クラスタリングとヒストグラム特徴を用いた類似曲検索システム
Music Search System By Clustering And Histogram Characteristic

坂場 悠平
Yuhei Sakaba
法政大学情報科学部ディジタルメディア学科
E-mail: E-mail: 07k1024@stu.hosei.ac.jp

Abstract

Since the memory capacity of mobile device is expanding, it is not rare that individual users to hold thousands or tens of thousands of music as a part of personal data. I propose the speedy method that is able to search similar music without information of name of songs or artists' characteristics from all the personal music. Former method that uses comparison of spectrum or sound wave is precise enough to search similar music, but there would be enormous calculation needs to done. From former research of chronological order active search method with using histogram characteristics and setting appropriate parameter, I will build clustering. As a result, the system was able to make similarities from all the music of ten thousand pieces in 2.7 seconds if characteristic of music was already extracted.

1 まえがき

本論文では，大量の楽曲を所有している場合に検索クエリとなるアーティスト情報や曲名情報が無く，重複して同一楽曲を所有してしまった場合や，同様にアーティスト情報や曲名情報を調べる際の，音響信号同士の類似性判定を高速に探索を行うシステムを提案する．

音楽データベースの内容一致検索を，より高速で正確に行う研究はすでにいくつか行われてきている．そのほとんどは，内容検索を目的として，音響情報のインデクシングや分類を試みたものである．このような研究は，時間領域や周波数領域における様々な特徴量に基づくものと，ワードスポッティングに基づくものに大別される．本論文で扱う探索方法はそれらとは異なり，類似度を求めたい複数の信号が具体的に与えられていることと，類似している信号区間がほぼそのままの形である (スペクトルの変動が小さい) ことを前提とする．音響信号の一致検索としては，相互相関関数のような波形のずらし照合などで解決できるが，長時間の信号や多数の信号を探索対象とする場合に，計算量の問題から長時間を要する問題があり，実用的な処理時間で，もれなく探索することは困難である．

本論では柏野氏らによって提案された時系列アクティブ探索法 [4] におけるスペクトル特徴におけるヒストグラム (以後はヒストグラム特徴) に注目する．このヒストグラム特徴には時系列と周波数帯域における局所的かつ大域的な情報が含まれる．また，類似度の定義にヒストグラムの重なり率を用いることで，簡単な計算によって類似度を算出できる．同一楽曲検索に最適なパラメータでヒストグラム特徴を抽出し，クラスタ構築を行うことで，同一楽曲を高速に探索するシステムを提案する．

2 現在の楽曲検索

既存の音楽配信サービスでの配信対象楽曲数は増加する一方であり，iTunesMusicStore では 100 万曲以上，着うたでは 10 万曲近くの楽曲の購入が可能である．

記憶領域の大容量化によって，個人で数千曲から数万曲を記録可能な携帯型音楽プレイヤーを持ち歩き，大量の楽曲を整理できずに属性情報の無い曲が混在することも少なくない．そこで，大量の音楽データの中からユーザーが必要としている楽曲を効率的に検索するシステムの重要性が高まっている．

音楽再生アプリケーションには，個々の楽曲に付与されているアーティスト名や楽曲タイトルなど，属性情報を利用した検索機能がほぼ例外なく実装されている．このような属性情報の検索機能は，ユーザーが必要とする楽曲を大量の音楽データの中から検索するために必要不可欠な技術であり非常に高速で正確であるが，属性情報の無い未知の楽曲を発見することはできない．

また今後さらに普及が広まるとされる音楽プレーヤー機能付きの携帯電話やスマートフォンにおいては，楽曲のタイトル・アーティスト情報などの入力でも手間がかかってしまう．

さらに著作権者の使用許諾のない楽曲の使用に関しても，聞いている人にとって未知の楽曲である場合，取り締まりは困難である．最近の事例として，2010 年に中華人民共和国で行われた上海万国博覧会のテーマソングとして中国人が作曲したとされる楽曲が，実際には日本人シンガーソングライターの著作物であることが発覚した，これは世界的に非難を浴びるものであり，楽曲を使用するにあたっての事前の確認や，同じ曲があるかどうかの判定を行っていればこのような大きな問題には至らなかったはずである．

2.1 content-based 音楽情報検索

検索システムは必要な情報に関するキーワードを入力とするものが一般的だが，音楽音響情報に対する検索では，検索キーを文字で表現することは難しい場合が多い．属性情報に基づく音楽情報検索の問題を解決するため，楽曲のコンテンツ，すなわち音響的な特徴量などに基づいて検索を行う”content-based 音楽情報検索”に関する研究が近年盛んに行われている．content-based 音楽情報検索技術であれば，従来の属性情報に基づく音楽情報検索と異なり，ユーザーが検索したい楽曲に関する情報を有していない場合でも，楽曲の検索が可能となり，音楽データの利便性が大きく向上する．content-based 音楽情報検索技術は，ユーザーのみならず，音楽コンテンツを提供するプロバイダからも発展が期待されている技術である．

現在進められている content-based 音楽情報検索の研究として厳密型音楽情報検索技術がある．厳密型音楽情報検索技術の目的は，大量の音楽データの中にある特定の楽曲を検索することである．厳密型音楽情報検索の主な実現例として，ユーザが検索したい楽曲の一部を発声することによりシステムに入力する「ハミング検索システム」と，楽曲の音源そのものをクエリとして入力し，入力された楽曲を検索する「音楽認識システム」が挙げられる [1]．このような音響信号に基づく照合をパターン・マッチングと呼ぶ．

パターン・マッチングに用いる音楽音響情報の特徴量や類似度の判定手法は多く提案されているが，探索結果は類似度順にソートされたランキングリストの形で与えられることが多い．つまりユーザーは類似検索技術を利用することで，必ずしも正確な検索キーを入力しなくとも，提示された複数の回答案の中

Supervisor: Prof. Katunobu Itou

から自分が意図していたものを選び出すことができる．

厳密型音楽情報検索技術を用いた代表的な実用例として Gracenote 社による MusicID が挙げられる．Gracenote 社は PC 上での音楽再生アプリケーションのユーザーから提供される情報を元に「CDDB」と呼ばれる世界最大級の音楽データベースを構築しており，ネットワーク環境があれば音楽情報を検索できる．

本研究では，ヒストグラム特徴を特徴量とした楽曲同士の照合と，ヒストグラム特徴の各指定周波数帯域の平均パワーを空間内の座標としたときの，座標間ユークリッド距離によるクラスタ構築による照合対象を限定した，楽曲の類似性判定をヒストグラム重なり率によって求める，同一曲の高速探索システムについて提案する．

3 提案手法

提案手法では，時系列アクティブ探索法 [4] におけるヒストグラム特徴を用いる．このヒストグラム特徴は，局所パターン [2] や音符の出現確率 [3][5] などの特徴量とは異なり，時系列ごとの周波数帯域の分布が含まれる．つまり，ヒストグラム特徴の重ね合わせ方法によって時系列における大域的かつ局所的な類似率複合することが可能となる．

比較対象として相互相関関数を挙げる．相関法を比較対象とした理由は (1) 信号同士の重ね合わせによって類似区間を求める，(2) 局所的なスポッティング検索ではなく，信号幅全てを用いた類似性判定判定を行う，この 2 点が一致検索技術として酷似している為である．相互相関関数を用いた場合，同一データに関しての検索精度は高いが計算量が膨大になる．相互相関関数の類似率判定の精度と，処理時間を比較対象として，ヒストグラム特徴を用いた性能評価の実験を行う．

3.1 ヒストグラム特徴

本研究におけるヒストグラム特徴について説明する．本研究で用いるヒストグラム特徴とは，柏野氏らによって提案された時系列アクティブ探索法に用いられている．図 1 にヒストグラム特徴の抽出手順を示す．ヒストグラム特徴は音楽音響情報の，時系列に沿った指定周波数帯域ごとのパワーの総和から求めた連続したヒストグラムの集合である．

まず音響信号を指定した時系列で分割する．時間幅は秒単位で指定することも出来るが，分割幅を時間ではなく，単に分割数を指定することもできる．前者の場合，例えばヒストグラム特徴の重なり率算出の際の重ね合わせのずらし幅を 20ms と細かく設定することで，より詳細な類似区間を検索可能となる．後者の場合だと詳細な類似区間は検索できないが，音響信号全体における類似率を前者に比べ高速に検索可能だ．

次に分割した波形を 2 次の IIR バンドパスフィルタを用いて指定周波数帯域ごとで出力する．出力波形ごとのパワースペクトルの総和をそれぞれヒストグラムのビンの値とする．図 1 に特徴抽出の一連の流れを示す．

3.2 ヒストグラムを用いた探索方法

図 2 に方法の概要を示す．はじめに全ての楽曲に対して図 1 のような処理を行う．すなわち，まず音響データに対して特徴抽出を行い，その特徴抽出したヒストグラム特徴を元にデータベースを作成する．楽曲一曲あたりのヒストグラム特徴に対して必要なデータベースの容量は 1kB から 20kB ほどである．計算量，探索時間が増すことは実際の利用を想定すると利便性に富んでいるとは言えない．ヒストグラムのビンの数が多いほどその数だけ重なり率を求めるための計算量が増す．

つまり実用的な時間かつもれなく探索できる特徴抽出と照合区間幅のパラメータを何通りもの実験を行い最適に設定する必要がある．

3.3 音響信号の類似率判定

音響信号の類似率判定は，ヒストグラムの照合に基づいて行う．ヒストグラム同士の類似度としては，ヒストグラム重なり率を用いる．時間分割された i 番目のヒストグラムにおけるヒストグラム重なり率 S_i は，次のように定義される．

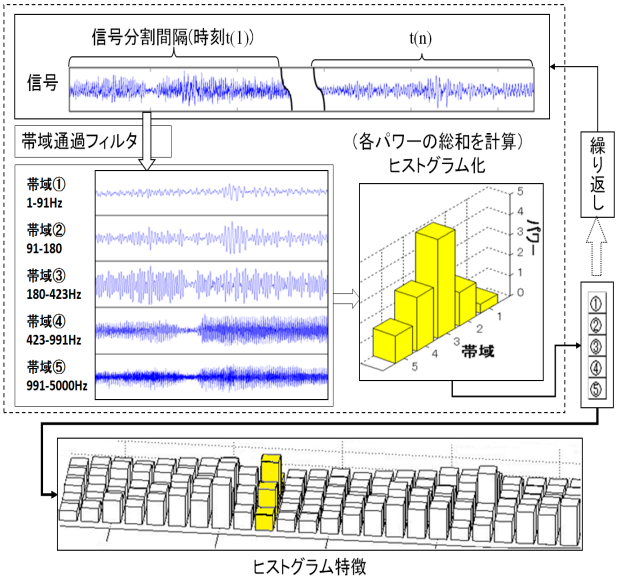

図 1　ヒストグラム特徴の抽出例

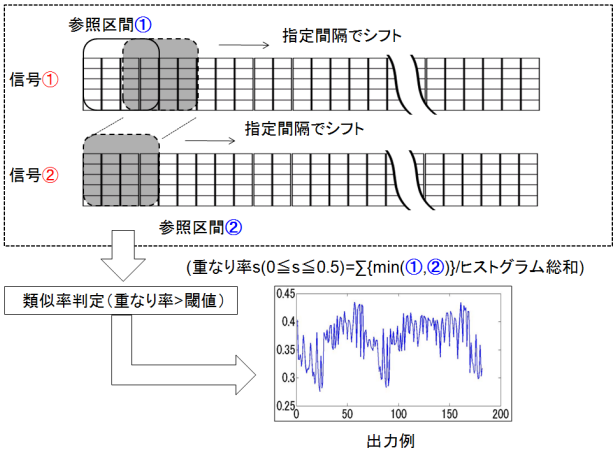

図 2　ヒストグラム特徴の照合例

$$S_i(h_i^R, h_i^I) = \frac{1}{D}\sum_{l=1}^{L} \min(h_{il}^R, h_{il}^I) \tag{1}$$

ここで h_i^R と h_i^I は，照合を行う互いの信号に対する i 番目の時間分割に対するヒストグラムであり，h_{il}^R と h_{il}^I はそれぞれのヒストグラムの l 番目のビンに含まれる要素数である．また，L はヒストグラムのビンの数，D_i は i 番目のヒストグラムの総度数である．参照区間の時間窓全体における類似度 S は，S_i を用いて次のように定義できる．

$$S(h^R, h^I) = \min_i(S_i(h_i^R, h_i^I)) \tag{2}$$

3.3.1 類似率と繊密計算区間

ヒストグラム特徴は特徴ベクトルの時系列を分類し累積したものであるため，ずらし照合を行う際，時間幅を長くしない限り両信号間の類似度が急激に変化することは無い．あらかじめの照合移動幅を長くし，一定の閾値を超えた瞬間の前後区間のみを細かく照合させることで，探索精度を十分に維持したまま数十倍の速度で探索が可能だ．つまり類似度の低い区間では照合を大きくスキップし，類似度が高い時点では繊密な類似度を計算するという適応的な動作となる．

表 1　実験に用いた計算機の仕様

ソフトウェア	Matlab:MathWorks 社
マシン OS	Microsoft:Windows7 Enterprise 64bit
CPU	Intel core i7 860(2.8 GHz)
メモリ	8GB

4　評価実験

パラメータを決定するにあたり，実験環境を表 1 に示す．

評価実験のために，ランダムに 60 秒から 372 秒の楽曲 10000 曲を用意した．

サンプル数を 10000 曲にした理由は，個人の保有楽曲数は多くても数千曲程度なことがほとんどだからである．

合計音響時間は 691 時間 51 分 27 秒，1 曲あたりの平均音響時間は約 249 秒である．この 10000 曲同士の類似性を全て判定する為には 10000 の階和から自らを引いた回数，つまり 49995000 回の照合が必要となる．いずれのサンプルも標本化周波数 11025Hz，量子化精度 8bit 直線，モノラルである．

4.1　相関法の探索時間評価

相関法を用いて探索に必要な時間をヒストグラム照合と同様の条件で計測した．ランダムに選んだ楽曲 100 曲の照合を 3 回繰り返し行った結果，平均で 9438 秒を要した．10000 曲の照合に必要な照合回数は 49995000 回であり，相関法を用いた場合，理論値では約 1103 日を要する．

4.2　ヒストグラム特徴の照合

各音響信号は 2 次の IIR バンドパスフィルタを用いて特徴抽出を行う．フィルタの中心周波数とヒストグラム特徴のビン数の決定は，フィルタリング後の各周波数帯域に含まれるパワーの総和が，全ての楽曲を平均して同等といえる量をとる．これは特徴抽出の際，全ての楽曲においてヒストグラムの各ビンの数値に偏りが生じることによる類似性判定の精度の低下を防ぐためである．

サンプリング周波数は 11025Hz だが，5500Hz や 1Hz に近い周波数帯域によっては音響情報 (パワースペクトルの総和) が少なく，特徴抽出した際に大きな変化が見られないことから，特徴が多く含まれる最適な範囲を決定する必要がある．

対数周波数軸上における分割する数についても検討する．対数周波数軸における分割数が多い場合，全ての楽曲に対して近い結果になってしまうという問題がある．また少なすぎても同様の結果を導いてしまう可能性がある為，最適な値を求めなければならない．

実際に 10000 曲のサンプルから，対数周波数軸を元に 1Hz から 5500Hz 間の各周波数帯域に含まれるパワーの割合を 100 帯域求めた．これによってパワーの平均が偏らないよう，表 2 のように分割する帯域幅の基準となる値を決定した．

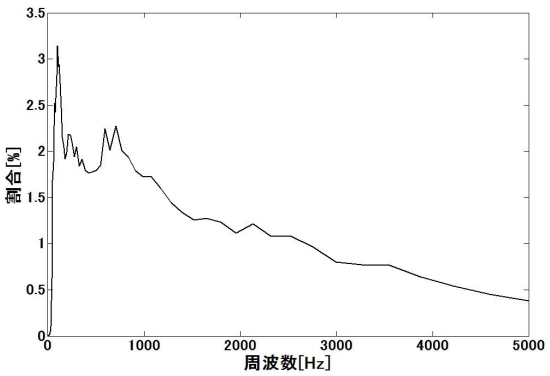

図 3　各周波数帯域に含まれるパワーの平均割合

表 3　抽出時間と探索時間

n	h(ms)	抽出 (s)	探索（s）	精度 (%)
3	1024	85	445	100
	512	123	1782	100
	256	220	7672	100
5	1024	135	448	100
	512	205	1790	100
	256	306	7128	100
7	1024	186	452	100
	512	286	1788	100
	256	530	7227	100

4.3　特徴抽出におけるパラメータ

最終的に最適なパラメータの決定を行った．決定実験は 2 通り行う．

まず特徴を抽出する際に，曲の長さによって十数 ms〜数千 ms で抽出するか，または時系列におけるデータ数を固定とするかである．前者であればより時系列ごとの詳細な検索が可能となり，類似区間が仮に短かったとしても検出可能である．後者であれば大域的なデータを照合させるため，前者と比べ照合回数が大幅に少ないため，高速な探索が可能となる．

ミリ秒単位で抽出を行う場合について説明する．分割数 n は「3,5,7」の 3 通り，時間軸上における分割間隔 h は「256ms, 512ms, 1024ms」の 3 通りの計 9 通りについての実験結果を表 4.3 に示した．実験内容はランダムに 100 曲を選出し，特徴抽出と全照合を繰り返し 10 回行った平均値で算出した．

各パラメータの最適な組み合わせは，探索速度と類似曲の検出精度によって決定する．この実験では対数周波数軸上における分割数を 3 通り，時間軸上における分割間隔を 6 通りの計 30 通りの評価を行った．

表 3 が実験結果である．音響データをメモリ上にロードしてから，メモリ上で処理が完了し，結果が出力されるまでを CPU 時間で計測した．照合回数の多いものから順に計測したところ，探索精度は全て 100% であり，探索速度がヒストグラム特徴の帯域分割数にほとんど左右されていない．また，同一曲の探索に関して全ての探索で 100% の精度となった．

次に特徴量の次元数を特徴抽出時にすべての楽曲に対して同一にする実験を行った．全て統一するメリットは，ヒストグラム照合の際，時系列における詳細な照合必要がないため，照合回数を最低限に抑えられる．分割数の少ない組み合わせから順番に精度の評価を行った．

今回の実験では周波数軸上 3 分割，時間軸上 12 分割で特徴抽出を行った場合に，精度 100% を維持したまま，最も速い探索結果が得られた．また同一楽曲でもサンプリング方法の違いにより，先頭や後尾に無音区間が存在するものがあったが，無音区間 3 秒程度までであれば同一曲の検出に頑健であることがわかった．

4.4　クラスタ構築による照合回数の削減

前節で得られたパラメータを元に抽出した 10000 曲分のヒストグラム特徴を元にクラスタ構築を行う．あらかじめ各楽曲の特徴量を分類することで，全探索における照合回数の削減を期待できる．ヒストグラム特徴の周波数帯域に含まれるパワースペクトルの総和平均を n 次元空間座標としたとき，空間内に配置された座標間の距離は特徴量全体の類似性となる．

今回の実験では非階層クラスタリングの代表的アルゴリズムである k-means 法を用いてクラスタ構築を行う．k-means 法を用いた理由は，全時間領域における各周波数帯域が属性ベクトルとして記述されているため，クラスタ構築が容易に行えると考えたためだ．まず各座標にプロトタイプを分けたいクラスタ数だけランダムに与え，クラスタに含まれるデータの重心を求める．そして各要素の所属するクラスタを最も近いクラスタに変更する．同様の処理を繰り返し，各クラスタに所属する要素の重心の移動距離の総和が最も少なくなったときにクラスタ

表 2　各帯域におけるパワー総和が均衡する帯域

分割数	始点	分割点 2	分割点 3	分割点 4	分割点 5	分割点 6	分割点 7	終点
2	1Hz	253Hz	5500Hz	-	-	-	-	-
3	1Hz	128Hz	546Hz	5500Hz	-	-	-	-
4	1Hz	99Hz	253Hz	767Hz	5500Hz	-	-	-
5	1Hz	91Hz	180Hz	422Hz	991Hz	5500Hz	-	-
6	1Hz	76Hz	128Hz	253Hz	546Hz	1175Hz	5500Hz	-
7	1Hz	70Hz	108Hz	180Hz	327Hz	647Hz	1279Hz	5500Hz

表 4　最終実験結果

照合方法	相関法 (理論値)	ヒストグラム照合	ヒストグラム照合 (クラスタリング後)
クラスタ構築 (s)	-	-	50
特徴抽出 (s)	-	6,139	6,139
照合時間 (s)	95,387,783	33,718	10,202
合計時間 (s)	95,387,783	39,857	16,391
照合回数 (回)	49,995,000	49,995,000	18,880,776
速度比	1(基準)	2,493	6,845
探索精度 (%)	100	100	100

構築が完了する．

今回の実験ではクラスタ数 7 まで実験を行った．クラスタ構築を繰り返し行なったところ，平均的に 10000 曲のヒストグラム特徴が 1400 前後の要素から成る 7 つのクラスタに分類されたが，厳密に同数程度の要素に分類はされなかった．またクラスタ構築に必要な時間は各要素のプロトタイプのクラスタによって上下するが，平均約 50 秒でクラスタ構築が可能であることがわかった．最終的にクラスタリングを行った際の実験結果は表 4 の通りである．

5　考察

今回の実験では，ヒストグラム特徴を抽出する際に，あらかじめ 10000 曲のサンプルから求めた対数周波数軸上におけるパワースペクトルの分布図 3 より，表 2 のような平均的な中心周波数を求めたことで，特徴抽出して得た一曲ごとのヒストグラム特徴に多様性を持たせることができ，探索精度 100% を維持したまま，特徴量を 36 次元まで圧縮することができた．また，特徴抽出時の時間領域と周波数領域におけるパラメータを固定し，照合回数を最小値に設定することで詳細区間の検索に比べ，同一曲の検索では大幅に探索時間を短縮できた．

クラスタ構築においては，周波数領域におけるパワーの総和平均から座標間ユークリッド距離を求め，k-means アルゴリズムを用いて 7 つのクラスタに分類することで，今回の実験ではもれなく探索ができ，クラスタリングを行う前と比べ約 2.7 倍の探索速度の向上を図れた．

サンプル数 10000 で行った全照合に要する時間は，クラスタリングを行った場合約 4.6 時間であった．これはまだ実用的な時間であるとは言えない．100% の精度を維持したまま照合の高速化を図るためには，クラスタ数を増やし，照合における探索漏れを引き起こさないために新たなクラスタリング手法を導入しなければならない．

6　あとがき

本論文では，複数の異なる音響信号同士の類似率から，重複楽曲を探索するシステムを提案した．提案法は，各信号の時系列ごとの指定周波数帯域のパワーの総和からヒストグラム化されたヒストグラム特徴を抽出し，そのヒストグラムの重なり率を類似度として判定するというものであった．あらかじめ特徴抽出し，クラスタ構築による参照区間の限定を行っておいた場合の探索に要する時間は 1 万曲あたり約 2.7 秒であった．実用上は，新たに加わった音響信号に対しても特徴抽出を行い，最も近いクラスタの重心を求め，クラスタに所属する特徴量のみを参照すればよい．

今後の課題として，一つ目に探索速度の向上が挙げられる．特徴量の次元数を更に圧縮することで探索速度の向上は見込めるが，異なる楽曲が同一曲として検出される可能性が大きい．サンプル数を更に増やし，多種多様なジャンルの楽曲を収集する必要もある．また最近では採譜技術を用いて波形データなどの音響信号から MIDI データのような音符号を取り出すことも可能となっており，MIDI データから楽曲のジャンルの自動分類を行う研究もある．つまり音響信号からジャンルによる自動分類も可能であると考えられ，照合数を減らすことも可能だ．

また今回の k-means アルゴリズムを用いた各周波数帯域上の平均パワー総和を空間上の座標としたときの座標間ユークリッド距離によるクラス分類では，探索対象曲が数十万から数百万曲となったときに，各クラスタ間の境界線付近の座標で探索漏れが起こることが考えられる．様々なクラスタリング手法を用いて何通りもの評価実験を行い，精度を維持したまま特徴量の次元数を圧縮の限界を見極めることが重要だ．

二つ目の課題に，探索対象曲が類似楽曲だった場合の問題がある．これはそもそもの類似楽曲の定義が曖昧であることが挙げられる．音のキーを上げたものや，歌い手が異なるもの，原曲にアレンジを加えテクノ調にしたものなど様々だ．つまり対象類似楽曲別のパラメータを決定し，それぞれに応じた照合区間や照合方法を変化させることで，探索速度や探索精度の低下を防げるのではないかと考えている．

参考文献

[1] 帆足 他，”楽曲配信サービスを支える音楽情報検索技術，”信学誌．88(7)，529-534，2005-07-01

[2] 辻 他，”曲の局所パターン特徴量を用いた類似曲検索・感性語による検索，”信学技報．SP，音声 96(565)，17-24，1997-03-06

[3] 獅々堀 他，”楽 Earth Mover’s Distance を用いたハミングによる類似音楽検索手法，”情処学論，48(1)， 300-311，2007-01-15

[4] 柏野 他，”ヒストグラム特徴を用いた音響信号の高速探索法 : 時系列アクティブ探索法，”信学論．D-II, 情報・システム，II-パターン処理 J82-D-II(9)，1365-1373，1999-09-25

[5] 山本 他，”ハミング検索における統計的方法の検討，”JSCS 大会論文集 (21)， 141-144， 2007-05

[6] 市川 他，”複数の音程特徴量によるハミング入力楽曲検索システムの高精度化，”情処学研報． [音楽情報科学] 2008(12)， 7-12， 2008-02-08

[7] 津田 啓夢．”音楽検索サービス「LISMO Music Search」”．ケータイ Watch．（オンライン），入手先，http://k-tai.impress.co.jp/cda/article/news_toppage/32745.html（参照 2011-01-28）．