社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－ＨＩ－122 (2)
2007－SLP－ 65 (2)
2007／2／9

ベクトル空間法を用いた相対的感性表現による音検索

相川 清明、谷島 加奈子

東京工科大学メディア学部
〒192-0982 八王子市片倉町 1404-1
aik@media.teu.ac.jp

あらまし 音の検索においては定量的な感性表現が難しい。そこで、選択された候補を基準としてインタラクティブに候補を検索する方法を考案した。ベクトル空間法を用いて、計測した多変量な情報を相手が理解しやすい言語表現で出力する手法については気象情報案内システムとして既に報告している。本報告では、音を感性表現のベクトルで表し、ベクトルの類似度を用いて検索する方法について述べる。感性表現による検索では、第１候補で満足できない場合がある。本報告では、提示された候補を基準として次の候補を探索する方法を提案する。本方法は映画の効果音検索などの応用が考えられる。

Vector-Based Sound Retrieval using Relative Queries in Emotional Representation

Kiyoaki Aikawa and Kanako Yajima

School of Media Science, Tokyo University of Technology
1404-1 Katakuracho, Hachioji, Tokyo 192-0982, Japan

Abstract Emotional representations can be more effective than conventional keywords such as genres and artist names in retrieving music. A vector-based Weather Reporter system has been reported for representing weather information in emotional phrases such as “refreshing” or “muggy” depending on the measured weather parameters. A sound retrieval system can be achieved by replacing the weather parameters into emotional parameters, the linguistic weather expressions into sounds, respectively. However, it is difficult to quantitatively represent the emotional features of a sound. This paper proposes a method for interactively retrieving sounds using relative queries regarding the found candidate as the current basis. This method can contribute to the sound effect retrieval for producing movies.

1. はじめに

人は五感で感じ取った情報を直感的にわかりやすい表現を用いて相手に伝えている。今までに、このメカニズムを、ベクトル空間法を用いて模擬する方法について報告してきた。その実現例が、温度や湿度などの気象情報を「さわやか」「うっとおしい」などの相手に直感的に理解しやすい表現に置き換えて出力するシステム、Weather Reporterである[1-6]。

Weather Reporterシステムでは、気象情報表現用語を温度や湿度などから構成されるベクトルで表現し、センサーから取り込んだ観測データに対する同じ構成のベクトルに最も近いベクトルを選び、それに対応する気象情報表現用語を出力する。ベクトルの近さを測る類似度は余弦距離を用いた。観測データのベクトルと選ばれた気象情報表現用語のベクトルの類似度の値は与えられた観測データと選ばれた気象情報表現用語の適合の度合いを示している。従って、類似度の値は、選ばれた表現を用いて良いかどうかの一種の確信度と考えられる。そこで、この類似度に応じて「かもしれない」などの確信度を表す表現を付加するようにした[1-2]。また、このとき、2つの気象情報表現用語のベクトルの和が観測データのベクトルに近い場合、2つの表現を併用して表現する機能を持たせた。このとき、2つの用語に対応するベクトル間の非類似度を求めて、接続詞を自動決定する方法についても報告した[5]。Weather Reporterシステムは地域に依存した表現への対応も可能なため、バイリンガルシステムや大阪方言システムについても検討を行った[3-4]。

本報告では、感性表現のベクトルを用い、ベクトル空間法により音を検索する方法について述べる。感性表現による検索では、第1候補で満足できない場合がある。第2候補を探す時、ベクトルの類似度が大きい順に候補を提示する方法が考えられる。また、感性表現を修正して再探索を行う方法も考えられる。しかし、音の感性特徴は定量的な表現が難しく、これらの方法では目的にたどり着けない可能性がある。そこで、一度選択された候補を基準としてインタラクティブに候補を探索する方法を考案した。ベクトルを用いた音楽検索の例はあるが[7]、ベクトルを用いた相対的な検索方法は報告されていない。

2. ベクトル空間法による音検索

今までに、ベクトル空間法を用いたWeather Reporterシステムについて報告してきた。この方法に基づき、気象情報表現用語を選ぶ代わりに音を選ぶように変更すれば、音検索に用いることができる。図1は音の感性特徴3種から構成されるベクトルを用いたベクトル空間法による音検索の模式図である。音の特徴としては、「楽しさ」「悲しさ」などが考えられる。

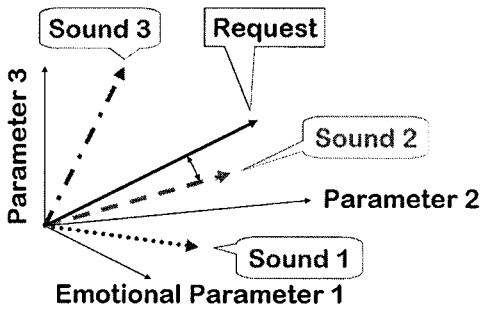


図1　ベクトル空間法による音検索

音の特徴はN次元空間中のベクトルとして表現される。従って、検索対象の音、検索要求は共にこのN次元空間中のベクトルとして表現される。本報告では検索要求を表すベクトルを検索ベクトル、音の特徴を表すベクトルをテンプレートベクトルと呼ぶ。

検索ベクトルをx、k番目のテンプレートベクトルをa_kとすると、検索ベクトルとk番目のテンプレートベクトルの類似度$s(x, a_k)$は以下の余弦距離で計測される。

$$s(x,a_k)=\frac{(x,a_k)}{|x\|a_k|}=\frac{(x,a_k)}{\sqrt{(x,x)(a_k,a_k)}}=\cos\theta_k \qquad (1)$$

ここで、θ_kは2つのベクトルの角度を表す。(x, a_k)は内積を表し、次式で与えられる。

$$(x,a_k)=\frac{1}{N}\sum_{j=1}^{N}x(j)a_k(j) \qquad (2)$$

ここで、Nはベクトルの次元数を表す。

検索要求に対し、最も類似度が高いテンプレートベクトルに対応する音が検索結果として示

される。なお、本報告では、Weather Reporter システムで用いた２つの表現の合成は用いない。

3. 相対的検索

図１において、最初の検索要求により、３つの音のうち音２に対応するテンプレートベクトルが、検索ベクトルとなす角度が最も小さく類似度が高いため、検索結果として選ばれたとする。これに満足できない場合、今までの類似度に基づく検索法では、次の候補を探す方法として以下の２つの方法がある。

① 類似度の順に候補を示す。

② 検索ベクトルを設定しなおして検索する。

感性表現のベクトル空間において、自分が設定した検索ベクトルと各音のベクトルの配置はイメージが描きにくい。従って、上記の２方法では目的の音を探しにくい可能性がある。そこで、一度選ばれた音のベクトルを基準とし、「選ばれた音よりパラメータ３が大きく、パラメータ２が小さいもの」などの相対的な要求により次の候補を検索する方法が考えられる。本報告ではこのような相対的な要求をベクトルで表したものを相対ベクトルと呼ぶことにする。

図２は相対的な検索の方法を模式的に描いたものである。この考えに基づいて検索を行うには、音２のベクトルを起点として相対ベクトル方向にあるすべてのテンプレートベクトルのうちから、最も類似度が高くなるベクトルを探索する。すなわち、１つの音のベクトルに対し、最初に選ばれた音２のベクトルと、正の実数倍された相対ベクトルの和のベクトルがなす角度が最も小さくなるように相対ベクトルの倍率を求める。全ての音についてこの演算を行い、もっとも小さい角度が得られる音を次候補として求める。図２では、その探索の結果、音３が次の候補として選択される様子を示している。

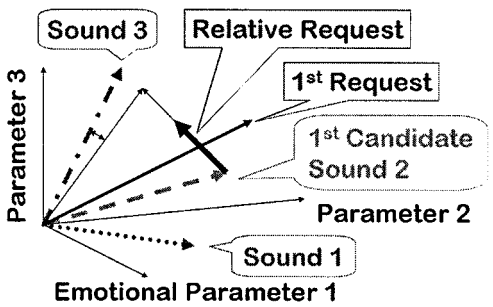


図２　ベクトル空間法による相対検索

次に、この探索法を定式化する。最初の検索の結果、１つのテンプレートベクトル a_k が選ばれているとする。これを基準にして、より「楽しい」、より「悲しい」、などのように相対的な探索を行うことを考える。相対ベクトルを r、正の実数倍率をλとすると、２つのベクトルの和は

$$a_k + \lambda r \tag{3}$$

により与えられる。最初に選択された音のベクトル a_k と相対検索ベクトル r が与えられているとき、類似度 $S(a_k+\lambda r,a_m)$が最大となる m 番目の音のテンプレートベクトル a_m を求めればよい。すなわち、M 個のテンプレートベクトル a_m について正の実数倍率λを変化させて最大の類似度を求め、その類似度が最大となる m を選べば良いことになる。a_m は以下の式で表される。

$$\begin{aligned} a_m &= \arg\max_{m,\lambda} s(a_k + \lambda r, a_m) \\ &= \arg\max_{m} z_m \\ z_m &= \arg\max_{\lambda} s(a_k + \lambda r, a_m) \end{aligned} \tag{4}$$

$$\begin{aligned} & s(a_k + \lambda r, a_m) \\ &= \frac{(a_k + \lambda r, a_m)}{|a_k + \lambda r||a_m|} \\ &= \frac{(a_k + \lambda r, a_k)}{\sqrt{(a_k + \lambda r, a_k + \lambda r)(a_m, a_m)}} \end{aligned} \tag{5}$$

ある m について類似度を最大化するにはλで微分した関数が 0 となるλを以下の式により求めればよい。

$$\frac{\partial s(a_k + \lambda r, a_m)}{\partial \lambda} \to 0 \tag{6}$$

ある m についての類似度 $s(a_k+\lambda r,a_m)$は

$$\begin{aligned} & s(a_k + \lambda r, a_m) \\ &= \frac{(a_k + \lambda r, a_m)}{|a_k + \lambda r||a_m|} \\ &= \frac{(a_k, a_m) + \lambda(r, a_m)}{\sqrt{(a_k, a_k) + 2\lambda(a_k, r) + \lambda^2(r, r)}|a_m|} \end{aligned} \tag{7}$$

で表されるので、

$$
\begin{aligned}
&\frac{\partial s(a_k+\lambda r, a_m)}{\partial \lambda} \\
&= \frac{1}{|a_m|}\frac{(r,a_m)}{\sqrt{(a_k,a_k)+2\lambda(a_k,r)+\lambda^2(r,r)}} \\
&- \frac{1}{2|a_m|}\frac{(a_k,a_m)+\lambda(r,a_m)}{\left((a_k,a_k)+2\lambda(a_k,r)+\lambda^2(r,r)\right)^{3/2}} \\
&\times\left(2(a_k,r)+2\lambda(r,r)\right) \\
&= \frac{1}{|a_m|}\frac{(r,a_m)\left((a_k,a_k)+2\lambda(a_k,r)+\lambda^2(r,r)\right)}{\left((a_k,a_k)+2\lambda(a_k,r)+\lambda^2(r,r)\right)^{3/2}} \\
&- \frac{1}{|a_m|}\frac{\left((a_k,a_m)+\lambda(r,a_m)\right)\left((a_k,r)+\lambda(r,r)\right)}{\left((a_k,a_k)+2\lambda(a_k,r)+\lambda^2(r,r)\right)^{3/2}}
\end{aligned}
\tag{8}
$$

となる。このうち、分子 n は

$$
\begin{aligned}
n &= (r,a_m)(a_k,a_k)+2\lambda(r,a_m)(a_k,r) \\
&+\lambda^2(r,a_m)(r,r) \\
&-(a_k,a_m)(a_k,r)-\lambda(r,a_m)(a_k,r) \\
&-\lambda(a_k,a_m)(r,r)-\lambda^2(r,a_m)(r,r) \\
&=(r,a_m)(a_k,a_k)+\lambda(r,a_m)(a_k,r) \\
&-(a_k,a_m)(a_k,r)-\lambda(a_k,a_m)(r,r)
\end{aligned}
\tag{9}
$$

であるから、分子が 0 となるλは

$$
\lambda = -\frac{(r,a_m)(a_k,a_k)-(a_k,a_m)(a_k,r)}{(r,a_m)(a_k,r)-(a_k,a_m)(r,r)} \tag{10}
$$

として求めることができる。このλを式(7)に代入すればテンプレートベクトル a_m についての最大類似度を求めることができるので、それを最大にする m を選べば良いことになる。

4. Sound Advisor システム

ベクトル空間法による相対的な検索の方法に基づき、MATLAB を用いて音検索を行う Sound Advisor システムを作成した。

4.1. 音の特徴表現

本研究では音の特徴を表す感性表現を表 1 に示す 8 種に絞った。テンプレートベクトルの値はアンケート調査により求めた。88 種の効果音を用意し、14 名の被験者に対して各音が 8 種の感性表現とどのような関係があるかをアンケート調査で調べた。回答は表 2 の 5 種類のうち 1 つを選ぶ形式で行った。

表 1.　音選択のための感性表現

感性表現	選択肢数
楽しさ	5
悲しさ	5
恐怖	5
落ち着き	5
怒り	5
不気味	5
明るさ	5
爽やかさ	5

表 2　音に対する感性表現のレベル

1	非常に遠い
2	やや遠い
3	どちらともいえない
4	やや近い
5	非常に近い

8 種の感性表現に対し、被験者の回答を平均して 8 次元のベクトルとする方法が考えられる。しかし、感性表現によっては図 3 のように、被験者の回答が特定の値に集中するものと、人により感じ方が異なり、分散するものがある。従って、感性表現毎の平均値のみを用いたのでは人によるばらつきの有無が反映できない。これに対して、分散も用いる方法が提案されている[7]。本研究では分布をそのまま用いることを考えた。すなわち、全ての被験者の回答を累積して 40=5×8 の長さのベクトル上での頻度分布としてテンプレートベクトルを作成した。この考え方では、「楽しさからは非常に遠い」、「悲しさに近い」などの 40 種類がすべてベクトルの要素となる。

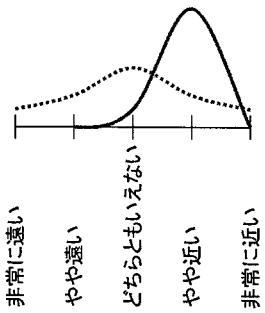


図 3　アンケート結果の分布の模式図

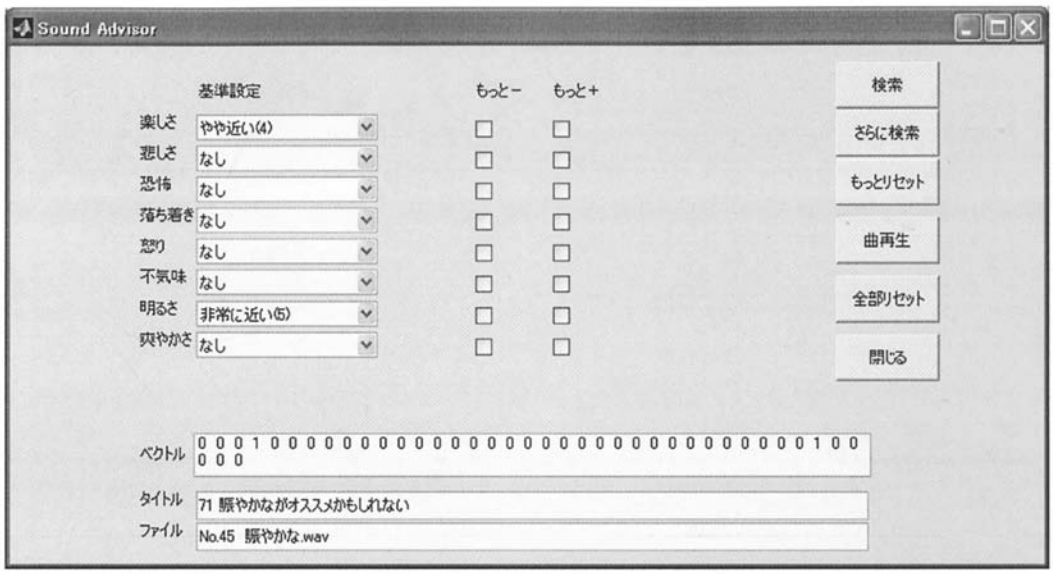


図４　Sound Advisor システムの GUI

4.2. GUI

図４に Sound Advisor システムの GUI を示す。左側には８種の感性表現が並ぶ。ポップアップメニュにより表２の５種から各感性表現のレベルを設定する。図５は１番目の感性表現を「どちらともいえない」に設定した場合を示している。右側の検索ボタンを押すと余弦距離による音検索が行われ、下部のタイトル欄に音のタイトルが、ファイル欄に音ファイル名が表示される。ベクトル欄には検索要求のベクトルが表示される。右下の曲再生ボタンを押すと音が演奏される。

相対ベクトルを設定するために「もっと＋」と「もっと−」のチェックボックスを設定した。「もっと＋」にチェックを入れると、まず図６左側上段のように相対ベクトルの「どちらともいえない」次元に１を設定する。「さらに検索」ボタンを押すと、この相対ベクトルで検索を行って結果を出力した後、図６左側中段のように右側の「やや近い」側に１の位置を変更する。再度「さらに検索」ボタンを押すと、１の位置はさらに右にシフトする。何もチェックが入っていない場合には、図６右側のように相対ベクトルは０が並んだまま変化は起こらない。

「もっと−」にチェックを入れた場合は、まず図７上段のように相対ベクトルの「どちらともいえない」次元に１を設定する。「さらに検索」ボタンを押すたびに相対的な検索を行った後、１の位置を左にシフトさせる。

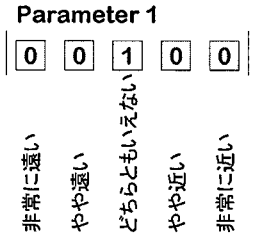


図５　ポップアップメニュにより設定される検索ベクトル。「なし」では全てが０となる。

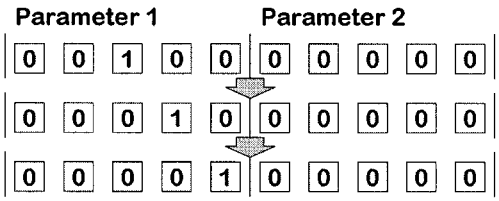


図６　「もっと＋」チェック時の相対ベクトル

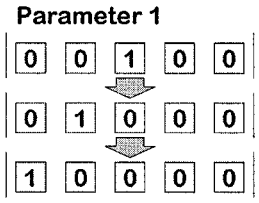


図７　「もっと−」チェック時の相対ベクトル

5. 評価実験

10 名の被験者に音検索システム Sound Advisor を使って音検索を行ってもらった。試行回数は 1 名あたり 3 回である。その結果、「楽しい」、「明るい」、「爽やか」を主な検索条件とし、ポップアップメニュの設定レベルが表 2 の 5 段階のうち 3 以上である場合は、92%（11/12）で検索された音がイメージしていた音に近いという結果となった。しかし、「悲しい」「恐怖」「落ち着く」「不気味」を主な検索条件とした場合は、イメージしていた音に近いという割合は 50%(7/14)であった。また、相対的検索により 60%(18/30)においてイメージした音に近づいたという回答を得た。

6. 結論

ベクトル空間法を用いた感性表現による音検索法の 1 手法を提案した。本方法の特徴は、一旦検索された音の候補を基準とする相対的な表現を用いて、より適切な候補を検索することである。初期の候補を求める場合も相対的な表現を用いた検索においても、ベクトルの近さを測る類似度としては一貫して余弦距離を用いている。評価実験の結果、相対検索を行うことにより、全体の 60%でより適切な候補が検索できた。本方法は映画への効果音の付加などの応用が考えられる。

謝辞

確信度表現部は、飯田朱美博士のプログラムを使用させて頂いた。ここに深く感謝する。

参考文献

[1] 飯田朱美, 上野嘉人, 松浦良平, 相川清明, “ベクトル空間法を用いたイメージを想起させるお天気表現システム”, 情報処理学会 第 109 回ヒューマンインタフェース・第 52 回音声言語情報処理共催研究会 HI109SLP52, pp. 113-18, (2004-07).

[2] Iida, A., Ueno, Y., Matsuura, R., Aikawa, K. “A Vector-based Method for Efficiently Representing Multivariate Environmental Information”, Proceedings of ICSLP 2004, pp.269-272, (2004-09).

[3] 飯田朱美, 相川清明, "Weather Reporter II 微妙なニュアンスを共通語と大阪方言で語るお天気表現システム", インタラクション 2005, C-327, (2005-02).

[4] 飯田朱美, 相川清明, "微妙なニュアンスを伝える表現を自動生成するバイリンガルお天気表現システム", 日本音響学会講演論文集, Vol. I, pp. 657-658, (2005-03).

[5] 飯田朱美, 相川 清明, "ベクトルの非類似度を用いて複数表現の接続詞を自動決定するお天気情報システム", 音声言語情報処理, Vol. 2005-SLP-57, No.24, pp. 141-146, (2005-07).

[6] Kiyoaki Aikawa and Akemi Iida, “Vector-based language generation for associatively evoking environmental images”. J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, pp.3038, (2006-11).

[7] Kimiko Ohta, Tadahiko Kumamoto, and Hitoshi Isahara, “Design of an impression-based music retrieval system”, J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, pp.3236, (2006-11).