

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5115966号
(P5115966)

(45) 発行日 平成25年1月9日(2013.1.9)

(24) 登録日 平成24年10月26日(2012.10.26)

(51) Int.Cl. F I
G06F 17/30 (2006.01)
 G06F 17/30 170E
 G06F 17/30 350C
 G06F 17/30 360Z

請求項の数 4 (全 28 頁)

<p>(21) 出願番号 特願2007-298695 (P2007-298695) (22) 出願日 平成19年11月16日(2007.11.16) (65) 公開番号 特開2009-123124 (P2009-123124A) (43) 公開日 平成21年6月4日(2009.6.4) 審査請求日 平成22年1月8日(2010.1.8)</p> <p>特許法第30条第1項適用 2007年8月1日 社団法人 情報処理学会発行の「情報処理学会研究報告 情報処研報 Vol. 2007 No. 81」に発表</p> <p>特許法第30条第1項適用 2007年11月9日 「日本経済新聞」に発表</p>	<p>(73) 特許権者 301021533 独立行政法人産業技術総合研究所 東京都千代田区霞が関1-3-1</p> <p>(74) 代理人 100091443 弁理士 西浦 ▲嗣▼晴</p> <p>(72) 発明者 藤原 弘将 茨城県つくば市東1-1-1 独立行政法人産業技術総合研究所 つくばセンター内</p> <p>(72) 発明者 後藤 真孝 茨城県つくば市東1-1-1 独立行政法人産業技術総合研究所 つくばセンター内</p> <p>審査官 吉田 誠</p>
---	---

最終頁に続く

(54) 【発明の名称】 楽曲検索システム及び方法並びにそのプログラム

(57) 【特許請求の範囲】

【請求項1】

複数の楽曲について、それぞれの楽曲の歌声の声質特徴量及び曲調特徴量と各楽曲の識別情報とを蓄積する声質及び曲調特徴量蓄積手段と、

前記識別情報と対応させて前記複数の楽曲データを蓄積する楽曲データ蓄積手段と、

前記複数の楽曲の中から一つの楽曲が選択されると、該一つの楽曲の前記声質特徴量とその他の前記複数の楽曲のそれぞれの前記声質特徴量とに基づいて、前記一つの楽曲と前記その他の複数の楽曲の声質類似度を計算する声質類似度計算手段と、

前記複数の楽曲の中から他の一つの楽曲が選択されると、前記他の一つの楽曲の前記曲調特徴量とその他の前記複数の楽曲のそれぞれの前記曲調特徴量とに基づいて、前記他の一つの楽曲と前記その他の複数の楽曲の曲調類似度を計算する曲調類似度計算手段と、

前記声質類似度計算手段により計算された前記声質類似度と、前記曲調類似度計算手段により計算された前記曲調類似度とに基づいて、前記一つの楽曲中の前記歌声と前記声質が類似する歌声を含み且つ前記他の一つの楽曲中の前記楽曲と曲調が類似する複数の楽曲についての複数の前記識別情報をディスプレイ上に表示する類似楽曲検索表示手段と、

前記ディスプレイ上に表示された前記複数の識別情報の表示から、1以上の前記識別情報が選択されると、前記選択された前記識別情報の表示に対応する前記楽曲データを前記楽曲データ蓄積手段に蓄積された複数の前記楽曲データから抽出して再生する楽曲データ再生手段とを備え、

前記類似楽曲検索表示手段は、前記一つの楽曲と前記その他の複数の楽曲の前記声質類

10

20

似度と前記他の一つの楽曲と前記その他の複数の楽曲の曲調類似度の前記曲調類似度の大小関係が判る表示様式により、前記複数の楽曲の前記識別情報を前記ディスプレイ上に表示するように構成されていることを特徴とする楽曲検索システム。

【請求項 2】

前記表示様式は、縦軸及び横軸の一方が前記声質類似度を示し、他方が前記曲調類似度を示し、一つの前記楽曲の前記識別情報を、該一つの楽曲の前記声質類似度の位置で軸と直交する第 1 の仮想線と該一つの楽曲の前記曲調類似度の位置で軸と直交する第 2 の仮想線の交差点またはその近傍に表示するように定められている請求項 1 に記載の楽曲検索システム。

【請求項 3】

複数の楽曲について、それぞれの楽曲の歌声の声質特徴量及び曲調特徴量と各楽曲の識別情報とを声質及び曲調特徴量蓄積手段に蓄積するステップと、

前記識別情報と対応させて前記複数の楽曲データを楽曲データ蓄積手段に蓄積するステップと、

前記複数の楽曲の中から一つの楽曲が選択されると、該一つの楽曲の前記声質特徴量とその他の前記複数の楽曲のそれぞれの前記声質特徴量とに基づいて、前記一つの楽曲と前記その他の複数の楽曲の声質類似度を計算するステップと、

前記複数の楽曲の中から他の一つの楽曲が選択されると、前記他の一つの楽曲の前記曲調特徴量とその他の前記複数の楽曲のそれぞれの前記曲調特徴量とに基づいて、前記他の一つの楽曲と前記その他の複数の楽曲の曲調類似度を計算するステップと、

前記声質類似度と前記曲調類似度とに基づいて、前記一つの楽曲中の前記歌声と前記声質が類似する歌声を含み且つ前記他の一つの楽曲中の前記楽曲と曲調が類似する複数の楽曲についての複数の前記識別情報をディスプレイ上に表示する検索表示ステップと、

前記ディスプレイ上に表示された前記複数の識別情報の表示から、1 以上の前記識別情報が選択されると、前記選択された前記識別情報の表示に対応する前記楽曲データを前記楽曲データ蓄積手段に蓄積された複数の前記楽曲データから抽出して楽曲データ再生手段により再生するステップとをコンピュータを用いて実行し、

前記検索表示ステップでは、前記一つの楽曲と前記その他の複数の楽曲の前記声質類似度と前記他の一つの楽曲と前記その他の複数の楽曲の曲調類似度の前記曲調類似度の大小関係が判る表示様式により、前記複数の楽曲の前記識別情報を前記ディスプレイ上に表示することを特徴とする楽曲検索方法。

【請求項 4】

複数の楽曲について、それぞれの楽曲の歌声の声質特徴量及び曲調特徴量と各楽曲の識別情報とを声質及び曲調特徴量蓄積手段に蓄積するステップと、

前記識別情報と対応させて前記複数の楽曲データを楽曲データ蓄積手段に蓄積するステップと、

前記複数の楽曲の中から一つの楽曲が選択されると、該一つの楽曲の前記声質特徴量とその他の前記複数の楽曲のそれぞれの前記声質特徴量とに基づいて、前記一つの楽曲と前記その他の複数の楽曲の声質類似度を計算するステップと、

前記複数の楽曲の中から他の一つの楽曲が選択されると、前記他の一つの楽曲の前記曲調特徴量とその他の前記複数の楽曲のそれぞれの前記曲調特徴量とに基づいて、前記他の一つの楽曲と前記その他の複数の楽曲の曲調類似度を計算するステップと、

前記声質類似度と前記曲調類似度とに基づいて、前記一つの楽曲中の前記歌声と前記声質が類似する歌声を含み且つ前記他の一つの楽曲中の前記楽曲と曲調が類似する複数の楽曲についての複数の前記識別情報をディスプレイ上に表示する検索表示ステップと、

前記ディスプレイ上に表示された前記複数の識別情報の表示から、1 以上の前記識別情報が選択されると、前記選択された前記識別情報の表示に対応する前記楽曲データを前記楽曲データ蓄積手段に蓄積された複数の前記楽曲データから抽出して楽曲データ再生手段により再生するステップとをコンピュータに実行させるように構成され、

前記検索表示ステップは、前記一つの楽曲と前記その他の複数の楽曲の前記声質類似度

10

20

30

40

50

と前記他の一つの楽曲と前記その他の複数の楽曲の曲調類似度の前記曲調類似度の大小関係が判る表示様式により、前記複数の楽曲の前記識別情報を前記ディスプレイ上に表示するように構成されていることを特徴とする楽曲検索性プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、声質が類似する楽曲を検索できる楽曲検索性システム及び方法並びにそのプログラムに関するものである。

【背景技術】

【0002】

近年、音楽検索が重要性を増してきている。携帯型音楽プレーヤやWeb上でのオンライン楽曲販売サービスの普及により、ユーザ達は膨大な量の楽曲から好みの楽曲を検索し、どんな時でもどんな場所でも聴きたい音楽を鑑賞出来るようになった。これにより、自分の好みの楽曲を手がかりに、まだ聴いたことのない楽曲を発見したいという要求が生じてきた。しかし、探したい対象の楽曲が未知で、「好み」のような曖昧な情報のみが手がかりの場合、従来の歌手名、ジャンル名等の書誌情報のみに基づく検索システムは無効だった。このような動向から、楽曲の内容に基づく音楽検索性システムに関して、多くの研究がなされてきた（非特許文献1乃至9）。

【非特許文献1】Aucouturier, J. - J. and Pachet, F.: Music Similarity Measures: What's the Use?, Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR2002), pp. 157 - 163 (2002).

【非特許文献2】Logan, B.: Content-Based Playlist Generation: Exploratory Experiments, Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR2002), pp. 295 - 296 (2003).

【非特許文献3】Allamanche, E., Herre, J., Hellmuth, O., Kastner, T. and Ertel, C.: A Multiple Feature Model for Musical Similarity Retrieval, Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR2003), pp. 217 - 218 (2003).

【非特許文献4】Berenzweig, A., Logan, B., Ellis, D. P. W. and Whitman, B.: A Large-Scale Evaluation of Acoustic and Subjective Music Similarity Measures, Computer Music Journal, Vol. 28, No. 2, pp. 63 - 76 (2004).

【非特許文献5】McKinney, M. F. and Breebaart, J.: Features for audio and music classification, Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR2003), pp. 151 - 158 (2003).

【非特許文献6】Tzanetakis, G., Gao, J. and Steenkiste, P.: A Scalable Peer-to-Peer System for Music Content and Information Retrieval, Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR2003), pp. 209 - 214 (2003).

10

20

30

40

50

【非特許文献7】Pampalk, E., Flexer, A. and Widmer, G.: Improvements of Audio-based Music Similarity and Genre Classification, Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR2005), pp. 628 - 633 (2005).

【非特許文献8】Flexer, A., Gouyou, F., Dixon, S. and Widmer, G.: Probabilistic combination of features for music classification, Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR2006), pp. 628 - 633 (2006).

10

【非特許文献9】Pohle, T., Knees, P., Schedl, M. and Widmer, G.: Independent Component Analysis for Music Similarity Computation, Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR2006), pp. 228 - 233 (2006).

【発明の開示】

【発明が解決しようとする課題】

20

【0003】

しかし、これらの研究で示された楽曲検索技術では、音楽の内容を表現するのに、主に楽曲の曲調を表現するMFCC、スペクトル重心、ロールオフ、フラックスなどの特徴量を用いており、歌声（ボーカル）の声質などのより詳細な内容を表現する特徴量は用いられていなかった。そのため従来は、声質が類似する歌声を含む楽曲の検索をすることができなかつた。

【0004】

本発明の目的は、声質が類似する歌声を含む未知の楽曲の検索をすることができる楽曲検索システム及び方法並びにそのプログラムを提供することにある。

【0005】

30

上記目的に加えて、本発明の他の目的は、声質と曲調が類似する歌声を含む未知の楽曲の検索をすることができる楽曲検索システム及び方法並びにそのプログラムを提供することにある。

【0006】

上記目的に加えて、本発明のさらに他の目的は、ある楽曲中の歌声の声質と類似する歌声を含み、しかも別の楽曲の曲調と類似する曲調の未知の楽曲を検索することができる楽曲検索システム及び方法並びにそのプログラムを提供することにある。

【0007】

上記目的に加えて、本発明のさらに他の目的は、声質または曲調が類似する未知の楽曲の選択が容易な楽曲検索システムを提供することにある。

40

【課題を解決するための手段】

【0008】

本発明の楽曲検索システムは、声質特徴量蓄積手段と、類似度計算手段と、類似楽曲検索表示手段と、楽曲データ再生手段とを備えている。声質特徴量蓄積手段は、複数の楽曲について、それぞれの楽曲中の歌声の声質特徴量と各楽曲の識別情報とを蓄積する。複数の楽曲は、歌声と伴奏音とからなる楽曲だけでなく、歌声のみから楽曲のいずれも含まれる。そして楽曲の歌声の声質特徴量を抽出するためには、歌声と伴奏音からなる楽曲では、伴奏音の影響を低減（伴奏音を抑制）した音響信号中から声質に関する特徴量（声質特徴量）を抽出する。抽出方法は、任意である。声質特徴量としては、声質の特徴を示すものであればどのようなものでも用いることができる。例えば、公知のLPCメルケプスト

50

ラム係数 (LPMCC) と F0 (歌手の歌い方を表現する特徴量で、F0 軌跡の時間変化係数) を声質特徴量として用いることができる。また「楽曲の識別情報」とは、楽曲を選択する際にディスプレイに表示する情報を含むものであり、例えば、曲名、歌手名、楽曲のジャンル等である。声質特徴量は、予め検索対象とするすべての楽曲について求めておき、声質特徴量蓄積手段に蓄積しておくことになる。なお新たな楽曲についても、声質特徴量の追加登録をできるようにしてもよいのは勿論である。

【0009】

そして楽曲データ蓄積手段には、複数の識別情報と対応させて複数の楽曲データを蓄積する。楽曲データは、楽曲データ再生手段によって再生する際に利用する音響信号に対応するデータである。

10

【0010】

類似度計算手段は、複数の楽曲の中から一つの楽曲が選択されると、該一つの楽曲の声質特徴量とその他の複数の楽曲のそれぞれの声質特徴量とに基づいて、選択された一つの楽曲とその他の複数の楽曲の声質類似度を計算する。すなわち選択された楽曲の音質特徴量と、その他の複数の楽曲の声質特徴量 (声質特徴量蓄積手段に蓄積された残りの楽曲の声質特徴量) のすべてとの間において、類似度を計算する。楽曲の選択は、例えば、ディスプレイ上に複数の楽曲の少なくとも一部の識別情報を表示し、表示させた識別情報を選択することにより、楽曲を選択するようにしてもよい。またディスプレイ上の特定の入力場所に、手入力により、楽曲の曲名を入力する等により楽曲を選択してもよく、選択方法は任意である。また類似度の計算方法も、任意であって、特定の類似度計算方法に限定されるものではない。類似度の計算には、例えば、2つの楽曲間の類似尺度としては、相互情報量の他、COS 尺度、距離尺度 EMD (Earth Mover's Distance) 等も用いることができる。

20

【0011】

類似楽曲検索表示手段は、類似度計算手段により計算された複数の声質類似度に基づいて、一つの楽曲中の歌声の声質に類似する歌声を含む複数の楽曲の複数の識別情報をディスプレイ上に表示する。類似する歌声を含む複数の楽曲の複数の識別情報をディスプレイ上に表示する表示様式は、類似度の大小関係が分かるものであればどのような様式であってもよい。例えば、類似度の大小関係が判る表示様式としては、類似度の大きい順番に識別情報を順番に並べてディスプレイ上に表示する様式や、類似度の大きい順番に識別情報をディスプレイ上に一つずつ順番に表示する様式や、ディスプレイ上に表示する複数の識別情報の大きさ (寸法) を類似度の大きさに比例して大きくなるように定める様式等を用いることができる。

30

【0012】

また楽曲データ再生手段は、ディスプレイ上に表示された複数の識別情報の表示から、1以上の識別情報が選択されると、選択された識別情報の表示に対応する楽曲データを楽曲データ蓄積手段に蓄積された複数の楽曲データから抽出して再生する。楽曲の再生により、検索結果を実際に確認することが可能になる。なお楽曲データ再生手段による楽曲データの再生は、楽曲データの先頭から再生してもよいが、いわゆる楽曲のサビ部分だけを再生するようにしてもよい。また楽曲中の複数の歌声部分から選択した適宜の部分から楽曲データを再生するようにしてもよい。

40

【0013】

本発明によれば、ユーザが気に入った楽曲中の歌声の声質と似た歌声を含む楽曲を、未知の複数の楽曲中から高い精度で検索することができる。

【0014】

声質に加えて、楽曲の曲調も類似する楽曲を選択できるようにしてもよい。その場合には、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量及び曲調特徴量と各楽曲の識別情報とを蓄積する声質及び曲調特徴量蓄積手段と曲調類似度計算手段とを用意する。ここで曲調特徴量を抽出するためには、歌声と伴奏音からなる楽曲では、歌声の影響を低減 (歌声を抑制) した音響信号中から伴奏音に関する特徴量 (曲調特徴量) を抽出する。

50

抽出方法は、任意である。曲調特徴量としては、伴奏音の特徴を示すものであればどのようなものでも用いることができる。例えば、M F C C、スペクトル重心、ロールオフ、フラックスなどの特徴量を用いることができる。

【 0 0 1 5 】

曲調類似度計算手段は、複数の楽曲の中から一つの楽曲が選択されると、一つの楽曲の曲調特徴量とその他の複数の楽曲のそれぞれの曲調特徴量とに基づいて、一つの楽曲とその他の複数の楽曲の曲調類似度を計算する。すなわち選択された楽曲の曲調特徴量と、その他の複数の楽曲の曲調特徴量（声質及び曲調特徴量蓄積手段に蓄積された残りの楽曲の曲調特徴量）のすべてとの間において、類似度を計算する。また類似度の計算方法も、任意であって、特定の類似度計算方法に限定されるものではない。曲調の類似度の計算にも、例えば、2つの楽曲間の類似尺度として、相互情報量を使用することができる。また類似尺度としては、相互情報量の他、C O S 尺度、距離尺度 E M D (Earth Mover's Distance) 等も用いることができる。

10

【 0 0 1 6 】

類似楽曲検索表示手段は、声質類似度計算手段により計算された声質類似度と、曲調類似度計算手段により計算された曲調類似度とに基づいて、声質が類似する歌声を含み且つ曲調が類似する複数の楽曲の複数の識別情報をディスプレイ上に表示する。声質が類似する歌声を含み且つ曲調が類似する複数の楽曲は、1つの楽曲の声質類似度及び曲調類似度の平均値を基準としてもよく、声質類似度及び曲調類似度のそれぞれに所定の重み付けをした値の平均値を基準として検索を行って、平均値の大きい順に検索された複数の楽曲の識別情報をディスプレイ上に並べるようにしてもよい。また縦軸及び横軸の一方が声質類似度を示し、他方が曲調類似度を示し、一つの楽曲の識別情報を、該一つの楽曲の声質類似度の位置で軸と直交する第1の仮想線と曲調類似度の位置で軸と直交する第2の仮想線の交差点またはその近傍に表示する表示様式により、平面上に複数の楽曲の識別情報を分散して配置したような状態で表示するようにしてもよい。すなわち声質類似度及び曲調類似度が共に同じ値で、この値がそれぞれ異なる複数の楽曲の識別情報は、縦軸と横軸の交点から縦軸と横軸の間の45度の角度方向に延びる仮想中央線上に並んで配置されることになる。この場合、声質類似度及び曲調類似度が共に最も大きい楽曲の識別情報が縦軸と横軸の交点に位置するようにすると、交点近傍の識別情報を選択すれば、声質及び曲調が共に類似した楽曲を素早く見つけることが可能になる。

20

30

【 0 0 1 7 】

声質の類似を検索する際に基準とする楽曲と曲調の類似を検索する際に基準とする楽曲とを異ならせてもよい。すなわち声質はある楽曲中の歌声（例えば一人の女性がロックを歌っている楽曲の歌声）の声質と類似し、曲調は他の楽曲（例えばコーラスグループがポップスを歌っている楽曲）の曲調と類似する楽曲を検索できるようにしてもよい。この場合に、声質及び曲調特徴量蓄積手段と、声質類似度計算手段と、曲調類似度計算手段、類似楽曲検索表示手段と、楽曲データ再生手段とを用いる。声質類似度計算手段は、複数の楽曲の中から一つの楽曲が選択されると、該一つの楽曲の声質特徴量とその他の複数の楽曲のそれぞれの声質特徴量とに基づいて、一つの楽曲とその他の複数の楽曲の声質類似度を計算する。また曲調類似度計算手段は、複数の楽曲の中から他の一つの楽曲が選択されると、他の一つの楽曲の曲調特徴量とその他の複数の楽曲のそれぞれの曲調特徴量とに基づいて、他の一つの楽曲とその他の複数の楽曲の曲調類似度を計算する。そして類似楽曲検索表示手段は、声質類似度計算手段により計算された声質類似度と、曲調類似度計算手段により計算された曲調類似度とに基づいて、一つの楽曲中の歌声と声質が類似する歌声を含み且つ他の一つの楽曲中の楽曲と曲調が類似する複数の楽曲についての複数の識別情報をディスプレイ上に表示する。ディスプレイ上への識別情報の表示様式は任意である。例えば、前述のように、縦軸及び横軸の一方が声質類似度を示し、他方が曲調類似度を示し、一つの楽曲の識別情報を、該一つの楽曲の声質類似度の位置で軸と直交する第1の仮想線と該一つの楽曲の曲調類似度の位置で軸と直交する第2の仮想線の交差点またはその近傍に表示する表示様式を用いることができる。この表示様式を用いると、声質と曲調が

40

50

類似する複数の楽曲を、類似の程度と共に一目で認識することができる。

【0018】

上記基本発明では、毎回類似度を計算しているが、予め類似度を計算しておき類似度蓄積手段に蓄積しておいてもよい。声質が類似する楽曲を検索する場合には、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量とその他の複数の楽曲の声質特徴量とに基づいて算出された各楽曲とその他の複数の楽曲との間の複数の声質類似度と、各楽曲の識別情報とを蓄積する類似度蓄積手段を用意する。そして類似楽曲検索表示手段は、複数の楽曲の中から一つの楽曲が選択されると、類似度蓄積手段から、該一つの楽曲とその他の複数の楽曲との間の複数の声質類似度を読み出し、複数の声質類似度に基づいて一つの楽曲中の歌声の声質と類似する歌声を含む複数の楽曲の複数の識別情報をディスプレイ上に表示するように構成する。このようにすると事前に類似度の計算を行っておくため、検索時間を早くすることができる。

10

【0019】

また声質と曲調が類似する楽曲を検索する場合には、類似度蓄積手段は、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量とその他の複数の楽曲の声質特徴量とに基づいて算出された各楽曲とその他の複数の楽曲との間の複数の声質類似度と、複数の楽曲について、それぞれの楽曲の曲調特徴量とその他の複数の楽曲の曲調特徴量とに基づいて算出された各楽曲とその他の複数の楽曲との間の複数の曲調類似度と、各楽曲の識別情報とを蓄積する。そして類似楽曲検索表示手段は、複数の楽曲の中から一つの楽曲が選択されると、類似度蓄積手段から、該一つの楽曲とその他の複数の楽曲との間の複数の声質類似度及び複数の曲調類似度を読み出し、複数の声質類似度及び複数の曲調類似度に基づいて一つの楽曲中の歌声の声質と類似する歌声を含み及び該一つの楽曲の曲調と類似する複数の楽曲の複数の識別情報をディスプレイ上に表示する。このようにすると事前に類似度の計算を行っておくため、検索時間を早くすることができる。

20

【0020】

本発明を、コンピュータが実行する方法の発明として特定すると、コンピュータは以下のステップを実行することになる。まず複数の楽曲について、それぞれの楽曲の歌声の声質特徴量と各楽曲の識別情報とを声質特徴量蓄積手段に記憶するステップを実行する。次に複数の識別情報と対応させて複数の楽曲データを楽曲データ蓄積手段に蓄積するステップを実行する。そして複数の楽曲の中から一つの楽曲が選択されると、該一つの楽曲の声質特徴量とその他の複数の楽曲のそれぞれの声質特徴量とに基づいて、該一つの楽曲と前記その他の複数の楽曲の声質類似度を計算するステップを実行する。更に類似度計算手段により計算された声質類似度に基づいて、一つの楽曲中の歌声の声質に類似する歌声を含む複数の楽曲の複数の識別情報をディスプレイ上に表示するステップを実行する。そしてディスプレイ上に表示された複数の識別情報の表示から、一つの識別情報が選択され、更に再生指令が入力されると、選択された識別情報の表示に対応する楽曲データを楽曲データ蓄積手段に蓄積された複数の楽曲データから抽出して楽曲再生手段により再生するステップを実行する。

30

【0021】

本発明は上記ステップをコンピュータに実行させるプログラムとしても特定することができる。

40

【発明の効果】

【0022】

本発明によれば、ユーザが気に入った楽曲中の歌声の声質と似た歌声を含む楽曲を、未知の複数の楽曲中から高い精度で検索することができる。またユーザが気に入った楽曲中の歌声の声質と似た歌声を含み且つ曲調が似た楽曲を、未知の複数の楽曲中から高い精度で検索することができる。更に、ユーザが気に入った一つの楽曲中の歌声の声質と似た歌声を含み且つ別の一つの楽曲の曲調と似た楽曲を、未知の複数の楽曲中から高い精度で検索することができる。

【発明を実施するための最良の形態】

50

【 0 0 2 3 】

以下図面を参照して本発明の楽曲検索システム及び方法の実施の形態について詳細に説明する。図1は、楽曲検索システムの一実施の形態の基本的な技術的思想を概念的に示す図である。図1に示す一実施の形態の楽曲検索システムでは、データベースに予め登録された楽曲(target songs)TS1~TSnの中から、ユーザが提示したクエリ(query)即ち選択した楽曲と類似した声質を持つ楽曲を検索し、ユーザに提示する。このシステムは、大きく分けてデータベース構築段階(楽曲の音響信号の分析及び声質特徴量の抽出)と実行段階(楽曲の検索)の2段階からなる。データベース構築段階では、検索対象の楽曲はダウンロードされたり、CDからリップングされた後にデータベースDBに貯蓄される。そして、このシステムは各楽曲を分析し、歌声即ち歌声の声質を表現する特徴量(声質特徴量)を抽出する。実行段階では、ユーザが好みの楽曲をクエリとしてシステムに入力すると、システムはクエリの楽曲を分析し、楽曲の歌声の声質を表現する特徴量を抽出する。そして、システムはクエリの楽曲と、データベース中のそれぞれの楽曲で、歌声の声質の類似度を計算し、高い類似度を持つ楽曲を検索結果として出力する。

10

【 0 0 2 4 】

図2は、図1の概念を具体化した第1の楽曲検索システム1の具体的構成を示すブロック図である。この楽曲検索システム1は、声質特徴量蓄積手段2と、声質類似度計算手段3と、楽曲データ蓄積手段4と、類似楽曲検索表示手段5と、楽曲データ再生手段6、制御手段7と、楽曲選択手段8と、ディスプレイドライバ9と、ディスプレイ10とを備えている。ディスプレイドライバ9は、入力信号に基づいて、ディスプレイ10を駆動して各種の表示をディスプレイ10上に表示するための動作をする。そして楽曲選択手段8は、ディスプレイ10上に表示される各種の入力画面において、楽曲を選択するためにユーザによって操作される手段であって、例えばパソコンであればカーソルを動かすためのマウス等の操作手段である。また制御手段7は、システムをコンピュータで実現する場合において各種のプログラムを実行してデータを処理する際の中心をなすものであり、各手段からのデータを信号処理してディスプレイドライバ9に対して入力信号として出力するものである。

20

【 0 0 2 5 】

声質特徴量蓄積手段2は、複数の楽曲TS1~TSnについて、それぞれの楽曲中の歌声の声質特徴量と各楽曲の識別情報とを蓄積するデータベースである。複数の楽曲には、歌声と伴奏音とからなる楽曲だけでなく、歌声のみからなる楽曲のいずれもが含まれる。声質特徴量蓄積手段2に蓄積するために、各楽曲の歌声の声質特徴量を抽出するためには、歌声と伴奏音からなる楽曲では、伴奏音の影響を低減(伴奏音を抑制)した音響信号中から声質に関する特徴量(声質特徴量)を抽出する。抽出方法は、任意であり、声質特徴量としては、声質の特徴を示すものであればどのようなものでも用いることができる。

30

【 0 0 2 6 】

本実施の形態では、声質特徴量蓄積手段2に蓄積する声質特徴量を以下のようにして抽出する。まず歌声の声質を表現する特徴ベクトルを計算するために、発明者が先に提案した歌手名同定手法における特徴抽出法を使用する。この手法は、「Fujihara, H., Kitahara, T., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Singer Identification Based on Accompaniment Sound Reduction and Reliable Frame Selection, Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR2005), pp. 329-336 (2005)」に開示されている。この手法を用いると、楽曲中に混在する伴奏音の影響を低減することができる。そのため、伴奏音と歌声が混在した状態をそのまま表現するMFC Cなどの特徴量と比較して、この特徴ベクトルは歌声の声質をよりよく表現することができる。

40

【 0 0 2 7 】

50

この特徴抽出手法は、伴奏音抑制、特徴抽出、高信頼度フレーム選択の3つの処理からなる。伴奏音抑制とは、伴奏音の影響を低減させるために、歌声の高調波構造に基づき歌声を分離・再合成する手法である。特徴抽出では、分離された歌声から特徴量を計算する。高信頼度フレーム選択では、抽出された特徴ベクトル列中から、歌声と信頼できる区間（フレーム）を選択し、歌声を含まない区間や伴奏音の影響を大きく受けている区間を除去する。

【0028】

まず伴奏音抑制について説明する。伴奏音抑制では、楽曲中のメロディを再合成することで、伴奏音の影響を低減させる。伴奏音抑制は、以下の3つの処理からなる。

【0029】

(1) 歌声メロディの基本周波数(F0)を、発明者の一人である後藤真孝が先に提案したPreFEst法(Goto, M.: A real-time music-scene-description system: predominant-F0 estimation for detecting melody and basslines in real-world audio signals, Speech Communication, Vol. 43, No. 4, pp. 311-329 (2004))を用いて推定する。

【0030】

(2) 推定された基本周波数F0に基づき、歌声メロディの高調波構造を抽出する。

【0031】

(3) 正弦波重畳モデルに基づき、メロディの音響信号を再合成する。

【0032】

歌声メロディの基本周波数を推定するために、前述のPreFEst法を用いる。PreFEst法は、制限された周波数帯域内で、混合音中の最も優勢な基本周波数F0を推定する手法である。多くの場合、歌声のメロディラインは中高域で最も優勢な高調波構造である。そのため、適切に周波数帯域を制限することで、PreFEst法を用いればメロディの基本周波数F0を推定することができる。

【0033】

次に、推定された基本周波数F0を用いて、基本周波数成分と高調波成分のパワーを抽出する。各高調波成分の抽出では、rcentの誤差を許容し、その範囲内で最大となる周波数成分のパワーを抽出する。時刻(t)におけるl次倍音(l=1, . . . , L)の周波数

【数1】

$$F_l^{(t)}$$

【0034】

と振幅

【数2】

$$A_l^{(t)}$$

【0035】

は

10

20

30

40

【数 3】

$$F_l^{(t)} = \operatorname{argmax}_F |S^{(t)}(F)| \\ (\overline{F}^{(t)}(1 - 2^{\frac{r}{1200}}) \leq F \leq \overline{F}^{(t)}(1 + 2^{\frac{r}{1200}}))$$

【数 4】

$$A_l^{(t)} = |S^{(t)}(F_l)|$$

10

【0036】

のようになる。ここで、 $S^{(t)}(F)$ は複素スペクトルを表し、

【数 5】

$$\overline{F}^{(t)} \text{ PreFEst}$$

【0037】

によって推定された基本周波数 F_0 を表す。後述する本実施の形態の実験例では、 r を 20 に設定した。

20

【0038】

最後に、正弦波重畳モデルを用いて、抽出された高調波構造から

【数 6】

$$F_l^{(t)} \text{ and } A_l^{(t)}$$

【0039】

のメロディの音響信号を再合成する。周波数が線形に変化するように、位相の変化は 2 次関数を用いて近似する。また、振幅の変化も線形近似する。再合成された音響信号 $s(k)$ は、

30

【数 7】

$$s(k) = \sum_{l=1}^L s_l(k)$$

【数 8】

$$s_l(k) = \left\{ (A_l^{(t+1)} - A_l^{(t)}) \frac{k}{K} + A_l^{(t)} \right\} \sin(\theta_l(k))$$

40

【数9】

$$\theta_l(k) = \frac{\pi(F_l^{(t+1)} - F_l^{(t)})}{K} k^2 + 2\pi F_l^{(t)} k + \theta_{l,0}$$

【0040】

のように表現される。ここで、 k は秒を単位とする時間であり、 $k = 0$ は時刻(t)に相当する。また、 K は時刻 t と $t + 1$ の間隔で、単位は秒である。 $\theta_{l,1}$ 、 $\theta_{l,0}$ は初期位相を表す。

10

【0041】

次に特徴量の抽出について説明する。再合成された音響信号から、以下の2種類の特徴量を抽出する。

【0042】

・LPCメルケプストラム係数(LPMCC)(徳田恵一, 小林隆夫, 今井聖: メル一般化ケプストラムの再帰的計算法, 電子情報通信学会論文誌A, Vol. J71-A, No. 1, pp. 128-131(1988))

音響信号の個人性を表す特徴量は、スペクトル包絡に含まれていることが知られている。そこで本実施の形態では、そのような包絡を表現するスペクトル特徴量としてLPMCCを使用する。LPMCCが、音楽のモデリングによく使われるメル周波数ケプストラム係数(MFCC)などと比較して、歌手の個人性をよく表現することは、発明者等が以前行った歌手名同定の実験で確認されている(藤原弘将, 北原鉄朗, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃博: 伴奏音抑制と高信頼度フレーム選択に基づく楽曲の歌手名同定手法, 情報処理学会論文誌, Vol. 47, No. 6, pp. 1831-1843(2006))。

20

【0043】

・F0

歌手の歌い方を表現する特徴量として基本周波数F0の軌跡の時間変化係数であるF0を使用する。F0については、発明者等が発表した論文に記載されているので説明は省略する[Ohishi, Y., Goto, M., Itou, K. and Takeda, K.: Discrimination between Singing and Speaking Voices, Proceedings of 9th European Conference on Speech Communication and Technology (Eurospeech 2005), pp. 1141.1144 (2005).]

30

【0044】

歌声は、ビブラート、オーバーシュートなど、短時間の時間変動が激しいので、この特徴量は、歌手の特性をよく表現すると期待される。

【0045】

次に高信頼度フレーム選択について説明する。PreFEST法では、メロディは単に最も優勢な基本周波数F0として定義されているので、再合成された音響信号は歌声が存在する区間では歌声を含んでいるが、間奏部などではその他の楽器の音を含んでいる。さらに、歌声が存在する区間でも伴奏音の音量が大きい場合は、再合成された歌声は大きく歪んでいる場合がある。そのため、ここで得られた特徴ベクトル列は、一部に信頼できない区間を含んでいる可能性がある。そこで本実施の形態では、高信頼度フレーム選択によって、そのような信頼できない区間を除去し、信頼できる特徴量のみを用いて類似度計算できるようにした。これを実現するため、歌声混合ガウス分布(GMM) G_{V} と非歌声GMM G_{N} の2種類のGMMを導入する。歌声GMM G_{V} は歌声を含む区間から抽出された特徴量で学習され、非歌声GMM G_{N} は間奏部などから抽出された特徴量で学習される。特徴量 x が与えられると、歌声/非歌声GMMの尤度

40

【数 1 0】

$$p(\mathbf{x}|\lambda_V)$$

【0 0 4 6】

と

【数 1 1】

$$p(\mathbf{x}|\lambda_N)$$

10

【0 0 4 7】

は、この特徴量がどの程度歌声（または歌声以外の音）に近いかを表現する。これらの尤度を用いて、特徴量 x が信頼できるかどうかを次式で判定する。

【数 1 2】

$$\log p(\mathbf{x}|\lambda_V) - \log p(\mathbf{x}|\lambda_N) \begin{array}{c} \text{reliable} \\ \geq \\ < \\ \text{not-reliable} \end{array} \eta$$

20

【0 0 4 8】

ここで η は閾値を表す。全ての楽曲に対して閾値を定数として定めることは困難である。なぜなら、伴奏の影響が全体に大きい楽曲では選択されるフレームが少なくなりすぎる場合があるからである。そこで実施の形態では、楽曲全体の特徴量中の 15% の特徴量が選択されるように、楽曲に応じて閾値を自動決定した。声質特徴量は、予め検索対象とするすべての楽曲について求めておき、声質特徴量蓄積手段 2 に蓄積しておくことになる。そこで本実施の形態では、すべての楽曲に関して、高信頼度フレームにおける声質特徴量を識別情報と共に声質特徴量蓄積手段 2 に蓄積している。なお新たな楽曲についても、声質特徴量の追加登録をできるようにしてもよいのは勿論である。「楽曲の識別情報」とは、楽曲を選択する際にディスプレイに表示する情報を含むものであり、例えば、曲名、歌手名、楽曲のジャンル等である。

30

【0 0 4 9】

楽曲データ蓄積手段 4 には、複数の識別情報と対応させて複数の楽曲データを蓄積する。楽曲データは、楽曲データ再生手段によって再生する際に利用する音響信号に対応するデータである。制御手段 7 は、ディスプレイドライバ 9 を介してディスプレイ 10 の画面上に楽曲選択のために使用する複数の楽曲の識別情報を表示させる。ユーザはマウス等の楽曲選択手段 8 を用いて、ディスプレイ 10 の画面上に表示された楽曲の識別情報から好みの 1 曲の楽曲をクエリとして選択する。なお楽曲選択手段 8 の構成は任意である。楽曲の選択は、例えば、ディスプレイ 10 上に複数の楽曲の少なくとも一部の識別情報を表示し、表示させた識別情報を選択することにより、楽曲を選択するようにしてもよい。またディスプレイ 10 上の特定の入力場所に、手入力により、楽曲の曲名を所定の入力窓に入力する等により楽曲を選択してもよく、選択方法は任意である。画面上の所定入力窓に好みの 1 曲の楽曲の識別情報（例えば曲面）をマニュアルで入力する場合には、キーボード入力装置が楽曲選択手段 8 の一部を構成することになる。そして楽曲選択手段 8 からの指令信号は制御手段 7 で処理される。

40

【0 0 5 0】

クエリとなる 1 曲が選択されると、類似度計算手段 3 は、選択された一つの楽曲の声質特徴量とその他の複数の楽曲のそれぞれの声質特徴量を声質特徴量蓄積手段 2 から読み出

50

して、選択された一つの楽曲とその他の複数の楽曲の声質類似度を計算する。すなわち類似度計算手段3は、選択された楽曲の音質特徴量と、その他の複数の楽曲の声質特徴量（声質特徴量蓄積手段に蓄積された残りの楽曲の声質特徴量）のすべてとの間において、類似度を計算する。類似度計算手段3における類似度の計算方法は、任意である。本実施の形態では、2つの楽曲間の類似度尺度として、相互情報量を使用する。

【0051】

本実施の形態では、2つの楽曲（XとY）間の類似度尺度として相互情報量を使用するために、楽曲の特徴ベクトルの分布を、混合ガウス分布モデル（GMM）を用いてモデル化する。そして楽曲から抽出された特徴ベクトル列から、EMアルゴリズムを用いて各楽曲に対して混合ガウス分布モデル（GMM）のパラメータを推定する。そして、楽曲Xと楽曲Yの間の類似度 $d_{CE}(X, Y)$ を、以下のように計算する。

【数13】

$$d_{CE}(X, Y) = \log \prod_i \frac{\mathcal{N}_{GMM}(x_i; \theta_X)}{\mathcal{N}_{GMM}(x_i; \theta_Y)} + \log \prod_j \frac{\mathcal{N}_{GMM}(y_j; \theta_Y)}{\mathcal{N}_{GMM}(y_j; \theta_X)}$$

【0052】

ここで、 x_i と y_j は、それぞれ楽曲X、楽曲Yの特徴ベクトル列を表現する。ただし、これは高信頼度フレーム選択により選択されている。 θ_X と θ_Y はそれぞれ楽曲X、楽曲Yの混合ガウス分布モデル（GMM）のパラメータで、 $\mathcal{N}_{GMM}(x; \theta)$ は、 θ をパラメータとする混合ガウス分布モデル（GMM）の確率密度関数を表す。

【0053】

類似楽曲検索表示手段5は、類似度計算手段3により計算された複数の声質類似度に基づいて、一つの楽曲中の歌声の声質に類似する歌声を含む複数の楽曲の複数の識別情報をディスプレイ10上に表示するための指令を制御手段7へと出力する。制御手段7はこの指令に基づいて、類似する歌声を含む複数の楽曲の複数の識別情報をディスプレイ10上に表示する。表示様式は、類似度の大小関係が分かるものであればどのような様式であってもよい。例えば、類似度の大小関係が判る表示様式としては、類似度の大きい順番に識別情報を順番に並べてディスプレイ上に表示する様式や、類似度の大きい順番に識別情報をディスプレイ上に一つずつ順番に表示する様式や、ディスプレイ上に表示する複数の識別情報の大きさ（寸法）を類似度の大きさに比例して大きくなるように定める様式等を用いることができる。

【0054】

また楽曲データ再生手段6は、ディスプレイ上に表示された複数の識別情報の表示から、楽曲選択手段8を用いて1以上の識別情報が選択されると、選択された識別情報の表示に対応する楽曲データを楽曲データ蓄積手段4に蓄積された複数の楽曲データから抽出して再生する。この楽曲の再生により、検索結果を実際に確認することが可能になる。複数の識別情報が選択される場合には、選択された順番に複数の楽曲データを再生してもよく、また声質類似度が大きい順番に複数の楽曲データを再生してもよい。

【0055】

なお楽曲データ再生手段6による楽曲データの再生は、楽曲データの先頭から再生してもよいが、いわゆる楽曲のサビ部分だけを再生するようにしてもよい。また楽曲中の複数の歌声部分から選択した適宜の部分から楽曲データを再生するようにしてもよい。

【0056】

図3は、図2に示した第1の楽曲検索システム1を、コンピュータを用いて実現する場

合に用いるプログラムのアルゴリズムを示すフローチャートである。コンピュータでプログラムを実行すると、準備段階として、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量と各楽曲の識別情報とを声質特徴量蓄積手段 2 に記憶し、また複数の識別情報と対応させて複数の楽曲データを楽曲データ蓄積手段 4 に蓄積する。そしてステップ S T 1 で複数の選択用楽曲の識別情報をディスプレイ上に表示する。そしてステップ S T 2 で複数の選択用楽曲の中から一つの楽曲が選択されるとステップ S T 3 へと進む。図 4 は、ディスプレイ 1 0 に表示される表示画面（検索結果を表示する画面）で実際に検索を行ったときの表示を示している。本実施の形態では、この表示画面を利用して好きな楽曲（クエリ）を選択する。その場合には、候補リスト表示部 L に表示された複数の楽曲の名称及び歌手の名前（識別情報）から好きな 1 曲の楽曲の識別情報をクリックすると入力窓部 W に選択された識別情報が表示される。そして再生ボタン S をクリックすると、ステップ S T 3 の内容即ち声質類似度の計算が開始される。なおこの表示画面における候補リスト表示部 L は、後述する検索結果を表示する表示部としても兼用される。候補リスト表示部 L に表示された楽曲に好みの楽曲がない場合にはスライドボタン S L をスライドさせて次の N 曲の楽曲を表示することができる。ステップ S T 3 では、選択された一つの楽曲の声質特徴量とその他の複数の楽曲のそれぞれの声質特徴量とに基づいて、該一つの楽曲とその他の複数の楽曲の声質類似度を計算する。そしてステップ S T 4 では、類似度計算手段 3 により計算された声質類似度に基づいて、一つの楽曲中の歌声の声質に類似する歌声を含む上位 N 曲の楽曲の識別情報をディスプレイ上に表示する。本実施の形態では、図 4 に示した表示画面の候補リスト表示部 L に表示せる複数の楽曲が、類似度の大きい順番に並び変えられて表示される。そしてステップ S T 5 で、ディスプレイ 1 0 上の候補リスト表示部 L に表示された複数の識別情報の表示から、一つの識別情報がクリックにより選択され、更に再生ボタン R P がクリックされて再生指令が入力されると、選択された識別情報の表示に対応する楽曲データを楽曲データ蓄積手段 4 に蓄積された複数の楽曲データから抽出して楽曲再生手段により再生する（ステップ S T 6）。候補リスト表示部 L に表示された複数の識別情報の中に、選択したい楽曲の識別情報が表示されていない場合には、ステップ S T 9 において、スライドボタン S L をスライドさせて次の N 曲の識別情報が表示される。また 1 曲の楽曲をステップ S T 6 において再生した後更に別の楽曲を再生する場合には、ステップ S T 8 へと進む。ステップ S T 8 では、候補リスト表示部 L の表示を初期の状態に戻すか否かの判定が行われる。初期状態に戻す場合には、ステップ S T 4 へと戻り、初期状態に戻さずに、現状の候補から再度選択する場合には、ステップ S T 5 へと戻る。

【 0 0 5 7 】

なおステップ S T 5 において、複数の楽曲を選択できるようにしてもよい。複数の楽曲を選択できるようにした場合には、ステップ S T 6 では選択された順番に複数の楽曲を再生するようにしてもよく、また類似度の大きい順番に複数の楽曲を再生するようにしてもよい。

【 0 0 5 8 】

図 4 の表示画面は、本実施の形態を用いて実際に実験を行った検索結果を表示しているので、この実験の条件と検索結果を説明する。まず歌声 G M M と非歌声 G M M の学習のために、「R W C 研究用音楽データベース：ポピュラー音楽」（R W C - M D B - P - 2 0 0 1）（後藤真孝，橋口博樹，西村拓一，岡隆一：R W C 研究用音楽データベース：研究目的で利用可能な著作権処理済み楽曲・楽器音データベース，情報処理学会論文誌，V o l . 4 5 ， N o . 3 ， p p . 7 2 8 - 7 3 8 （ 2 0 0 4 ））から、2 5 曲を選択して使用した。そしてシステムの検索対象データベースとしての楽曲データ蓄積手段 4 中には、R W C - M D B - P - 2 0 0 1 で歌声 / 非歌声 G M M の構築に使わなかった 7 5 曲を登録した。図 4 の運用例では、女性歌手「緒方智美」が歌う「P R O L O G U E」（R W C - M D B - P - 2 0 0 1 N o . 7）がクエリとして入力されている。クエリの楽曲が与えられ、ユーザが検索ボタン S を押すと、類似度の計算に 2 0 秒程度の時間がかかった後に、候補リスト表示部 L に検索結果として楽曲が上位から順に並べられて出力される。この例で

は、検索結果として順位、楽曲名、歌手名、類似度の順に表示される。ユーザが検索結果リスト中の楽曲をクリックして再生ボタン R P をクリックするか楽曲の表示をダブルクリックすると、楽曲データ再生手段 6 にその楽曲が読み込まれ、再生が始まる。その他の楽曲をクエリとして入力した場合でも、ほとんどの場合で上位の 10 曲に選ばれた楽曲はクエリの楽曲と歌声の声質が似ていると感じられた。例えば、図 4 の例では、上位の 21 曲はクエリと同様に女性歌手の曲であった。また、図 4 中の上位 15 曲の歌声の声質はクエリの楽曲と類似していた。クエリの楽曲の歌手である「緒方智美」の別の楽曲は、上位の 1 番目、2 番目にランクしていた一方で、比較的下位の 10 番目、12 番目にもランクしていた。これは、10 番目の曲と 12 番目の曲に関しては、歌唱スタイルがクエリの曲や 1 番目、2 番目の曲と異なっていたことが原因である。一方、検索結果の下位にランクされた楽曲については、順位は楽曲の類似度を正しく反映していないように思われた。例えば、図 4 の例では、画面上のプレイリストの範囲外で 22 位より下位では男性歌声と女声歌声が混在していた。しかし、通常の検索用途では上位の 10 曲程度を提示すれば十分であり、問題にはならない。キーとなる楽曲が、複数の歌手によって歌われる楽曲や、歌声に深くエフェクトがかけられている曲では、検索結果が直感的な類似度を反映していない場合もあった。

【0059】

次に、従来の内容に基づく音楽検索手法である MFCC を用いた検索システムと比較して、本実施の形態の歌声の声質に基づく楽曲検索システムを評価する。比較の対象とする従来の検索システムでは、特徴量として MFCC を使用し、特徴抽出の際に伴奏音抑制、高信頼度フレームを行わなかった。類似度の計算は本実施の形態の場合と同様にした。

【0060】

実験手順について説明する。まず本実験には、6 人の大学生（女性 4 人、男性 2 人）が被験者として参加した。被験者は、音楽の専門的な教育は受けていない。まず、被験者は図 5 のような画面が提示され 3 つの楽曲を聴くことが指示される。これらの 3 曲は、クエリの楽曲（図中の曲 X）、本実施の形態の検索システムにより最上位にランクされた楽曲（図中の曲 A または曲 B）、従来の検索システム（比較手法）により最上位にランクされた楽曲（図中の曲 B または曲 A）から構成されている。そして、被験者は曲 A と曲 B のどちらがよりクエリの楽曲に似ているかを判定する。被験者は曲 A と曲 B のどちらの楽曲が本実施の形態のシステムによる検索結果かは知らされておらず、曲 A と曲 B の順はランダムに決定されている。また、被験者は 3 つの楽曲を好きな順番で何度でも繰り返し聴くことが出来る。RWC 研究用音楽データベース：ポピュラー音楽（RWC-MDB-P-2001）中から、ジャンルや歌手の性別が多様になるように留意しながら図 6 に示すように 10 曲のクエリを選択した。図 6 に示す表において、3 桁の数字は RWC 研究用音楽データベース（RWC-MDB-P-2001）中の楽曲番号を表す。それぞれのクエリを入力した際の、本実施の形態の検索システムと従来の検索システム（MFCC）による最上位の検索結果が同じ行に記されている。また表中の「日」は日本語の歌詞であることを表しており、「英」は英語の歌詞であることを表している。

【0061】

実験では、それぞれのクエリの曲について、被験者に以下の 2 つの質問をした。

【0062】

質問 1：歌声の声質に関して、曲 A、曲 B のどちらがより曲 X に似ていると思いますか？

質問 2：曲全体の音質に関して、曲 A、曲 B のどちらがより曲 X に似ていると思いますか？

実験結果を図 7 と図 8 に示す。図 7 は、質問 1（ボーカールの声質）に対する回答結果を示しており、図 8 は質問 2（楽曲全体の音質）に対する回答結果を示している。即ち図 7 は、それぞれの楽曲について、被験者が本実施の形態の検索システムと従来の検索システムの検索結果のどちらの声質のほうがクエリの楽曲と類似しているかとの割合を表す。6 人の被験者の 10 曲に対する回答のうち 80% が、本実施の形態の検索システ

10

20

30

40

50

ムの検索結果の歌声の声質のほうによりクエリの楽曲と類似していると答えた。一方図 8 に示すように、曲全体の音質の類似度に関しては、70%の回答が従来の検索システムの検索結果がよりクエリに似ていると答えた。これらの実験結果から、本実施の検索システムによれば、伴奏音の影響が低減されて、歌声の声質の類似度に基づいて楽曲が検索できていることが確認された。実際、図 6 の表を見ると、従来の検索システムのように特徴量として MFCC を用いた場合は、10 曲中 4 曲でクエリと異なる性別の楽曲を出力しているが、本実施の形態の検索システムではそのような問題は発生していない。

【0063】

また、本実施の形態の検索システムの検索結果は、単に歌声の声質の類似度だけでなく、歌唱スタイルの類似度の観点からも類似している例が見られた。例えば、RWC-MDB-P-2001の#053の楽曲がクエリとして使用された場合、本実施の形態の検索システムの検索結果と従来の検索システムの検索結果が共にクエリと同じ歌手の楽曲である。ところが、6人の被験者中の5人が、本実施の形態の検索システムの検索結果のほうにクエリと類似していると判断している。実験によって、本実施の形態の検索システムを、75曲の楽曲を用いて実際に運用し、声質の類似度に基づいて楽曲を検索するのに有効であることが確認できた。また本実施の形態の検索システムの検索結果が実際に声質の類似度に基づいていることを確かめるため、6人の被験者により被験者実験を行い、本実施の形態の検索システムの有効性を確認できた。なお本実験からは、2つの楽曲の類似尺度として、相互情報量を用いると、効果的であることが確認できた。一方で、相互情報量は計算の際に全ての特徴量を使用するため、計算時間が長くなったり、また必要な記憶容量が大きくなるという問題がある。しかしながら使用するコンピュータの演算速度が速くなり、また使用するメモリの価格が下がれば、この問題は解消するので、十分に実用可能である。なお類似尺度としては、Earth Mover's Distance (EMD) (Berenzweig, A., Logan, B., Ellis, D. P. W. and Whitman, B.: A Large-Scale Evaluation of Acoustic and Subjective Music Similarity Measures, Computer Music Journal, Vol. 28, No. 2, pp. 63-76 (2004)) など、その他の類似尺度を導入すれば、計算時間を削減することができる。さらに、本実施の形態の楽曲検索システムをその他の音楽検索手法と統合することで、より柔軟で幅広い楽曲検索システムを開発することも可能である。

【0064】

図 9 は、本発明の楽曲検索システムの第 2 の実施の形態の構成を示すブロック図である。本実施の形態は、第 1 の実施の形態の特徴である声質が類似する歌声を含む楽曲の検索に加えて、楽曲の曲調も類似する楽曲を選択できる。図 9 においては、図 2 に示した第 1 の実施の形態の構成と同様のブロックに、図 2 に示したブロックに付した符号の数に 100 の数を加えた数の符号を付してある。そして機能が同じブロック（手段）に関しては、説明を省略する。この楽曲探索システム 101 では、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量と各楽曲の識別情報とを蓄積する声質特徴量蓄積手段 102 に加えて、各楽曲の曲調特徴量と識別情報とを蓄積する曲調特徴量蓄積手段 111 を備えている。なお声質特徴量蓄積手段 102 と曲調特徴量蓄積手段 111 とは一つの蓄積手段によって構成することができるのは勿論である。本実施の形態では、声質特徴量蓄積手段 102 と曲調特徴量蓄積手段 111 とによって声質及び曲調特徴量蓄積手段が構成されている。また本実施の形態では、曲調類似度計算手段 112 を更に備えている。曲調特徴量としては、例えば、MFCC (Mel-Frequency Cepstrum Coefficient)、スペクトル重心 (Audio Spectrum Centroid)、ロールオフ (Roll Off)、フラックスなどの特徴量を用いることができる。

【0065】

曲調類似度計算手段 112 は、楽曲選択手段 108 によって複数の楽曲の中から一つの楽曲が選択されると、一つの楽曲の曲調特徴量とその他の複数の楽曲のそれぞれの曲調特

10

20

30

40

50

微量とに基づいて、一つの楽曲とその他の複数の楽曲の曲調類似度を計算する。すなわち選択された楽曲の曲調特徴量と、その他の複数の楽曲の曲調特徴量（曲調特徴量蓄積手段 1 1 1 に蓄積された残りの楽曲の曲調特徴量）のすべてとの間において、類似度を計算する。類似度の計算方法も、任意であって、特定の類似度計算方法に限定されるものではない。曲調の類似度の計算にも、例えば、2つの楽曲間の類似尺度として、相互情報量の他、COS 尺度、距離尺度 EMD (Earth Mover's Distance)等を使用することができる。

【 0 0 6 6 】

類似楽曲検索表示手段 1 0 5 は、声質類似度計算手段 1 0 3 により計算された声質類似度と、曲調類似度計算手段 1 1 1 により計算された曲調類似度とに基づいて、声質が類似する歌声を含み且つ曲調が類似する複数の楽曲の複数の識別情報をディスプレイ 1 1 0 上に表示する指令を制御手段 1 0 7 に出力する。制御手段 1 0 7 は、この指令を受けてディスプレイ 1 1 0 上に複数の楽曲の複数の識別情報を表示する指令をディスプレイドライバ 1 0 9 に出力する。声質が類似する歌声を含み且つ曲調が類似する複数の楽曲を検索する場合には、1つの楽曲の声質類似度及び曲調類似度の平均値を基準としてもよく、声質類似度及び曲調類似度のそれぞれに所定の重み付けをした値の平均値を基準として検索を行って、平均値の大きい順に検索された複数の楽曲の識別情報をディスプレイ 1 1 0 上に図 4 の表示と同様の様式で並べるようにしてもよい。また図 1 1 に示すような、縦軸 Y A 及び横軸 X A の一方（図 1 1 では横軸）が声質類似度を示し、他方（図 1 1 の例で縦軸）が曲調類似度を示す表示様式を用いることができる。図 1 1 は、図 1 1 の右下に示した実際の表示様式の一部の領域（四角で囲んだ領域）を拡大して示したものである。この表示様式では、一つの楽曲の識別情報を、この一つの楽曲の声質類似度の位置で横軸 X A と直交する第 1 の仮想線 P L 1 と、曲調類似度の位置で縦軸 Y A と直交する第 2 の仮想線 P L 2 の交点またはその近傍に表示する。図 1 1 の例では、丸の図形の内部に識別情報として歌手名と楽曲名とが表示されている。図 1 1 のような表示様式を採用すると、各楽曲の識別情報を平面上に分散して配置したような状態で表示することができる。図 1 1 に示す様式で検索結果を表示する場合には、声質類似度及び曲調類似度が共に同じ値で、この値がそれぞれ異なる複数の楽曲の識別情報は、縦軸 Y A と横軸 X A の交点から縦軸 Y A と横軸 X A の間の 4 5 度の角度方向に延びる仮想中央線 P C 上に並んで配置されることになる。この場合、声質類似度及び曲調類似度が共に最も大きい楽曲の識別情報が縦軸 Y L と横軸 X A の交点に位置するようにすると、交点近傍の識別情報を選択すれば、声質及び曲調が共に類似した楽曲を素早く見つけることが可能になる。

【 0 0 6 7 】

図 1 1 の表示様式（インターフェース）では、クエリとなる楽曲と検索対象の他の複数の楽曲の、歌手の声質の類似度と曲調の類似度を平面上に可視化することで、ユーザはクエリの楽曲と歌手の声質及び曲調が似ている曲を直感的な操作で探すことが可能である。また図 1 1 において、図中の円は、それぞれ一つの楽曲を表現している。左上の領域中の楽曲（この図では「西一男」の「永遠のレプリカ」）は、ユーザが提示したクエリとなる楽曲である。図中央部の大きな領域には、検索対象の楽曲を横軸方向は歌声の類似度の順に、縦軸方向は曲調の類似度の順にプロットした平面（以後類似度平面と呼ぶ）の一部が表示してある。類似度平面中では、左部に表示されている楽曲ほどクエリの楽曲に歌手の声質が類似し、上部に表示されている楽曲ほどクエリ楽曲と曲調が類似している。

【 0 0 6 8 】

また横軸 X A に沿って配置されたスライダ S L 1 及び縦軸 Y A に沿って配置されたスライダ S L 2 は類似度平面中のどの領域を現在表示しているかを表しており、スライダ S L 1 及び S L 2 を動かすことで任意の領域を表示することが可能である。また、領域中の任意の点をドラッグすることによっても、類似度平面中の表示領域を変更することが可能である。図右下部の、多くの白い点がある小さな領域は、類似度平面全体の中の検索対象の楽曲の分布を表しており、一つ一つの白い点が各楽曲に対応する。その中の薄い小さな四角の領域が現在図中央部に表示している範囲を表している。ユーザはこの領域を直接クリックすることでも、類似度平面中の表示領域を変更することが可能である。

10

20

30

40

50

【 0 0 6 9 】

また、図中央部の領域には類似度平面の一部が表示されているが、その際の縮尺は自由に変更することが可能である。つまり、倍率を小さくすることで、一画面に多くの楽曲を表示することも可能である。また、倍率を小さくすることで、一画面に表示される楽曲の数を少なくし、検索結果を見やすくすることも可能である。倍率の変更は縦軸と横軸で独立して行うことが可能である。

【 0 0 7 0 】

楽曲を表現する円をユーザがクリックするとその楽曲の再生を開始する。再生中の楽曲を表す円は他の円と比べて区別できるように描画され、再生中の位置が円の周囲に表示される帯状表示等でわかるように表示される。図 1 1 の例では、「風戸ヒサヨシ」の「言えない」が再生中で、現在約 4 0 % 程度再生済みであることを表している。なお、楽曲を再生する際には、ユーザが楽曲を探しやすいように、サビ区間のみを再生することも可能である。この場合、サビ区間自動検出手法によりサビ区間をあらかじめ検出しておけばよい。

【 0 0 7 1 】

図 1 0 は、図 9 に示した第 2 の楽曲検索システム 1 0 1 を、コンピュータを用いて実現する場合に用いるプログラムのアルゴリズムを示すフローチャートである。コンピュータでプログラムを実行すると、準備段階として、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量と各楽曲の識別情報とを声質特徴量蓄積手段 1 0 2 に記憶する。またそれぞれの楽曲の曲調特徴量と各楽曲の識別情報とを曲調特徴量蓄積手段 1 1 1 に蓄積する。さらに複数の識別情報と対応させて複数の楽曲データを楽曲データ蓄積手段 1 0 4 に蓄積する。そしてステップ S T 1 0 1 で複数の選択用楽曲の識別情報をディスプレイ上に表示する。そしてステップ S T 1 0 2 で複数の選択用楽曲の中から一つの楽曲が選択されるとステップ S T 1 0 3 へと進む。ステップ S T 1 0 3 では、選択された一つの楽曲の声質特徴量とその他の複数の楽曲のそれぞれの声質特徴量とに基づいて、該一つの楽曲とその他の複数の楽曲の声質類似度を計算する。そして次にステップ S T 1 0 4 へと進み、ステップ S T 1 0 4 では、一つの楽曲とその他の複数の楽曲の曲調類似度を計算する。そしてステップ S T 1 0 5 では、類似度計算手段 1 0 3 により計算された声質類似度と曲調類似度計算手段 1 1 2 により計算された曲調類似度とに基づいて、一つの楽曲中の歌声の声質に類似する歌声を含み且つ曲調が類似する上位 N 曲の楽曲の識別情報をディスプレイ上に表示する。そしてステップ S T 1 0 6 で、ディスプレイ 1 1 0 上の候補リスト表示部 L に表示された複数の識別情報の表示から、一つの識別情報がクリックにより選択され、選択された識別情報の表示に対応する楽曲データを楽曲データ蓄積手段 1 0 4 に蓄積された複数の楽曲データから抽出して楽曲再生手段により再生する（ステップ S T 1 0 7）。候補リスト表示部 L に表示された複数の識別情報の中には、選択したい楽曲の識別情報が表示されていない場合には、ステップ S T 1 1 0 において、次の N 曲の識別情報が表示される。また 1 曲の楽曲をステップ S T 1 0 7 において再生した後更に別の楽曲を再生する場合には、ステップ S T 1 0 9 へと進む。ステップ S T 1 0 9 では、候補リスト表示部 L の表示を初期の状態に戻すか否かの判定が行われる。初期状態に戻す場合には、ステップ S T 1 0 5 へと戻り、初期状態に戻さずに、現状の候補から再度選択する場合には、ステップ S T 1 0 6 へと戻る。なおステップ S T 1 0 6 において、複数の楽曲を選択できるようにしてもよい。この第 2 の実施の形態によれば、ユーザが好む楽曲中の歌声の声質と声質が似ており、しかもその楽曲と曲調が似た未知の楽曲に巡り合うことが可能になる。

【 0 0 7 2 】

なお声質の類似を検索する際に基準とする楽曲と曲調の類似を検索する際に基準とする楽曲を異ならせてもよい。すなわち声質はある楽曲中の歌声（例えば一人の女性がロックを歌っている楽曲の歌声）の声質と類似し、曲調は他の楽曲（例えばコーラスグループがポップスを歌っている楽曲）の曲調と類似する楽曲を検索できるようにしてもよい。図 1 2 は、このような検索を実現するための本発明の楽曲検索システムの第 3 の実施の形態の構成を示すブロック図である。また図 1 3 は、図 1 2 の実施の形態をコンピュータを用い

10

20

30

40

50

て実現する場合に用いるプログラムのフローチャートを示す図である。本実施の形態の楽曲検索システムでは、ある一つの楽曲中の歌声の声質に類似する歌声を含み、しかも他の一つの楽曲の曲調と類似する曲調の楽曲を検索することができる。図12においては、図9に示した第2の実施の形態の構成と同様のブロックに、図9に示したブロックに付した符号の数に100の数を加えた数の符号を付してある。そして機能が同じブロック(手段)に関しては、説明を省略する。この楽曲探索システム101では、楽曲選択手段208が声質のための楽曲と曲調のための楽曲の選択に利用される。

【0073】

第3の実施の形態でも、声質及び曲調特徴量蓄積手段(202及び211)と、声質類似度計算手段203と、曲調類似度計算手段212と、類似楽曲検索表示手段205と、楽曲データ再生手段206とを用いる。声質類似度計算手段202は、複数の楽曲の中から楽曲選択手段208により一つの楽曲M1が選択されると、該一つの楽曲M1の声質特徴量とその他の複数の楽曲のそれぞれの声質特徴量とに基づいて、一つの楽曲とその他の複数の楽曲の声質類似度を計算する。また曲調類似度計算手段212は、複数の楽曲の中から他の一つの楽曲M2が選択されると、他の一つの楽曲M2の曲調特徴量とその他の複数の楽曲のそれぞれの曲調特徴量とに基づいて、他の一つの楽曲M2とその他の複数の楽曲の曲調類似度を計算する。そして類似楽曲検索表示手段205は、声質類似度計算手段202により計算された声質類似度と、曲調類似度計算手段212により計算された曲調類似度とに基づいて、一つの楽曲M1中の歌声と声質が類似する歌声を含み且つ他の一つの楽曲M2中の楽曲と曲調が類似する複数の楽曲についての複数の識別情報をディスプレイ210上に表示する指令を制御手段207に出力する。ディスプレイ210上への識別情報の表示様式は任意である。例えば、図14に示すように、図11の様式と同様に、縦軸YA及び横軸XAの一方(この例では横軸)が声質類似度を示し、他方(この例では縦軸)が曲調類似度を示し、一つの楽曲の識別情報を、該一つの楽曲の声質類似度の位置で横軸XAと直交する第1の仮想線と該一つの楽曲の曲調類似度の位置で縦軸YAと直交する第2の仮想線の交差点またはその近傍に表示する表示様式を用いることができる。この表示様式を用いると、一つの楽曲M1と声質が類似し、他の楽曲M2と曲調が類似する複数の楽曲を、類似の程度と共に一目で認識することができる。

【0074】

図14の表示様式(インタフェース)は、図11と比べると、クエリの楽曲M1及びM2を表示する図中の左上の領域が2つになっている。そして歌声の類似度を計算する際のクエリ楽曲M1と曲調の類似度を計算する際のクエリ楽曲M2を異なる楽曲にすることができる。これにより、ある楽曲の歌声は好きだが曲調はあまり好きでないという場合でも、検索対象の楽曲の中から歌声と曲調が両方好きな楽曲を探してることができる。この場合も操作方法は図11の場合と同様であり、検索結果を類似度平面にプロットすることにより、好みの楽曲を直観的な操作で探し出すことが可能である。

【0075】

コンピュータで図13に示すアルゴリズムのプログラムを実行すると、準備段階として、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量と各楽曲の識別情報とを声質特徴量蓄積手段202に記憶する。またそれぞれの楽曲の曲調特徴量と各楽曲の識別情報とを曲調特徴量蓄積手段211に蓄積する。さらに複数の識別情報と対応させて複数の楽曲データを楽曲データ蓄積手段204に蓄積する。そしてステップST201で複数の選択用楽曲の識別情報をディスプレイ上に表示する。ステップST202で複数の選択用楽曲の中から声質選択用の一つの楽曲(クエリ)が選択されるとステップST203へと進む。ステップST203では、選択された一つの楽曲の声質特徴量とその他の複数の楽曲のそれぞれの声質特徴量とに基づいて、該一つの楽曲とその他の複数の楽曲の声質類似度を計算する。次にステップST204で、複数の選択用楽曲の中から曲調選択用の他の一つの楽曲(クエリ)が選択される。そしてステップST205で、選択された他の一つの楽曲の曲調特徴量とその他の複数の楽曲のそれぞれの曲調特徴量とに基づいて、該他の一つの楽曲とその他の複数の楽曲の曲調類似度を計算する。次にステップST206では、

10

20

30

40

50

類似度計算手段 203 により計算された声質類似度と曲調類似度計算手段 212 により計算された曲調類似度とに基づいて、一つの楽曲中の歌声の声質に類似する歌声を含み且つ他の一つの楽曲と曲調が類似する上位 N 曲の楽曲の識別情報をディスプレイ上に表示する。そしてステップ ST 209 で、ディスプレイ 210 上の候補リスト表示部 L に表示された複数の識別情報の表示から、一つの識別情報がクリックにより選択され、選択された識別情報の表示に対応する楽曲データを楽曲データ蓄積手段 204 に蓄積された複数の楽曲データから抽出して楽曲再生手段により再生する（ステップ ST 210）。候補リスト表示部 L に表示された複数の識別情報の中には、選択したい楽曲の識別情報が表示されていない場合には、ステップ ST 213 において、次の N 曲の識別情報が表示される。また 1 曲の楽曲をステップ ST 210 において再生した後更に別の楽曲を再生する場合には、ステップ ST 211 からステップ ST 212 へと進む。ステップ ST 212 では、候補リスト表示部 L の表示を初期の状態に戻すか否かの判定が行われる。初期状態に戻す場合には、ステップ ST 206 へと戻り、初期状態に戻さずに、現状の候補から再度選択する場合には、ステップ ST 209 へと戻る。なおステップ ST 209 において、複数の楽曲を選択できるようにしてもよい。またこの例では、ステップ ST 204 において曲調選択用の楽曲が選択されない場合には、ステップ ST 207 へと進んで、曲調の選択を中止するかどうか判定される。ステップ ST 207 で曲調の選択が中止されると、ステップ ST 208 へと進み、先に選択した声質だけが類似する歌声を含む楽曲の検索だけが実行される。すなわちステップ ST 208 では、声質類似度が上記 N 曲の識別情報がディスプレイ 210 上に表示される。その後は、前述のステップ ST 209 ~ ステップ ST 213 と同様である。

10

20

【0076】

この第 2 の実施の形態によれば、ユーザが好む楽曲中の歌声の声質と声質が似ており、しかもその楽曲と曲調が似た未知の楽曲に巡り合うことが可能になる。

【0077】

上記各実施の形態では、毎回類似度を計算しているが、予め類似度を計算しておき類似度蓄積手段に蓄積しておいてもよい。図 15 は、予め類似度を計算しておき類似度蓄積手段に蓄積しておく、第 4 の実施の形態の構成を示すブロック図である。図 15 においては、図 2 に示した第 1 の実施の形態の構成と同様のブロックに、図 2 に示したブロックに付した符号の数に 300 の数を加えた数の符号を付してある。声質が類似する楽曲を検索する場合には、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量とその他の複数の楽曲の声質特徴量とに基づいて算出された各楽曲とその他の複数の楽曲との間の複数の声質類似度と、各楽曲の識別情報とを類似度蓄積手段 313 に蓄積する。類似楽曲検索表示手段 305 は、複数の楽曲の中から一つの楽曲が選択されると、類似度蓄積手段 313 から、該一つの楽曲とその他の複数の楽曲との間の複数の声質類似度を読み出し、複数の声質類似度に基づいて一つの楽曲中の歌声の声質と類似する歌声を含む複数の楽曲の複数の識別情報を制御手段 307 及びデジスプレイドライバ 309 を介してディスプレイ 310 上に表示する。このようにすると事前に類似度の計算を行っておくため、検索時間を早くすることができる。

30

【0078】

図 16 は、図 15 に示した第 4 の楽曲検索システム 301 を、コンピュータを用いて実現する場合に用いるプログラムのアルゴリズムを示すフローチャートである。コンピュータでプログラムを実行すると、準備段階として、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量とその他の複数の楽曲の前記声質特徴量とに基づいて算出された各楽曲とその他の複数の楽曲との間の複数の声質類似度と、各楽曲の識別情報とが類似度蓄積手段 313 に蓄積される。そしてステップ ST 301 で複数の選択用楽曲の識別情報をディスプレイ上に表示する。そしてステップ ST 302 で複数の選択用楽曲の中から一つの楽曲が選択されるとステップ ST 303 へと進む。ステップ ST 303 では、類似度蓄積手段 313 に蓄積された声質類似度に基づいて、一つの楽曲中の歌声の声質に類似する歌声を含む上位 N 曲の楽曲の識別情報をディスプレイ上に表示する。以下のステップ ST 30

40

50

4乃至ステップST308は、図3のステップST5からステップST9と同様になる。

【0079】

また声質と曲調が類似する楽曲を検索する場合、類似度蓄積手段313には、複数の楽曲について、それぞれの楽曲の歌声の声質特徴量とその他の複数の楽曲の声質特徴量とに基づいて算出された各楽曲とその他の複数の楽曲との間の複数の声質類似度と、複数の楽曲について、それぞれの楽曲の曲調特徴量とその他の複数の楽曲の曲調特徴量とに基づいて算出された各楽曲とその他の複数の楽曲との間の複数の曲調類似度と、各楽曲の識別情報とを蓄積しおけばよい。そして類似楽曲検索表示手段305は、複数の楽曲の中から一つの楽曲が選択されると、類似度蓄積手段313から、該一つの楽曲とその他の複数の楽曲との間の複数の声質類似度及び複数の曲調類似度を読み出し、複数の声質類似度及び複数の曲調類似度に基づいて一つの楽曲中の歌声の声質と類似する歌声を含み及び該一つの楽曲の曲調と類似する複数の楽曲の複数の識別情報をディスプレイ310上に表示する。

10

【0080】

図17は、声質特徴量の類似度と曲調特徴量の類似度の両方を類似度蓄積手段313に蓄積する場合の楽曲検索システム301を、コンピュータを用いて実現する場合に用いるプログラムのアルゴリズムを示すフローチャートである。図16に示したフローチャートと比較すると、ステップST403において、曲調類似度を含めている点で相違するだけで、ステップST403を除くステップST401～ステップST408の内容は、図16のステップST303を除くステップST301～ステップST308と実質的に同じである。

20

なお上記第3の実施の形態と同様に、声質と曲調のクエリ楽曲をそれぞれ異なる楽曲とする場合にも、図15に示した実施の形態と同様に、事前にすべての類似度を計算して類似度蓄積手段313に蓄積しておき、図15の実施の形態と同様に、検索を行ってもよいのは勿論である。

【図面の簡単な説明】

【0081】

【図1】本発明の楽曲検索システムの一実施の形態の基本的な技術的思想を概念的に示す図である。

【図2】図1の概念を具体化した第1の楽曲検索システムの具体的構成を示すブロック図である。

30

【図3】図2に示した第1の楽曲検索システムをコンピュータを用いて実現する場合に用いるプログラムのアルゴリズムを示すフローチャートである。

【図4】ディスプレイに表示される表示画面（検索結果を表示する画面）で実際に検索を行ったときの表示を示す図である。

【図5】被験者に提示する問題を示す画面の図である。

【図6】10曲のクエリを選択した結果を示す表である。

【図7】質問1（ボーカルの声質）に対する回答結果を示す図である。

【図8】質問2（楽曲全体の音質）に対する回答結果を示す図である。

【図9】本発明の楽曲検索システムの第2の実施の形態の構成を示すブロック図である。

【図10】図9に示した第2の楽曲検索システムを、コンピュータを用いて実現する場合に用いるプログラムのアルゴリズムを示すフローチャートである。

40

【図11】検索結果の表示様式の一例を示す図である。

【図12】本発明の楽曲検索システムの第3の実施の形態の構成を示すブロック図である。

【図13】図12の実施の形態をコンピュータを用いて実現する場合に用いるプログラムのフローチャートを示す図である。

【図14】図12の実施の形態による検索結果の表示様式を示す図である。

【図15】予め類似度を計算しておき類似度蓄積手段に蓄積しておく、第4の実施の形態の構成を示すブロック図である。

【図16】図15に示した第4の楽曲検索システムを、コンピュータを用いて実現する場

50

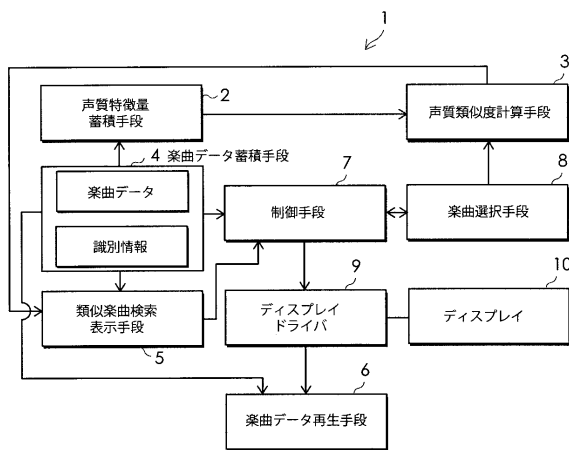
合に用いるプログラムのアルゴリズムを示すフローチャートである。

【図17】図15に示した第4の楽曲検索システムを、コンピュータを用いて実現する場合に用いる別のプログラムのアルゴリズムを示すフローチャートである。

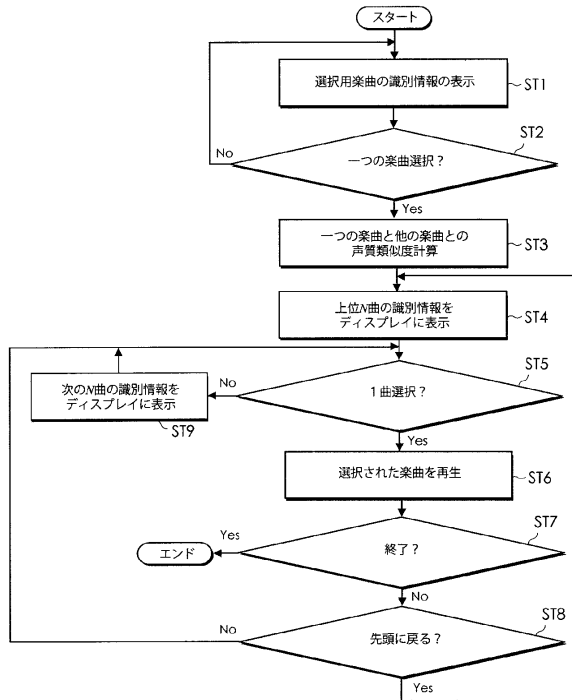
【0082】

- 1 楽曲検索システム
- 2 声質特徴量蓄積手段
- 3 声質類似度計算手段
- 4 楽曲データ蓄積手段
- 5 類似楽曲検索表示手段
- 6 楽曲データ再生手段
- 7 制御手段
- 8 楽曲選択手段
- 9 ディスプレイドライバ
- 10 ディスプレイ

【図2】



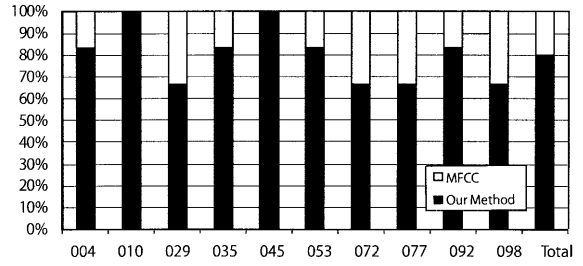
【図3】



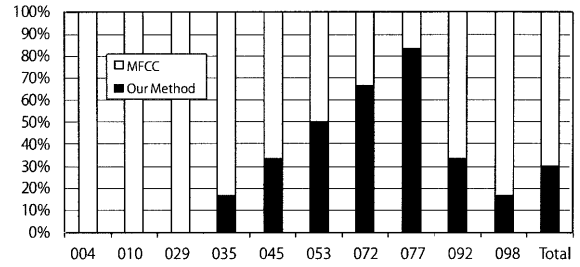
【図6】

クエリーの楽曲				検索結果 (最上位の楽曲)						
#	歌手名	性別	言語	曲調	手法	#	歌手名	性別	言語	曲調
004	風戸ヒサヨシ	男	日	ブラックソウル	Mfcc	031	永山雄一	男	日	ハウス
010	ブレイキーズ	女	日	ダンスミュージック	Mfcc	082	井口雄也	男	日	ハードロック
029	西一男	男	日	ブラックソウル	Mfcc	016	吉井弘美	女	日	ソウル
035	吉井弘美	女	日	メタルロック	Mfcc	054	漣	女	日	J-POP
045	森元誠介	男	日	ストリートフォーク	Mfcc	017	吉井弘美	女	日	J-POP
053	漣	女	日	J-POP	Mfcc	012	西一男	男	日	ブリテンジャズロック
072	森元誠介	男	日	ニューミュージック	Mfcc	036	西一男	男	日	メタルロック
077	服部まきこ	女	日	J-POP	Mfcc	094	Betty	女	英	ポップス
092	Betty	女	英	ダンスミュージック	Mfcc	090	井口雄也	男	日	ロック
098	JeffManning	男	英	リハブソウルサウンド	Mfcc	042	森元誠介	男	日	J-POP
					Mfcc	062	漣	女	日	J-POP
					Mfcc	014	漣	女	日	J-POP
					Mfcc	071	吉井弘美	女	日	ニューミュージック
					Mfcc	076	森元誠介	男	日	ストリートフォーク
					Mfcc	071	吉井弘美	女	日	ニューミュージック
					Mfcc	067	服部まきこ	女	日	J-POP
					Mfcc	024	風戸ヒサヨシ	男	日	ダンスミュージック
					Mfcc	086	Betty	女	英	ダンスミュージック
					Mfcc	009	西一男	男	日	ブリテンジャズロック
					Mfcc	085	JeffManning	男	英	モータウン

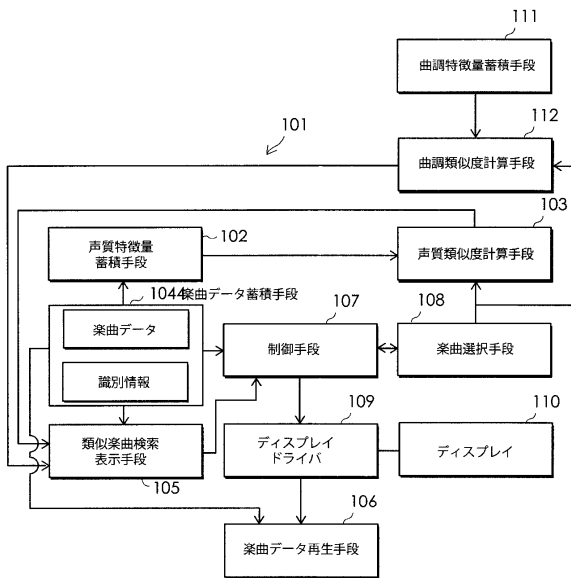
【図7】



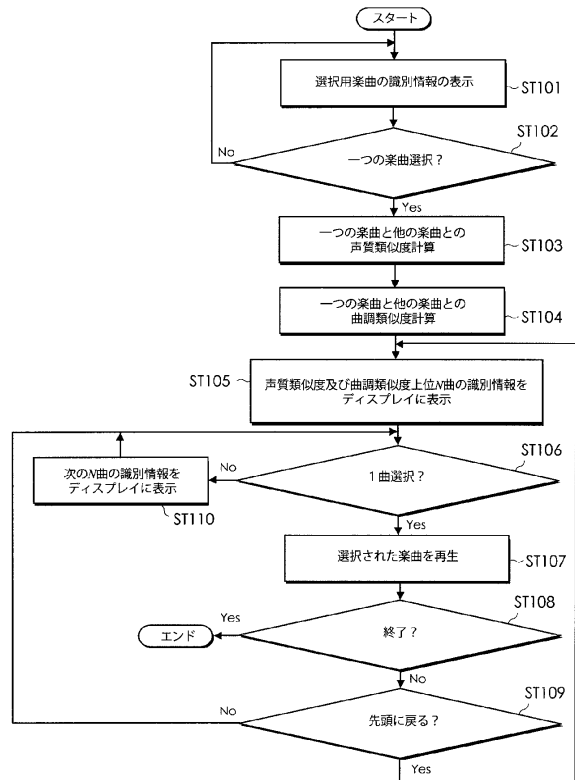
【図8】



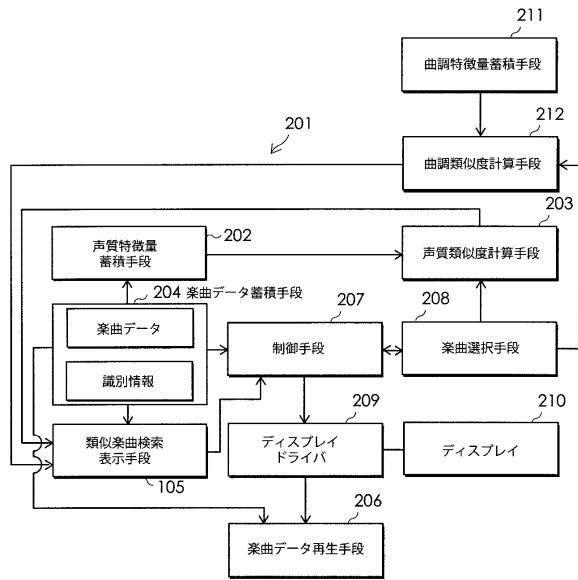
【図9】



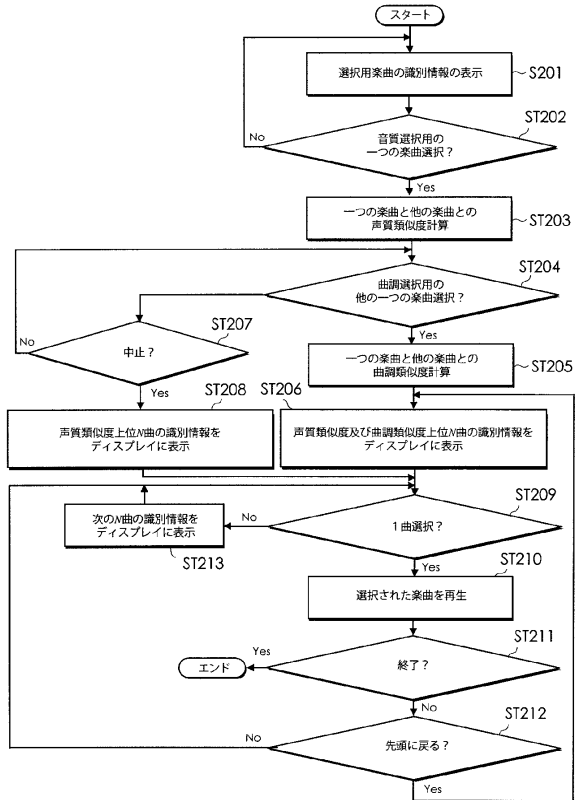
【図10】



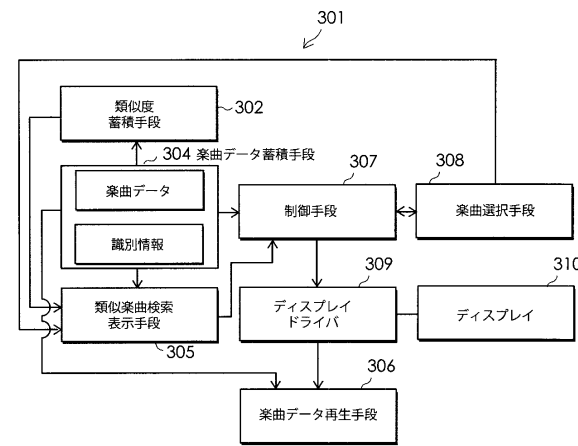
【図12】



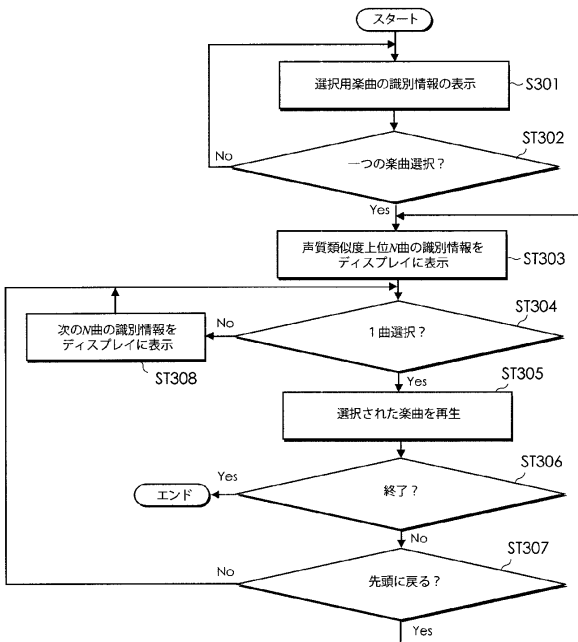
【図13】



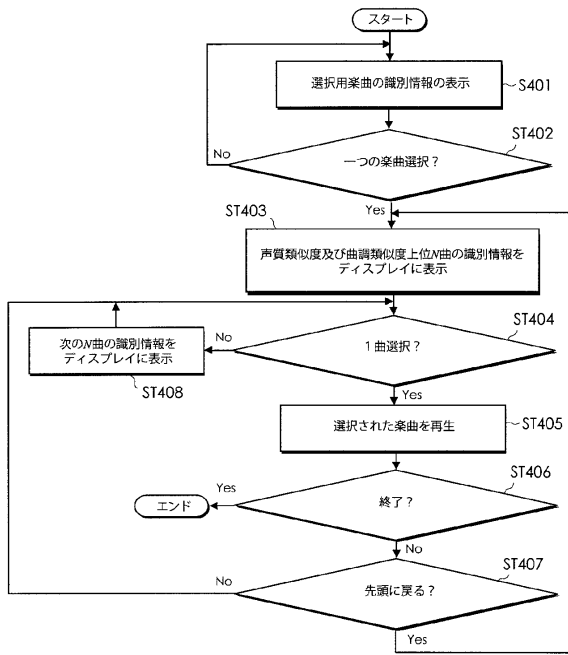
【図15】



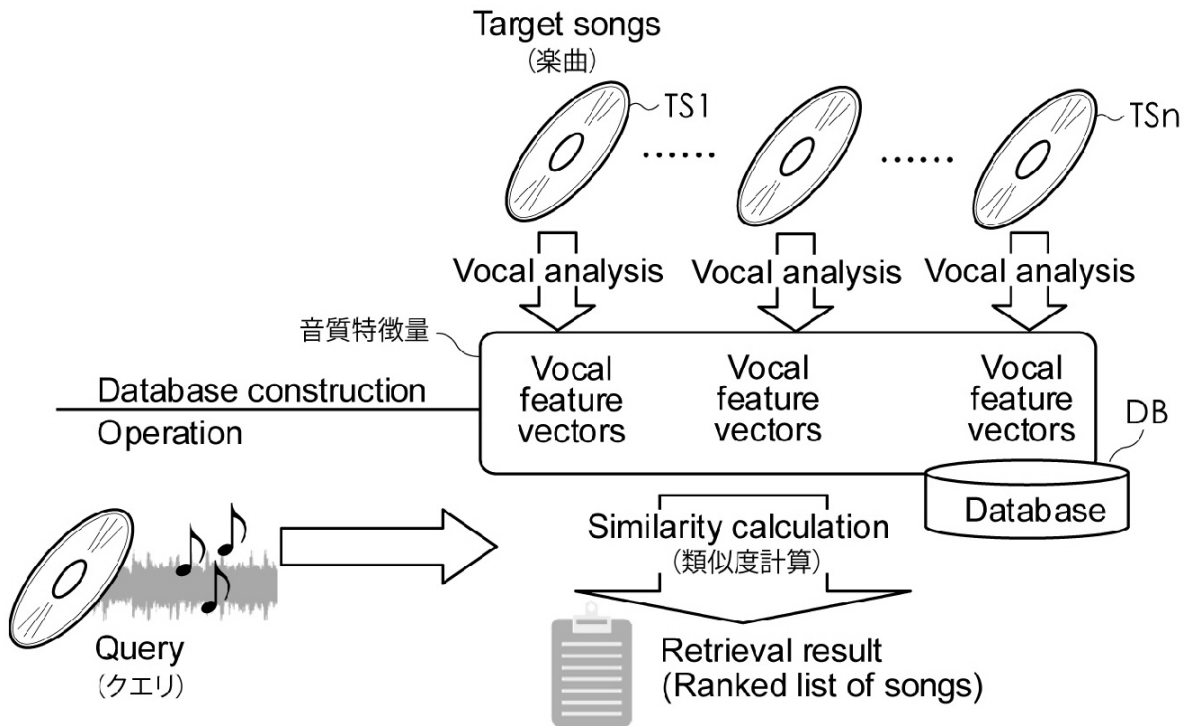
【図16】



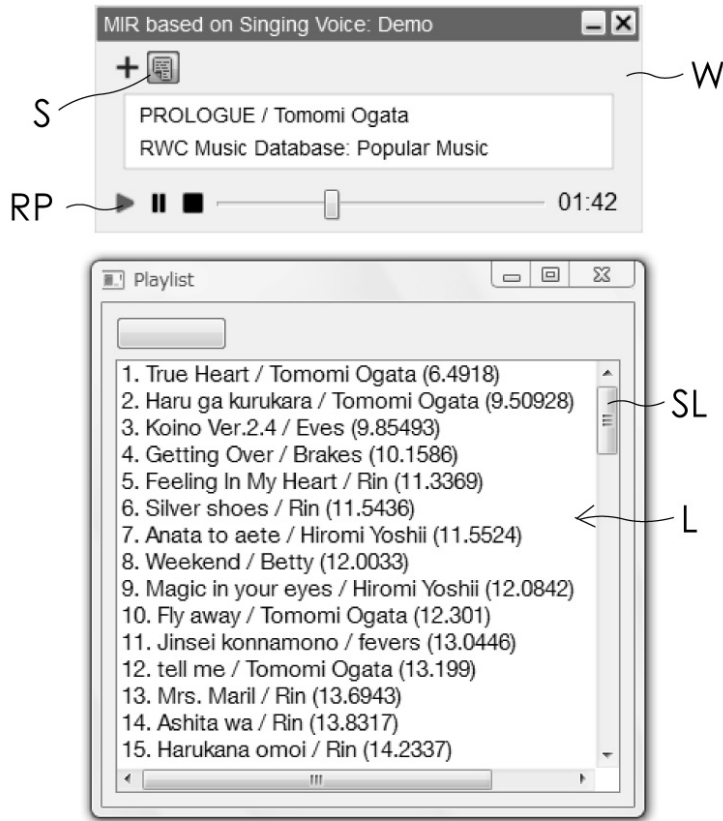
【 図 17 】



【 図 1 】



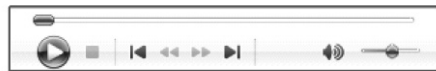
【 図 4 】



【 図 5 】

問題1/10

以下の3曲を聴いて、以下の質問にAまたはBで答えて下さい。



曲X



曲 A

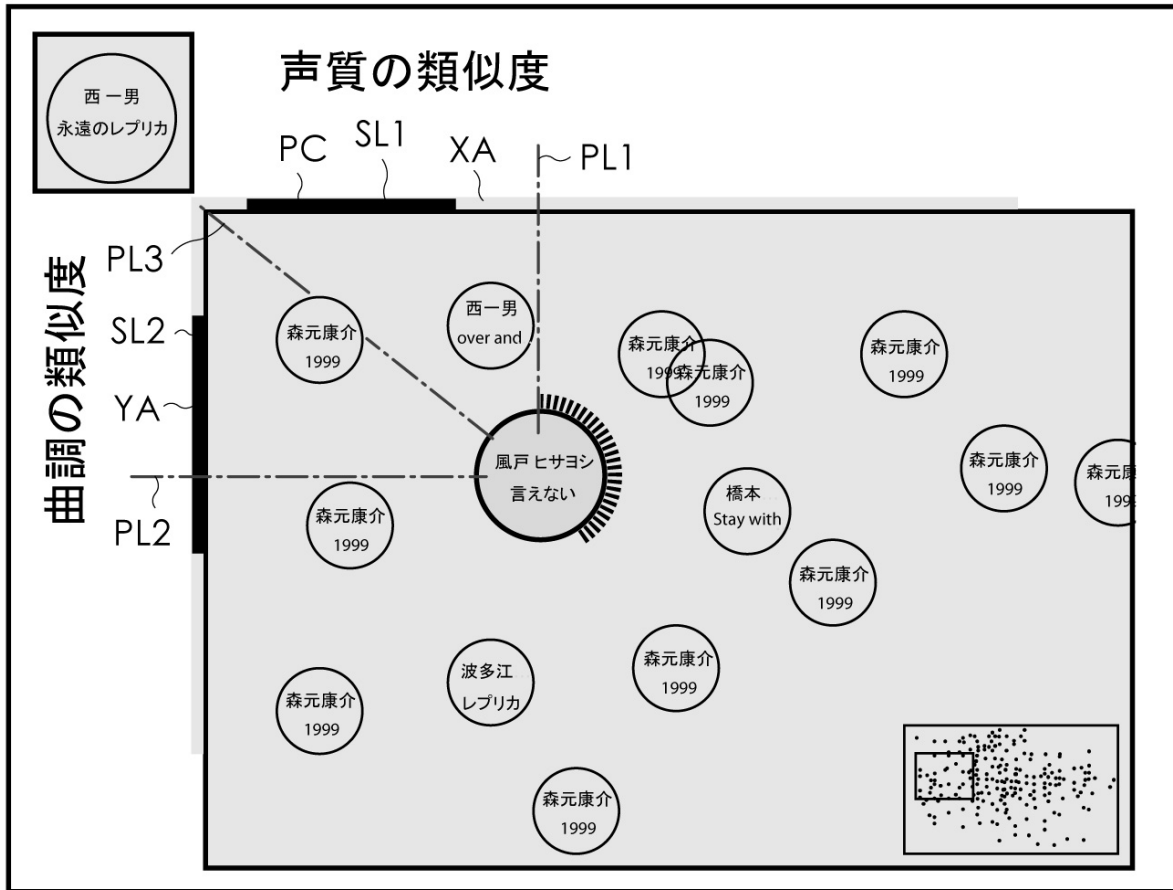


曲 B

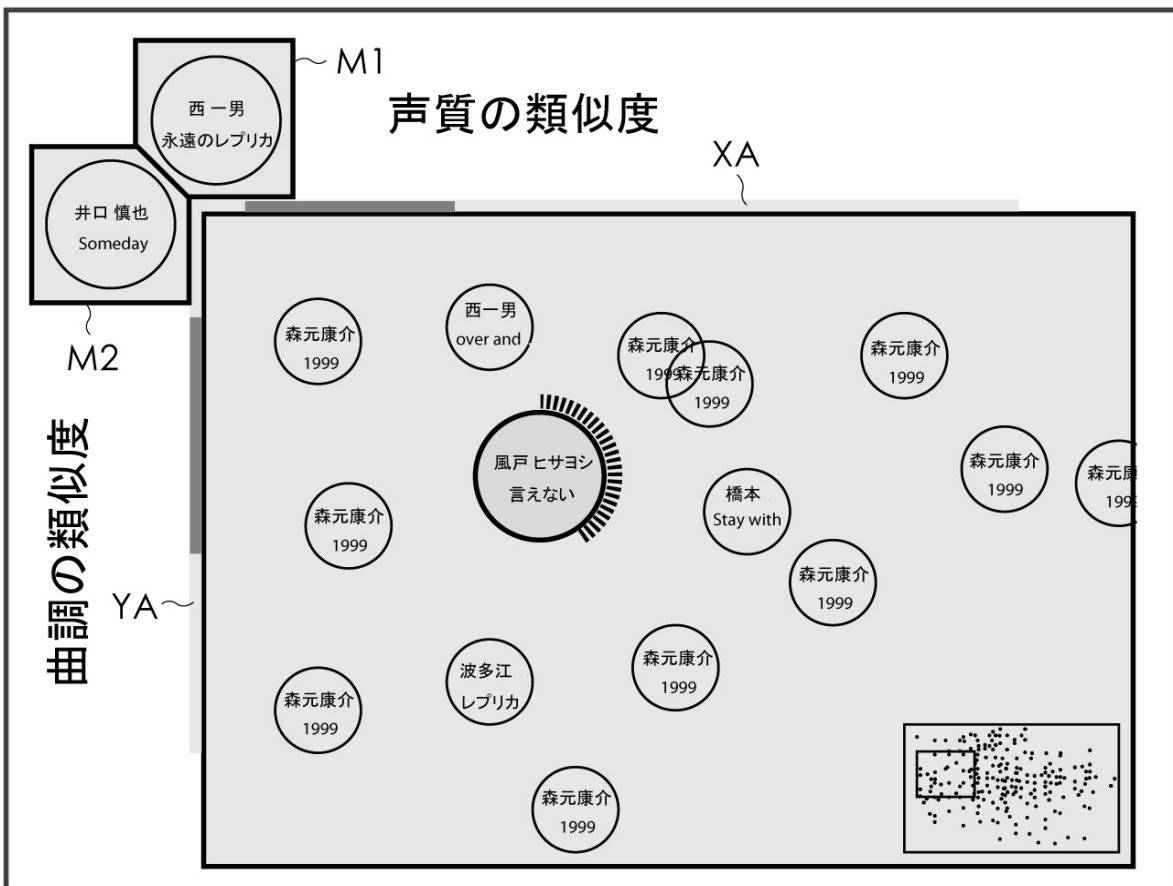
質問1: ボーカルの声質に関して、曲A、曲Bのどちらがより曲Xに似ていると思いますか？

質問2: 曲全体の音質に関して、曲A、曲Bのどちらがより曲Xに似ていると思いますか？

【図11】



【図14】



フロントページの続き

(出願人による申告)平成19年度独立行政法人科学技術振興機構「音楽デザイン転写・音響信号理解に基づく音インタフェース」委託研究、産業技術力強化法第19条の適用を受ける特許出願

- (56)参考文献 特開2002-014974(JP,A)
特開2004-152110(JP,A)
国際公開第2006/075432(WO,A1)
特開2000-285141(JP,A)
宮沢 祐光, 好みの楽曲を選定するための楽曲類似度測定方法, 第68回(平成18年)全国大会講演文集(2) 人工知能と認知科学, 社団法人情報処理学会, 2006年 3月 7日, 2-181~2-182ページ

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30