K-091

日本語指文字認識・合成用コーパスの構築
Construction of a Corpus for Japanese Manual Alphabet Recognition and Synthesis

江本 祐太† 宮島 千代美† 伊藤 克亘† 武田 一哉†
Yuta Emoto Chiyomi Miyajima Katsunobu Itou Kazuya Takeda

1. はじめに

現在，難聴に苦しむ人は全国で30万人，軽度のものも含めば，600万人以上もの人がいる．ろう者の積極的な社会参加には，健聴者との間のコミュニケーション支援が重要である．しかし，コミュニケーション手段の1つである手話を習得することは容易ではなく，ろう者においても手話の習得率は20%足らずと言われており，学習支援が望まれる．この問題を解決するための1つの手段として，手話学習システムを始めとした手話認識・合成システムが期待されている [1]-[7].

手話認識の研究は国内外を問わず盛んに行われており，圧縮連続DP照合を用いた手法 [3] や統計的モデルである隠れマルコフモデル (HMM: Hidden Markov Model) を用いた手法 [1][2] により，高い認識率が得られている．本研究で対象とする指文字 [8] の認識に関しては，ニューラルネットを用いた手法 [5] や手の関節角をコード化して認識を行う手法 [6] などが報告されているが，これらは手形状が区別できる指文字のみを対象としており，手形状のみでは区別できない動作を伴う指文字の認識には至っていない．また，後藤らのベイズ識別を用いた手法 [7] では，動作を伴う指文字を含む，連続した指文字の認識を行い，94.4%の指文字認識率が得られているが，認識対象が50単語という比較的小規模な実験であった．

本研究では，大語彙連続指文字認識および，任意のテキストから指文字画像を合成することを目的とする．音声の認識・合成の研究においては，HMMなどの統計的なモデル化の有効性が確認されており [9]，そのモデル化の際には，音韻のバランスのとれたコーパスを使用する必要がある [10][11]．しかし，指文字の大語彙コーパスは存在しないため，今回，連続指文字認識・合成研究用コーパスを構築した．

また，作成した学習・評価用セットをもとに指文字のデータを収録し，このデータを用いてHMMに基づく連続指文字認識および合成 [4] を行った．

2. 手話・指文字

手話は，ろう者間およびろう者・健聴者間のコミュニケーション手段であり，身振りを基本に単語を記号化したものである．指文字は，日本語の50音一字一字に対応した手指記号であり，手話会話の中では，手話に存在しない地名や人名などの固有名詞や，比較的新しい外来語などで手話がまだ確立していないものを表現するために，手話と組み合わせて用いられる．本研究で対象とする指文字 (81文字) は，手の静的形状で文字の定まる静止指文字 (41文字) と，手形状と動きで文字の定まる動作を伴う指文字 (40文字) に分類される．図1の例のように，濁音は清音の指文字を右に，半濁音は上に，拗音・促音などの小文字はその指文字を手前に動かすことで表現する．また，“も”，“の”，“ん” や “ー”(長音) なども動作を伴う指文字である．長音については，例えば “たー” と表す場合，“た”，“ー” と2文字の指文字で表現する方法と，“た” の指文字を下に移動させる表現方法があるが，本研究では前者の表現方法を用いる．

図1: 指文字の例「ダーウィン」

3. 指文字テキストコーパスの構築

3.1 単語の選定方法

バランスのとれたテキストコーパスを設計する方法として，情報エントロピーを評価尺度とするテキスト選択法が提案されている [10][11]．一般に，音声認識のための統計モデルの学習には，音素を最小単位とし，音韻のバランスがとれるよう，子音 (C) と母音 (V) から成る三つ組CVC・VCVのバランスを考慮して構築されたコーパスが利用されている．指文字のモデル化においては，文字 (L) が最小単位であり，連続する指文字の並びのバランスを考慮するため，Lの二つ組LLを基本として考える．また，単語コーパスは収録負荷を考慮すると，できるだけ少ないテキストに多くのLLがバランスよく含まれることが望ましい．そこで，本研究では，LLの出現確率のバランスがよくなるような，最小限のサイズのテキストセットを選定した．単語の選定手順は以下の通りである．

(i) 基となる単語データベースの中で1度しか登場しないLLを含む単語 (唯一語) を全て採用する．

(ii) 新しいLLを最も多く含む単語を採用する．
但し，該当する単語が複数ある場合は，LLでのエントロピーを最大化する単語を採用する．

(iii) (ii) を新しい並びがなくなるまで，または規定単語数に達するまで繰り返す．

(iv) 選定された単語セット中より，日本語としてこなれない，指文字でスムーズに読むことが困難な単語を除外し，さらに (ii)，(iii) を繰り返す．

但し，エントロピー S は，選定されたテキストセット中で n 番目の文字の二つ組 (LL) が出現する確率を p_n，N を登場したLLの総数として，次式で与えられる．

$$S = -\sum_{n=1}^{N} p_n \log_2 p_n \quad (1)$$

†名古屋大学大学院情報科学研究科

表 1: 使用データベース

データベース名	選定に用いた単語	総単語数
毎日新聞 (1991〜2002)	頻出人名・地名・カタカナ語	15,307
郵便番号データベース	市町村郡名	3,675
姓名読みふりファイル	日本人の名	24,623
西洋人名辞書	西洋人名	47,299
合計		82,994

表 2: 選定単語の例

評価用セット	学習用セット
1. アーカンソー	1. アーメダバード
2. アイダホ	2. アイヅタカダマチ
3. アイヅワカマツ	3. アオヤナギ
4. アイルランド	4. アボカド
5. アインシュタイン	5. アケボノバシ
⋮	⋮
998. ワタヌキ	900. ワハシー
999. ワトソン	901. ワルトシュミット
1000. ワヘイ	902. ワンヤンアクダ

上記の手順でまず評価用セット 1,000 単語を選定し，さらに評価用セットに含まれる LL をカバーするように，同じエントロピー基準で学習用セット 902 語の選定を行った．

3.2 コーパス構築に用いたデータベース

2 節で述べたように，指文字は主に固有名詞や外来語などの手話にない単語を表現するのに用いられるため，表 1 に示した各データベースから固有名詞や外来語を抽出し，コーパス作成に用いた．また，実際の認識・合成の評価の際には，使用頻度の高い単語における評価が重要となるため，評価用セットは，実用上出現頻度の高い人名・地名・カタカナ語 (計 15,307 語) から選定した．ここで選定されなかった語に，市町村郡名，日本人の名，西洋人名を加えた計 81,944 語 (重複を除く) から，学習用セットを作成した．

3.3 選定結果

選定された単語の例を表 2 に示す．また，全データ (82,944 単語) に出現する LL の総数は 5,344 通りであり，各単語セットが全データに含まれる LL をどれだけカバーしているか (カバー率) を表 3 の第 4 列に示した．全データ中から無作為に選んだ 1,000 語とのカバー率を比較することにより，提案手法により選定された単語が，バランスよく選定されていることが確認された．

また学習用セットは，評価用セット中の LL を 98.7%カバーした．

表 3: 選定結果

単語セット	単語数	含まれる LL の類数	全データ中の LL のカバー率
評価用	1,000	2,316	43.3 %
学習用	902	2,708	50.7 %
ランダム	1,000	1,544	28.9 %

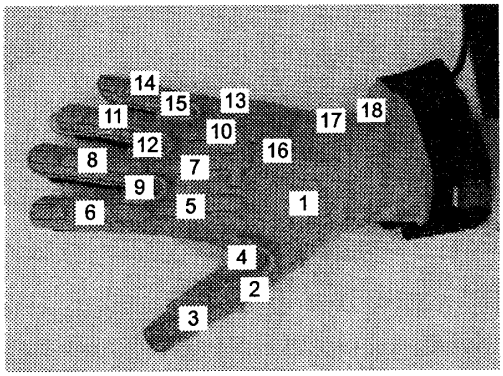

図 2: 角度センサーの位置

4. データ収録

3 節で作成したテキストセットに基づき，3 次元位置センサー (3SPACE FASTRAK) および手形状測定装置 (Cyber Glove) を用いて，手首の位置・方向 (6 次元) および手指の関節角度 (18 次元) を収録した [4]．Cyber Glove のセンサーの位置 (18 箇所) を図 2 に示す．収録は共同研究者の森らによって行われ，健聴者 1 名分の指文字のデータとともに，指文字動作時に発声した音声，および確認用ビデオも同時に収録された．

なお，今回は 1 名のみの収録であったが，今後手話を母語とするろう者も含め，データベースを拡大する予定である．

5. HMM に基づく連続指文字認識・合成

HMM は統計的なモデル化手法であり，音声・画像の認識や合成などでその有効性が示されている [1][4][9]．認識では各モデルによって出力系列が生起する尤度を求め，最大尤度を与えるモデルを選んで，これを認識結果とする．

5.1 関節角データによる認識

まず，予備実験として，18 次元の手指の関節角 (θ) のみを特徴量として用い，学習用データにより指文字単位の HMM を学習し，評価用データについて認識実験を行った．実験条件を表 5 に示す．

5.1.1 実験 1：状態数・動的特徴量に関する評価実験

手指の関節角 θ より，1 次・2 次の動的特徴量 ($\Delta\theta$, $\Delta^2\theta$) を算出し，HMM の状態数および用いる特徴量をそれぞれ変化させ，文字単位で認識実験を行った．

表 4: 実験条件

標本化周波数	100 Hz
HMM の状態数	3, 10, 20, 30, 40, 50
HMM の混合数	単一ガウス分布
特徴量	関節角 θ (18 次元), θ, $\Delta\theta$(36 次元), θ, $\Delta\theta$, $\Delta^2\theta$(54 次元)
辞書サイズ	1,000, 1,902, 15,307
学習データ	1, 1/2, 1/4, 1/6
縮小方法	選定手法，無作為

5.1.2 実験 2：辞書サイズに関する評価実験

単語辞書として評価セットの 1,000 単語，学習セットを含めた 1,902 単語，および表 1 における頻出語 15,307 単語を用い，単語単位で認識実験を行った．但し，HMM の状態数および用いる特徴量は実験 1 において最も高い認識率が得られたものとした．

5.1.3 実験 3：学習データ量に関する評価実験

エントロピー最大化基準でのテキスト選定法の有効性を確認するために，学習データを 3.1 節の選定手法で，1/2, 1/4, 1/6 の量に縮小し，認識実験を行った．但し，唯一語の選定は行わないものとした．学習データから無作為に抽出した単語で同様の実験を行い，提案手法により縮小した場合との認識性能を比較した．HMM の状態数および用いる特徴量は実験 1 において最も高い認識率が得られたものとし，認識単語辞書のサイズは 15,307 単語とした．

5.1.4 実験結果

実験 1 では，図 3 に示すように，認識率は特徴量として θ , $\Delta\theta$((36 次) および θ , $\Delta\theta$, $\Delta^2\theta$(54 次) を用い，HMM の状態数を 30 とした場合が最も認識率が高く，76.8%の文字認識率が得られた．また，動的特徴量 ($\Delta\theta$, $\Delta^2\theta$) を加えることで，少ない状態数においても，高い認識率が実現可能となることが分かった．

実験 2 の結果を図 4 に示す．辞書サイズを 15,307 単語とした場合の認識率は 95.8%であり，評価用単語セットの 1000 語を辞書とした場合 (99.4%) と比較すると，辞書サイズを大語彙なものとしても，3.6%の低下で抑えられていることが分かる．

さらに図 5 に示したように，本手法により選定した学習データは，ランダムに選定した場合に比べ，高い認識率が保持できた．これにより，本選定手法の有効性が確認された．

5.1.5 考察

表 5 に，評価用セット中に 10 回以上登場する文字のうち，認識率が低かった文字について，その認識結果を示した．表 5・図 6 に示したように，“び” を “ひ”，“ひ” と誤るなど，濁音・半濁音・拗音・促音などの同一の手形状で動作を伴う文字において誤認識が多かった．また，“あ” と “た”，“み” と “ゆ” などの手首の向きのみが異なる静止文字の誤りもみられた．これらの誤りについては，位置センサーから得られる手首の位置・方向データを特徴量として併用することで改善できると考えられる．

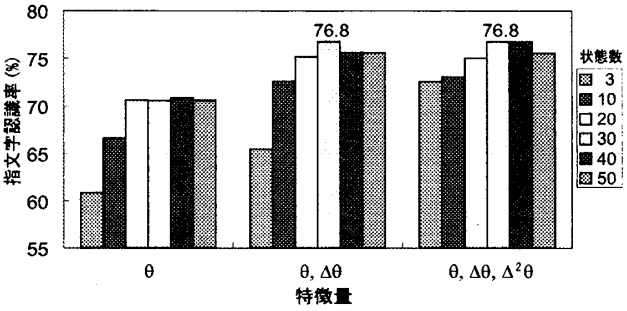

図 3: 状態数・特徴量の変化における文字認識率

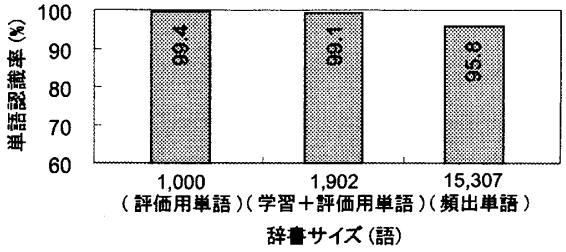

図 4: 辞書サイズと単語認識率

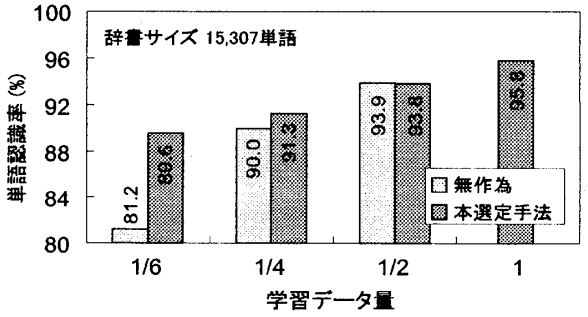

図 5: 学習データの縮小による認識率の変化

5.2 実験 4：関節角および位置情報を用いた認識

5.2.1 実験条件

2 つのセンサー間でサンプリング周波数が異なるため，区分的 3 次エルミート補間 [12] により補間を行い，サンプリング周波数を 100Hz に統一し，特徴量を初期統合した．実験条件を表 6 に示す．ただし，HMM の状態数は 10，30，50 とし，予備実験より，用いる特徴量は関節角，位置データとそれぞれの 1 次動的特徴量とした．

5.2.2 実験結果

実験結果を図 7 に示す．位置・方向データの統合により，文字認識において 18.1%，単語認識において 2.1%の認識率向上が見られた．これは，同一の手形状で動作を伴う文字および，手首の向きのみが異なる静止文字に対する認識誤りが大きく改善されているためである．実際，

表 5: 関節角データのみの場合の認識誤り例

正解文字		主な認識誤り
ぴ	→	び , ひ
ぺ	→	べ
ぱ	→	ば, は, DEL
ぞ	→	ー, ん, そ ぞ, DEL
ぽ	→	ぼ, ぽ, DEL

DEL: 削除誤り

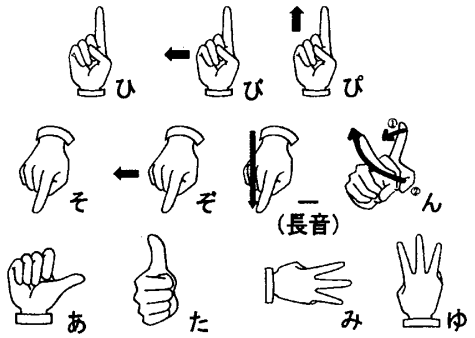

図 6: 手形状が同一で動作・向きのみが異なる指文字の例

これらの認識誤り数は，804 文字から 73 文字と 10 分の 1 以下に減少した．

5.3 連続指文字合成

共同研究者の森らにより，本研究で設計したコーパスに基づいて作成されたデータベースを用いて HMM に基づく連続指文字合成が行われ [4]，その結果，自然な指文字の動画像が合成可能であることが確認された．これにより，動画像合成用の統計モデルの学習にも，今回構築したコーパスが有効であることが確認された．

6. まとめと今後の課題

本研究では，統計的なモデル化手法に基づく連続指文字認識および合成を行うための，テキストコーパスを構築した．コーパスに従い収録された指文字の手形状及び手首の位置・方向データを用いて，HMM に基づく認識実験を行った結果，大語彙においても 97.9%と高い認識率が得られた．更に，共同研究者が本コーパスを用いて指文字動画像合成を行った結果，自然で滑らかな指文字動画像を合成できることを確認した．これらの結果から，今回構築したコーパスの有効性が確認された．

今後，コンテキスト (前後の文字環境) を考慮したモデル化や，指文字の記述法に基づくモデル化などにより，認識率の改善が図れると考えられる．また，ろう者・手話経験者など多くの被験者によりデータ収録を行い，データベースを充実させることも課題として挙げられる．

表 6: 関節角・位置情報を統合した場合の認識実験

標本化周波数	100 Hz
HMM の状態数	10, 30, 50
HMM の混合数	単一ガウス分布
特徴量	θ, $\Delta\theta$(36 次元) θ, $\Delta\theta$, z, Δz(48 次元)
辞書サイズ	15,307 単語

θ: 手指の関節角 (18 次元)
z: 手の位置・方向 (6 次元)

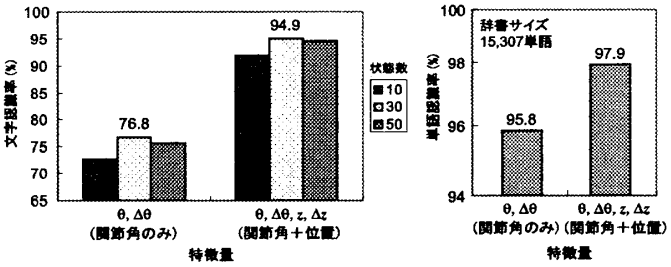

図 7: 認識率の比較

参考文献

[1] T. Sterner et al.“, Real-time American sign language recognition using desk and wearable computer based video, ” Proc. IEEE, vol.20, no.12, pp.1371–1375, Dec 1998.

[2] B. Bauer et al.“, Relevant features for video-based continuos sign language recognition, ” Proc. IEEE-FGR'00, pp.440–445, Mar. 2000.

[3] 佐川 浩彦, 酒匂 裕, 大平 栄二, 崎山 朝子, 阿部 正博, “圧縮連続 DP 照合を用いた手話認識方式”, 電子情報通信学会誌, vol.J77-D-II, no.4, pp.753–763, Apr. 1994.

[4] 森健史, 南角吉彦, 宮島千代美, 徳田恵一, 北村正 “隠れマルコフモデルに基づく指文字動画像生成”, 情報科学技術フォーラム, Sept. 2005(発表予定).

[5] 内田 雅文, “手形状認識と手話への応用”, 電気学会論文誌, vol. 114-C, no. 10, pp.995–999, Oct. 1994.

[6] 高橋 友一, 岸野 文郎, “手振り認識方法とその応用”, 電子情報通信学会論文誌, vol.J73-D-II, no.12, pp.1985–1992, Dec. 1990.

[7] 後藤 岳志, “動作を伴なう指文字を含む連続指文字認識”, 電気関係学会北陸支部連合大会, p. 369, 1996.

[8] NPO 手話技能検定協会,“手話技能検定公式テキスト”, Oct. 2003.

[9] 鹿野 清宏, “音声認識システム”, オーム社, 2001.

[10] 産業技術総合研究所, “研究用音声データベース”, http://www.aist.go.jp/RIODB/db066/

[11] H. Kuwabara, et al.“, Construction of a large-scale Japanese speech database and its management system, ” Proc. ICASSP, vol.1, pp.560–563, May 1989.

[12] Ricardo A. D. Zanardini“, Piecewise cubic hermite interpolation,” Feb. 2005.