

Multimedia-Messaging-Technologien auf der Basis von MPEG-7

Jörg Heuer

Uwe Rauschenbach

André Kaup

{Joerg.Heuer, Uwe.Rauschenbach, Andre.Kaup}@mchp.siemens.de
Siemens Corporate Technology, Information and Communications
81730 München

Zusammenfassung

In diesem Beitrag werden Anforderungen an eine zukünftige asynchrone Multimediakommunikation vorgestellt, Lösungsansätze basierend auf dem Metadaten-Standard MPEG-7 erläutert und eine prototypische Realisierung beschrieben. Wesentliche Rahmenbedingungen aus dem beschriebenen Szenario sind die intuitive Generierung von Nachrichten mit Multimediaelementen und eine Kommunikation zwischen mobilen Endgeräten unterschiedlicher Gerätegattungen. Ein Lösungsansatz für diese Problematik basierend auf einer Metadatenbeschreibung der Nachricht inklusive der Mediadaten nach dem neuen Standard MPEG-7 wird ausführlich vorgestellt. Zudem wird aufgezeigt, wie mit MPEG-7 verwandten Technologien bei der Abfrage des Messaging-Centers die Nachrichtendarstellung an die Nutzerpräferenzen, das Endgerät und dessen Netzanbindung angepaßt werden kann.

1. Einleitung

Bereits heute werden die ersten Multimediaelemente wie animierte Bilder in SMS-Nachrichten verwendet. Allerdings sind dies meist vordefinierte Multimediaelemente, die dem Nachrichtensender wenig Freiheiten lassen, eine Nachricht nach seinen Bedürfnissen zu gestalten. Durch zunehmende Übertragungsbandbreite in den mobilen Kommunikationsnetzen und Rechenkapazität auf den Kommunikationsendgeräten wird die Möglichkeit, Multimediaelemente für die Kommunikation zu erzeugen, zu übermitteln und darzustellen mit den nächsten Gerätegenerationen zunehmen. Aufgrund der ermöglichten Funktionsvielfalt werden sich bezüglich der Funktion unterschiedliche Gerätegattungen ausprägen. Ein Beispiel hierfür sind tastaturlose PDAs mit drahtloser Netzanbindung. Jedoch ist heute eine asynchrone Kommunikation zwischen Geräten unterschiedlicher Gattungen nur bei Verwendung textueller Nachrichten möglich. Um eine zunehmende Isolierung der Kommunikation bezüglich der Endgeräte zu vermeiden, muß an die asynchrone Multimediakommunikation die folgende Anforderung gestellt werden:

- (1) Nachrichten inklusive der enthaltenen Multimediaelemente müssen mit geringem Aufwand an zur Verfügung stehende Netzwerkressourcen und Endgeräte angepaßt werden können.**

Bereits heute sind multimediale Nachrichten in Form von E-Mails möglich. Jedoch hat dies bisher keine Verbreitung gefunden, da die unbearbeiteten

Multimediaelemente nur eine geringe Informationsdichte aufweisen und schwierig vom Nachrichtensender zu erstellen sind. Hieraus leitet sich für die Endgeräte die folgende Anforderung ab:

- (2) Multimediamnachrichten müssen intuitiv erstellt werden können.
Um eine Nachricht anschaulich zu übermitteln, muß der Nachrichtensender Media- und Nachrichtenelemente verknüpfen und deren Bedeutung in der Nachricht beziehungsweise zueinander spezifizieren und übertragen können.**

Unabhängig von der Netzanbindung oder den Eigenschaften der Geräte hat der Nachrichtempfänger abhängig von der momentanen Situation unterschiedliche Anforderungen an die Präsentation von Nachrichten. Beispielsweise wird man in Eile nur eine Kurzübersicht der in der Messsagebox enthaltenen Nachrichten anfordern, während man in Ruhe eine detailliertere Übersicht zur Auswahl relevanter Informationen bevorzugt.

- (3) Die Übertragung und Präsentation der Multimediamnachrichten, die von einem Messagecenter empfangen worden sind, müssen an die zum Abfragezeitpunkt existierenden Nutzerpräferenzen angepaßt werden.**

Basierend auf diesen Anforderungen aus Sicht der Applikation werden im nächsten Abschnitt technische Lösungsmöglichkeiten diskutiert. Anschließend werden die Komponenten einer ersten prototypischen Implementierung dieser Multimedia-Messaging-

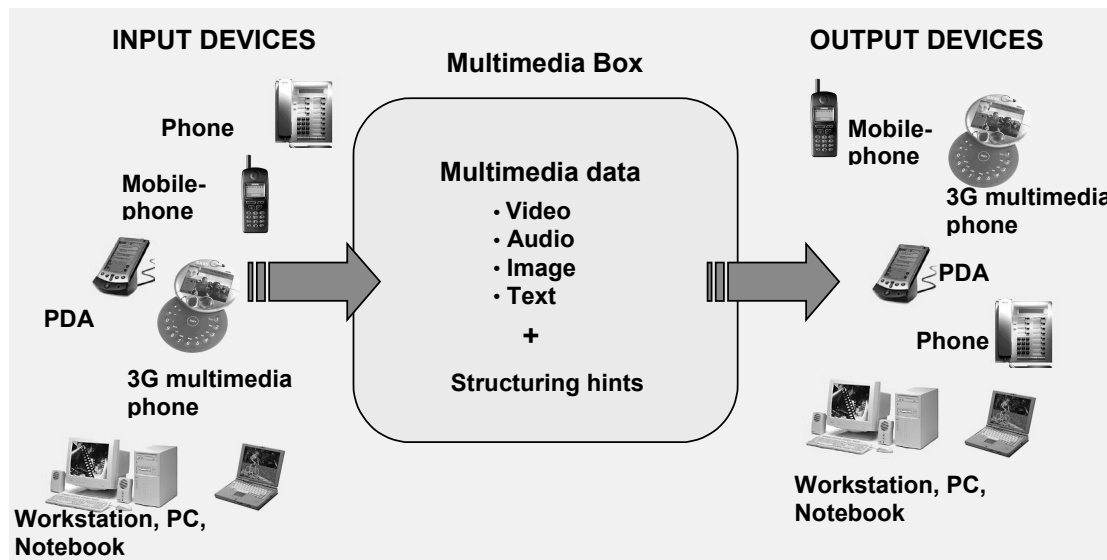


Abb. 1: Das Messaging Szenario.

Applikation vorgestellt und abschließend ein Ausblick gegeben.

2. Lösungsansätze

Ausgehend von Anforderung (1) läßt sich wie in Abb. 1 dargestellt eine grundlegende Architektur der asynchronen Kommunikation beschreiben: Entsprechend der Eigenschaften des Sendegerätes wird eine Nachricht erstellt, die neben Mediaelementen auch strukturierende Metadaten enthält. Diese wird an ein Message Center gesendet und dort gespeichert. Bei Abruf der Nachrichten einer Messagebox werden diese mit Hilfe der strukturierenden Metadaten auf das abfragende Endgerät angepaßt.

Lösungen für die Adaption von Multimediadokumenten an das abrufende Gerät wurden mit der InfoPyramid [1] und bezüglich einer dynamischen Adaption schon in [2] vorgestellt. Diese Ansätze definieren ein Maß für den Informationsverlust, nach dem eine optimale Adaption einzelner Mediaelemente im Multimediadokument durchgeführt wird. Eine Medienadaption basierend auf dem Metadatenstandard MPEG-7 [3,4] für Multimediadaten entspricht dem Konzept der InfoPyramid, beinhaltet jedoch zudem sogenannte Transcoding Hints, die eine schnellere Transcodierung ermöglichen. Während die ersten zwei Lösungsansätze für den Informationsabruf eine automatische Adaption der Mediaelemente auf der Serverseite beschreiben, bietet MPEG-7 für das beschriebene Kommunikationsszenario den Vorteil, daß die Kommunikation der verwendeten Metadaten Ende 2001 standardisiert werden wird und somit die Adaption clientseitig gesteuert werden kann.

Die Anforderungen aus Sicht des Nachrichtensenders (2) erfordern eine Unterstützung durch das Endgerät für die Generierung der Nachricht. Dies kann beispielsweise die

automatische Berechnung einer zeitlichen Segmentierung sein, die dem Anwender für die weitere Kommentierung vorgeschlagen wird. Die Nachricht kann gemäß der Vorstellungen des Nachrichtensenders strukturiert sein und mehrere Mediaelemente mit unterschiedlicher Bedeutung enthalten. Dies ist beispielhaft in Abb. 2 dargestellt. Die generierten Metadaten, zum Beispiel die Unterteilung und Kommentierung einer Videoszene, müssen standardisiert übertragen werden, damit das Empfangsgerät unabhängig vom Hersteller des Gerätes diese Information darstellen kann. Für die Übertragung der Metadaten in Bezug auf die Struktur der Nachricht und die Semantik der Nachrichtenelemente eignet sich das Codierungsformat, das innerhalb von MPEG-7 standardisiert wird.

In Bezug auf die Anforderungen aus Sicht des Empfängers (3) ist zu beachten, daß die Codierung der Metadaten innerhalb von MPEG-7 auf XML-Schema Definitionen beruht. Aufgrund dessen können Nachrichten an die momentanen Bedürfnisse des Empfängers nicht nur basierend auf den Formaten der Mediaelemente angepaßt werden, wie es heute bei protokollorientierten Adaptionen schon möglich ist, sondern mit Hilfe von XML-basierten Technologien eine Adaption abhängig von der Semantik des bestimmten Mediaelements spezifiziert werden. So kann beispielsweise ein Audiokommentar im Format adaptiert übertragen werden, da Audiokommentare dem Empfänger wichtig sind. Jedoch wird ein Audioelement, das die Einleitung repräsentiert, nicht als wichtig eingestuft und deshalb zunächst nicht adaptiert bzw. übertragen.

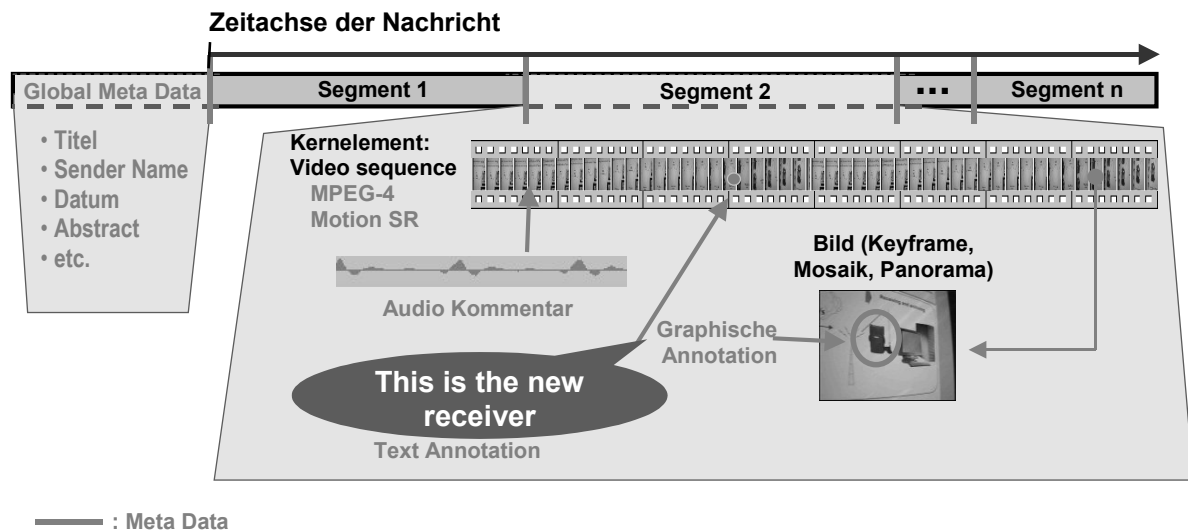


Abb. 2: Struktur einer Multimedia Nachricht, die mehrere Mediaelemente, wie beispielsweise Video, Audio, Bilder und Text, mit unterschiedlicher Bedeutung enthält.

3. Das MultiMedia Message Center – Die M³-Box

Die diskutierten Lösungsansätze sind in die Entwicklung einer ersten prototypischen Implementierung des beschriebenen Kommunikationsszenarios eingeflossen [5]. Im folgenden wird näher auf die enthaltenen Komponenten eingegangen. Hierzu wird in diesem Abschnitt zunächst das Nachrichtenformat und die technologischen Komponenten eines simulierten Endgerätes beschrieben. Die Komponenten für die Nachrichtenspeicherung in der M³-Box werden anschließend diskutiert. Abschließend wird auf die Realisierung der Adaption beim Nachrichtenabruf eingegangen.

3.1. Die Nachrichtenerstellung

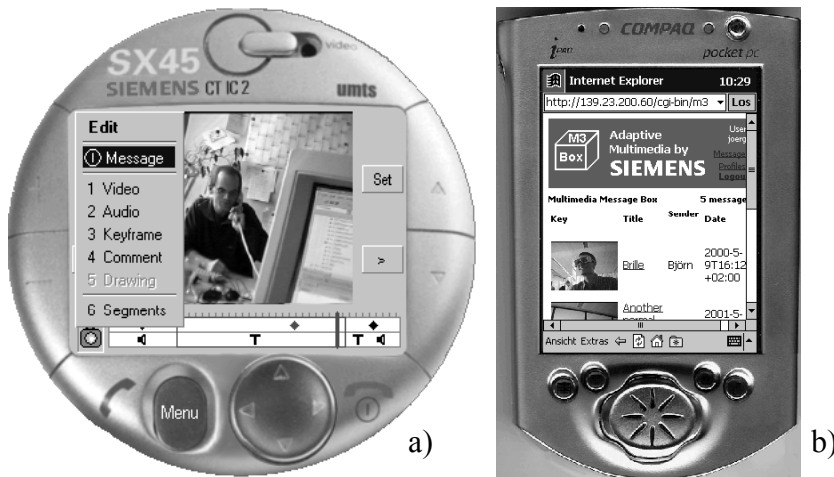
Meist wird der Bedarf eine Nachricht zu senden durch Ereignisse in der Umwelt des Nachrichtensenders hervorgerufen. Daher ist es hilfreich den Prozeß der Nachrichtenentstehung zu betrachten, um ein Bild von der verwendeten Nachrichtenstruktur zu gewinnen. Entsprechend der Initiierung der Nachricht stellt die Repräsentation des erwähnten Ereignisses bzw. Anlasses der Nachricht ein Kernelement der Nachricht dar. Dieses Kernelement ist ein Mediaelement, das beispielsweise als Video (s. Abb. 2) übertragen wird. Als Mediaelemente werden in der hier vorgestellten Realisierung Video, Bilder, Audio und Text betrachtet. Die Nachricht des Senders besteht jedoch im allgemeinen nicht nur aus dem Kernelement, sondern zum großen Teil aus den Informationen, die sich auf dieses Kernelement beziehen. Dies sind unter anderem Kommentare, die in Form unterschiedlicher Mediaelemente aufgenommen wurden bzw. vorliegen, wie beispielsweise als Audio- oder

Textkommentar. Um den Nachrichtempfänger anzusprechen und den Kontext einer Nachricht aufzuzeigen, wird zudem beispielsweise eine Einleitung zum Nachrichten Kern hinzugefügt. Dementsprechend setzt sich eine Nachricht aus mehreren Mediaelementen zusammen, die in der Nachricht unterschiedliche Bedeutung haben können. Diese Bedeutung ist dem Nachrichtenersteller bekannt und ist neben der Übermittlung der Mediaelemente in Form von Metadaten zu übertragen, um sie für die Adaption von Nachrichten nutzen zu können.

Die Mediaelemente, die die Bedeutung von Kommentaren haben, können sich hierbei nur auf Teile des Kernelements der Nachricht beziehen. Beispielsweise kann sich ein Audiokommentar nur auf ein Videosegment beziehen, das mit einem einzelnen Keyframe abstrahiert wird. Dementsprechend sind neben der Bedeutung der Mediaelemente auch die Relationen zwischen den Mediaelementen als Nachrichtinhalt zu übermitteln (s. Abb. 2).

Während das gesamte Mediakernelement als Hintergrundinformation interessant sein kann, ist für die Verdeutlichung der Nachrichten eine abstrahierte Repräsentation des Mediakernelements oft hilfreich. So kann die Zusammenfassung eines Mediakernelements für die Kommentierung effizienter genutzt werden. Strukturelle Informationen, die eine Zusammenfassung der Kernelemente erlauben, sollten daher ebenfalls Bestandteil der Metadatenbeschreibung der Nachricht sein.

Neben den aufgeführten Metadaten, die die Bedeutung und Struktur der Nachricht beinhalten, können Informationen über das Codierungsformat der Mediadaten als Metadaten abgespeichert werden. Dies ermöglicht eine



**Abb. 3: a) Nachrichtenerstellung mit einem Videophone
b) Angepaßter Nachrichtenabruf mit einem Palm Computer**

Filterung oder das Anstoßen einer Transcodierung, ohne den Mediadatastrom zuvor zu decodieren.

Entsprechend der heutigen textbasierten Nachrichten sind ebenfalls Metadaten wie der Name und die Adresse des Nachrichtensenders, das Datum und die Uhrzeit der Sendung zu übertragen.

Um eine Kommunikation zwischen Endgeräten unterschiedlicher Hersteller zu ermöglichen, ist es notwendig, daß die Nachricht in einem standardisierten Format übertragen wird. Während für die Mediaelemente Standards wie MPEG-4 zur Verfügung stehen, wird eine standardisierte Übertragung von Metadaten erst durch den kommenden Standard MPEG-7 ermöglicht. In einer ersten prototypischen Implementierung wurden die beschriebenen Metadaten basierend auf [6] in Instanziierungen der folgenden *Description Schemes* codiert:

- *Segment DS*, zur Beschreibung der Struktur der Mediaelemente,
- *StructuredAnnotation DS*, zur Beschreibung von Kommentaren,
- *Summarization DS*, zur Beschreibung von abstrahierten Repräsentationen der Mediaelemente.

In der Realisierung eines Endgerätes, mit dem Multimediamnachrichten aufgenommen werden können, ist die Generierung der Metadaten jedoch transparent. Dies bedeutet, daß der Nutzer durch den Editierungsprozeß einer Nachricht von der Nutzerschnittstelle des Endgerätes geleitet wird, ohne daß er Zugriff auf die hierbei vom Endgerät generierten Metadaten erhält. Dies soll eine intuitive Erstellung der Nachricht unabhängig von der verwendeten Codierung der Metadaten ermöglichen. In Abb. 3a) werden dem Nutzer beispielsweise in einem Menü verschiedene Editierungsoptionen gegeben. Je nach Auswahl werden

dem im folgenden generierten Nachrichtenelement entsprechend der getroffenen Auswahl Metadaten vom Gerät hinzugefügt.

Trotz dieser Führung des Nutzers durch den Prozeß der Nachrichtenerstellung kann die Generierung von Metadaten beispielsweise in Bezug auf die Strukturierung der Nachricht zeitaufwendig sein. Hier ist es erforderlich, daß das Endgerät den Nutzer in der Eingabe der Nachricht unterstützt. Betrachtet man das Beispiel, daß das Kernelement der Nachricht ein aufgezeichnetes Video ist, so kann das Gerät das Video automatisch segmentieren und

Keyframes selektieren und dem Nutzer als Vorauswahl zur Verfügung stellen. Der Nachrichtenersteller kann diese Auswahl nutzen, um das Kernelement der Nachricht zu kommentieren, ohne zuvor explizit aus dem gesamten Video Szenen auszuwählen. Eine solche automatische Segmentierung und Keyframe Selektion basierend auf der Trajektorie der globalen Bewegung wurde für Eingabegeräte realisiert. Im Gegensatz zu Segmentierungsalgorithmen, wie sie für Rundfunksendungen Anwendung finden, werden hierbei nicht Schnitte oder Überblendungen detektiert, sondern Segmente aufgrund der Änderungen in der Kamerabewegung festgelegt und Keyframes aufgrund des repräsentativen Bildinhaltes bezüglich des Segments ausgewählt.

Neben der geräteunterstützten Generierung der Nachricht ist auch die Darstellung der Nachricht bei der Erstellung an das verwendete Gerät zu adaptieren. Bei dem in Abb.3a) dargestellten Gerät ist dies beispielsweise durch eine kompakte Repräsentation der bisher editierten Nachrichtenstruktur im unteren Bereich der Anzeige geschehen. Hierbei wird der Teil der Nachricht bezüglich Ihrer Struktur detailliert dargestellt, der gerade generiert bzw. editiert wird.

3.2. Die Nachrichtenspeicherung in der M³-Box

Die im vorhergehenden Abschnitt beschriebene Nachricht setzt sich aus mehreren Mediaströmen und einem Metadatenstrom zusammen. Diese werden in einem Strom codiert über eine TCP/IP-Verbindung zur M³-Box gesendet. Eine solche Übertragung wird auch in [7] beschrieben, ist zur Zeit jedoch noch nicht standardisiert.

Der an den Empfang der Nachricht anschließende Prozeß der Nachrichtenspeicherung kann in drei Verarbeitungsschritte aufgeteilt werden:

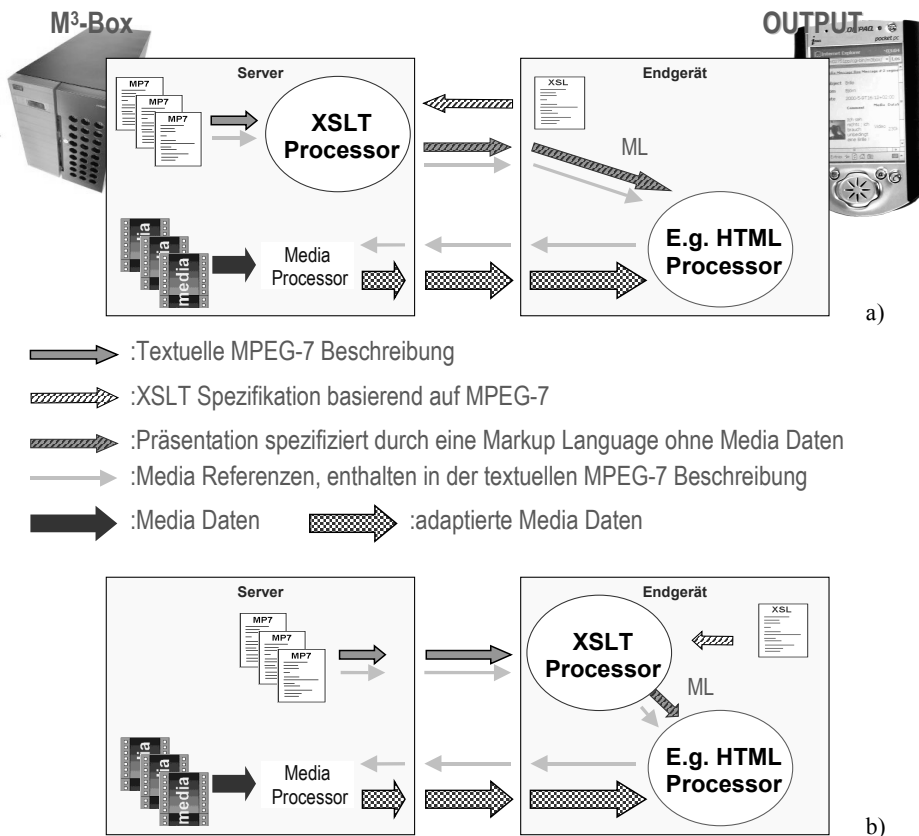


Abb. 4: a) Serverseitig und b) Clientseitig gesteuerte Adaption der Nachricht

- Extraktion von ergänzenden Metadaten,
- Speicherung der Metadaten und
- Transcodierung und Speicherung der Mediadaten.

Sind bei dem Empfang von Nachrichten Metadaten beispielsweise zu den Formaten der Mediadaten in der Nachricht oder der Struktur von Mediadaten nur teilweise vorhanden, da diese vom Nachrichtensender nicht vollständig generiert worden sind, so können diese nachträglich ergänzt werden. Zudem werden weitere Metadaten wie beispielsweise der Zeitpunkt des Empfangs der Nachricht hinzugefügt.

An die Speicherung der Metadaten werden in diesem Szenario drei Anforderungen gestellt:

- eine schnelle Speicherung, da kontinuierlich Nachrichten empfangen werden können.
- einen schnellen Abruf der gesamten Metadaten einer Nachricht, um diese für die Adaption der Nachricht auf einem ausgebenden Endgerät filtern zu können,
- eine schnelle Suche in den Metadaten, um das Auffinden von bestimmten Nachrichten auch bei Anfragen mehrerer Nutzer zu ermöglichen.

Die ersten zwei Kriterien werden bereits durch eine geeignete Speicherung der Metadaten in einem

Filesystem, wie es in [5] beschrieben ist, erfüllt. Jedoch steigt die Suchkomplexität linear mit dem Metadatenvolumen. Eine geeignete Indexierung der Metadaten erlaubt, diese Komplexität zu verringern. Da jedoch im Vergleich zu anderen Applikationen im allgemeinen die Speicheroperation wesentlich häufiger erfolgt als die Suche in dem Datenbestand, sind Indexierungs- bzw. Datenbanklösungen zu verwenden, die ein schnelles Laden der Datenbank ermöglichen. In der Implementierung des Prototyps wird daher die Datenbank dbXML [8] verwendet, die eine Indexierung der MPEG-7 Metadaten Elemente bei schneller Datenspeicherung und Abruf der gesamten Metadaten einer Nachricht erlaubt.

Bei der Speicherung der Mediadaten ist der schnelle Abruf der Mediadaten gegenüber dem Speicherplatzbedarf zu optimieren.

Die Strukturbeschreibung mit den Metadaten stellt eine Indexierung der Mediadaten dar. Entsprechend kann der Zugriff auf Teile der Mediadaten, die mit den Metadaten indexiert sind, bei der Speicherung der Mediadaten optimiert werden. Dies kann beispielsweise erreicht werden, indem indexierte Keyframes in einem Video zur Speicherung extrahiert werden oder das entsprechende Frame im Video für eine schnelle Extraktion bei Abruf als intracodiertes Frame gespeichert wird. In dem realisierten Prototypen wurde letzteres Verfahren realisiert.

3.3. Der Nachrichtenabruf aus der M³-Box

Beim Abruf der Nachrichten ist es erforderlich, diese entsprechend dem verwendeten Endgerät, der Netzwerk-anbindung und den Nutzerpräferenzen anzupassen. Basierend auf den strukturellen Metadaten ist es möglich, die Präsentationsform der Nachricht anzupassen und die Mediadaten zu adaptieren. Hierbei kann die Präsentation der Mediaelemente in Abhängigkeit von deren Bedeutung in der Nachricht angepaßt werden. Die Formatinformation der Mediaelemente kann auf der anderen Seite genutzt werden, um eine Formatkonversion entsprechend der Vorgaben des Endgerätes durchzuführen.

Eine Adaption der Nachricht kann - wie in Abb.4 dargestellt wird - server- oder clientseitig gesteuert werden. Wird die Adaption serverseitig gesteuert, so sind die Endgerät-, Nutzer- und Netzwerkpräferenzen an den Server zu übermitteln. Entsprechend der Anforderung einer Kommunikation der Metadaten hat auch die Übermittlung der Präferenzen standardkonform zu erfolgen, um die Interoperabilität zwischen Geräten unterschiedlicher Hersteller sicherzustellen. Bei der clientseitig gesteuerten Adaption werden zunächst die MPEG-7-Metadaten der Nachricht an das Endgerät übertragen. Basierend auf den MPEG-7-Metadaten können anschließend die angepaßten Mediadaten beim Messagecenter angefordert werden.

In der Implementierung der M³-Box wurden die Präferenzen mit Hilfe von *Extensible Style Sheet Language* (XSL) Transformationsanweisungen [9], die auf den MPEG-7-Metadaten angewendet werden, formuliert. Dies ermöglicht ausgehend von den MPEG-7 Metadaten Präsentationen für Markup Language basierte Browser wie HTML-Browser zu generieren. Da mit MPEG-7 der XML Dialekt auf den die XSL-T Anweisungen operieren festgelegt ist, können diese allgemein gültig formuliert werden und die Adaption der Nachrichten wie in Abb.4 dargestellt sowohl server- wie auch clientseitig gesteuert werden.

In der Implementierung des Szenarios wurden XSL-T Anweisungen für unterschiedliche Präferenzen und Endgeräte generiert. Die MPEG-7 Metadaten werden über eine HTTP Verbindung von einem Apache Web Server [10] zur Verfügung gestellt. Bei initialer Anmeldung auf dem Server werden die Präferenzenbeschreibungen übermittelt oder bereits in vorhergehenden Sessions verwendete Präferenzen ausgewählt. Die resultierende, adaptierte Präsentation einer Nachricht auf einem Palm-Computer ist beispielsweise in Abb. 3b) dargestellt. In den Präferenzen wurde hierbei beispielsweise spezifiziert, daß der Inhalt der Messagebox in drei Hierarchien angezeigt werden soll: die in der Messagebox enthaltenen Nachrichten im Überblick mit einem einzelnen Keyframe pro Nachricht, die einzelne Nachricht im Überblick bezüglich der enthaltenen Nachrichtensegmente und eine detaillierte Darstellung eines Nachrichtensegments. Entsprechend der MPEG-7-Beschreibung der Segmente in der *Summarization DS* können die einzelnen Mediaelemente in der Darstellung an das Endgerät angepaßt werden. So kann beispielsweise ein Video durch mehrere Standbilder oder als Mosaik dargestellt werden, wenn das Endgerät aufgrund der limitierten Rechenleistung ein Video nicht wiedergeben kann.

4. Ausblick

Zur Zeit ist ein erster funktionierender Prototyp implementiert. Wie beschrieben basiert dieser bezüglich der Metadateninformation auf dem Standard MPEG-7. Entsprechend wird die Beschreibung des abrufenden Endgerätes und der Netzwerkverbindung bezogen auf die geforderten Eigenschaften der Mediadaten in MPEG-7 codiert übertragen. Eine präzisere Anpassung auf Seiten des Servers, zum Beispiel auf die zur Verfügung stehende Bandbreite, kann jedoch durchgeführt werden, wenn ergänzend zu den MPEG-7 Metadaten die Beschreibung der Endgeräte- und Netzwerkeigenschaften übertragen wird. Hierzu soll beispielsweise der Vorschlag zur Composite Capabilities / Preference Profiles (CC/PP) [11] innerhalb des W3C weiter untersucht werden.

Literatur

- [1] Smith J. R., Mohan R., Li C., Scalable Multimedia Delivery for Pervasive Computing, In Proc. ACM Intern. Conf. Multimedia '99, pp 130-139, Orlando, November 1999
- [2] Boll S., Klas W., Wandel J., A Cross-Media Adaptation Strategy for Multimedia Presentations, In Proc. ACM Intern. Conf. Multimedia '99, pp 37-46, Orlando, November 1999
- [3] MPEG-7: Context, Objectives and Technical Roadmap, V.12, ISO/IEC JTC1/SC29/WG11/N2861, <http://www.cselt.stet.it/mpeg/>, July 1999.
- [4] P. Salembier, J. R. Smith, MPEG-7 Multimedia Description Schemes, IEEE Trans. on Circuits and Systems for Video Technology, Vol 11, No 6, June 2001.
- [5] Heuer J., Casas J. L., Kaup A., Adaptive Multimedia Messaging Based on MPEG-7 – The M³-Box, Proc. Second International Symposium on Mobile Multimedia Systems & Applications, Delft, The Netherlands, 9-10 November 2000.
- [6] MPEG-7 Multimedia Description Schemes WD v.3.0, ISO/IEC JTC1/SC29/WG11/N3411, <http://www.cselt.stet.it/mpeg/>, May 2000.
- [7] O. Avaro, P. Salembier, MPEG-7 Systems: Overview, IEEE Trans. on Circuits and Systems for Video Technology, Vol 11, No 6, June 2001.
- [8] The dbXML Group L.L.C, dbXML Core Server, <http://www.dbxml.org>, Juni 2001
- [9] Extensible Style Sheet Language – an Overview, <http://www.w3.org/Style/XSL/Overview.html>.
- [10] The Apache XML Project, <http://xml.apache.org/>
- [11] Composite Capabilities/Preference Profiles: Requirements and Architecture, <http://www.w3.org/TR/CCPP-ra/>, July 2000.