

環境音・楽音受聴時における擬音語の特性分析と 擬音語曖昧検索への応用

山田祐雅^{†1} 小坂直敏^{†1}

概要: われわれは、マルチメディアコンテンツ制作の支援ツールとして、環境音・楽音のデータベースに対して音の検索・管理・探査を行うことができる「電子音色辞書」を構築してきた。同システムは、分類やキーワードなどの音情報の他に擬音語を保持することや、インターネットを通じて誰でも電子音色辞書サーバにアクセスし、電子音色辞書の全機能を利用できる点が特徴である。擬音語検索ではユーザが登録時に入力した擬音語から検索することができる。しかし擬音語は個人内でも個人間でも多義的に記述され、システムは擬音語の多義性を考慮していないため、検索結果に正しく反映されない。これらの多義的な擬音語文字列の平均値と標準偏差を求め、擬音語の表記実験を行った。環境音・楽音を対象とした音の連続性・繰り返し性・階層性・音韻の長さ・言語性の各要因について、レーベンシュタイン距離を応用した集合中央文字列と集合中央値を用いることによる、平均的な擬音語と標準偏差を分析し、個人内・個人間の変動を明らかにした。また、擬音語曖昧検索への応用手法を述べる。

1. はじめに

近年、インターネットや高性能コンピュータの普及により、個人でDAW (Digital Audio Workstation)などのソフトウェアを用いてのDTM (DeskTop Music)や動画制作、コンピュータ音楽などのマルチメディアコンテンツ制作が盛んに行われている。制作の際には、さまざまな効果音やサンプリング音源を大量に保有するデータベースから、適切な音素材を効率的に選択し使用することが重要である。

われわれは、マルチメディアコンテンツ制作の支援ツールとして、環境音・楽音のデータベースに対して、音の検索・管理・探査を行うことができる「電子音色辞書」を構築してきた[1]。同システムでは、音の管理や擬音語検索などの音検索を行うことが出来る音色ブラウザ機能を有す。また、3次元空間上へ音オブジェクトを球体で配置した3D音色表示機能[2]を有す。配置は、人の聴感特性を考慮した。これにより、音探査を行うことが出来る。電子音色辞書は、分類やキーワードなどの音情報の他に擬音語を保持することや、インターネットを通じて誰でも電子音色辞書サーバにアクセスして音素材を共有できることが特徴である。電子音色辞書の外観を図1に示す。

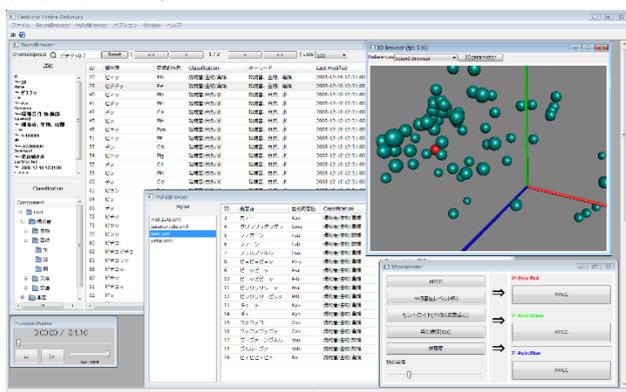


図1 電子音色辞書の外観

一般に、擬音語には、同一のサンプル音に対して、個人間のみならず個人内でも異なる表記をする、という多義性がある。しかしシステムの表示の便宜上、1つの音を1つの擬音語で表すことが望ましい。本稿では、電子音色辞書に登録されている擬音語を登録擬音語、検索者が検索する擬音語を検索擬音語と表記する。これまで、音の登録者が同時に登録擬音語を入力してきたが、検索者は音の検索時に登録擬音語とは別の検索擬音語を入力することが一般的である。その結果、完全一致による検索では、所望の音を検索することが非常に難しくなる。特に、「トントン」、「コケッコウ」のような辞書項目以外の音は検索が非常に困難である。

図2に具体的な例を挙げる。音の登録者が雷の音を“ズドバーン”と、ティンパニの音を”ドーン”と登録する。雷の音が欲しい検索者は擬音語検索で”ドーン”と入力するが、ティンパニの音が検索結果に含まれ、所望の音を見つけない。

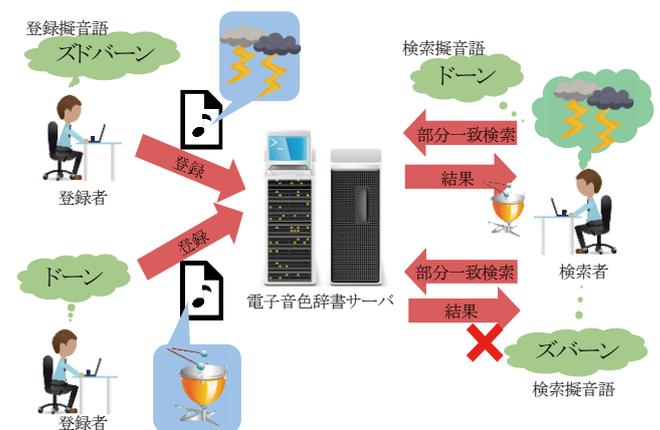


図2 擬音語の多義性による検索失敗例

擬音語の多義性による検索の失敗を防ぐため、検索擬音語と登録擬音語間の編集距離の閾値を手動で指定する擬音語曖昧検索機能を実装した。曖昧検索で用いる数値を根拠のあるものとするため、音の連続性、繰り返し性、階層性を要因とした、環境音・楽音受聴時における擬音語表記実

^{†1} 東京電機大学
Tokyo Denki University

験を行った。本稿では、編集距離による擬音語の個人内・個人間変動の実験結果を報告し、実験結果を用いて閾値を自動指定することができる擬音語曖昧検索への応用手法を述べる。

2. 擬音語曖昧検索

登録されている擬音語と異なる擬音語入力に対して、正解の音データを検索するために、2つの擬音語の距離に依存した手法を考える必要がある。離散的な記号の距離については編集距離の一種であるレーベンシュタイン距離がよく知られている[3]。レーベンシュタイン距離には、編集方法を追加・変更した Damerau-Levenshtein 距離[4]や Compression distance[5]、String edit distance with moves[6]などの拡張がある。本研究では一般的で扱いやすいレーベンシュタイン距離を用いて、検索擬音語と登録擬音語が一定距離内の場合は、検索結果に含むシステムを考える。

レーベンシュタイン距離は、2つの文字列がどの程度異なるかを数値で表す手法である。文字の挿入、置換、削除によって1つの文字列を別の文字列に変形するのに必要な手順の最小回数として与えられる。レーベンシュタイン距離は、バイオインフォマティクス分野で、DNAにおける挿入、置換、欠失の3つの方法による変化に注目し、DNA配列同士の類似性を判断するためにも応用されている[7]。

一般的に、距離は2つの文字列の最大文字列長で正規化する。レーベンシュタイン距離を求めるアルゴリズムは多数あるが、基本は動的計画法が用いる。文字列長 n と文字列長 m の文字列間の距離を求める計算時間は $O(nm)$ である。図3に曖昧検索のフローチャートを示す。

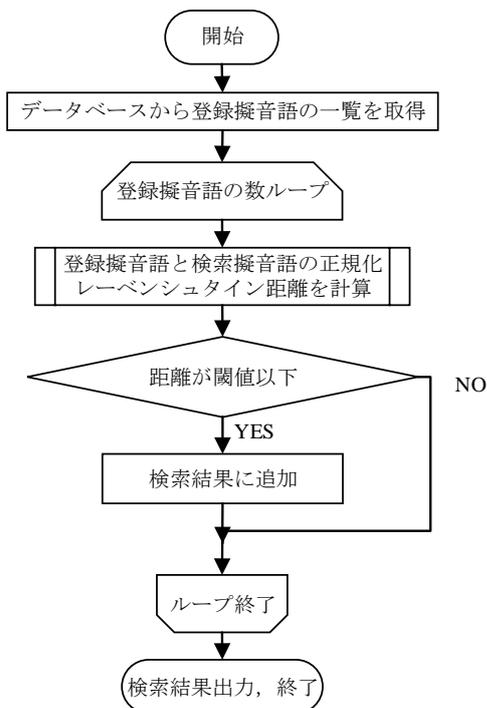


図3 曖昧検索のフローチャート

3. 擬音語表記実験

3.1 実験目的

閾値を指定することによって曖昧検索を行うことができる機能を実装した。しかし、検索毎の閾値入力は効率が悪い。検索したい音や検索擬音語の特徴によって、自動で閾値を指定するシステムを考えたい。そこで、いくつかの環境音や楽音に対する擬音語の文字列を、集合中央値、集合中央文字列、標準偏差を用いて分析を行い、音の違いによる個人間変動・個人内変動、編集距離の平均・標準偏差と閾値との関係を定量的に明らかにすることを目的とする。

3.2 集合中央値と集合中央文字列

集合中央値と集合中央文字列は、T. Kohonen[8]により、文字列認識における中央値の文字列を求める方法が提案された。集合中央文字列により、平均的な擬音語を求めることができる。以下に中央値を求める方法を概説する。

文字列集合 $S = \{S_1, S_2, \dots, S_i, \dots, S_I\}$ ($i = 1, 2, \dots, I$) を考える。この集合の中から一つの文字列 S_1 と、文字列 S_1 と文字列 S_2 のレーベンシュタイン距離を $LD_{(S_1, S_2)}$ と表すとすると、文字列集合における文字列 S_i と他の文字列との正規化下レーベンシュタイン距離の平均 MD_i は以下のように求まる。

$$MD_i = \frac{1}{I-1} \times \sum_{n=1}^I LD_{(S_i, S_n)}, \quad (n \neq i)$$

すべての文字列に対して計算し、すべての総和の中での最小値を集合中央値 MD_{min} 、集合中央値を持っている文字列を集合中央文字列(Set Median String)と定義する。図4に集合中央値と集合中央文字列の図解を示す。図4は $I = 5$ のときを表し、破線より実線の平均距離が短いので、 S_4 が集合中央文字列となる。

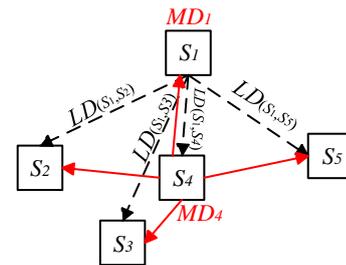


図4 集合中央値と集合中央文字列の例

3.3 実験概要

現在、電子音色辞書に登録されている音は約2000個ある。フィールドレコーディングによって収録した音や、効果音CDから実験的に登録した音を含む。これらの音ファイルは、量子化ビット数16bit・サンプリング周波数44.1kHzまたは48kHzのwavファイルである。本実験では、このデータベースのうち56個を刺激音とした。表1に、被験者の年代と人数を、表2に実験で使用した刺激音56種類の音の分類とサンプル数を示す。

表1 被験者の年代と人数

	日本人男性	日本人女性
20代	15	5
30代	1	0

表2 刺激音 56 種類の音の分類とサンプル数

大分類	中分類	小分類	サンプル数
環境音	生物	鳥類	12
		獣類	4
		虫	1
	自然	水	8
	天候	雷	1
	交通	飛行機	1
		電車	2
	打撃音	金属	3
		生活	家庭
	その他		3
楽音	打楽器	打楽器	5
		ベル	1
		その他	1
	弦楽器	打弦楽器	1
		擦弦楽器	1
	管楽器	木管楽器	1
		金管楽器	1
音声	無意味音声	外国語	6

被験者は、ランダムで提示された 56 種類の刺激音を十分な休憩を取りながら受聴し、擬音語をひらがなで表記する。ある音に対して、被験者全員から得られた擬音語を擬音語集合と定義する。また被験者のうち、30 代男性 1 名と 20 代女性 1 名は、この実験を十分な休憩を取りながら 13 回繰り返し、個人内変動の特性も調査する。

個人内変動は、以下の手順で求める。

1. 各音に対して平均的な擬音語を求めるために、ある被験者から得られた擬音語から、集合中央値と集合中央文字列を求める。
2. 各音に対して、1 で求めた集合中央文字列と、ある被験者から得られた擬音語との正規化レーベンシュタイン距離を求め、その値と集合中央値から標準偏差を求める。集合中央値を MD_{min} 、集合中央文

字列を SMS 、繰り返し数を N 、ある被験者から得られた擬音語を S_i ($i = 1, 2, \dots, N$)、文字列 S_1 と文字列 S_2 との最大文字列長で正規化したレーベンシュタイン距離を $LD_{(S_1, S_2)}$ とすると、標準偏差 σ は以下の式で求まる。

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (LD_{(S_i, SMS)} - MD_{min})^2}$$

個人間変動は、以下の手順で求める。

1. 各音に対して平均的な擬音語を求めるために、擬音語集合から、集合中央値と集合中央文字列を求める。
2. 各音に対して、1 で求めた集合中央文字列と擬音語集合の各擬音語との最大文字列長で正規化したレーベンシュタイン距離と、集合中央値から標準偏差を求める。集合中央値を MD_{min} 、集合中央文字列を SMS 、被験者の数を M 、ある音の擬音語集合の各擬音語を S_i ($i = 1, 2, \dots, M$)、文字列 S_1 と文字列 S_2 との正規化レーベンシュタイン距離を $LD_{(S_1, S_2)}$ とすると、標準偏差 σ は以下の式で求まる。

$$\sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M (LD_{(S_i, SMS)} - MD_{min})^2}$$

実施の際に、1 つの音を何度も受聴することができる点、聴こえに忠実に擬音語にする点、繰り返しがある音の場合は () * の記号で括る点について指示した。これは閉包 (closure) を表す記号である。クリーネ閉包、接続閉包ともいう。例えば、ひよこの鳴き声で「びよびよびよびよ」と聴こえた場合、「(びよ)*」と表記する。同じ音の繰り返しは何回聴こえたか回数を問うものではなく、どのように音を知覚したかを調査するためである。本実験では閉包があった場合、距離計算の際に 1 回聴こえた、として処理した。

表3 刺激音の各要因と水準

	要因	水準数	水準
刺激音	連続性	2	連続音／孤立音
	繰り返し性	2	あり／なし
	音韻の長さ	5	1 モーラ、2 モーラ、3 モーラ、4 モーラ、5 モーラ以上
	言語性	2	なし／習得経験のないその他の言語
	階層性	2	あり／なし

個人間変動, 個人内変動どちらも, 音の連続性, 繰り返し性, 音韻の長さ, 言語性, 階層性の要因について調査した. 表3に, 刺激音の各要因と水準を示す.

3.4 各要因の定義

孤立音は, ひとまとまりの音の印象を与える音と定義する. これは, 音声言語のように一意に定義できるものではない. 特に物を叩く, 打つ, 振るなどの動作の際に発せられる音はひとまとまりに音を感じやすく, 孤立音となりやすい.

逆に, 孤立音が連結し, かつそれにより元の孤立音が変形されて滑らかに繋がる知覚を与えるものを連続音と定義する. 孤立音と連続音の関係は, 音声言語の孤立発声と連続発声との関係と同様である. いずれも実現するためのプランニング上は, 連続系は孤立系と同一であるが, 連続系ではそのプランの実現上調音器官が時々刻々物理的に変形している. 一方環境音の場合は, 実際にその音を発する振動体が時々刻々変化しているかどうかより, 聴感上そのような知覚を与える音を連続音と定義する. 水を攪拌する音は, 水中で手を振る動作が連続的に起こり, 「チャ」「ポ」などの孤立音が繋がっている印象を与えることから, 連続音とする. 電子音のように, 実際に対応する物理振動が存在しない音であっても, 時間とともに変化する知覚を与える音を連続音とする.

繰り返し性は, ある時間区間を元に繰り返しによって全体の音が構成されている音の性質を表す. ドリルの「ドドドド」などがその例である.

音韻の長さは, モーラ数で表す. モーラは, 音韻論上の一定の時間的長さを持った音の文節単位である. 日本語の場合, 基本的に仮名1文字が1モーラである. しかし, 「あ」や「や」などの小文字については, その前の仮名1文字と合わせて1モーラである. また, 音節と異なり, 長音, 促音, 撥音は独立して1モーラである. たとえば, 「こけこっこー」は, 「こ」「け」「こ」「っ」「こ」「ー」の6モーラとする.

階層性は, ある時間区間を切り出すと1つの孤立音に聴こえ, その音が高速に繰り返されて1つの音として知覚することができる音の性質を表す. 例えばタンバリンのトリルの音である.

3.5 実験結果

3.5.1 個人内変動

30代男性1名と20代女性1名を対象に個人内変動の調査した. 表4に個人内変動を分析した優劣表を示す.

3.5.2 個人間変動

表5に個人間変動を分析した優劣表を示す.

表4 個人内変動を分析した優劣表

優			劣		
内容	正規化レーベンシュタイン距離	σ	内容	正規化レーベンシュタイン距離	σ
孤立音	0.33	0.17	連続音	0.38	0.16
連続音かつ繰り返しあり	0.36	0.16	連続音かつ繰り返しなし	0.38	0.16
孤立音かつ繰り返しなし	0.32	0.16	孤立音かつ繰り返しあり	0.35	0.21
無意味音声	0.14	0.06	非音声	0.36	0.16

表5 個人間変動を分析した優劣表

優			劣		
内容	正規化レーベンシュタイン距離	σ	内容	正規化レーベンシュタイン距離	σ
孤立音	0.53	0.11	連続音	0.62	0.11
連続音かつ繰り返しなし	0.60	0.11	連続音かつ繰り返しあり	0.65	0.11
孤立音かつ繰り返しなし	0.50	0.09	孤立音かつ繰り返しあり	0.71	0.02
無意味音声	0.36	0.12	非音声	0.59	0.12

3.6 考察

実験結果から, 以下の6点が考えられる.

- 個人内変動と個人間変動共に, 無意味音声は, 非音声と比べて平均距離が短くなった. このことから, 無意味単語であっても音声は弁別能力が高いことがわかる.
- 個人内変動では, 特に各要因による差異は見られなかった. このことから擬音語の個人内変動は正規分布に従うと仮定して, 非音声は0.04~0.68(95%信頼区間), 無意味音声は0.02~0.26(95%信頼区間)であることがわかる.
- 孤立音は, 連続音と比べて平均距離が短くなったことから, 連続音を擬音語に置き換えることは難しいことがわかる. また, 連続音と孤立音との平均距離は0.09離れていることがわかった.
- 電子音色辞書の登録音は非音声の環境音・楽音を対象としているため, 曖昧検索時の閾値を0.59に設定す

れば50%の確率で、0.71に設定すれば84%の確率で
所望の音が見つかることがわかった。

5. モーラ数では、音の要因を説明することができなかった。このことから、音から擬音語に表記する際に、音韻論的な単位を考えることは無意味であることがわかった。
6. 本実験の計算方法では、繰り返し性のある音の擬音語表記の処理をする際、1回聴こえたとした。その結果、最適な距離計算はできなかった。例えば、「びちよばちよ」と「(びちよ)*」では、繰り返しを1回とするとレーベンシュタイン距離は3となるが、繰り返しを2回とするとレーベンシュタイン距離は1となる。つまり、対応する文字から繰り返し回数を考慮するなど、高度な処理を行うことによって、より距離を短くすることができる。その処理は動的計画法の応用によって実現できるのではないかと考える。

3.7 今後の曖昧検索への指針

擬音語の多義性の問題を扱うために、その擬音語上の距離によりその特性を論じてきた。本来多義的であっても、システム表記上は代表的な一つの擬音語を提示することが電子音色辞書での閲覧上も便利である。そこで、今後は利用人も擬音語の登録を可能とし、システムはこれまでに登録された全擬音語から算出された集合中央文字列を記すことを検討している。曖昧検索は、この文字列との距離により検索を行う予定である。

4. おわりに

本研究では、擬音語の多義性による電子音色辞書での擬音語検索の性能向上のため、レーベンシュタイン距離に基づく擬音語曖昧検索機能を実装した。この機能では検索毎に閾値を手動で指定する必要があるため、自動で閾値を指定するシステムを考えた。そのために、レーベンシュタイン距離に基づく擬音語の知覚特性を調査した。その結果、音声言語は弁別能力が高いこと、連続性のある音を擬音語に置き換えることは難易度が高いこと、モーラ数は擬音語の知覚特性には反映しないこと、閾値を0.59に設定すると50%の確率で、0.71に設定すると84%の確率で所望の音が見つかることを示すことができた。今後の展開として、所望の音の特徴によって曖昧検索の閾値を自動で設定できるシステムを考えていきたい。

参考文献

- [1] 山田祐雅,小坂直敏: C++によるクロスプラットフォーム化した電子音色辞書の構築, 情報処理学会研究報告, Vol.2014, No.57(2014-MUS-103), pp.1-6, 2014.

- [2] Naotoshi Osaka, Yoshinori Saito, Shinya Ishitsuka and Yasuhiro Yoshioka, "An electronic timbre dictionary and 3D timbre display," Proceeding of ICMC 2009, pp. 9-12, Montreal, Aug., 2009.
- [3] V.I. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals", Soviet Physics Doklady, vol. 10, no. 8, pp. 707-710, 1966.
- [4] F. J. Damerau, "A technique for computer detection and correction of spelling errors," Communications of the ACM, vol. 7, pp. 171-176, 1964.
- [5] G. Cormode, M. S. Paterson, S. C. Sahinalp, and U. Vishkin, "Communication complexity of document exchange," 1999.
- [6] G. Cormode and S. Muthukrishnan, "The string edit distance matching problem with moves", In Proceedings of the 13th Annual Symposium on Discrete Algorithms, pp. 667-676, 2002.
- [7] K. A. Majorek, S. Dunin-Horkawicz, K. Steczkiewicz, A. Muszewska, M. Nowotny, K. Ginalski, et al., "The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification," Nucleic acids research, vol. 42, pp. 4160-4179, 2014.
- [8] T. Kohonen, "Median strings," Pattern Recognition Letters, vol. 3, pp. 309-313, 1985.