

音声入出力技術の利用動向

Utilization Trends of Voice Technology

栗原利彦* Toshihiko Kurihara
花木将守* Shôji Hanaki

最近の音声合成及び認識技術は、ここ数年来飛躍的に伸び、既に通信、コンピュータの端末から家庭電気品、教育機器、ゲーム、玩具などへの応用が拡大してきている。

音声は人間にとって最も自然かつ容易な情報入出力手段であり、この意味から幅広いニーズがもともと存在していたと言えるが、ここにきて急激に応用機器が拡大した理由は、現代のエレクトロニクス技術にマッチした技術体系をもっている点にあると考えられる。

本稿では、最近の音声入出力技術の特徴と応用例として音声応答システムや音声認識装置及び話者認識装置を挙げそれぞれの紹介と実用上の問題点に触れ、今後の応用分野と課題について述べる。

1 緒言

近時、新しいマンマシンインタフェース用として、音声の入出力技術の応用が盛んに行なわれるようになってきた。

このマンマシンインタフェースに音声出力装置を用いたセンタ形装置として、電話予約システム、勝馬投票券電話投票システム、銀行の振込連絡照会システム、食品製造業を中心としたオーダエントリシステムなどの音声応答システムが各方面で数多く利用されている。

また、最近のLSI化音声合成チップの出現により、センタ周辺機器中心の応用からしだいに端末機器、家庭電気品などへと利用範囲が広がってきており、日立読上算練習器などがその例で、身近かで数多く使われ始めてきている。

一方、音声入力装置には特定話者を対象とした単語認識装置が荷物の行先別仕分けや検査データなど、手作業を伴うデータ入力分野に適しており、日立製作所では全社的規模で生産技術関連部門に導入した多くの実績をもち、その効果を発揮している。また、不特定話者を対象とした単語認識装置は電話回線を通して不特定多数の話者が利用するようなシス

テムに適しており、銀行での電話サービスなどに音声応答システムと組み合わせて利用されている。これら応用分野のいっそうの拡大と発展のため、連続発声認識、小形・低価格化、多語彙の認識などの技術課題について研究開発を進めている。

図1に音声入出力技術の利用動向を示す。

2 音声入出力技術への期待

日常生活での人間相互間の通信の大部分が音声によって行なわれており、人間にとって最も自然かつ容易な情報入出力手段であるこの音声を、マンマシンインタフェースとして用いることにより、次に述べるような多くの利点が考えられる。

- (1) 音声は人間にとって最も自然な情報手段である。
- (2) 情報伝達速度が速く、理解が容易である。
- (3) 情報伝達時(音声による入出力時)手と目が自由である。
- (4) 人間の五感の中でも特に音声は割込順位が高いため、注意を喚起できる。
- (5) ワイヤレスマイクの利用により、広い範囲を動いて情報

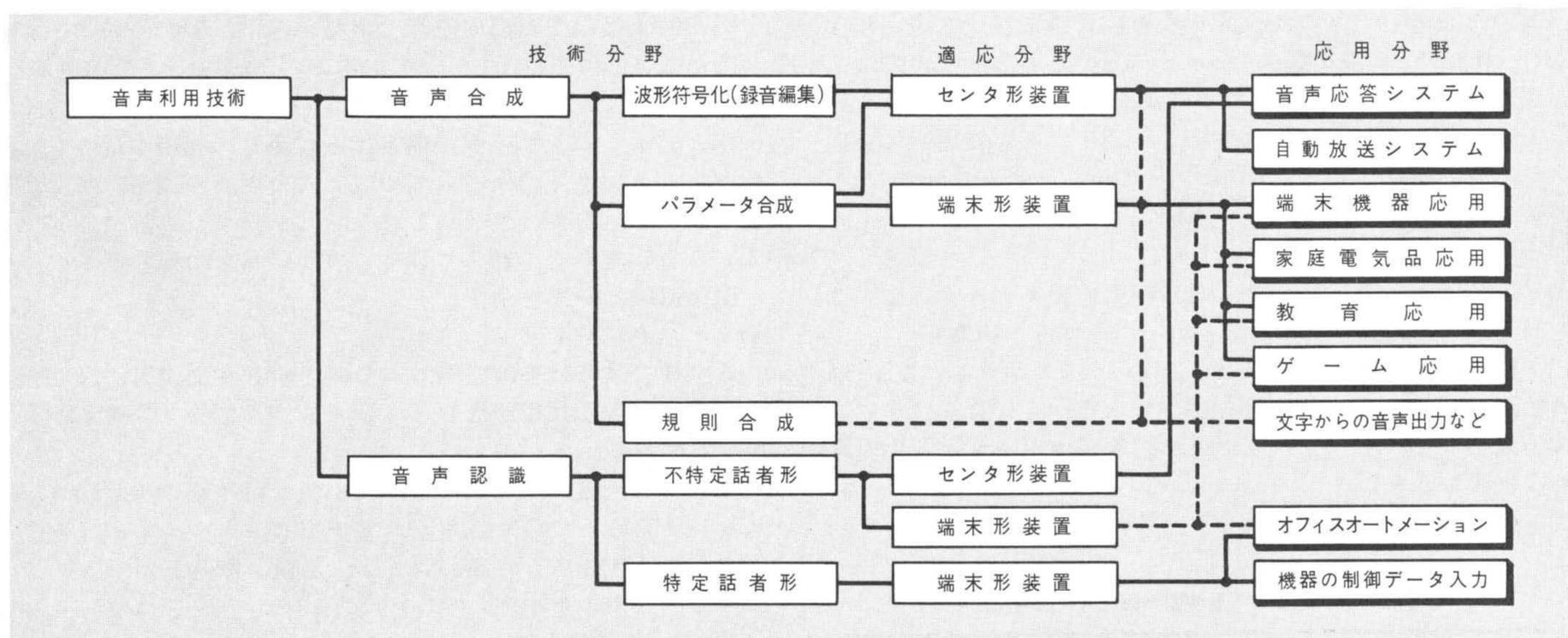


図1 音声入出力技術の利用動向 音声入出力技術の応用分野を図にまとめた。点線は現在研究・開発中のものを示す。

* 日立製作所通信機事業部

の入力ができる。

- (6) 自由な姿勢がとれる。
- (7) 特別の訓練を必要としない。
- (8) 電話の利用により全国オンラインが可能である。

このような多くの利点をもつ音声によるマンマシンインタフェース技術は、現時点では人間のもつ機能に比較し、まだ十分とは言えないのが現状である。したがって、実用化されている音声入出力装置は、技術的に実現された機能に合わせた用途を選択し利用され始めている状況である。

3 音声出力技術

3.1 音声合成

音声合成の種類は次の3方式に分類できる。

(1) 波形符号化方式(録音編集)

音声波形をアナログ又はデジタルで記録しておき、必要に応じて読み出しつなぎ合わせて音声波形を作り出す方式で、最も実績のある方式であり多重応答能力が高く、回線当たりの経済性を図ることができる。

音声信号の情報圧縮度が小さいため、記憶媒体は大形となるが、処理が簡単で音声品質が優れており、PCM(Pulse Code Modulation:パルス符号変調)などの変調方式が使用されている。

(2) パラメータ合成方式(分析合成)

音声を分析して、情報圧縮されたパラメータの形でデジタル記録しておき、再生時に合成器を利用して音声波形を再合成し作り出す方式で、PARCOR(Partial Autocorrelation:偏自己相関係数)方式で代表される音声合成用LSIが開発され各分野に導入されている。

この方式は音声信号の情報圧縮度が大きく小形化でき、高い音声品質が得られる特長がある。

(3) 規則合成方式

任意の単語や文章を合成し、音声出力しようとするもので、無限の語彙を発生することが可能であるが、現状では合成音声の品質が低い欠点がある。

表1に音声合成方式の比較を示す。

3.2 センタ形装置への応用

音声出力の応用例の代表的なものに音声応答システムHIVORS(Hitachi Voice Response System)がある。この音声応答システムは、コンピュータの処理結果を人間の音声に変換して出力するもので、端末にはプッシュホンを使用して完備された電話網により「いつでも」「どこからでも」「だれでも」がコンピュータと対話できるシステムを実現している。

この音声応答システムの特長は、

- (1) 電話機からの押しボタン信号がそのままデータとしてコンピュータに入力できること(センタ側の省力効果)。
- (2) 電話機をそのまま端末として用いることができること(経済性)。
- (3) 電話網を利用しているため、広範囲のオンライン化が容易に実現できること(システムの利用拡大)。
- (4) 押しボタン信号に対する入力確認を音声で行なうので、

表1 音声合成方式の比較 各種音声合成方式の機能比較を示す。

方式名	波形符号化(録音編集)	パラメータ合成	規則合成
波形又はパラメータメモリ	大	中	小
処理	簡単	やや複雑	複雑
語彙数	中	大	無限
多重度	大	小	小
音声品質	優	優	良

表2 銀行システムの入出力手順 振込連絡・残高照会システムのうち、振込連絡の一例を示す。

プッシュホン入力		順序	出力音声
入力項目	入力内容		
		(呼出し)	こちらは〇〇銀行テレホンサービスセンタです。お振込みの連絡をしますので、よろしければ暗証番号をどうぞ。
暗証番号	××××#	↙	〇件のお振込みがございます。 〇番〇〇様から(小切手で)〇〇〇円 以上〇件、合計は〇〇〇円です。 確認コードをどうぞ。
確認コード	0#	↘	ありがとうございました。

正確な入力を得られること(正確さ)。

- (5) 完全自動化により、24時間サービスが可能であること(サービス向上)。
などがある。

音声応答システムは、各方面で既に多数稼働している。その代表例を次に述べる。

(1) 予約業務¹⁾

日本国有鉄道では、「みどりの窓口」で活躍している座席予約システムと音声応答システムを接続して、昭和50年3月から「電話予約システム」としてプッシュホンから新幹線の座席予約サービスを関東一円で実施している。このほか、予約業務として航空券やホテル、病院などの予約、また企業内で会議室や応接室の予約など、広範囲な利用が検討されている。

(2) 勝馬投票券電話投票システム²⁾

日本中央競馬会では、競馬場や場外発売所の混雑緩和などを図るために、プッシュホンから勝馬投票券を購入する「電話投票システム」を、昭和51年10月から東京、大阪、名古屋、神戸など、全国各地に導入している。電話投票の加入者はあらかじめ中央競馬会と利用契約を結んだ後、自宅のプッシュホンから音声応答システムと会話しながら勝馬の投票を行なう。レース終了後、購入金、払戻し金などは指定銀行の普通預金口座にオフラインで記帳される。

(3) 銀行業務

昭和54年後半から都市銀行で事務合理化、顧客へのサービス向上などを目的として、音声応答システムによる「振込連絡・残高照会システム」が導入され、その他の都市銀行から地方銀行へと広まってきている。顧客はあらかじめ利用申込みをしておくことにより、振込入金や取立入金などの連絡、また預金残高や振込内容の照会などをコンピュータが音声で回答するシステムである。表2に銀行でのサービスの操作手順のうち振込連絡の一例を示す。

(4) オーダエントリシステム

この種の業務は多種多様であるが、現在注文受付、注文変更、売上報告、在庫照会などに多くの音声応答システムが利用されている。

従来は販売店から電話により人手で受け付けパンチし、コンピュータ処理していた業務が、音声応答システムにより自動化された。販売店から商品コード、数量、配送日時などのデータを、音声のガイダンスによりプッシュホン(又はプッシュボタンアタッチメント)から入力し、このデータをホストコンピュータで処理することにより、生産の合理化、省力化などを図っている。また、不動産などの情報を一元管理し、代理店や特約店などから不動産売買情報の照会用としても使用されている。

表3 日立音声応答システムの製品系列 製品系列の主要仕様を示す。

項目	形式	HIVORS-1300	HIVORS-1200	HIVORS-1100	HIVORS-1000	HIVORS-900
基本方式		録音編集方式				
記録方式		デジタル(PCM)記録方式				
音声素片		文節、単語又は単音節				
単位語の長さ(s)		1.1	0.8	0.5	0.8	0.5
収容語彙数		512~2,048	256~2,048	768~1,536	512~2,048	16~256
音声記憶媒体		磁気ドラム	固定ヘッド磁気ディスク			ICメモリ
同時処理回線数		128~768	32~256	4~16	4~20	4~30
適用例		日本国有鉄道の電話予約 日本中央競馬会の電話投票などの大形システム	銀行の大都市集中の大形システム	オーダエントリ在庫照会などの中・小形システム	銀行・オーダエントリなどの中・小形システム	オーダエントリなどの小形システム

このような多岐にわたるユーザーの使用状況や各種アプリケーションに対応するため、表3に示す製品系列をもっている。また、最近開発した比較的小規模向けの音声応答装置(HIVORS-1000形)の外観を、図2に示す。

このほか、PARCOR方式を採用した画面別音声ファイル装置⁴⁾や、波形符号化方式を採用した駅の自動放送装置⁵⁾など、広く応用されている。

3.3 端末形装置への応用

日立製作所では、日本電信電話公社の指導のもとに昭和54年10月国内初の音声合成用LSIを開発した。このLSI化音声合成チップの出現により、音声を操作の誘導などの補助手段として用いたコンピュータ端末への応用や、ATM(自動銀行端末)、自動販売機などへの応用が容易となり、急速に普及してきた。ここでの音声出力は次に述べるような役割を果たし、その威力を発揮している。

- (1) 操作手順の誘導
- (2) 入力情報や操作結果の確認
- (3) 操作ミスや異常発生時の警告

また高信頼性、小形化が望めるため、車載用機器や連続使用するエレベーター、エスカレーターの案内放送⁶⁾、家庭電気品、教育機器などにも応用され始めている。

4 音声入力技術

4.1 音声認識

音声認識は、次に述べるように分類できる。

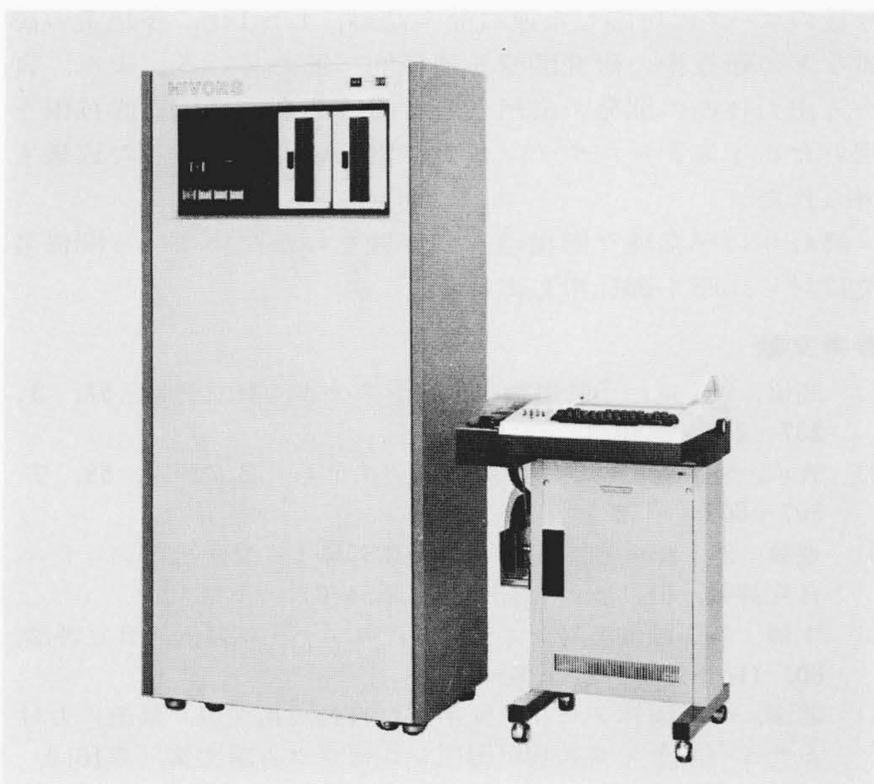


図2 HIVORS-1000形音声応答装置の外観 HIVORS-1000形音声応答装置の外観を示す[外形寸法：幅750×奥行450×高さ1,600(mm)]。

表4 音声認識による銀行システムの入出力手順 振込連絡・残高照会システムのうち、残高照会の一例を示す。

音声入力		順序	出力音声
入力項目	入力音声		
センタ呼び出し	×××-××××	→	こちらは〇〇銀行テレホンサービスセンタです。加入者番号をどうぞ。
加入者番号・確認	ビッ「」ビッ「」…ビッ「」	↙	加入者番号は〇〇〇〇〇〇ですね。暗証番号をどうぞ
	ビッ「ハイ」又は「イイエ」	→	
暗証番号・確認	ビッ「」ビッ「」…ビッ「」	↙	暗証番号は〇〇〇〇ですね。依頼コードをどうぞ
	ビッ「ハイ」又は「イイエ」	→	
依頼コード・確認	ビッ「ニイ」ビッ「サン」	↙	残高照会ですね。そのままお待ちください。〇〇日の残高は〇〇〇円です。確認コードをどうぞ
	ビッ「ハイ」又は「イイエ」	→	
確認コード	ビッ「ゼロ」	↙	ありがとうございました。

注：「」は音声を示す。

- (1) 発声内容による分類
 - (a) 離散発声^{*}された単音節や単音の認識
 - (b) 離散発声された単語の認識
 - (c) 連続発声された単語の認識
 - (d) 自然会話的に発声された音声の認識
- (2) 発声者による分類
 - (a) 特定(登録)話者の音声の認識
 - (b) 不特定話者の音声の認識

このほか、話し手がだれであるかを認識する話者認識がある。音声認識装置の技術的に困難な原因は、調音結合(隣接する音素相互の影響)や音声パターンの個人差による変動などがあり、現状では技術的に実現された機能に合わせた範囲の用途を選定し実用化されている。

4.2 センタ形装置への応用

不特定話者の音声認識に適した応用分野として代表的なものに、電話からの音声入力がある。この不特定話者の音声認識装置と音声応答システムを組み合わせることにより、回転ダイヤル電話機から数字などのデータを音声で入力し、コンピュータの処理結果を音声で聞く理想的な情報システムの構成が可能になり、音声応答サービスの利用範囲がますます拡大できる。

このような利用形態の具体例として、銀行で「振込連絡残高照会システム」に音声応答システムと音声認識装置とを組み合わせ、プッシュホンだけでなく、回転ダイヤル電話機から音声で入力するサービスが始められている。表4に銀行でのサービスの入出力手順のうち、残高照会の一例を示す。音声に

* 離散発声とは、一語一語区切って発声することを言う。

よる入力語としては、基本の数字(「ゼロ」「イチ」「ニイ」……「キウウ」)のほか、補助的に使われる言葉(「ハイ」「イイエ」「モウイチド」「トリケシ」など)の合計16語が使われる。また、発声の仕方は、離散発声であり、音声と音声の区切りの意味で「ピッ」という挿入音を発声者に聞かせている。

4.3 端末形装置への応用

特定(登録)話者の音声認識装置は、入力を特定話者に限定することで音声パターンが個人差により影響されないため、高い認識率を実現できるもので、既に各方面で実用化されている。両手を使っての作業で、手を止めずに指示や検査データなどを音声でコンピュータに入力するシステムに使用され、その適用分野は多岐にわたっている。日立製作所では、音声認識装置の導入を進めるに当たって、本社で統一的に推進を図り、研究所と共同で各工場の生産設備への応用を図り、システムの実用化を推進している。具体例として、工場では、ファクシミリ製造ラインの総合検査作業に導入しており、(1)検査作業の合理化、(2)データ集計、記述、品質管理、処理の機械化などにより省力化の効果を挙げている。また別工場では、電子顕微鏡の検査作業に導入することにより、(1)半暗室内での検査作業の合理化、(2)記述作業をなくし、報告書作成の機械化などにより省力化を図っている。そのほか日立製作所各工場では、音声認識装置の応用システムを計画中であるが、主なものとして(1)検査工程の合理化、(2)自動倉庫などの荷物仕分用制御機器への音声入力、(3)各種データのコンピュータへの入力などのアプリケーションに実用化される。こうした社内での実用化に際し、ソフトウェアの開発を行ない、エンドユーザーが音声認識システムを導入する場合、ソフトウェア作成の負担を低減できるようにした(詳細は、本特集号の別稿「音声入力システム用問題向き言語の開発」参照)。

前記の音声認識装置は単語認識装置とも言われるもので、入力された音声の内容を識別するものである。これに対して、話し手がだれであるかを識別する音声認識装置として話者認識装置がある。この話者認識装置は、原理的には単語認識と同様の処理を行なうが、多数の人の登録パターンの中から本人の音声パターンを選び出し結果を出力する。話者認識装置

は、機密部署への出入管理などに使用することができ、次のような特長がある。

- (1) 何も携帯しなくてよい(鍵、磁気カードなどが不要)。
- (2) 紛失、盗難、複製などの心配がない。

図3に話者認識装置の外観を、表5に主な仕様を示す。この装置には、誤り率の低減と処理時間の短縮のためテンキーによる暗証番号との併用方式を、また、音声の経年変化対策として自動更新方式などが採用されている。

5 音声入出力の新しい応用分野

以上述べたように、音声の入出力技術は既に各分野で利用され始めており、マンマシンインタフェースとして今後も更に新しい応用分野が開けてくるものと思われる。

音声入力装置と音声出力装置を組み合わせることにより、質問・回答・確認の形で人間が機械と対話することができるようになり、新しい応用分野として、今後次のような分野にますます利用されていくものと思われる。

- (1) 電話での応用
 - (a) 各種の案内、通知、照会などの業務
 - (b) 各種の予約、問合せ業務
- (2) 作業現場での応用
 - (a) 機械の制御、指示
 - (b) 検査データなどの入力
- (3) オフィスオートメーションでの応用
 - (a) 情報の検索
 - (b) 音声タイプライタ
 - (c) オフィスコンピュータの入力

これらの新しいニーズにこたえてゆくためには、更に解決しなければならない幾つかの課題がある。電話での応用に関しては連続音声認識の問題、作業現場での応用に関してはLSI化による小形化・低価格化、オフィスオートメーションでの応用に関しては多語彙の認識といった技術の開発が待たれる。日立製作所では研究所の協力のもとにこれらの技術の研究開発に鋭意努力中である。

6 結 言

音声入出力技術の特長と応用例について紹介したが、更に今後のニーズに対応した連続発声認識、LSI化、多語彙の認識などの新技術の研究開発を積極的に進めている。また、音声入出力技術の開発の応用として実時間音声分析合成技術を用いたボイスメールやデジタル電話の開発といった成果も得られた。

終わりに平素種々御指導、御鞭撻をいただいている関係各位に対し、厚く御礼申しあげる。

参考文献

- 1) 高橋, 外: 電話予約用音声応答システム, 日立評論, 57, 3, 237~242 (昭50-3)
- 2) 吉沢, 外: 勝馬投票用音声応答システム, 日立評論, 59, 7, 597~602 (昭52-7)
- 3) 遠藤, 外: 音声応答装置“HIVORS”による受注出荷システム, 日立評論, 61, 6, 435~438 (昭54-6)
- 4) 土師, 外: 画像応答システムの音声ファイル装置, 日立評論, 60, 11, 819~822 (昭53-11)
- 5) 北澤, 外: 固体メモリを使用した自動放送装置, 鉄道におけるサイバネティクス利用国内シンポジウム論文集, 第16回, 79~83 (昭54-11)
- 6) 弓伸, 外: 音声合成技術の現状とその応用, 日立評論, 63, 6, 405~410 (昭56-6)



図3 話者認識装置の外観 HRS-10話者認識装置の外観を示す〔外形寸法: 幅430×奥行450×高さ300(mm)〕。

表5 話者認識装置の主な仕様 話者認識装置の主な仕様を示す。

項 目	仕 様
形 式 名	HRS-10話者認識装置
登 録 人 員	150人
入 力 語 長	約0.5~1.0秒
サイレントインターバル	0.5秒以上
学 習 回 数	1回又は3回/パターン