

手話・身振りインタフェース構築の現状と課題

鎌田一雄

宇都宮大学工学部情報工学科
〒321 宇都宮市石井町2753

あらまし： 人の身体動作を認識する、あるいは、逆に映像として身体の動作を合成／提示する機能を備えたインタフェースは、コミュニケーション環境整備の上で、その重要度は今後とも増していくと考えられる。ここでは、ジェスチャーとことば（手話）の要素であるサインを対象とした映像処理（認識と合成）技術の開発、システム構築において考えるべきいくつかの事項を、機器の利用者である人間の特性、および目指す機器の位置づけを中心に、概説する。手話・身ぶりの認識／合成・提示するインタフェース機能を持つシステムは、社会生活の中で色々な波及影響を与えることが予想される。特に、手話を日常生活で使用する聴覚に障害を持つ人達へのコミュニケーション支援の役割は大きい、他の障害者の支援の可能性も大きい。

Issues on Designing Gesture and Signed Language Interface

Kazuo Kamata
Department of Information Science,
Utsunomiya University
2753 Ishii-machi Utsunomiya, 321 Japan

Abstract: This paper overviews certain issues that we must consider in realizing gesture and signed language recognition/generation functions for a interface. In the first part of the paper, we show certain fundamental features that differentiate gestures and signed languages. In the latter part, we show some issues that we need to take into account in developing technologies for the interface design. Among them are; structures of signed languages, and purpose or aim of the research.

1. まえがき

コミュニケーション環境の整備は、情報処理・通信技術の進歩とともに多様化、高機能化の道を歩んできている。対面会話環境における物理的、時間的な制約から解放された新しい形態の会話（メディアコミュニケーションとも呼ばれている）へと移りつつある。

現在までに、種々の研究機関等で人間の身ぶりを認識、あるいは合成することを目指した研究が行われている。ここでは、映像メディアの利用を想定したメディアコミュニケーション環境実現において、重要な役割を果たすことが期待されている人間の身体動作を機械が認識する、あるいは身体動作を映像として合成/提示する機能を持つインタフェース構築の基礎について概説する。特に、身体動作を身振りとして、さらには言語的な機能をそれらの動作が持つ手話の2つに焦点を合わせ、これらの動作を介して情報交換が可能な環境を実現するために考えるべき課題を述べる。しかしながら、画像・映像処理の技術、あるいは処理手法を直接の対象とはせず、このような機能を有する機器を利用するユーザ（人間）と、開発される機器（あるいは機器開発の目的・目標）の社会における位置付け（社会環境）、の2つの視点から技術開発の過程で考えておくべき項目をいくつか選んで述べることにする。

2. いくつかの課題

2. 1 基本的な事項

(1) 身ぶりの研究は、古い歴史を持っている。ここでは、コミュニケーションにおける身体動作などの解析などの分野には特に立ち入らない [17,18]。ノンバーバルコミュニケーション (nonverbal communication) という呼び名で規定されている研究対象が、この分野では大きな位置を占めている。ところが、この広く使用されている用語が、手話 (signed languages) に代表される音 (音響信号) を使用しないことば (silent languages) に対する考え方に、多分に曖昧さを与えているようにも感じる。ことばのモダリティ (modality)、機能から見ると、以下のような分類が手話を的確に捉えている [4]。

表1. ことばの分類

| Mode / Function | Linguistic | Nonlinguistic |
|-----------------|---|---------------------------------|
| Vocal | spoken languages sound | paralinguistics stress, etc. |
| Nonvocal | <u>signed languages</u> manual/nonmanual behaviours | paralinguistics gestures |

(2) また、手話と音声言語とのことばの表出・受容の違いについては、表2のようにまとめることができる [7]。音声語はその単語構成に恣意性 (arbitrariness, conventionality) があることを大きな特徴としている。手話の単語 (後述の sign words) には写像性が強いことが知られているが、図1に示すように写像性と恣意性は連続したものである [12]。

表2. 音声語と手話との比較

| | Spoken language | Sign language |
|-----------------------|--|--|
| Instrument Production | Air, larynx, tongue, etc. Articulated speech phonemes of a language | Hands, body, etc. Articulated signing cheremes of a language |
| Reception Units | Ears Morphemes, words, etc. | Eyes Morphemes, signs etc. |
| Combination | Grammar of a given spoken language | Grammar of a given sign language |
| Universal features | Only certain types of structures occur. Each spoken language uses a subset of these structures. Features also occur in sign languages. | Only certain types of structures occur. Each sign language uses a subset of these structures. Features also occur in spoken languages |

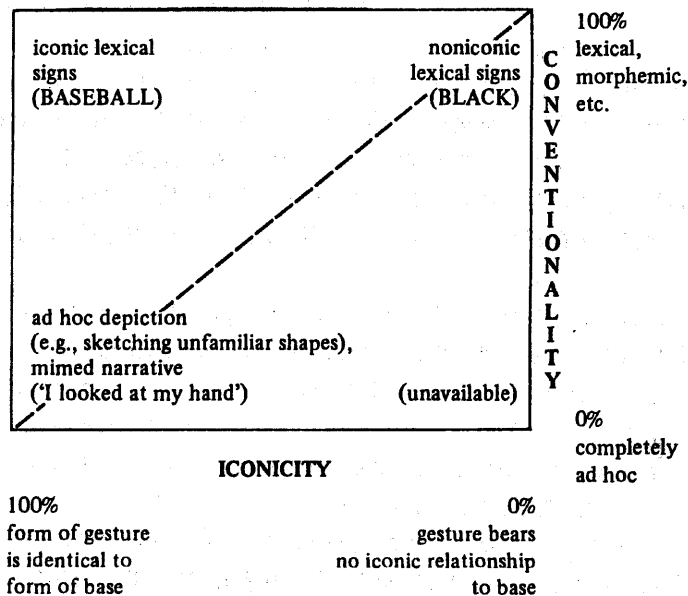


図1. ことばの写像性と恣意性

(3) *Gesture* と *Sign* という2つの用語を論ずるためには、記号論(*semiotics*)などの基礎も必要となるかとも思われる [6]。特に、用語 *Sign* は、色々な意味に(色々な定義の下に)使用されているが、本文では、手話の単語(*sign word*)を表す言葉として用いることにする。これに対して、*Gesture* はコミュニケーションで使用される身体動作すべてを意味するものとする。*Gesture* 中の特別な集まり(部分集合)が *Sign* を構成している。

2. 2 考えるべき課題

(1) 動作の構成要素は何か? (ことばの仕組みI)

よく知られているように音声に関する研究には多くの研究成果があり、ことばの要素の定義と(フォルマント周波数などの)音響信号の物理的に計測可能な特徴量との対応も、かなり明らかになっている部分がある。身ぶり・手話の認識などにおいても、このようなことばの要素と(計測が可能な)物理量との対応を明確にすることが必要となる。

手話の場合は、ことばの構成要素(音素、あるいは単語などの)が何であるか(何を構成要素と定義できるか)が、当然のことながら、重要である。手指動作(*manual behaviours*)の中では、手の形態(*hand configuration*)、手話動作の位置(*place of articulation*)、動き(*movement*)、身体と手との向きの関係(*orientation of the hands in relation to the body*)の4つに分けることができる [5, 13]。なお、向きを除いた3つを構成要素(パラメータ)とする考え方もある [19]。

これらの手指動作については、言語学の分野で多くの議論があり、共通的な認識が得られている部分もある。しかしながら、手の形態の中における手の形(*handshapes*)については、かなりの解析が行われ、最も整理が進んでいる。動きについては、まだまだ十分とはいえない部分が多いようである。後述するように、手話における動きが持つ意味は非常に重要である(項目(5)参照)。

認識と合成処理のどちらを検討の対象とするかによっても、決めるべき構成要素には違いが生じる。認識の場合は、それぞれの単語間との違いが物理的に計測可能な要素を用いて明確に記述されることが必要である。これに対して、合成(映像提示)の場合には、ことばの知覚に必要な特徴(要因/要素)の詳細な表示(記述)が(機械で制御が可能な構成要素を用いて)できることが要求される。

ところで、*Gesture* あるいは *Sign* の認識、合成の際に考えなければならない他の要因として、顔の表情などの(手話でいうところの)非手指動作(*nonmanual behaviours*)がある。手話で使われる非手指動作の(機能的な)解析は、いくつか記号化されているものもあるが、残念ながらほとんど進んでいないという状態である。コンピュータによる顔(*facial expressions*)の研究は、多様な分野で議論され、研究が進められている。このような研究、技術開発は言語学の分野でもツールとして期待されている [2, 8]。

Gesture を対象とした認識・合成の分野では、もちろん適用分野/目的にも強く依存するが、構成要素と動作に対する意味付けはシステムの利用者が作り出したもの、ときにはシステム設計者が作り出すものが利用できることもある。このような研究開発環境は、ことばとしての *Sign* を対象とした研究である認識、合成処理の場合とは大きく異なることを理解する必要がある。*Gesture* を対象とするならば、例えば、腕の(分離性に優れた)動作のみを用いた指示動作

(pointing) 等でインタフェースを構築することも可能である。

sign の場合には、形態的な構造についても検討しなければならない。手話の場合は、ことばの表出者(手話者)前面を中心とする手話空間(signing space)での空間的な動作を利用する。このようなことばの表出特性にも因ると考えられるが音声語よりも形態的な変化が、より大きな役割を果たしている傾向がある [9, 10, 16] .

(2) 対象とすることばの形態はどれか？ (対象言語)

手話の認識・合成(映像の生成)技術開発では、対象とすることば(対象言語)を決めなければならない。「どれにするか？」は、後述のように実際に機器を使用する人達(ユーザ)との関係から非常に重要な意味を持つ。

ろう者が日常生活で使用している手話は、それ自体音声語とは独立したことばであることは、広く認められている事実である。しかしながら、社会で大多数の人達が使用することばは音声語である。これら2つのモダリティがまったく異なることばの接触が社会で起きている。そこには音声言語の場合と同じように多様な言語形態が存在することは容易に類推できる。よく使われる説明図であるが、音声語と手話との接触で構成されることばは、ことばの連続体(language continuum)を構成する [3] .

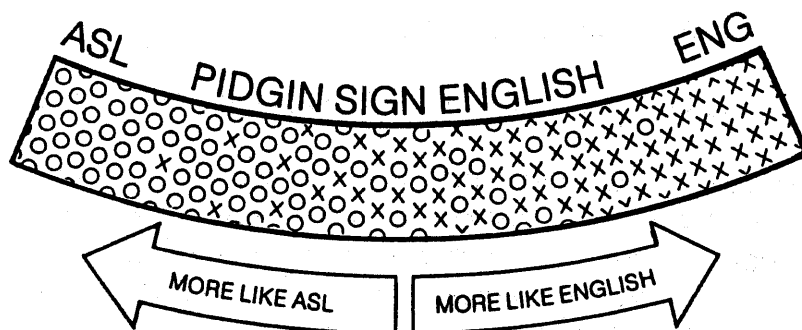


図2. ことばの連続体

ことばの接触から生まれるピジン手話(pidgin sign language, PSL)を使用する人達も多い。また、社会生活の中でフォーマルな場では多数集団のことばに近いことばを使用し、インフォーマルな場では日頃のことばを使用するというダイグロシア(diglossia)と呼ばれる現象も存在している [11] .

このような社会(言語的な)環境の下では、ろう者の本当のことばを見るのは難しいともいわれている [11] . ろう者は相手のことばの特性、能力などに応じて使い分けていることもよく知られた事実である²¹。認識対象とする言語の特徴/形態を十分に知らないままに処理技術の開発を

²¹ 手話学習のサークル(教室)などで手話を教えているろう者は、教室内で聴者に対して使用する手話と、ろう者と話をする時の手話とは、まったく形態が異なることがよくある。

進めることは難しい。また、合成映像に現れることばの形態が、想定した利用者のことばのそれと大幅に異なるならば、理解が難しいという状況も起きる。これらの現象は、「設計者と利用者とのギャップ」とも呼ぶこともできるが、実生活の中に開発機器を根付かすためには、幅広くかつ深い検討が必須である。

(3) 何のために構築するか？ (技術の社会適応性)

最近の手話の認識／合成に関する研究は、工学の分野に新しい研究分野を拓くという意義を認め多くの研究が実施されているようである。手話認識・合成機能を備えたインタフェース (人間と機械 (機械の後ろの人間も含む) との情報のやりとり) は、さらには、障害を持つ人達を含めた多くの人達に対する日常生活におけるコミュニケーション支援、および情報 (発信・獲得) 環境の整備という目標が挙げられている。

技術開発が目指すものが本当に何であるかということが明確でないと、特に福祉の領域では、当初の目的を十分に果たせないものとして終わってしまうことが危惧される。これまでに、技術開発としては成功しながら、どうしても、当初に掲げていた福祉への適用 (目的) が果たせないままに終わったものは多い。身ぶり、手話の認識あるいは合成の機能を持つシステムが、社会生活のどのような局面で活用されることを目指しているか、また、実効性が十分にあるか、という検討は常に必要である。

なお、利用者の多数が障害を持つ人達である場合には、その機器の使い勝手 (機器のユーザ・インタフェース) が、従来にも増して大きな課題となる。

(4) どこまで機能を実現すればよいか？ (期待と実現技術・機能とのギャップ)

手話、身ぶりの認識／合成は、ことばの認識／合成の1つである。音声言語の機械による処理 (自然言語処理) を考えるならば、完全な処理はかなり難しいことが容易に理解できる。次の問いは、音声言語を対象とした (関連した) すべての研究において考えなければならない：

「インタフェース／システムが実現すべき機能において、どの程度の (処理) 能力を有するならば、 (限定された使用環境のもとで) 実用的なシステムとみなしてもらえるか？」

特に、新しい分野である身ぶり、手話については共通的に判断する基準すら確立されていない状況にある。しかし、システムがそれなりの存在意義を得るためには、たとえ個別的であろうとも、システム評価は必須である。また、評価に基づいた妥当で受け入れられる使用環境の設定も必要である。

(5) ことばの知識は十分か？ (ことばの仕組みII)

項目 (1) に述べた形態素レベルの構成要素の外に、ことばの文法に関する知見がことばの認識においては必須である。手話の表現上の豊かさから考えると、機械による音声語を対象とした自然言語処理 (書きことばの処理) が検討してきた経緯から見ると、かなり高度なレベルを最初から考えなければならない。会話音声の認識については、最近、活発な研究が行われている [14]。手話を対象とする機械による認識では、このような分野の知見が技術開発の最初から要求されている。

手話の語順に関する検討を考えると、単純に手指動作による単語表現の特徴だけから検討することが難しいことは容易に理解できる。顔の表情などの非手指動作の果たす役割を十分に考慮しなければ語順の検討も十分に実りある結果を得ることは難しい [8]。

また、手話単語の品詞に関する知見も、単語の機能を解析する上で重要な項目である。アメリカ手話(American sign language, ASL)の検討結果からは、動詞と名詞を区別する特徴として動きの様態の違いがあることが分かっている [20]。

3. あとがき

本稿では、Gesture, Sign の利用を想定し、その認識/合成・提示機能を備えたインタフェース構築で必要となる、身体動作の構成要素、利用環境など技術開発過程で十分に検討すべき課題の中からいくつかを選んで説明した。映像処理などの要素技術は、かなり高性能・高機能なものを我々は持っていると考えられる。しかし、コミュニケーション形態の多様性、多機能化をさらに進めるためには利用者である人の特性などを十分に把握し、優れた整合性を実現することが、今後の研究では必須な要因である。また、よくいわれる開発者と利用者とのギャップの解消も、これを実際に実現する手法も含めて、これまで以上に重要な要因となる。社会生活における価値観の変化などを考慮すれば、今後の技術レベル(技術力)の大きな評価尺度となると予想できる。

本文の内容は、映像処理などの技術的要因を抑え、人間的な側面から見たときに必要な課題の説明を試みた。この分野での研究・検討に、多少でも参考になれば幸いである。

なお、手話全般に関しては、文献 [1, 15] などを参照されたい。

謝辞

日頃、障害者のコミュニケーション支援、社会環境に関連する検討で、ご協力、ご支援を頂いている、関係各位に深く感謝する。

文献

本文を書くために直接参考としたものに限って記載している。下記の文献以外に、数多くの関連する研究論文、著書等がある。手話に関する研究論文はデータベース化されており、ネットワーク上で検索できるようになっている (<http://www.sign-lang.uni-hamburg.de/bibweb/>)。ここから最新の手話(手話学)に関連する文献情報を得ることができる。

- [1] Armstrong, D.F., Stokoe, W.C., and Wilcox, S.: Gesture and the nature of language, Cambridge University Press(1995).
- [2] Backer-Schenk, S.: The facial behavior of deaf signers: Evidence of a complex language, American Annals of the Deaf, 130, 4, pp.297-304(Oct. 1985).
- [3] Baker-Schenk, C. & Cokely, D.: American sign language- A teacher's resource text on grammar and culture, Gallaudet University Press

- (1980).
- [4] Fischer, S.D.: アメリカ手話における言語. 非言語的特性、信学技報 HCS96-39 (1997-01).
- [5] Friedman, L.A.: Formational properties of American sign language, in Friedman, L.A. (Ed.): On the other hand, New perspectives on American sign language, pp.13-56, Academic Press (1977).
- [6] 池上嘉彦: 記号論への招待、岩波新書258 (1984).
- [7] Kyle, J.K., and Wool, B.: Sign language - The study of deaf people and their language, Cambridge University Press (1985).
- [8] Liddel, S.K.: American sign language syntax, Mouton Publishers (1980).
- [9] Liddel, S.K.: Think and believe: Sequentiality in American sign language, Language, 60, 2, pp.372-399 (June 1984).
- [10] Liddel, S.K.: Structures for representing handshape and local movement at the phonemic level, in Fischer, S.D., and Siple, P. (Eds.): Theoretical issues in sign language research, pp.37-65, The University of Chicago Press (1990).
- [11] Lucas, C., and Valli, C.: Language contact in the American deaf community, Academic Press (1992).
- [12] Mandel, M.: Iconic devices in American sign language, in Friedman, L.A. (Ed.): On the other hand, New perspectives on American sign language, pp.57-107, Academic Press (1977).
- [13] Newkirk, D., and Battison, R.: Properties of symbols in a silent language, in Klima, E., and Bellugi, U.: The signs of language, pp.35-66, Harvard University Press (1979).
- [14] 音声文法研究会編: 文法と音声、くろしお出版 (1997).
- [15] Sacks, O.: Seeing voices, Picador (1990).
邦訳は、佐野正信訳: 手話の世界へ、晶文社 (1996).
- [16] Sandler, W.: Temporal aspect and ASL phonology level, in Fischer, S.D., and Siple, P. (Eds.): Theoretical issues in sign language research, pp.7-35, The University of Chicago Press (1990).
- [17] 荘巖舜哉編: ノンバーバル行動の実験的研究、川島書店 (1993).
- [18] Schere, K.R., and Ekman, P. (Eds.): Handbook of methods in nonverbal behavior research, Cambridge University Press (1985).
- [19] Stokoe, W. Jr., Casterline, D.C., and Croneberg, C.G.: A dictionary of American sign language on linguistic principles, Gallaudet College Press (1965).
- [20] Supalla, T., and Newport, E.: How many seats in a chair, in Siple, P. (Ed.): Understanding language through sign language research, pp.91-132, Academic Press (1978).