

多数の小領域スペクトログラムの探索に基づく背景音楽の高速探索法

永野 秀尚^{†a)} 柏野 邦夫^{†b)} 村瀬 洋^{†*c)}

A Fast Search Algorithm for Background Music Signals Based on the Searches for Numerous Small-Region Spectrograms

Hidehisa NAGANO^{†a)}, Kunio KASHINO^{†b)}, and Hiroshi MURASE^{†*c)}

あらまし 音声に音楽に重畳した音響信号（蓄積信号）の中から，参照信号として与えられた音楽が音声の背景に含まれる区間（目的区間）を見つけ出すことを背景音楽探索と呼ぶ．本論文では，この背景音楽探索を高速に行う手法について述べる．背景音楽探索では，探索の目的となる目的区間においても，他の音声の重畳により，その音響信号は参照信号と著しく異なり，参照信号を蓄積信号と直接照合する手法での探索は困難である．そこで，参照信号のスペクトログラムを時間周波数領域において多数の小領域に分割し，これらの小領域スペクトログラムと類似度の高い小領域スペクトログラムをもつ時点を蓄積信号のスペクトログラムから探索し，その探索結果を統合することで，目的区間を探索する手法を提案する．本手法は，参照信号の各小領域スペクトログラムと類似度の高い小領域スペクトログラムのみを，蓄積信号のスペクトログラムから，高速な探索手法を用いて探索することで，背景音楽探索全体を高速に行うことを特徴とする．実験では，各小領域スペクトログラムの探索に時系列アクティブ探索法（TAS）を用いることで，約 30 分の蓄積信号からの 15 秒の目的区間の背景音楽探索を 1 秒未満で行うことができた．

キーワード 音響信号探索，背景音楽探索，混合音響信号，小領域スペクトログラム

1. ま え が き

音や映像のメディア情報の増加，多様化により様々なメディア情報の探索技術が必要とされている．我々は特に，時系列探索，すなわち，長時間の音や映像の信号（蓄積信号）またはそのデータベース（DB）と，探したい音や映像の信号（参照信号）が探索キーとして与えられたとき，蓄積信号中の参照信号に類似する区間を探し出す探索技術について研究を進めてきた [1]～[4]．そして，特に，本論文では，音楽に音声に重畳した蓄積信号中から参照信号として与えられる音楽が音声の背景に含まれる区間（目的区間）を見つけ出す背景音楽探索を高速に行う手法について述べる．背景音

楽探索においては，ナレーションなどの他の音声の音楽への重畳がありえるものとする．そのため，目的区間では，参照信号として与えられる音楽と同一の音が背景に鳴ってはいるが，前景である音声が重なっているために，音響信号そのものは参照信号と大きく異なりうる．背景音楽探索は，例えばラジオ放送中や TV 放送中で背景音楽（BGM）として使われた音楽の検出に応用することができる．この場合，各放送を録音したものを蓄積信号として蓄積しているとする．そして，ある音楽の CD の一部を参照信号とし，蓄積信号中の目的区間を検出することで，その音楽が BGM として使われた区間を検出することができる．また，他の応用としては，BGM として使われている音楽音響信号を探索キーとするビデオ検索システムなども考えられる．

音楽検索についてはこれまで多くの研究がなされてきたが，それらはその目的により二つに大別できる．一つは，演奏者やアレンジが異なるが同じ曲の検索をねらう類似音楽探索 [4]～[10] であり，もう一つは音楽 CD などと信号レベルでほぼ同一の音楽の検索をねら

[†] 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所，厚木市

NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi-shi, 243-0198 Japan

* 現在，名古屋大学大学院情報科学研究科

a) E-mail: nagano@eye.br1.ntt.co.jp

b) E-mail: kunio@eye.br1.ntt.co.jp

c) E-mail: murase@is.nagoya-u.ac.jp

う一致音楽探索である．そして，この一致音楽探索は，目的区間の音響信号自体が参照信号とほぼ同一である場合の検索を行う全景音楽探索と，目的区間において音声の背景に流れる音楽が参照信号と信号レベルで同一である場合の検索を行う背景音楽探索に分けられる．このうち，全景音楽探索は，参照信号と信号レベルで同一の音響信号区間を蓄積信号中から検出する同一信号探索 [1], [2], [11], [12] により実現できる．そして，この同一信号探索は，参照信号を蓄積信号上でずらしながら蓄積信号上のすべての時点で照合するずらし照合法（全探索法と呼ぶ）により実現できるが，より高速な手法として時系列アクティブ探索法（time-series active search, 本論文では以下 TAS と表記する．付録 1. 参照）が提案されている [1]．TAS ではスペクトログラムの照合においてヒストグラム特徴を用いることと，類似しないスペクトログラムとの照合を省略し，探索区間をスキップすることで，スペクトログラムの相関を用いた全探索法（相関法）よりも高速な探索が行える．そして，TAS を特徴ひずみに対し頑健にする手法 [2] や参照信号中の部分区間の探索に用いる応用手法なども提案されている [12]．しかし，これらの同一信号探索の手法は，探索の目的となる区間では，音響信号そのものが参照信号とほぼ同一であることを前提としており，背景音楽探索のように，他の音声など，参照信号以外の音が重畳している場合には探索精度が低下する．

最近，このような背景音楽探索のための手法として自己最適化スペクトル相関法（self-optimized spectral correlation method, 本論文では SSC と表記する）が提案されている [13]．SSC は背景音楽同定のための手法であるが，この手法を用いると，全探索に基づいた背景音楽探索が可能である．SSC においては参照信号のスペクトログラムを時間周波数領域で小領域のスペクトログラムに分割し，この小領域スペクトログラムと蓄積信号のスペクトログラム中の小領域スペクトログラムとの類似度を，スケールパラメータの最適化により最大化された相関とする．本論文ではこの最大化された相関で得られる小領域スペクトログラムの類似度（小領域類似度）を特に小領域相関類似度と呼ぶ．そして，参照信号のスペクトログラムを分割して得られる各小領域スペクトログラムについて，全探索法により，この小領域スペクトログラムと，蓄積信号中の各時点における小領域スペクトログラムとの小領域相関類似度を計算する．そして，これらの小領域相関類似

度をスケール及び蓄積信号中の小領域スペクトログラムの時点に関する投票空間への投票により統合し，蓄積信号中の目的区間を探索する．この投票は，蓄積信号中の同じ時点で，同じスケールで類似する小領域スペクトログラムの小領域相関類似度を，そのスケールと時点における類似度として積算することで行う^(注1)．その結果，あるスケールと時点において積算された類似度が大きい場合，この時点に目的区間があったとする．このように，SSC においては小領域ごとの照合で選ばれたスケールパラメータを投票法により選び，重畳する音声の影響が少ない時間周波数領域で最適化されたスケールパラメータを用いて，参照信号と照合区間全体の類似度を測っていると言える．しかし，この SSC では，その計算量の大きさから，長時間の蓄積信号からの探索には時間がかかるという問題点があった．

そこで，本論文では，背景音楽探索の高速な手法を提案する．提案の内容は 2 点に要約される．第 1 点は，参照信号の小領域スペクトログラムへの分割による背景音楽探索の同一信号探索への帰着である．すなわち，提案手法においては，信号のスペクトログラムを小領域スペクトログラムに分割し，参照信号と同一ではないが，部分的には同一となる小領域スペクトログラムを多く含む区間を探索することと考える．実際，TV 放送などで音楽が BGM に用いられる場合の前景音響は多くの場合音声などであり，時間方向においても息つぎや間などにより，空白が存在する．また，音声による音響は周波数方向においても偏在し，音楽音響の周波数領域をすべて覆うことはまれなので，このようなアプローチは有効と考えられる．提案の第 2 点は，このような一致する小領域スペクトログラムの組合せを TAS により高速に見つけ出す点である．これにより相関法に比べ高速な一致個所の検出が可能と考えられる．

上記 2 点の提案内容を，処理の手順に沿ってまとめると以下ようになる．提案手法では，参照信号のスペクトログラムを多数の小領域スペクトログラムに分割し，各小領域スペクトログラムについてそれに類似する小領域スペクトログラムをもつ時点を蓄積信号から TAS により高速に探索する．そして，この類似す

(注1): この投票は，参照信号の小領域スペクトログラムの参照信号内での時点を考慮したものである．例えば，参照信号の先頭から t_r の時点の小領域スペクトログラムと蓄積信号中の時点 t_s の小領域スペクトログラムとの小領域相関類似度が s で，スケールパラメータが a のとき， s は，時点-スケールの投票空間内で，時点が $t_s - t_r$ ，スケールが a の類似度として投票され積算される．

表 1 提案手法の位置づけ

Table 1 Classification of previous methods and ours.

	全景音楽探索 (同一信号探索)	背景音楽探索
全探索法	相関法	SSC
スキップ探索法	TAS	提案手法

る小領域スペクトログラムが検出された時点に基づき、その時点の小領域スペクトログラムの類似度を投票法により統合し、背景音楽探索を行う。なお、提案手法では、TAS の適用に伴い、小領域スペクトログラムの類似度には小領域スペクトログラムのヒストグラム特徴の重なり率（小領域ヒストグラム類似度）を用いる。

提案手法における参照信号のスペクトログラムの小領域スペクトログラムへの分割と投票法による小領域類似度の統合の基本的な発想は SSC に共通するものがある。しかし、SSC では、参照信号の各小領域スペクトログラムについて、全探索法により、蓄積信号中の全時点での相関を計算し、その結果を統合しているのに対し、提案手法では、参照信号の各小領域スペクトログラムのみを TAS により高速に検出したのち、それら検出された小領域スペクトログラムの類似度のみを統合することで探索結果を得るという違いがある。このように、部分一致探索に帰着させることにより、背景音楽探索を高速に行えることが提案手法の特長である。各手法の位置づけを表 1 に示す。

以下、2. で目的とする背景音楽探索と提案手法について述べ、3. で実験結果について述べる。そして、4. でまとめる。

2. 背景音楽探索

2.1 目的

本論文で目的とする背景音楽探索は、図 1 にあるように、蓄積信号 (stored signal) 中の参照信号 (reference signal) で与えられる音楽音響が他の音声の背景に含まれる区間 (segment) をすべて見つけ出すことである。なお、この探索の目的とする区間を目的区間と呼ぶ。また、信号中の各区間は、その先頭の時点 (point in time) で区別する。

2.2 提案手法

図 2 に提案手法の概要を示し、以下に処理手順を述べる。

(提案手法)

まず、図 2(a) にあるように、参照信号、蓄積信号

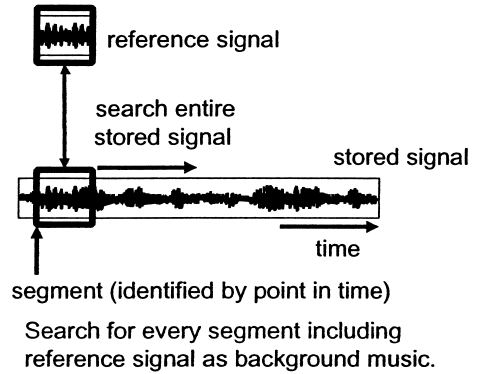


図 1 目的とする背景音楽探索
Fig. 1 Overview of background music retrieval.

の各音響波形信号からパワースペクトルを抽出し各々のスペクトログラム (spectrogram) を得る。そして、参照信号のスペクトログラムから一定のサイズの小領域のスペクトログラムを等間隔に切り出す。この小領域のスペクトログラムは周波数軸方向、時間軸方向に、各々一定の個数のスペクトログラム上の点をもとのスペクトログラムから切り出すことで得る。このような小領域のスペクトログラムを小領域スペクトログラム (small-region spectrogram) と呼ぶ。そして、先頭の時点が t_i 、周波数帯が w_m の参照信号中の小領域スペクトログラムを F_{t_i, w_m} と書く。また、蓄積信号のスペクトログラム中の、先頭の時点が t 、周波数帯が w_m の、 F_{t_i, w_m} と同じサイズの小領域スペクトログラムを G_{t, w_m} と書く。また、参照信号のスペクトログラム中の小領域スペクトログラム F_{t_i, w_m} を切り出した参照信号中のすべての時点 t_i の集合を T_R ($T_R = \{t_1, t_2, \dots\}$) とし、すべての周波数帯 w_m の集合を W ($W = \{w_1, w_2, \dots\}$) とする。なお、各小領域スペクトログラムのパワー値は、各小領域スペクトログラムごとに正規化する^(注2)。

次に、図 2(b) にあるように、参照信号中の各 F_{t_i, w_m} について、蓄積信号から F_{t_i, w_m} に類似する時点を、周波数帯 w_m において探索する。この探索は TAS を用いて行われる。なお、ここで、 F_{t_i, w_m} に類似する時点とは、 F_{t_i, w_m} と G_{t, w_m} の小領域類似度 $s^p(F_{t_i, w_m}, G_{t, w_m})$ が小領域探索しきい値 s_{th}^p を超えるような時点 t である。なお、本手法では、この類似する小領域スペクトログラムをもつ時点の探索に

(注2): 小領域内の各時点においてその時点の各周波数でのパワー値を小領域内のその時点のパワー値の平均値で正規化する。

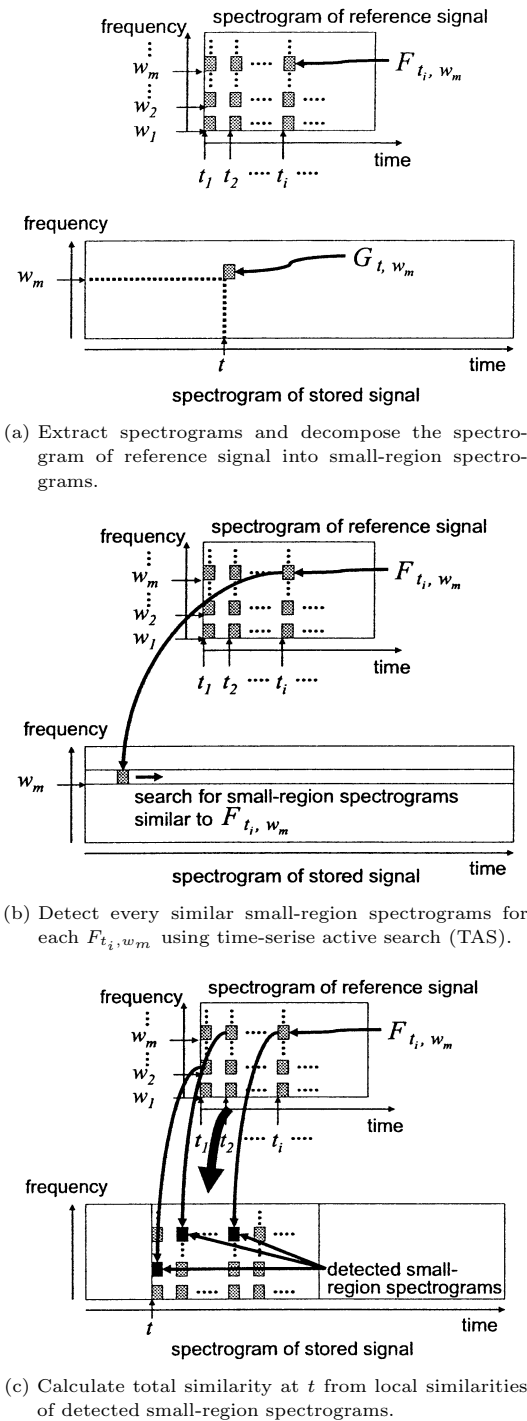


図2 提案手法の概要

Fig. 2 Overview of proposed search method.

TAS を用いるため、小領域類似度 $s^p(F_{t_i, w_m}, G_{t, w_m})$ としては F_{t_i, w_m} と G_{t, w_m} の間のヒストグラム重なり率 (付録 1. 参照) を用いる。ヒストグラム重なり率による小領域類似度を特に小領域ヒストグラム類似度と呼ぶ。

そして、図 2(c) にあるように、上記の各小領域スペクトログラム F_{t_i, w_m} についての探索結果をもとに、蓄積信号中の各時点 t について、小領域類似度を統合し、 t における参照信号との類似度 (区間類似度) $S(t)$ を、

$$S(t) = \frac{1}{|T_R|} \sum_{t_i \in T_R} \left(\max_{w_m \in W} (s^p(F_{t_i, w_m}, G_{t+t_i, w_m})) \right) \quad (1)$$

として求める。ここで、 $|T_R|$ は T_R の要素数を表す。なお、 F_{t_i, w_m} についての探索の結果、蓄積信号の時点 t において、 G_{t+t_i, w_m} が F_{t_i, w_m} に類似する小領域スペクトログラムとして検出されなかった場合、すなわち、

$$s^p(F_{t_i, w_m}, G_{t+t_i, w_m}) \leq s_{th}^p \quad (2)$$

の場合、式 (1) においては、

$$s^p(F_{t_i, w_m}, G_{t+t_i, w_m}) = 0 \quad (3)$$

とする。実際の探索においては、 G_{t+t_i, w_m} が F_{t_i, w_m} に類似する小領域スペクトログラムとして検出された場合のみ、式 (1) において $s^p(F_{t_i, w_m}, G_{t+t_i, w_m})$ を積算する。また、式 (1) においては、

$$\max_{w_m \in W} (s^p(F_{t_i, w_m}, G_{t+t_i, w_m})) \quad (4)$$

として、 $s^p(F_{t_i, w_m}, G_{t+t_i, w_m})$ について、その値が最大となるような周波数帯 w_m を全周波数帯の集合 W から選んでいる。これは、参照信号中の同じ時点の複数の異なる周波数帯の小領域スペクトログラムについて、これらに類似する小領域スペクトログラムとして、蓄積信号中の同じ時点の異なる複数の周波数帯の小領域スペクトログラムが検出された場合に、最も小領域ヒストグラム類似度が高い周波数帯、すなわち、重畳する音声中最も無音に近く、参照信号への重畳が少ないと考えられる周波数帯を選択するためである。

以上のようにして得られた区間類似度をもとに、区間類似度 $S(t)$ が探索しきい値 S_{th} よりも大きいような時点 t を先頭とする区間に参照信号が検出されたと

する (終)

□ 提案手法におけるスペクトログラムの小領域スペクトログラムへの分割については、その時間軸方向の長さがスペクトルの抽出間隔に比べて短すぎると、小領域スペクトログラム内のスペクトルの抽出数が少なくなり、小領域ヒストグラム類似度の信頼性が下がる。その結果、背景音楽探索における探索精度の低下を招く。しかし、小領域スペクトログラムの時間方向の長さが長すぎると、小領域スペクトログラムの照合において、前景音響の影響を受ける確率が高くなり、やはり探索精度が低下する。また、周波数軸方向への分割については前景音響の影響を避けるためにはより細かく分割することが望ましいが、その場合、小領域スペクトログラムの数が増えることにより、探索時間が増加するという問題がある。周波数軸方向への分割は探索精度との兼ね合いとなるが、探索精度を保ちながら、できるだけ少ない分割数にすることが望ましい。本論文では、これらの小領域スペクトログラムへの分割のパラメータは予備実験により決定することとし、3. で示すパラメータのようにした。

3. 実 験

3.1 探索精度

提案手法を用いた背景音楽探索の探索精度の比較実験について述べる。提案手法の探索精度の評価のため、(a) SSC, (b) 従来の TAS, (c) 提案手法の三つの手法で背景音楽探索実験を行い、各々の探索精度を比較評価した。(a) においては参照信号を小領域スペクトログラムに分割し、小領域スペクトログラムを蓄積信号と全探索により照合し、蓄積信号のすべての時点での小領域相関類似度を時点-スケールの投票空間に投票した。そして、投票空間において積算された類似度 (SSC における区間類似度とする) が探索しきい値より大きい時点を先頭とする区間を探索結果とすることとした。(b) においては、まず、スペクトログラムを分割しない従来の TAS によって、参照信号に類似する区間 (参照信号とのヒストグラム重なり率による類似度が 0.6 より大きい区間) を蓄積信号から探索する。そして、更に、この TAS による検出結果に対してしきい値処理を行い、ヒストグラム重なり率による類似度 (b) における区間類似度とする) が探索しきい値よりも大きい区間を探索結果とすることとした。また、(c) における小領域探索しきい値は 0.6 とした。

実験では、約 30 分の音楽の音響信号に音声の音響

信号を計算機上で混合したものを蓄積信号とした。各信号はサンプリング周波数 44.1 kHz, 16 bit 量子化のものを用いた。ここで、音楽音響信号にはポップスとロックの録音を用いた。また、音声音響信号には、DJ のトーク及びアーティストの対談を録音したものを用いた。ただし、この音声音響信号については、1 秒以上の無音区間がないように編集した。そして、音声音響信号の音楽音響信号への混合においては、TV 放送などで音楽が BGM として使われた場合を想定し、両信号の電力比 (混合比と呼ぶ) を 5, 0, -5, -10, -15, -20 dB とした。そして、各混合比において探索実験を行った。なお、ここで、混合比は

$$10 \log_{10} \left(\frac{\text{音楽音響信号の平均電力}}{\text{音声音響信号の平均電力}} \right) \quad (\text{dB}) \quad (5)$$

である。混合比が負数ということは音楽音響信号の方が音量が小さいことを意味する。参照信号はもとの音楽音響信号からランダムに切り出した 15 秒間の信号 100 個とした。そして、各参照信号について、もとの音楽音響信号から参照信号を切り出した時点と同じ蓄積信号中の時点をこの参照信号に対する目的区間とし、蓄積信号から探索した。探索により検出された区間を検出区間と呼ぶ。なお、本実験においては、ある区間 t が検出され、 t と t からのずれが前後 15 秒未満の範囲内の検出区間のうち、 t の区間類似度が最大であった場合、この範囲においては t のみが検出区間として検出されたとした。すなわち、区間類似度が前後 15 秒未満の範囲で極大かつ探索しきい値より大きい区間のみを検出区間とした。そして、検出区間のうち、目的区間からのずれが前後 15 秒未満の検出区間があればそのうちの一つを正しい検出区間とし、他のすべての検出区間は正しくない検出区間とした。また、目的区間からのずれが前後 15 秒未満の位置に検出区間が検出された場合、この目的区間は検出されたとし、探索結果に含まれたとした。

探索精度については、検出された検出区間のうちの正しい検出区間の割合 (正答率) と、目的区間のうち探索結果に含まれた割合 (検出率) の 100 回の試行における平均値が等しくなるように探索しきい値を各混合比の場合で調整し、等しくなったときのこれらの割合の値を探索精度とした (正答率, 検出率及び探索精度については付録 2. も参照のこと)。なお、各混合比においては、100 回の試行を通して探索しきい値は一定である。

次に、(a)、(b)、(c)の各手法について、実験に用いた他のパラメータを述べる。なお、パラメータの設定においては、すべての手法で、探索に用いる周波数帯域は500 Hzから2 kHzとし、スペクトルの周波数軸方向の抽出点数は28とした。この帯域は、携帯電話などの音声通信を目的とした帯域の狭い音響信号での探索を想定したものである。また、(a)と(c)の参照信号の小領域スペクトログラムへの分割については、両手法で同じ個数の小領域スペクトログラムを切り出すこととし、予備実験による探索精度と探索速度の予備評価から、周波数軸方向は四つに分割し、小領域スペクトログラムの参照信号からの切り出し間隔は600 msとした。

なお、(a)と(c)の比較においては音響特徴の抽出法、抽出間隔、分析窓長、並びに、小領域スペクトログラムの時間方向の長さ、小領域スペクトログラムへの周波数軸方向への分割の仕方についても、同じ条件で行うことを検討すべきである。そこで、(a)では文献[13]に従い、FFTで音響特徴を抽出し、小領域スペクトログラムへの周波数軸方向の分割は等間隔に行うこととし、(c)でも同様とすることを検討した。しかし、(c)の場合の探索精度を評価した予備実験において、(c)で、(a)と同様に、音響特徴の抽出にFFTを用い、小領域スペクトログラムへの分割は周波数軸上で等間隔に分割する場合と、音響特徴の抽出には中心周波数を対数軸上で等間隔に配置した帯域通過フィルタバンクを用い[1]、小領域スペクトログラムへの分割は対数軸上で等間隔に分割するように工夫した場合の探索精度を比較したところ、後者の方がより良い探索精度を得られた。そのため、(c)では後者の方法をとった。そしてこれらの条件の下、音響特徴の抽出間隔及び分析窓長、小領域スペクトログラムの長さは、(a)と(c)の場合で、各々、探索精度が最大になるよう実験的に定めた。これは、(a)と(c)で、各々、探索精度が最大となるパラメータが異なるためである。そして、(b)においては(c)と同じ音響特徴を用いることとした。

以下、実験に用いたパラメータを詳細に述べる。(a)では文献[13]に従い、FFTにより音響波形信号からスペクトルを抽出した。なお、FFTの分析窓長は50 msとし、5 msごとにスペクトルを抽出した。そして、周波数軸上で500 Hzから2 kHzの間に等間隔にある28個の周波数において得られたFFTの振幅を実験に用いるスペクトルとした。そして、参照信号の小領域ス

ペクトログラムへの分割については、周波数軸方向には、上記の28個の出力を7点ずつ、四つに分割し、時間軸方向には、長さ50 msとすることとした。一つの小領域スペクトログラムには周波数軸方向で7点のスペクトルが時間軸方向に10組あることになる。そして、このような小領域スペクトログラムを参照信号から600 msおきに切り出した。(b)の場合は、音響特徴の抽出には上述のように帯域通過フィルタバンクを用いた。帯域通過フィルタは計28個のフィルタを500 Hzから2 kHzの帯域に対数軸上で等間隔に配置して用いた。そして、各フィルタで2 msの時間区間の分析を0.5 msごとに行い、各時間、各周波数におけるパワーを抽出し、各時間の28個のフィルタの出力をその時点のスペクトルとした。なお、(b)では参照信号の小領域スペクトログラムへの分割などは行わず、蓄積信号中から参照信号に類似する区間を単に従来のTASを用いて探索した。また、(c)では(b)と同じ音響特徴を用いた。そして、参照信号の小領域スペクトログラムへの分割については、周波数軸方向には、上記の28の帯域通過フィルタの出力を周波数の順に七つずつに分けた4区間への分割とし、小領域スペクトログラムの時間軸方向の長さは100 msとした。すなわち、各小領域スペクトログラムは七つの帯域通過フィルタから得られる100 ms分、200個所の分析によるスペクトログラムとした。そして、このような小領域スペクトログラムを参照信号から600 msおきに切り出した。表2にこれらのパラメータをまとめる。

各手法の各混合比における探索精度を表3に示す。混合比が5 dBの場合、(c)の探索精度は100%であり、検出もれも過剰検出もないことが分かる。また、各混合比において(c)の探索精度が最も良いことが分かる。

表2 実験パラメータ

Table 2 Parameters for experiments.

音響特徴抽出	(a) SSC (従来法)	(b) TAS (従来法)	(c) 提案手法
抽出法	FFT	帯域通過フィルタバンク	
分析窓長	50 ms	2 ms	
抽出間隔	5 ms	0.5 ms	
特徴抽出点数 (周波数軸上)	28		
小領域への分割	(a) SSC (従来法)	(b) TAS (従来法)	(c) 提案手法
分割の有無	あり	なし	あり
小領域の長さ	50 ms	—	100 ms
周波数軸方向	4等分	—	4等分(対数軸上)
切り出し間隔	600 ms	—	600 ms

表 3 探索精度 (%)
Table 3 Search accuracy (%).

混合比 (dB)	(a) SSC (従来法)	(b) TAS (従来法)	(c) 提案手法
5	99.8	93.0	100.0
0	98.0	88.1	99.0
-5	92.9	81.0	96.8
-10	72.0	72.9	94.8
-15	44.9	64.0	90.3
-20	24.8	-	73.2

しかし、(a)、(c) のいずれの場合においても混合比が小さくなるにつれて、探索精度が低下している。このような BGM が非常に小さい場合に探索精度を上げるためには、周波数軸方向において、より細かい小領域スペクトログラムへの分割が必要になると思われる。なお、(b) の場合は各混合比で探索精度が低く、背景音楽探索への適用は難しいと考えられる。また、(b) の混合比が -20 dB の場合では、しきい値を 0.6 とした TAS の段階での検出もれが多く、この TAS の検出結果に対するしきい値処理において探索しきい値を調整しても、検出率と正答率が等しくなることはなかった。また、(a) においては、音響特徴の抽出間隔を短くすることで精度の向上を図ることも考えられるが、予備的に行った実験では探索精度の向上は見られず、探索時間は増加した。

3.2 探索速度

次に提案手法における探索速度の評価実験について述べる。本実験では、提案手法の探索速度を、背景音楽探索が可能な SSC、及び、提案手法で小領域スペクトログラムの探索に全探索法を用いた場合と比較して評価した。SSC と提案手法について、探索時の実験条件は 3.1 と同じである。小領域スペクトログラムの探索を単純に全探索を用いて行う場合は、各小領域スペクトログラムを蓄積信号上のすべての時点において照合し、小領域ヒストグラム類似度を求めることになる。以下、提案手法で小領域スペクトログラムの探索に全探索法を用いた場合をヒストグラム全探索法と呼ぶ。ヒストグラム全探索法のその他の実験条件は提案手法と同じである。

表 4 に、3.1 の背景音楽探索において、SSC、ヒストグラム全探索法、提案手法の 1 回の背景音楽探索で行われた小領域スペクトログラムの照合回数の平均値を示す。3.1 で述べたように、いずれの混合比においても、提案手法の小領域探索しきい値は 0.6 である。なお、SSC と、ヒストグラム全探索法については、小

表 4 30 分の蓄積信号の探索における平均照合回数
Table 4 Average number of matchings. (Stored signal: 30 m)

混合比 (dB)	SSC (従来法)	ヒストグラム全探索法 (従来法)	提案手法
5	3.3×10^7	3.3×10^8	3.4×10^6
0	同上	同上	3.4×10^6
-5	同上	同上	3.3×10^6
-10	同上	同上	3.3×10^6
-15	同上	同上	3.3×10^6
-20	同上	同上	3.3×10^6

領域スペクトログラムを蓄積信号のスペクトログラム上のすべての時点で照合するため、混合比による照合回数の変動はない。SSC とヒストグラム全探索法の照合回数をみると、ヒストグラム全探索法の場合の音響特徴の抽出間隔が SSC の場合の 1/10 であるため、ヒストグラム全探索法の場合の照合回数が SSC の場合の照合回数の 10 倍になっていることが分かる。しかし、ヒストグラム全探索法と提案手法では、音響特徴の抽出間隔は同じであるが、提案手法では TAS により、不要な時点での照合を省略しながら小領域スペクトログラムに類似する時点を探るため、提案手法では照合回数を削減することができる。表 4 からは、いずれの混合比においても、提案手法の照合回数はヒストグラム全探索法の照合回数の約 1% にまで大幅に削減されていることが分かる。そして、その結果、提案手法の照合回数は SSC の照合回数の約 10% となっている。SSC における小領域スペクトログラムの一度の照合にかかる計算時間と、ヒストグラム全探索法、提案手法における小領域スペクトログラムの一度の照合にかかる計算時間は異なるが、照合回数の観点からは、提案手法により、より高速な背景音楽探索が可能になることが考えられる。なお、提案手法においては混合比が小さくなるにつれて、照合回数が減少する。これは、前景の音声信号が大きくなるにつれ、音楽音響信号である参照信号の小領域スペクトログラムを蓄積信号と照合したときの小領域類似度が小さくなり、TAS におけるスキップ幅が大きくなるためと考えられる。

次に、1 回の背景音楽探索にかかった CPU 時間の平均値を表 5 に示す。表 4 にある照合回数の削減にともない、表 5 においても提案手法の探索時間がヒストグラム全探索法の場合の約 2% まで大幅に削減されていることが分かる。そして、提案手法の探索時間は SSC の約 3% であることが分かる。以上の実験から、

表 5 30 分の蓄積信号の探索における平均探索時間 (CPU time)

Table 5 Average search time. (CPU time, stored signal: 30m)

混合比 (dB)	SSC (従来法)	ヒストグラム全探索法 (従来法)	提案手法
5	30.3 s	48.5 s	0.86 s
0	30.6 s	48.5 s	0.86 s
-5	30.7 s	48.3 s	0.86 s
-10	30.7 s	48.4 s	0.85 s
-15	30.7 s	48.3 s	0.85 s
-20	30.7 s	48.2 s	0.85 s

提案手法が背景音楽探索を高速に行えることが分かる。

なお、表 4 において、提案手法の照合回数はヒストグラム全探索法の照合回数の約 1% であるが、表 5 において、提案手法の探索時間はヒストグラム全探索法の探索時間の約 2% となっている。これは、提案手法の方が、1 回の照合において、蓄積信号上の小領域スペクトログラムのヒストグラムの作成にかかる計算時間が長いからである。ヒストグラム全探索法の場合は、蓄積信号上で小領域スペクトログラムの照合位置を一つずらしたとき、蓄積信号上の小領域スペクトログラムのヒストグラムは、一つ前の時点のヒストグラムから 1 要素を取り除き、そして、新たに 1 要素を加えることで作成できる。提案手法では 1 回の小領域スペクトログラムの照合を行う際に、前に照合を行った時点のヒストグラムから、TAS によるスキップ幅だけ要素を入れ換えたヒストグラムを作成する必要がある。

また、表 4 において、ヒストグラム全探索法の照合回数は SSC の照合回数の 10 倍であるが、表 5 において、ヒストグラム全探索法の探索時間は SSC の探索時間の約 1.6 倍である。これは、ヒストグラム全探索法の各照合における小領域ヒストグラム類似度の計算時間が SSC の各照合における小領域相関類似度の計算時間よりも小さいためと考えられる。また提案手法の照合回数は SSC の照合回数の約 10% であるが、同じ理由により、提案手法の探索時間は SSC の探索時間の約 3% にまで削減されている。なお、表 5 の探索時間には、前処理 (音響波形信号からのスペクトルの抽出など) にかかる時間は含まれていない。また、CPU 時間の計測には PC (CPU: Pentium 4 2GHz, L2 キャッシュ 256 kByte, OS: Red Hat Linux 7.2) を用いた。

4. む す び

本論文では、音響信号の部分一致に着目した、高速な背景音楽探索の手法を提案した。本手法は参照信号のスペクトログラムを小領域スペクトログラムに分割し、各小領域スペクトログラムに類似する小領域スペクトログラムをもつ時点を蓄積信号のスペクトログラムから TAS により高速に探索し、検出された時点の小領域類似度を統合し、その結果から参照信号を含む区間を検出する。この類似する小領域スペクトログラムの探索に TAS を用い、小領域スペクトログラム同士の照合回数を削減することで、高速な背景音楽探索が可能になる。実験では小領域スペクトログラムの探索に全探索を用いた場合に比べ、探索時間を約 2% にまで短縮することができた。そして、その結果、15 秒の参照信号を用いた約 30 分の蓄積信号からの探索を 1 秒未満で行うことができた。今後は、より音声の重畳にロバストな背景音楽探索のための参照信号の小領域スペクトログラムへの分割法と小領域類似度の統合法の検討を行いたい。

謝辞 日ごろ御指導を頂く NTT コミュニケーション科学基礎研究所の菅村昇所長、片桐滋主席研究員、牧野昭二部長に感謝する。また日ごろ御協力を頂く同研究所メディア認識研究グループの諸氏に感謝する。

文 献

- [1] 柏野邦夫, ガビンズミス, 村瀬 洋, “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), vol. J82-D-II, no.9, pp.1365-1373, Sept. 1999.
- [2] 黒住隆行, 柏野邦夫, 村瀬 洋, “時系列アクティブ探索のための特徴ひずみに頑健な確率ディザボータイング” 信学論 (D-II), vol. J84-D-II, no.8, pp.1817-1825, Aug. 2001.
- [3] 木村昭悟, 柏野邦夫, 黒住隆行, 村瀬 洋, “グローバルな枝刈りを導入した音や映像の高速探索” 信学論 (D-II), vol. J85-D-II, no.10, pp.1552-1562, Oct. 2002.
- [4] H. Nagano, K. Kashino, and H. Murase, “Fast music retrieval using polyphonic binary feature vectors,” Proc. ICME2002, vol. I, pp.101-104, Lausanne, Switzerland, Aug. 2002.
- [5] 蔭山哲也, 高島洋典, “ハミング歌唱を手掛かりとするメロディ検索” 信学論 (D-II), vol. J77-D-II, no.8, pp.1543-1551, Aug. 1994.
- [6] A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith, “Query by humming: Musical information retrieval in an audio database,” Proc. ACM Multimedia '95, pp.231-236, San Francisco, USA, Nov. 1995.
- [7] L.A. Smith, R.J. McNab, and I.H. Witten, “Sequence-based melodic comparison: A dynamic programming approach,” in Melodic Similarity :

Concepts, Procedures, and Applications, vol.11 of Computing in Musicology, ed. W.B. Hewlett and E. Selfridge-Field, pp.101-117, The MIT Press, England, 1998.

[8] 園田智也, 後藤真孝, 村岡洋一, “WWW 上での歌声による曲検索システム,” 信学論 (D-II), vol.J82-D-II, no.4, pp.721-731, April 1999.

[9] N. Kosugi, Y. Nishihara, S. Kon'ya, M. Yamamuro, and K. Kushima, “Music retrieval by humming – Using similarity retrieval over high dimensional feature vector space,” Proc. IEEE PACRIM '99, pp.404-407, Victoria, Canada, Aug. 1999.

[10] 橋口博樹, 西村拓一, 張 建新, 滝田順子, 岡 隆一, “モデル依存傾斜制限型の連続 DP を用いた鼻歌入力による楽曲信号のスポッティング検索,” 信学論 (D-II), vol.J84-D-II, no.12, pp.2479-2488, Dec. 2001.

[11] 杉山雅英, “セグメントの高速探索法,” 信学技報, SP98-141, Feb. 1999.

[12] 西村拓一, 水野道尚, 小木しのぶ, 関本信博, 岡 隆一, “アクティブ探索法による時系列データ中の一致区間検出 参照区間自由時系列アクティブ探索法,” 信学論 (D-II), vol.J84-D-II, no.8, pp.1826-1837, Aug. 2001.

[13] 安部素嗣, 西口正之, “背景音楽同定のための自己最適化スペクトル相関法,” 信学技報, PRMU2001-209, Jan. 2002.

付 録

1. 時系列アクティブ探索法 (TAS)[1] の概要

TAS の概要を図 A・1 に示す。TAS は、蓄積信号のスペクトログラムから、参照信号のスペクトログラムとのヒストグラム重なり率による類似度がしきい値 θ より大きいスペクトログラムをもつ参照信号と同じ長さの区間を探索する。

まず、スペクトログラム X と Y のヒストグラム重なり率を説明する。ここで X と Y は、時間方向と周波数方向で同じサイズのスペクトログラムである。はじめに、スペクトログラム上の各時点のスペクトル特徴を正規化^(注3)した上で、ベクトル量子化 (VQ)^(注4)により符号化し、各スペクトログラムについて、VQ 符号の系列 (code string) を作成しておく。次に、ヒストグラム重なり率の計算であるが、各スペクトログラムについて上記の VQ 符号の出現回数を数え上げたヒストグラム (ヒストグラム特徴) を作成する。そして、 h^X と h^Y を X と Y のヒストグラム特徴をとすると、 X と Y のヒストグラム重なり率 $S_h(h^X, h^Y)$ は

The segment beginning at time t is detected if and only if the similarity $S(t)$ is greater than a threshold value θ .

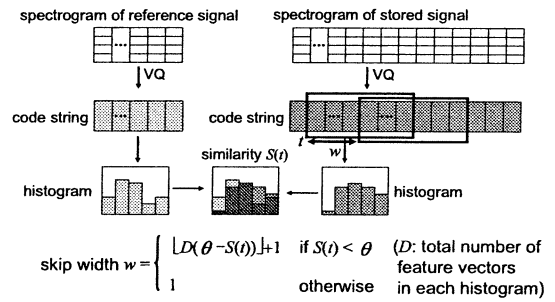


図 A・1 TAS の概要
Fig. A・1 Overview of time-series active search (TAS).

$$S_h(h^X, h^Y) = \frac{1}{D} \sum_{i=1}^L \min(h_i^X, h_i^Y) \quad (A.1)$$

として計算される。なお、ここで、 h_i^X と h_i^Y は、それぞれ h^X と h^Y の i 番目のビンに含まれる度数であり、 L はヒストグラムのビンの数、 D は一つのヒストグラムの総度数である。

TAS においては、上記のヒストグラム重なり率をスペクトログラムの類似度用いる。そして、参照信号のスペクトログラムと蓄積信号中の区間 t のスペクトログラムのヒストグラム重なり率を $S(t)$ としたとき、 t で照合の後、 $S(t)$ を用いて次の照合位置までのスキップ幅 (skip width) w を図 A・1 のように計算し、照合位置を w だけずらして照合を行い、新たにスキップ幅を計算する。ただし、 $[x]$ は x を超えない最大の整数を表す。TAS では、これを繰り返していくことで、探索を行う。そして、照合した区間でのヒストグラム重なり率が θ より大きい場合、その区間を検出する。TAS では、このようにしてスキップにより照合回数を削減しながらもヒストグラム重なり率がしきい値 θ より大きい区間をもらさず探索できる。

2. 探索精度について

本論文の探索においては、探索しきい値 S_{th} を大きくすれば正答率は上がり、検出率は下がる。そのため、探索精度は、探索しきい値を調整し、両割合の平均値が同じになったときの両割合の値で測ることとした。また、このとき、正しい検出区間の数を高々一つとすることで、(検出率 - 正答率) が探索しきい値に関する広義の単調減少関数となる。その結果、

(注3): 本論文では (注2) で述べたように、各時点のスペクトル特徴の各要素を、それらの平均値で正規化する。

(注4): 本論文では、VQ の符号帳の学習には LBG アルゴリズムを用いた。

(検出率の平均値 - 正答率の平均値) も探索しきい値に関する広義の単調減少関数となり、両割合の平均値が等しくなるときの探索しきい値を容易に求めることができる。

(平成 15 年 5 月 23 日受付, 9 月 24 日再受付)



永野 秀尚 (正員)

平 6 阪大・基礎工・情報卒。平 8 同大大学院修士課程了。同年 NTT に入社。現在、NTT コミュニケーション科学基礎研究所研究主任。パターンの認識・理解、情報検索に興味をもつ。情報処理学会、IEEE 各会員。



柏野 邦夫 (正員)

平 2 東大・工・電子卒。平 7 同大大学院電気工学専攻博士課程了。工博。同年 NTT に入社。現在、NTT コミュニケーション科学基礎研究所主任研究員。この間、平 12 から 1 年間英国ケンブリッジ大学客員研究員。音響信号の認識・分離・探索、及び情報統合の研究に従事。メディア情報を対象とする信号処理及び知識処理に興味をもつ。平 5 情報処理学会奨励賞、平 9 人工知能学会全国大会優秀論文賞、平 11 日本音響学会粟屋潔学術奨励賞、平 13 本会ソサイエティ論文賞、平 14 本会業績賞各受賞。情報処理学会、日本音響学会、人工知能学会、IEEE 各会員。



村瀬 洋 (正員)

昭 53 名大・工・電子卒。昭 55 同大大学院修士課程了。同年日本電信電話公社(現 NTT)に入社。以来、文字・図形認識、コンピュータビジョン、マルチメディア認識の研究に従事。平 4 から 1 年間米国コロニア大客員研究員。平 13 NTT コミュニケーション科学基礎研究所メディア情報研究部部長。平 15 名古屋大学大学院情報科学研究科教授。現在に至る。工博。昭 60 本会学術奨励賞、平 4 電気通信普及財団テレコムシステム技術賞、平 6 IEEE-CVPR 国際会議最優秀論文賞、平 7 情報処理学会山下記念研究賞、平 8 IEEE-ICRA 国際会議最優秀ビデオ賞、平 13 高柳記念電子科学技術振興財団高柳記念奨励賞、平 13 本会ソサイエティ論文賞、平 14 本会業績賞、平 15 文部科学大臣賞各受賞。情報処理学会、IEEE 各会員。