

論文 / 著書情報
Article / Book Information

論題(和文)	音声情報処理とヒューマンインタフェース
Title	
著者(和文)	古井貞熙
Author	SADAOKI FURUI
出典(和文)	日本音響学会誌50巻1号, , , pp. 53-55
Journal/Book name	, , , pp. 53-55
発行日 / Issue date	1994, 1

小特集—学会誌 50 巻を迎えて—

音声情報処理とヒューマンインタフェース*

古井 貞 熙 (NTT ヒューマンインタフェース研究所古井特別研究室)**

1. はじめに

音声情報処理, すなわち, 音声認識, 合成, 符号化などの研究には長い歴史があるが, この 25 年間に大きな進歩を遂げ, これまでの多くの技術的積み上げが, 今まさに花を開こうとしている。

音声の符号化に関しては, 最近日本において世界に先駆けて, 自動車及び携帯電話用のいわゆるハーフレートの符号化方式の標準化が行われ, わずか 4 kbps 程度の情報量で, 高い品質の音声伝送できるようになっている。これに要する計算量は 18 MOPS 程度と, 25 年前には想像もできなかった莫大な量であるが, 今やこれが 2~3 個の汎用 DSP, あるいは最新のワークステーション上のソフトウェアで実時間で実行できるようになっている。符号化技術そのものと共に, ハードウェア技術の進歩のおかげである。

音声の合成に関しても, 種々の合成規則の進歩と共に, 記憶容量や処理速度の大幅な向上をフルに利用した方法が, 最近活発に研究されている。例えば, 調音結合の効果による音素や音節のスペクトルの変動をクラスタ化して, 多数のパターンを蓄えておき, それらを接続して音声を生成することにより, 高い品質の合成音声を得られるようになっている。

大語彙の連続音声を不特定多数の人が発声した音声を認識するためには, 極めて複雑で大量な処理が必要である。しかし, この分野においても, 統計的モデルを中心とする大きな技術的進歩と, コンピュータの高速化により, ワークステーションによる実験システムの実現が可能になってきた。

2. 音声情報処理の階層構造

音声の認識, 符号化, 合成の処理の過程と相互関連を, 情報量に関連付けて図-1 に示す。この内, 符号化について実現されているのは, 波形符号化と分析合成までで, 認識符号化から上はまだ実現されていない。認識については, 対話音声の認識の研究が活発になるに従って, 意味概念を積極的に取り扱う音声理解の研究が盛んに行われるようになってきた。合成においても, 人間とコンピュータとの対話システムを実現するためには, システムがユーザに伝えるべき情報を, 対話の状況に応じて適切な言葉に変換する必要があるので, 概念からの音声合成の研究が始められている。

この階層構造は, 音声に含まれる言語的情報のみに注目したものであるが, 音声には, 個人性情報, 情緒性(感情)情報なども含まれており, 我々の生活の中で重要な役割を果たしている。個人性情報を用いた話者認識の技術も, 最近大きく進歩している。音声に含まれる話し手の感情を自動的に読み取ったり, 感情を付与した音声を合成したりする技術も, 将来必要となるであろう。

3. 音声情報処理とマルチメディア

これからはマルチメディアの時代である, と言われるが, 音声の重要性は今後も変わらないであろう。音声のみによる情報伝達には限界があるが, 映像や図表の組み合わせによって, 情報伝達や対話の効率を大きく上げることができ, そこにマルチメディアの効果がある。人間とコンピュータとの対話システムの基本的構成として, コンピュータへの情報入力には, 音声認識, 手書き文字認識, タッチパネル, 更に補助的手段としてマウスやキーボードが考えられる。一方コンピュータからの情報の出力には, 図表, 文字, 映像, 合成音声な

* Speech information processing and human interface.

** Sadaoki Furui (Furui Research Laboratory, NTT Human Interface Laboratories, Musashino, 180)

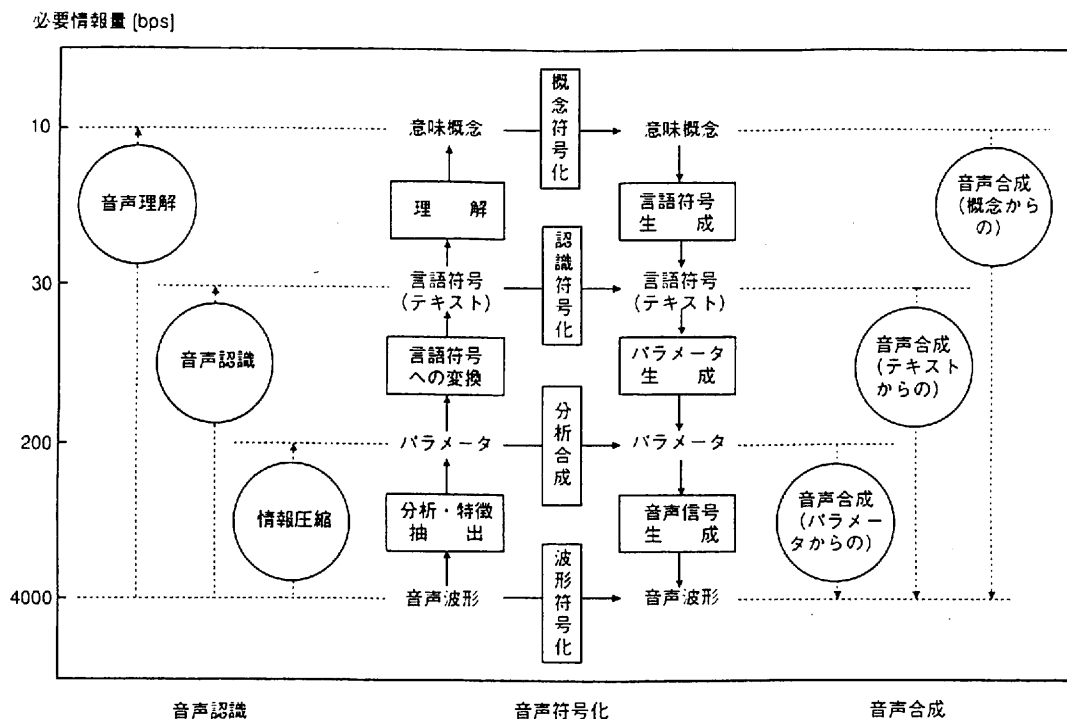


図-1 音声情報処理の階層構造

どが考えられる。最も基本的で、人間にとって自然かつ便利な手段は、音声認識による入力と、文字及び図表による出力の組み合わせであろう。大局を表現することが得意な図形、画像などによる表示をまず行って、細かいところは言語情報などで説明するのが理想的な姿であろう。

そのようなシステムの用途としては、データベース(情報)検索、種々の案内、予約、注文、電子秘書、ロボット、障害者の補助などがあり、将来的には自動翻訳電話も考えられる。それらを実現するための究極の情報入力手段は、限りなく人間に近い能力を有し、なおかつ幾つかの点で人間の能力を越えることができる音声認識理解システムであろう。人間の能力を越える方法としては、より早い、より正確、より賢い、より安い、より気楽に使える、など種々の側面が考えられる。この実現のための研究課題の詳細は省略するが、一つ重要な課題を上げるとすれば、自己学習及び適応化の機能であろう。すなわち、タスク(話題)や話者が変わって、構文、語彙、話し方、言葉の意味、スペクトル特徴などが変化したとき、あるいは認識誤りを生じたときなどに、自動的に新しい知識を獲得し、変化に適応して、正しい認識ができるようになる機能の付与が重要な課題であ

る。

4. 人にやさしいインタフェース

人間は必ず何等かの形でミスをするものである。書かれた文章をアナウンサーが読み上げる場合ですら、読み間違いということがある。ましてや原稿なしで話しているときには、書き言葉の文法に照らして正しく話していることは少ない。このような誤りを伴う発声をしたときに、それを受け入れてくれるような、インタフェースが必要である。こう書くと、対話音声の認識はかなり難しいのではないかと考えられるが、2歳や3歳の子供でもやっていることであるから、発想を転換して、書き言葉を対象とした従来の文法規則などから離れた新しい規則を見出せば、かえって易くなる可能性もある。対話は二者で内容を確認しながら進めていくものなので、インタフェースがうまくできていれば、一つ一つの文は比較的単純になるとも考えられる。

更に、システムを用いている途中で、どうしたらよいか分からなくなったような時に、親切にかつ適切に助けてくれるインタフェースが必要である。学習機能を持っていて、個々人の知識や思考過程を吸収したり、予測したりしてくれる、個人

に適応的なインタフェースも必要である。これにより、人間の知的活動を支援するようなインタフェースが可能となると期待される。

5. む す び

これまで、音声情報処理の研究の進展に関して、本学会の果たしてきた役割は極めて大きい。今後も研究成果の発表、討論、意見交換、切磋琢磨の

場として、本学会が重要な役割を果たしていくことを期待する。また、今後いっそう、日本と外国との国際交流や国際貢献が期待されるようになると予想されるので、日本からの情報の発信源としての、国際会議、ワークショップ、英文論文誌の活発化などにも、本学会が積極的に取り組んでいく必要がある。