E-048

ライブ演奏への応用を目的とした音楽情報アラインメントの特徴量別精度検証

佐藤博志 †　村岡洋一 ‡
† 早稲田大学基幹理工学研究科　‡ 早稲田大学理工学術院

1　はじめに

本研究では, ライブ演奏で起こりうる雑音や波形の非線形伸縮といった現実的な問題を想定し, それらに対する耐性が高い特徴量を実験によって調査する.

一般に, 音響システムの優劣は物理的・電気的な特性を基準に判断されるが, 実際は耳に入った音をそのまま聴いているわけではなく, 他の情報や過去の経験と照らし合わせるなどの脳内処理を経た後に, 最終的に音として認識されることが近年の研究によって確認されている. すなわち, ライブ演奏における感動をより強くするためには, 音楽それ自体を単体として提供するだけではなく, 過去の体験に基づいた映像と同時に提供する必要がある.

しかし, 現代のライブシーンにおいては, 従来から音響的な特質及び技術が重視される一方で, 人間の認識を構成するもう一方の要素である映像に対する認識が高いとは言えない. 特にクラシックコンサートでは以前から音楽のみが観客に提供されている状態である. これに対し, ポップミュージックに関しては音楽と映像が同時に提供される試みが前者に比べ目立つが, その際に提供される映像の多くはライブ特有の映像であることが多く, 臨場感を演出しているものの決して過去の記憶や体験に訴える映像を提供できているわけではない. そこで, 音楽と強く関係を持つ映像と, ライブ演奏で流される音楽を自動的に同期させる新規のシステムを提案する.

今回はアラインメント精度の検証と同時に映像同期システムも実装することで, ライブ演奏への応用の可能性も検証した.

2　提案手法

本システムでは, 事前準備として, ある楽曲と, その楽曲に雑音を加えるなどの加工を施した楽曲を用意する. 次にこれらの楽曲すべてに対し,HTK が提供する異なる特徴抽出方法を用いて, それぞれ特徴量を抽出する. 次に, クリーンな音楽 (参照データ) と, 雑音等が乗った音楽 (比較データ) の双方の特徴ベクトルを動的時間伸縮アルゴリズムによって解析し, 時系列に沿った対応を取ることで, 特徴量別にアラインメント精度を検証した. この時, 動的時間伸縮に関して次の方法で経路探索を行った.

$$g(i,j) = d(i,j) + \min \begin{bmatrix} g(i-1,j) \\ g(i-1,j-1) \\ g(i,j-1) \end{bmatrix}$$

また, 経路探索に行われる特徴量は, 同時刻における参照-比較データ間の特徴ベクトルの距離と定義した.

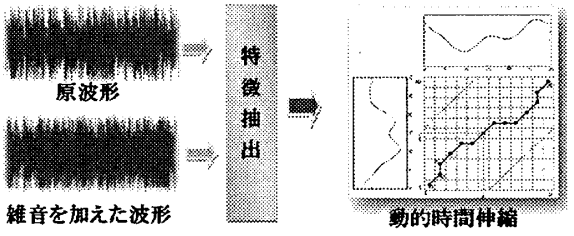

図 1: 提案手法

3　実験データ

使用した楽曲は,An der Schonen Blauen Donau(Blue Donub Walts), op.314 である. 雑音に対する耐性を測るために, これに歓声及び拍手を加えた. またそれとは別に, 空白に対する耐性を測る為に, 曲の途中に空白を挿入したデータも用意した. 更に, 演奏速度が異なる場合を想定して, 波形の非線形な時間伸縮を行ったデータを用意した. これは, 参照データを 3 つの部分に分け, それぞれの区間における再生速度を等倍速,1.5 倍速,0.6 倍速したものである.

また, 本実験に用いた特徴量は, 次の通りである.

MFCC(mel-frequency cepstral coefficients):12 次元
LPC(linear prediction filter coefficients):64 次元
LPCC (LPC cepstral coefficients) :12 次元
MELS(linear mel-filter bank channel outputs):24 次元
FBANK(log mel-filter bank channel outputs):80 次元

これに関して, 上記の特徴ベクトルの構成は, 各特徴量 (上記の次元数)+Δ Power (左と同等の次元数) +Energy (1 次元) とし, またその他のパラメータに関しても, サンプリング周波数:16 kHz, プリエンファシス:0.97, 分析窓:Hamming 窓, 分析窓長:25 ms, 窓間隔 10 ms, 周波数

分析等:メル間隔フィルタバンク, フィルタバンク:24 チャネルとして統一した. これは, それぞれの特徴抽出方法に対して平等にアラインメント精度を計測するためである.

4 精度検証比較実験とその結果

本実験において, 平均時間誤差と最大時間誤差の両面から定量的な誤差判定を行った.

雑音:歓声などのノイズによる時間誤差		
特徴量	平均誤差 [秒]	最大誤差 [秒]
MELS24	0.0000505	0.06
MFCC12	0.0027	0.11
FBANK80	0.0355	0.46
LPC64	0.11	0.16
LPCC12	0.18	1.63

表 1: 雑音への耐性

空白に対する耐性への誤差判定は, 時間誤差と視覚的な判断の両面から行った. その結果,FBANK は時間誤差は短かったが, 空白に突入しても空白とは認識できずに, 空白が終わってからその時間が空白であったことを認識した為, これを視覚的に不可とした.

空白挿入によって生じた時間語差		
特徴量	時間誤差 [秒]	視覚的判断
MELS12	0.000145	○
MFCC12	0.000145	○
LPC64	0.00192	○
FBANK80	0.00101	×

表 2: 空白への耐性

波形の非線形伸縮に関しては, 正解データとの面積差を用いてこれを計算した. 単位は HTK 尺度の時間である為, ここでは省略する.

時間伸縮による誤差	
特徴量	正解データとの面積差
MFCC12	10.208
LPCC12	10.761
MELS24	12.181
FBANK80	17.592

表 3: 非線形伸縮時のミスアラインメント

5 アプリケーションの概要

1. 音楽同期再生
   参照データの時刻を入力することで, 参照データの音楽と, その時刻に対応する比較データの音楽を再生する機能である. これによって, 正確にアラインメントが取れているかどうかを聴覚的に判断することができる. (音楽アラインメントツール"MATCH"[3] にも同様の機能が存在する.)

2. 映像同期システム
   本システムでは, 静止画を連続的に表示する手法で映像同期を実装した.1 秒間あたり 30 フレームの静止画を動画から切り出してバッファリングし, アラインメント結果から計算された速度比率をスレッドの待機時間と関連付けることで, 映像へと再編した.

6 まとめと今後の課題

以上の結果から, 雑音と空白挿入に対して最も優れた特性を持つのは,MFCC とゼロ平均正規化した線形メルフィルタバンク (※ゼロ平均正規化をかけなかった場合, 線形メルフィルタバンクだけでは雑音に対する耐性が低く, アラインメントを取ることが難しかった) であり, また波形の時間伸縮に対する耐性は, 正規化の有無に関わらず MFCC が最も高い精度であることが判明した.

このことから, ライブコンサートへの適用が期待できる特徴量は, 今回の範疇に於いて MFCC であると結論づけられる. これを用いて, 実用へ向けた改変と拡張を行うことで, 冒頭に述べたリアルタイム音楽映像システムの実現を目指す.

参考文献

[1] 後藤正孝, 橋口弘樹, 西村拓一, 岡　隆一, RWC 研究用音楽データベース：ポピュラー音楽データベースと著作権切れと著作権切れデータベース・クラシック音楽, 情報処理学会論文誌, Vol2001, No.103, pp. 35-42 (2001).

[2] K.Ishii, K.Hoashi, K.Matsumoto, J.Katto (2006). ユーザ嗜好に基づく音楽情報検索システムのための学習データ抽出方法. 情報処理学会第 67 回全国大会.

[3] S.Dixon, G.Widmer(2005). MATCH: A Music Alignment Tool Chest. 6th International Conference on Music Information Retrieval (ISMIR 2005), London, England, September 2005, pp 492-497. (Best Poster Award)