

## ユーザの嗜好に適応させた選曲システムの提案

彦坂健太郎<sup>†</sup> 谷口徹<sup>††</sup>  
蒼田雅彰<sup>††</sup> 白井克彦<sup>†††</sup>

本研究では、ユーザに煩わしい操作を要求することなく、ユーザの好みに合わせた自動的に選曲を行うシステムの開発を目指し、主観的な好みと楽音の音響的特徴との関連性を基にユーザの気分に対応する楽曲を楽曲再生時のオンライン学習により適応的に選択する手法を提案する。具体的にはユーザがある楽曲を「聴きたくない」としたときに特徴量空間においてそれらの楽曲からの距離が遠い楽曲を選曲するアルゴリズムを適用する。今回実験によって従来のランダム再生と比較しこの手法が有効であることが確認された。

### A method of automatic music selection adapted for user's preference

KENTARO HIKOSAKA,<sup>†</sup> TORU TANIGUCHI,<sup>††</sup> MASAOKI HONDA<sup>††</sup>  
and KATSUHIKO SHIRAI<sup>†††</sup>

In this research, we are aiming for development of a music selection system adapted for user's feeling without disturbing operation. We propose an adaptive automatic music selection method based on the relationship between the acoustic features of music and the subjective user feeling. Concretely, when a user makes a decision that they don't want to listen some music, we apply an algorithm that choose a music which is furthest from the music. In this paper, we could confirm efficacy of this means compare with random selection.

#### 1. はじめに

近年、ハードディスクの大容量化やネットワークの高速度により我々はPCをオーディオプレイヤーとして使用し、大量の音楽を入手、蓄積し、再生することが可能となった。また、携帯型ハードディスクプレイヤーの普及により、手軽にどこへでも大量の音楽を持ち出すことが可能になった。今後このような音楽再生の様式の変化はますます進むと考えられる。

このように大量の音楽を扱いながら再生することが可能になると、「聴く曲を自ら選曲したくないが、何か適当な音楽を流したい」、「ランダムに再生して意外性のある選曲で楽しみたいが、今の気分には合わない曲は省きたい」といった欲求がでてくると考えられる。しかしながら、従来のものと比較して現在の音楽再生のインターフェイスはほとんど変わっておらず、これらの要求に十分応えているとは言いがたい。現在の音楽再生インターフェイスを見てみると、このような要求に応えるための機能として、ユーザが聴きたい曲の一群をあらかじめ指定し、保有しておく「プレイリスト機能」や、楽曲を完全にランダ

ムな順序で再生する「ランダム再生機能」などが用意されている。しかし、「プレイリスト機能」では最初にユーザが手動でプレイリストに楽曲を登録しなければならず、そのプレイリストは変わることはないため意外性がなく、そのプレイリストに飽きてしまうとまた別のプレイリストを手動にて作成していかなければいけないなど手間のかかるものである。一方「ランダム再生機能」ではユーザが楽曲を指定する手間はかからないものの、ハードディスク内の楽曲が、そのジャンルやアーティストの曲調に関わらずランダムに選曲されるためハードロックの後にクラシック、といったような流れの悪い選曲になりがちであるし、ユーザが今聴きたい曲が再生されるとは限らない。

#### 2. 関連研究と本研究の目的

自動選曲についての関連研究として、ユーザの複数のプレイリストを集合として扱い、プレイリストの評価重み付けを利用者の嗜好に合わせて動的に変更し、プレイリストの重みから各楽曲の優先度を計算するアルゴリズムの研究<sup>1)</sup> などがある。しかしメタデータなどによる選曲では、年代、ジャンル、アーティスト、経験など細かい設定を行うことができる反面、ユーザが多くのこと考えたり、データの付加を行うといった複雑な操作をすることを要求されてしまう。また、十分なメタデータが蓄えられ有効に活用できるようになるまでの時間が長く、気

<sup>†</sup> 早稲田大学大学院 理工学研究科

Waseda university graduate school faculty of science and engineering

<sup>††</sup> 早稲田大学 スポーツ科学部

Waseda university faculty of sport sciences

<sup>†††</sup> 早稲田大学 理工学術院

Waseda university faculty of science and engineering

分の変化に応じた選曲や意外性を与える選曲ができないなどといった問題がある。

そこで我々は音響特徴量を用い、ユーザに複雑な操作を要求することなく且つユーザの気分に応じ適応的にユーザに合う楽曲を選曲していくシステムを提案する。この選曲問題は音響特徴による一種の楽曲分類・識別問題と捉えることができる。このような問題を扱った研究としてパワースペクトルから抽出される情報により音楽のジャンル分類を行う研究<sup>2)</sup>、パワースペクトルから抽出される音響の情報による類似度と主観的な類似度との評価を行う研究<sup>3)</sup>などが行われている。しかし、未だ音響特徴量を有効に用いて、変化するユーザの気分に合わせて選曲を行うシステムや再生インターフェイスは十分に発展しているとは言えない。

本研究では、対象としてユーザの音楽プレイヤーやPCのHDDに入っている数千曲程度の楽曲ファイルを扱うことを想定する。HDDに入る楽曲はユーザが聞きたいと思った楽曲であるという前提の下、その中からユーザの再生時の気分合う楽曲を選曲することのできるシステムを提案する。

ユーザの状況として一般的な音楽プレイヤーで、「再生」「次の曲」などの標準的な操作のみしか行うことができないことを想定する。ユーザが一連の音楽を再生する中で、「この曲は今聴きたくないのでスキップする」という通常通りの簡単な判断をすることにより、その次からの選曲に対してその判断を反映させ、選ばれた「聞きたくない曲」と類似度の高い曲を選曲しないことを目指す。逐次的にユーザの選択を選曲に反映させるためユーザの聞きたくない曲が気分や状況により変わった時も、そのことを選曲に反映させることができる。また本システムは楽曲の音響的特徴に基づいて識別を行うため、アーティストやジャンルなどといった楽曲に付けられているメタデータに依らない選曲を行うことができる。

本手法は、例えばポータブルプレイヤーなどの限定的なユーザインターフェイスにおいても、従来のインターフェイスを変更することなく実現することが可能であることは実用上の利点である。また、本研究の応用例として、音響信号の分析による楽曲のリコメンドシステムの実現や楽曲検索に適用されることが期待できる。

図1にシステムのイメージを示す。最初に楽曲から複数の特徴量を抽出する。さらにユーザが聞きたくないと選んだ楽曲の特徴量と、残りの楽曲コンテンツ群の特徴量に選曲アルゴリズムを適用することによりユーザの趣向に合わせた選曲を実現する。以降では3章において各特徴量抽出について、4章において選曲アルゴリズムについて、5章で実験について、6章においてまとめと今後の課題について述べる。

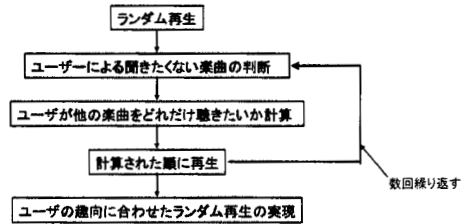


図1 システムイメージ

### 3. 特徴量抽出

まず最初に識別アルゴリズムに必要な音響信号の特徴量を抽出することが必要となる。音響分析に用いる特徴量として多くの候補が挙げられるが、選曲アルゴリズムでは特徴量空間における分布を用いて計算を行うため、特徴量を適切に選択することは非常に重要な作業である。その性能向上のため有効な特徴量とは楽曲の主観的・客観的特徴をよく捉え且つ、特徴量空間上でユーザの気分によるまとまりに分離され分布するような特徴量であると考えられる。

楽曲の特徴としてはリズム、テンポ、コード進行、ピッチ情報、曲構造情報などが考えられるがそのような特徴は複雑なものであり、計算により導きだすことは非常に困難である。しかし、それらの特徴に関連した特徴量を用いることにより、楽曲の性質が適切に表せると考えている。今回は調査した複数の特徴量の中から結果が良好であったものを用いた。特徴量は次の3種類である。

- (1) リズム、テンポなどのビート情報 (ビートスペクトル)
- (2) 音色情報 (平均 MFCC)
- (3) パワー情報 (パワーヒストグラム)

以下に各特徴量の詳細を示す。

#### 3.1 ビートスペクトル

まず初めに楽曲を特徴づける大きな要素であると考えられるビート情報の抽出について検討することにした。リズム情報を抽出する有効な手法の一つとして音響信号の短時間特徴量の自己相関を用いて、リズム情報であるビートスペクトルという特徴量を算出する手法<sup>4)</sup>がある。

この手法は楽曲内の類似度を用いてビートを算出する手法であり、帯域制限などを行わないため、ドラムやベースなどリズム楽器を含んでいない楽曲や音量の小さい部分を多く含む楽曲などにも広く適応することができる。そのため楽曲間のリズムの類似度を明らかにする研究<sup>5)</sup>や、リズムの類似度を用いて楽音の検索を行う研究<sup>6)</sup>や、楽音の分布を音響的特徴に基づいて可視化する研究<sup>7)</sup>などに用いられている。

##### 3.1.1 ビートスペクトルの計算手法

ビートスペクトルの計算手順は以下の通りである。

- (1) 音響信号のパラメタライズ

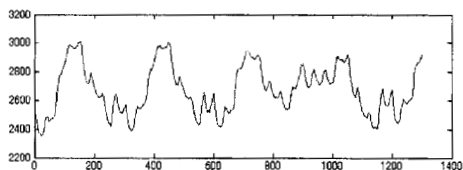


図2 rockのビートスペクトル

- (2) フレーム類似度の計算
- (3) 距離マトリクスの作成
- (4) ビートスペクトラムの導出

音響信号をパラメタライズする手法としては様々な方法が挙げられるが、ここでは対数パワースペクトルを用いる。窓長は256点、シフトサイズは128点とした。ここで用いる音響信号は16bitで量子化、22kHzでサンプリングされた信号である。よってフレーム長はおよそ11msとなる。今回ビートスペクトルは10秒毎に時間窓10秒で抽出を行った。例としてある時間のrock楽曲のビートスペクトルを図2に示す。図をみるとビートスペクトルの形状が繰り返し構造になっており、rockの比較的ビートの繰り返しが強いという特徴捉えているものと考えられる。

### 3.2 長時間MFCC

楽曲の音色を表す特徴量として、スペクトル包絡を表わす特徴量であるMFCC(mel-frequency cepstral coefficients)<sup>8)</sup>を一楽曲に渡って平均した長時間MFCCを用いた。

#### 3.2.1 長時間MFCCの算出方法

まず、一楽曲全体にわたって、窓長4096点、シフト幅1280点で全てのフレームのMFCCを算出する。MFCCの計算時のフィルタバンク数は40であり、求められる次元数は13次元である。算出された全てのMFCCの平均を求める。これをその楽曲の長時間MFCCとする。

### 3.3 パワーヒストグラム

楽曲のパワーの構成を表す特徴量として短時間パワーのヒストグラムを用いた。これは時間毎のパワーをヒストグラムにした特徴量で、これにより楽曲の盛り上がりの激しさなどを表すことができ、楽曲の構造的な特徴をとらえることができるのではないかと考えた。

#### 3.3.1 パワーヒストグラムの算出方法

まず、一楽曲全体にわたって一定の時間間隔ごとに短時間パワーを算出する。次に、それらのパワー値によって10分割のヒストグラムを作成する。これを10次元の特徴量として用いた。

この特徴量においてはパワーをとる窓長により、パワー値が変化するため特徴量も変化すると考えられる。よってこの時間の設定が適切かどうかという問題があるが、今回は複数調べた時間間隔の中で一番良好な結果が得られた0.8秒に設定した。

## 4. 選曲アルゴリズム

この章では以下に選曲アルゴリズムの手順について述べる。

ここで $\mathbf{x}(i)$ は楽曲*i*の特徴量ベクトル、 $s(i, j)$ は楽曲*i*の*j*回目の選曲におけるスコアである。

- (1) 特徴量の抽出・保持  
全ての楽曲について前述の3つの特徴量を抽出し、正規化した上で各楽曲について保持する。
- (2) 楽曲スコアの初期値設定  
 $n$ 個の全楽曲に対して同じ初期値 $s(i, 1)(i = 1, 2, \dots, n)$ を与える。
- (3) ランダム再生  
 $n$ 曲の楽曲のうち、ランダムに1曲を選択し、再生する。これを(4)の指定があるまで繰り返す。
- (4) 聞きたくない楽曲指定  
ユーザが聞きたくないと思った楽曲が*j*回目の選曲において出てきた時点で、その楽曲が嫌いと指定される。その時の楽曲インデックスを“hate”とし、(5)に進む。
- (5) 楽曲スコアの再計算  
全ての楽曲について、(5)において指定された楽曲の特徴量空間上の位置ベクトル $\mathbf{x}(\text{hate})$ と他の各楽曲の位置ベクトル $\mathbf{x}(i)$ との距離 $d(i)$ を計算し、その逆数を楽曲スコアからそれぞれ減算する。  

$$d(i) = |\mathbf{x}(\text{hate}) - \mathbf{x}(i)| \quad (i = 1, 2, \dots, n) \quad (1)$$

$$s(i, j) = s(i, j-1) - \frac{1}{d(i, j)} \quad (i = 1, 2, \dots, n) \quad (2)$$
- (6) 選曲  
選曲されていない楽曲の中で計算されたスコアが最も高い $\max s(i, j)$ である楽曲が選曲される。もしそのような楽曲が複数あるような場合はランダムにいずれかを選択する。
- (7) 以降繰り返し  
(4)から(6)までを繰り返す。

この手法では、(1)、(2)が選曲アルゴリズムの前処理、(3)が完全なランダム再生、(4)~(7)が選曲アルゴリズムの適用部となる。次章からこのアルゴリズムを基に行った実験について示す。

## 5. 実験

本手法の有効性を確かめるため実験を行った。

### 5.1 分析用楽曲データ

各被験者に、被験者自身のPCや携帯音楽プレイヤー中に保持している楽曲データの中から落ち着きたい時にききたい曲(class1)、通常の気分の時に聞きたい曲(class2)、気分を高揚させたい時に聞きたい曲(class3)の3classで楽曲を各30曲ずつ挙げてもらった。

被験者A、Bにはジャンルやアーティストの指定は行

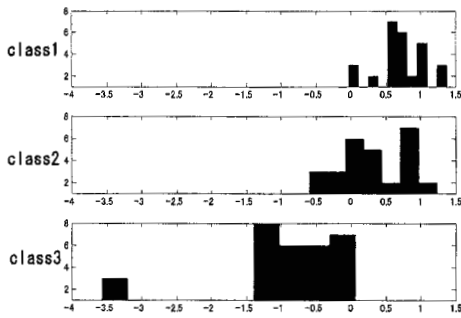


図3 ビートスペクトルの平均値のヒストグラム

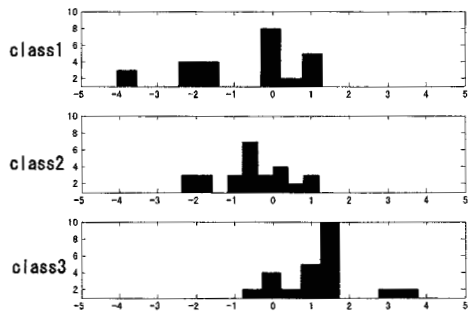


図4 MFCC12次元の第一主成分の値のヒストグラム

わなかった。楽曲のジャンルは特に指定せず自由に選んでもらったが、結果として被験者 A は洋楽の rock, pops, テクノなど、またはインストメンタル、被験者 B では洋楽、邦楽の rock または pops, jazz, classic を含むインストメンタルの楽曲を選択した。被験者 C はジャンルがロックである同じアーティストの楽曲 90 曲である。

なおここで用いられた楽曲は全て量子化数 16bit, サンプリング周波数 22kHz, モノラルである。

## 5.2 特徴量分析

抽出された各特長量がどのような性質を示すか、ということ調べるため分析を行った。分析に用いたデータは被験者 B 一人分のデータである。

### 5.2.1 ビートスペクトルの分析

3.1.1 において得られたビートスペクトルは高次元な特徴量なので、どのような形に縮約するかという問題がある。今回は複数の縮約方法を試した。その中でビートスペクトルが同じ楽曲内での類似度を用いる手法であることから、ビートスペクトルの値そのものがビート構造の特徴を示すという考えの下、ビートスペクトル全てについての平均値を特徴量とする方法が最も良い結果が得られたため、そちらを採用した。図 3 に各 class ごとに特徴量の値をヒストグラムで表したものを示す。なお特徴量は平均が 0, 分散が 1 となるように正規化されている。図 3 を見ると class 間で重なっている部分はあるものの、おおむね class により分散していることがわかる。

### 5.2.2 長時間 MFCC の分析

今回得られた長時間 MFCC において、直流成分である 1 次元目を除き、2 次元目から 13 次元目までの合計 12 次元を特徴量として用いた。図 4 に各 class ごとに、12 次元の特徴量に対して主成分分析を行い、第一主成分の値をヒストグラムで表したものを示す。なお特徴量は各次元ごとに平均が 0, 分散が 1 となるように正規化されている。表を見ると class1 と class2 は重なっている部分があるが大きい、class3 はやはり他の class より、値が違うことがわかる。

### 5.2.3 パワーヒストグラムの分析

今回得られた合計 10 次元のパワーヒストグラムに対し

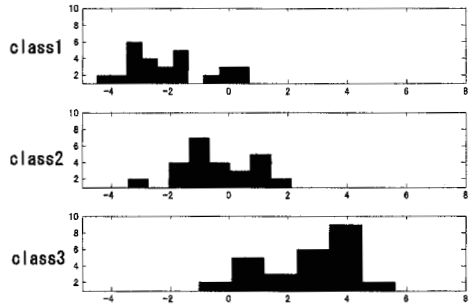


図5 パワーヒストグラムの第一主成分のヒストグラム

主成分分析を行い、class ごとに第一主成分の分布をヒストグラムとして示した (図 5)。なお特徴量は各次元について平均 0, 分散 1 となるように正規化されている。表を見ると完全に分かれるまではいかないが、class ごとのヒストグラムの重なりが小さいことがわかる。

### 5.2.4 全特徴量の統合

システムに特徴量を適用するにあたって、ビートスペクトルとパワーヒストグラムについては平均 0, 分散が 1 となるように各次元ごとに正規化を行った。しかし、長時間 MFCC については各次元で表す特徴の重要度が異なるためこのような正規化を行うことは適切ではないと考えた。そこで上記の実験手順において実験的に重み  $w$  を定めた。被験者 A, B, C について  $w$  を変えながら各被験者の結果において基準線と総合評価の曲線との間の面積を計算しそれらの合計が最大となった  $w=14$  を設定した。

被験者 B についての特徴量セットについて主成分分析を行い、横軸に第一主成分、縦軸に第二主成分をとりそれらの分布を調べた。結果を図 6 に示す。この分布を見ると各 class 重なっている部分はあるものの大まかには class ごとの分布が形成されていることがわかる。よってこのシステムで用いられる特徴量は正規化したビートスペクトル 1 次元、重みを付けた長時間 MFCC12 次元、正規化したパワーヒストグラム 10 次元の合計 23 次元である。

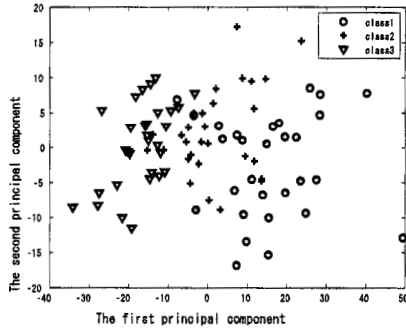


図 6 特徴量セットの分布

### 5.3 アルゴリズムの適用実験

統合された特徴量のセットをアルゴリズムに適用した時の性能を調べるため、実験を行った。

#### 5.3.1 実験手順

まず被験者 1 人分のデータを準備する。4 章で述べた選曲アルゴリズムにおいては (1) でランダム再生を行い、ユーザが気分によって「聴きたくない」と選んだ楽曲により逐次的に選曲を適応させていく。しかし今回の実験では、楽曲に付加された class により性能を評価する。

現実での使用状況を再現するため、ここでは先に述べた 3class のうち 1 つを「今聴きたくない」曲であると仮定して 4 章のアルゴリズムを適用した。

各被験者、各 class について 30 セットの試行を行う。各試行は「聴きたくない」class のうちの 1 曲から開始され、その被験者が選択した全 class 計 90 曲を再生し終わるまで続ける。この 30 回×3 セットの選曲において、選曲位置毎の聴きたくない楽曲の累積出現数の平均をとり、それをアルゴリズムの性能の 3 つの class についての総合評価とし、嫌いな楽曲が一樣に等出現率で 30 曲出現した場合の直線を基準線、つまり完全にランダムな選曲の場合の直線として比較する。また class1, 2, 3 それぞれの class のみについても調査した。

#### 5.3.2 実験結果

各被験者の実験結果を図 7, 8, 9 に示す。横軸が選曲を行った回数、縦軸がその中で聴きたくない指定した class に属する楽曲が選曲された累積数である。各図において総合評価 (all)、class1 のみを平均した評価 (class1)、class2 のみを平均した評価 (class2)、class3 のみを平均した評価 (class3) を示した。参考として嫌いな楽曲が一樣に等出現率で 30 曲出現した場合である基準線 (base)、最初一曲の判断のみで聴きたくない class の楽曲が最後まででてこなくなった場合、つまり最良の結果 (best)、最初に聴きたくない楽曲が全て出現しその後他の楽曲が出現する最悪の結果を想定した場合も示した (worst)。評価としては完全にランダムであると仮定した場合である基準線より評価線が下まれば下まわるほど評価が高い

ということになる。結果を見ると被験者 A, B, C 全てほぼ全域で基準線を下回っていることがわかる。最も評価が高いのは被験者 B であり、被験者 A, C については同程度であるということがいえる。被験者 B においてはかなり高い性能が得られた。さらに特に被験者 C において、ほとんど似た楽曲である同じアーティストの楽曲についても基準線を下回ったことは、本手法の有効性を示すものであるといえる。class ごとの評価を見ると class3 についての性能が若干良好であることが見て取れるが明確な差はなかった。被験者 A については音楽の種類が多岐に渡り、ある種類の音楽の中での聞きたい気分と他の種類の音楽の中での聞きたい気分が同じ言葉で表現されていても、必ずしも一致せず特徴量空間の形成が困難であった可能性がある。

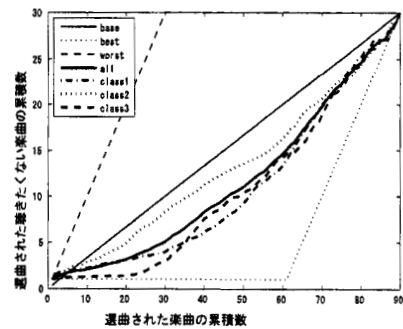


図 7 被験者 A の結果

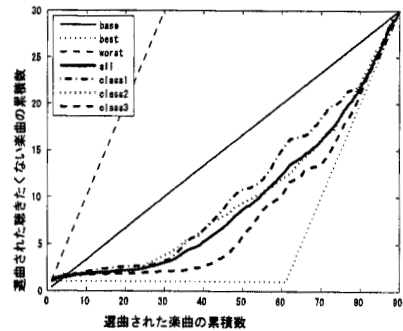


図 8 被験者 B の結果

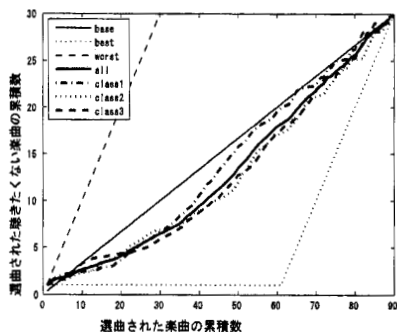


図9 被験者Cの結果

## 6. まとめと今後の課題

今回新しい楽曲選曲システムを提案し、楽音からの特徴量抽出、アルゴリズムの適用を行い、本手法の有効性を確認した。今後さらにその他の有効な特徴量を探すと共に、アルゴリズムの改善を行い、少数サンプルから頑健な学習を行うことができるシステムを目指していく。

## 謝 辞

本研究の一部は、早稲田大学理工学総合研究センターの研究課題「自発的コミュニケーション機構を有するマルチモーダルヒューマンインタフェースの研究」、平成17年度科学研究費基盤研究(B)課題番号17300066「対話状況に応じた自発的コミュニケーション機構の研究」によるものである。ここに記して謝意を表する。

## 参 考 文 献

- 1) 暦本純一, “UniversalPlaylist: 利用者の嗜好に動的に適合するメディア再生機構”, インタラクシオン2005, 2005.
- 2) G. Tzanetakis, P. Cook, “Musical Genre Classification of Audio Signals”, IEEE Transactions On Speech And Audio Processing, VOL. 10, NO. 5, July2002
- 3) A. Berenzweig, B. Logan, D. Ellis, B. Whitman, “A Large-Scale Evaluation of Acoustic and Subjective Music Similarity Measures” Fourth International Conference on Music Information Retrieval, 26-30 October 2003
- 4) J. Foot, S. Uchihashi. “The Beat Spectrum: A New Approach to Rhythm Analysis,” in Proc. International Conference on Multimedia and Expo 2001.
- 5) J. Paulus and A. Klapuri, “Measuring the similarity of rhythmic patterns” Proc. ISMIR 2002, pp. 150-156 (2002).
- 6) J. Foote, M. Cooper and U. Nam, “Audio retrieval by rhythmic similarity,” Proc. ISMIR 2002, pp. 265-266 (2002).
- 7) E. Pampalk, A. Rauber, and D. Merkl.

“Content-based Organization and Visualization of Music Archives,” Proceedings of the 10th ACM International Conference on Multimedia (MM’02), pp 570-579, Juan-les-Pins, France, December 1-6, ACM.

- 8) Davis, S. and P. Mermelstein, “Comparison of Parametric Representations for Monosyllable Word Recognition in Continuously Spoken Sentences,” IEEE Trans. on Acoustics, Speech and Signal Processing, 1980, 28(4), pp. 357=366.