3N-7

OnomaTree: 擬音語と木構造を併用した環境音検索インターフェース

清水 敬太† 北原 鉄朗‡ 駒谷 和範‡ 尾形 哲也‡ 奥乃 博‡
† 京都大学 工学部情報学科 ‡ 京都大学大学院 情報学研究科 知能情報学専攻

1. はじめに

環境音（本稿では音声と楽音以外の音全般をさす）は，効果音という形で映像作品や音楽作品にしばしば登場し，作品に盛り上がりやアクセントをつけるなど一定の役割を果たしている．近年は，このような作品制作のために，さまざまな効果音を収録したCDが多数市販されており，制作者は膨大な効果音コレクションから目的のものを選ぶことができる．しかし，付与されている説明は十分ではなく，音源（what sound it is）は分かってもどのような音か（how it sounds）は聴いてみないとわからないことが少なくない．本稿では，このような状況を鑑み，環境音を効率的に検索できるシステムについて検討する．具体的には，はじめに音源名で絞り込みを行い，それでも多数のデータが該当するという状況を想定し，同じ音源（ここでは「水」）によるデータが多数含まれたデータベースに対する検索を扱う．

2. 環境音の検索

2.1 従来研究

環境音の検索については，環境音を扱った研究自体が少ないこともあって，研究事例は非常に少ない．そのようななかで，和気らの研究 [1] は，環境音検索の先駆的研究といえる．和気らは，人間が音を表すときに用いる表現として「波形の説明」（擬音語および音の高さ・長さの説明）「音源の説明」（音源や発音状況の説明）「主観の表現」（形容詞的表現）の3種に着目し，各説明方式の有効性を被験者実験を通じて分析した．そして，この3種の説明方式を用いた環境音検索システムを構築した．林ら [2] は，擬音語と音色のパラメータの対応関係をあらかじめ人手で記述しておくことで，擬音語による環境音の検索を実現した．石原ら [3] は，環境音にそれを表現する擬音語を自動的に付与する手法を開発し，擬音語をクエリーとした環境音検索システムを構築した．田口ら [4] は，擬音語では環境音の特徴を十分に表現できないことを指摘し，擬音語に音高の遷移や時間的構造を追加したXMLタグを設計し，その自動付与に取り組んだ．

これらの研究の共通点は，個々の音をどう表現するか，あるいは，クエリーと個々の音とのマッチングをどう取るかに焦点を当てており，ユーザがもつ環境音コレクションにどのような音があるのか，コレクション全体を大まかに表現する方法は扱っていなかったことである．この機能は，次節で述べるように，環境音検索において大変重要であると考える．

2.2 環境音コレクション全体の可視化と絞り込みに基づく検索方式

我々は，次の理由により，環境音検索において，検索対象である環境音コレクションにどのような音があるのかを大まかに可視化することが重要と考える．

- **ユーザが求める音がそもそもそのコレクションに含まれているのかをまず知ることができる．**
  擬音語に基づく環境音検索では，擬音語が必ずしも一意に定まらないため，そもそも自分が求める音が存在しないのか，クエリーの作り方が悪いのか迷うことがしばしばある．
- **ユーザが今聴いている音がコレクション全体のどこに位置付けられるかを知ることができる．**
  これにより，たとえば目的の音に近いけど少し異なる音が見つかったときに，より目的に近い音を探す余地があるか判断の目安とすることができる．
- **明示的にクエリーを作らずに絞り込みで検索できる．**
  擬音語は人によって表現に差ができることから，ラベラー（自動付与の場合は学習データのラベラー）によって擬音語表現に“くせ”ができると考えられる．そのため，何もヒントがない状態から適切なクエリーを作るのは容易ではない．全体の可視化表現から徐々に絞り込んでいくことで，明示的にクエリーを作らずに目的の音に達することができる．

そこで，次章では，環境音コレクション全体を木構造として可視化し，徐々に絞り込むことで目的の環境音へ達することができるインターフェースを提案する．

3. OnomaTree: 擬音語と木構造を併用した環境音検索インターフェース

以上の議論に基づき，以下のように環境音検索インターフェースを設計した．

3.1 全体像

本インターフェースでは，環境音をその音響的類似性に基づいて階層的に分類し，木構造として可視化する．つまり，この木構造においては，葉の部分に個々の環境音が配置され，根は環境音コレクション全体を表し，それ以外の節点は，音響的に類似した環境音を集めたグループ（クラスタと呼ぶ）を表す．そして，各々の葉には，対応する環境音を表現する擬音語をラベルとして付与し，葉以外の各節点にはその子孫がラベルとしてもつ全擬音語の集合を付与する．

この木構造に基づいたインターフェースの画面を図1に示す．木構造の表示は，よくディレクトリ構造の表示などに用いられるインターフェースと共通となっており，容易に使いこなせると考えられる．また，クラスタに付与されている擬音語の集合のすべての要素を表示すると表示が煩雑になってしまうので，そのクラスタのなかで最も多くの環境音に付与されている擬音語を代表ラベルとし，代表ラベルのみを表示する．その他の擬音語は，マウスカーソルを代表ラベル上にポイントしたときに表示されるようにした．また，代表ラベル，その他の擬音語ともに，対象クラスタ内にその擬音語が多く含まれているほど大きな文字で表示されるようにした．また，葉をクリックした場合には対応する環境音を，葉以外の節点をクリックした場合には，そのクラスタの代表音を再

Onomatree: Interface for Retrieving Environmental Sounds Using Onomatopoeia and Tree Structure: *Keita Shimizu, Tetsuro Kitahara, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto University)*

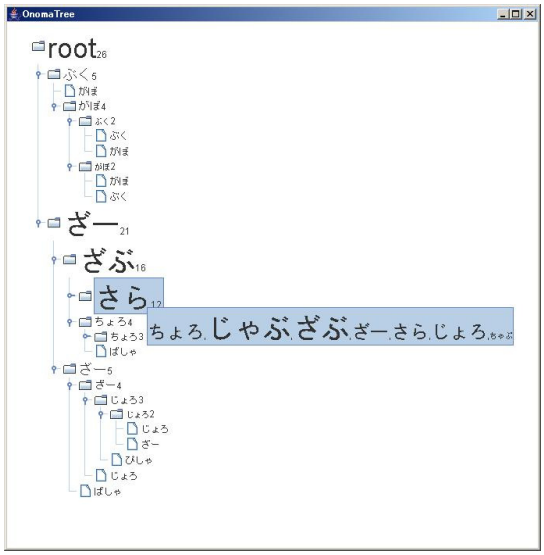

図 1: 設計した環境音検索インターフェース

生できるようにし，耳でも音を確認できるようにした．ここで，代表音は特徴空間内で対象クラスタの重心から最も近い音とする．このように擬音語による目での確認と試聴による耳での確認の両方ができるようになっているのは，次の理由によるものである．前者は，クラスタに含まれるすべての擬音語を一度に表示できる（一覧性）が，擬音語だけでは音をイメージしにくい．一方，後者は音そのものを試聴するのでわかりやすい（具体性）が，一度には 1 つしか試聴できない．そのため，両方を併用することでお互いの欠点を補えると考えたためである．

3.2 木構造の作成手法

木構造は，階層的クラスタリングを用いて作成する．これは，環境音間の特徴空間内での距離を算出し，距離が近い環境音対を 1 つのクラスタにまとめるという処理をすべての環境音が 1 つのクラスタになるまで再帰的に繰り返す手法である．ここでは，群平均法（ある対象からクラスタまでの距離を，そのクラスタ内の各要素までの距離の平均とする手法）を用いる．特徴空間には，音楽音響信号処理の研究 [5, 6] で用いられている Intensity，Brightness，Rolloff，Bandwidth などの 33 個の特徴量を主成分分析（累積寄与率 95%）で次元圧縮したものを用いる．

3.3 擬音語の設計

日本語で用いられる擬音語は多種多様であり，すべての擬音語を扱うのは事実上不可能であるため，本研究では以下のように擬音語の構造に制限を設ける．

本研究では，擬音語は**語基**と**音韻構造変化**に分けられると考える [7]．たとえば，「ぼた」という語基に対して反復あるいは促音の付与という音韻構造変化を加えることで「ぼたぼた」や「ぼたっ」などの擬音語を得ると考える．語基は 1 モーラあるいは 2 モーラの音素列とし，音韻構造変化としては反復，長音化，促音・撥音の付与を考える．語基は音そのものの特徴を表すのに対し，音韻構造変化は音の鳴り方（同じ音が反復的に鳴るのか，継続的に鳴るのか，単発的になるのかなど）を表していると考えられる．つまり，音韻構造が異なっていても語基が同じであれば音そのものの特徴は類似していると考え，語基のみを扱うものとする．

さらに，語基を子音と母音に分けると，子音が共通であれば音の印象は類似していると考える．たとえば「べちゃ」「びちゃ」「ばちゃ」はいずれも類似した印象をもつと考えられる．そこで，音響的特徴からは子音のみを決定し，母音はトップダウンに定める．

3.4 擬音語ラベルの自動付与

上記のように，音響信号からは語基の子音のみを決定する．子音の決定には，通常の音声認識で用いられるメル周波数ケプストラム係数（MFCC）と隠れマルコフモデル（HMM）を用いる．ただし，たとえば「びちゃ」と表されるような音は，通常の音声と異なり，最初に「び」に相当する音があって，その後に「ちゃ」に相当する音があるとは考えにくい．そこで，2 モーラの語基（の子音）も 1 つのクラス（「びちゃ」であれば「b-ch」）とみなして学習・認識を行う．さらに，日本語において一般的に用いられない語基は認識対象から除外する．その結果，認識すべきクラス数は 47 個となった．

母音については子音（2 モーラなら子音列）決定後，その子音（列）の下での各母音の出現頻度に基づいて決定する．これは，本来はコーパスからの統計的な分析によって決めるのが望ましいが，現時点で大規模な擬音語コーパスが存在しないので，各子音（列）に対する最瀬の母音（列）を人手で与えた．

4. 実装

上で述べたインターフェースを Java を用いて実装した．擬音語の学習・識別には HTK，階層的クラスタリングには Matlab を用いた．HMM の学習には，市販の効果音 CD から得た水の音約 200 例を用いた．

HMM の学習に用いたのとは別の水の音 26 個を用意し，本インターフェースで検索できるように処理を行った．階層的クラスタリングでは，波の音 3 つが最も葉に近い節点で 1 つのクラスタとなっていたり，水が激しく流れる音 4 つも最も葉に近い節点で 1 つのクラスタになるなど，音の類似性を適切に反映する結果となった．擬音語については，いくつか不適切なものがあったが，本インターフェースの有用性を妨げるほどではなかった．

5. おわりに

本稿では，木構造と擬音語ラベルによって効率的に環境音を検索できるインターフェースを提案した．今後は，得られた木や擬音語の妥当性の定量的評価およびインターフェースの試用実験を行う予定である．

謝辞 本研究の一部は，学術振興会科学研究費補助金，21 世紀 COE プログラムの支援を受けた．また，特徴量抽出には西山正紘氏のプログラムを利用した．

参考文献

[1] Wake, S. and Asahi, T.: Sound Retrieval with Intuitive Verbal Descriptions, *IEICE Trans. Inf. and Syst.*, **E84-D**, 11, pp.1568–1576 (2001).

[2] 林, 徳原: 映像コンテンツの音響製作を効率化する為の効果音検索—擬音語による効果音検索手法の提案—, 第 65 回情処全大, 4T8A-2 (2003).

[3] 石原他: 環境音を対象とした擬音語自動認識, 人工知能学会論文誌, **20**, 3, pp.229–235 (2005).

[4] 田口他: 擬音語・抑揚・リズムに基づく環境音記述用 XML タグの設計と自動付与, 第 68 回情処全大, 3L-7 (2006).

[5] 西山他: マルチメディアコンテンツにおける音楽と映像の調和の分析, 第 69 回情処全大 (2007).

[6] Lu, L. *et al.*: Automatic Mood Detection and Tracking of Music Audio Signals, *IEEE Trans. Audio Speech Lang. Process.*, **14**, 1, pp.5–8 (2006).

[7] 田守: オノマトピア—擬音・擬態語の楽園, 勁草書房 (1993).