E-003

擬音語 HMM に基づく音場ディクテーションシステムの評価

Evaluation of the Acoustic Sound Field Dictation System with Hidden Markov Model Based on an Onomatopoeia

林田 亘平† 溝口 遊† 小川 純平‡ 森勢 将雅‡ 西浦 敬信‡ 山下 洋一‡
Kohei Hayashida Yu Mizoguchi Junpei Ogawa Masanori Morise Takanobu Nishiura Yoichi Yamashita

1．まえがき

環境音はこれまで雑音として扱われ音声認識の前処理などにおいて除去すべき対象として研究が行われてきた．しかし近年，環境音を含めた音環境の理解に注目が集まっている[1]．音声だけでなく環境音を含めて音場をディクテーションできれば，音環境の正確なアーカイブが可能となる．そして，長時間信号の中から異常音をテキスト情報で高速に検索可能となるため，防犯システムなどでの需要が高まっている．本稿では，実環境の音場理解を目的として，音場ディクテーションシステムのためのテキスト情報検索の利便性を考慮した環境音識別法について検討する．環境音識別の従来研究として，三木らの HMM(Hidden Markov Model)を用いた方法がある[2]．この方法は HMM [3] を用いて各環境音を個々にモデル化し識別を行う．しかし環境音は無数に存在するため，実世界に無数に存在する環境音を全て個々にモデル化することは困難である．また，音環境のアーカイブにおいて検索の利便性を考慮した場合，類似音源をまとめて検索できることが望ましい．そこで本稿では擬音語 HMM を用いた環境音識別法を提案し，モデル数の削減を図る．

2．擬音語 HMM を用いた環境音識別法の提案

自然界の様々な音を模倣して作られた言葉は擬音語と定義される[4]．近年，音とそれを表現する擬音語と人間の感覚の関係が研究されている[5]．これらの研究は擬音語の文字表記から音源を想起可能であり，音の大きさ・高さ・音色などの特徴を推定可能であることを示している．そこで類似した音源を擬音語によって分類し 1 つのモデルで表現する擬音語 HMM を用いた環境音識別法を提案する．擬音語 HMM を用いて環境音識別を行うことでモデル数を削減し，識別結果から音源を容易に想起可能となる．

3. 擬音語カテゴリの策定

本稿ではモデルの学習及び識別実験に技術研究組合 新情報処理開発機構(RWCP: Real World Computing Partnership)が作成した実環境音声・音響データベースの非音声ドライソース(RWCP-DB)[6]を環境音として使用した．まず，RWCP-DB の各音源と擬音語との対応関係を明らかにするためアンケート調査を行った．被験者は男性 15 名，女性 2 名とした．アンケートは RWCP-DB 内の音源を提示し，各種類ごとに 12 個の擬音語を選択肢として与え，もっとも音源を表現している擬音語を選択するという方法で行った．各音源に対する擬音語の選択肢は，比屋根らの RWCP-DB に対する分類[7]を参考にした．アンケートの結果から本研究では表 1 に示す RWCP 標準カテゴリの環境音を表 2 に示す 33 個の擬音語に分類した．

4. 提案法の評価実験

提案法の評価実験として，まず HMM を用いて音源を個々にモデル化する従来法と擬音語 HMM を用いた提案法について環境音識別精度の比較を行った．そして従来法・提案法について識別結果から音源を想起可能であるかを確認するため，実環境で収録した環境音を使用して主観評価実験を行った．

4. 1 環境音識別実験と結果

提案法と従来法の環境音識別精度の比較を行った．従来法は，表 1 に示す RWCP 標準カテゴリを用いて識別を行い，提案法は表 2 に示す擬音語カテゴリを用いて識別を行った．環境音識別の実験条件を表 3 に示す．HMM の学習及び実験には RWCP-DB から，サンプル数が十分な 86 種類を使用した．

図 1 に従来法・提案法それぞれの誤識別率を示す．図 1(a), (b)から提案法は各カテゴリの誤識別率が従来法と比較して改善されていることがわかり，提案法の有効性が確認できる．図 1(b)において，提案法の誤識別率が高かったものは，正解カテゴリ「シュー」に対して「ザー」への誤識

表 1 RWCP 標準カテゴリ

cherry1	cherry2	cherry3	magno1	magno2	magno3
teak1	teak2	teak3	wood1	wood2	wood3
bank	bowl	candybwl	coffcan	colacan	metal05
metal10	metal15	pan	trashbox	caes1	case3
dice1	dice2	dice3	bottle1	bottle2	china1
china2	china3	china4	cup1	cup2	particl1
particl2	pump	spray	file	sandpp1	sandpp2
saw1	saw2	aircap	sticks	cap1	cap2
clap1	clap2	claps1	claps2	snap	bells5
coin1	coin2	coin3	coins1	coins4	book1
book2	crumple	tear	castanet	drum	horn
maracas	ring	string	whistle1	whistle2	whistle3
buzzer	clock2	phone1	phone4	pipong	clock1
coffmill	doorlock	dryer	mechbell	padlock	punch
shaver	stapler				

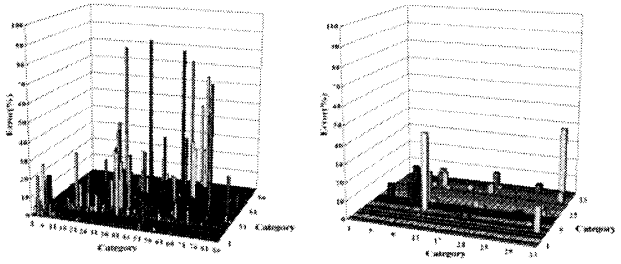
表 2 擬音語カテゴリ

バシ	ブー	チャッ	チャリン	チッ	チリン
ガン	ガサ	ゴー	カチャ	カン	カラン
カララ	カタ	カタカタ	カタタ	キュッ	パーン
パキ	パン	ピー	ピンポン	ピピピ	ピリリ
ピロロ	ポーン	シュッ	シュー	チン	トン
ザッ	ザー	ジリリ			

† 立命館大学大学院 理工学研究科
‡ 立命館大学 情報理工学部

表 3 環境音識別の実験条件

標本化周波数	16 kHz
特徴ベクトル	計 33 次元 MFCC 16 次元 + ⊿MFCC 16 次元 + ⊿パワー 1 次元
状態数	8 状態(left-to-right)
HMM	128 混合
学習データ	RWCP-DB 内 86 種の音源, 各 50 サンプル
実験データ	RWCP-DB 内 86 種の音源, 学習に未使用の各 50 サンプル



(a) 従来法　(b) 提案法

図 1 誤識別率

別, 正解カテゴリ「カラン」に対して「カン」への誤識別であった. これらは前述の擬音語カテゴリ策定のために実施したアンケートの結果においても, 回答に同様のばらつきがあった. そのため, 人間の聴覚で識別の困難な音源は, 計算機による識別も困難であるということがわかる.

4.2 主観評価実験と結果

次に, 実環境音の識別結果から音源を想起可能であるか確かめるため, 主観評価実験を行った. 実験には実環境で収録した音源を使用した. 音源は表 4 に示す 20 種類を各 5 サンプルずつ収録した. そして従来法・提案法を用いて識別を行い, どちらも識別率が 60 % 以上となる結果を識別結果として使用した. そして, 識別結果から音源を想起可能であるかを「想起できる」「どちらともいえない」「想起できない」の 3 段階での評定尺度法を用いて主観評価を行った. 被験者は男性 13 名, 女性 2 名とした.

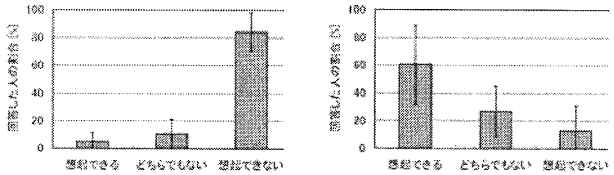
従来法と提案法の主観評価結果を図 2 に示す. 結果は, 各選択肢に解答した人の割合の平均と標準偏差を示している. 図 2(a), (b)から従来法と比較し提案法は実環境で収録した音源に対し, より識別結果から音源を想起可能であることがわかる.

5 考察

提案法において想起できないと回答した人の割合が多かった音源として「引き出しを閉める音」, 「引き戸を閉める音」, 「扉を閉める音」, 「窓を閉める音」がある. これらの音源は閉まる途中の音と閉まった瞬間の衝撃音の 2 音節で構成されている. そのためこれらの音源は, 1 音節の擬音語を用いた提案法では十分にモデル化が行えていないため, 想起できないと回答した人の割合が多くなったと考えられる. 従って, 2 音節以上の擬音語によって表現される複合環境音についてモデルを作成する必要があると考えられる.

表 4 実環境音の音源リスト

携帯電話の着信音	ペンを床に落とす音
コップをテーブルに置く音	エンターキーを叩く音
コップをスプーンで叩く音	足音
コップをスプーンで混ぜる音	椅子に座る音
カンをテーブルに置く音	扉を閉める音
スプーンを床に落とす音	窓を閉める音
引き出しを閉める音	携帯のマナー音
鈴の付いた鍵を取り出す音	紙を丸める音
ホッチキスで紙を綴じる音	紙を破る音
引き戸を閉める音	スプレーを噴射する音



(a) 従来法　(b) 提案法

図 2 主観評価結果

6 おわりに

本研究では擬音語 HMM を用いた環境音識別法を提案した. 環境音識別実験の結果, 従来法よりも提案法の識別率が向上した. また主観評価実験の結果, 提案法は実環境で収録した音源の識別結果から音源を想起可能であることがわかった. 以上の結果から提案法は, 音場ディクテーション結果から音源を想起可能な環境音識別が可能であると考えられる. しかし, 複合環境音に対し提案法は十分なモデル化が行えていないことがわかった. そのため, 今後は複合環境音についてモデル化を行う必要がある. また, 今回はアンケートによって主観的に環境音を擬音語カテゴリに分類したが, 今後は音響空間上での距離を考慮するなど, 客観的な尺度を検討する必要がある.

7 謝辞

本研究の一部はグローバル COE, 科研費による研究助成を受けた.

8 文献

[1] 奥乃博 他, “コンピュータサイエンスから見た聴覚の情景分析, ” 日本音響学会誌, Vol. 50, No. 12, pp. 1017—1022, 1994.

[2] 三木一浩 他, “HMM を用いた環境音識別の検討, ” 信学技報, SP99-106, pp. 79—84, 1999.

[3] 徳田恵一 他, “HMM による音声の合成の基礎, ” 信学技報, SP2000-74, pp. 43—50, 2000.

[4] 浅野鶴子 編, “擬音語・擬態語辞典, ” 角川書店, 1978.

[5] 川井敬二 他, “環境音の印象評価構造に関する研究, ” 日本音響学会誌, Vol. 60, No. 5, pp. 249—257, 2004.

[6] 新情報処理開発機構 技術研究組合 実環境音響サブワーキンググループ, “実環境音声・音響データベース報告書, ” 1998.

[7] 比屋根一雄 他, “単発音のスペクトル構造とその擬音語表現に関する検討, ” 信学技報, SP97-125, pp. 65—72, 1998