

直感的な音データ検索/編集システムの開発

和氣早苗 旭敏之
wake@obp.cl.nec.co.jp

NEC 関西 C&C 研究所
〒540 大阪市中心区城見 1-4-24

誰でも簡単に、自分のほしい音を得ることができるツールの開発を目的とし、音検索システムと音編集システムを開発した。音検索システムでは、すでに報告した心理的実験に基づき、擬音語、音源名、主観、という音を表現する3種類のキーワードを用いて検索することとした。そのうちの擬音語検索では、擬音語の表現/表記のゆれに対応するため、擬音語間の類似度を評価する方式を提案する。音編集システムでは、口まねで入力した音声を教師音とし、対象音の振幅およびピッチエンベロープを加工する。

An Intuitive Retrieval and Editing System for Sound Data

Sanae Wake Toshiyuki Asahi

NEC Kansai C&C Research Labs.
1-4-24, Shiromi, Chuo-ku, Osaka 540, Japan

A sound retrieval system and a sound editing system were developed in order that anybody can get one's favorite sound easily. The sound retrieval system provides three types of retrieval keywords; onomatopoeias, sound source names and adjectives. A technique to cope with the diversity of onomatopoeia expression is proposed, that measures the similarities between onomatopoeias. The sound editing system uses voice-sound as input, and its amplitude-envelope and pitch-envelope are used as parameters to edit the sound.

1 はじめに

昨今、ソフトウェアや携帯機器、ホームページ等様々な場において音が利用されている。このような中、“与えられた音”をそのまま使うだけでなく、ユーザ自身が独自の音を創作したり使用したりすることへのニーズも高まりつつある。

我々は、誰でも簡単に音を扱うことができるツールの開発を目指している。イラスト集から絵を選び、目的に合わせて大きさや構成に変更を加え、プレゼンテーション等に用いることはもはや通常になっているが、それと同様な手軽さで、音を扱う環境の構築が目標である。今回、その一環として、音検索システムおよび音編集システムを構築した。音データベースの中から最も自分のイメージに近い音を検索し、その音に対して直感的な編集を行う、という利用形態である。

2 検索システム

本章では、音の3つの側面を利用した音検索システムについて紹介する。

近年、音データの供給も多くなっている。しかし、データ集から音を選ぶには、音の題名や説明文などを手がかりとして、一つ一つ音を聞きながら音を探さなければならないのが現状である。音は複数のデータを一度に聴取することができないため、データ量が多くなるに伴いこの作業も多大となることが問題である。こういった問題に対して、音響の特徴をキーとした音響による音データ検索も提案されている^[1]。我々はユーザにとって容易で、直感的な入力手法を明らかにするためまず、“人間は頭の中でイメージした音をどのように表現するのか”を調査する実験を行い、その結果を反映して音検索システムを開発した。

2-1 音の表現方法に関する実験

音を相手に説明するという状況において、人間はどのような表現を用いるのかを観察した^[2]。

2-1-1 音の分類とサンプル音の選出

実験に用いるサンプル音の選出にあたり、音を音楽、音声、環境音、という3つに分類、さらにそれぞれを符号音と副次音に分けた。符号音とは

“意図をもって発せられる音”、副次音とは“ある事象の結果生じる音”であるとした。これら6カテゴリーから17種類のサンプル音を選出した。

2-1-2 実験の方法

実験は14名(7ペア)の被験者に対し行った。被験者Aはサンプル音聴取後、被験者Bに対して聞いた音を説明する。被験者Bは説明を聞いて音を想像する。途中必要に応じて質問をしてもよい。

2-1-3 音表現の3要素

実験の結果から、音の説明に用いられた言葉は次の3タイプに分類できた。

(1) 波形の説明

“聞こえ”そのものを表現しようとするもので、“物理量の説明(音の長さや高さ)”と“擬音語”に分けられる。このうち“物理量の説明”は楽音など高さや長さを表現しやすい音に対しての利用が多かったが、“擬音語”は音の種類を問わず全般に対して用いられた。

(2) 音源の説明

発音源や発音状況に関しての説明。何の音か、という説明の他、いつ(時間)発せられるか、どこで発せられるか等の説明も見られた。

(3) 主観の表現

感想を表現するもの。形容詞や～的といった表現が用いられる。特に非現実な音に対して多く用いられた。

以上、説明の方法は“音源の説明”が最も多く、それが困難な音に関しては“波形の説明”や“主観の表現”が用いられた。これらの利用頻度は約7:2:1であった。

以上の結果を基に、音検索システムの検索キーとして、擬音語、音源名、主観、の3種類を利用することとした。(波形の説明における“物理量の説明”は音の種類が限定されるため採用しなかった。)

2-2 音データベース

音データベースには自然音、生活音、機械音、合成音など814音を蓄積し、個々の音に、以下に示す、擬音語、音源名、主観に関するキーワードおよび数値をあらかじめラベルとして付加した。

(1) 擬音語

慣用的な擬音語（ワンワン等）に加え、音を比較的忠実に仮名に置き換えた擬音語（ウーウォン等）の双方を登録した。1つの音に対して擬音語は唯一決まるものではないため、最大8種類まで登録可能とした。

(2) 発音源

発音に作用した音源名であるが、録音時に実際に使われたものに加え、音を聞いて類推される発音源も登録する。例えば、実際はある虫の声でも鳥の様にも聞こえれば鳥も登録する。最大3種類まで登録可能とした。

(3) 主観

予め10種類の主観ワードが選定されており、これらに対して数値がつけられている。この数値は予め行われたデータ収集の結果であり「何割の人が音を聞いてこの主観ワードが当てはまると感じたか」を示す数値である。例えば、表1に示す例では、SIND-015.wavを聞いて“迫力のある”と感じた人が5割いることを示す。主観ワードの選出、およびデータ収集に関しては、2-3-3項にて詳細を説明する。

また音はタイトルも持つ。ユーザに対する音の提示はタイトルで行う。以上で説明したラベルの例を表1に示す。これらのラベル付け作業は、手作業で行った。なお、音ファイルはwavファイル形式で保存されている。

2-3 検索の方法

検索システムのユーザインタフェースを図1に示す。ユーザは、擬音語、音源名、主観、のうち

ファイル名	SIND-015.wav	
タイトル	呼ばれて飛び出てジャジャジャーン！	
擬音語	バランラン、ビャンビャン、ガァンワンワン バロンロン、ガランワンワン、ビヤランラン	
発音源	鉄筋、ピアノ、弦楽器	
主観	うるさい … 3.33	迫力のある … 5.00
	金属性の … 5.00	かたい … 3.33
	にこった … 1.67	陽気な … 1.67
	美しい … 0	にぶい … 1.67
	ものたりない … 0	おだやかな … 0

表1. 音データのラベル例

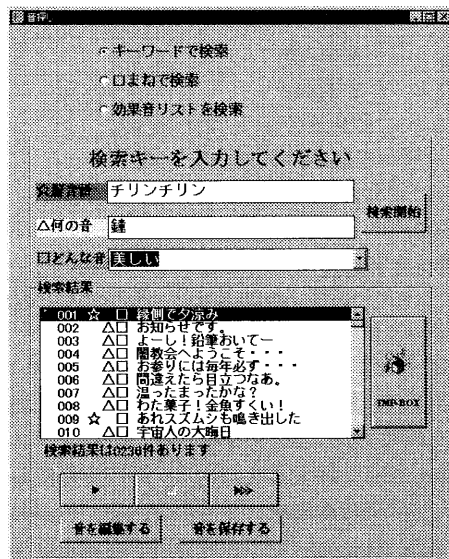


図1. 検索システム

少なくとも1種の検索キーを入力する。システムはそれぞれの検索キーでデータベースに対して検索を行うが、検索キーとラベルの一致度に応じて各音データに対し検索得点を与える。検索方法、検索得点については次項より示す。それぞれの検索における得点の合計が検索の総合得点となり、総合得点の高い音データから順にユーザに提示される。

2-3-1 擬音語検索

2-3-1-1 擬音語

擬音語は音の“聞こえ”を表現する一手法であり、日本語の音韻体系に合致した言葉である^[3]。ところが、擬音語のバリエーションは非常に多様である。同じ音を表すにも個人差や曖昧さから何種類もの文字列で表現することが可能である。逆に、文字列としては異なっても似た音を想起できるものは数多くある（例：パチャン、ピション等）。

このような擬音語のバリエーションに対応するためには、検索キーとして入力される擬音語とデータベースのラベルに記述されている擬音語との間で単純なキーワード一致検索を行うだけでは不十分である。今回擬音語の特徴を考慮し、バリエーションを自動的に作り出し、それに基づいて検索を行う類似擬音語検索方式を提案する。この方法では擬音語の類似度に応じて得点が決定される。

条件	例	得点
検索キーと完全一致	カラン	10
検索キーが部分一致	カランコロ	9
検索キーのある1文字(但し、1文字目以外)を削除した残りの文字列が部分一致	カラリ カン、 カンコロ	6
検索キーの1文字目を削除した残りの文字列が部分一致	バラ	4

表2. マイナスワン検索(例は入力が“カラン”のとき)

2-3-1-2 類似擬音語検索方式

まず、検索の基本となるマイナスワン検索について説明する。この方法は、検索キーでラベルを検索した後、検索キーのある一文字を削除してできた文字列で検索を行う方法である。検索の方法と得点、検索キーが“カラン”であったときの例を表2に示す。

このマイナスワン検索だけで検索を行った場合、幾つかの問題点が生じる。その問題を回避するため、また検索の精度を上げるため、実際は次に示す擬音語の特徴を反映して検索、得点付けを行う。

(1) 短擬音

2文字以下の検索キーに対してマイナスワン検索を適用すると、ある仮名1文字をラベルに含むデータが全て検索されてしまうため、これらについては、文字数を条件に入れた検索を行う。例えば検索キーが2文字であった場合(例:ゴン)完全一致検索の後、いずれかの仮名を含む2文字のラベル(ゴツ、ガン等)、また2文字を含む3文字のラベル(ゴツン等)などを検索ヒットとする。

(2) 小文字擬音

擬音語で用いられる小文字は通常の仮名に比べて音響的に曖昧さが大きいと言える。例えば、ピチャとカチャでは違う音を想起させるのに対し、ピチャ、ピチュ、ピチョ、からは音響的に似通った音を想起するであろう。マイナスワン検索のルールのみでは小文字も通常の仮名と同様に一文字として処理されてしまう。今回先に述べた小文字と通常の文字の音響的影響の差を考慮し、小文字の仮名が異なってもヒットとする条件を加えたマイナスワン検索を行う。

(3) 繰り返し擬音

タラララ、パラパラなど同じ文字または文字列

が複数回繰り返されると、文字列としての繰り返し回数に依存せずに検索を行う。まず、擬音語中の繰り返し部分を判定し、ある文字または文字列が2回以上繰り返されることを示す繰り返し記号を用いた形に文字列を変換してから(例:タラララ→タ(ラ)*)、マイナスワン検索を行う。

以上の特徴およびそれらの組み合わせから、実際には合計26の条件によって11段階(0点～10点)の得点が個々の音データに与えられる。

2-3-2 音源名検索

2つ目の検索である音源名検索は、単純なキーワード一致検索である。検索キーとして入力された文字列がデータベースの音源名ラベル部分に存在すれば10点、存在しなければ0点が音データに対して与えられる。

2-3-3 主観検索

主観検索の検索キーは選択性である。予め用意された10種類の主観ワードからユーザは適するものを選択する。データベースにおいて個々の音データは、それぞれの主観ワードに対して0～10の得点を保持する。この得点は“何割の人が音を聞いてその印象を持つか”を示す値である。

2-3-3-1 主観ワードの選定

音に対して用いられる形容詞について、その印象の類似性を分析した難波ら^[4]の研究結果を参考にし、10種類の形容詞を選定した(表1の主観欄に記載)。

2-3-3-2 主観値データ収集

10代～30代の男女6名に対し次の方法でデータ収集を行った。

- 音データ全814音の個々について、10種類の主観ワードそれぞれの印象が当てはまるかどうかを答える。

全データ収集後、個々の音のそれぞれの主観ワードについて当てはまると答えたものが何割いるかを算出し、その値を主観値としてデータベースのラベルに記載した。

2-3-3-3 主観値を用いた検索

ユーザは10種類の主観ワードから1つを検索キーとして選択する。検索において、データベースラベルのその主観ワードに対応する主観値が音

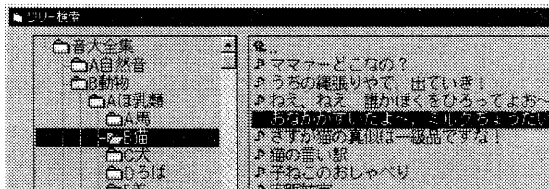


図2. ツリー検索

データの検索得点となる。すなわち、多くの人が検索キーの印象をもつ音ほど高得点が与えられる。

2-3-4 その他の検索機能

2-3-4-1 口まね擬音検索

検索キーの入力方法として、音声による擬音語入力もサポートする。音声認識装置で音声を文字列に変換することで、音声入力を可能としている。

2-3-4-2 ツリー検索

検索機能だけではデータベース全体が把握しにくい欠点があるため^[5]、音データを音源にて大分類、中分類、小分類に分類し、ツリー構造でタイトルを表示するインタフェースも用意した(図2)。

3 編集システム

本章では、直感的な編集の方法として、音(音声)による音の編集機能を備える音編集システムについて述べる。

3-1 直感的編集

音の編集はこれまでにシンセサイザー等で実現されてきたが、多くの音響パラメータの指定が必要であり、思い通りの音を制作するには音響的知識と熟練を必要とした。音響パラメータを直接指定しない方法としては、人工生物学的アプローチをとったもの^[6]などが提案されている。我々は、特に音の大きさと高さの変化に注目することとした。ある音色を持つ編集対象音に対して、口まね音声を入力としてエンベロープ編集を行う。口まねは直感的であり、さらに大きさと高さを同時に入力できる方法だと考える。ユーザは口まね音声で大きさの変化と高さの変化を指示するが、音の長さやリズムも大きさの変化として同様に入力できる。なお、口まね音声の代わりに別の音や音ファイルを入力とすることもできる。また教師エンベロープを手書き描画する機能も持つ。

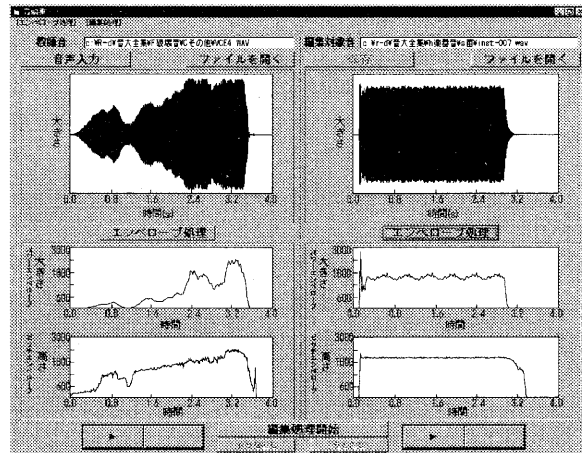


図3. 編集システム

音編集システムのユーザインタフェースを図3に示す。編集対象音が右に、教師音(編集の指針となる音、音声など)が左に図示されている。上のグラフは原波形、中央はパワーエンベロープ(パワーの包絡線)、下はピッチエンベロープである。教師音の振幅エンベロープおよびピッチエンベロープを鋳型として、編集対象音のそれを加工する。(編集には振幅を利用するが、表示にはより直観的なパワーエンベロープを用いている。)

3-2 振幅編集

教師音の振幅エンベロープを鋳型として編集対象音の振幅エンベロープを加工する。相似編集方式と強制編集方式の2種類の編集方式を提供する。

3-2-1 振幅エンベロープの抽出

フレーム長32ms、フレーム周期10msでフレーム内の各サンプル点から平均振幅を算出する。

3-2-2 振幅エンベロープ編集

(1) 相似編集方式

編集対象音のエンベロープの特徴を残しながら加工を行う方法で、教師音の振幅エンベロープに従って編集対象音のボリュームを上げ下げするような効果が得られる。

○編集対象音の波形データを $f(t)$ 、教師音の振幅エンベロープを $x(t)$ 、編集後の波形を $F(t)$ としたとき、それらの関係は次式で表される。

$$F(t) = x(t) \cdot f(t)$$

(2) 強制編集方式

編集対象音のエンベロープが教師音のエンベロープと同形になるように加工する。

○教師音、編集対象音の両振幅エンベロープを比較し、倍率 $(m(t))$ を算出する。

○編集対象音に倍率 $(m(t))$ をかけ、波形を加工する。

$$F(t) = m(t) \cdot f(t)$$

○編集対象音の振幅エンベロープが閾値以下の区間（無音区間）は、無音区間と同じ長さだけさかのぼった波形を無音区間に補完してから上記の処理を行う。

3-3 ピッチ編集

教師音のピッチエンベロープを鋳型として、編集対象音のピッチエンベロープを加工する。編集の方式として、BPF(バンドパスフィルタ)方式とFFT方式と波形伸縮方式の3種類を提供する。

3-3-1 ピッチエンベロープの抽出

エンベロープ抽出の対象が音声の場合は、LPC残差の自己相関分析を、対象が効果音の場合は、零交叉分析を用いてピッチ抽出を行う。

3-3-2 ピッチエンベロープ編集

(1) BPF方式

バンドパスフィルタを用いて、編集対象音をイコライジングする。通過周波数は教師音のピッチエンベロープより求め、処理は単位時間毎に区切って行う。ノイズ系の音に利用できる。

(2) FFT方式

フーリエ変換で周波数軸上に展開された波形を伸縮し、逆フーリエ変換で時間軸上に戻す。処理は単位時間毎に区切って行う。伸縮率は教師音、編集対象音の両ピッチエンベロープより算出する。

(3) 波形伸縮方式

時間軸上で波形をそのまま伸縮する。時間的な過不足は削除/補完で対応する。処理は単位時間毎に区切って行う。伸縮率は教師音、編集対象音の両ピッチエンベロープより算出する。

4. まとめ

誰もが簡単に音を探し、音を編集することのできる環境の提供を目的とし、音検索システムと音編集システムを開発した。

音検索システムは、擬音語、音源名、主観という音の3つの側面からの検索を可能とした。この事は、音響的に似ている音、音源が同じ音、印象が似ている音、という3点からの音のグルーピングを行ったとも言え換えられ、ユーザの音検索の目的に応じて効率のよい検索が可能になる。しかしデータベースの構築に多大の作業を要するなど、まだ解決すべき問題も残っている。

音編集システムは、音による音の編集ということで、口まね音声を目指したり、音Aを音Bのように変更、などという編集の指定が可能になる。音響的に特別な知識がなくとも使えるという点、同じエンベロープを持つ音を容易に制作できるという点、タイミングや長さの指定がしやすいので、画像の動きにあった音を作りやすいなどの長所があると考えている。現在はまだ試作の段階であるが、今後評価を行い、有効性を実証して行く。

【参考文献】

- [1] 斉,村松,橋本,"音響データベースのマルチモーダル検索の試み-音響によるデータ検索-",情報処理学会研究報告 96-MUS-16, pp51-56, 1996
- [2] 和氣他"効果音検索システム:音の表現方法に関する実験と考察",第48回情報処理学会全国大会論文集, 5P-1, 1994
- [3] 田守育啓, 日本語オノマトペの研究, 神戸商科大学経済研究所 pp1-2
- [4] S.Namba, S.Kuwano et al., "Verbal expression of emotional impression of sound: A cross-cultural study", J Acoust Soc Jpn E, Vol.12, No.1 pp19-29, 1991
- [5] 和氣, 旭, "擬音語による効果音データの検索", 第52回情報処理学会全国大会論文集, 5D-1, 1996
- [6] A. Horner, J.Beauchamp et al., "Timbre Breeding", Proc. of ICMC'93, pp-396-397, 1993.