

AnnoTone : 高周波音の映像収録時埋め込みによる編集支援

鈴木 良平 坂本 大介 五十嵐 健夫 *

概要. GPS 座標値や演技の良し悪し、試合状況といった注釈情報をビデオ録画時に埋め込んでおくことによって、カット処理やタイミングに合わせたエフェクトの生成などビデオ編集を支援する手法を提案する。具体的には、注釈情報を表すデジタルデータを人間には聴き取りづらい高域の音声信号に変換し、その信号をビデオカメラの近くに置いたデバイスのスピーカーから送出する。その音声信号を映像とともに記録しておき、編集時にその情報を抽出することで、編集を支援する。本手法で埋め込まれる音声信号は完全に不可聴ではないが、フィルタにより音質を大きく落とすことなく音声信号から取り除くことができる。また、注釈信号埋め込みの信頼性や音声データの変換に対する注釈信号の耐久性を、実際の利用場面を想定した実験により評価し、本手法が高い実用性を備えることを検証した。本論文では提案手法を使用した複数の映像編集アプリケーションを例示し、注釈情報を用いることで実現される様々な新しい映像制作技法の可能性を議論する。

1 はじめに

YouTube をはじめとした動画投稿サイトおよび高性能なデジタルカメラの普及により、映像を撮影し共有する行為は、多くの人々にとって日常的な楽しみの一つになりつつある。実際、動画投稿サイトで検索を行うと、楽器や料理のレクチャービデオやツーリングを記録した映像など、数多くの一般ユーザーによって製作されたコンテンツを見ることができる。こうした映像共有の大衆化を背景として、撮影された映像から手軽に魅力的な映像コンテンツを作り出せるような、非プロフェッショナル向けの映像編集ツールへの需要が高まっている。

映像コンテンツの制作において、演技の良し悪しや外部騒音、映っている人物などといった撮影時の文脈的情報を反映するような編集や特殊効果の追加は、映像内容の見直しや別途用意した記録との照合が必要となるために時間が掛かる工程である。また、映像に撮影場所を示す地図を重畳するために GPS 測定値を利用したい場合など、カメラに搭載されていないセンサー由来のデータを用いる場合には、編集時に映像シーケンスとセンサーデータ系列を対応付けする必要がある、やはり大きな手間が掛かる。本研究は、ビデオ撮影時に映像に直接注釈データを埋め込む手段を提供することで、このような文脈的情報を用いた編集を支援することを目的とする。

通常のビデオカメラは 44.1 kHz のサンプリングレートでの録音が可能であり、約 22 kHz までの周波数の音声を記録することができる。一方、人間の

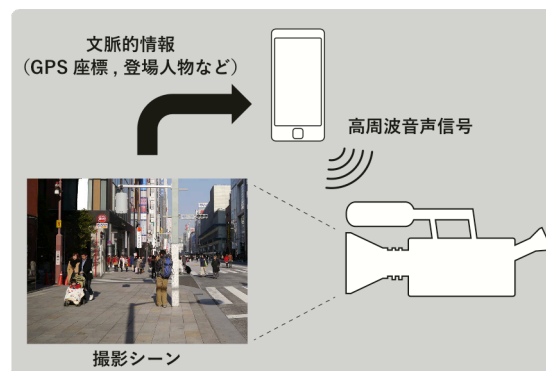


図 1. AnnoTone の概念図。シーンに含まれる文脈的情報をスマートフォンで音声に変換しビデオに吹き込む。

可聴域は 20 Hz から 20 kHz と言われており、個人差はあるものの、多くの人には 18 kHz 程度までしか聴き取れないことが知られている[1,2]。したがって、ビデオカメラは人間が聴き取れない程度の高周波音を記録できるため、デジタルデータをこの帯域の音声信号に変調することで、本来の音声コンテンツを損ねることなく撮影データ中に保存することが出来る。本稿ではこうしたビデオカメラの特性を利用し、映像撮影時の文脈的情報をスマートフォンによって取得した後、高周波音に変換してビデオカメラのマイクから吹きこむことで、録画映像の音声トラックに注釈情報を埋め込む手法を提案する(図 1)。

今回我々は、注釈埋め込みのためのスマートフォン用ライブラリおよび注釈抽出のためのソフトウェアをセットとしたシステム“AnnoTone”を開発し、これを実際の映像制作ワークフローに導入して複雑な編集を簡単に行えるようにすることで、本手法の有用性を検証した。また、本手法が実用に耐える信頼性を持つことを実験で評価するとともに、埋め込

Copyright is held by the author(s).

* Ryohei Suzuki, Daisuke Sakamoto, Takeo Igarashi, 東京大学

まれる注釈信号がコンテンツの品質を大きく落とさないことを、ユーザーテストにより検証した。

2 関連研究

2.1 音声透かし

音声データに対して、存在を知覚することが難しい形態で情報を埋め込む技術は、音声透かし (Audio Watermarking) と呼ばれ、Bender らによる位相コーディングやスペクトラム拡散などの基本的な手法の提案[3]を皮切りに 1990 年代後半より広く研究が行われてきた。音声透かし技術の多くは映画や音楽への著作権情報の埋め込みによる海賊版利用の防止を目的に設計されているため、既成の音声データに対してオフラインに埋め込み処理を行う利用形態が主である。一方、近年ではライブパフォーマンスの不正録画防止等の目的で、音声収録と透かし埋め込みとを同時に行う手法が研究されており、例えば Tachibana による Sonic Watermarking[4]は、演奏と透かし信号を別々のスピーカーから発音させて空气中で混合することで、会場で録音したデータへの情報埋め込みを実現するものである。本稿の手法もこれに近い方法を用いており、聴取対象のコンテンツは通常通りマイクで録音しつつ、スマートフォンから注釈信号を発して空气中で混合している。

音声透かし技術は知財保護の他、広告や放送媒体でのデバイス間通信のためにも用いられている。Matsuoka らは、音声信号を加工することで音声放送に URL や文字情報を重畳する acoustic OFDM[5]という技術を開発した。リスナーはスマートフォンなどのレシーバを用いて放送音声から楽曲情報などを取得することが出来る。類似の技術は美術館展示やコンサートなどでのインタラクティブ要素の導入にも利用されている。森勢らによる Clairvoyant[6]は、高音域帯を用いて楽曲に情報を埋め込み、鑑賞者のスマートフォンに歌詞等を呈示するシステムである。Hirabayashi らはライブハウスにおける演奏者と聴衆とのインタラクションのために、会場のスピーカーと手持ちのスマートフォンとの通信を可能にする Cryptone[7]というシステムを提案している。このシステムは高音域の正弦波の重ねあわせにより識別子を表現する DTMF を実装しており、本稿で提案する手法ではこれを任意のデータを表現できる通信プロトコルへと拡張したものを利用している。

本研究は、以上のような音声透かし技術を、録画時に生成した注釈情報を利用したビデオ編集支援という異なる目的に利用する点に新規性がある。

2.2 ビデオ注釈

動画ファイルに注釈情報を添付するためには、

Adobe XMP[8]などのメタデータファイルを用いる手法が一般化されているが、本稿の提案手法はこうした特別な形式のファイル処理に対応していないビデオカメラでも利用できるという長所があり、より多様な利用ケースに対応することが出来る。

Watanabe らによる WillCam[9]は、撮影時の位置や気温、環境音や撮影者の表情、写真中の重要なオブジェクトといった情報を保存する機能を備えたデジタルカメラであり、写真自体に含まれない文脈的情報を記録して閲覧や編集に活かすという点で提案手法と目的を共有している。一方、WillCam は専用のハードウェアを用いて写真データを直接操作することで、画像データに注釈情報を合成したりメタデータを書き込んだりするのに対して、提案手法は撮影データに対して直接アクセスせずに、不可聴な音声を重畳することで映像データに情報を埋め込む点で相違がある。また、WillCam は写真を対象としているのに対して AnnoTone は映像を対象とする。

ContextCam[10]は、ホームビデオのアーカイブ制作を主な目的とした、注釈付き映像を撮影することのできるビデオカメラである。ContextCam には GPS レシーバや超音波エコーなど多くのセンサーが搭載されており、撮影時のデータを記録するとともに機械学習を適用することで「誕生日会」のような高次の場面状況までを録画映像に埋め込むことができ、本提案と同様の目的を実現している。一方、ContextCam も WillCam 同様に特別なハードウェアを用いたシステムで、利用可能なシーンは限定されている。これに対して AnnoTone は任意のビデオカメラと組み合わせ利用可能であり、埋め込むデータの種類もスマートフォンで扱える限り自由に選ぶことが出来るという点で、性質に違いがある。

この他、Sony 社の HDR-AS30V/B[11]を始めとするアクションカムには GPS センサー等を備え、映像と同時に記録した後に閲覧や編集に利用できるものが存在する。本手法はこうした特別な機能を備えない既存のビデオカメラに対して、必要に応じて同様の機能を追加することを可能とするものである。

3 システムの構成

AnnoTone システムは、信号生成器、検出器および除去器によって構成される。生成器は Android 用のソフトウェアライブラリとして提供され、センサーやユーザーが入力したデジタルデータを高周波信号に変換、スピーカーから発音する。検出器は Java プログラムとして提供され、映像ファイルを解析して埋め込まれた注釈情報を抽出して出力する。検出器は独自の映像編集ソフトウェアを開発する際に組み込んで用いることができる他、既存のオーサリン

グソフトウェアとデータを連携して利用することも可能である．除去器は 17 kHz をカットオフ周波数とするローパスフィルタであり，完成した映像コンテンツから本システムがデータコーディングに利用する高周波帯域の音声信号を除去する機能を持つ．

本システムの基本的な利用シナリオは次のようになる．ユーザーはスマートフォンをゴムバンドやテープなどを用いてビデオカメラに固定し（図 3），信号生成器を組み込んだアプリを起動する．その後，必要に応じてアプリを操作しつつ映像撮影を行い，注釈の埋め込みを行う．収録後，PC にインポートされたビデオは映像編集アプリケーションに読み込まれる．この際，検出器が注釈情報を抽出してアプリケーションに提供することで，ユーザーはカット処理の自動化など編集支援の恩恵を受ける．編集が完了したタイミングで，除去器が映像データから注釈信号を除去し，完成版の動画ファイルが得られる．



図 3. AnnoTone 使用時のハードウェア構成の例

4 データコーディング手法

本セクションでは，提案手法におけるデジタルデータの音声信号への変換方式を説明する．本方式は高周波音を用いた Dual-tone multi frequency (DTMF) 信号を採用しているが，この基本アイデアは先行研究[7,12]で紹介されており，本研究で開発した手法は，このアイデアを汎用的なデータ表現プロトコルへと拡張したものである．

4.1 高周波数帯での音声 DTMF

AnnoTone は，DTMF と呼ばれる変調手法を用いて，入力されたデジタルデータをほぼ不可聴な音声信号に変換する．AnnoTone の単位信号は，17.6 kHz から 20 kHz までの 7 種類の周波数からなるサブキャリアのうち 2 つの単音の重ねあわせで表現され（図 4），この組み合わせによって情報を表現する．可能な組み合わせは 21 通り（ $7C_2$ ）存在するが，このうち 16 通りを使用することで，単位信号あたり 4 ビットの情報を表現する．現在の実装では単位信号の長さは 10 ミリ秒に設定されており，400 bps のグロス通信レートを実現している．

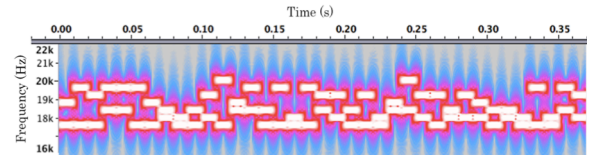


図 4. 16 バイト長の注釈信号のスペクトログラム例．

4.2 パケット構造

各注釈データは直列化され，単位信号の連続からなる一つのパケットに構造化される．パケットは定まった値を示すスタートフレームから開始し，これに続く長さフレームがペイロード長を指示する．第三シグナル以降はペイロード領域が指定された長さだけ続き，注釈の種類を示す ID と CRC-8 チェックサムで終わる（図 5）．単位信号あたり 4 ビットの情報を表現できるため，ペイロードはバイトごとに低位 4 ビットと高位 4 ビットの 2 つの単位信号に変換される．ID は 1 つの映像ファイル中に複数種類の注釈（GPS 座標と加速度値など）を埋め込んだ際，それらを区別するために用いられる．

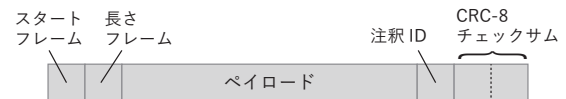


図 5. 注釈信号のパケット構造の模式図．

4.3 信号の生成と抽出方法

DTMF 信号は次のようなプロセスで生成される．まず，注釈データがバイト配列に直列化される．次いでバイト配列は単位信号配列に変換されるとともにペイロード長，ID およびチェックサムが付与され，パケットを構成するデータストリームが準備される．このストリームから次の式に従って波形を生成する．

$$s(t, m) = w(t) \cdot a(m, d \left\lfloor \frac{t}{0.01} \right\rfloor) \cdot \sin(2\pi t \cdot f(m)) \quad (1)$$

ここで $s(t, m)$ は m 番目のサブキャリアの t 秒時点でのサンプル値に対応する． $f(m)$ は m 番目のサブキャリアの周波数 (Hz) であり， $d(n)$ は配列中の n 個目の単位信号の値である． $a(m, x)$ は DTMF 変調関数であり，次のテーブルの (m, x) 番目の要素に対応する．

表 1. DTMF 変調関数

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
7						1				1					1	
6					1				1					1		
5				1					1				1			1
4			1					1				1				1
3		1					1					1	1	1	1	
2	1						1	1	1	1	1					
1	1	1	1	1	1	1										

$w(t)$ は単位信号の境界部分での低周波雑音の発生を抑えるためのエンベロープフィルタであり，次のよう定義したものをを用いた．

$$w(t) = \left(0.5 - 0.5 \cos \left(2\pi \frac{t \bmod 0.01}{0.01} \right) \right)^{0.8} \quad (2)$$

注釈信号 $x(t)$ は最終的に各サブキャリアの信号値を足し合わせることで生成され、スマートフォンに搭載されたスピーカーから発音される。

DTMF は周波数偏移変調 (FSK) の亜種として見なすことが出来るため、抽出は FSK 用の一般的な復調手法を用いて行うことが出来る。現在の実装では、各サブキャリア周波数のバンドパスフィルタを適用してから、適応的二値化を行なっている。

5 映像編集支援手法とワークフロー

本セクションでは、3つの AnnoTone を組み込んだ映像編集アプリケーションの例および、既存の映像オーサリングツールへの組み込み手法を説明する。

5.1 収録時映像編集

映像の編集プロセスでは、制作者はシーンの削除やつなぎ合わせなどの作業を行うが、これには何度も映像を確認する必要がある。例えばビデオ講義の収録を行うケースでは講師が説明を間違える度に撮り直しを行う必要があるが、撮影後には大量の動画断片が生成されることになる。これらを選別し一本の映像にまとめるのには大変な手間が掛かる。

今回我々は、映像の撮影時に出演者の演技正否を逐次記録し、編集時にこの情報を用いて自動的に成功した部分だけ繋ぎあわせた動画を生成できるアプリケーション (図 6 右) を実装した。カメラ操作者が図 6 左に示すスマートフォンのボタンを押すことで、各時点での演技の正否が埋め込まれる。本アプリケーションを用いると、演技に失敗してもビデオカメラの録画停止・再開操作を繰り返す必要なく、録画状態を維持したままで納得できるまで演技することができる。これにより、大量のファイルの管理や編集に必要な手間を省くことが出来る。

5.2 試合状況の自動画像オーバーレイ

多くの映像コンテンツは、状況説明や補足的情報を提供するために、オーバーレイ画像を用いている。例えば球技の中継映像には、スコア数などの表示ボードが常に表示されている。こうしたオーバーレイを映像内容と同期して手作業で作成するのは手間が掛かるが、映像収録中に試合状況など必要な情報を注釈しておくことで、自動生成が可能となる。

我々は、チェスの試合映像を撮影する際に棋譜を記録しておくことで、編集時に全自動で盤面状況の CG アニメーションをオーバーレイできる (図 7 右) アプリケーションを開発した。カメラ操作者がチェスゲームを模したスマートフォン画面上の UI (図 7

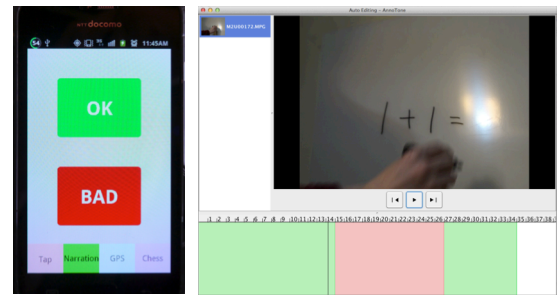


図 6. 左：演技正否入力用 UI. 右：編集用アプリケーション。

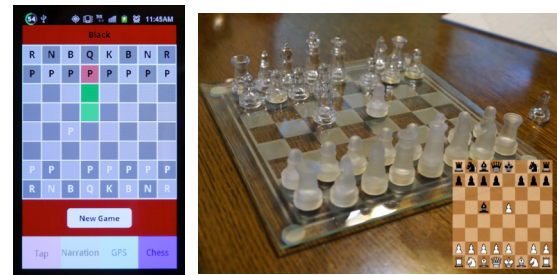


図 7. 左：棋譜入力用 UI. 右：グラフィックが付与された映像。



図 8. GPS データにより地図がオーバーレイされた映像。

左) を操作すると、標準的なチェス棋譜データ (FEN 形式) の注釈が動画に埋め込まれる。

5.3 GPS 座標を用いた映像編集

旅行ビデオやツーリング記録をはじめ、カメラが移動しながら映像が撮影されることはよくあるが、こうしたビデオでは各瞬間のカメラの位置情報が内容において重要な意味を持つ。我々は映像に数秒おきに GPS 座標値を埋め込むことで、ビデオと位置情報の系列を対応付け、地図画像および移動の軌跡を自動的に映像に合成するアプリケーションを開発した (図 8)。本システムは録画映像から GPS 座標系列を抽出し、Google Maps API から必要な地図画像を自動的にダウンロードして合成に用いる。

5.4 Adobe After Effects との連携

ここまでの事例では、専用の映像編集アプリケーションを開発することで注釈を用いたが、多くのユーザーは市販の映像編集ツールを用いており、AnnoTone も既存のワークフローに組み込んで利用することが望ましい。幾つかの編集ツールには、

映像効果の生成等を自動化するエンドユーザー用のスクリプティング機能が備わっているが、抽出された注釈情報をこうしたスクリプトから参照できるようにすることで、注釈情報を用いた映像編集に既存のノウハウを活用することができる。

今回我々は、Adobe After Effects (AE) に信号検出器の出力を読み込む拡張機能を作成し、各注釈を JSON データを含むキーフレームに変換することで、AE の標準機能を用いた映像制作に注釈データを利用できることを検証した。例えば、車載カメラ映像に埋め込んだ GPS 座標値等の注釈情報をキーフレームとして読み込み、これと AE で作成した速度計のアニメーションをスクリプトにより対応付けするなど、文脈情報を用いた特殊効果を含む映像コンテンツを効率的に作成することが可能である。

6 性能評価

提案手法の有用性を示すために、注釈埋め込み方法の信頼性およびコンテンツの品質維持を検証する性能評価を実施した。本評価においては家庭用ハンディカム (Sony NEX-VG30H) とスマートフォン (Samsung Galaxy S) をハードウェアに用い、音声は 44.1 kHz 16-bit ステレオ PCM 形式で記録した。

6.1 データレートの上限評価

大きいデータを注釈信号として埋め込むためには単位信号長を短くする必要がある。一方、信号長が短いほどにデコードのエラー率が上がるため、十分なデコード成功率 (CDR) を確保するためにはデータレートに上限が存在する。6 ミリ秒から 10 ミリ秒まで単位信号長を変えながら、それぞれ 4 種類の音響環境条件で録音実験を行い、デコードの成功率を測定した。音響環境としては 1) 静かな空間、2) 公共空間、3) ロック音楽のような大音量環境、および 4) 電子音が干渉する可能性がある電子音楽の環境である。図 9 が示すように、単位信号長 10 ミリ秒では全ての環境条件で十分な信頼性が得られているが、これより短くなると CDR が顕著に低下した。

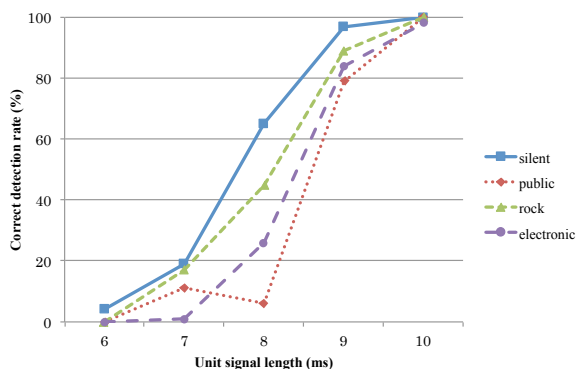


図 9. 単位信号長に対するデコード成功率の変化

6.2 メディアフォーマット変換への耐性

映像編集工程ではメディア形式の変換が行われることがあるが、最終工程で信号除去を行うまで注釈情報が維持されることを検証するため、形式変換による注釈情報の損失率を測定した。実験には MP3, Ogg Vorbis および AC-3 (Dolby Digital) の三種類のフォーマットを用い、それぞれ 128 kbps から 320 kbps まで四種類の音質で音声ファイルを圧縮した後、抽出可能であった注釈の割合を計測した。

結果、図 10 に示されている通り、Ogg Vorbis および AC-3 形式では中程度 (192 kbps) 以上の音質で圧縮した際にはほぼ完全に信号が維持されることが分かった。一方、MP3 では高音質設定でもなおデコード失敗率が高くなったが、これは高音域成分を除去しやすい圧縮アルゴリズムの特性によるものと考えられる。これらの観察より、適切な形式を選択することで変換を行っても注釈を損なわずにデータ運用が可能であると言える。

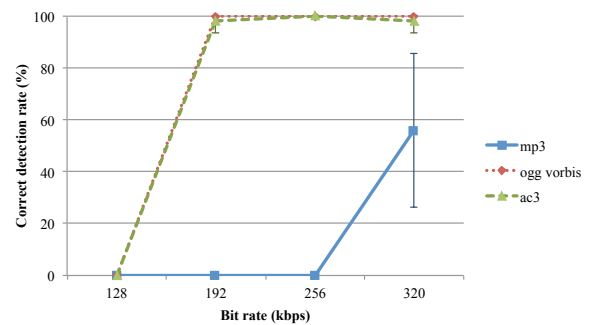


図 10. メディア形式変換に対するデコード成功率の変化。

6.3 注釈信号による音質変化の主観評価

注釈信号がコンテンツの音質を損ねることがないかを検証するため、5 名の被験者 (男性 4 名、女性 1 名、平均 36.2 歳) の協力によりユーザーテストを実施した。被験者には、注釈信号を複数埋め込んだ四種類の音声データ (無音・公衆騒音・ロック音楽・電子音楽) をヘッドフォンで聴き、不自然な音を感じた瞬間に手元のボタンを押すよう指示を行った。聴取時の音量は予め被験者が快適に感じる大きさに自由に調節してもらい、各音声データについて正しく知覚された注釈信号の個数を集計した。また、信号除去器の効果を検証するために、除去処理済みの音源についても同様の実験を行った。

図 11 の結果の通り、概ね知覚された注釈信号の割合は 20% 程度であり、聴取体験に大きい影響を及ぼすことはなかったが、過去の研究事例 [2] の通り、若い人ほど気づきやすい傾向が見られた。また、信号除去器を用いることでほぼ完全に知覚を不可能にできる事が検証できた。これらの結果より、提案手法を映像コンテンツの制作において使用することで、

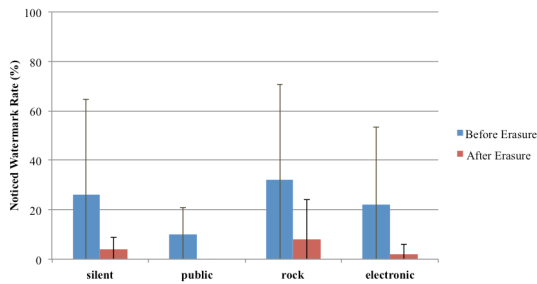


図 11. 被験者により気づかれた注釈信号の割合。

作品の品質を落とすことはない結論できる。

6.4 デコード速度

現在の Java による信号検出器の実装では、標準的な PC (2 GHz Intel Core i7, 8GB RAM) で約 11 倍速のデコード速度を得られているが、SIMD の利用等のプログラム改善によりさらなる高速化も期待でき、実用上差し障ることはないと考えられる。

7 まとめと今後の課題

本稿では、デジタルデータを不可聴音に変換して記録することで収録時に注釈情報をビデオに埋め込む手法および、これを抽出して編集に利用する手法を提案し、文脈情報を利用した動画編集の支援の可能性を議論した。提案手法を利用することで、特別なカメラやハードウェアを購入できない一般ユーザーでも様々な文脈の情報を用いて魅力的なコンテンツを効率的に制作できると期待される。

本研究の今後の発展としては、Cryptone のように大型スピーカーから注釈信号を発音することで、公衆空間での映像撮影の拡張に利用できると考えられる。例えばスタジアム等に設置されたスピーカーから注釈を出力することで、スポーツ試合等を撮影した大量の動画の同期や対応付けを行い、ウェブサイトに掲載された動画群からの多視点映像の自動生成などに用いることができる可能性がある。

一方、提案手法の問題点として、実験結果が示しているように影響は大きくないものの、音声トラックに注釈を記録することで一部の鋭敏な視聴者には不快感を及ぼす可能性があることや、埋め込みのデータレートがあまり大きくないために応用の幅が制限されることが挙げられる。これらの問題については、今回は実装が簡単な高音域 DTMF を変調に用いたが、代わりに洗練された音声透かし手法を活用することで改善が望めると考えている。

また、搭載カメラの大幅な性能向上のため、最近ではスマートフォン単体で映像が撮影されることが多いが、こうしたケースでは本提案手法を用いるよりも、Adobe XMP などのメタデータに対応したカ

メラアプリを用いた方が便利である場合が多いと思われる。しかし、GoPro をはじめリアルカメラの需要は画質や機能などの点から依然として確固としたものであり、スマートフォンの普及が提案手法の有用性を否定するものではないと考えている。

参考文献

- [1] K. Ashihara, K. Kuraakata, T. Mizutani, and K. Matsushita, "Hearing threshold for pure tones above 20 khz," *Acoustical science and technology*, 27, 1, pp. 12–19, 2006.
- [2] P. G. Stelmachowicz, K. A. Beauchaine, A. Kalberer, and W. Jesteadt, "Normative thresholds in the 8 to 20khz range as a function of age," *J. Acoust. Soc. Am.*, 86, 4, pp. 1384–1391, 1989.
- [3] W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for data hiding," *IBM systems journal*, 35, 3.4, pp. 313–336, 1996.
- [4] R. Tachibana, "Audio watermarking for live performance," In *Electronic Imaging 2003*, pp. 32–43, 2003.
- [5] H. Matsuoka, Y. Nakashima, T. Yoshimura, and T. Kawahara, "Acoustic ofdm: Embedding high bit-rate data in audio," In *Advances in Multimedia Modeling*, pp. 498–507, 2008.
- [6] 森勢将雅, 山下洋一. モスキート音に基づく電子透かしを用いた歌詞連動型音楽試聴システムの提案. 情報処理学会研究報告[音楽情報科学], 96(4), pp. 1-6, 2012.
- [7] M. Hirabayashi and M. Shimizu, "Cryptone: Interaction between performers and audiences with inaudible dtmf sounds," In *SIGGRAPH Asia 2012 e-tech*, 5:1–5:4, 2012.
- [8] Adobe XMP: Adding Intelligence to Media. <https://www.adobe.com/products/xmp/>.
- [9] K. Watanabe, K. Tsukada, and M. Yasumura, "Willcam: A digital camera visualizing users' interest," In *CHI '07 Extended Abstract*, pp. 2747–2752, 2007.
- [10] S. N. Patel and G. D. Abowd, "The contextcam: Automated point of capture video annotation," In *UbiComp 2004*, pp. 301–318, 2004.
- [11] Action Cam with GPS - HDRAS30V/B Review <http://store.sony.com/compact-pov-action-cam-zid27-HD-RAS30V/B/cat-27-catid-All-Action-Cam>
- [12] R. Healy and J. Timoney, "Digital audio watermarking with semi-blind detection for in-car music content identification," In *AES 36th International Conference: Automotive AudioSound in Motion*, 2009.