

1Z-04 動画像と環境音による音響生成を用いたインタラクティブシステム

高橋勝吾 鈴木健嗣 橋本周司
早稲田大学理工学部 応用物理学科

1. はじめに

動画像から背景音楽を自動生成するシステムがいくつか提案されている^{[1],[2]}。本稿では、人間のパフォーマンスシーンを観察することにより音響を生成する、インタラクティブなパフォーマンスシステムを提案する。本システムでは、動画像情報だけでなく、環境音の物理的特徴も抽出し、リアルタイムで音楽的特徴と結び付けることで周辺環境を反映するような音響の自動生成が可能である。

一方、一般のユーザに加え、音楽家などのユーザが容易にシステムをデザインできる環境を提供することも重要である。本システムは、ソフトウェアとしてMaxのみを利用することにより、ユーザである音楽家や演出家が、Macintosh上で容易に音楽環境をデザインできるシステムでもある。これより、ユーザのより豊かな創作活動を支援する環境を構築することができる。

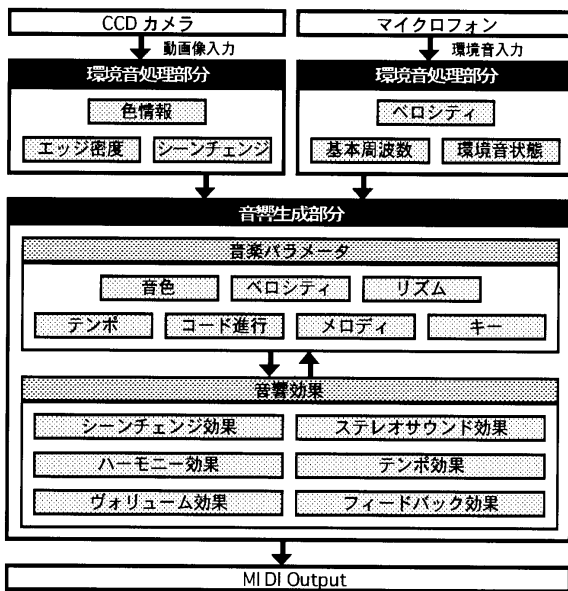


図1 システム概要

Interactive System to Generate Music from Moving Image and Environmental Sound
Shogo Takahashi, Kenji Suzuki and Shuji Hashimoto
Dept. of Applied Physics, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

2. システム構成

2.1. システム概要

システムはMacintoshのMAX上で実装されており、図1のように動画像処理、環境音処理、音響生成処理の3つの部分により構成されている。システムの構成を図2に示す。

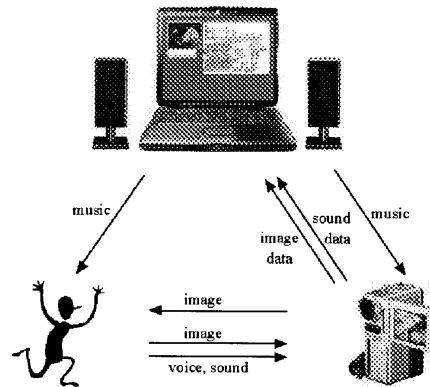


図2 システム構成

2.2. 動画像処理部分

動画像情報は、30フレーム/秒で取得した各フレーム(320 x 249 pixel)を $M \times N$ のエリアに分割し、各エリア毎に特徴量を抽出する。

1) 色情報取得

時刻 t における R, G, B それぞれの輝度総和を取得する。これより HSL (色相, 彩度, 明度) 情報を計算によって求める。

2) エッジ密度取得

時刻 t における R, G, B 毎の輝度勾配をエッジ情報と考え、その総和を求める。

3) シーンチェンジ

上記の2つのフレーム内の特徴量のフレーム間差分値にしきい値を設けてシーンチェンジを検出する。

2.3. 環境音処理部分

マイクより入力された環境音から、以下に示す特徴量を抽出する。

1) ペロシティ

11kHzでサンプリングした16bitの音響データ256個を1フレームとして、このフレーム内での平均振幅を求める。

2) 基本周波数

各フレーム毎にケプストラム法によりピッチを検出する。

2.4. 音響生成部分

ここでは上記の2つの処理部分によって抽出された動画像と環境音の特徴量を音楽的な特徴量と結び付け、音響を生成する。これらの関係はシステムが持つものと、ユーザとの相互作用によって決定されるものがある。

1) 音響生成

対応関係は任意に設定できるがここでは、メロディを色情報、ペロシティをエッジ密度に対応させている。曲のキーは抽出された基本周波数に基づいて決定され、コード進行は準備された5つのパターンから選択される。またドラムパターンには、6つの異なるテンポを持つ72種類のMIDIファイルを利用している。

作曲の基本的ルールは音楽理論に基づいているがユーザが新しいルールを付け加えることも可能である。決定された音楽パラメータは、音色、テンポ、リズム、コード進行などは8小節毎に、メロディとペロシティはリアルタイムでMIDI音源に出力される。

2) 音響効果

システムはカメラの前に物体が現れる、またはマイクに何か入力があると、音響の生成を始める。また、以下に示すような効果を用意した。

例えばシーンチェンジ効果は、シーンチェンジが検出されると全ての音楽パラメータが一新される。また、入力画像中の物体位置により、出力音響にステレオ定位効果が現れる。さらに歌声を検出するとその音声にハーモニーを加える。本システムはユーザの意志だけでなく、システム自らが出力した音響によっても強く影響され、ユーザとシステムの相互作用により主導権をやりとりしているような対話型のパフォーマンスを行なうことが可能である。

3. 実験と考察

実験では、ダンサーがシステムの前で踊ったり、音を立てたり、声を出したりすることで、パフォーマンスを行った。画像は1 x 3のエリアに分割した。図3にその様子を示す。

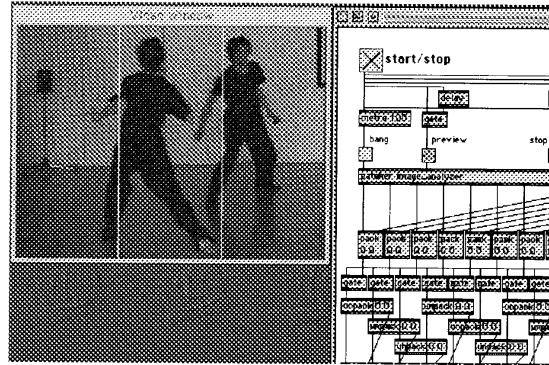


図3 MAXアプリケーションを用いた実験の様子

今回の実験では、動作と音響出力の対応の再現性に若干の問題があったが、ダンサーとのパフォーマンスにおいてシステムが十分な即応性があることが確認でき、新しい音楽環境を生成することが可能であることが判った。

4. あとがき

このシステムの応用例として、ロボットインタフェースに用いたものも報告済みである^[3]。

このようなインタラクティブなパフォーマンスシステムでは、従来の工学的側面についての評価に加え、ユーザの満足度や有効性などの評価もまた重要であり、今後の重要な課題である。

参考文献

- [1] J. Nakamura, T. Kaku, K. Hyun, T. Noma, and S. Yoshida: "Automatic Background Music Generation based on Actors' Mood and Motions", The Journal of Visualization and Computer Animation, Vol.5, pp.247-264, (1994)
- [2] Naoyuki Onoe, Dingding Chang, and Shuji Hashimoto: "Background Music Generation Based on Scene Analysis", in Proc. of International Computer Music Conference '96 pp.361-362, (1996)
- [3] Kenji Suzuki, Takeshi Ohashi, and Shuji Hashimoto: "Interactive Multimodal Mobile Robot for Musical Performance", in Proc. of International Computer Music Conference '99, (1999)