung zugeführt.

Werden Programme mit Hilfe einer Capture-Software gespeichert, so ist entscheidend, in welchem Format (z. B. AVI, JPEG, Real-Player-Format oder MPEG bzw. CDI) abgespeichert wird und ob ggf. „Time-Stamps“ vom System eingefügt und angewandt werden. Weiterhin hängt es von der Leistungsfähigkeit des Prozessors, der Grafikkarte bzw. des Hardwaredecoders und von der Eigenschaft der Software des Medienplayers ab, mit welchem Bild/Ton-Versatz das Material wiedergegeben wird. Die Erfahrung zeigt, dass bezüglich Bild/Ton-Versatz der optischen Wiedergabe (Auflösung, Bildwechselfrequenz) große Kompromisse eingegangen werden müssen und dass dies nur eingeschränkt mit einem FS-Gerät verglichen werden kann. Dies gilt auch für das Abspielen von CDI-Material.

Durch die immer größere Leistungsfähigkeit des Internet ist es inzwischen möglich auch Programmsequenzen bzw. Programme direkt abzuspielen (z. B. Tagesschaubeiträge). Dies entspricht der Wiedergabe mit einem Medienplayer und ist somit u.U. mit Abstrichen bezüglich der Qualität und dem Bild/Ton –Versatz möglich.

5. Möglichkeiten der messtechnischen Erfassung der Laufzeitdifferenzen zwischen Audio und Video

Zur Messung des Bild/Ton-Versatzes außerhalb des Programms (offline) stehen spezielle Video- und Audiosignale mit entsprechenden Messverfahren zur Verfügung. Eine parallel zum Programm (online) durchzuführende kontinuierliche Messung mit ggf. automatischer Korrektur kann heute schon realisiert werden.

5.1 Online-Messung des Bild/Ton-Versatzes

5.1.1 Korrelationsverfahren

Das Korrelationsverfahren wurde in Zusammenarbeit vom IRT mit der Fachhochschule Würzburg-Schweinfurt entwickelt.

Für dieses Verfahren wird an der Signalquelle aus dem Tonsignal ein charakteristisches Signal mit möglichst niedriger Bitrate abgeleitet. Diese Information wird unter Ausnutzung der redundanten Bereiche entweder im aktiven Bereich oder (auch theoretisch) im Austastbereich des Bildsignals transportiert und erfährt so die gleiche Verzögerung wie das Bildsignal selbst. Auf der Empfangsseite wird dann die Laufzeitdifferenz zwischen Bild- und Toninformation mit Hilfe der charakteristischen Information bestimmt, die einerseits an der Quelle erzeugt und im Videosignal mit übertragen wird und andererseits empfangsseitig aus dem übertragenen Tonsignal in gleicher Weise generiert wird. Die für den Vergleich geeignete Maßnahme ist die Korrelation.

Das Verfahren wurde für die Übertragung von Bild- und Tonsignalen in analoger Umgebung entwickelt und bisher nicht in die digitale Welt transferiert. Es ist sehr wohl denkbar, dieses Verfahren auch auf der digitalen Ebene anzuwenden. Da das Tonsignal bereits in digitaler Form vorliegt, müsste hier nur noch ein für das Tonsignal charakteristisches Signal, z. B. das MSB, als Datensignal zusammen mit den Bilddaten übertragen und auf der Empfangsseite mittels Korrelation mit dem MSB des Tonsignals verglichen werden.

5.1.2 Watermarking-Verfahren

Die Bezeichnung „Watermarking“ (Wasserzeichen) bezieht sich auf die "versteckte" Übertragung von Daten mit relativ geringer Bitrate innerhalb des aktiven Bildbereichs eines Videosignals. Hier erfolgt diese Codierung durch unterschiedliche, vom Auge "tolerierte" Rauschmuster. Üblicherweise werden Daten in der Austastlücke eines Videosignals übertragen, können dabei aber durch Austastvorgänge bzw. Sync-Regeneration etc. verloren gehen. Für bestimmte Anwendungen - z.B. zur Herkunftskennzeichnung, aber auch für Anwendungen zum Aufrechterhalten der korrekten zeitlichen Bild/Ton-Zuordnung - muss demgegenüber ein zwangsweises Verbleiben der Daten beim Videosignal gewährleistet sein.

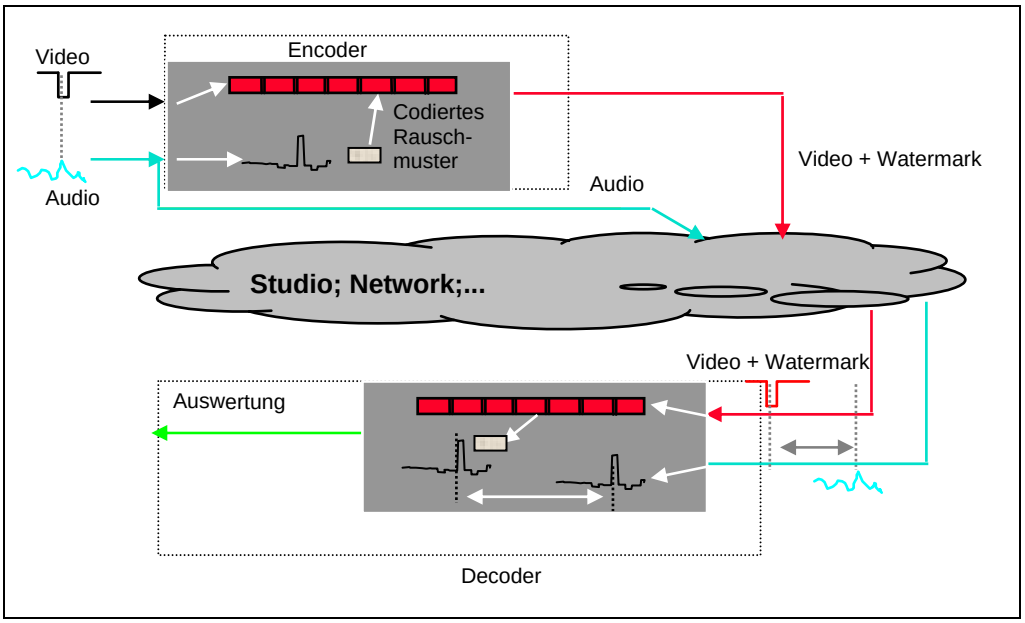

Bild 5: Blockschaltbild eines Bild/Ton-Versatz-„Watermarking“ Verfahrens

Das ankommende Videosignal und die zugehörigen Audiosignale werden einem Encoder zugeführt. Vom Audiosignal wird ein vereinfachtes, quantisiertes, den aktuellen Tonverlauf "charakterisierendes" Signal abgeleitet, dessen verschiedenen Quantisierungsstufen unterschiedliche Rauschmuster zugeordnet werden. Diese Rauschmuster werden dem Bild so überlagert, dass sie unsichtbar bleiben.

Am Ende einer Übertragungskette werden sowohl das mit dem Wasserzeichen versehene, übertragene Videosignal als auch das übertragene Audiosignal einem Decoder zugeführt. Der Decoder leitet von den übertragenen Audiosignalen in gleicher Weise wie beim Encoder wieder das, den Tonverlauf charakterisierende Signal, ab. Auch aus dem übertragenen Videosignal wird das Wasserzeichen extrahiert und decodiert. Aus der Korrelation dieser beiden "charakterisierenden" Signale kann der Audio/Videoversatz festgestellt und ggf. korrigiert werden.

Wasserzeichenverfahren, die in aufeinanderfolgenden Halb- oder Vollbildern das Wasserzeichen gegenphasig zufügen, können unter Umständen nachfolgende Datenreduktionscoder in andere Betriebarten (Field/Frame-modus) zwingen, was zu höheren Bandbreitenbedarf führen kann.

5.1.3 Video-/Audioanalyser (QuMax2000, K-Will Corporation)

Dieses Gerät analysiert Video- und Audiosignale in Echtzeit und erkennt Fehler wie z.B. Signalausfälle, eingefrorene Videobilder oder Tonaussetzer. An zwei oder mehr Stellen der Produktions- bzw. Übertragungskette werden für das Videosignal und das Audiosignal markante Information erzeugt. Aus dem Vergleich der über IP übertragen markanten Informationen von zwei Stellen kann der Bild-/Tonversatz ermittelt werden.

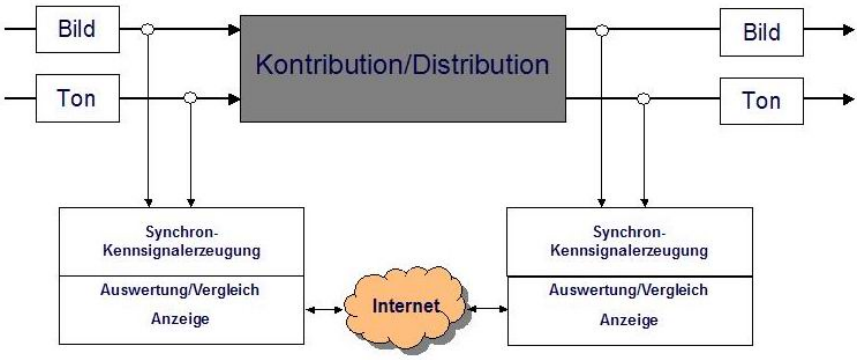

Bild 6: Prinzip der Vergleichsmessung

Das System muss für die Analyse keine Manipulationen am Videomaterial durchführen, arbeitet in Echtzeit und ist somit ein Onlineverfahren.

Das Messverfahren kann nur einen zusätzlichen Bild/Tonversatz zwischen den zwei Messpunkten ermitteln und eignet sich nicht für Standbilder, Testtöne und Stille.

5.1.4 Verfahren mit Bild- und Sprachanalyse (LipTracker™, Pixel Instruments Corporation)

Dieses Verfahren analysiert die Lippenbewegungen und das Audiosignal einer sprechenden Person und berechnet daraus den momentanen Bild-/Tonversatz. Voraussetzung hierfür ist allerdings, dass die sprechende Person direkt in eine Kamera blicken muss, damit die Lippenbewegungen von dem Gerät erfasst und analysiert werden können.

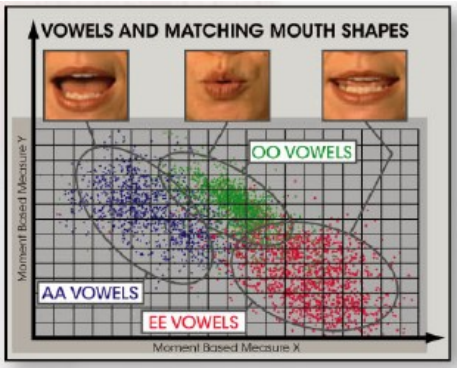

Bild 7: Zuordnung von Vokalen zur Lippenbewegung

Der große Vorteil besteht darin, dass hierfür keine Zusatzsignale mit übertragen werden müssen, und es an jeder Stelle der Produktions- und Übertragungskette eingesetzt werden kann. Es handelt sich also um ein „Online“-Verfahren.

Die generelle Funktionsweise lässt sich mit Hilfe von Bild 6 vereinfacht aufzeigen. Zu Beginn der Messung versucht das System zuerst ein Gesicht, dann den Mund der sprechenden Person zu erfassen. Im folgenden Messverlauf achtet es besonders auf die Mundbewegungen, die durch die Aussprache von Lauten wie, „A“, „U“, „I“, „S“, „V“, „B“, … besonders markante Bewegungen hervorrufen. Parallel dazu führt das System eine entsprechende Analyse des ankommenden Audiosignals durch und versucht aus diesen beiden Informationen den momentanen Bild-/Tonversatz zu ermitteln. Für die Erkennung dieser Laute im Audiosignal wird eine patentierte Technik verwendet, die ohne Einlernphase auskommt und dadurch unabhängig vom Sprecher ist.

Das System funktioniert bei Gesichtern die zum Zuschauer zugewendet sind (z.B. Nachrichtensprecher/in). Für Sportereignisse, Musikbeiträge (Orchester) und synchronisierte Filme und Beiträge ist das Verfahren nicht geeignet.

5.2 Offline-Messung

5.2.1 SmartLips LipSync management system (Broadcast Project Research)

Von einem Gernerator wird sowohl ein Helligkeitsimpuls als auch ein in zeitlicher Lage definierter Tonimpuls in regelmäßigen Abständen am Set erzeugt. Die von einem Mikrofon bzw. von einer Kamera aufgenommenen und über eine Übertragungsstrecke versandten Audio- und Video-Impulse werden am Empfangsort mit dem Auswertegerät über einen Lautsprecher und einen Videomonitor analysiert. Akustische und optische Laufzeiten vor dem Mikrofon/der Kamera und hinter den Monitoren werden dabei mit erfasst.

5.2.2 Messverfahren mit Tektronix VM 700

Auch der Tektronix VM 700 bietet einen Messmodus zur Messung des Bild/Ton-Versatzes an. Voraussetzung für diese „Offline“-Messung ist eine Signalquelle, die eine Bildsequenz abgibt, in der die Sequenz mit schwarzen Bildern durch einen weißen Bildimpuls unterbrochen wird, der mit einem Tonburst gekoppelt ist. Um die Messung zu ermöglichen, darf die Sequenz außer diesen Impuls keinen weiteren Weißimpuls (z. B. Weißimpuls in der Austastlücke) enthalten. Die erzielbare Genauigkeit mit diesem Verfahren liegt bei ± 200 µs bei einem maximalen Messbereich von ± 2 sec, kann aber die beiden Frontenden des Übertragungskanals (Kamera und Videomonitor) nicht mit erfassen.

Da der Tektronix VM 700 ein relativ aufwendiges und teures Messsystem ist, kann das System nur bedingt als Betriebsverfahren eingesetzt werden. Als Testsignalgeber kann der Tektronix TG2000, oder von einer MAZ abgespielte Testsequenzen verwendet werden.

5.2.3 Messverfahren mit „VALID“ der Firma Pro-Bel

Der Begriff „VALID“ steht für „Video Audio Line-up & IDentification“.

Das System besteht aus einem Generator und einem Reader. Der Generator erzeugt ein spezielles Testbild mit entsprechenden Audiosignalen. Die Audiosignale können entweder separat über eine externe Leitung oder als „embedded Audio“ übertragen werden.

Mit dem Testsignal können außer dem Bild-/Tonversatz auch noch weitere Parameter wie z.B. Audioaussteuerung und -kanalzuordnung ermittelt werden.

Das Messgerät kann den ermittelten Versatz in Millisekundenschritten angeben und hat hierbei, laut Hersteller, eine Messgenauigkeit von ±1 ms.

Die Messergebnisse können von der Auswerteeinheit bei Bedarf in das SD-/HDSDI-Signal eingeblendet werden.

Bild 8: VALID-Signal mit Auswertung

5.2.4 Messverfahren der Firma OmniTek

Omnitek bietet ein auf einer PC-Plattform aufgebautes Analysesystem für SDI-Signale an. Ein Bestandteil der Omnitek Applikationen ist die Option der Laufzeitmessung.

Der Analsyser kann an den zwei SDI/HDSDI-Eingängen mit einem speziellen Omnitek-Testsignal folgende Messungen durchführen:

Laufzeit Videosignal A gegen Videosignal B
Laufzeit embedded Audio A gegen embedded Audio B
Laufzeit Videosignal gegen embedded Audio

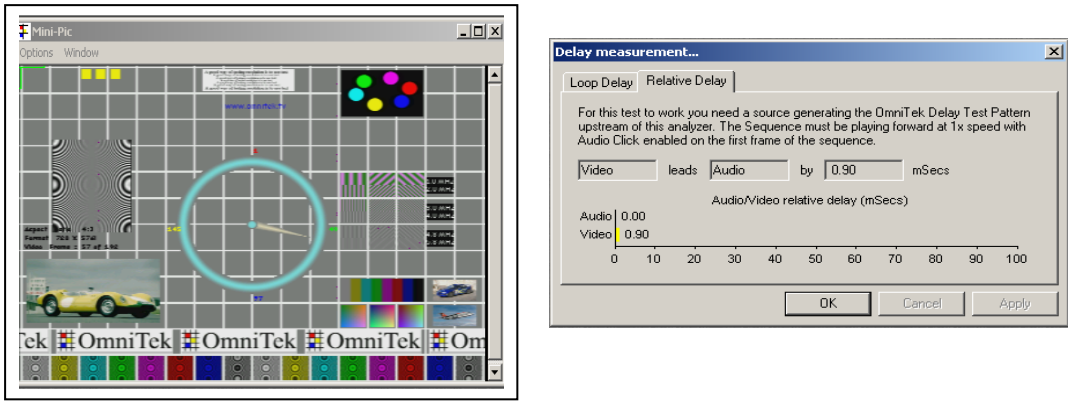

Bild 9: Omnitek-Signal mit Auswertung

5.2.5 Messverfahren mit Tektronix WFM 71xx/WVR 61xx

Tektronix bietet die Messung des Bild/Tonversatzes als numerische Ausgabe bei Verwendung einer speziellen Testsequenz (Helligkeitssprung + Tonburst)

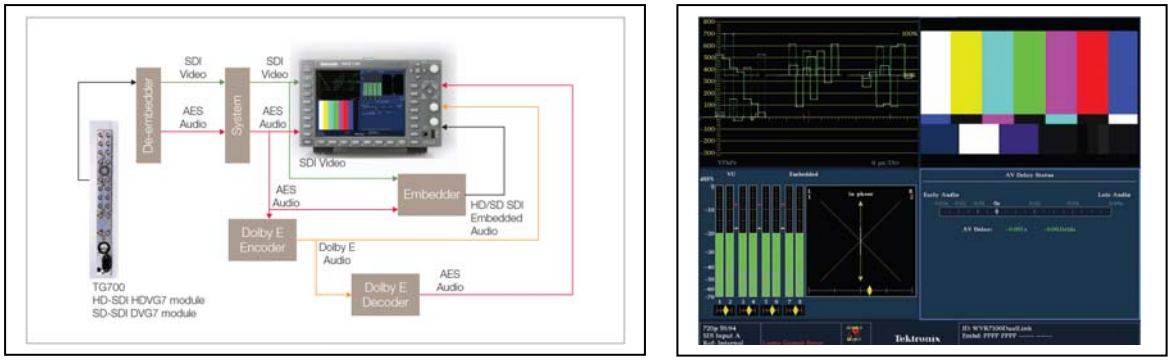

Bild 10: Tektronixverfahren mit Auswertung

5.2.6 EBU-Testsequenz

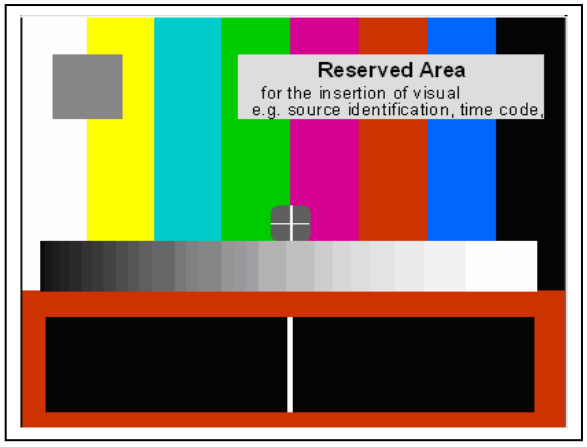

Bild 11: Testsequenz EBU Tech 3305

Die Testsequenz EBU Tech.3305 ermöglicht eine subjektive Beurteilung des Bild/Tonversatzes

5.3 Absolute Laufzeiten

Zur Analyse des Bild/Ton-Versatzes kann man die Laufzeiten für die Signalwege von Audio und Video auch getrennt ermitteln. Wichtig ist eine Abschätzung der Signallaufzeit, um bei der Wahl des Testsignals eine Mehrdeutigkeit zu vermeiden. Als mögliches Testsignale bieten sich für die Messung der Videolaufzeit bietet sich z. B. das Bounce-Signal (mit entsprechender Impulswechselfrequenz) an. Sowohl das Eingangs- als auch das Ausgangssignal wird mit einem Oszilloscop unter Beachtung der richtigen Triggerung dargestellt und analysiert. Auf der Tonseite sind z.B. durch Tonbursts Laufzeitmessungen möglich.

Verfügt man über einen Videogenerator, dessen Bounce-Signal sich extern von einem Audio-Mess-Equipment triggern lässt, so kann man das Video- und das Audio-Signal an der „Ausgangsseite“ direkt miteinander darstellen und so den Bild/Ton-Versatz ermitteln. In jedem Fall sollte auch hier vorher eine Laufzeitabschätzung erfolgen, um Messfehler auszuschließen. In der EBU Tech 3283 ist ein ähnliches Messverfahren beschrieben.

6. Reduzierung des Bild/Ton-Versatzes

Generell lässt sich mit der üblichen Leitungspegelung vor der Sendung/Aufzeichnung, die mit einer Signalgröße -9 dBr (-3dBu, -18 dBFS) und der Frequenz 1 kHz sowie einem Farbbalken-Testsignal erfolgt, keine Aussage über die Bild/Ton-Beziehung machen. Nur durch die Übertragung eines eindeutig zeitgleich definierten Video- und Audiosignals lässt sich diese erkennen und es lassen sich gegebenenfalls im Vorfeld Maßnahmen ergreifen.

6.1 Abschnittsweises Kompensieren

Wie dem Kapitel 4 zu entnehmen ist, ist durch den Einzug der Digitaltechnik die Verzögerungsproblematik wesentlich vergrößert worden. Da relevante Verzögerungszeiten überwiegend im Videosignal auftreten, ergeben sich nahezu keine statistisch verteilte Kompensationen mit den entsprechenden Audiolaufzeiten.

Um in der Gesamtheit die in den Standards genannten Toleranzen einhalten zu können, sollte bei jedem Produktionsschritt und Übertragungsabschnitt der Bild/Ton-Versatz berücksichtigt und wenn möglich vorhandene Laufzeitendifferenzen ausgeglichen werden (Vergl. Kapitel 3.2). Es sollte vermieden werden, dass ein Gerät oder ein Produktionsabschnitt bereits den Toleranzbereich „aufbraucht“, der über die gesamte Produktions- und Übertragungskette eingehalten werden sollte.

6.2 Konsequente Beachtung bei Aufnahme und Bearbeitung

Wie in Kapitel 4.3.2 und 4.3.3 beschrieben, kann schon bei der Aufnahme und bei gewissen Produktionsschritten durch bewusstes Vorgehen der Bild/Ton-Versatz in gewissen Grenzen reduziert werden **und dies teilweise ohne jeglichen finanziellen Aufwand**. Schulungs- und Fortbildungsmaßnahmen für die betreffenden Mitarbeiter sollten dieser Thematik gerecht werden.

Bei einer Streckenüberprüfung muss auch die A/V-Beziehung mit einem Offline-Messverfahren (z.B. mit VALID-Signal) überprüft werden .

6.3 Planung

Bei der Planung eines digitalen Studios werden allzu leichtfertig Videosynchronizer zur Laufzeitanpassung eingesetzt. Eine konsequente Laufzeitplanung kann die Notwendigkeit von diesen Verzögerungselementen reduzieren, um dann den Aufwand bzw. die Notwendigkeit für Audiodelays in Grenzen zu halten.

Es ist darauf zu achten, dass bei allen möglichen Signalwegen an den Übergabepunkten der Bild/Tonversatz kompensiert ist. Besonderer Augenmerk ist dabei auf die Verzögerungszeit von Dolby-E-Codecs zu legen.

Siehe auch Technische Richtlinie zur Herstellung von Fernsehproduktionen.

6.4 Synchronisation der Taktgeber mit GPS

Werden alle Videotakte der Rundfunkanstalten, z. B. der ARD, von einer zentralen Referenz synchronisiert, so stehen die Takte in einem festen Bezug zueinander. Die Folge ist, dass mit konstanten Phasenbeziehungen der Eingangssignale (isochrone Signale) gearbeitet werden kann. Daraus resultiert, dass die in 4.4.1 beschriebenen Phänomene wie Weglassen bzw. Verdoppeln von Bildern und die dazugehörigen Tonstörungen vermieden werden. Dies gilt auch für die Synchronisation von Ü-Wagen bei Live-Sendungen. Insbesondere sollte gleiches auch bei der Übertragung von DolbyE-codierten Audio-Surround-Signalen erfolgen, da dann die resultierenden „Klicks“ nicht mehr durch die Asynchronität auftreten können.

6.5 „Timestamps“ für zukünftige digitale Systeme

Bei digitalen Übertragungsverfahren wie MPEG2 kann durch die Einführung von Timestamps eine synchrone Übertragung von Video und Audio erreicht werden. Der digitale SD-Video-Standard ITU-R BT 601 und der HD-Video-Standard ITU-R BT 709 beinhaltet im Gegensatz zum Audio-Standard derzeit keine genormten Möglichkeiten zur Mitführung von Timestamps. Für das Audiosignal kann nach EBU Tech. 3250 bzw. nach AES 3∗ ein sample-genauer Timecode bereitgestellt werden. Bei zukünftigen Entwicklungen und Festlegungen sollten Mechanismen zur automatischen Korrektur des Bild/Ton-Versatzes eingebunden werden.

7. Anmerkungen

Nochmals sei darauf hingewiesen, dass die in der ITU-R BT.1359 angegebenen Werte mit Nachrichtensprecherinnen in Japan, der Schweiz und in Australien ermittelt wurden und ***nicht*** Szenen (z. B. Sport- oder Musikbeiträge) betrachten, die wesentlich kritischer sein können. **Aus diesem Grund ist für ein allgemein gültiges Umfeld die EBU-R 37 (2007) mit den enger spezifizierten Werten zu Grunde zu legen.**

∗ Timecode in EBU Tech. 3250 im Channel-Status-Bit: Byte 18-21;Time-of-day sample address code (32-bit binary)

Es werden Untersuchungen folgen, die das tatsächliche Bild/Ton-Verhältnis der gesamten Signalkette bis zum Zuschauer aufzeigen. Aus den Ergebnissen sind entsprechende Maßnahmen abzuleiten und zu realisieren.

Selbst wenn sorgfältig bei der Produktions- und Übertragungskette auf den Bild/Ton-Versatz geachtet wird, sind nicht alle Fehlerquellen vollständig kompensierbar.

8. Literatur

[1] Diplomarbeit „Laufzeitdifferenzen zwischen Video- und Audiosignalen bei der digitalen Bildbearbeitung“ von Herrn Gregor Schmid / Fachhochschule Wiesbaden 04.07.1995

Institut für Rundfunktechnik
Floriansmühlstraße 60
80939 München
www.irt.de
Tel. +49 (0) 89 | 323 99 - 204
Fax +49 (0) 89 | 323 99 - 205
presse@irt.de

Registergericht München Eintrag Abteilung
B Band 65 Nr. 5191