

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。再生される音声を検索キーとして、データベースからBGMを特定し、特定した楽曲情報から関連情報を検索するという技術課題である。26年度の目標は、混合音を構成する各構成音の混合比率の影響を受けにくい音響特徴量を設計することであった。また、その特徴量は、データベースと検索キーとの間で高速比較可能な表現形式の実現が必要である。25年度に設計したクロマバンク特徴量とクロマスペクトラム特徴量をベースに新たな特徴量（バイナリクロマスペクトル）を設計し、混合比率に影響を受けにくい音響特徴を開発した[1]。この特徴は、1秒当たり720bitという極めて少ない情報表現で表現できる。クロマバンク特徴量やクロマスペクトラム特徴量と比較して、32分の1程度に情報表現を効率化し、かつ混合音の検索性能を向上している。

情報圧縮しているにも関わらず検索性能が向上する理由は、音量の大きい周波数帯域の情報と音量の小さい周波数帯域の情報が、他の音の影響を受けにくい性質を情報圧縮によって実現したためである。圧縮した情報表現は、音の情報をビット列で表すことから、文字情報検索アルゴリズムを適用可能となり、連続信号の検索に比べ処理の高速化が容易になった。

下表にバイナリクロマスペクトルを用いた検索精度を示す。信号の相対振幅が1.0, 0.1, 10.0倍であっても精度が変わらないことを確認できる。つまり再生ボリュームの影響を受けにくい特徴がある。また雑音との混合比-5, 0, 5, 10 dBで比較すると、混合比の影響で精度が低下するものの、影響を受けにくい（従来法では、0%となる）特性も併せ持つ。検索キー長は、長いほど検索精度が向上する傾向がある。極めて短い2秒の検索キーを用いても0dB条件で50%の精度を達成できた。10秒であれば、0dBでも95%を達成できる。

以上の様に、混合音から必要な情報を抽出し高精度にパターンマッチングする枠組みを構築した。今後は大規模なデータベースを構築しアプリケーションを開発したい。

	信号の振幅（相対値）	1倍		0.1倍		10倍	
		10s	2s	10s	2s	10s	2s
音声と 楽曲の 混合比	-5dB	52%	17%	52%	17%	52%	17%
	0dB	95%	50%	95%	50%	95%	50%
	5dB	98%	74%	98%	74%	98%	74%
	10dB	100%	88%	100%	88%	100%	88%

[1]高橋 徹, "混合音を検索キーとした音楽検索のための高速特徴量比較手法の検討", 日本音響学会 2015 年春季研究発表会, 3-5-3, 日本大学, March, 16-18, 2015 (発表日 3/18).