解　説

音声情報処理技術の動向†

中　田　和　男††

1. 音声情報処理とは

音声波形のあらわす情報（広義の音声情報）は，単にその言語的内容に関するものだけではなく，話し手が男か女か，また誰であるか，どんな気持で話しているか（心理的な側面と生理的な側面の両方を含む），といったことから，さらに話し手がどの地方の出身者か，どのような背景のもとに育った人か，といったことまでも含んでいる．

これらの多様な情報の中で，通常，言語的内容だけが話し手によって意図的に音声波形に担わされた情報である．

他の情報は，音声波形がある特定の話し手によって発生されるものであることによって，必然的にではあるが附随的に，音声波形が背負っている情報である．

これらの情報の中で，現在ある程度まで科学的に取り扱うことのできるのは，言語的内容に関するもの（狭義の音声情報）と話し手の同定に関するもの（話者情報）だけである．したがって，本解説で扱う音声情報の範囲も上記の2つが主である．

音声情報処理における情報処理の内容は，図-1 に示すように，音声情報の符号化である．音声が人間と人間の間の情報伝達の手段として使われるとき，その符号化は能率的な符号化 (low bit rate coding) もしくは雑音や妨害に強い符号化 (robust coding) となる．音声を人間と機械の間の情報授受の手段として使おうとするとき，その符号化は，音声の自動認識による文字化と，文字表記から音声波形への音声合成による変換となる．

この音声情報処理を達成する技術の基礎は，図-2 に示すように，音声分析技術であり，波形符号化技術，

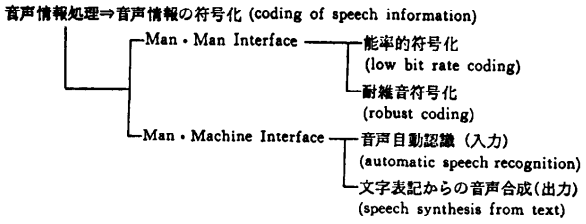

図-1　音声情報処理の主要内容

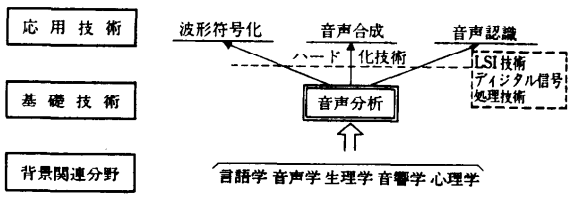

図-2　音声情報処理技術の説明

音声合成技術，音声認識技術はその応用技術である．

音声とは言語の情報を音響的な物理現象として実現（表出）するときの符号の体系であり，その符号化プロセスが人間の発話運動であり，復号化プロセスが人間の音声知覚，認知過程である．したがって音声情報処理の研究は，必然的に言語学，音声学，生理学，音響学，心理学といった幅広い学問的背景を必要とし，さらに装置のハード化に当たっては，ディジタル信号処理技術，LSI 技術の知識を必要とする．本質的に学際的な分野である．

2. 音声情報処理の必要性

音声情報は本来人間と人間の間の情報伝達の手段として，人類進化の歴史とともに発達したものである．

電気通信技術の進歩により，音声通信は電話を中心とした形で現代社会に組み込まれている．通信技術の発展につれてディジタル化がすすみ，音声，画像，データといった情報の内容に立ち到ることなく，すべて同一形式の信号として取り扱おうとする統合ディジタル通信網の計画がすすめられている．

† Technical Trends of Speech Information Processing by Kazuo NAKATA (Department of Applied Physics, Faculty of Engineering, Tokyo University of Agriculture and Technology).
†† 東京農工大学工学部応用物理学科

音声波形をそのまま単純にディジタル化すると，サンプリング定理と量子化雑音による信号対雑音比の観点から，64Kビット/秒を必要とする．これは9.6Kビット/秒を標準とするデータの伝送速度からみて数倍の情報量であり，簡単に同一信号として統合することはできない．また，経済的な通信のためにもこの情報量を（本来の情報を失うことなく）低減することが必要である．

また計算機を主体とする情報処理の発展につれて，人間と機械のインタフェースの重要性がクローズアップされてきた．それにともなって本来人間にとってもっとも自然で自由な「音声」を機械とのインタフェースとしても使えるようにしたいという要求が高まってきた．

前者の要求にこたえるのが，波形符号化技術による音声の情報圧縮であり，後者の要求にこたえるのが，音声認識，音声合成技術による音声の自動入出力である．高度な波形符号化技術は，音声の分析・合成技術へと必然的に発展し，さらに，音声認識→文字への符号化→音声合成という形態も想定され，これらの技術は互いに深く関連し合って現在の音声情報処理技術を形作っている．

3. 音声情報処理への関心

この数年，音声情報処理とくにその実用的な応用としての音声合成と音声認識に人々の関心が高まっている．その背景には，線形予測理論にもとづく音声分析技術の飛躍的な進歩と，その基礎のうえに立つ新しい音声合成，音声認識技術の展開があり，またそれらの理論的な成果を具体化するためのハードウェア技術とくにディジタル信号処理とLSI技術の進歩があった．その引き金となったのがテキサスインスツルメンツ社による音声合成LSIの開発と，その商品化としてのSpeak & Spellの発売であった．

商用MOS・LSI技術と，新しい音声合成技術であるLPC (Linear Predictive Coding：線形予測符号化）方式との結び付きによる商品化の成功という図式の延長上にVLSI技術と新しい音声認識技術との結び付きによる音声認識装置の製品化と実用化という期待が見通され，人々の関心を引き付けている．

情報処理システムにおけるマン・マシン・インタフェースとしての音声利用には，つねに潜在的ニーズが存在しており，技術的，コスト的に可能だということになればニーズはどっと顕在化するわけで，それだけ人々の関心も高まっている．

4. 音声情報処理の現状

それでは音声情報処理あるいは処理技術の現状はどの位まで進んでいるといえるか，その詳細については本解説につづく各論で述べられるところだから，ここではごく概観的に把握するに止めておく．

1）波形符号化技術　　図-3に示すように，32Kビット/秒のレベルではADPCM (Adaptive Differential PCM：適応差分PCM）方式が国際標準として固まりつつある．

しかし，電話音声より高品質の通信用（ステレオ音楽伝送などをも含めて）としてはAPC-AB (Adaptive Predictive Coding-Adaptive Bit Allocation：適応的ビット配分適応予測符号化)方式が有力である．16Kビット/秒レベルではAPC-ABとADM (Adaptive Delta Modulation) 方式が共存しており，前者は高品質用，後者は専用通信用と分野を分かち，ADMおよびその改良方式は装置が簡単なことを利点としている．

9.6Kビット/秒レベルとなると，分析・合成方式とならざるをえず，LPC（実用的にはPARCOR：Partial auto-Correlation（偏相関）方式，またはLSP：Line Spectrum Pair（線スペクトル対）方式），ベクトル量子化方式などが主となり，2.4Kビット/秒以下ではLSPとなる．

2）音声合成　　実用化されているものはほとんどすべて分析・合成方式であるが，最近いわゆる純粋合成（または規則合成）方式といって，文字表記だけからある程度自然に聞こえる音声を作り出す研究も進んでいる．米国では一部製品化されているが，日本ではまだ製品化されたものはない．これは技術レベルのちがいというより，音質に対する日本人の要求のきびしさに原因があるものと思われる．

3）音声認識　　特定話者，限定語彙，離散発生の

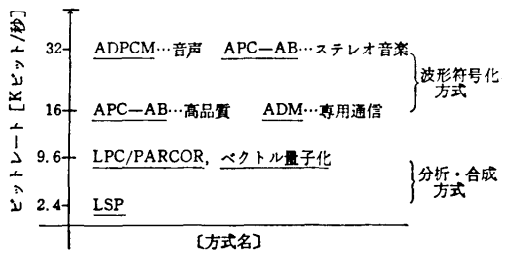

図-3　情報圧縮方式の位置づけ

単語認識レベルのものは内外で市販されており，実用化の成功例も数多く報告されている．特定話者，限定語彙であっても，それを登録しかえることによって装置の適応性を高めており，限られた使い方には十分役立つものとなっている．

最近の動向として，不特定話者の，とくに電話を介した，音声の認識が銀行の情報サービスシステムなどを中心に実用化されつつある．これは従来押しボタンダイヤルで行われていた情報の入力を直接音声によって行おうとするものであるが，現在のところ語彙は0から9までの数字とハイ，イイエなどごく少数の制御語に限られている．

5. 情報処理システムと音声入出力

音声入力の利点として，とくにキーボード入力との比較で挙げられるのは，次のような点である．

i）人間にとって自然であり，入力操作に訓練を要しない．

ii）手と目が入力動作から解放されて自由になり，物品の取り扱い，測定作業などを入力動作と並行して行える．

iii）ワイヤレスマイクの使用によって，ある範囲で動きまわりながら入力ができる．

iv）入力速度が速い．

v）入力がオンライン（情報発生地点）でただちに行われ，認識結果の確認（音声応答や視覚表示による）が容易であり，総合して入力情報の信頼性と能率が高くなる．

また音声出力の利点として，とくに視覚表示との比較で挙げられるのは，次の諸点である．

i）作業をしながら聞くことができる．

ii）必要な時に注意を喚起し，人間への割り込み機能が高い．

iii）内容の理解が容易であり，すぐに対応することができる．

iv）ワイヤレス・レシーバによって動きながらでも情報を受けることができる．

v）動作の指示，入力の確認などに適している．

以下これらの利点をいくつかの代表的な応用例について具体的に説明する．

5.1 生産ラインとデータの音声入出力

現在の音声認識装置実用化例の多くが生産ラインにおけるデータの音声入力用である．とくに部品や製品の検査工程におけるデータの音声入力である．

手と目がデータの入力作業から解放されているから，ベルトコンベヤ上を流れてくる部品を手にとって検査し，必要な測定を行いながら，その結果，たとえば部品番号とその合否や測定値そのものを音声でデータ入力することができる．その結果によって部品は仕分けされ，合格品はコンテナに格納される．米国 General Electric 社の家電品の部品検査の例では，キーパンチによるデータ入力の場合にくらべて，部品検査のスループットが約 30% 向上したと報告されている．

また必要な情報が，その発生地点で，オンライン的にただちに入力されることから，次のような点も大きな利点として指摘されている．

i）合格部品がコンテナに一パイになったところで指示を与えることにより，部品リストがただちに印刷されて出力され，それを切り取って伝票としてコンテナに貼れば，そのまま次の工程に進めることができる．

ii）部品の不良率が即時に算定され，それがあるレベルを切ると，その情報はただちに製造ラインにフィードバックされ，製造条件の改善が行われる．

また自動車の生産における最終検査などでは，検査する人がある範囲で動きまわりながら検査することが必要であり，ここでもワイヤレスマイクと FM 中継器（ベルトに装着する）による音声入力の利点が発揮される．

音声出力は，音声認識による音声入力の結果の確認に使われるだけでなく，複雑な配線や加工を必要とする生産ラインで，作業者に適確な作業指示をつぎつぎと与えるのにも使用され，効果があがっている．

これらの実例を通じて，将来の生産工場では，音声入出力は生産工程の一要素としてその中に一体として組みこまれ，生産性向上のために必要不可欠のものとなるだろうと予言されている．

5.2 OA システムと音声入出力

日本における OA システム実用化のポイントの1つが日本語情報の入出力，ことにその簡易な入力手段の開発にあることは周知の通りである．ここでもまた音声入力実用化への期待は大きい．

しかし，この場合の音声認識技術は，現行音声認識技術とは全く異なったアプローチを必要とする．

すなわち，現行の音声認識技術は単語を単位とした有限語彙の音声認識技術であるが，日本語情報入力のために必要な音声認識技術は，任意の日本語音声を認識できる無限語彙音声認識か，一歩ゆずっても，10^3

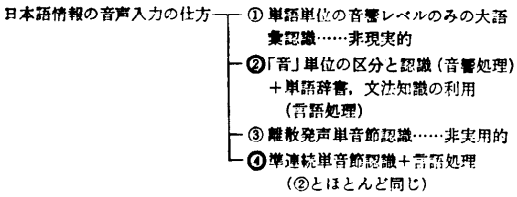

図-4　日本語情報の音声入力の仕方

語以上の大語彙音声認識技術である．それには図-4に示すような可能性があるが，音響的な分析結果を，単語を単位として，そのまま標準パタンとして使用する方式では，標準パタンメモリの量からいっても，またその各単語標準パタンと比較照合を行う処理量，処理回数からいっても，10^3 語以上をあつかうことは現実的ではない．

そこで単音節（カナ文字1字に対応する発音）のような基本的な「音（オン）」の単位を考え，その系列として単語や句，文を認識するというアプローチをとらざるをえない．しかし音（オン）のレベルの認識だけで十分な結果がえられるとは考えられないから，単語辞書や文法の利用が必要となる．音（オン）の単位のとり方によって，SPLIT（擬音素パタン）法，音素（VおよびC）単位，単音節（Vまたは CV）単位，VCV, CVC 単位（Cは子音を，Vは母音をあらわす）などの変形がありうる．

一方，より原始的に，ちょうどカナタイプを打つように，単語や文を音節単位に区切って離散発声し，これを一音一音認識していくという考え方もありうる．

しかし，この直接的な離散発声単音節認識の形では日本語情報の音声入力の実用化は難しい．その理由として，a）単語や文を離散単音節として発声することは人間にとって大へん不自然であり，自然で自由だという音声入力の利点がほとんど失われてしまう．b）入力速度が速いという利点も大幅に失われる．c）技術的にも単音節認識だからやさしいとはいえない．などの点をあげることができる．

単音節認識の形で実用化するためには，話し手の意識のうえでは離散発声であるが，物理的にはほとんど連続発声に近く，ただ一音一音がユックリとまたハッキリと発音されているといった準連続的な発声を許すレベルにまで技術を高めなければならない．

日本語の音声入力の場合，特定の人，あるいは特定の業務に限定すれば，使われる単語や熟語の数が 10^3 をこえることはあまりないと考えられる．ただその 10^3 語の内容は人が変わり，業務が変われば変わると考えなければならない．それを汎用にカバーしようとすると $3\sim5\times10^3$ 語が必要ということになる．

たとえば，単語辞書は文字や発音記号などによって表記されており，語彙の変更に応じて（人間がその単語をいちいち発音してみなくても）容易に内容を追加，変更できるようにしておく．そして新しい発音や未知の話者が加わったとき，音響レベルの音（オン）の標準パタンのみを現実の音声波形（特定話者の実際の発音）から切り出して作れるようにしておくことが望まれる．

日本語情報の音声入力は音声認識にとって大へん魅力的なマーケットだけれども，現状ではいま一つ技術的にギャップがあるといわざるをえない．このギャップをどう埋めていくか，音声認識技術の今後の大きな課題であり，成功すれば成果の大きな研究課題の一つである．

5.3　ゲームと音声入出力

マイコン応用の今後の大きなマーケットにゲームがある．この中には，いわゆる娯楽のゲームもあるが訓練や教育用といったものも含めて考える．音（オト）の出力はゲームにすでに様々な形でとり入れられ，ゲームを一層エキサイテングなものにするのに役立っている．

しかし音声出力はごく一部いわゆる単語，計算の練習や，将棋，囲碁の訓練用のようなものを除いてはまだあまり使われていない．

いわゆるゲームでは，ゲームセンタのような環境条件からいって音声入力はなじまないという意見もある．

Speak & Spell の例からみてもわかるように，訓練や教育用の機器には音声入出力がもっと使われてよいはずで，それを拒んでいるものはコストではないかと思われる．

5.4　情報システムの Security と音声

データバンクへの情報の蓄積と，システムのネットワーク化による広域利用とが発達普及してくると，情報システムには2つの意味で security（安全の確保）が問題になってくる．1つは，そのシステム自身の安全を守るために，その所在場所へ出入りする人の身元を厳重にチェックし管理しなければならない．もう1つは，その情報を利用できる人（データにアクセスして情報をとり出せる人）を制限し，データをとられている人の個人の秘密（プライバシ）を守らなければならない．その基本は個人の同定技術である．

通常は，各個人に秘密の暗証コードが与えられており（ID 番号），それを端末から入力したり，その番号

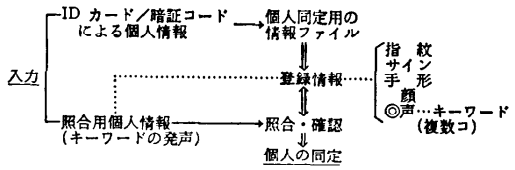

図-5　声による話者照合システムの原理

を記録した磁気カードをセンサに入力することによって個人を同定している．しかしこのような手段はいずれも盗用が可能であり，安全確実な手段とはいえない．そこで指紋，サイン(筆蹟)，手の形，顔(画像)，声など人間個体と分離することができず，真似することもむずかしい情報によって，ID 番号（カード）によって同定された個人と端末ユーザを照合し，確認をとるという方法が望ましい．この中で，入力センサが簡単で処理時間が短く，しかも個人照合率の高いものとして声によるものが注目されている．これが，音声による話者照合 (speaker verification) 技術である．

この場合，**図-5** に原理を示すように，必ず別の手段によって個人を同定し，その人と同一人であるか否かを音声によって照合するという形にシステムを作ることが必要である．技術的には，照合に当たって，予め登録されているもの（キーワード）と言語的に同じ内容の音声を利用できる場合と，そうでない場合とがあり，当然後者の方が難しい．また照合率を高めるために，複数コの単語をキーワードとして使用できるようにすることが実用上大切である．

安全がよく守られ，プライバシに関して比較的関心の低い日本では，まだこの技術の重要性が十分認識されていないように見受けられるが，米国では軍事や企業の情報の安全と機密保持のためにシステマティックに研究開発が進められており，音声の利用が一番実用性が高いと評価されている．技術的な問題点の一つとして，声の個人的特徴の経時的変化による照合率の劣化が指摘されているが，その実態については電電公社武蔵野電気通信研究所による精力的な研究結果があり，音源逆フィルタや複数キーワードの使用によって克服することができる．また毎日使う場合には参照音声パタンを適時更新するという対策も考えられており，十分実用にたえうる．音声による話者照合の技術は，今後音声認識におとらない重要さをもってくるものと考えられる．

6. 音声情報処理技術の今後の展開

以上，技術的な内容そのものについては各論の解説にゆずって，ユーザの立場から使い方を主に音声情報処理技術の内容と現状について概観してきた．つぎに今後の展開についてトピックス的な内容を追って説明する．

6.1 VLSI 技術と音声認識

音声認識装置実用化の大きなポイントが，その機能をおとすことなく，そのハードの小形・低電力・低コスト化にあることはいうまでもない．それを実現する技術的な手段として LSI，ことに VLSI 技術との結合がクローズアップされている．

現在，すでに**図-6**に示すような機能ブロック別 LSI 化（音声認識装置の LSI シリーズ化）と，それによる小形化が進められており，近い将来 1 ボード化まで進むものと思われる．しかし機能ブロック別の LSI 化では，ブロック間のインタフェースにかなりの IC 回路を必要とし，思ったほどの小形化低コスト化にはなりにくいという状況もある．そこで VLSI 技術によって，**図-7** に示すように，標準パタンメモリは別として，1 チップ化を実現し，完全に 1 ボード化することが望まれる．チップコストとマーケットサイズは鶏と卵の関係にあり，商品としてなにに量産化の突破口を見出すかが非常に重要なポイントになる．

米国の VHSIC (Very High-Speed Integrated Circuits) 開発計画でも音響信号，音声情報処理を目的とした高速信号処理プロセッサの試作は主要な開発項目の 1 つに挙げられている．1980 年代前半が開発競争の時代，1980 年後半が本格的 VLSI 音声認識チ

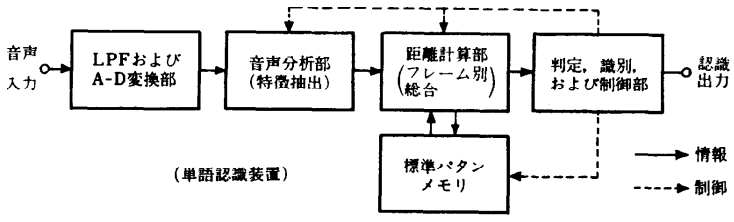

図-6　音声認識装置の機能ブロック別構成

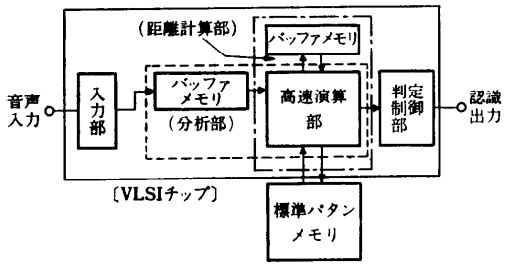

図-7 VLSI 音声認識装置のブロック図

ップの実用化時代とみることができよう.

6.2 音声情報のメール化

近い将来，音声認識に劣らぬマーケットが期待されるが音声情報のメール化サービスである.

音声情報のメール化とは，音声情報を符号化して一旦記憶し，必要に応じて分配・再生することを意味し，voice storage and forward service と呼ばれている．電話（電気通信）が音響信号による音声伝達の距離の限界を取り払った進歩とすれば，メール化は音声通信の実時間性（一過性）の制約を取り除いた機能拡張ということができる.

もちろん，従来の録音技術によって原理的には音声信号の一過性の制限は取り除かれているが，いわゆる通信網と結合して，それと一体となって記憶機能を発揮するには，磁気録音ではコスト的にもハード的にも不可能に近かった．しかし最近の音声情報圧縮技術，とくに分析・合成技術の進歩と半導体メモリの大容量化，低コスト化によって，メモリとして半導体メモリを使うことがコスト的に可能な範囲に入り，高速のランダムアクセスメモリで音声情報を記憶することが現実的となり，本格的なメール化が可能になった.

メール化によって次のような新通話サービスが可能になる.

ⅰ）話中通話を一時記憶し，通話終了時にそれを呼び出し再生する.

ⅱ）指定相手に指定時刻に記憶メッセージを送出する（会議案内や召集メッセージ通話の自動発呼など).

ⅲ）国際通話の時差の解消（一方的なメッセージ伝達のみに限る).

ⅳ）通話者間の了解のもとで通話の内容を記録する.

ⅴ）留守中や夜間のメッセージを記録し，受信者の要求により適時再生する.

音声情報のメール化は，単に新しい通話サービス機能をもたらすのみでなく，OA システムの中に音声情報をとりこむことをも可能にするものである．音声情報は，最終的には，音声から文字へ，文字から音声へ，という形で OA システムの中にとりこまれるべきものと思う．しかしそれが技術的にそう簡単にできるとは思われない以上，情報圧縮符号化の形でのメール化が当面の姿であろう.

音声のメール化と基本的には全く同じ分析・合成技術による低ビットレート伝送方式（9.6Kビット/秒以下，できれば 2.4Kビット/秒程度までの低ビット化）は，ディジタル電話という形でも実用化されつつある．ここでは 9.6Kビット/秒のデータ回線で音声を，できれば複数チャネル伝送し，あるいはファックスやデータ情報と音声情報を同時に伝送し，データ回線利用の効率をあげ，等価的な低コスト化を狙っている.

また音声のパケット伝送も試みられている.

これらに共通の技術は，低ビットレートで品質劣化のできるだけ少ない音声の分析・合成方式の開発とそのハード化である．合成側については各種方式について LSI 化によるハード化はすでに製品化されており，問題は分析側の実時間処理とその LSI 化である．高速信号処理プロセッサを２コ使った構成の PARCOR 方式分析部のハード化例が報告されている．今後こうした製品（voice digitizer と総称されている）のマーケットの拡大が期待されている.

6.3 Telephone transaction

従来，押しボタン電話機の押しボタンを入力手段とし，音声応答による音声を出力手段とする国鉄の電話座席予約システムや銀行の情報案内（残高照介や振込案内）システムが一部実用化されてきた．しかし押しボタン電話機の普及が十分でなく，押しボタン入力にかえて音声認識による音声入力を実現したいという要求が強まり，銀行などでは一部実用化されている．技術的にみたとき，このシステム実用化のポイントは不特定話者の電話音声認識の実現である.

この点，技術的に十分に解明されて対策が立てられているのではなく，むしろニーズに迫まられて便宜的な解決が計られているところが多い．したがって認識語彙が極端に少なく（10 数字と数語の制御語の計 16 語程度），離散発声に限られている.

もし，不特定話者，電話音声で連続発声数字の認識ができるということになれば，押しボタン入力よりは便利となるから，こうした電話機を入出力端末とした情報サービス，通信販売，口座間の金額のやりとりな

どがかなり自由に行われるようになる．こうしたサービスを一括して telephone transaction（電話取引き）と呼んでいる．電話網という既存のネットワークと電話機という簡便な入出力端末を利用するものであり，広い用途が期待される．

問題は不特定話者の連続発声電話数字音声をどの位の認識率で認識できるかである．n 桁の連続数字音声がすべて正しく認識され，音声入力の目的が一度で達成される確率Pは，一桁の数字が誤認識される確率を ε としたとき，$P=(1-\varepsilon)^n \approx 1-n\cdot\varepsilon$，$\varepsilon \ll 1$ である．したがって相当な高認識率でなければならない．$n=4$，$\varepsilon=0.03$（1桁数字音認識率 97%）として，$P \approx 1-4\times 0.03=.88$ すなわち4桁連続発声では9割を割る認識率になってしまう．

現在のところ，特定話者の連続発声数字者，不特定話者の離散発声1桁数字音の認識は一応技術的に可能である．その最高認識率をそれぞれ 97% とみて，その組み合せの認識率が最悪単純に積であらわされるとすれば，4桁数字では，$.97\times(.97)^4 \approx 0.86$ となる．

もちろんシステム的に，誤認識やリジェクトのときの再入力の方法について色々人間工学的な配慮がなされる．

それにしても現在の技術の延長，組み合せで，実用的な telephone transaction システムを作ることができるかどうか今後の大きな研究課題である．

オフィスなどの専用業務における telephone transaction システムでは，音声認識装置を特定話者用として LSI で小形化して端末に組みこんでしまうという方向も将来方向としては一考に値するのではないかと思われる．

6.4 INS と音声情報圧縮伝送

電話通信網（アナログ信号伝送），データ通信網（ディジタル信号伝送），画像通信をディジタル信号形式で統合し，統合ディジタル通信網として画像，音声，データを一体とした高度の情報サービスを提供する将来計画が INS (Information Network System) 計画として電電公社によって進められている．ここでは音声情報もディジタル化されて伝送される．その際，INS への移行に当たって問題となるのは既存電話網との互換性 (compatibility) である．互換性には2つの意味がある．1つは信号形式としての（アナログ信号とディジタル信号の）共存性互換性であり，もう1つは品質の等レベル維持である．

前者のためには，波形符号化，分析・合成符号化を含めて，高品質低ビットレートの音声情報伝送技術の確立が必要であり，APC-AB 方式や PARCOR, LSP 方式の開発はそのための技術開発である．一方後者のためには復号化された音声波形信号の品質の客観的評価量として，品質の主観評価とよい相関を示す物理的尺度の確立が必要である．ADPCM, APC-AB などは波形符号化方式の系統に属する符号化方式であり，従来の品質評価のわく組みの中で処理することができ，物理尺度としても SNR（信号対雑音比）を拡張した segmental SNR（フレーム別 SNR の平均値）で比較することができる．分析・合成方式となると波形歪としての SNR ではうまく評価できないが，フレーム別パワスペクトルの差の平均で定義されるスペクトル歪と MOS 値（平均オピニョン値：主観的品質評価の客観的測定値）はよい対応を示す．しかし規則合成音声の品質となると，品質の内容自体が人間の肉声による音声の場合と相当に異なったものとなるため，従来の品質評価のわく組みであつかうのは難しい．新しい品質概念の確立がまず必要である．

電話通信のわく外にある音声通信（移動無線，船舶通信，専用通信など）は既存通信との互換性の制約がないから，低ビットレートの新伝送方式を容易に実用化しうる分野と考えられる．その中間に自動車電話がある．

6.5 音声情報処理と人工知能

音声が言語を物理的に表現するための符号体系と考えられる以上，音声情報処理が自然言語処理と深い関係にあることはいうまでもない．音声認識や文字からの音声合成（規則合成）が，単語辞書や文法規則の形で自然言語の情報を活用しなければ十分なものとならないのがそのよい実例である．

さらに進んで，限定されたタスクであっても，音声理解 (speech understanding) という形で音声認識を考えるとき，対象についての知識の表現やその検索，それにもとづく推論による不確さの推定，認識結果の予盾の検出，あるいは最終的な目的（たとえば航空機の座席予約）を達するための人間・機械間の Q・A システムにおける質問や確認文章の生成など，現在の人工知能研究の中心課題である知識の獲得，表現，利用の具体例となっている．いいかえれば，音声が言語と不可分のものである以上，音声情報処理を音響的な信号処理のレベルから少しでも発展させようとすれば，それは必然的に言語処理，知識の表現，利用といった人工知能の中心課題と関連する．いやむしろ人工知能

の問題の具体的展開そのものとなる．最新の人工知能とくに自然言語処理の研究成果を背景に，今後音声情報処理の中でとくに発展の期待される分野である．

以上最近の音声情報処理技術について，研究者の立場よりは一般ユーザの立場から，技術の内容，現状，今後の発展方向，可能性などについて説明した．研究者の立場からの技術上の問題点についての筆者の見解は別に発表の機会をえた*．参考としてあわせて読んで頂ければ幸いである．

結論として，音声情報処理技術全体として，ある限定された条件の下では十分実用にたえる機能を，しかるべきコストで提供しうるまでには発達している．しかし人間と同じようなフレキシブルな機能を達成するには，まだまだ本質的な研究開発が必要であるといわなければならない．

* 中田：音声認識の基礎，東北大学応用情報学研究センタ10周年記念シンポジウム予稿，pp. 1-8 (2月，1983).

参考文献

個々の技術内容については，各論およびその参考文献に挙げられているので，ここではまとまった参考書，主として著書をあげておく．

1. 外国著書

1) Fant, G.: Acoustic theory of speech production, 's-Gravenhage: Mouton & Co., 2nd ed. (1970).
2) Flanagan, J.L.: Speech analysis, synthesis and perception, Springer-Verlag (1965).
3) Markel, J.D. and Gray Jr., A.H.: Linear prediction of speech, Springer-Verlag (1976).
4) 論文特集号
   a) Digital Signal Processing, Pro. IEEE, 63-5 (Apr. 1975).
   b) Man-Machine Communication by Voice, Pro. IEEE, 64-4 (Apr. 1976).

2. 国内著書

1) 大泉，藤村監修：音声科学，東大出版会 (1972).
2) 三浦種敏他：聴覚と音声（新版）(1980).
3) 中田和男：音声，コロナ社 (1977).
4) 新美康永：音声認識，共立出版 (1979).
5) 斉藤，中田：音声情報処理の基礎，オーム社 (1981).
6) 電子通信学会：LSI 応用，コロナ社 (1982).

(昭和58年3月17日受付)