

関連情報自動検索機能付き映像再生システムの開発

Development of video playing system overlaying automatically collected Information

高橋 徹(TAKAHASHI Toru)

本研究の課題は、テレビやDVDなどのビデオ映像に関連する情報を視聴画面上に自動的に表示可能な映像再生システムを開発することである。映像中の出演者やBGMなどに関するテキストまたは画像などの可視情報を関連情報と呼んでいる。本研究のキーアイデアは、関連情報を自動取得するために、ビデオ映像の音声チャンネルを利用する点である。本年度は、ビデオ映像の音声チャンネルに乗せられている音声信号を解析し、新たな特徴量を設計した。音声信号の断片を検索キーとして、当該区間の特徴をリアルタイムにデータベースに問い合わせ、問い合わせ結果を関連情報として選択する方法を検討した。

技術的課題は、数秒程度の短い音声信号断片から、その信号を構成する音楽や音声を聞き分けることである。従来から、音声認識・話者識別・楽曲検索などが研究されているものの、本課題に直接適用することは困難である。なぜならば、テレビやDVDなどのビデオ映像の音声信号は、音声や楽曲など複数の音源が混合された信号であり、従来法の適用条件「入力信号は非混合信号」という条件を満たさないためである。

平成25年度の具体的な目標は、「音声信号から音声信号の類似検索手法を開発すること」であり、以下3つの副目標を設定し、課題に取り組んだ。

- (1) 音声ファイルとBGMファイルを収集し、擬似TV番組音声を作成し、システム評価用データベースを構築
- (2) 擬似TV番組音声から、楽曲データを検索する手法を開発
- (3) 音源分離、ノイズ除去、ミッシングフィーチャ理論を導入した音響特徴量を設計

これら3つの副目標を実現し、開発目標となるシステムのベースラインシステムを開発した。評価用データベースを構築したことから、ベースラインシステムの検索性能を計測可能となった。今後のシステム改良による、評価結果と比較可能となった。

副目標(1)に対して、音声認識用のコーパスから読み上げ文章を収集し、音声ファイルを用意し、音楽CDから音楽ファイルを用意し、これらを任意の混合比で混合する擬似TV番組音声生成システムを開発した。このシステムにより、100音声発話と100楽曲を3種類の混合比率で混合した擬似TV番組音声データベースを構築した。

副目標(2)に対しては、累積フレーム間の重み付きユークリッド距離を用いた類似度を設計することにより、混合音を検索キーとした場合においても、適切な楽曲を頑健に検索できることを確認できた。

副目標(3)に対しては、従来の音声認識や楽曲検索に用いられていた音響特徴量を本研究課題用に改良し、クロマバンク特徴量およびクロマケプストラム特徴量を開発できた。

以上の報告通り、計画通りに研究を遂行できたと考える。