

## 手話マルチモーダルデータベース

竹内 勝, 佐川浩彦

技術研究組合 新情報処理開発機構 (RWCP) マルチモーダル機能日立研究室

手話認識研究においては、手動作・表情情報の把握、手話認識アルゴリズムにおけるパラメータの決定、手話文法解析のためのデータベースが必要となる。手話データの解析には、手話発話の映像データのみでは不十分であり、映像データに加えて、詳細な手動作情報を得るためにセンサグローブデータと手話情報の表現内容、表現部位、表現区間を示すタグ情報を統合したデータベースを構築した。これにより、映像データのみでは得られない情報を解析し、手話認識方式の高度化が可能となる。本稿ではデータベースのデータ体系と自然な発話データ収集のための発話指示方法と照明方法について報告する。

## Sign Language Multi-modal Database

Masaru Takeuchi, Hirohiko Sagawa

Multi-modal Function Hitachi Laboratory, RWCP

Sign language database is necessary to develop sign language recognition technologies. Since hand gesture and facial expression represents words and grammatical information, it is essential that the database include manual motion data and facial expression change data. In this paper, the database scheme and the high quality data acquisition are described. The constructed database comprises natural conversation data which integrates video, signal by globe based input device and label which indicate what type of motion are performed and meaning of motion. In the future, we expect that many researchers use this database for the progress of sign language understanding.

### 1. はじめに

聴覚障害者の自然な手話発話において手話単語は、手話の辞書や手話教育用のテキスト記載された手話単語の動作が直接表現されない場合が頻繁に起こる。また、手話は単に手話単語を逐次的に並べて単語情報を表現するだけではない。手指動作の変化は文法情報をも表現する<sup>(1)</sup>。さらに、手話は手指動作だけではなく、頭部の動き、口形などほぼ上半身全体を使って情報伝達を行う<sup>(2)</sup>。本稿では、頭部の動き、表情など、手話における手指動作以外の情報表現を行う動作を非手指動作とよぶ。

このように聴覚障害者の手話は手指動作により手話単語を表現するだけでなく、手指動作と非手指動作に基づく情報を通じて文を表現している。本稿では質の異なる（情報）という意味でマルチモーダル（情報）という用語を用いる。

筆者等は、認識対象を文法情報・非手指動作情報にまで拡げ、マルチモーダル情報を統合することにより、聴覚障害者の自然な手話発話を認識することが出来る手話認識方式の研究を行っている<sup>(3), (4)</sup>。研究過程においては、聴覚障害者による自然な手話発話に関する映像データの収集と解析が必要となる。こうした過程で作成される手話発話データを手話マルチモーダルデータとよぶ。

本稿では、手話発話データベースのデータ体系、自然な手話発話データの収集方式について述べる。

## 2. 手話マルチモーダルデータベースの体系と利用目的

手話マルチモーダルデータベースの体系を図1に示す。

## データベースは

- (1) センサグローブを装着しない状態での手話文映像データ
  - (2) センサグローブを装着した状態でのセンサグローブ信号データおよび手話文映像データ
  - (3) センサグローブを装着した状態でのセンサグローブ信号データおよび手話単語映像データ

により構成される。ここでセンサグローブとはバーチャルリアリティなどで使用される手袋型の手指動作の入力装置である。センサグローブはビデオレートと同じ 1/30 秒間隔で掌の位置・方向と各指の関節の曲げ角を収集する。

(1) では手指動作情報を検討するための正面および横方向からの上半身の映像データ、非手指動作

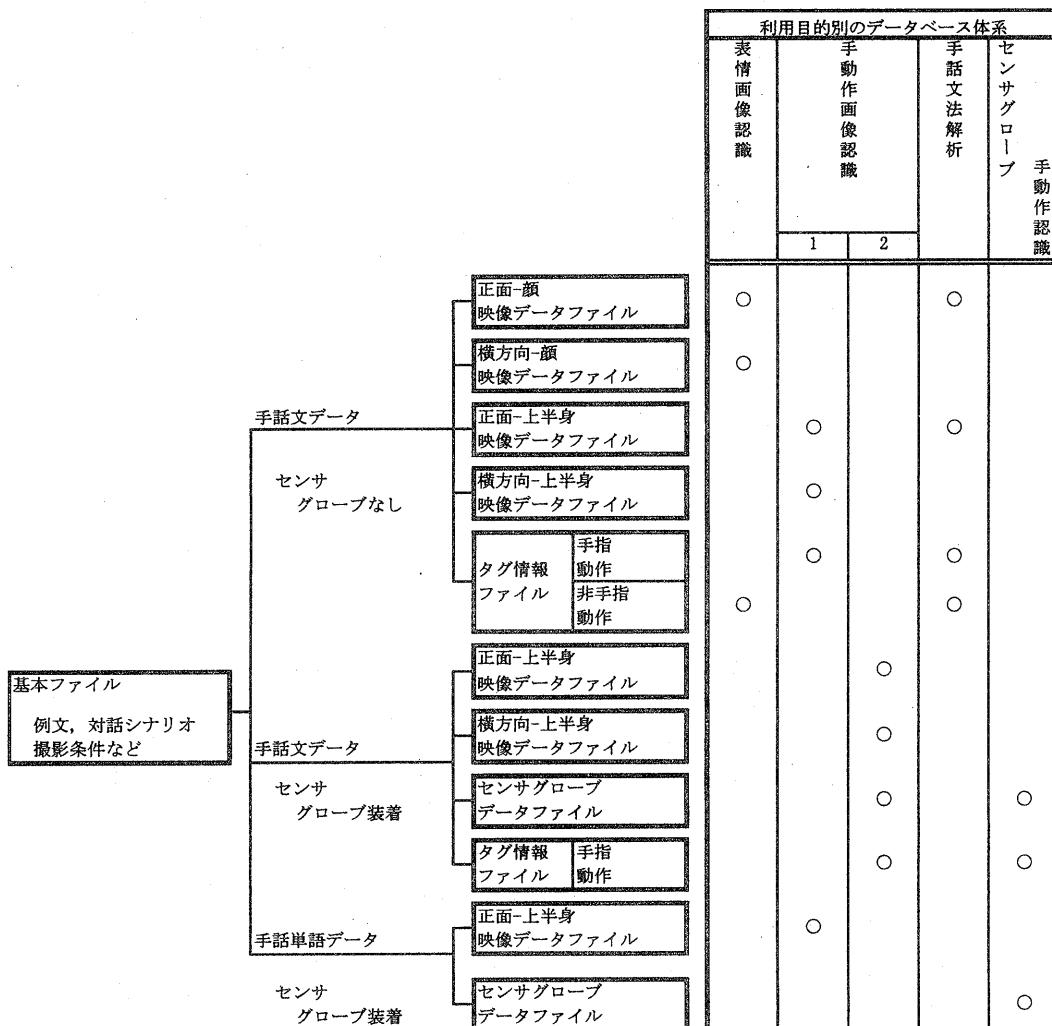


図1：手話マルチモーダルデータベース体系と利用目的

情報を検討するための正面および横方向からの頭部の映像データを撮影している。また、(2)、(3)では正面および横方向からの上半身の映像データを撮影している。

そして、収集した映像データは JPEG 形式で計算機に入力し、複数の映像データ間の同期情報とタグ情報を作成した。タグ情報としては、手話発話において、どの体の部位（表現部位）をもちいて、どのような手指動作・非手指動作情報（表現内容）が、発話中のどの時点で開始され、終了したか（表現区間）を示す情報を与えている（図 2）。

手話マルチモーダルデータベースは、映像情報だけでなく、タグ情報ファイル、センサグローブデータファイルを持ち、対応付けていることに特長がある。これにより、表情画像認識、手動作画像認識、手話文法解析、センサグローブによる手指動作認識など、以下のような、多目的利用のための体系をもつ。

- (1) 手指動作、非手指動作に関するタグ情報ファイルには認識対象となる手話表現がどの範囲で提示されたかが明示されている。これにより、正面と横方向の顔映像データファイルを用いた手動作画像認識や表情画像認識アルゴリズムの検証を行うことができる。特に手動作画像認識については、センサグローブデータにより、映像データと手の位置・方向、各指の関節角との関係を詳細に検討することも可能である。
- (2) 正面と横方向の顔映像データファイルと上半身映像データファイルを参照し、タグ情報ファイルを用いて、文法規則の抽出や仮説の検証など手話文法解析研究用データベースとしての利用が可能である。
- (3) センサグローブによる手指動作認識研究においては、タグ情報ファイルの援用により手動作の詳細な解析を行うことができる。

### 3. 手指動作・非手指動作に関するタグ付け方式・内容に関する検討

収集した複数（4種類または2種類）の映像データは JPEG 形式で計算機に入力する。そして、複数の映像データの同期をとるための情報と手指動作・非手指動作のタグ情報を作成する。

図 2 にタグ情報ファイルの内容の一例を示す。映像データには、先ず、聴覚障害者と手話通訳者によりオフセット値が与えられ、複数の映像データの時間的な同期がとられる。そして、手指動作・非手指動作が表現されているフレーム番号を確認し、その表現区間を定め、当該動作のタグ付けを行う。

実際にタグ情報ファイルの作成を行うと、表現区間の開始点（フレーム番号）および終了点（フレーム番号）を正確に定めることが非常に困難であるため、タグが示す情報をその内部に含んでいればよいという緩い条件の下で表現区間を定めている。

手指動作についてはその手話単語としての意味をタグとして与えている。図 2 の手指動作のタグにおける PT1, PT3 は各々一人称、三人称を差すポインティング動作である。

現在のところ手話の文法規則は仮説の域を出ないものが多い。このため、非手指動作については文法機能の名称ではなく、非手指動作自体の名称を与えている。また、タグ情報ファイルにおける非手指動作は表現部位と表現内容を示す略称で与えている。

図 2, 167 フレームから 184 フレームまでの頭の向きにおける「あよか」は頭部（あ）が横（よ）に傾く（か）非手指動作を表す。また、326 フレームから、338 フレームまでの頭の向きにおける「あうな」は頭部（あ）が頷く（うな）非手指動作を表し、これは肯定文文末を示しているものと考えられる。

(同期情報)						
オフセット値1:	-72	頭向き:	あうな	185	192	
オフセット値2:	-73	頭向き:	あうな	265	276	
オフセット値3:	-134	頭向き:	あよか	296	313	
オフセット値4:	-75	頭向き:	あうな	326	338	(肯定文文末)
(表現部位、表現内容)		頭向き:	あうな	405	414	
手動作: ある	113	124	頭向き:	あ下	434	445
手動作: 夢	166	178	頸:	ア上	114	127
手動作: 言う	184	190	眉毛:	ましか	299	313
手動作: 黒沢	219	239	瞼:	めまば	121	129
手動作: P T 3	241	250	瞼:	めまば	164	170
手動作: 監督	256	269	瞼:	めまば	188	192
手動作: 映画	279	284	瞼:	めまば	266	272
手動作: 字幕	294	304	瞼:	めまば	289	295
手動作: 希望	307	313	瞼:	めまば	328	333
手動作: P T 1	317	321	瞼:	めまば	410	415
手動作: 後で	419	423	瞼:	めまば	424	426
手動作: 連絡	426	432	瞼:	めまば	440	446
手動作: お願い	436	442	視線:	し下	241	247
頭向き: あよか	167	184	口形:	くふく	113	125

(表現区間)

図2: タグ情報ファイルの内容

表1に非手指動作情報に関するタグの一覧を示す。備考として、その非手指動作が表現しているものと推定される機能を記載した。

非手指動作のタグ情報を概観すると、タグの表記のゆれが生じることと、発話者の癖や生理現象と手話における非手指動作との区別がなされていないという、タグ付け方式に関する問題点が明らかになる。

頭部の動きのタグについては、例えば、同一の非手指動作を「あうな(肯き)」と記する場合と「ア引(頸引き)」と記するというように、動作の表現内容の表記にゆれがみられる。典型的な頷く動作と頸を引く動作とは異なった動作であるが、頸を引く方向が下方に向い、引いた状態からすぐにもとの状態に戻ると頷く動作との区別がつきにくくなる。このために表記にゆれがみられるものと考えられる。

眉を上げる動作と目を開く動作が同期する場合や、頸を下げる動作と眉を上げる動作と口形「才」を提示する動作が同期する場合など、複数の非手指動作が連動して表現される例もある。前者は目を開く動作に付随して、顔の構造上、必然的に眉を上げる動作が起きたものと考えられる。一方、後者の表現においては顔の構造上の必然性はなく、これら3つの動作の組み合わせで疑問文を示す文法情報を表現しているものと考えられる。

次に、発話者の癖や生理現象とみなせる非手指動作の区別を行うことが必要であることも明かになった。発話者によって頷きの頻度に違いがみられる。この頷きは文法情報を発信しているわけではなく、単に発話者の癖であるものと思われる。高頻度に頷く発話者の場合でも、肯定文文末の頷きは発話者の癖と思われる頷きに比べ動作が大きく、ゆっくりと提示されている。一方、瞬きはかなり頻繁に起こるが、それが非手指動作であるのか生理現象であるのかの区別は難しい。

表1：タグ情報一覧

No.	部位	タグ	内容	備考
1	頭向き:	あうな	頭傾き	助詞, 時間継続長, 肯定文文末, 単語の区切り
2	頭向き:	あたふ	頭を縦に振る	
3	頭向き:	あよか	頭を横に傾ける	単語の区切り, 考える
4	頭向き:	あよふ	頭を横に振る	否定, 疑問
5	頭向き:	あよむ	頭横向き	
6	頭向き:	あ下	頭を下げる	依頼, 疑問, 文を切る, よびかけ
7	顎:	ア引	顎を引く	
8	顎:	ア下	顎を下げる	
9	顎:	ア上	顎を上げる	助詞, 時間継続長
10	顎:	ア前	顎を前に出す	
11	体向き:	か前	体を前傾させる	
12	眉毛:	ましか	眉をしかめる	
13	眉毛:	ま上	眉上げ	疑問, 話題化, 依頼
14	頬:	ほふく	頬を膨らませる	
15	瞼:	めうす	うすめ	感情表現, 強調表現
16	瞼:	めばた	瞬き	
17	瞼:	め開	目を開く	疑問, 話題化, 依頼
18	瞼:	め閉	目を閉じる	
19	視線:	しなう	視線斜め上	考えている状態
20	視線:	しなし	視線斜め下	
21	視線:	しよ	視線横	
22	視線:	し下	視線下	
23	視線:	し上	視線上	
24	視線:	し対	視線対象	
25	口形:	くイ	イ	
26	口形:	くエ	エ	
27	口形:	くオ	オ	疑問
28	口形:	くパ	パ	過去, 不可能
29	口形:	くふく	口を膨らませる	疑問
30	口形:	く引	口を引き締める	
31	口形:	く開	口を開く	
32	口形:	く舌	舌を出す	
33	口形:	く尖	口を尖らす	
34	口形:	く閉	口を閉じる	

#### 4. 聴覚障害者の自然な発話データ収集方式の検討

聴覚障害者の自然な発話データの収集のために発話指示方法と照明方法について検討した。

##### (1) 発話指示方法

聴覚障害者に発話してもらう方法として、(1)自由会話方式、(2)例文発話方式、(3)対話シナリオ方式、(4)ビデオ参照方式の4つの方式を検討した。なお、(2)～(4)については、表2の発話内容、3

名の発話者についてのデータを収集し、収集方式の検討に用いている。

自由会話方式は、二人以上の聴者、または、聴者と手話通訳者に自由に会話をするよう指示し、その状況を撮影する方式である。例文発話方式は、聴者に日本語文で発話内容を提示し、発話状況を撮影する方式である。対話シナリオ方式は、二人の対話形式でのデータ収集方式であり、対話内容を説明した上で、プロンプターと呼ばれる対話相手と聴者の会話状況を撮影する方式である。ビデオ参照方式は始めに自然な手話発話が現れるまで例文発話を繰り返し、次に、自然な例文発話の映像データを参照しながら聴者が発話をを行い、その状況を撮影する方式である。

収集するデータには発話の自然性が要求される。一方、認識実験や発話内容の解析の対象となる情報は何度か繰り返して発話してもらうこと（これを繰り返し発話とよぶ）も要求される。以下、4つのデータ収集方式に関して発話の自然性と繰り返し発話について検討する。自由会話方式は発話の自然性は自明であるが、繰り返し発話をしてもらうためには、事前に、双方のまたはどちらか一方の発話者に発話内容を限定する指示を与える必要がある。

一方、例文発話方式は繰り返し発話は可能であるが、発話の自然性は満たされないことがある。

聴覚障害者には手話を母語とする人と、後者は日本語を母語とする人がいる。手話を母語とする聴覚障害者は手話で思考を行う。3名の手話発話者は手話を母語とするが日本語についても十分な理解力をもつ。このため、日本語で例文を示し、その内容を手話で発話してもらう例文発話方式では、日本語の表現の影響を受け、手話による思考が阻害され、自然な発話と差が生じる場合が多い。

そのため、収集する発話内容を示す日本語文の前後に発話内容に合わせた日本語文を追加し簡単なシナリオを作成する（表3）。作成したシナリオに基づき日本語文で提示した内容を手話で発話させることが自然な発話を収集するための1つの方策であると考える。このデータ収集方式を対話シナリオ方式とよぶ。この方式は、発話内容は日本語例文であらかじめ提示することができるため、繰り返し発話が可能になる。また、手話による対話の中でデータ収集を行うため、手話による思考を維持したままの発話も可能となる。

表2：データ採取法式の検討に用いた例文

No.	手話単語	発話内容
1	産まれる	昨日、妻に子供が産まれたんです。
2	英語	外国の友達がこれから引越してくるので、手続きのための英語の資料はありますか。
3	鉛筆	この書類は鉛筆で書いても大丈夫ですか。
4	女	すみません。女性問題を考えるの、公演の会場は、どちらですか。
5	事務	公務員試験の申し込みをしたいのですが、事務職の年齢制限は、いくつですか。
6	上手	すみませんが料理を上手になりたいので近くにやすく習える料理の教室がありましたら教えてください。
7	過ぎる	すみませんが、ずっと家を空けたままなので、この通知は昨日見たばかりでもう期限が過ぎてしましましたが、お願ひできますか。
8	夢	「夢」という黒沢明監督の映画に字幕をつけて欲しいです。
9	理解	何とか理解できました。ほかの質問があります。
10	理由	成人式の講演に手話通訳が付かないなんてどういう理由があつてのことなんですか。

表3：対話シナリオの例

プロンプター 発話者	市民センターの映画会で観たい映画はありますか。 「夢」という黒沢明監督の映画に字幕をつけて欲しいです。
プロンプター 発話者	相談してみます。 後日連絡を下さい。

ビデオ参照方式も自然な発話を撮影する上では有効であり、繰り返し発話も可能である。しかしながら、参照用データの真似をする、という行為が自然性を阻害する要因になることが懸念される。

図3に手話マルチモーダルデータベースのデータ収集・作成フローを示す。手話マルチモーダルデータベースでは聴覚障害者の自然な手話発話の撮影は対話シナリオ方式をベースとし、ビデオ参照方式を援用している。はじめに、会話例文から対話シナリオを作成し、発話者へ対話内容を説明する。その対話シナリオの練習を行い、参照用データを作成し、データ収集に入る。そして、収集したデータの自然性を検討する。このとき、必要に応じて参照用データとの比較を行う。データ収集は自然な手話発話であると判断されるまで繰り返し行われる。

## (2) 照明方法

聴覚障害者の手話発話の手指動作は動きが速く、高速のシャッター速度での撮影が必要となる。シャッター速度が速くなると光量が少くなり画面が暗くなる。このため、照明機材を使用する必要が生じる。手話マルチモーダルデータベースのデータ収集方式においては、撮影した映像を再生し、自然な発話が行われた聴覚障害者が判断するまで繰り返す。このため、撮影時間が長くなる傾向がある。通常、映像データの撮影にはタンクスティン灯型の照明装置や白熱灯（デイライト）型の照明装置を用いる。被験者に直接光りを当てずに間接光を用いたり、すりガラス等で光を拡散させた場合でも、室温が上昇する。特に、照明装置に近い位置において、直接光が当たる撮影対象となる被験者付近の温度が上昇し、発汗等により被験者の負担が増加し、長時間の撮影が行えない。

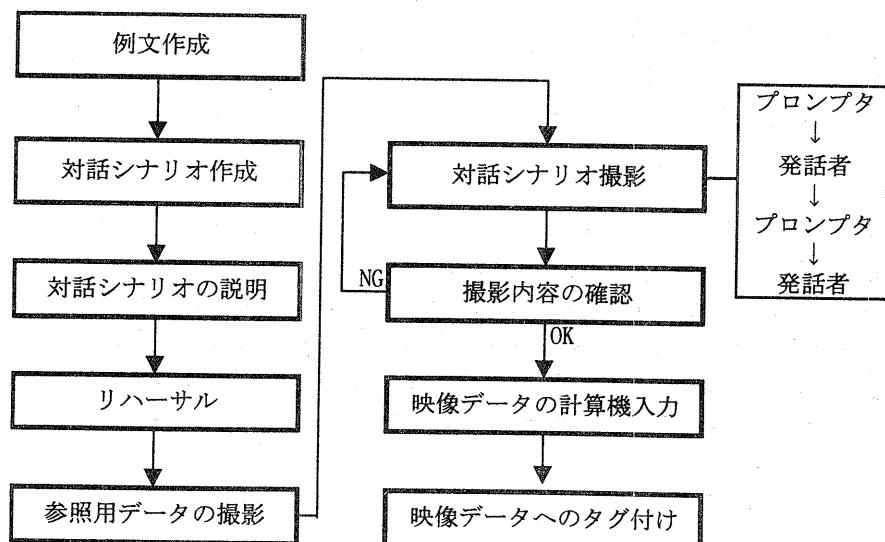


図3：手話マルチモーダルデータベースのデータ収集・作成フロー

このため発熱量の少ない安定器付きの高周波蛍光灯(75,000Hz)の使用を検討した。高周波蛍光灯は数万ヘルツで明滅する光源である。一般に高速のシャッター速度の下では光源の明滅の影響を受け動画像における1枚ごとの画像の明るさに違いが生じる場合がある。画像処理・認識を行う上では1枚ごとの画像の明るさに違いがないことが望ましい。

高周波蛍光灯を使用し、シャッター速度を変化させ撮影を行った。撮影に使用したデジタルビデオカメラは通常(オート設定時)はシャッター速度が1/60秒で毎秒30枚の映像を撮影する。絞りをF2.0に固定し、シャッター速度が1/60秒の場合は画面の明るさの変化はなく、明るさ自体も十分であるが、手話の手指動作にはかなりのぶれが生じており、動画像中の1枚の静止画像を逐次処理していくタイプの画像認識アルゴリズムで使用には向かない。シャッター速度が1/1000秒の場合は画面の明るさの変化はなく、光量も十分であり、手指動作にもぶれが生じていない。シャッター速度が1/10000秒になると多少の画面の明るさの変化が生じ、光量は不十分になる。

手話マルチモーダルデータベースでは1/1000秒のシャッター速度を用いて撮影を行っている。

## 5. おわりに

画像認識や手話文法解析等、多目的に利用可能であることを特長とする手話発話データベースのデータ体系と高品質データの収集方法について検討し、以下の結果を得た。

- (1) 手話マルチモーダルデータベースのデータ体系は映像データ、センサグローブデータ、タグデータを統合している点に特長がある。本データベースにより、聴覚障害者の自然な手話発話の手動作・表情画像認識や手話文法解析が可能となる。
- (2) 聴覚障害者の自然な発話に関する実験用データ収集方式として、対話内容を説明した上で、プロンプターとよぶ対話相手と聴覚障害者の会話状況を撮影する対話シナリオ方式を考案し、長時間にわたる映像データ収集の際の照明として、高周波蛍光灯を使用した場合、1/1000秒のシャッター速度で手動作のフレームごとの明るさの変化がない映像データが収集可能であることを確認した。

今後は、手話文法規則の抽出を行い、手話文法を利用した手話認識方式の研究・開発を行う。手話文法規則の抽出には大規模な聴覚障害者の自然な手話発話に関する映像データの収集、タグ情報の作成が必要となる。本データベースの体系に従って、映像データの収集、タグ情報の作成を行っていく。

## 参考文献

- (1) 市田靖弘: 日本手話の文法と語彙, 日本語額, Vol. 13, pp. 25-35(1994)
- (2) 木村晴美, 市田泰弘: はじめての手話, 日本文芸社(1995)
- (3) H. Sagawa and M. Takeuchi: Methods to Describe and Recognize Sign Language Based on Gesture Components Represented by Symbols and Numerical Values, Knowledge-Based Systems(1998).
- (4) H. Sagawa and M. Takeuchi: A Method for Analyzing Spatial Relationship between Words in Sign Language Recognition, The 3rd Gesture Workshop(1999).