

[招待講演]
福祉と音声研究者

市川 壱[†]

†千葉大学 〒263-8522 千葉市稻毛区弥生町1-33

E-mail: †ichikawa@faculty.chiba-u.jp

あらまし 代替手段を用いる障害者の心的負担は大きく、単にアクセシビリティの視点だけでは実質的に使えない。ユーザビリティの配慮が重要である。対面対話は最も基本的コミュニケーション形態であり、ユーザビリティの実現に重要な視点を与えると考えられる。その意味でも、対話言語研究開発の先駆者である音声研究者の福祉分野への貢献が期待される。対話言語は、揮発性でありながら、何万語もある語彙の世界の中で、瞬時に単語などをセグメントし、切り出された言葉の相互の関係（意味）を理解し、円滑に話者が交代してゆくメカニズムが存在している。また言語情報だけでなく、誰がどのような気持ちで話しているかなどの情報も重要な役割を果たしている。これらの機能はいわゆるプロソディ情報が担っている。手話（視覚言語）や指點字（触覚言語）も音声言語と同様に揮発性の言語であり、これらプロソディの機能についても論じる。

キーワード 福祉、揮発性、プロソディ、対話音声、手話、指點字

Assistive Technology and Speech People

Akira ICHIKAWA[†]

† Chiba University

E-mail: †ichikawa@faculty.chiba-u.jp

Abstract Usability is a very important view point of the assistive technology for persons with disabilities. It is considered that any utterance of dialogue, spoken language, has functions enabling real-time understanding to hear itself and to control dialogue smoothly in spite of its volatile characteristics. Prosody plays a very important role not only in spoken dialogue (aural language) but also in sign dialogue (visual language) and finger Braille (tactile language). A native user of a language may detect word boundaries in an utterance and estimate its sentence structure immediately and smooth exchange of turns using prosody. Speech people are expected to play a role in the domain.

Keyword assistive technology, volatile, prosody, dialogue language, sign language, finger Braille

1. はじめに

2001年WHO総会で採択された人間の生活機能と障害の分類法「国際生活機能分類 ICF」では[1]、障害=活動制限、社会的不利=参加制約が、環境すなわち社会によって引き起こされるという非常に重要な観点に基づき規定されたものである。情報環境は現在社会環境の中で非常に大きな部分を構成している。

また、情報発信権や情報取得権など、情報環境への参加を可能とすることは、基本的人権であるという認識が国際的のも認められている。

その情報環境との情報の交換の中心は言語情報であるが、現代社会では書記言語の役割が大きい。しかし、日常生活における最も基本的なメディアは音声と言ってよいだろう。障害者・高齢者のコミュニケーション支援においては、音声のような使い慣れた対話型メディアに対する直接的利用要求は高い。したがって、音声研究開発者の対話型メディアである音声に関する知見や音声技術開発の経験は大きな役割を果たしており、また将来に対する期待も大きい。

コミュニケーション関連の障害者支援における主要な課題の一つとして、障害者が情報を取得可能とするためのチャネルとして、代替チャネルを用いなければならないことがあげられる。しかし、代替手段を用いざるを得ない

ということは極めて負担の大きい操作であり、単にアクセシブルというだけでは実用上全く不十便であり、健常者の機器以上にユーザビリティを考慮しなければならない。書記言語と他の対話型のメディアとの変換や、その時のそのモダリティ特性の持つ長所と短所及び障害者・高齢者の知的活動の特性を十分把握し対応することが重要である。この側面への音声研究者の今後の積極的な取り組みも期待される。

また、逆に手話など障害者の分野から音声研究にヒントとして与えられる知見も多い。

2. 障害者・高齢者

2. 1 日本の障害者

心身障害は大きく身体障害、知的障害、精神障害に分類され、わが国では18歳以上の障害を持つ人の概数（障害者手帳保持者）は、身体障害352万人（内、肢体不自由180万人、内部障害86万人、聴覚・言語障害36万人、視覚障害31万人）、知的障害46万人、精神障害約258万人である。総人口比は約4%である。

さらに、上記の障害を複数持つ（重複障害）の人も多い。たとえば重複身体障害は18万人（上記に含まれている）で、そのうち視覚障害と聴覚・言語障害を重複するいわゆる盲ろうの人は1万7千人、聴覚・言語障害と肢体不自由重複は5万人、視覚障害と肢体不自由重複は2万9千人、肢体不自由と内部障害重複は5万1千人、その他の重複が3万3千人である（[2]ほかより推定）。

障害について考えるとき、上記の障害別分類のほか、その程度、障害の部位、先天性か後天性かなどを考慮する必要がある。また、その人のこれまでの生育環境、教育環境なども考慮に入れる必要がある。

2. 2 高齢者

65歳以上を「高齢者」と定義しており、75歳未満を前期高齢者、75歳以上を後期高齢者と呼んでいる。一般に高齢化率が7%を超えた社会を高齢化社会、14%を超えたときには高齢社会、21%を超えた社会を超高齢化社会といふ。¹

2007年1月時点での65歳以上の人口は全人口の21%を超え、日本は世界で最初に超高齢社会に突入した。

¹ 高齢化社会の定義は、国連が高齢者の問題について報告した際に、欧米で社会的に高齢者の問題がクローズアップされた人口比が概ね7%であったことを参考に提案され、その後、その倍の14%以上を高齢社会、3倍の21%以上を超高齢社会と習慣的にと呼ぶようになったとされている。

3. 特定領域「情報福祉の基礎」から

我々は新たな研究領域として「情報福祉学」という名称を与え文科省科学研究費特定領域研究「情報福祉の基礎」（領域代表 市川熹）による研究が平成16年度から18年度にわたって行った。その結果、以下のようなことが明らかになった。

一般的のヒューマンインターフェースの開発では、多くの場合、使用効率のような客観的尺度や、一般ユーザにとってどのように見えるかといった主観尺度を用いて評価され、開発されている。しかし、高齢者や障害者は、その障害の内容や程度が多様であるだけではなく、各ユーザの経緯や経験などの相違から、個人ごとの違いは極めて大きい。したがって、その機器を実際の場面で使用するときに、どのような負担を負っているかは、一人ひとりにより大きく異なる。

このようなユーザは、視覚や聴覚、肢体など様々に障害があり、代替手段によらざるをえない。そのため、その負担は健常者以上に大きく、単に使える（アクセシビリティ）だけでは実用上不十分であり、夫々の個人的状況に応じた使いやすさ（ユーザビリティ）を配慮したインターフェースの実現が望まれる。

認知心理学では「メンタルワーカロード」（心的負担）という概念があり、これが、障害の種類を問わず機器の使いやすさを表現できる可能性があると考えられる。ユーザ自身が機器使用時に自分が感じている心的負担を評価する主観的評価手法である。したがって、「メンタルワーカロード」という新たな共通の基本的視点に基づき、高齢者・障害者の問題を効果的に解決するための方向を示し、その体系的に開発を行うことには大きな意義がある。

個々のユーザの「心的負担」を軽減し、ユーザビリティを改善するためには、機器の見え方や操作の容易さだけでは不十分である。そこで扱われるメディアが実時間でやり取りされる言語情報として、認知的に最適化されているか、例えば手話を単に画像としてではなく、視覚的自然言語として理解しやすいか、円滑な対話を実現するための情報が具備されているかというようなことも重要な視点となる。さらに、言語というと、無意識のうちに書き言葉の性質を前提に発想してしまうが、最もプリミティブなコミュニケーションの形が実時間対話であるように、高齢者・障害者のコミュニケーション手段の形は実時間対話であり、そこでの言葉の本質は書き言葉とは大きく異なり、認知的性質も非常に異なる。ユーザビリティの視点からは欠かすことのできない視点である。

このように、音声に関する知見や経験は情報福祉の立場

から極めて重要であることがわかる。

4. 音声技術の福祉対応の現状

4. 1 音声出力

(1) 視覚障害支援

情報が紙で提供される時代には、視覚障害者は他の人に読んでもらう必要があった。電子データで提供され、音声合成技術や点字ディスプレイが開発されると、それらを利用して自立して情報を獲得する道が開けた。しかしGUIの多用やHPからの情報が増大している現在では、それらは2次元的に情報が表示されており、また画像情報も多用されているため、再び大きな問題となっている。画面表示の情報に対して、音声は1次元の揮発性の情報であり、単純に変換するだけでは、情報取得には大きな負担がかかる。

(2) 発声障害支援

脳性麻痺などで発話が困難な場合はVOCALと呼ばれる携帯用の機器を利用し、スイッチから文字を入力して合成音声に変換するものが色々開発されている。

(3) 人工喉頭

声帯を癌などで切除され、発声が困難な人には人工喉頭と呼ばれる機器を喉の外壁にあて、声帯波形の代行を行い、発声支援を行う[3]。

(4) 失声者の感情音声合成

単に合成音声で出力するのではなく、失声前の音声を分析し、個人性や感情の音質も再合成する試みも行われている[4]。

4. 2 音声入力

(1) 聴覚障害支援、盲ろう者支援

総務省は2007年までにニュースやスポーツ中継などの生番組等を除くテレビ番組に100%字幕を付加する目標に設定している。そこで、放送音声を認識し、字幕に変換付与する技術の開発が進めてきた。この技術は学校における講義をビデオに収録し、字幕を付けて聴覚障害学生が利用できるシステムなどの応用が可能である。

また、視聴覚重複障害者は文字も見ることができないため、音声認識結果を触覚情報である点字や指點字に変換する支援技術も開発が進められている。

(2) 脳性麻痺音声の認識

発声が不安定のため、聞き取りが困難な音声を認識し、聞きやすい合成音声に変換する技術の開発が期待される[5]。しかし、個人差も大きく、当事者は緊張すると発声がその都度様々に変動するため現状技術では認識性能が不足している。音声認識技術と強化学習などとの組み合わせた手法の開発など課題は大きい。

5. 音声関連技術と知見の応用

特に手話においては、同じく揮発性の対話メディアとしての音声と共通の課題が存在しており、先行する音声の研究開発手法は参考になる項目が多い。

5. 1 品質評価手法と符号化技術

音声品質の評価方法は、手話の画像伝送の品質評価に応用されている。しかし、手話では音韻レベルの単位の解明が未だ十分行われていないことや、空間に並列して情報が配置されている（両手や顔、身体などの動作が並行し組み合わされている）ことなどのために、現状では明瞭度の評価は現状では難しく、了解度試験が中心である。

このような手法を参考に、手話画像伝送における情報圧縮技術の開発が進められている[6]。

5. 2 伝送技術

音声の通話品質の課題の一つに、伝送系の遅延の影響の問題がある。手話伝送においても、画像処理の負荷が重いこともあり、遅延の影響の評価が必要になる。この課題に対する評価手法として音声通信における遅延評価手法が参考とされている[7]。

5. 3 手話の合成規則開発

音声の規則合成技術の開発では、音韻性の実現だけではなく、音韻の時間配置やアクセント、イントネーションなどプロソディの規則の開発が重要である。

手話の規則合成（アニメやCG）においても、聴者にとって読み取りやすい手話の合成には時間構造は重要である。平山は音声の時間構造の認知的基準点を見出す為に、メトロノームとの同期発声を分析した[8]。これを参考に、周期的振動と同期した聴者の手話を分析し、手話文の時間構造規則の手がかりを得ようという試みがなされている。

5. 4 手話CGの品質評価

合成音声の品質の良さの一つの視点として、どの程度余裕を持って楽に聞き取れるかというものがあり、その評価法として二重課題法が用いられる。二重課題としては音声の聴覚チャンネルとバッティングしないように視覚情報を用いることが多い。しかし聴者の場合は視覚で手話を見ており、聴覚は使えないでの、2課題をどのように設定するかに工夫を要するのが実態である。（最近では、二重課題法ではなく、NASA-TLXのような心的負荷の評価法[9]の利用の検討が進められている）。

5. 5 認識手法

HMM法利用の検討が進められている。手話では未だ音韻に相当する単位に関する知見が十分ではないため、模索の段階と言つてよいと思われる。

5. 6 認識性能評価手法

音声対話システムの評価法を参考に、単語認識率や理解率、完了率などの利用が検討されている。

5. 7 対話処理

情報の受け手の状態が送り手の情報表現に影響を大きく及ぼすことが見られる。例えば子供や外国人に話す場合は、相手の理解の程度を勘案して自然に話し方が変わることは極普通に見られる現象である。

このような対話の持つ特徴を活かすことは、障害者支援技術のユーザビリティを向上する上で大きなヒントを与えるものと期待される。対話形式の持つ可能性の例としては、当事者が必要な情報を問い合わせ、徐々に対象の持つ構造を脳内に再構築できる点である。

例えば、表は2次元構造で、各欄の情報は、縦および横の見出しの欄の内容と関係付けないと理解が出来ない構造になっている。数式も分数など複雑な2次元構造の上に、日本語と基本的表現が大きく異なる。このような表や数式を単純に一方的に音声に変換しても、内容を理解することは非常に困難である。ユーザが対話を通じて必要な情報を引き出し、再構築することが期待される[10]。

5. 8 文字盤入力形式における応用

肢体不自由者は文字盤操作が困難であり操作回数の少ない入力方法などが検討されている。例えば音韻の出現頻度情報やバイフォンモデルなどの応用である[11]。

6. 実時間メディアとしての音声

何故連続している音韻の列の中から言葉と言葉の境目が直ちに判るのか、何故脳の中に記憶されているであろう何万語、何十万語もある単語から特定の言葉が直ちに取り出せるのか、何故取り出された単語と単語の関係（文の構造）が直ちに判るのか、など、いずれも非常に不思議である。余裕を持って予測できるような仕掛けと、効率的な心内辞書のアクセス法が存在することが予想される。

また、単に情報が取得できたり、情報を発信できるだけでなく、会議などにおける物事の決定にリアルタイムで参加が可能になることも、人権や人格を尊重する上で重要な事項であろう。

しかし後天的に障害となり、手話や点字、指点字などの手段を迅速に使いこなせない当事者も多く、事前の希望聴取や決定事項の事後における了承の打診などが現状といえよう。会議システムの実現が必要である。

そのためには、音声による会議などでどのように情報が交換され、合意が形成されてゆくのか、速度に大きなズレがある場合、どのような問題が生じるのか、といった基本的性質の解明が、このための技術開発の基礎となる。

これらの性質を明らかにし、応用できれば、障害者や高齢者にとって負担の極めて小さいインターフェースの実現が期待される[12]。

6. 1 音声情報の再考

音声の持つ情報として、従来下記のような分類がなされてきた。

ア. 言語情報

イ. 周辺言語情報

ウ. 非言語情報

この分類は、発声された音声から最終的に得られる情報（伝達内容情報）といえよう。一方このような情報の知覚理解過程を支援する情報（伝達支援情報）も同時に存在し、伝達される情報の円滑な獲得を可能にしていると考えられる。

また、話し手の発声と平行してその音声は聞き手により聴取され、理解が進行し、反応が現れ、またその反応に応じて発声が影響されるという、ダイナミックな制御を行う情報が存在すると考えられる。話者交替の制御などの対話制御なども含む情報である。

A. 伝達内容情報 伝えられた内容

B. 伝達支援情報 實時間理解支援

伝達内容構造情報

伝達内容構造予告情報

話者交替予告（対話進行支援）情報

音声情報				備 考	
伝達内容		伝達支援情報			
従来 分類	伝達内容 情報	伝達内容 構造情報	予告情報		
言語 情報	音韻・音節 情報	音節構造 情報	調音結合	コ ミ ュ ニ テ イ	
	語彙レベル 情報	語彙・アクセ ント筋 情報	アクセント筋 分節予告情報		
周辺 言語 情報	文レベル 情報	文構造情報	係り受け予告 情報	個 人	
	レトリック 情報 (強調等)	レトリック 情報	? (未検討)		
—	—	—	話者交替予告 情報		
非言語 情報	感情情報	感情表現構 造情報	? (未検討)	個 人	
	個人性情報	? (未検討)	? (未検討)		

注：太字斜体の情報は主にプロソディによる。

われわれは、この伝達支援情報Bには様々なレベルの予告情報が含まれ、聞き手の知覚や理解の負担を軽減する非常に大きな役割を果たしていると考えている。ユーザビリティを上げるためにには、この情報Bの解明と活用が極めて重要である。また、心的辞書への高速なアクセス構造やコンテンツ構造の配慮も重要である。

6. 2 伝達支援情報

前節のような視点から、われわれは音声のような実時間対話言語には、仮説として、対話参加者が予測可能な階層的予告機能が存在するものと考える。例えば

- a. 話者交替の予告機能
- b. 文及び文章構造の予告機能
- c. 語彙レベル分節位置予告機能
- d. 後続音韻の予告機能

などである。

a～c はプロソディが重要な役割を果たしているであろう。また d は調音結合などが、その役割を担っているのではないか。

(1) 話者交替の予告機能

文の構造や発話終了か継続かの情報がイントネーションの中に先行して予告的情報が存在していること[13][14]などの事実が解かつてきている。

発話末の 1 アクセントフレーズ (A P) における基本周波数 F_0 、パワー、時間長などの韻律 (プロソディ) 特徴量のみを使用して、決定木学習により話者交替／継続の判別実験を行った。結果として、発話末 1 モーラに関するパラメータの影響を除いた場合と含めた場合とでは、正解率には殆ど差が無く、夫々 80% と 83% (クローズド実験) となつた。これらの結果は言語情報を全く用いていないことを考慮に入れると、ある程度高い精度で予測可能とする情報が文末以前のプロソディに存在することを示唆している[13]。

F_0 モデルのパラメータ α , β も変数として同時推定する手法を検討した。 F_0 をフィルタの応答として指數関数でモデル化したもので、入力のインバ尔斯の振幅を A_p とし、応答波形の減衰特性を α としてフレーズ成分を表現するものである (β はアクセント成分のパラメータ)。 A_p と α はフレーズの頭の部分で決まるものである。従来は α や β は定数として仮定されていた。

α , β を変数とした F_0 モデルのパラメータ推定は多パラメータを有した最適化問題であるが、A-b-S 处理に遺伝的アルゴリズムを利用することによって効率的な最適化を行うことが可能である。提案手法は初期値を得るための処理と最適化処理の 2 段階の処理を共に遺伝的アルゴリズムによる A-b-S を用いて行った。提案手法の精度をシミュレーションと実音声の実験によって検討した結果、良い精度が

得られた。この手法を用いて対話データにおいて話者交替の有無とパラメータの関係を統計的に分析し、 F_0 モデルのフレーズ成分のパラメータ α および A_p の影響を検討した。このとき、話者交替の起こる発話の場合には α は大きな値をとり A_p は小さな値をとるという結果が得られている[14]。

(2) 文及び文章構造の予告機能

音声の基本周波数のパターンを折れ線で近似し、その傾きと長さ、隣接する折れ線の終点と始点の差の値を用いて、音声の文構造が良く推定できる[15]。これを改良し、隣接する音声区間の間の無音区間の時間長もパラメータに加え、連続する 3 つの音声区間の係り受け関係を推定する手法により[16]、学習外の読み上げ音声に対し、約 80% の精度が得られている。

プロソディから文意の構造を推定する場合に、実時間理解に有利な構造があるのではないか、という考え方から、先行する音声区間の情報のみから、後続の音声区間の情報を使うことなく、後続の音声区間に係り受け関係を推定する可能性を検討し、後続の情報を使った場合と殆ど遜色の無い率 (約 76%) での推定に成功している。後続区間との間の長さや、基本周波数の差の代わりに、先行区間の最後の拍の長さと基本周波数の勾配を用いた。その場合に比較すると、正解率はわずかに 4% の低下にとどまっている。

(3) 語彙あるいはアクセント節の分節位置予告機能

アクセントの役割は、主にセグメンテーション機能である。自然な言葉と同じ調子で発声された無意味な言葉を聞かせて、何處で切れると判断されるかという認知実験を行った。

3 + 6 モーラ, 4 + 5 モーラ, 5 + 4 モーラ, 6 + 3 モーラなどからなる有意味音声と同じ調子で、無意味な言葉「なななななななな」を発音して、何處で切れるのが自然かという実験を行った。結果は 90% 以上の確率で、アクセント節の切れ目と一致する位置で切れると判断された。音声の音韻の情報が無くても、アクセントフレーズの切れ目が解かるのである[17]。

一方、アクセント型の推定を同時にを行うと、推定成績は 60% 台と意外に低い。アクセント情報がセグメンテーションとアクセント型推定と別処理となっている可能性が高い (9 月の日本心理学会で畠野から報告予定)。このことは、アクセント節の開始部分の情報 (β の値など) からアクセント節の尾部の位置を予告している可能性を示しているものと思われる。この点については今後畠野らと検討を進める計画である。

(4) 後続音韻の予告機能

調音結合などにもその機能が存在するのではないかと予

測される。実際、単独に発声した音節を録音しつなぎ合わせた合成音声は極めて聞きづらい。おそらくは次の音の予測が困難であるとか、一つの纏まった単語パターンとしての把握が困難である、あるいはその双方、というようなことと思われる。

6. 3 感情・個人性

対面対話では、単に言葉という形でのコミュニケーションだけではなく、感情や、誰が話しているかといった個人性などの情報も極めて重要な役割を果たしている。これらの情報がセットになって、豊かなコミュニケーションが成り立っている。

個人性（身体的・生理的特徴や言語的習慣）や心理的特長（感情などの現れ方など[18]）などはプロソディやスペクトル特性などに現れる。

（1）感情

感情の表れは、生命維持などのレベルから文化習慣のレベルまで極めて多様である。それらの感情が生理的・身体的フィルタを通って発声器官などの身体動作の影響の下に音声に反映される。その主な影響はプロソディに現れる[18]。生理的緊張などを考慮すると声帯波形にも大きな影響が予測され、その結果スペクトル形状などにも反映されると考えられるが、その関係は未だ必ずしも明らかではない。

障害者や高齢者が自身の体調や心理状態を的確に表現することが困難な場合があるが、プロソディなどを通してその情報を把握することが期待される。

（2）個人性

相手が誰で、どのような状況でいるかも、対話時の相手からの情報を理解判断する上で重要な働きをしているものと考えられる。

6. 4 心内辞書へのアクセス

これまで提案されている多くのモデルは、音韻認識を得て心的辞書にアクセスするものが多い[19]。しかし実時間性や言語障害の事例、幼児の言語発達の様子などを考えると、単語やアクセントフレーズでセグメントされたスペクトル特系列パターンのようなものから直接アクセスしていくのではないかと思われる。

6. 5 伝達内容情報とコンテンツ構造

（1）非一覧性とHPビルダ

自治体など公共の組織のHPはアクセシブルである事が求められる。しかし、残念ながらHPを作製する大部分の人はアクセシブルなHPの備えるべき条件を把握していない[20]。実際アクセシビリティ Webに関する規定の各項目を遵守していても、われわれの調査によれば必ずしもユーザビリティは良くはなっていないのが現状である[21]。

したがって、音声化を想定したアクセシブルなHPを作製支援するHPビルダの実現が期待される。そこでは揮発性のメディアである音声の性質や、マジックナンバ7と呼ばれる認知的特性を反映するものであることが望まれる。

（2）仮名漢字変換

日本語は漢字の導入により音韻の種類が少なくなり、漢語の利用もあいまって同音異義語が多くなっている。

このため、視覚障害者にとっては、仮名漢字変換する場合に、適切な漢字を選択することに大きな課題を抱えている。現状では、「詳細読み」という手法により、同一漢字を用いる単語をヒントに漢字変換する方法が主流となっているが、先天性の視覚障害者はもともと漢字を見たことがないため、同一かどうかの判断が難しいことがある。この対策として、読みと意味から選択する方法が提案されている[22]。

視覚の利用が可能な先天性の聴覚障害者も実はその漢字の読み方の音声を聞いた事がないため、妥当な読み方が分からず、仮名入力が出来ず、仮名漢字変換が困難な例も存在する。

6. 6 対話の構造

対話においては、最低限話題の展開が円滑に行われることが重要である。

（1）二人対話

二人対話では、基本的には相手の発話を呼応するので、話題の展開には大きな問題は生じないだろう。その上で、実時間での話者交替や情報伝達が可能であることが理想的である。

（2）複数対話（会議）

複数対話では、参加者に情報の授受の速度に差がある場合、遅い者が話題の展開に取り残されることが最大の問題である。コミュニケーション障害者がこれに相当する。話題の展開に追従できるように、対話の流れを制御する工夫が必要になる。発話権の円滑な制御の保障も必要である。

われわれは、外出が難しく、触覚情報でのコミュニケーションを余儀なくされている視聴覚重複障害者（盲ろう者）が実時間合意形成に参加可能なインターネット利用会議システムを試作している[23]。

7. 障害者用メディアの研究から音声研究へのヒントの例

7. 1 音声と手話、指点字の比較研究から

音声と同じく揮発性である手話や指点字も、実時間での対話が可能である事実から推論すると、これら手話や指点字にも音声のプロソディに相当する情報が存在するものと

予測される。これらの比較検討から、逆に音声に関する視点や理解が進む側面が存在する。

(1) 手話のプロソディ

日本手話は音声と同様に対話型自然言語であり、実時間でのコミュニケーションが可能な性質を備えている。手話にも音声のプロソディに相当する情報が存在すると考えられる。手話では、基本周波数は存在しないが、顔の表情[24]や身体の動きなどの非手指動作（NMS）と、時間構造、動作の大きさの時間パターンなどがその情報を担っているとかんがえられる。これらを手話のプロソディと呼ぶことに対する[25]。日本語対応手話²も日本語音声と類似した時間構造のプロソディが存在する。

手話や指点字でも、音声と同様に[26]、プロソディ機能に対応すると思われる情報を外すと、文の理解は約80%から50%程度に低下することが観測される。

聞き易い音声合成の実現に的確なプロソディの付与が重要なように、CGなどによる手話動画合成には、手話のプロソディの的確な付与が欠かせない。

手話動作から文を構成する手話単語間の係り受けの構造を推定することも試みられている[8]。

手話においても、予測機能が存在している事を示唆する結果が報告されている。一定の手話文章の中に挟まれた手話単語が何であるかを、その手話単語の始まりからどこまで見たら解かるかを調べた。手話動画を途中で打ち切ったものを作成し、どのような単語かを予測させた（ゲート法）[27]。打ち切り時点を順次遅くし、どこで正解となるかを調べたのである。先天の聾者は殆ど単語の始まった時点（単語全体の約15%）で正解を得ているが、手話通訳者は約26%必要であり、手話の初心者は約60%（手指の形が完成した時点）を見ないと正解できなかった。

この結果は、手話を母語としているものは、単語が始まつた時点でどのような単語が示されるかが予測できることを示している。

(2) 指点字のプロソディ

指点字は、盲ろう者のためのコミュニケーション手段の一つである。点字の出来る視覚障害者から盲ろう者になった方が主に使用している[25]。障害者の6本の指を点字タ

² 日本の手話：日本における手話には、日本手話と日本語対応手話、中間型手話などがあると言われている。日本手話は主に先天性の聴覚障害者が用いている手話で、日本語とは異なる自然言語であるが、文法などは未解明の点が多い。日本語対応手話は主に日本語を母語としてきた人が病気などで聴覚障害となり、主に日本語の順序で手話単語を並べて表現される。中間型手話はそれらの混合的なもので、手話の学習の進み具合や、相手により表現の混合の程度が変わる。

イプライタに見たて、その指に点字を打って情報を伝えるもので、実時間性に優れた方法である。熟達した盲ろう者は1分間に300から350字を読み取ることが出来ると言われている。点字のコード系を利用しているおり、点字は記号化され6点の組み合わせのため、熟練度を無視すれば相対的に学習が容易な方法である。

実時間性に優れていることから、「指点字」にもプロソディがあるという仮説を立て、指点字の分析を行った。文節の間の時間も、文の構造が理解しやすいように、関係の深い文節間ほど短く、関係の遠い文節間ほどゆっくり打たれている[28]。

指点字では、その言語系は音声と基本的に同じであるから、音声に近い時間構造でパターン化されるのではないかと予測される。実際指点字のベテランの人のパターンを見てみると、アクセントフレーズ中ではほぼ200ms程度のリズムで、アクセントフレーズ間では400ms程度以上である。

(3) 手話と音声の身体動作の比較

音声発声時にもうなずきなど様々な身体動作を伴う。手話の身体動作との比較により、その共通性と相違点を観察することから、対話の構造をより詳しく調べることが期待される[29]。

7. 2 失語症の研究から

失語症は様々なタイプがあり、その解析からヒトの音声処理のメカニズムを知るヒントが多数存在する。

手話者の失語症にも音声の場合と同じ現象が観測されている。

7. 3 手話認識技術から

手話は両手の動作や表情、身体動作などが平行して表される言語であり、その機械認識には並行認識技術が必要である。この技術は未開発であるが、実現されれば音声のスペクトルとプロソディの同時処理や、マルチモーダルな情報認識など、幅広く応用されよう。

8. おわりに

代替手段を使わざるを得ない障害者や加齢に伴う機能低下などのある高齢者にとってコミュニケーションのための心的負担は高く、ユーザビリティへの配慮は非常に重要である。

ヒトにとってのコミュニケーションのもともとプリミティブな形態は対話形式であるが、揮発性という心的負担が大きいという欠点を持っている。

しかし自然な対話音声はそれを補う伝達支援情報を含んでおり、その有効活用がユーザビリティを改善する上で極めて有効と考えられる。伝達支援情報には、伝えたい情報

を実時間で理解可能にする情報（伝達したい情報の構造情報や、その構造を予告する情報）や、相手の状態（個人性や心理状態）を伝える情報が主にプロソディに含まれている。

この情報は、障害者が用いる手話や指点字などの揮発性メディアにも含まれているが、それらの性質は未解明の部分が大きい。

音声の研究者は、その経験を活かし、この性質を解明し、活用することなどを通して、情報福祉の世界に貢献することが期待されている。

参考文献

- [1] 厚生労働省:「国際生活機能分類—国際障害分類改訂版一」(日本語版);
<http://mhlw.go.jp/stf/seisaku/seisaku-000010000000000000.html> (2003)
- [2] 平成13年身体障害者実態調査, 厚生労働省 (2006.9)
- [3] 伊福部達, “福祉工学の挑戦—身体機能を支援する科学とビジネス”, 中公新書1776, 中央公論新社 (2004)
- [4] A. Iida, et. al., “Communication aid for non-vocal people using corpus-based concatenative speech synthesis,” Proc. Eurospeech 2001, pp. 2401–2409 (2001)
- [5] 松政宏典, 他, “情報家電操作における脳性麻痺構音障害者の音声認識評価”, 電子情報通信学会福祉情報工学研究会資料, WIT2007-7, pp. 33–38 (2007.5)
- [6] 中園薰, “手話動画像通信に関する研究”, 千葉大学学位論文 (2006.3)
- [7] 寺内美奈, 他, “遅延手話対話における話者交替時の信号表出に関する解析的検討”, 電子情報通信学会福祉情報工学研究会資料, WIT-2007-11 (2007.5)
- [8] 平山望武, 他, “日本手話における時間構造の分析”, ヒューマンインタフェース学会論文誌, Vol. 3, No. 3, pp. 9–14 (2001.8)
- [9] 芳賀繁, “メンタルワーカロードの理論と測定”, 日本出版サービス (2001.7)
- [10] 藤原敦史, 他, “表および数式の音声化の検討”, ヒューマン・インターフェース・シンポジウム論文集, pp. 643–648 (1997.10)
- [11] 森大毅, “連続音声認識の手法を応用した走査型文字入力方式”, 日本音響学会春季講演論文集1-Q-31 (2007.3)
- [12] 市川薰, 手嶋教之, “福祉と情報技術”, オーム社 (2006.9)
- [13] 大須賀智子, 他, “音声対話での話者交替／継続の予測における韻律情報の有効性”, 人工知能学会論文誌, Vol. 21, No. 1, pp. 1–8 (2006.1)
- [14] 木村太郎, 他, “遺伝的アルゴリズムによるF0モデルパラメータ推定法と話者交替分析への適用”, 電子情報通信学会音声研究会資料, SP2006-82, pp. 37–42 (2006.12)
- [15] 小松昭男, 他, “韻律情報を利用した構文推定およびワードスポットによる会話音声理解方式”, 電子情報通信学会論文誌, Vol. J71-D, No. 7, pp. 1218–1228 (1988)
- [16] Ohsga, et. al., “Estimating Syntactic Structure from Prosody in Japanese Speech,” IEICE Trans. Inf. & Syst., Vol. E86-D, No. 3, pp. 558–564 (2003.3)
- [17] Hatano, T., “Prosody Based Speech Segmentation,” In Proc. The 5th International Conference of the Cognitive Science (ICCS2006), pp. 103–104 (2006)
- [18] 市川薰, 他, “合成音声の自然性に関する実験的考察”, 日本音響学会秋季講演論文集, pp. 95–96 (1967)
- [19] 紺野加奈江, “失語症言語治療の基礎”, 診断と治療社 (2001.9)
- [20] 「公共分野におけるアクセシビリティの確保に関する研究会」報告書の公表,
http://www.soumu.go.jp/s-news/2005/051215_1.html
- [21] 飯塚潤一, 他, “視覚障害者のウェブサイトの検索効率と心的負担に関する考察”電子情報通信学会福祉情報工学研究会資料, pp. 55–60, (2007.3)
- [22] 西田昌史, 他, “意味情報を利用した視覚障害者が連想しやすい仮名漢字変換手法”, 日本音響学会春季講演論文集, 2-8-8, pp. 343–344, (2007.3)
- [23] 宮城愛美, 他, “発言権を考慮した指點字と文字による会議システムの構築”, 電子情報通信学会論文誌D, Vol. J90-D, No. 3, pp. 732–741 (2007.3)
- [24] 市川薰, 他, “手話における“顔”のはたらき”, 情報処理学会コンピュータビジョンとイメージメディア研究会資料, 2005-CVIM-148, pp. 67–72 (2005.3)
- [25] 市川薰, “人と人をつなぐ声・手話・指點字”, 岩波書店 (2001.10)
- [26] 北原義典, 他, “音声言語受容における韻律効果の検討”, 電子情報通信学会創立70周年記念総合大会1339, (1987)
- [27] 市川優子, 他, “認知科学的手法による手話読み取り特性的検討”, 日本国際学会第22回大会予稿集, 5. 3. pp. 71–74 (1996.6)
- [28] 宮城愛美, 他, “指點字のプロソディの分析”, ヒューマン・インターフェース学会論文誌, Vol. 1, No. 3, pp. 35–40 (1999.8)
- [29] 堀内靖雄, 他, “日本手話の後続うなづきの機能に関する検討”, 電子情報通信学会福祉情報工学研究会資料, WIT-2007-12 (2007.5)