Schutz von Multi-Media Content vor der Extraktion einzelner Tonspuren

Marc Ihle

Department of Media Engineering and Technology, German University in Cairo, Ägypten, Email: marc.ihle@ieee.org

Einleitung

Bei der Speicherung und Übertragung von Multi-Media-Content werden zunehmend einzelne Tonspuren als Objekte separat gespeichert oder übertragen. Dies bietet dem Kunden einen Mehrwert durch die Möglichkeit der Generierung von Variationen. Insbesondere ermöglicht es dem Nutzer, etwa bei Computerspielen, interaktiv das Mischungsverhältnis zwischen den Tonspuren/Objekten zu steuern.

Bei der Codierung der Signale können verschiedene Codecs zum Einsatz kommen. Die Spanne reicht von Codecs bei denen die Tonspuren einzeln codiert werden (z.B. via MP3 [1]), über paarweise gemeinsame Codierung (Joint Stereo) bis hin zu einer gemeinsamen Codierung vieler getrennter Objekte, etwa mittels „Spatial Audio Object Coding" (SAOC) [2]. Bei dem letztgenannten System werden die einzelnen Objekte entsprechend Abbildung 1 gemeinsam encodiert, im Endgerät wieder decodiert und im Block „Renderer" abgemischt.

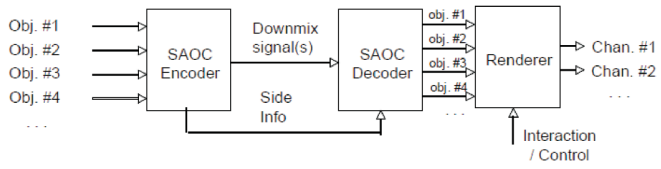

Abbildung 1: Blockdiagramm der SAOC-Codierung [3]

Neben der intendierten Nutzung der Tonspuren im Endgerät können damit auch Karaoke-Versionen abgeleitet oder die Singstimme eines Mediums mit den Background-Spuren eines anderen Mediums kombiniert werden. Solche Vermischungen werden als „mashup music" bezeichnet [4]. Der Autor geht davon aus, dass diese zum Teil nicht-intendierten Nutzungsmöglichkeiten viele Autoren und Produzenten davon abhält, mehrkanaligen Content in Form von Audio-Objekten für die derzeit verfügbaren Medien anzubieten.

Ein Schutz des geistigen Eigentums besteht üblicherweise darin, die einzelnen Objekte zu verschlüsseln. Somit müssen diese zunächst im Endgerät wieder entschlüsselt werden, bevor sie mit anderen Objekten gemischt werden können. Damit hat der Urheber die Möglichkeit zu bestimmen, wer wann welches Objekt nutzen kann. Wird einem Kunden die Nutzung einzelner Objekte gestattet, hat dieser uneingeschränkte Möglichkeiten, diese in beliebigem Mischungsverhältnis oder separat zu nutzen. Selbst wenn ein Wiedergabesystem nur intendierte Mischungsverhältnisse gestatten sollte (etwa über zusätzliche Randbedingungen, die dem Renderer aus Abbildung 1 übergeben werden), so liegen im Wiedergabesystem zwingend die dekodierten Objekte vor und können somit auch leicht für nicht-intendierte Zwecke verwendet werden.

Eine Verschlüsselung der einzelnen Tonspuren hat daher nur sehr begrenzten Nutzen. Dieser Beitrag stellt ein Verfahren vor, mit dem nur die Wiedergabe von zulässigen Kombinationen einzelner Tonspuren in guter Klangqualität möglich ist. Ein Rekonstruktionsversuch von nicht-intendierten Kombinationen führt dagegen zu einer deutlichen Herabsetzung der Klangqualität, so dass derartige Rekonstruktionen nicht sinnvoll verwendet werden können.

Beschreibung des Systems

Das im folgenden vorgestellte System besteht aus zwei Teilen. Abbildung 2 zeigt das Blockdiagramm des im Encoder realisierten Teils. Es codiert die Objekte so, dass die oben genannte Rekonstruktion nicht-intendierter Kombinationen verhindert wird. Der zweite, in Abbildung 3 dargestellte Teil, ist im Decoder angesiedelt und dient zur Rekonstruktion der zulässigen Kombinationen.

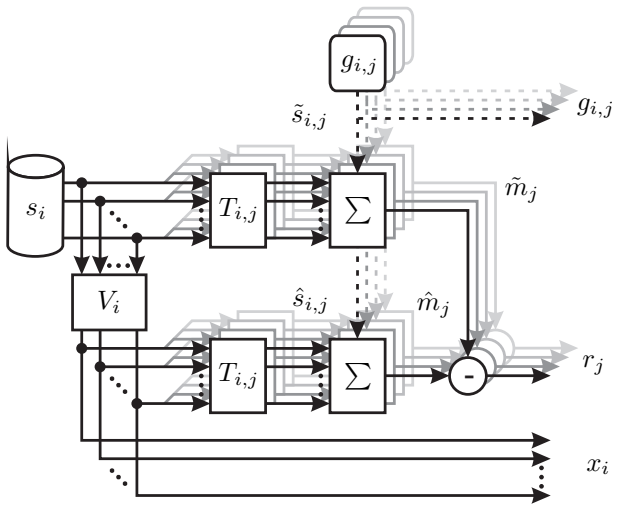

Abbildung 2: Blockdiagramm des Encoders

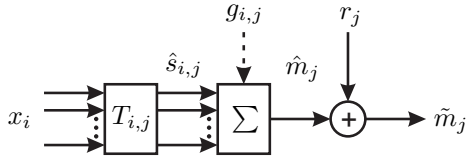

Abbildung 3: Blockdiagramm des Decoders

Im Encoder werden zunächst aus den einzelnen Objekten s_i Varianten $\tilde{s}_{i,j}$ abgeleitet, die nach psychoakustischen

Gesichtspunkten nicht oder nur schwer vom Original unterscheidbar sind. Diese Varianten werden jeweils durch eine Transformation $T_{i,j}$ abgeleitet. Diese könnte im einfachsten Fall z.B. durch eine Phasenverschiebung realisiert werden. Eine solche, lineare Transformation bietet jedoch keinen wirksamen Schutz vor einem Rückschluß auf die Objekte. Sinnvoller ist es daher, eine nicht-lineare Transformation zu verwenden. Sie kann beispielsweise Teil eines verlustbehafteten Musik-Encoders sein.

Im nächsten Schritt werden in den mit $\sum$ gekennzeichneten Blöcken Abmischungen $\tilde{m}_j$ aus sämtlichen zulässigen Mischungsverhältnissen $g_{i,j}$ berechnet:

$$\tilde{m}_j = \sum_{i=0}^{I-1} g_{i,j} \cdot \tilde{s}_{i,j}$$

Diese Abmischungen $\tilde{m}_j$ unterscheiden sich wegen der Verwendung unterschiedlicher Transformationen $g_{i,j}$ selbst bei geringen klanglichen Unterschieden soweit voneinander, dass eine Rekonstruktion der zugrundeliegenden Objekte s_i oder den klanglich äquivalenten Varianten $\tilde{s}_{i,j}$, etwa durch Linearkombination der Abmischungen $\tilde{m}_j$, nicht möglich ist.

Denkbar wäre nun, dass sämtliche zulässige Abmischungen $\tilde{m}_j$ einzeln komprimiert und gespeichert bzw. übertragen werden. Dies führt jedoch bei einer großen Zahl zulässiger Varianten zu einer sehr großen Datenmenge. Sinnvoll ist es daher, die Redundanz der Varianten bei der Codierung zu nutzen. Daher werden im Folgenden Gemeinsamkeiten x_i der Abmischungen $\tilde{m}_j$ nur einmal codiert. Dies geschieht, indem die Objekte s_i zunächst über eine weitere Transformation V_i verfälscht werden und anschließend codiert werden. Die Verfälschung, im einfachsten Fall durch Hinzufügen von Quantisierungsrauschen, ist dabei notwendig, da ansonsten im Decoder die Objekte s_i verlustfrei direkt aus x_i rekonstruiert werden könnten. Im nächsten Schritt erfolgt eine Transformation von x_i über $T_{i,j}$ und Abmischung mit $g_{i,j}$ zu $\hat{m}_j$, was der Dekodierung der Signale m_j entspräche, wären die Eingangssignale der Transformation $T_{i,j}$ nicht durch V_i verfälscht. Die Differenzsignale $r_j = \tilde{m}_j - \hat{m}_j$ werden daher im Decoder benötigt, um jedes beliebige $\tilde{m}_j$ verlustfrei rekonstruieren zu können. Die für alle Varianten benötigten Differenzsignale lassen sich mit geringer Datenrate codieren. Die benötigte Datenrate ist abhängig von den durch V_j verursachten Abweichungen und, im Falle von verlustbehafteter Codierung, der geforderten Rekonstruktionsqualität.

Im Decoder wird entsprechend Abbildung 3 in der Regel stets nur eine erlaubte Abmischung decodiert. Die Signale x_i werden dazu mittels der passenden Transformation $T_{i,j}$ zunächst in die Signale $\hat{s}_{i,j}$ überführt. Diese weichen von den Signalen s_i (auf Grund der Funktion V_i im Encoder) soweit ab, dass sie für eine direkte, nicht-intendierte Rekonstruktion der Signale s_i nicht geeignet sind. Erst nachdem die Signale $\hat{s}_{i,j}$ im nächsten Schritt im intendierten Mischungsverhältnis kombiniert wurden, und mit dem geeigneten Residuum r_j korrigiert wurden, ergibt sich ein optimal rekonstruiertes Mischsignal. Wird ein falsches Mischungsverhältnis oder ein falsches Residuum verwendet, ergibt sich zwangsläufig ein klanglich degradiertes Ausgangssignal.

Ausblick

Das in diesem Artikel beschriebene Verfahren wurde zunächst mittels einfacher linearer und nicht-linearer Transformationen für $T_{i,j}$ und V_i getestet. Insbesondere wurden Varianten mit Phasenverschiebungen, Quantisierungen, sowie additivem weißen Rauschen untersucht. Zur Weiterentwicklung des Systems beabsichtigt der Autor die Integration des Verfahrens in bestehende, verlustbehaftete Audio-Codecs. Zudem wird ein Fokus auf der Optimierung der Transformationen $T_{i,j}$ und V_i in Hinblick auf eine geringe Datenrate bei gleichzeitiger Sicherstellung einer hinreichenden Qualitätsabsenkung bei nicht-intendiertem Gebrauch der Objekte liegen.

Zusammenfassung

Das hier vorgestellte Verfahren zum Schutz von Multimedia Content, ermöglicht Autoren und Produzenten die Kontrolle der Verwertung des dem Kunden zur Verfügung gestellten Audio-Materials. Ein besserer Urheberschutz ist damit insbesondere für objektorientiertes Multimedia-Material möglich. Mit dem beschriebenen System lassen sich die intendierten Kombinationen in voller Qualität rekonstruieren, nicht jedoch einzelne Tonspuren. Da lediglich die Unterschiede zwischen den intendierten Kombinationen zu den Mischungen der Einzel-Tonspuren codiert werden, ergibt sich eine ausreichende Datenreduktion, womit das Verfahren auch bei der Bereitstellung vieler Kombinationsmöglichkeiten effizient ist.

Literatur

[1] ISO/IEC, Int. Std. 11172-3: Coding of moving pictures and associated audio for digital storrage media at up to about 1.5 Mbit/s, Part 3: Audio. (1993)

[2] Terentiev, L.; Falch, C.; Hellmuth, O.; Herre, J.: Efficient Parametric Audio Coding for Interactive Rendering: The Upcoming ISO/MPEG Standard on Spatial Audio Object Coding (SAOC). NAG/DAGA (2009), 1115-1118

[3] Brandenburg, K.: Vorlesungsskript „Aktuelle Projekte der MPEG: MPEG -A, -B, -C, -D, -E, -V", URL: `http://www.tu-ilmenau.de/fakei/fileadmin/template/i/mt/Multimedia_Standards/VL_MMS_14_MPEG-A-V.pdf`

[4] Anderson, M.: Media: The Mash Monsters, IEEE Spectrum (2008), 22-23,