



Title	音声信号特徴に基づく映像要約再生手法に関する研究
Author(s)	大塚, 功
Citation	大阪大学, 2014, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/34516">https://hdl.handle.net/11094/34516</a>
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、<a href="https://www.library.osaka-u.ac.jp/thesis/#closed">大阪大学の博士論文について</a>をご参照ください。

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

## 論文内容の要旨

氏名 ( 大塚 功 )	
論文題名	音声信号特徴に基づく映像要約再生手法に関する研究
論文内容の要旨	
<p>本研究は、放送番組の録画再生装置において、録画した番組の速覧を実現するための映像要約再生手法の研究に取り組んだものである。研究成果は民生用AV機器製品に適用することを目標にしており、そのため分析に要する処理負荷が低い音声信号から特徴を抽出し、所望とする映像の重要シーンを高い確度で検出できる映像要約再生システムを提案する。</p> <p>録り貯めた膨大な映像コンテンツを速覧するためのアプリケーションとしては、重要なシーンを連続視聴する映像要約と、見たいシーンを素早く探す映像検索などがあり、これらの機能を実現するにはコンテンツ内の各シーンの重要度合を録画再生システムが自動推定して把握している必要がある。民生用AV機器の限定されたハードウェアリソースで本機能のリアルタイム処理を実現するには、重要度合推定のための処理負荷の軽減が課題となる。そこで本研究では映像信号と比べてデータ量が少ない音声信号に着目し、音声信号からローレベルな符号的特徴を抽出してハイレベルな内容情報に置き換えて重要度を推定し、ハイライトシーンを特定する手法を提案した。</p> <p>音声信号の符号的特徴には、MDCT係数値を採用した。アナウンサの絶叫や歎声など、ハイライトシーンに特有な音声のMDCT係数値を混合ガウス分布(GMM)でモデル化した音声クラスと尤度が高いシーンに高い重要度レベルを与えることで、スポーツ番組のハイライトシーンを自動検出する手法を提案した。また、音楽シーンのMDCT係数値をモデル化した音声クラスと尤度が高く且つ左右チャンネルのオーディオエネルギー差分値が高いシーンを抽出することで、音楽番組の楽曲シーンを自動検出する手法を提案した。さらに、MDCT係数値から算出したオーディオエネルギーと映像のシーンチェンジの出現傾向から放送本編と本編以外のシーン境界を自動判別する手法を提案した。これらの手法を実現するためのフレームワークを構築し、アプリケーションとしての機能の有効性と実用化の可能性を示した。</p>	

## 論文審査の結果の要旨及び担当者

氏 名 ( 大塚 功 )		
	(職)	氏 名
論文審査担当者	主査	教授 西田正吾
	副査	教授 飯國洋二
	副査	教授 佐藤宏介

## 論文審査の結果の要旨

本論文は、放送番組の録画再生装置において、録画した番組の速覧を実現するための映像要約再生手法の研究に取り組んだものである。研究成果は民生用AV機器製品に適用することを目標にしており、そのため分析に要する処理負荷が低い音声信号から特徴を抽出し、所望とする映像の重要シーンを高い確度で検出できる映像要約再生システムを提案している。

本論文では映像信号と比べてデータ量が少ない音声信号に着目し、音声信号からローレベルな符号的特徴を抽出してハイレベルな内容情報に置き換えて重要度を推定し、ハイライトシーンを特定する手法を提案している。具体的には、音声信号の符号的特徴には、MDCT係数値を採用するとともに、アナウンサの絶叫や歎声など、ハイライトシーンに特有な音声のMDCT係数値を混合ガウス分布(GMM)でモデル化した音声クラスと尤度が高いシーンに高い重要度レベルを与えることで、スポーツ番組のハイライトシーンを自動検出する手法を提案している。また、音楽シーンのMDCT係数値をモデル化した音声クラスと尤度が高く且つ左右チャンネルのオーディオエネルギー差分値が高いシーンを抽出することで、音楽番組の楽曲シーンを自動検出する手法を提案している。さらに、MDCT係数値から算出したオーディオエネルギーと映像のシーンチェンジの出現傾向から放送本編と本編以外のシーン境界を自動判別する手法も提案している。また、これらの手法を実現するためのフレームワークを構築し、アプリケーションとしての機能の有効性と実用化の可能性を示している。

以上のように、本論文は実用性の高い音声信号に基づく映像要約再生手法を提案しており、博士（工学）の学位論文の価値があるものと認める。