

## 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 informatio

高橋 徹 (TAKAHASH 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った。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。

出演者音声と BGM が重なって放送される音を音声と BGM に別けて情報を取得することを考えた。従来は音源分離 [1] という手法を用いて、音声と BGM の混合信号をそれぞれ個別音源に分離し、分離された信号に対して音声認識と音楽の楽曲認識を行う方法が用いられてきた。しかし、音声認識や楽曲認識には、信号を高精度に分離する必要がある。手掛かりなしに音源分離するブラインド音源分離では、高い精度で音声や楽曲に認識することが困難であるという課題があった。不十分な精度での音源分離は、最終的に音声認識や楽曲認識の精度の向上が見込めないという問題を内包している。

本研究では、音源を分離せず混合音から直接音声や楽曲を認識する手法を開発 [2] した。音源を分離せずに目的の音声または楽曲を認識する枠組みを提案し、その枠組みで高い認識精度を得るために、バイナリクロマスペクトルは、(1)音量に不変な特徴 (2)音色の特徴であるスペクトルの山と谷を保存する (3)他の音が重畳しても山と谷の構造を保ちやすい、という 3 つの特徴を併せ持つよう開発した。この特徴量空間上に距離を定義し、音声と楽曲が混合した音源から個々の音源を分離せずに楽曲を高精度に認識することを可能にした。一般に音源分離は計算コストが高い問題があり、音源分離をせずに処理可能な方法は、テレビ機器などの組み込み製品に向いているという長所がある。

音声と楽曲の混合音を入力とし、その混合比率に依存することなく高精度に楽曲の曲名と再生位置を検出することが可能になったことから、テレビや DVD などの音声トラックをモニタリングし、リアルタイムに BGM の曲名とその楽曲の再生位置をとらえることが可能となった。一旦、曲名を取得できれば、曲名を検索キーとして作曲者や作詞者をインターネット経由で取得できる。さらに作曲者を検索キーとして作曲者のプロフィールや活動を検索可能であり、テレビや DVD を再生しながら、そのコンテンツに関連のある情報をリアルタイムに画面に表示させることが可能になった。

[1] Xun Chen, et al, "Joint Blind Source Separation for Neurophysiological Data Analysis," IEEE Signal Processing magazine, May, 2016.

[2] 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.