論文 / 著書情報
Article / Book Information

論題	大語彙手話認識のための動素決定木クラスタリング
著者	安中哲也, 篠田浩一
出典	第19回 画像センシングシンポジウム，，，pp. IS3-18-1 to IS3-18-8
発行日 / Issue date	2013，6
Note	第19回画像センシングシンポジウム講演論文集より転載

大語彙手話認識のための動素決定木クラスタリング

安中 哲也† 篠田 浩一‡
†東京工業大学情報工学科 ‡東京工業大学大学院情報理工学研究科計算工学専攻
E-mail: annaka@ks.cs.titech.ac.jp† shinoda@cs.titech.ac.jp‡

Abstract

手話を動素と呼ばれる細かい動きの単位に分解し，HMM によりモデル化して自動認識を行う研究が進められている．従来法では，学習データに対して動素 HMM のパラメータ数が多くしばしば過学習に陥る，未知の動素 HMM を含む単語が精度良く認識できない，という問題があった．そこで本研究では，大語彙音声認識で用いられる決定木クラスタリングを用いて，動素 HMM の状態クラスタリングを行うことで，動素 HMM の状態数を削減する．未知の動素 HMM に対してのパラメータ推定も可能になる． 414 単語を学習に用い，未知動素を少なくとも 1 つ含んだ単語のみで構成された 55 単語を用いた認識と，学習単語も加え合計 469 単語を用いた認識の，2 種類の評価実験を行った．提案手法は，Data-Driven のクラスタリング手法より，各々の実験において認識率が 4.5 ポイントと 9.2 ポイント上回った．

1 はじめに

近年，障がい者の社会進出が進んでおり．健常者と障がい者とのコミュニケーション手段が望まれている．そのひとつとして，コンピュータによる手話の自動認識の技術がある． 過去の手話認識の研究は，その多くが手話の観測方法や特徴量に関するものであった．しかし近年，比較的安価な Depth カメラを用いて手の位置を高精度に認識できる技術が確立された [1] ことから，手話の観測方法や特徴量に関する問題点は解決されつつある．したがって，手話認識の残る課題は認識アルゴリズムであり，実用化のためには大語彙認識に拡張可能な手法であることが必要である．

手話も言語の一種とみなすことができる．そのため音声を認識する際に用いられる隠れマルコフモデル (Hidden Markov Model, HMM) を手話認識に応用する研究が多い [2][3][4]．手話単語を HMM でモデル化する場合，手話単語ごとに HMM を構成している場合がほとんどである．実際に用いられる手話単語は数千種類ある．この手法だと，認識語彙数の増加に伴ってモデル数も増加するため，学習サンプル数を十分な量用意することが困難である．また，学習データに存在しない単語を認識することができないという問題点もある．

そこで佐藤 [5] らは，Vogler[4] らの研究を参考に新たな手話動素単位を定義するとともに，手話を「手の位置・動き」「手首の向き」「手の形」の 3 種類の時系列で表現した．そして，要素ごとに動素 HMM を構成し，各構成要素の HMM の出力尤度の重み付き和を用いる認識手法を提案した．しかし，この手法は，動素の総状態数が多く，しばしば過学習に陥る．

本稿では，動素の HMM の状態数を減らすために，決定木クラスタリングを用いる手法を提案する．一動素あたりの学習サンプル数が増え安定な学習が行われる．また，未知の動素 HMM のパラメータの推定が可能になり，未知の動素を含む単語も認識できるようになる．

以下，第 2 章では従来手法における認識アルゴリズムを，第 3 章では従来法の課題を，第 4 章ではその解決手法としての決定木クラスタリングと適用方法について述べる．そして第 5 章で評価実験のコンディションとその結果を，第 6 章で結論を述べる．

2 動素を用いた手話認識

佐藤ら [5] の動素単位を用いた手話認識について説明する．

2.1 概要

手話を右手・左手各々の「手の位置・動き」「手の形」「手首の向き」の計 6 種類の要素の時系列データの組みで表現し (例：表 1)，それぞれの要素ごとに動素 HMM を構成する．そして各構成要素の HMM の出力尤度の重み付き和を用いて認識結果を得ている (例：図 1)．

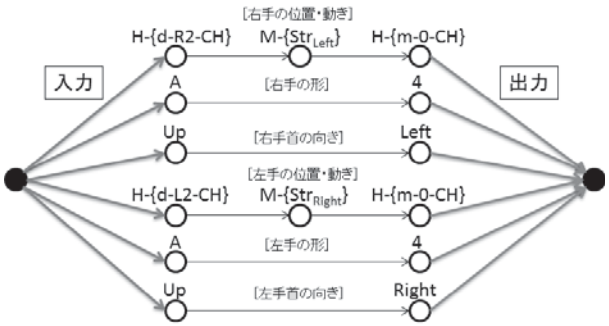

図 1　“afraid” のパラレル HMM．1 つのノードが 1 つの HMM を表す．

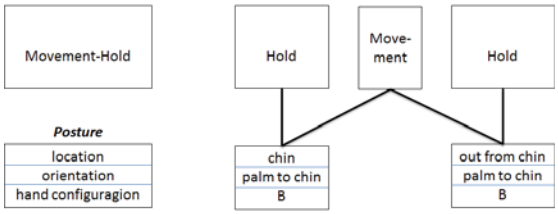

図 2　“Thank you”の Movement–Hold モデル

2.2　動素単位

手話単語を動素の時系列データの組で表現するにあたり，動素を定義する必要がある．佐藤 [5] は Liddel[8], Vogler[4] らの動素単位を参考に，以下の **Movement–Hold モデル**を用いている．

Liddel ら [8] は，アメリカ手話を **Movement–Hold モデル**という形式で記述した．これは，手話単語を **Hold**（静止）と **Movement**（動き）の繋がりによって表すものである．Hold にはそれぞれ **Posture** と呼ばれる素性が付属しており，この素性に「手の形」や「手の位置」といった手話単語の詳細が記述される．例として，アメリカ手話の “Thank you” の利き手側の Movement–Hold モデルを図 2 に示す．この例では Posture において **location**（手の位置），**orientation**（手の向き），**hand configuration**（手の形）の 3 つを記述している．

また，この研究では手の位置を表す座標系として手話話者を中心とした円柱座標系を採用している．図 3 は手話話者を上から見下ろしたものである．手の位置の定義の詳細を以下に示す．

- 手話話者からの距離は，話者付近 [*p*], 話者からひじの長さ程度離れた位置 [*m*], さらにそれよりも少し離れた位置 [*d*], 完全に腕を伸ばした位置 [*e*] の 4 段階で定義する．
- 手話話者に対する角度に関しては，手話話者の目の前を [*0*] とし，そこから右に約 30 度毎に [*R1*], [*R2*], [*R3*] と定義する．左の場合も同様に [*L1*], [*L2*], [*L3*] と定義する．
- 高さに関しては，体の部分を基準に考える．*MO*(口), *CH*(胸), *AB*(腹) のような表記で定義される（図 4 参照）．

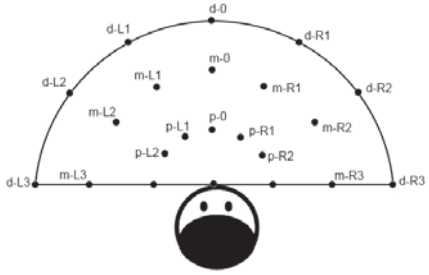

図 3　手の位置の極座標表現

以上 3 つの組み合わせで位置を表現する．例えば ***m-R1-CH*** は，胸の高さで角度は正面から約 30 度右，体からひじの長さ程度離れた位置を表すことになる．また，高さのみの表記の場合はその体の位置を直接指すことにする．

Vogler らは Liddel らの手話記述法を簡略化し，動素単位を定義している [3]．彼らは手の動きに関して M, H, X の 3 種類の連続したシンボルにより手話を表現している．ここで M は **Movement** で動きを, H は **Hold** で静止位置を表す．また，新たな記法である X は位置を表すのは H と同じだが静止はしないことを表し，主に手話の開始位置を表現する際に用いる．

表 1 に手話単語を Movement–Hold モデルで表した例を示す．例えば，“I” の手話は “$\boldsymbol{X\text{-}\{p\text{-}0\text{-}CH\}\ M\text{-}\{str_{Toward}\}\ H\text{-}\{CH\}}$” と表されるが，$X\text{-}\{p\text{-}0\text{-}CH\}$ が手話の開始時における右手の位置が胸の少し前であること，$M\text{-}\{str_{Toward}\}$ がそこから手前に右手を動かすこと，$H\text{-}\{CH\}$ が右手が胸の位置で静止することを示している．

2.3　手の動き・位置

「手の位置・動き」の記述法については上述の Movement–Hold モデルを採用する．Hold モデルでは Vogler[4] らの定義をほぼそのまま用いる．ある位置で

表 1　Vogler らの手話記述例（右手 Movement–Hold）

Sign	Transcription
I	X-{p-0-CH} M-{str_{Toward}} H-{CH}
man	H-{FH} M-{str_{Down}} M-{str_{Toward}} H-{CH}
woman	H-{CN} M-{str_{Down}} M-{str_{Toward}} H-{CH}

静止することを表す場合は H, ある位置から静止せずに開始することを表す場合は X とする．位置の表現は Liddel[8] らの円柱座標系の位置定義を参考にし，高さに関して簡略化したものを用いる（図 4）．手の動き，すなわち Movement モデルも，Vogler[4] らの定義を参考にし，以下の 7 種類を考える．

- **直線運動 (Straight)：**
  手をまっすぐに移動させることを表す．動かす方向は前 (Away), 手前 (Toward), 上 (Up), 下 (Down), 右 (Right), 左 (Left) の組み合わせで表現する．例えば $M_{StraightDownRight}$ は，手を右下に動かすことを表す．
- **短距離直線運動 (ShortStraight)：**
  上記の直線運動よりも短い運動を表す．動かす距離が 10cm 程度の時にこちらを用いる．動かす方向に斜めは考えない．
- **円運動 (Round)：**
  手の円弧運動を表す．最低でも 90 度以上動かす場合にこの表記を用いる．円運動の軸の方向（水平：Horizontal, 垂直：Vertical, 奥行き：Depth）及び回転方向（時計回り：Veering, 反時計回り：Backing）の組み合わせで表現する．
- **局所振幅運動 (ShakeIterate)：**
  手首による手の振幅運動を表す．振幅の方向を，水平 (Horizontal), 垂直 (Vertical), 奥行き (Depth）のどれかで表現する．
- **局所円運動 (ShakeCircle)：**
  手首による手の円運動を表す．円運動の軸の方向を水平 (Horizontal), 垂直 (Vertical), 奥行き (Depth）のどれかで表現する．
- **手首の回転運動 (Twist)：**
  手首を回転させるだけの場合に用いる．
- **短距離直線連続運動 (Swing)：**
  短距離直線運動の連続で表される手話単語が多いので，別に定義した．例えば，$M_{SwingUpDownUp}$ は {$M_{ShortStraightUp}$ $M_{ShortStraightDown}$ $M_{ShortStraightUp}$} と同値である．

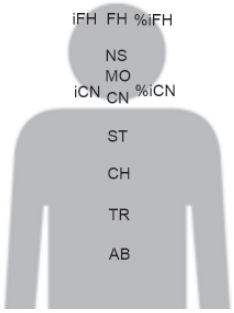

図 4　体の位置のシンボル定義（提案手法）

それぞれの手話単語の「手の位置・動き」は Movement と Hold の並びで表現される．定義例を表 2.3 に示す．

2.4　手の形

「手の形」は Stokoe の定義 [7] を簡略化した 13 種類に分類する．これらは指文字の ‘A’, ‘B’, ‘C’, ‘H’, ‘I’, ‘K’, ‘S’, ‘V’, ‘Y’, ‘1’, ‘4’, ‘8’, ‘9’ である．これ以外の指文字は近いものに分類する．「手の形」は手話単語の中で同じ形のままであるとは限らない．例えば “understand” という手話の場合，最初は指文字 ‘S’ の形であるが最後には指文字 ‘1’ の形になる．このような場合には，シンボル列によって「手の形」の変化を記述する．記述例を表 2.3 に示す．

2.5　手首の向き

「手首の向き」は手の位置が同じでも異なる場合が存在するため，手話認識において手話単語を区別する情報となる．「手首の向き」は前 (Away), 手前 (Toward), 上 (Up), 下 (Down), 右 (Right), 左 (Left) の 6 種類のみとする．それ以外の方向については 6 種類のうち最も近いものにする．手話単語中での「手首の向き」の変化を，シンボル列によって表現する．

3　従来法の課題

佐藤らの研究 [5] での動素定義では，手話の構成要素全てについて理論的に取りうる動素の種類数は合計で 550 ある．内訳は「手の動き・位置」が 531 種類,「手の形」が 13 種類,「手の向き」が 6 種類となっている．

「手の動き・位置」の動素の 531 種類のうち，実際に使われている動素は，258 単語のコーパスにおいて，右手で 163 種類，左手で 112 種類にとどまる．また，コー

表 2　手話記述例（右手の位置・動き）

Sign	Transcription
meet	X-{m-$R1$-ST} M-{$Straight_{Left}$} H-{m-0-ST}
day	X-{m-$R1$-NS} M-{$Round_{DepthAxisBacking}$} H-{m-$L1$-CH}
cool	X-{m-$R1$-ST} M-{$ShakeIterate_{Depth}$} H-{m-$R1$-ST}
apple	X-{iCN} M-{$Twist$} H-{iCN}
school	X-{m-0-ST} M-{$Swing_{DownUpDown}$} H-{m-0-CH}

表 3　手話記述例 (右手の形)

Sign	Transcription
I	1
my	B
mother	4
understand	S 1

パスで使われている動素を見ると，出現回数が数回となるものが多い．認識率を押し下げる要因になっている.「手の動き・位置」について，動素の種類が過剰で，過学習が起き，認識性能が低下している．そこで「手の動き・位置」についての過学習を防ぎ，HMM の複数の状態間でそのパラメータを共有して，動素あたりの学習量を増やすことで認識率が向上することが期待される．なお,「手の形」「手の向き」は種類数が少ないため，過学習が起きていないと考えられる．

4　動素の決定木クラスタリング

動素の HMM の状態の共有関係を定めるために，主にトライフォン HMM を使った音声認識で用いられる決定木クラスタリングを用いる．

4.1　アルゴリズム

決定木クラスタリングは，トップダウン型のクラスタリング手法である．はじめ，HMM の状態をひとまとめに扱い，事前に定義する「質問リスト」をもとにクラスタを分割していく，

事前に木の分割方法を記述した「質問リスト」を用意する.「質問リスト」の質問は，

「円運動かつ，時計回りか？」

「短距離直線運動かつ，水平方向への運動か？」

「空間を 6 分割したときに，あるエリア X に手が存在するか？」

「空間を 48 分割したときに，あるエリア Y に手が存在するか？」

などである．動素の類似性による分割方法が記述され，全ての動素は，各質問に対して yes, または no の答えをもつ．

初め，全てのトライフォンは木のルートノードにまとめられている．これは，全ての HMM の状態がひとつに統合され一つのガウス分布で表現されていることに相当する．次に木構造の「葉」の部分に質問を適用し,「葉」を二分割しそれぞれに対しガウス分布をあてはめ，そのパラメータを推定する．このとき，二分割されたあとの分割前後の尤度差を最大にする質問を選択する．このような尤度最大規準で質問を選択し，初めは大雑把にクラスタリング，あとになるにつれ細かいクラスタリングを行う．二分割されたノードは新たな「葉」となる．これは HMM の状態数が増えたことに相当する．こうした二分割を繰り返し，二つの子ノードの尤度の和の親ノードの尤度からの増分が，予め定められたしきい値を上回ったところで状態の分割を止める．そのときの HMM の状態の共有関係をもとに，新たな HMM パラメータを最推定する．

4.2　動素を用いた手話認識への応用

トライフォンを用いた音声認識では，決定木の分割の際「右の音素が鼻音であるか?」といった，左右の音素の種類による分割を行なっている．決定木クラスタリングをモノフォンを用いた手話認識に適用させるには，分割方法を工夫する必要がある．本研究では，手話の動素の種類そのもので分割方法を指定する．動きに関しては，動作の類似性に着目した．位置に関しては．空間的に分割したときに，手がどのエリアに含まれるかに着目した．動きと位置それぞれに対し，大きな分割から小さな分割まで，適切と思われる質問を用意した．表 4 に質問リストを載せる．

決定木クラスタリングの適用例をあげる，前述 2.2 の動素のうち，$MstrU$，$MstrD$，$MstrR$，$MshkC$，$Mtwi$，$MrndHB$，$MswiUDU$ の 7 つを分割する．それぞれ,「鉛直上方向への直線運動」「鉛直下方向への直線運動」「水平右方向への直線運動」「局所円運動」「手首回転運動」「水平方向反時計周り円運動」「上下上短距離直線運動」を表す．

図の「直線運動の動素か？」という質問に対して，それぞれの動素は以下の yes/no のどちらかの答えを持つ．

yes $MstrU$, $MstrD$, $MstrR$

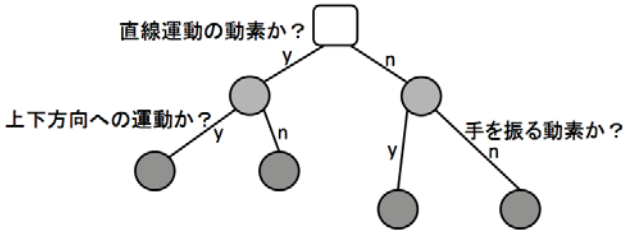

図 5　決定木クラスタリングにおける決定木の構造

no $MshkC$, $Mtwi$, $MrndHB$, $MswiUDU$

これで，もともと 1 つの状態だった 7 つの動素が，3 つの動素を持つ yes グループと 4 つの動素を持つ no グループの 2 つの子ノードに分割されたことになる．この分割を繰り返し，図 5 の決定木を適用させた結果が以下のようになる．

グループ 1：$MstrU$, $MstrD$
グループ 2：$MstrR$
グループ 3：$MswiUDU$
グループ 4：$Mtwi$, $MrndHB$

5　評価実験

5.1　評価方法

以下の 2 通りの，手話の孤立単語認識実験により提案手法の評価を行った．

55 単語の認識実験

手話話者 1 名が 414 単語について 4 回づつ収録したデータを全て学習に用い，55 単語について 4 回づつ収録したデータを全てテストに用いた．テストに用いた 55 単語は全て，学習データに含まれておらず，また，学習データにない動素を少なくともひとつ以上含んでいる．

469 単語の認識実験

手話話者 1 名が 414 単語について 4 回づつ収録したデータを全て学習に使い，上で述べた未知の 55 単語に学習データを加えた 469 単語について，それぞれ 4 回づつ収録したデータを全てテストに用いた．

なお，比較のため，Data-Driven のクラスタリング手法を用いたときの認識結果も示した．Data-Driven の

表 4　決定木クラスタリングで用いた質問の一覧:動き

動素の種類	質問
円運動	円運動か？
	かつ時計回りか？
	かつ反時計回りか？
	かつ軸が水平方向か？
	かつ軸が鉛直方向か？
	かつ軸が奥行き方向か？
局所運動	局所運動か？
	かつ局所振幅運動か？
	かつ局所円運動か？
	かつ軸が水平方向か？
	かつ軸が鉛直方向か？
	かつ軸が奥行き方向か？
直線運動	直線運動，または短距離直線運動か？
	直線運動か？
	短距離直線運動か？
	右方向への直線運動か？（右斜め方向も含む）
	左方向への直線運動か？（同上）
	上方向への直線運動か？（同上）
	下方向への直線運動か？（同上）
	奥方向への直線運動か？（同上）
	手前方向への直線運動か？（同上）
	右方向への直線運動か？（右斜め方向は除く）
	左方向への直線運動か？（同上）
	上方向への直線運動か？（同上）
	下方向への直線運動か？（同上）
	奥方向への直線運動か？（同上）
	手前方向への直線運動か？（同上）
短距離直線運動	短距離直線連続運動か？
	かつ軸が水平方向か？
	かつ軸が鉛直方向か？
	かつ軸が奥行き方向か？
手首回転運動	手首の回転運動か？

表 5　決定木クラスタリングで用いた質問の一覧:位置

動素の種類	質問
位置	空間を 6 分割したときどこに位置するか？ （距離 1 分割/角度 3 分割/高さ 2 分割）
	空間を 12 分割したときどこに位置するか？ （距離 2 分割/角度 3 分割/高さ 2 分割）
	空間を 24 分割したときどこに位置するか？ （距離 2 分割/角度 3 分割/高さ 4 分割）
	空間を 36 分割したときどこに位置するか？ （距離 4 分割/角度 3 分割/高さ 3 分割）
	空間を 42 分割したときどこに位置するか？ （距離 2 分割/角度 7 分割/高さ 3 分割）
	空間を 48 分割したときどこに位置するか？ （距離 2 分割/角度 3 分割/高さ 8 分割）
	空間を 48 分割したときどこに位置するか？ （距離 4 分割/角度 3 分割/高さ 4 分割）
	空間を 56 分割したときどこに位置するか？ （距離 2 分割/角度 7 分割/高さ 4 分割）
	空間を 84 分割したときどこに位置するか？ （距離 4 分割/角度 7 分割/高さ 3 分割）
	空間を 96 分割したときどこに位置するか？ （距離 4 分割/角度 3 分割/高さ 8 分割）
	空間を 96 分割したときどこに位置するか？ （距離 1 分割/角度 3 分割/高さ 2 分割）
	空間を 112 分割したときどこに位置するか？ （距離 2 分割/角度 7 分割/高さ 8 分割）
	空間を 112 分割したときどこに位置するか？ （距離 4 分割/角度 7 分割/高さ 4 分割）
	あごに手を触れる位置か？
	頭の頂点に手を触れる位置か？

プロパティ	スペック
深度画像解像度	640 × 480
フレームレート	30 fps
深度観測可能距離	1.2〜3.5 m （実用範囲）

表 6　Microsoft Kinect のスペック

クラスタリングは，はじめ HMM の状態をモノフォンの種類ごとに別々として扱い，HMM の状態の平均ベクトル間のユークリッド距離を規準に状態をまとめていく，ボトムアップ型の手法である．

5.2　データベース

手話の収録には Depth カメラの 1 つである Microsoft Kinect（表 6）を用いた．Microsoft Kinect からは，手話者の骨格推定の情報と深度画像が得られる．(図 6)

手話話者 1 名が 469 単語について 4 回収録し，のべ 1,876 単語のデータベースを作成した．なお，本研究では Shotton らの手法 [1] によって手話話者の頭・首・肩・ひじ・手の 3 次元座標の推定を行う．推定の精度を目視で確認し，不適当なサンプルはデータから取り除いた．

図 6　Kinect から得られる深度画像と骨格推定結果

5.3　特徴量

手の位置・動き

手の 3 次元座標とその時間差分の計 6 次元のベクトルを用いた．なおこの時の座標は，首の座標を中心とした座標系に変換したものを用いた．

X 系	3
H 系	10
M-{ShortStraight} 系	4
M-{Shake} 系	15
M-{Swing} 系	15
上記以外の M 系	8
それ以外	10
手の形	10
手首の向き	10

表 7　手の位置・動きの HMM の状態数

手の形

手の位置の座標からある一定距離にある領域を手の領域とし，その領域を XY 平面に射影する．射影した 2 次元形状の Hu モーメント [9] を計算して得られる 7 次元のモーメント及びその時間差分の計 14 次元のベクトルを用いた．なお，Hu モーメントはスケール変化や回転に対して不変であることが知られている．

手首の向き

ひじの座標から手の座標へ向かう 3 次元ベクトル及びその時間差分の計 6 次元のベクトルを用いた．

状態数

動素を left-to-right 型の skip なし HMM でモデル化する．動素は種類によって複雑さが異なるため，予備実験を行い各々の動素 HMM の状態数を表 7 のように設定した．

5.4　動素 HMM

動素 HMM の混合数を 4，各混合数における学習回数を 9 回とした．クラスタリングの前には，各動素を H（静止位置）・X（開始位置）の種類ごとに分類し，異なる種類の動素の混同を防いだ．事前実験から，M から始まる動素（手の動き）はクラスタリングをすると認識率が下がることが分かったため，M から始まる動素はクラスタリングをしない．これは，M から始まる動素の種類数が少ないためと考えられる．

統合時の手話の尤度の要素間の重みは，それぞれを 0, 0.25, 0.5, 0.75, 1 と変化させて，最も良い単語正解率を達成するものを選択した．クラスタリングの尤度は，0 から 60 まで，5 刻みで変化させた．尤度の値を大きくすると，得られるひとつひとつのクラスタも大きくなる．なお，実験には HTK[10] を用いた．

クラスタリングの種類	単語認識率	クラスタリング後 HMM 総状態数	クラスタリング前 HMM 総状態数
決定木	73.6 %	777	1507
Data-Driven	69.1%	1099	1507

表 8　55 単語の認識実験　結果まとめ

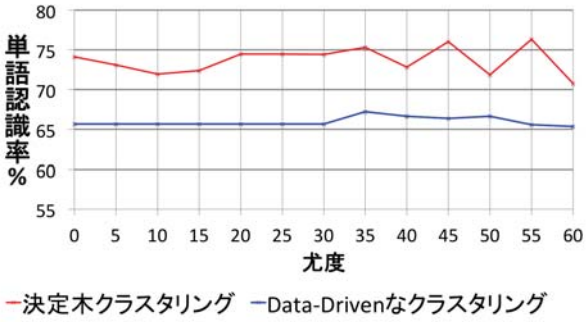

図 7　各クラスタリングによる認識率の推移

5.5　実験結果

5.5.1　55 単語の認識実験

未知動素を 1 つ以上含む，未知単語 55 をテストに用いて認識を行う実験の結果を 8 に示す．決定木クラスタリングを用いた方が，Data-Driven のクラスタリングを用いた場合よりも単語認識率が 4.5 ポイント上回った．

5.5.2　469 単語の認識実験

決定木クラスタリングを用いた際の，単語認識率の推移のグラフを図 7 に示す．縦軸は単語認識率，横軸はクラスタリングの際の尤度である．認識率が最大となったときの，動素の H モデルにおける決定木の構造を図 8 に示す．決定木クラスタリングを用いた方が，Data-Driven のクラスタリングを用いた場合よりも単語認識率が 9.2 ポイント上回った．

6　おわりに

動素 HMM を用いた手話認識において，従来法では過学習が見られた．そこで本研究では，音声認識で使われる決定木クラスタリングを，動素を用いた手話認識に応用した．これにより，過学習を防ぎ頑健なモデル選択ができるようになり，また，未知動素を含む未知単語も認識できるようになった．未知動素を含む未

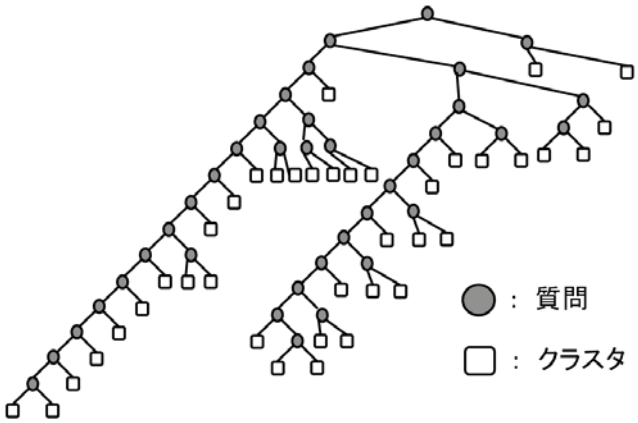

図 8　認識率が最大となったときの決定木の構造

知の 55 単語の認識では，Data-Driven のクラスタリングと比べ 4.5 ポイントの優位性がみられ，学習に使った 414 単語に未知の 55 単語を加えた計 469 単語の認識では，Data-Driven のクラスタリングと比べ 9.2 ポイントの優位性がみられた．数値的な近さを規準にまとめていく Data-Driven のクラスタリングよりも，動きの類似性に着目してまとめていく決定木クラスタリングの方が優れている，ということが示された．また，動素を用いた大語彙手話認識の，語彙拡張の可能性が示すことができた．

今後の課題として，左右の動素への影響を考慮したトライフォンモデルを導入すること，両手を同時に動かす・手を交互に動かす・等の動素の同期等の制約を考慮すること，認識精度の低い「手の形」の検出方法の向上，過学習の起きない動素定義の見直し，実用化に向けての連続手話認識への拡張　が考えられる．

参考文献

[1] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, “Real-Time Human Pose Recognition in Parts from Single Depth Images,” *In Proc. of CVPR,* 2011.

[2] 有賀 光希, 酒向 慎司, 北村 正, “日本手話の音韻構造を考慮した HMM に基づく手話認識,” 電気情報通信学会技術報告, PRMU2010-112, pp. 127-132, 2010.

[3] C. Vogler and D. Metaxas, “Toward Scalability in ASL Recognition: Breaking Down Signs into Phonemes,” *In Proc. of Gesture Workshop,* pp. 211-224, 1999.

[4] C. Vogler and D. Metaxas, “A Framework for Recognizing the Simultaneous Aspects of American Sign Language,” *Computer Vision and Image Understanding,* vol. 81, pp. 358-384, 2001.

[5] 佐藤 新，篠田 浩一, “手話素単位を用いた大語彙手話認識,” 電気情報通信学会技術報告, PRMU2011-222, vol. 111, no. 430,pp. 155-160, 2012,

[6] 神田 和幸, 基礎から学ぶ手話学, 福村出版, 2009.

[7] W. C. Stokoe, “Sign Language Structure: An Outline of the Visual Communication System of the American Deaf,” *Studies in Linguistics: Occasional Papers* (No. 8), 1960.

[8] S. K. Liddel and R. E. Johnson, “American Sign Language: The Phonological Base,” *Sign Language Studies,* 64:195-277, 1989.

[9] M. K. Hu, “Visual Pattern Recognition by Moment Invariants,” *IRE Trans. Info. Theory,* vol. IT-8, pp. 179-187, 1962.

[10] HMM Tool Kit (HTK), `http://htk.eng.cam.ac.uk/`.