環境音の擬音語変換における音素決定曖昧性の解消

石原 一志[†] 服部 佑哉[‡] 中谷 智広[‡‡] 尾形 哲也[†] 奥乃 博[†]
[†]京都大学大学院情報学研究科 [‡]京都大学工学部情報学科 [‡‡]NTT コミュニケーション科学基礎研究所

1. はじめに

本研究の目指す環境音の擬音語変換は, 環境音を擬音語というシンボルに結びつけることで, 様々な音を人間・計算機共有の情報として扱えるようにすることが目的である. たとえば, 擬音語変換を通じた環境音データの高度なアーカイブ化や検索の実現, コミュニケーション用インタフェースの対話戦略への活用などが期待できる. このような信号・記号変換は, 最近始まった人生の総アーカイブ化プロジェクトである LifeLog[1)]でも重要な課題である.

環境音の擬音語変換を行う上での最大の問題点は,『環境音を表す擬音語表現の音素は聴取者や状況に依存して一意に定まらない』という問題である. たとえば, 同じ猫の鳴き声であっても「にゃーん」と聞く人も「みゃーん」と聞く人もいる. このため環境音を適切に表す擬音語ラベルを記述することは難しい. そこで, 単一の擬音語ではなく, 複数の擬音語候補からなる擬音語集合をラベルとして付与する. 上記の例では,「にゃーん」と「みゃーん」からなる擬音語集合をラベルとする. 適切な擬音語集合とは,『聴取者の表現する様々な擬音語のうち代表的なものを含み (再現性), かつ, どの聴取者の擬音語とも一致しない擬音語は含まない (適合性)』集合である. このような集合をラベルとして学習・認識を行うことで, システムは複数の擬音語候補から各ユーザの聞こえ方に合った適切な擬音語を選択することが可能となる.

しかし, 通常の音素でこのような擬音語集合のラベルを表現することはできない. そこで, 環境音のための特殊な音素集合を従来の日本語の音素を利用して設計する. たとえば, /p/とも/f/とも聞こえる音に対して/pf/という新音素を定義する, などである. この音素集合は, 日本語の音素を用いて設計してあるため日本語音素への変換は容易である. 区別のため, 音声言語の音素を「音声素」, 環境音の音素を「環境音素」と呼ぶ. 曖昧性が強い環境音素は, 多くの音声素へマッピングを持ち, 多数の擬音語を生成する.

本稿では, 曖昧性の強さが異なる 4 種類の環境音素集合を提案し, 評価実験によりどの環境音素が人間の聞き方と一致しているかを調べる. なお, 簡便のために対象音は単発音に限定し, 音素決定処理のみを扱う (図 1). 複雑な環境音であっても服部ら[2)]の手法により単発音に切り分けることができる.

2. 音素決定システムの概要

環境音の音響データと環境音素で構成するラベルを用いて HMM 学習を行い, 認識結果にもとづいて複数の擬音語候補を出力する. 音響ファイルは RWCP の環境音データベース内の単音節音 6011 サンプルを用いる. 特徴量には 16 次元の MFCC とパワー, それらの一次微分からなる 34 次元特徴量を採用した. フレーム長とフレーム間隔はそれぞれ 50ms/10ms である. 学習の混合数は 16 とした. MFCC のメル周波数やフィルタバンク分析は人間の聴覚特性に合わせて設計してあるため, 本実験の特徴量としても適切であると考える. 実際, 予備実験において, HMM は同定木を用いた手法などよりも高精度で, MFCC は LPC や FBANK[3)]などの特徴量よりも高精度で認識を行うことを確認している. これらのパラメータは, さまざまな条件下で行った予備実験 (全 396 条件) にもとづいて決定した.

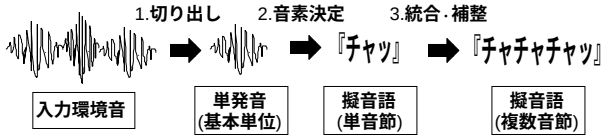

図 1 擬音語変換の処理過程

表 1 曖昧性の異なる 4 種類の環境音素 (PG: 音声素グループ)

表現	音素集合	排他性	認識例	認識例 2
$\mathcal{A}$	既定	不可	u-exp ao N	fric u: Q
$\mathcal{B}$	既定	可能	k-t ao N	s-sh u: Q
$\mathcal{C}$	動的決定	—	k o N, t a N	sh u: Q
$\mathcal{D}$	単一音素	—	k a N	sh u: Q

3. 曖昧性許容度が異なる環境音素集合

本章では曖昧性の強さが異なる 4 種類の環境音素の設計方針を述べる (表 1). 音声素をグループ化した『$\mathcal{A}$:音声素グループ』, $\mathcal{A}$ よりも曖昧性が少ない『$\mathcal{B}$:細分化音声素グループ』と『$\mathcal{C}$:複数音声素』, そして, 音声素をそのまま環境音素として利用する『$\mathcal{D}$:音声素』である. $\mathcal{A}$, $\mathcal{B}$ では環境音素を構成する音声素の組み合わせは既定だが, $\mathcal{C}$ は動的に決定する. $\mathcal{D}$ は音声素そのものである. また, $\mathcal{B}$ は音声素が複数グループに属することを認めるが, $\mathcal{A}$ は許可しない (排他性).

3.1 重複学習

HMM 学習を行う際に, 環境音素で完全に表現できない曖昧性に対しては**重複学習**を行うことで解決を図る. 重複学習とは, 一つの音響ファイルに対して複数のラベルを正解として学習することを指す. たとえば,「たっ」とも「たん」とも聞こえる音があり, 促音と撥音の両方を含む環境音素が無い場合には, 両方のラベルで学習を行う. 重複学習は $\mathcal{A}$ と $\mathcal{D}$ の学習, および $\mathcal{B}$ の一部の例外の学習に適用する.

3.2 環境音素 $\mathcal{A}$: 音声素グループの設計

人が混用しやすい音声素ごとにグループ化を行い, 環境音素として利用する. このグループを**音声素グループ**と呼ぶ. 表 2 は, 聴取実験において同じ音を表すのによく用いられた子音同士を同じ音声素グループとしてまとめて分類したものである[4)]. この分類は摩擦音・鼻音などといった調音方法にもとづく音声素分類とほぼ一致している. これにより, 音声素グループは発音状況にもとづく環境音素であるといえる.

音声素グループは 4 種類の環境音素の中で最も曖昧性の許容度が大きいため, 再現性が高く適合性が低い. そのためアーカイブのタグとしての利用には向いているが, 擬音語に変換する上では不適切な音声素へのマッピングが多く現れる.

Disambiguation in Phoneme Determination of Sound-Imitation Words for Environmental Sound Recognition *by Kazushi Ishihara, Yuya Hattori (Kyoto Univ.), Tomohiro Nakatani (NTT), Tetsuya Ogata and Hiroshi G. Okuno (Kyoto Univ.)*

表 2 聴取実験にもとづいた音声素グループ分類 [$\mathcal{A}$]

1. m,n,my,ny (“nasal” (鼻音)) 2. s,sh,j,z (“fric” (摩擦音))
3. h,f,hy (“hf” (摩擦音)) 4. w,y (“semiv” (半母音))
5. b,d,g,by,dy,gy (“v-exp” (有声破裂音))
6. p,t,k,py,ts,ch,py,r (“u-exp” (無声破裂音, 流音))

表 3 細分化音声素グループ [$\mathcal{B}$] (子音は出現頻度順)

t, k-t, b, p, t-ch, f-p, t-p, z-j, k, g, r, k-p, k-t-ch, b-d,
j, t-ts, ts-ch, s-sh, d-g, b-d-g, w, sh-j, k-t-r, k-g, t-d, ch,
sh, ao, a, i, u, e, o, ao:, a:, i:, u:, e:, o:, N, Q, QN

表 4 評価実験結果 (%) [PG: 音声素グループ]

評価	再現率	適合率	評点
$\mathcal{A}$: PG	81/140(57.9)	27/104(26.0)	—
$\mathcal{B}$: 細分化 PG	79/140(56.4)	26/36(72.2)	3.48
$\mathcal{C}$: 複数音声素	76/140(54.3)	22/26(84.6)	3.66
$\mathcal{D}$: 音声素	56/140(40.0)	17/22(77.3)	3.57

($\mathcal{A}$ は生成する候補が非常に多いため評点の評価は行っていない)

3.3 環境音素 $\mathcal{B}$: 音声素グループの細分化

各音声素が複数のグループに属することを認めることで, $\mathcal{A}$ とは異なる音声素グループを設計する. たとえば, /k/とも/t/とも聞こうるが, それ以外の音声素としては聞えない子音に対して/k-t/というラベルを与える. $\mathcal{A}$ の音声素グループと区別してこれを**細分化音声素グループ**と呼ぶ. 本稿で使用した音響データ中に現れた細分化音素グループを表3に示す. 細分化音声素グループは, $\mathcal{A}$ よりも曖昧性が少ないため, 音声素へマッピングした時に不適切な音声素が現れにくい.

その一方で, 音声素グループを細分化することでクラス数が増大し, 学習サンプル不足のクラスが頻出する. 本研究では, 学習サンプルが充分にあるクラスに対してはそのまま学習を行い, 学習サンプルが少ないクラスに対してはクラスを構成する音素でそれぞれ重複学習を行うことでこの問題を解決した. たとえば, サンプル数が少ない/p-w/は, /p-w/として学習を行わず/p/と/w/のラベルでそれぞれ学習を行う.

3.4 環境音素 $\mathcal{C}$: 音声素の階層的絞り込み

音声素グループ認識を行い, 認識結果であるグループに属する全音声素に対して「対象音を表す音声素として妥当」であるかを判定する. そして, 妥当とした全ての音声素を用いて擬音語を生成する. この手法は音声素グループのようにあらかじめ音声素の集合が設計してあるわけではなく, 音声素の組み合わせは毎回動的に決定する. 妥当性の判定は音声素HMM の尤度を用いた閾値判定により行う.

この手法は音声素グループを認識することで発音状況を特定してから音声素を絞り込む手法と考えることもでき, 一種の階層的手法であるといえる. そのため前段階である音声素グループ認識での誤りが全体の精度に大きく影響する. 一方で, 音声素グループ認識の精度が充分に高ければ, 絞り込み段階で誤ってもそれほど不適切な音声素は選択されない.

3.5 環境音素 $\mathcal{D}$: 音声素レベルでの重複学習

音声素をそのまま環境音素として利用する. これは曖昧性を持たない環境音素であり, 他の環境音素との比較として用いる. 生成する擬音語は一つであるため, ユーザの聞こえ方に合わせて擬音語を選択することはできない. なお, 学習における曖昧性は重複学習を行うことで解決した.

4. 評 価 実 験

20 種類の環境音とその認識結果を用いて, 7 名の被験者に対して評価実験を行った. 被験者は対象音を聞き, 自分の聞こえ方に従って擬音語 (回答) を記述する. 記述する擬音語は複数でも良い. 次に, 3 節の認識器を用いて生成した擬音語候補 (認識結果) に対して 1 点 [不適切] から 5 点 [適切] の評点を与える. 再現性は, 認識結果と一致する回答を持つサンプルの割合で評価する. 適合性は, 聴取者の回答と一致する擬音語を持つ認識結果の割合で評価する. これらの評価は, 1 章で述べた適切な擬音語集合の評価から設計した.

評価実験の結果を表 4 に示す. 擬音語を一つしか生成しない $\mathcal{D}$ は再現率が低く全体の 4 割程度である. 一方で, 曖昧性を許容する $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ の環境音素は, $\mathcal{D}$ よりも 15%程高い再現率を示している. 逆に, 曖昧性の度合いが強い $\mathcal{A}$ は不適切な擬音語を多く生成するため適合率が低い. 全体としては, 再現率・適合率が共に高く, さらに被験者の評点平均が最も高い『$\mathcal{C}$: 複数音素』が, 最も人の聞こえ方に一致した環境音素であると考える.

本実験のような主観にもとづく評価は聴取者に依存して変動するのが普通である. 実際, 評価実験において, 1 点 (不適切) の評価を受けながら, 他の被験者からは 5 点 (適切) の評価を受けたサンプルは全体の約 3 割程を占めていた. そこで, 以下の URL で本評価実験に用いたデータと認識結果, 聴取者の回答を公開する. (http://winnie.kyoto-u.ac.jp/~ishihara/onomatopoeia.html)

5. おわりに

本稿は, 環境音の擬音語変換における音素決定の曖昧性問題に対し, 4 種類の環境音素を設計して解決を図った. 評価実験によりそれぞれの表現のもつ曖昧性と妥当性について, 人間の聞き方と比較・検討を行った. その結果, 音声素よりも曖昧性があり, 音声素グループよりも曖昧性が少ない『$\mathcal{C}$: 複数音素』の表現が, 人間の許容する曖昧性に近いことが分かった.

今後の課題として, 環境音データの不足の問題や, 複数の擬音語候補からユーザの聞こえ方に合った音声素選択処理の設計などがある. これらの問題を統合的に解決するために発達論的コミュニケーションシステムを設計する.

本研究の一部は科研費と 21 世紀 COE, および NTT との共同研究の支援を受けた. また, 研究に関して御助言をくださった NTT の南泰浩氏, 中村篤氏, 京都大学学術メディアセンターの坪田康氏に深く感謝する. また, 評価実験に協力くださった, 北原氏を始奥乃研究室の各位に感謝する.

参 考 文 献

1) LifeLog, http://www.darpa.mil/ipto/program/lifelog/
2) 服部 他: 連続環境音の繰り返し構造の認識, 第 66 回情報処理学会全国大会論文集, 2004.
3) HTKBOOK, http://htk.eng.cam.ac.uk/
4) 石原 他: 聴者依存性に着目した環境音の擬音語へのシンボルグラウンディング, 日本ソフトウェア科学会第 20 回大会論文集, 2003.