

生活環境音を記録し音響的特徴を用いて要約する インタフェースの提案

大塚 昭徳[†] 伊丹 徳重[‡] 坂倉 美保[‡] 富塚 清史[‡] 大川 茂樹^{†,‡}

千葉工業大学 [†]大学院情報科学専攻 [‡]情報ネットワーク学科

1. はじめに

近年、コンピュータ技術や移動体通信、インターネット等に代表されるネットワーク技術の発達に伴い、時間と場所を選ばずとも、音声や動画などの大容量のメディア情報を効率的に利用できるようになった。しかし、膨大な情報の中から必要な情報のみを検索して利用するためには、高度なパターン認識処理や知識処理が不可欠である^[1]。

音声認識や画像認識の技術が発展し、最近ではかなり大容量で複雑な音声や画像に対してもある程度の認識性能を達成できるようになったことから、さらなる応用システムとして、例えば長時間の音声や動画から重要な部分を抽出（要約）する取り組みが多く行われている。対象を音声に限っても、データベースへのアクセスを簡便にするためのインデキシング、会議録の自動生成や放送番組のダイジェスト生成などの目的での「音声要約」の研究が盛んに行われて始めている^{[2][3]}。

我々の研究室でも、従来、音声認識に関する研究を進めてきたが、本研究では対象を音声から「生活環境音」に広げた場合の要約原理を検討し、新たな音インタフェースを提案する。図1に本インタフェースの概念を示す。

音声では言語情報が要約の重要な手がかりとなるが、対象が一般の音データになると言語的単位の定義が難しく、音響的特徴が重要な手がかりとなる。したがって、本研究ではまず、生活環境音に含まれるどのような情報が抽出できれば要約できたことになるのかを検討する。次

に、どのような音響的特徴が効果的な生活環境音要約に役立つかを検討し、最後に、実際の音データに対する要約シミュレーションを行う。

2. 生活環境音を要約するインタフェース

本研究が対象とする「生活環境音」には、我々の身近に存在するあらゆる音情報が含まれる。音声や音楽のようにそれ自体が構造的な情報を持つ音もあれば、機械や乗り物の音や自然界の音など、聞く立場によって情報の重要度が変化する（ノイズとなる）音もある。したがって、要約の目的は要約する側の立場や見方によって様々であるともいえる。以下にこのようなインタフェースの具体例をいくつか示す。

- ① 生活を「記憶」する音インタフェース：
我々の日常体験を大容量のハードディスク等に音で記録しておき、人間の補助記憶手段として利用する研究^[4]に応用する。
- ② 記憶障害者への支援：
肢体障害や知覚障害に比べ、記憶障害や認知症への工学的支援はあまり行われていない。①と同じ見地から、記憶障害者の行動を音で記録し、行動の検索等に利用する。
- ③ スポーツなどのハイライト生成：
スポーツ中継のアナウンス音声や観客の歓声からハイライトシーンのダイジェストを作成する。
- ④ 放送番組一覧の自動生成：
TV やラジオ（音のみに着目する場合は特にラジオ）の放送音源から、放送内容のカテゴリ分類を自動的に行う。

その他、音の特性を生かした応用例として、夜間など画像が用いにくい状況での防犯や、動物の活動記録、作業音（工場内の音・調理音など）からの状況記録などが挙げられる。

3. 生活環境音の分類

あらかじめ様々な生活環境音を収録し、含まれる音のカテゴリ分類を行う。音声における単語や音素のように、一般的に定義可能なラベル情報が存在しないため、人間が音を聞いたとき

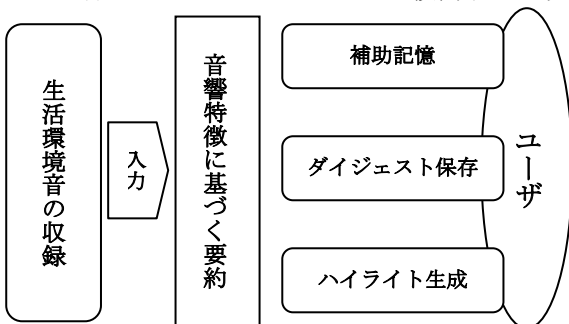


図1：生活環境音要約インタフェースの概念

の主観に基づき階層的な分類を行った。その結果の一部を図2に示す。カテゴリは5階層とし、最上位層2種類（人工音および自然音）、最下位層136種類の音カテゴリを定義した。

LAYER1	LAYER2	LAYER3	LAYER4	LAYER5
artificial	machine	vehicle	train	door run engine brake
			car	door run engine brake
			bus	door run
.
.
.
natural	animal	human	action	step hand write eat trun cough
			voice	laugh voice
.
.
.

図2：生活環境音の分類（部分）

4. 音響的特徴を用いた音要約

次に、実際に収録した生活環境音に対して音響特徴量を計算し、音要約を行う。本研究で用いた特徴量は、(1)音信号のパワー値、(2)20次の自己相関係数、(3)スペクトログラム中から切り出した時間一周波数セグメントのマトリクス量子化コードの3種類である。

(1)は一般的な音区間の同定に、(2)は周期性を持つ音の抽出に、(3)は音声や音楽のような周波数構造を持つ音の分類に有用と考えている。

5. 要約のシミュレーション

まず、生活環境音として、ある学生が自宅から大学まで通学する際の音をデジタルデータとして約60分間収録し、それを4名の被験者に聞いてもらい先に定義した音カテゴリ分類に従って各時刻（ここでは10秒単位）へのラベル付けを行った。これらを整形してカテゴリラベル付き元データとする。次に、各ラベルと対応する音響特徴量を計算した。

現時点では、全ての音カテゴリの自動的な抽出識別は非常に困難であることから、先に定義した階層的音カテゴリの一部について、手動操作で特徴量との対応付けを行い、以下のシミュレーションに利用した。

約350秒間の音データについて、音のパワー値及び音カテゴリラベルを図示したタイムチャートを図3に示す。図中網掛け部分が、この場

合の「理想的な」音要約結果となる。

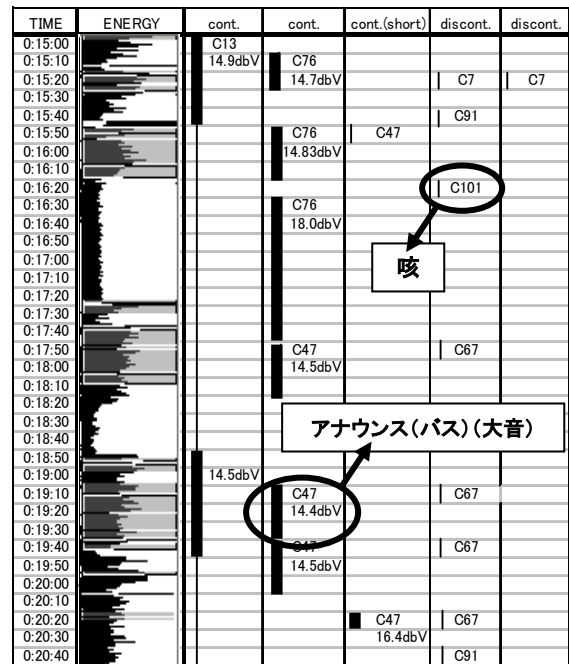


図3：音要約のシミュレーション結果（例）

6. 考察・まとめ

本研究では、様々な生活環境音を含む長時間の音データに対して、音響的特徴を用いて要約を行う新しいインタフェースを提案した。現状では、あらゆる種類の音の識別は困難であり、また、図3からも見ることができるよう、人はロンバート効果のように周囲の音によって認知感覚がずれてしまう。このことから、音カテゴリの持つ情報をいかに定量的に評価するかが、このインタフェースを有用にする上でのポイントになると思われる。

今後は、さらに多くの音データを用いて、実際の音要約を実現するための特徴抽出、カテゴリ識別、要約処理技術の検討を進めたい。

文 献

- [1] 要約コンテンツ自動生成・配信機能を備えた映像・音楽ストリーミング配信ソリューション,
<http://www.ntt-east.co.jp/release/0406/040624a.html>
- [2] 相沢 他, “ウェアラブル映像の構造化と要約:個人の主観を考慮した要約生成の試み,” 信学論, Vol.J86-D-II, No.6, pp.807-815, (2003)
- [3] 堀, 古井, “音声要約技術の現状とこれから”, 信学技報, SP2003-171, pp.7-12 (2004)
- [4] MYCOM PC WEB, “Pervasive 2004 から見えてくるユビキタス未来像,”
<http://pcweb.mycom.co.jp/articles/2004/06/30/pervasive/>