

## 1.7 音声端末の概要

最近では音声認識システムが一般のパソコンに組み込まれる例が見られるようになった。しかし、認識精度のレベルは別にしても問題は要求される動作環境である。次節で取り上げた2つのソフトウェアはいずれも、必要メモリ 32M バイト以上、空きハードディスク容量 60M バイト以上、インテル社製 Pentium プロセッサのクロック周波数 100 ~ 133MHz を推奨している。しかし、デモンストレーションでは 200 MHz のパソコンを使用していたが、それでも変換はスムーズではなかった。

Windows95 華やかりし2年前にパソコンを購入したユーザの内、必要メモリ 32M バイトをクリアする機種を保有するユーザはかなり少ないであろうし、クロック周波数でもひっかかるであろう。

音声認識システムを必要とするのは、初心者ユーザ、障害者、或いは手術中の医師や物流分野等ハンドフリーが要求される分野である。その人達に音声認識が定着するのは、最低でも標準実装メモリが 64M バイト、プロセッサ・クロック 300MHz が標準装備となる頃のことではないかと思われる。

いずれにしても、音声認識が益々インターフェイス手段として重要度を増して来るのは確かであろう。MIT メディアラボのニコラス・ネグロポンテ教授が GII (Global Information Infrastructure) ジュニアサミット'95 での招待講演でも、コンピュータの小型化に伴ってグラフィカルなインターフェイスではなく、言語で操作できるようになるだろうと予想している。言語以外では小さな機械とのインターアクションが出来ないとの見解である。ネグロポンテ教授は最終的にはコンピュータは靴のかかたとにセットされると予想している。<sup>14)</sup>

一方、パソコン上でテキストを声で読み上げる音声合成機能を持ったソフトも数多く登場している。音声合成も自然性を追求していくとまだまだ問題は多々あるが、商品化に際しては一応機能的に問題が少ないので、音声認識よりも普及している。

### 1.7.1 音声認識・合成ソフトウェアの実例

#### 1) 日本 IBM VoiceType Dictation 3.0

VoiceType Dictation 3.0 は話者も選ばず、学習機能により使うほど認識率が向上するというふれ込みである。声によりプログラムを操作したり、日本語テキストの入力を行なう。例えば「プログラムリスト」と発声すればリストが現れ、「プログラム名+実行」と発声すれば目的のプログラムが起動する。終了するときには「閉じる」と発声する。「印刷」「ファイル」「編集」等のメニューも発声によって操作可能である。又、「私」「は」「駅」「に」「行く」の様に言葉と言葉の間に間をとる離散的な読み上げ文章を、前後の文脈を類推しながら漢字仮名混じり文に変換する。

論文の下書き等に使用しているユーザの感想では、認識性能は思ったよりも良い印象を受けている。但し多少の練習が必要であり、ある程度大きい声でマイクに向かって発声する必要があり、静かな環境では周囲に迷惑をかけることになる。

IBM は、更に離散的発声ではなく、自然な話し方の様な連続発声による入力を可能にし

た音声認識ソフト ViaVoiceGold を発表しているが、ハード環境はプロセッサ・クロック 200MHz 以上、ハードディスク容量 150M バイト以上と一段と厳しくなっている。<sup>15)</sup>

## 2) NEC “しゃべっていいメール”

“しゃべっていいメール”も話者を選ばずに、音声で文章を入力すると自動的に漢字混じりの文章に変換してくれたり、届いたメールを読み上げてくれる。又、音声でのパソコン操作も可能である。

しかし、実際に試してみるとデモストレータが気の毒なくらい変換待ち時間が長く、誤変換が多い。言葉を切って話す等入力にコツがあるようで、慣れればもう少し良くなるのであろう。

文章の読み上げの方は文意も良く判かり問題は少ない。

## 3) 富士通 “おしゃべりメイト”

“おしゃべりメイト”は文脈から構文を理解し、男性の声もしくは女性の声で自然なイントネーションとアクセントで読んでくれる。

### 1.7.2 音声認識の応用例

音声認識は、パソコンだけでなく色々なマルチメディアの入力手段として活用されつつある。ここでは特徴のある例を幾つか上げる。

#### 1) 超小型 PHS 端末機

図 1 - 14 に NTT が開発した超小型 PHS 端末機を示す。音声認識によりダイヤルする機能によって操作ボタンを減らして小型化を実現している。登録機能により「天気予報」と発声すれば“177”、「会社」と発声すれば勤務先にダイヤルすることが出来る。図 1 - 14 のものはアンテナ、バッテリー、音声認識機能を親機に別に用意していた。これを基に平成 9 年 12 月には一体化した世界最小・最軽量の図 1 - 15 の様な実用機が完成している。内蔵マイクとスピーカでの拡声通話及びイヤホンマイクを使つての通話が可能である。イヤホンマイクは、空気中を伝わる音と発声者の頭の骨を伝わる骨導音との混合比を制御して工事現場の騒音下での通話を可能にしている。腕時計電話が現実のものになりつつある。

<sup>16)</sup>

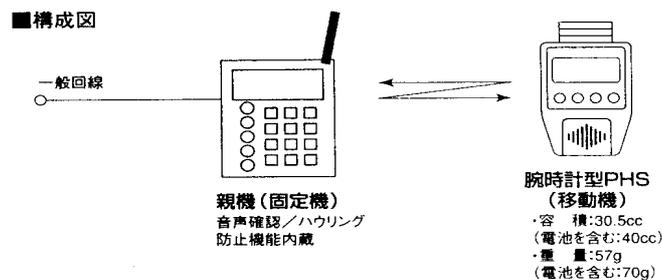


図 1 - 14 超小型 PHS 端末試作機

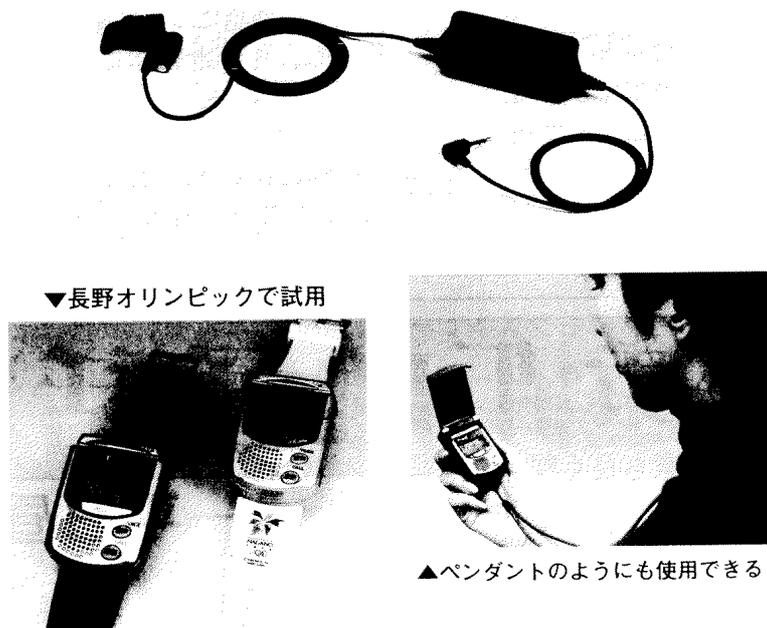
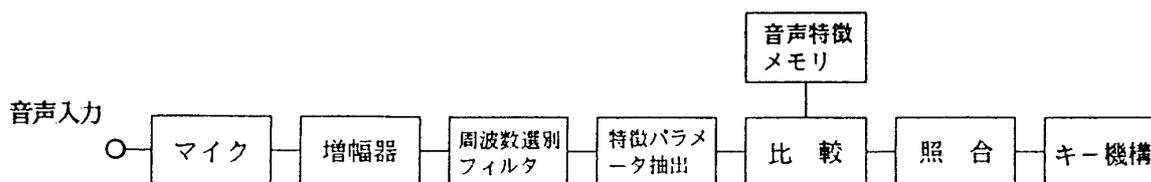


図 1 - 15 その実用機とイヤホンマイク

### 2) 音声キーシステム

(株)ユーシンと千葉工業大学の城戸教授他は、音声キーシステムを開発している。図 1 - 16 の様に、特徴パラメータであるタイムスペクトルパターンの抽出を個人的特徴が大きい周波数帯域のみに限定し、処理量を大幅に減少している。発声時間・判定時間ともに 1 秒で認識率は 93%、自動車の盗難防止や出入管理等に向け実用化を進めるといふ。音声キーはネットワーク上での個人認証への応用等、将来的に大きな市場になる可能性がある。



信号処理ブロック図

図 1 - 16 音声キーシステム

### 3) ゲーム機

家庭用ゲーム機にもコントローラの代わりに音声認識を使うものが現れた。

富士通の“南の島のフィンフィン”では、パソコン上で動物を飼う際の動物へのメッセージ入力手段として使われている。又、任天堂は NINTENDO 64 対応のもので、ヘッドホンタイプの小型マイクに向かって「走れ」「止れ」としゃべって画面のキャラクタを動かしたり、「かわいいね」「おはよう」等の認識も出来るゲームを 98 年秋に発売予定であるという。<sup>17)</sup>

#### 4) 番組のリクエスト・検索

NHK 放送技術研究所では、音声認識を用いて番組のリクエストや検索を行なう図 1 - 17 の様なシステムを開発している。音声から母音を利用してキーワードのみを検出するシステムであり、自由な言回りで話し掛けることが出来る。<sup>18)</sup>

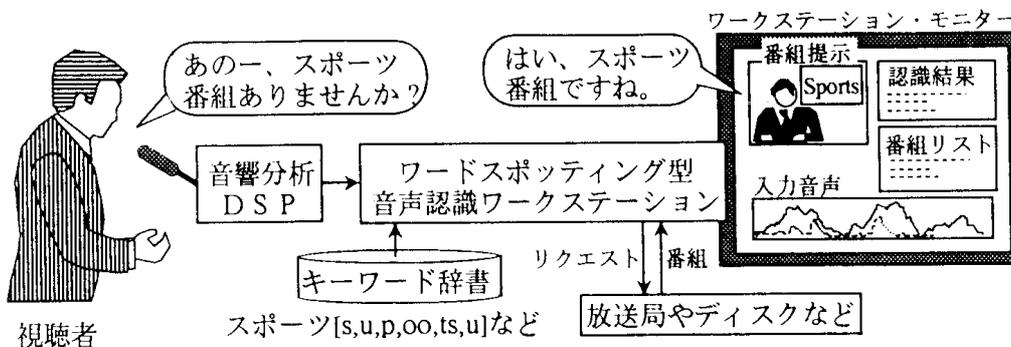


図 1 - 17 番組のリクエスト・検索

その他、(株)アニモの“ボイスクリニック”の様に音声によりストレス度や性格を診断するものも現われている。又、音声対話によるデータベース検索と応答等も研究されている。<sup>19)</sup>

ただ、単純な音声認識は別として一般の人が日常的に使用する場合には、学会レベルでは音声認識はまだ十分に実用化に達していないという見解が強いことも確かである。

情報処理学会音声言語処理研究会での参加者(80名強)を対象にした「あなたは、なぜ、これまで音声認識技術が期待するほど使われてこなかったと思いますか? 選択肢の中に当てはまると思うものがあれば、複数回答でお答え下さい」という質問に対する集計結果を表 1 - 1 に示す。<sup>20)</sup>

その結果から「音声認識に関してはまだ基本性能が十分でなく、諸性能がまだ不足」との見解が示されている。100%の認識率が期待出来ない音声認識をいかに使いこなすか、その利用法の研究が少ないことを指摘し、又、現段階では音声認識の応用分野は他に入手手段がない場合、身体障害の場合、作業中に処理する場合、或いはカーナビゲーション等の車載システムであることも指摘している。

- (a) 認識率が低い。音声認識は技術が未熟である。まだ使いものにならない。……………36 票
- (b) 自由発話音声認識ができるまでは本格的に使えない。いまの連続音声認識は使いものにならない。……23 票
- (c) 話者の間で認識性能の差が大き過ぎる。認識率が低い話者がいるのでは使えない。……………13 票
- (d) 扱える語彙サイズ(語数)がまだ小さ過ぎる。……11 票
- (e) 語彙制約があるのでは使えない。任意の音声・文字変換できなければ不自由。……………4 票
- (f) ワードスポッティング技術が肝要。……………18 票
- (g) リジェクト力が弱いことが最大の問題。……………40 票
- (h) 雑音や発声変形や回線変動などに対するロバストネスが不足。……………46 票
- (i) ヒューマンインタフェースが未熟だから使いにくい。……………41 票
- (j) うまい対話制御が重要。これがうまくいっていないから使われない。……………22 票
- (k) 人は機械に向かって話すのには抵抗がある。音声認識技術は本質的に嫌われる。……………7 票
- (l) 音声認識誤りがどのように起こるのか、どう発声すれば避けられるのか分からない。この不透明感がユーザにとって最も辛い。……………41 票
- (m) キーボードなどの他手段に比べて入力効率が決して良くない。……………17 票
- (n) 音声認識機能とアプリケーションのインタフェースが確立していない。……………24 票
- (o) コストの問題。現状では、音声認識技術は高価過ぎる。……………11 票
- (p) まだ速度が十分でない。……………6 票
- (q) マイクフォンが口の近くにしなければ動作しないのでは応用が限られる。……………10 票
- (r) 音声認識の応用について知恵が足りない。現在の技術レベルでも工夫すれば使えるはず。……………37 票
- (s) 言語処理が問題。音響処理はかなりのレベルに達しているが言語モデルあるいは言語処理(構文解析、意味理解、状況、知識、社会常識)が遅れている。……………18 票
- (t) 音声認識への要求条件が厳し過ぎる。もっと育てるつもりで使えるところから使うべきだ。……………0 票
- (u) 音声認識を使う人間を訓練する必要がある。キーボードだって一日で使えるようにはならない。……………6 票
- (v) 音声認識の使い方の社会的コンセンサスがまだない。……………11 票
- (A) 音声研究者が実用化など考えていないからだ。この技術は行ける、という確信がなければ実用化はできない。……………22 票
- (B) 提供されるサービスが十分に知的(intelligent)である必要がある。「対話」が知的であるかではなく「サービス」が知的であることが肝要。……………21 票
- (C) 次の3つの条件がまだうまく噛み合っていない。
  - I. 音声認識理解の性能と機能がある程度のレベル(応用による)に達していること
  - II. システムに組み込む人と言う意味でのユーザが容易に組み込める為のサポートツールが用意されていること
  - III. 音声認識理解を利用することを前提とした応用技術が開発されていること(例えば、音声認識を前提としたCAI技術など)……………34 票
- (D) 長年の研究成果があまり開かれたものになっていない。市場を立ち上げるためには、基礎技術以外にも製品やサービスとして完成させるアイデアや技術や根性や資本が不可欠。……………7 票
- (E) 音声認識というものは既に人間が非常な夢を描き、理想が先行してしまった技術であるために、なかなか便利であると認めて貰えない。発想の転換が必要。……4 票
- (F) メンタルモデル(システムがユーザにどのように見えているか)とシステムイメージのギャップを埋める手段が確立していない。……………36 票

表 1 - 1 音声認識技術が期待するほど使われてこなかった原因 アンケート調査

一方、最近では音声に重点をおいたマルチモダールインターフェイスが各社で研究され、人間とコンピュータとの自由な対話形態の実現を目指している。

例えば図 1 - 18 に示すのは日立中央研究所で開発されたマルチモダールウィンドウシステム(MMoWinS)である。音声とペンによるポインティングによってパソコンのウィンドウ操作を可能としている。情報統合部で音声とペンの入力情報を統合し、ユーザの操作を解釈してコマンドに変換し、ウィンドウマネージャに実行させる様なシステムである。

21)

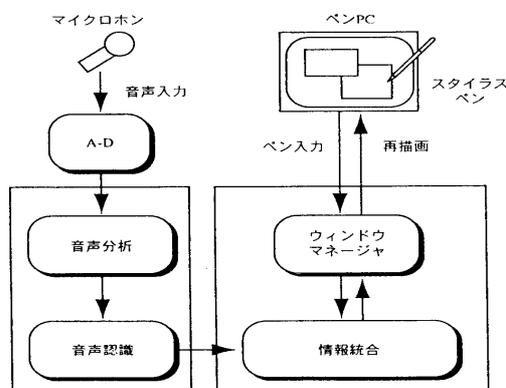


図 1 - 18 マルチモダールウィンドウシステム

将来的には音声認識 + ジェスチャ、更には表情、視線等を用いたマンマシンインターフェイスが盛んになっていくであろう。