

研究報告

環境音の 3D 空間上への表示方法に関する研究
 STUDY ON ENVIRONMENTAL SOUNDS DISPLAY
 METHOD IN 3D SPACE

津田 浩利
 Hiromichi Tsuda
 東京電機大学
 Tokyo Denki University

小坂 直敏
 Naotoshi Osaka
 東京電機大学
 Tokyo Denki University

概要

当研究室では、マルチメディアコンテンツ制作者への支援を目的として、環境音を対象とした音の検索および合成システム「電子音色辞書」を構築している。この GUI (Graphical User Interface) として、一つの音を一つのオブジェクトとし、これを 3D 上の球として視覚表現してきた。これらの 3D 上の配置と聴感上の印象とを対応させるため、どのような音響特徴量を用いどのように表現したらよいかを考察した。

1. はじめに

音素材の使用は、音楽や音コンテンツ制作において重要な位置を占める。我々はマルチメディアコンテンツ制作者の支援システムとして、環境音を対象とした音の検索および合成システム電子音色辞書 [1] を構築している。同システムでは GUI として 1 つの音 (音脈) をデータベース上に登録し、これを一つの 3D 空間上の球として表示している。また同表示では、音の知覚と対応のよい 3 次元空間上への音配置を目的としている [2]。

現在の 3D 表示は、12 次元 MFCC [3] を音響特徴量とし、それを主成分分析によって 3 次元に圧縮し、それを 3 次元空間上に表示している。MFCC は音声認識や楽音同定などに用いられている特徴量で、この方法で音色の知覚的な距離に対応した表示を期待したが、聴感上明らかに違うものが重なって表示されることもあり、まだ性能は不十分である。これは、楽音、音声という区分よりさらに広い音を扱っていること、MFCC という特徴量に限定していること、および、次元圧縮による情報の欠落などが原因と考えられる。これらの複合した問題を解決する第一歩として、任意の環境音が楽音、音声以上に広範であることを鑑み、パワーによる大分類を考え、その分類毎に表示する方法について検討する。図 1

は現在のシステムでの 3D 表示であり、分類ごとに色分けで表示している。

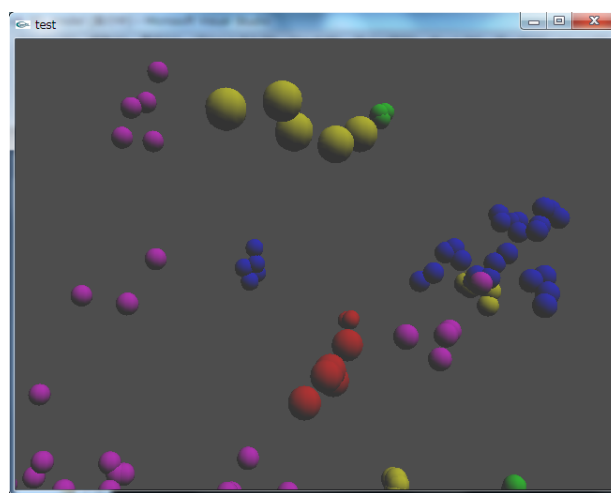


図 1. 現在のシステムでの 3D 表示

2. パワーの分類による提案手法

本稿では、環境音として楽音は含むが、複数声部の音楽音と電子音は除き、継続時間が数秒以内の音を対象とした。すなわち、音脈として物理的に一つの発音体を音源とする音波形全体を対象とする。

今回は、12 次元 MFCC を主成分分析で 3 次元に圧縮した特徴量を基本的統計量として、文献 [4] のように分類したときの各分類の統計量について調べ、今後の分類法や特徴量の選択を考える際の基礎資料とする。文献 [4] では、音を単発音と複数音に大分類し、さらに表 1 に示すようにそれぞれ 3 グループ、2 グループに分類している。環境音の構成について、音源は RWCP 実環境音声・音響データベース [5] よりサンプリング周波数 48kHz、16bit 量子化のものを使用した。表 1 に実験に用いた環境音の内訳を示す。5 分類にそれぞれ 6 種類の音

を設定し、1 種類に 5 サンプル揃え、全体で 150(6 × 5 × 5) サンプルのデータとした。

表 1. 実験に用いた環境音の内容

	分類	音の種類 (各 6 種)
(A) 単発	(1) 持続音	ガスピレーの噴射音, ドライヤーの音, 他 4 種
	(2) 減衰音	自転車ベル, ガラスコップを叩く, 他 4 種
	(3) インパルス性雑音	ホッチキスで紙を綴じる, カスタネット, 他 4 種
(B) 複数	(4) 繰り返し	携帯電話の着信音, 目覚まし時計, 他 4 種
	(5) 不規則	紙を握り潰す, コインを落とす, 他 4 種

3. 実験結果

3.1. 異なる音間の距離と同一音の分散との比較

まず、各分類の中の 6 種類それぞれについて、同一種の音で 5 サンプルに対する分散を求める。これは再現性について調べていることになる。表 2 は持続音の例である。これらの平均を σ_1 とする。さらに同一分類で各種間の分散を求める。これらの平均を σ_2 とする。最後に分類間の分散を求め、これを σ_3 とする。

表 2. 持続音の 3 次元圧縮 MFCC に対する分散

音の種類	1	2	3	4	5	6
第 1 主成分	0.09	0.01	0.05	0.11	0.06	0.09
第 2 主成分	0.16	0.03	0.05	0.13	0.08	0.06
第 3 主成分	0.03	0.01	0.06	0.11	0.18	0.07

1 : ドライヤー音 2 : ブザー音 3 : スプレー音
4 : シェーバー音 5 : ゼンマイ音 6 : ホイッスル

表 2 の結果から、各種類の分散は小さいため、同一音の繰り返しは特徴量として利用可能であることがわかる。次に分離度を 2 種類を求める。まず、同一音の繰り返しを一つのクラスとして、各分類内の 6 種のクラスについての分離を考える。すなわち、クラス内分散を σ_1 、クラス間分散を σ_2 としたときの分離度 R1 を求める。次に、上位階層の分類を考える。表 1 の分類をクラスとみて、クラス内分散を σ_2 、クラス間分散を σ_3 としたときの分離度 R2 を求める。それらを表 3 に示す。

3.2. 考察

表 3 より、 σ_1 は小さく同一種で密集しており、また、分離度 R1 が大きく同一分類内でも密集していることがわかる。これは、同一音の再現性があることを示してい

表 3. 分離度 R1 と R2

	σ_1	σ_2	σ_3	R1	R2
第 1 主成分	0.12	0.83	0.70	6.87	0.84
第 2 主成分	0.11	0.63	0.50	5.75	0.79
第 3 主成分	0.17	0.87	0.44	5.22	0.50

る。しかし、分離度 R2 は小さいことから各分類同士があまり分離していないことがわかる。つまり 3D 表示した場合、全分類が同じ個所に密集していることがわかる。これらより 3D 上に表示した場合、5 つに大分類されたものが MFCC では明確に分類できていないといえる。これは図 1 において同一色別に明確にグルーピングされていないことに対応する。このためすべての環境音を 3 次元上に表示することは限界であることを確認できた。一方、この大分類は聴感上明らかに異なり、また、簡単にパワーで分類できる。そこで今後はパワーによる大分類を行った上で 3D の表示を検討していく。

4. まとめ

環境音について、12 次元 MFCC を主成分分析で 3 次元に圧縮した特徴量を基本的統計量とし、3D 表示を行った際の分離度を調査した。結果より知覚的な 3 次元表示をすべての環境音に対して行うことに限界があり、そのためパワー分類ごとに分けて表示を行う必要があることがわかった。今後の展開は、大分類のさらなる細分化と、その他の音響特徴量の利用を検討する。

5. 参考文献

- [1] 小林, 小坂, 春季音学講論 pp. 449-452, 2008.3
- [2] 吉岡, 他, 情処学会音情研報告 pp. 53-58, 2009.02.
- [3] 安藤, 電子情報通信学会, pp. 28-30, 2003.
- [4] 根本, 小坂, 音学講論 1-19-16, 2011.9.
- [5] RWCP 実環境音声・音響データベース, Real World Computing Partnership, 1998-2001, <http://tosa.mri.co.jp/sounddb/>

6. 著者プロフィール

津田浩利 (Hiromichi TSUDA)

2013 年東京電機大学大学院在籍。電子音色辞書を中心とした音響信号処理応用の研究を行っている。

小坂直敏 (Naotoshi OSAKA)

本会報 早坂将昭 「遠隔動作を光と音として表現するサウンドインスタレーションの制作」参照