

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月23日現在

機関番号：13903

研究種目：基盤研究(C)

研究期間：2010～2012

課題番号：22500506

研究課題名(和文) 統計モデルに基づく手話動作のサブユニットモデル自動生成と大語彙連続手話認識

研究課題名(英文) Subunit modeling for Japanese sign language recognition based on stochastic model

研究代表者

北村 正 (KITAMURA TADASHI)

名古屋工業大学・工学研究科・教授

研究者番号：60114865

研究成果の概要(和文)：本研究では、単語レベルで学習された多数のHMMを用い、各HMMの状態パラメータの分類から手話間の共通要素を集約する新たな手法と、個人差などによる手話動作の差異に頑健に対応する手話認識手法を提案した。その際、手話では手の位置、運動、局所情報という性質の異なる要素の同時的な組合せから多様な動作を表す、手話の音韻構造に着目し、これらの音韻要素と両手の独立性を考え、個別に学習・分類することで効率的に共通要素の集約がより効率化される。これにより単語全体のモデルパラメータ数削減と、それに伴う相対的な学習データ数の増加が見込まれる。RWC手話データベースを用いた単語認識実験では、本提案によるサブユニットの導入と音韻構造を考慮することにより、我々の手法が手話単語認識精度の改善に寄与することが確認された。

研究成果の概要(英文)：In this study, an important issue for modeling sign is that how to determine the constituent element of sign (i.e., subunit) like "phoneme" in spoken language. We focused on special feature of sign language that Japanese sign language is composed of three types of phonological elements which is hand local information, position, and movement. In this paper, we propose an efficiently method of generating subunit using multi-stream HMM which is correspond to phonological elements. 100 and 307 isolated sign recognition experiment has confirmed the effectiveness of our proposed method.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,900,000	570,000	2,470,000
2011年度	900,000	270,000	1,170,000
2012年度	500,000	150,000	650,000
総計	3,300,000	990,000	4,290,000

研究分野：総合領域

科研費の分科・細目：人間医工学

キーワード：ヒューマンインターフェース、手話認識、手話動作サブユニット

1. 研究開始当初の背景

人間のコミュニケーションにおいて、会話などの音声言語情報とともに、身振りや表情のような視覚情報の果たす役割は極めて重

要であり、円滑な意思疎通には欠かすことができないものである。場合によっては音声よりも重要な役割を果たすことがあり、視覚言語としての手話は聴覚障害者のコミュニケ

ーションに不可欠なものである。

近年の音声認識・合成技術の発展に伴い、音声で機械と対話する情報システムの実用化が進む中で、ユニバーサルアクセスという観点からも、手話を認識・合成するための技術が必要とされている。

しかしながら手話を対象とする国内外の研究では、手話データ収集の困難さや動画処理による計算量の多さなどの問題があり、定まった知見がまだ十分に得られていない。各国の手話を中心として手話認識研究が行われているが、手話単語単位などの限定的なものも多く、大規模な連続手話単語認識や、文章レベルの手話認識の研究はまだ少ない。その原因の一つとして、手話では音声言語と比べて言語資源的な知識の集約がまだ十分ではなく、同時にそれらの知見を手話認識の枠組みに活用できていないことが考えられる。従って、手話認識ではまだ小規模な語彙による実験や評価が中心で、実用性という観点では十分な成果が得られていないのが現状である。

このような視点から、実用的な連続手話認識を実現するための取り組みとして、手話の動作を構成するためのサブユニット（音声言語でいう音素に相当する、語彙間で共通の基本動作単位）に基づいて、サブユニットによる手話動作のモデル化・認識を行う試みが我々のグループやドイツの RWTH などでも検討されてきた。これまでの我々の取り組みでは、音声認識で主流となっている隠れマルコフモデルの枠組みを手話動作のモデル化手法に応用する中で、手話表現を構成する基本的な動作要素に分解することで、限られた手話動画データから様々なサブユニットを自動的に定めて複数の語彙の中でそれらを共有する効率的なモデル化手法を提案してきた。本申請では、サブユニットモデルの高精度化と、大語彙での認識に適した効率的な認識アルゴリズム等を開発し、実用的な手話認識システムの実現にむけた研究を展開する。

2. 研究の目的

これまでの研究では、サブユニットの導入によってモデルパラメータの共有化による頑健性や認識率の改善に寄与する一方で、手話表現が細部で微妙な違いをもつことから、個々のサブユニットが過剰に分類される傾向が問題として残されている。これは、個々のモデルの推定精度が低下するほか、大語彙タスクに適用する際の解探索の非効率化を招き、認識率低下に直結すると考えられる。そのため、限られた学習データをより効率的に用いたモデル学習手法が必要と考えた。また、手話研究には研究用データベースに限られるほか、近年開発が進んでいる深度センサにより3次元情報を計測することが容易にな

りつつある。今後の大規模データベースの作成を推進するため、手話データベースの構築手法を検討する必要がある。

そこで本申請では、主に下記にあげる研究課題に取り組み

(1) 手話言語の特性に基づいたサブユニットモデルの自動生成：従来手法を拡張し、手話の言語的な構造を取り入れることで、手話の音韻分類に基づいたサブユニットの自動生成手法を開発する。

(2) 認識アルゴリズムの効率化：大規模な認識タスクではサブユニット数も膨大となり、複雑なサブユニット列が構成されるが、これは個人の差異が大きい手話では問題となる。そこで、個人差の違いに対応するため、複数の遷移パターンを考慮したサブユニットネットワークの自動生成手法を開発する。

(3) 大規模タスクに必要なデータベース設計方法の確立：これまでの音声研究で培ってきたデータベース構築の経験を生かし、文章セット作成や語彙選定など、手話の特性に適合したデータベース設計法を模索し、今後のさらなる大規模データの収集にむけた方法論を確立する。

3. 研究の方法

(1) 言語的特性を考慮したサブユニットモデリング

手話は空間における身体動作によって表現され、音声とは異なる音韻構造があると言われている。神田らによる日本手話音韻表記法によれば、手話は手のかまえ（掌の型や局所運動など）部門・位置部門・運動（手の移動方向など）部門という3種類の要素があり、これらの要素が両手で同時的に出現することで多様な動作を表現すると言える。

そこで、図1のような手話動作のサブユニット自動生成の考え方を拡張し、局所情報・位置・運動のストリームを個別に学習し、音韻の種類別にクラスタリングを行う。これにより、他のストリームに依存せず類似動作を効率的に分類することができ、分類効率の改善によるさらなる総パラメータ数削減と認識率向上が期待できる。

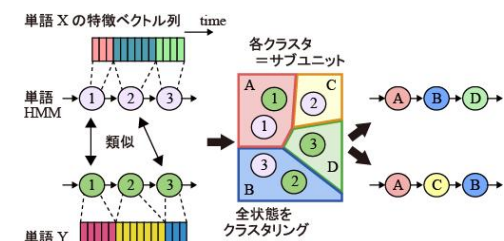


図1. HMMを用いたサブユニットの自動生成

ここで、両手の動作は異なるが、構成要素であるサブユニットは両手で共有可能と考え

られるため、サブユニットのモデルパラメータは片手分のみとなる。

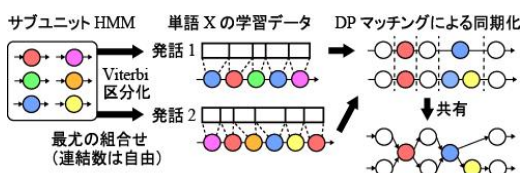


図 2. 多様な動作に対応可能な状態遷移の生成手法

(2) 手話動作の違いに対応した状態遷移構造の生成

手話動作は物事のイメージを表現したものであることが多いため、意味の解釈に支障がないレベルで同一単語の動作表現が異なることがあり、特に話者や文脈が異なる場合にその影響が大きいと考えられる。よって、単語 HMM の状態構造としては、状態のスキップを持たない left-to-right 構造よりも、よりパターンの変動に柔軟に対応可能なモデル構造が必要である。本研究では、同一単語における動作の変動がサブユニットの連鎖パターンの変化であるとみなし、図 2 のようにして、学習データ中に存在する多様な動作表現に対応したサブユニットの遷移構造を生成する。まず、発話ごとにサブユニット列を Viterbi アルゴリズムにより定め、DP マッチングにより時間的同期をとった上で、発話間でサブユニットの共有化を行う。これにより、共通部分からどの発話のパターンにでも分岐して遷移可能であるような単語のモデル構造が生成でき、学習データ中に存在しない異なる状態遷移パターンを含んだ未知データを柔軟に認識可能となる。

(3) 手話データベースの構築

手話技能検定初級程度の手話に含まれる基本的な語を対象とし、以下のステップにより、抜粋された 548 語から音韻のバランスをできるだけ一定に保つように単語数を削減し、337 単語を選定した。

- ・手話技能検定初級の単語を抜粋
- ・手話辞典を参考に分析・調整
- ・分析結果を基に音韻の出現頻度を算出
- ・各音韻を網羅し、出現頻度を維持するように単語を削減

先行研究で発表されている神田らによる日本手話の音韻表記法に基づいて各単語を分類した。また、形式的な分類だけでなく、実用的な観点から手話の音韻の出現頻度に調整するため、専門家が分類を行った先行研究の知見も参考にした。選定された 337 単語を用いて、手話技能検定初級のテキストに見られる平易な構造の文章を 500 文生成し、デー

タベースの文章セットを作成した。

収録装置として、TOF カメラ (MESA SR4000) と CCD カメラ (PGR GRAS-03K2M/C) を同時に用い、手話通訳士 1 名により収録を行った。収録されたデータは発話単位ごとに切り出し、手話単語ラベルなどのラベリング作業を行った。

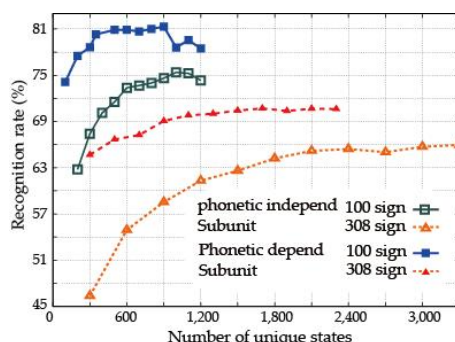


図 3. 音韻構造に基づいたサブユニットモデルの効果 (提案手法と従来手法の単語認識率の比較)

4. 研究成果

言語構造を考慮したサブユニットモデルの最適化手法、および手話動作の差異に対応する手法 (課題 1 および 2) に関して、RWC 手話データベース (語彙数 308、男女 4 名からなる動画像データベース) を用いて手話単語認識による評価実験を行った。まず、図 3 では、従来手法の音韻構造を用いないサブユニットモデルと、本研究の音韻構造を用いたサブユニットモデルを比較した。この結果では、総パラメータ数削減と認識率向上が示された。各音韻のサブユニット数が 500 のとき認識率が 80% を上回っている。従来法と比較して認識率を大幅に改善できる一方で、モデルのもつ総パラメータ数を大幅に削減できることが確認でき、モデルをより効率的に作成できるとともに、認識率の改善にも寄与することが示された。

また、手話動作の違いに対応する状態遷移構造の最適化 (課題 2) に関する実験結果を、図 4 に示す。柔軟な遷移構造を用いたことにより、従来法である left-to-right 型の単純な遷移構造にもモデルと比較し、未知話者の動作に対する認識率が全ての条件において向上したことがわかる。ここでの最高認識率は約 72 % となった。

3 次元手話データベースの構築のため、手話データベースの文章セットを音声データベース構築の経験と言語的な知見を用いて設計した。作成された 500 文による文章セットに基づいて、深度センサと CCD カメラを併用して図 5 のように手話通訳士 1 名による手話データベースを収録した。収録したデータ

は発話単位で切り出し、手話単語のセグメンテーションを行った。今後の大規模手話データベースの構築にあたり、様々な知見やノウハウを獲得することができた。また、このデータベースを用いて、奥行き情報を用いた手話認識実験を行い、奥行き情報を用いることの効果が確認できた。

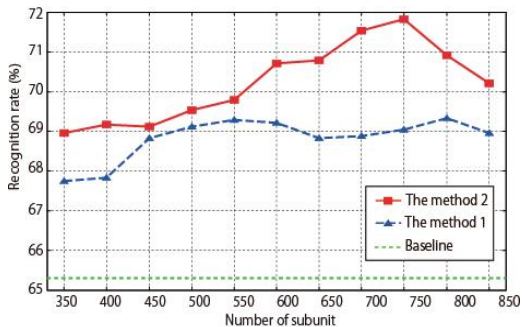


図 4. サブユニットネットワークの改善の効果（提案法と従来法による単語認識率の比較）

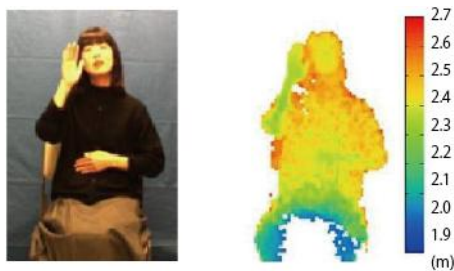


図 5. 3次元手話データベースの例

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 0 件）

〔学会発表〕（計 12 件）

- ① 酒向, 北村, “MSD-HMM による隠れ領域を考慮した手話動作のモデル化”, 電子情報通信学会技術研究報告, Vol. 109, No. 467, pp. 37-42, Mar. 2010.
- ② 有賀, 酒向, 北村, “HMM に基づく人物や動作の多様性を考慮した手話認識手法”, 電子情報通信学会技術研究報告, Vol. 110, no. 53, pp. 55-60, May. 2010.
- ③ 有賀, 酒向, 北村, “HMM に基づく手の動きと形状特徴のストリーム別クラスタリングを用いた手話認識手法”, 第 9 回情報科学技術フォーラム, K-057, Sep.

2010.

- ④ 有賀, 酒向, 北村, 日本手話の音韻構造を考慮した HMM に基づく手話認識, 電子情報通信学会技術研究報告, Vol. 110, No. 56, WIT2010-56, pp. 127-132, Oct. 2010.
- ⑤ K. Ariga, S. Sako and T. Kitamura, “HMM-based Sign Recognition in Consideration of Motion Diversity”, Proc. of 4th International Universal Communication Symposium (IUCS), pp. 257-260, Oct. 2010. (査読有)
- ⑥ 森口, 酒向, 北村, “3次元データベースを用いた HMM 手話認識”, 電子情報通信学会総合大会, ISS-P-112, Mar. 2011.
- ⑦ 酒向, 有賀, 北村, “日本手話の音韻構造を考慮した HMM に基づく手話認識の改善”, 電子情報通信学会技術研究報告, Vol. 111, No. 58, WIT2011-8, pp. 41-46, May. 2011.
- ⑧ 酒向, 有賀, 北村, “音韻構造を考慮した HMM 手話認識における音韻ストリーム重みの最適化手法の検討”, 電子情報通信学会技術研究報告, Vol. 111, No. 174, WIT2011-25, pp. 25-30, Aug. 2011.
- ⑨ 森口, 酒向, 北村, “TOF カメラによる連続手話認識に関する検討”, 第 10 回情報科学技術フォーラム, K-018, pp. 753-754, Sep. 2011.
- ⑩ 森口, 酒向, 北村, “TOF カメラによる 3次元情報を用いた手話認識に関する検討”, 電子情報通信学会福祉情報工学研究会, Vol. 111, No. 424, WIT2011-51, pp. 1-6, Jan. 2012.
- ⑪ 小山, 酒向, 北村, “TOF カメラによる手話の言語的特性を考慮した手話認識”, 情報処理学会第 75 回全国大会, pp. 4-192-193, 2013-03.
- ⑫ S. Sako, T. Kitamura, “Subunit Modeling for Japanese Sign Language Recognition Based on Phonetically Depend Multi-stream Hidden Markov Models”, UAHCI/HCI 2013, Part I, LNCS 8009, pp. 548-555. Jul. 2013. (査読有)

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：

国内外の別：

○取得状況（計0件）

名称：

発明者：

権利者：

種類：

番号：

取得年月日：

国内外の別：

〔その他〕

ホームページ等

<http://www.mmsp.nitech.ac.jp/>

6. 研究組織

(1)研究代表者

北村 正 (KI TAMURA TADASHI)

名古屋工業大学・工学研究科・教授

研究者番号：60114865

(2)研究分担者

酒向 慎司 (SAKO SHINJI)

名古屋工業大学・工学研究科・助教

研究者番号：30396791

(3)連携研究者

()

研究者番号：