

高齢者・障害者用メカトロ福祉機器の開発

～音声認識技術とボイスチェアの試作～

鎌田 英博, 綿貫 幸宏
安河内義明, 田栗 匡
吉成 哲, 吉田 佳史*

Development of Welfare Equipment Applied motion-technology for aged or handicapped person

Hidehiro KAMATA, Yukihiro WATANUKI
Yosiaki YASUKOUCHI, Tadashi TAGURI
Satoshi YOSINARI, Keiji YOSHIDA

抄 録

音声認識技術を応用して会話によるマンマシンインターフェースを開発した。これは人の話を装置側で聞くために音声認識機能を、話しかけるために録音再生機能を有する LSI を使用し、人と会話しながらその希望を認識し入力する装置である。本システムでは特定話者対応型の音声認識 LSI を使用したため小型化を実現し、機器内への組み込みが可能となり、かつコントローラ部のコストダウンも実現できた。さらにこのシステムを利用して人と会話しながら自らの姿勢を変動させる椅子（以下ボイスチェアと呼ぶ）を試作した。椅子本体はフレーム構造とし軽量化を図り、また駆動部では電動シリンダー一本で椅子の座部が後方に沈みながら背が起き上がる形式の効率良い機構を実現するなど使用者サイドに立った装置の設計を提案した。

1. はじめに

高齢者が住み慣れた地域や家庭でできるだけ自立し、安心して暮らすと共に積極的に社会参加することができるよう、これらを支援する技術や各種福祉機器の開発が求められている。

高齢者や障害者の利用を想定した機器の開発に当たっては、高度な技術を製品の機能性向上へつなげるばかりでなく、よりヒューマンフレンドリーなコンセプトを掲げて進める事も重要である。このためのマンマシンインターフェースに関する研究は福祉機器の分野以外にも様々な手法が提案されてきた。しかし、これらの研究によって提案されている作業は研究室レベルのものも多く、必ずしも十分な性能を得られるとは限らない。市場性やコストを考慮すると実用化のための応用研究が必要となっている。

そこで従来より福祉機器への導入が期待されている音声認識技術を応用して音声会話によるマンマシンインターフェースを開発した。これは人の話を聞くために音声認識機能を、

* ㈱市村製作所

話しかけるためには録音再生機能を使用し、人と会話しながらその希望を入力する装置である。本システムでは特定話者対応型の音声認識 LSI を使用したため小型化を実現し、機器内への組み込みが可能となり、かつコントローラ部のコストダウンも可能となった。

さらにこの装置を利用して人と会話しながら自らの姿勢を変動させる椅子（以下ボイスチェアと呼ぶ）を開発した。椅子本体は溶接によるフレーム構造とし軽量化を図った。また駆動部では全ての機構を椅子の座部に配置し、リクライニング、肘掛けの上下、椅子全体の回転のためにアクチュエータを制動し、さらにアクチュエータ駆動の防音効果も検討した。

本報告では作成した音声認識・再生装置の基本構成および音声会話システムの内容を示し、これを利用して開発したボイスチェアの構造や作動方法について報告する。

2. 音声認識技術

人の声をマンマシンインターフェースとして利用した代表的なシステムは自動翻訳装置であり、現状での使用方法は国際会議などでの同時通訳を目的としている。また、航空券な

どのチケットを電話で予約する場合に利用する試みもなされている。両者とも不特定話者に対応して開発されているがその内容はある専門分野の会話に限定されたものであり、大型のシステムを駆使したものである。したがって、この装置自体の小型化は当面現実性が薄いため、特定話者の認識に限定した装置でコンパクト化を実現し、かつコストダウンによる実用化を目指したものである。

製作した装置は音声認識用に TC8861F, TC8865F-01 (㈱東芝), 64K ビット DRAM の 3 チップで特定話者離散単語音声認識システム⁽⁶⁾を、また TC8835N で音声録音再生システムを作成した。さらにこの 2 系統のシステムを統合し、ホストに 8 ビットマイクロコンピュータボード (UECZ04, ㈱梅沢無線電機) を使用した音声会話システムを開発してボイスチェアの制御部へ組み込んだ。

2.1 認識 LSI によるシステム

認識時のデータのの流れの概要を図 1 に示す。認識 LSI のホストインターフェース回路は 8 ビットのパラレルデータバスから読み込み・書き出しが出来るため、ホストシステムは認識 LSI の全ての制御をこのデータバスを介して行う。

TC8861F はマイクから入力された信号をプリアンプで増幅する機能を持つ。増幅率の調整は pout 端子出力が発声時には 5v に環境音のみの場合には 0v に近似させる。一方、認識時の入力レベルは音量調節表示信号としてホストシステムから確認することもできる。

入力した音声はバンドパスフィルタなどで変換された後 A-D 変換されて出力される。この際、しきい値レベルを変更することで信号に含まれる環境ノイズを削除している。

TC8865F は音声区間の検出、ホストシステムインターフェ

ース、辞書の作成、認識 (複合類似度計算) 等の機能を持つ。ホストシステムはホストインターフェースを通して認識 LSI システムのコマンドおよびデータの読み書きを行うことができる。この時には認識 LSI のステータスレジスタのステータス信号 HWRDY および HRRDY を利用してハンドシェイクを行う。HWRDY が “1” であれば認識 LSI がホストからの書き込みを待っている状態であり、HRRDY が “1” のときには認識 LSI からリターンコードなどの出力を準備している。この HRRDY および HWRDY レジスタの確認は、それぞれサブルーチンプログラムを作成して迅速な対応を可能にした。

辞書の作成方法を図 2 に示す。登録作業は対象単語番号を指定したあと 1 単語につき 3 回の発声によって完了する。もしリターンコードで登録エラーが確認できた場合には再登録を行う。

図 3 にバンク設定の方法を示す。登録単語を特定のバンク (群) に指定して分類すると、指定したバンクの単語群のみが認識対象となる。その結果、毎回全単語を対象に認識する事と比較して処理時間を短縮できる。

環境騒音 (dB) およびマイクからの距離 (cm) に関する単語別の認識結果を表 1 に示す。この実験の範囲内ではマイクからの距離や騒音レベルの影響が極端に認識結果に影響を与えていない。この表の中で●印はエラーでその下に添記した数値が第 1 位で出力した単語番号である。また、○印は正常に認識したことを示し、() 内の数値は第 2 位となった単語の番号である。したがって、この表に示す様なデータを参考にして類似度が高くエラーを発生し易い単語を削除すればシステムの認識性能を向上させることが可能と思われる。

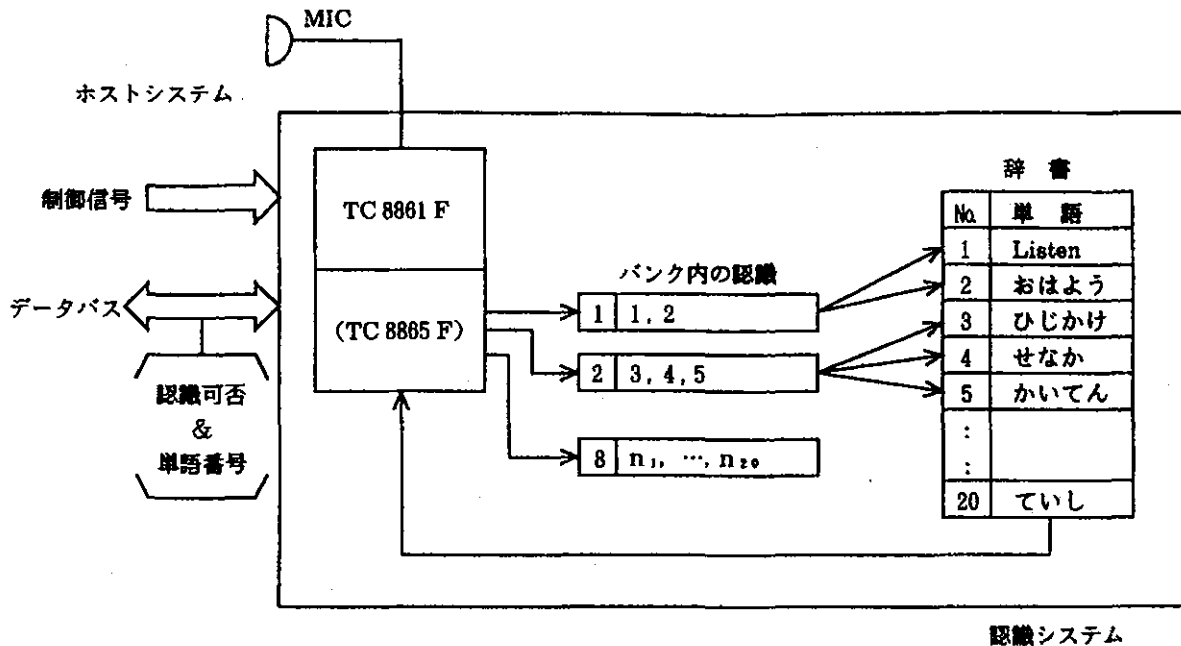


図 1 認識時のデータの流れ

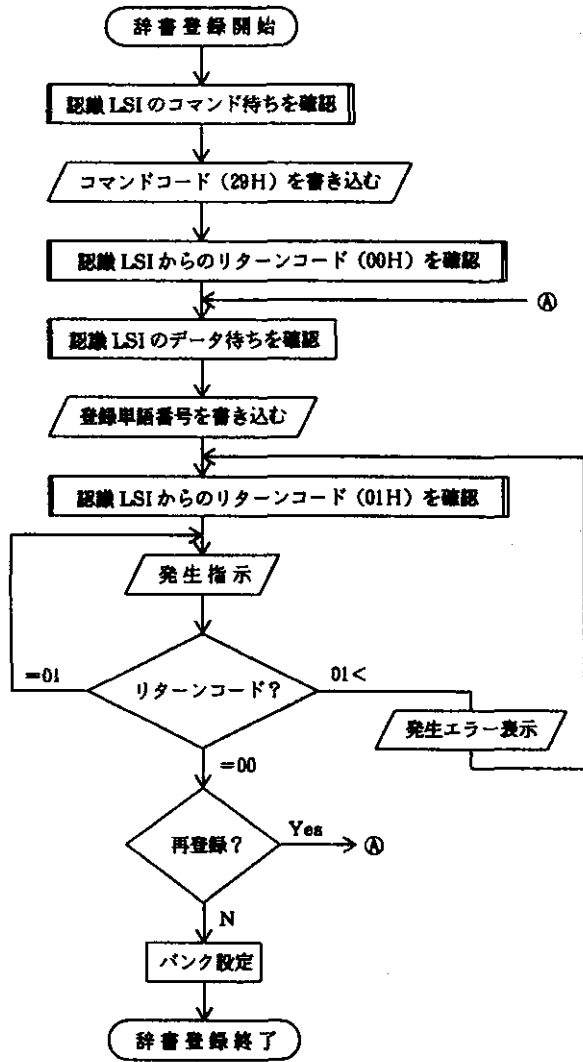


図2 辞書の作成方法

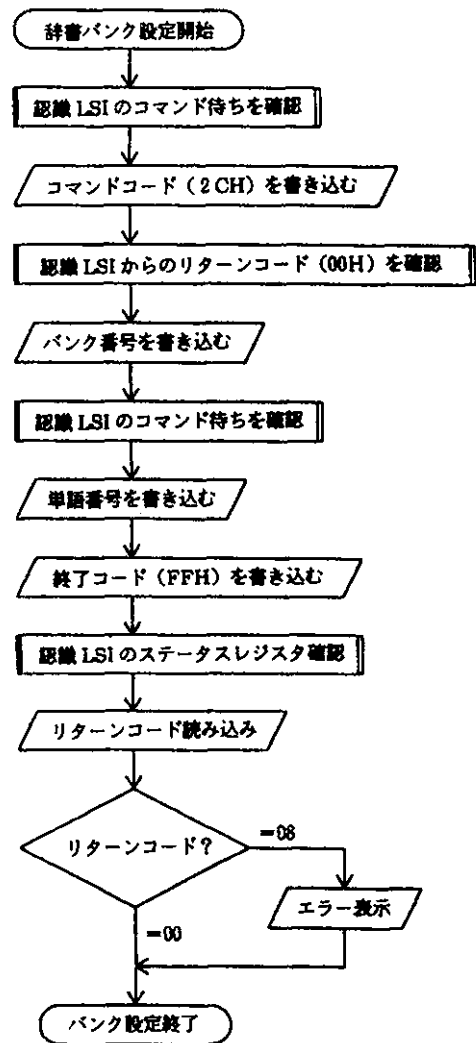


図3 バンク設定の方法

表1 単語別の認識測定結果

対象単語	37dB				46dB				55dB			
	5	20	50	100	5	20	50	100	5	20	50	100
1. う え	○ (2)	○ (3)	○ (5)	● (9)	○ (3)	○ (3)	○ (3)	○ (3)	○ (3)	○ (3)	○ (9)	○ (3)
2. し た	○ (7)	○ (6)	○ (7)	○ (6)	○ (7)	○ (6)	○ (7)	○ (7)	○ (7)	○ (7)	○ (7)	○ (7)
3. アップ	○ (1)	○ (1)	○ (1)	● (4)	○ (8)	○ (8)	○ (8)	○ (4)	○ (8)	○ (8)	○ (7)	● (4)
4. ダウン	○ (9)	○ (9)	○ (9)	○ (9)	○ (9)	○ (9)	○ (9)	○ (9)	○ (9)	○ (9)	○ (9)	○ (9)
5. み ぎ	○ (1)	○ (1)	○ (6)	○ (6)	○ (1)	○ (1)	○ (6)	○ (7)	○ (1)	○ (1)	○ (6)	○ (6)
6. ひだり	○ (7)	● (9)	● (9)	● (4)	○ (2)	○ (7)	○ (9)	○ (4)	○ (7)	○ (7)	○ (3)	○ (4)
7. ライト	○ (8)	○ (6)	○ (6)	○ (9)	○ (8)	○ (6)	○ (6)	○ (6)	○ (8)	○ (8)	○ (6)	○ (9)
8. レフト	○ (7)	○ (3)	○ (3)	○ (1)	○ (3)	○ (3)	○ (3)	○ (3)	○ (3)	○ (3)	○ (3)	○ (4)
9. ま え	○ (4)	○ (4)	○ (4)	○ (4)	○ (1)	○ (1)	○ (1)	○ (4)	○ (1)	○ (1)	○ (1)	○ (1)
10. うしろ	○ (2)	○ (2)	○ (2)	○ (2)	○ (2)	○ (2)	○ (2)	○ (2)	○ (2)	○ (2)	○ (2)	○ (2)

※○は認識した場合、●は認識しなかった場合である。

図4に認識の方法を示す。辞書登録処理で予め登録されている辞書データの中から入力音声と類似した波形を検索する。このサブルーチンでの認識結果の出力は13バイトであるが、その内、ヘッダと第一位単語番号のみをメモリへ格納することで処理時間短縮を図った。また、音声入力信号には必ず環境騒音が含まれているが、ノイズ対策のため次の機能が装備されており用途に応じて有効に利用できる。例えば、認識段階では認識コマンド指示前の約1秒間の平均ノイズか入力直後のノイズのどちらかを騒音レベルとして採用し、認識動作の判断に利用する。このフィルターは騒音状況に応じて選択可能であるためコマンドの特性とも併せて使用時期を検討した。

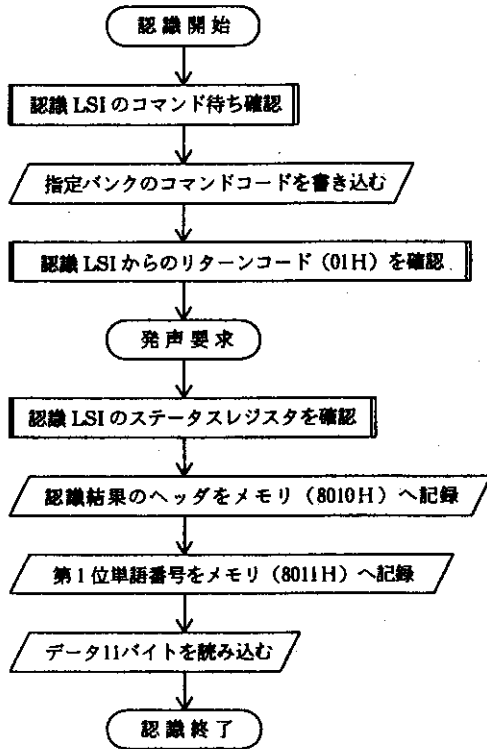


図4 認識の方法

2.2 音声録音再生 LSI によるシステム

音声録音再生用 LSI TC8835N は ADM (Adaptive Delta Modulation) 方式による録音・再生を行う LSI である。この LSI のコントロールは CPU とマニュアルの両者を選択できるが、本システムでは CPU コントロールモードを選択した。この方式ではホストシステム (Z80) の I/O ポートと LSI の D0, D1 端子を接続し 4 種類のコマンドと 2 ビットのステータスレジスタを利用して音声再生録音作業を行う。その特徴は次の様である。

- ① 音声データメモリとして最大 1M ビット DRAM を 2 個まで接続可能である。
- ② 最大 16 フレーズの録音・再生が可能である。
- ③ 最大録音再生時間は 1M DRAM が 2 個の時約 128 秒である。
- ④ 2 種類のビットレート (16k or 22kbps) を選択することにより音質や録音・再生時間の変更が可能である。
- ⑤ パッケージは SDIP28 ピンと SOP28 ピンの 2 種類から選択できる。
- ⑥ 5V 単一電源動作である。

図 5 に録音・再生時のデータの流れを、図 6 に録音作業の方法を示す。フレーズ番号 (録音する語句の番号) を指定した後、録音を開始する。録音中に停止するときは、REC STOP コマンドを実行する。このとき録音終了時のアドレスカウンタの内容がインデックスエリアに記録され動作を終了する。この録音作業ではフレーズ内容の変更や追加の場合は録音時

間を検討して、他のフレーズに影響を及ぼさないように注意する必要がある。

再生動作は録音時に設定したフレーズ番号を指定して行う。再生を終了時にアドレスカウンタはストップアドレス + 1 となるため、このアドレスを参考に別の単語を指定し、任意の時間に再生音を発声させることが出来る。

したがってユーザの好みに応じた声で日常親しんでいる単語を登録することが可能となり、会話システムに安心感や信頼感を与えることが出来た。

なお、本 LSI は電源投入後のシステムイニシャライズが必要である。この場合 RESET, RD, DO, D1, WR のそれぞれを急峻に立ち上げる必要があるが、電源の立ち上がりの遅れなど何等かの原因でリセットが行われにくい状況も発生したため、リセット端子のみをハード的にイニシャライズする必要があった。

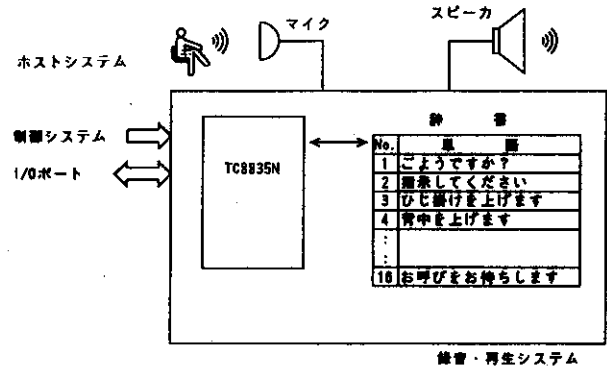


図5 録音・再生時のデータの流れ

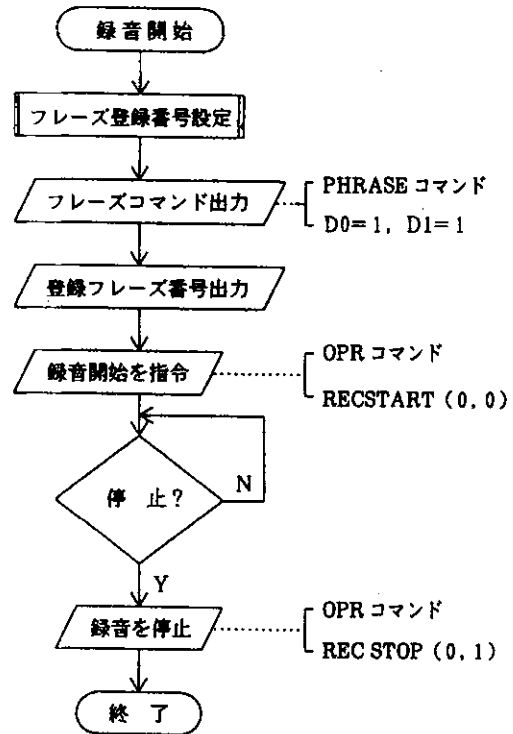


図6 録音の方法

3. ボイスチェア的设计・製作

3.1 機構部の設計

写真 1,2 にボイスチェアのフレーム構造の外観を、表 2 にはボイスチェアの仕様を示す。本装置は溶接によるフレーム構造となっており一定の強度を保持しつつ軽量化を図った。また、駆動部はリクライニング機能、肘掛けの上下動、椅子全体の回転機能で構成されている。

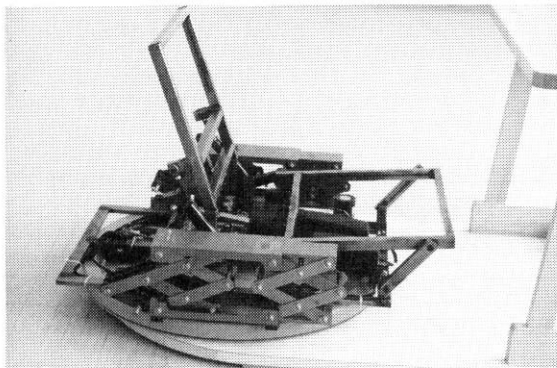


写真 1 ボイスチェアのフレーム構造(1)

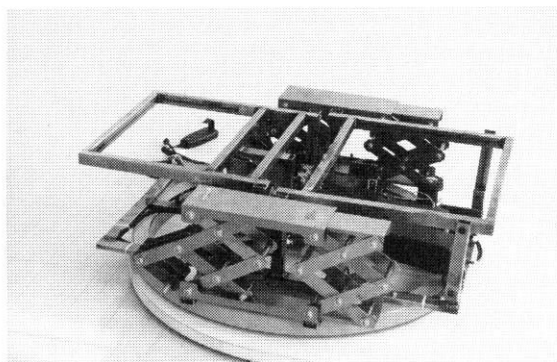


写真 2 ボイスチェアのフレーム構造(2)

表 2 ボイスチェアの仕様

寸 法	φ1200mm×H1400mm
リクライニング角度	0°~80°
座面傾斜角	-5°~20°
座面回転角	240°
肘掛け上下機構	ストローク 200mm で片側づつ単独動作可能
ウレタンクッション厚	t=80~100mm (厚めにする)
スイッチ部	音声指令、手動操作の 2 系統

本体(ソファ部)はラウンジチェアとしての利用を考慮し、フルフラットとなる。さらにパーソナルでの使用の場合はリクライニングとし、肘に可動性を持たせ、ベース部は回転できる構造とした。リクライニング部は座面に傾斜角度を持た

せ、マットと背の干渉を防ぐため座面が前方に逃げながら背が起き上がる機構である。この機構は一本の電動シリンダとリンク機能を使用して効率よく座り易い姿勢に変えることが出来る。

肘掛け部はパンタ機構とし、シリンダを横置きにして 1 リンクによる簡易パンタで上下動作を行う。この機構は取付空間の省スペースが可能である反面、構造が複雑となった。

本体回転機構は市販の回転ガイドを利用し、モータ軸に取り付けたゴムローラで移動するものであり、最高回転数 2 RPM である。

本体機構部品を保護し、美観を向上させるためと座面や背もたれ面のクッション性をよくするため、カバーリングの検討を行った。写真 3, 4 に最終仕上がり状態を示す。

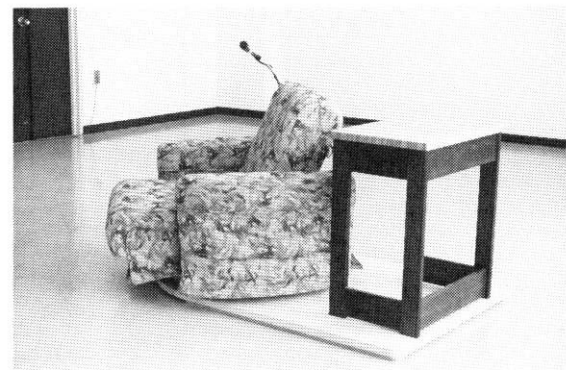


写真 3 ボイスチェア最終仕上がり(1)

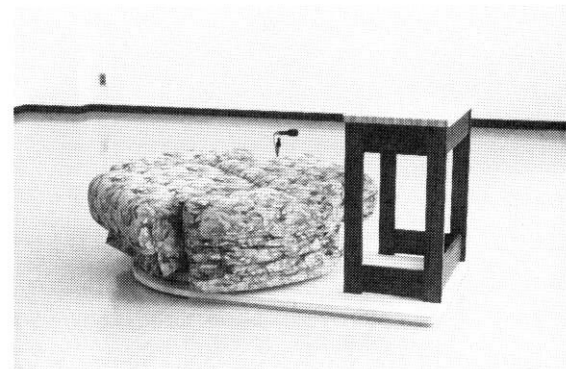


写真 4 ボイスチェア最終仕上がり(2)

3.2 音声による会話システム

人の話を聞いたり、人に話しかけたりしながら、使用者の希望を入力する会話システムを考案し、その入力結果に応じて自らの姿勢を変動させるボイスチェアシステムを開発した。このボイスチェアのように話者の反応を窺いながら動作するシステムでは機器の駆動が指令と異なると不具合が生じるため、高い認識率を維持する必要がある。しかし、認識 LSI はその機能も限定されているため、実用化のためにアルゴリズム構築の段階で有効な手法を考案して認識率の向上に努め

た。図7は会話システムの流れをPADで示したものである。ここで、メインルーチンの認識ブロックではバンク1に登録されている単語「Listen」を認識するまで待機している。「Listen」を認識した直後、ボイスチェアは指示を求める言葉（『指示して下さい』）を発声する。その言葉を受けてユーザーは希望する動作（バンク2に登録してある単語）を発声する。例えば「肘掛」と「上げて」を発声するとボイスチェア

は、ユーザーに対し、確認の言葉（『肘掛を上げます』）を告げて動作を開始する。希望の位置で「停止」と発声すると肘掛けを上げる動作が終了する。

もし、ボイスチェアの動作後、位置を僅かに変更したい場合には、微少動作を指定する単語（「もう少し」）により、その位置を訂正することもできる。

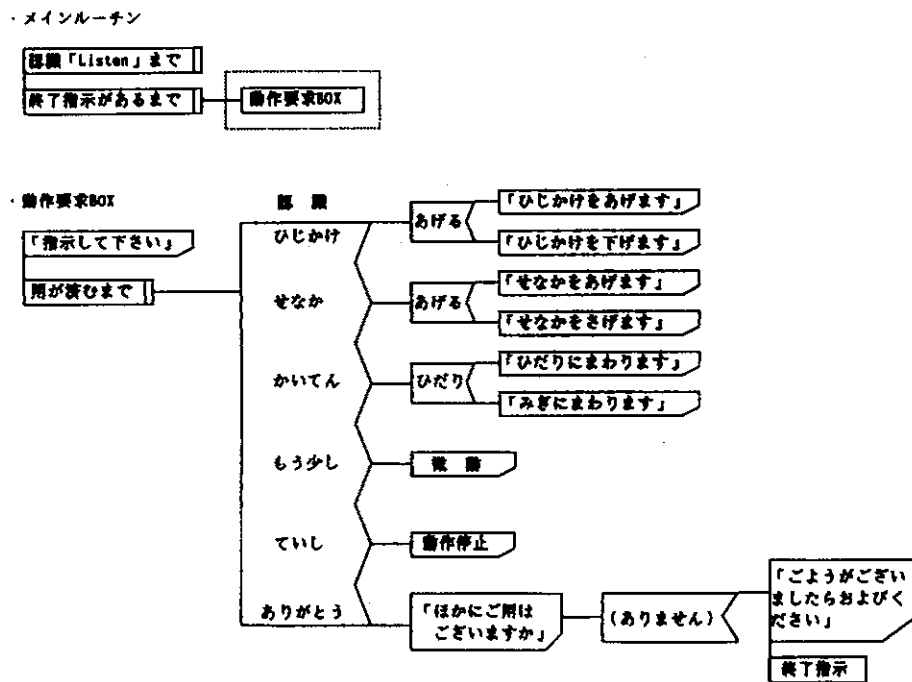


図7 会話の方法

4. まとめ

会話システムによるマンマシンインターフェースとそれを内蔵したボイスチェアを開発しそのシステム構成および動作機構を示した。その特長は以下の様になる。

- 1) Z80 ボードとLSIのハンドシェイクを可能な限り単純化し会話の反応時間を小さくするなど、きれのよい会話を目指した。
- 2) 本システムでは同一単語を複数のバンクへ登録できる。したがって、不慮の事故を防止するため、『停止』や『ストップ』などの単語を必要な箇所に多数登録した。
- 3) 本システムは特定話者を対象にした認識方式を採用しているが、辞書登録単語数に余裕があれば、複数の人の声を登録し、グループ話者対応型として使用可能である。しかし、移動の命令に付いてはユーザー以外の人の声に反応しない様にするべきであるが、停止指令などにはグループ話者認識を採用することとした。
- 4) 認識率を向上させるためにバンク機能の利用や類似単語

の削除などの対策を採用した。

本研究は心身の健康の保持、機能回復、住居居住空間の向上に有効な福祉機器の開発を意図したものである。開発した会話システムはヒューマンサイドテクノロジーとして人が日常使っている会話パターンや単語を使用してアルゴリズムを構築した。またメカトロ技術を活用したチェアにこのシステムを搭載することで安心感や親近感を与える事が出来た。本研究でこれらハイテクを積極的に活用し、ボイスチェアの製品化に向けた技術集積がなされたことは意義深いものと思う。今後はこのプロトタイプを基にフィールドテストを継続し、練度の高い製品として完成させたい。

高齢者・身障者用機器開発は日常の生活空間に必要な機器をユーザーサイドで改良していく事が基本であり、地域の生活環境との調和や機器の保守の面からも道内の製造業が主体となり得る分野である。さらに、近年の消費者ニーズの傾向を考慮するとインテリア製品群にもハイテクの要素を付加した製品が新規な需要を促す可能性もある。

本研究は産学官の共同研究として行ったが、道内の製造業

や関連装置産業が培った製造技術を製品開発へも活用しようとするものであり、融合化技術を基盤とした新規性あるプロジェクトである。今後の発展を期待したい。

なお、研究を推進するに当たってご尽力いただいた関係者の方々に厚く御礼申し上げるとともに、本体のカバーリングに快くご協力いただいた(株)ヤマチファニチャーの方々に深く感謝の意を表します。

参考文献

- 1) 新田恒雄他;電子情報通信学会, SP91-114, P49-56, (1990)
- 2) 正井康之他;日本音響学会講演論文集, P13-14, 平成3年10月
- 3) 小林晃一他;第6回リハ工学カンファレンス講演集, P441-444, (1991)
- 4) 山田智一他;電子情報通信学会, SP91-26, P65-72, (1991)
- 5) 松井和子;日本音響学会講演論文集, P13-14, 平成5年3月
- 6) 半導体データブック “音声 LSI”, 株式会社東芝