

音声制御ロボットシステムの研究開発

宮城職業能力開発短期大学校 中 田 英 次
青森職業能力開発短期大学校 佐々木 隆 幸

A Study on Robot System with Speech Recognition

Eiji NAKADA, Takayuki SASAKI

要約

本文は、人間の音声による指示命令で作動するロボットシステムを研究開発した報告書である。ここでのロボットは定型作業を自動的に行う従来のロボットとは違い、人間が指示命令した言葉を聞き取り、しかもその言葉どおりに動くというロボットである。つまり、聴覚を備えたロボットシステムの研究開発である。

研究開発の進め方として、最初に構築するロボットシステムを認識部と制御部に分割する。認識部は人間の音声認識するための音声認識部であり、制御部はロボットアームを構成するモータを精密にしかも確実に制御するロボットアーム制御部である。それぞれを研究開発した後、認識部と制御部を統合結合して音声制御ロボットシステムの構築完成とする。

構築したシステムの能力は次のとおりである。ロボットアームの作動制御用として設定した15個の単語に対して、正しく聞き取ったかどうかの音声認識の正解認識率は96.7%、また聞き取ってからロボットアームが作動開始するまでの演算処理時間は最大で1秒以内、そしてロボットアームが移動する最小距離は0.1mmである。

今回研究開発した音声制御ロボットシステムが、システムの教育訓練、デジタル信号処理およびデジタル制御のそれぞれの学習教材に有効活用できるものと期待している。

I はじめに

モータの技術は著しく向上している。「ただ回ればよい」から「回転速度、回転角をいかに精度よく高速に制御するか」という技術に移行してきている。

一方、コンピュータ技術の向上も著しい。「いかに高速に計算させるか」から「いかに知能をもたせるか」という人工知能の技術開発の段階に突入している。

そこで、これら二つの関連技術を組み合わせた音声制御ロボットシステムを研究開発することにした。それは、定型作業を自動的に行う従来のロボットシステムとは違い、人間の音声でロボットアームを随意に動かしたり止めたりすることができる音声制御ロボットシステムの構築である。それについて報告する。

II 音声制御ロボットシステム

構築した音声制御ロボットシステムの概要を図1に示す。モータ個数はロボットアーム関節の個数と同数

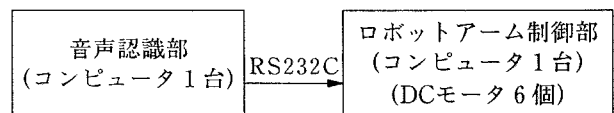


図1 音声制御ロボットシステム

の6個で、DCモータを使用する。コンピュータ台数は人間の音声認識するための音声認識部とロボットアームを制御するためのロボットアーム制御部にそれぞれ1台ずつの計2台使用する。コンピュータの使用言語は、前者がC言語、後者がBASIC言語とする。2台のコンピュータはRS232Cで接続する。

III ロボットアーム制御方法

ロボットアームの回転速度と回転角を迅速に確実に制御するためには、DCモータとその制御回路を一对とする組がロボット関節の個数だけ必要になる。しかも、それらが互いに協調作動するように制御されなけ

ればならない。それがロボットアーム制御部の役割である。

その制御を可能にするための考え方を以下に示す。簡単にするため、システムが一组のDCモータと制御回路で構成されるロボットアームとする。そのときのブロック図を図2に示す。

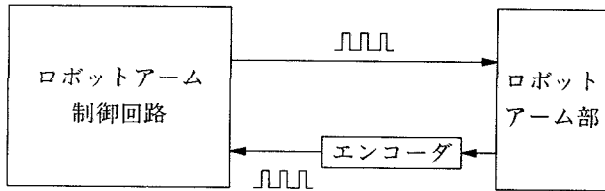


図2 ロボットアーム制御部

ロボットアーム制御部の入出力関係を明確にするため、以下に伝達関数を調べる。

1 ロボットアームの制御回路

ロボットアーム制御回路のブロックは、さらに図3のように機能分解される。

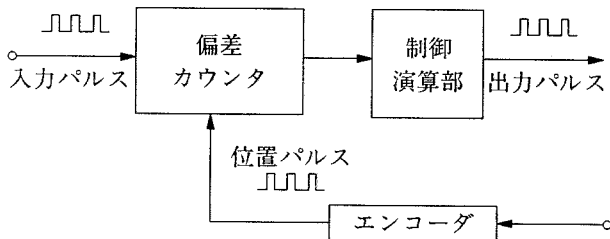


図3 ロボットアーム制御回路

図3のz変換を求める。入力パルス、その入力パルスに追従したロボットアームの現在位置を示した位置パルスおよび制御演算部のz変換をそれぞれX(z)、N(z)、D(z)と仮定すると、出力パルスのz変換Y(z)は

$$Y(z) = D(z) \cdot (X(z) - N(z))$$

と表せる。

次に、図2のロボットアーム部のブロックは図4のように機能分解される。

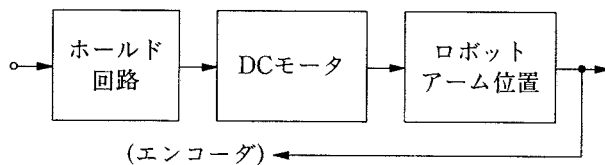


図4 ロボットアーム部

図4のz変換を求める。最初に、個々の伝達関数を求めると、ホールド回路、DCモータおよびロボットアーム位置の伝達関数はそれぞれ

$$(1 - e^{-\Delta T \cdot s}) / s, K_m / (1 + T_m \cdot s), K_r / s$$

となる。これら3つの伝達関数をまとめ、そのz変換をU(z)とすると、

$$U(z) = Z \left\{ \frac{1 - e^{-\Delta T \cdot s}}{s} \cdot \frac{K_m}{1 + T_m s} \cdot \frac{K_r}{s} \right\}$$

$$= K_m K_r \cdot \frac{a \cdot z^{-1} + b \cdot z^{-2}}{1 - c \cdot z^{-1} + d \cdot z^{-2}}$$

となる。

ただし

$$\left\{ \begin{array}{l} s, z : \text{複素数} \\ \Delta T : \text{パルス周期} \\ T_m : \text{機械的時定数} \\ K_m, K_r : \text{定数} \\ a = \Delta T - T_m + T_m \exp(-\Delta T / T_m) \\ b = T_m - (\Delta T + T_m) \exp(-\Delta T / T_m) \\ c = 1 + \exp(-\Delta T / T_m) \\ d = \exp(-\Delta T / T_m) \end{array} \right.$$

とする。

ロボットアーム位置は出力になるばかりではなく、エンコーダで位置パルスに変換され図3の偏差カウンタに負帰還される。

以上のことから、図2に示すロボットアーム制御回路からロボットアーム部までの全体の伝達関数W(z)は

$$W(z) = \frac{U(z) \cdot D(z)}{1 + U(z) \cdot D(z)}$$

とまとめられる。

この伝達関数W(z)は、入力パルスに対してロボットアームの位置がどう応答するかを表現したものである。

また、この式を変形すると、ロボットアームに対して任意の応答を設定できるようになる。すなわち、目標とする応答の伝達関数が任意のW(z)になるように設定するには、あらかじめD(z)を

$$D(z) = \frac{W(z)}{(1 - W(z)) \cdot U(z)}$$

と設定すればよい。

したがって、目標とする伝達関数W(z)を作り出すためには、上式を満たすように制御演算部の伝達関数D(z)をコンピュータプログラムで作成すればよい。

2 ロボットアームの作動可能範囲

ロボットアームが作動できる3次元空間の範囲を調

べる。図5のような球座標を設定する。

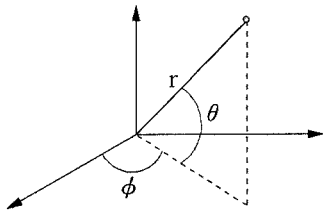


図5 設定した球座標

ϕ 方向にはロボットアーム全体が左右に $\pm 150[^\circ]$ 回転するだけであるので、簡単にするため、 r と θ を次元とする複素平面上で作動可能部分を調べる。

今回のロボットアームのモデルを図6に示す。点Pの位置 z は、 i を虚数単位として

$$z = r_1 \exp(i\theta_1) + r_2 \exp(i(\theta_1 - \theta_2)) + r_3 \exp(i(\theta_1 - \theta_2 + \theta_3))$$

と求まる。

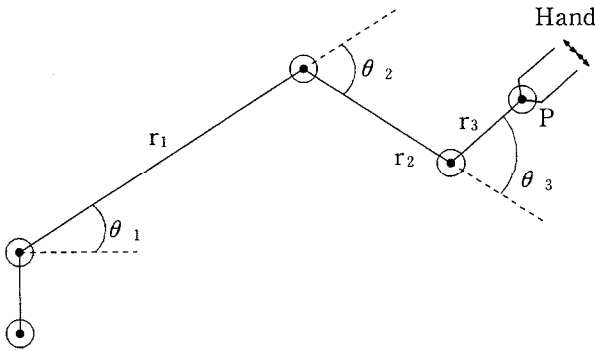


図6 ロボットアームモデル

ロボットアームの各部の長さと同転角度は次の通りである。

$$\begin{cases} r_1 = 250 \text{ mm} & -30^\circ \leq \theta_1 \leq 100^\circ \\ r_2 = 160 \text{ mm} & 0^\circ \leq \theta_2 \leq 110^\circ \\ r_3 = 72 \text{ mm} & -90^\circ \leq \theta_3 \leq 90^\circ \end{cases}$$

ロボットアームの複素平面上での作動可能部分は図7のようになる。

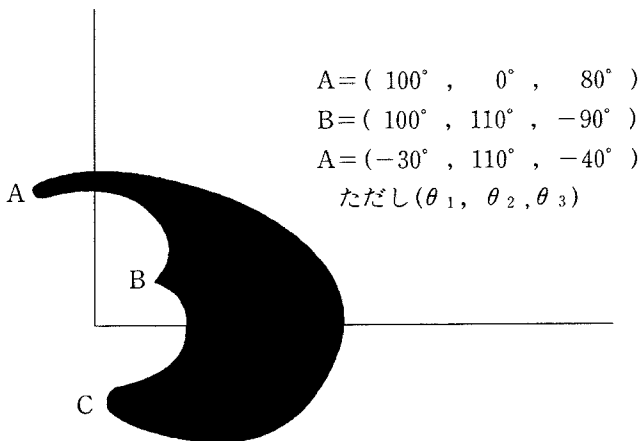


図7 ロボットアーム作動部分

したがって、3次元の空間におけるロボットアーム全体の作動可能範囲は図7の作動可能部分が、さらに ϕ 方向に $\pm 150[^\circ]$ 回転する範囲となる。

IV 音声認識方法

図1に示した、人間の音声を認識するための音声認識部を機能分解したのが図8である。音声認識部は、入力音声进行分析する音声分析と、事前に登録してある音声との比較照合を行う音声識別の二つのブロックで構成される。

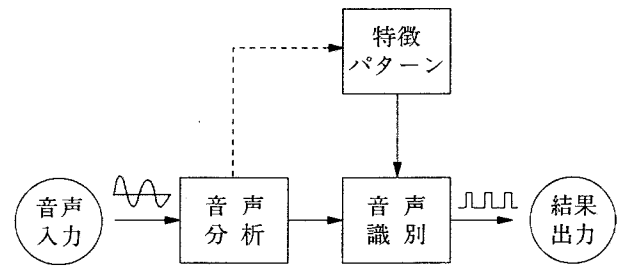


図8 音声認識部

1 音声分析

音声分析ブロックをさらに機能分解すると、図9のようになる。入力音声をA/D変換した後の波形を $x(n \Delta T)$ (n :整数、 ΔT :サンプリング周期、以後 ΔT を省く)とする。その $x(n)$ を N (N :データ個数)の時間ごとに切り出し、窓関数を乗ずる。使用した窓関数 $f(n)$ はハニング窓⁽¹⁾で

$$f(n) = 0.5 - 0.5 \cos(2\pi n/N)$$

で示される。窓関数は切り出された音声波形の両端での不連続性を小さくする働きをもっている。

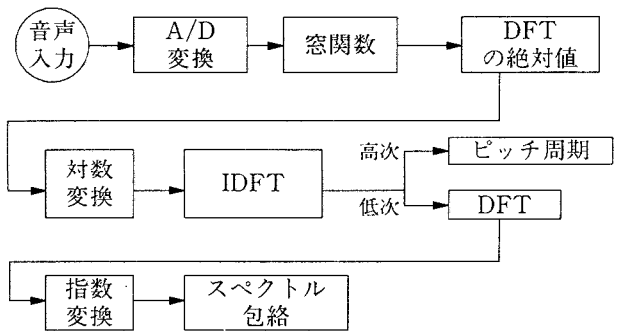


図9 音声分析

次に、窓関数を乗じた切り出し音声波形の一つを離散フーリエ変換(DFT)し、その絶対値をとる。それを $|X(k)|$ ($k=0, \dots, N-1$)とすると、

$$|X(n)| = \left| \sum_{k=0}^{N-1} f(k) \cdot x(k) \exp(-i 2\pi nk/N) \right|$$

となり、この $|X(n)|$ は周波数軸上に等間隔に N 本の

線スペクトルを形成する。

さらに、 $|X(n)|$ の対数をとる。この目的は音声スペクトル $|x(n)|$ を声帯の音源スペクトル特性 $|S(n)|$ と声道の共鳴スペクトル特性 $|H(n)|$ の和の形に分離することにある。すなわち、

$$\log |X(n)| = \log |S(n)| + \log |H(n)|$$

つづいて離散フーリエ逆変換(IDFT)を行い、それを $c(n)$ とする。

$$c(n) = -\sum_{k=0}^{N-1} (\log |X(k)|) \exp(i 2 \pi nk/N)$$

n の値が大きい部分(高次部分)に鋭いピークがあれば、それを声帯のピッチ周期とする。逆になければ、ピッチ周期はなしとする。

また、 n の値が小さい部分(低次部分)には声道の特性に関係したスペクトル情報が多く含まれているので、 $c(n)$ の低次部分についてさらに離散フーリエ変換を行い、指数変換すると、声道のスペクトル包絡が得られる。

2 特徴パラメータ抽出と特徴パターン登録

切出し音声波形をスペクトル分析した結果から、ピッチ周期に関するパラメータ1個、スペクトル包絡からその特徴を表すパラメータとして15個、計16個のパラメータを抽出する。これが切出し音声波形のデータとなり、一つの16次元ベクトルとしてまとめられる。

一つの入力音声波形は数個、数十個の切出し音声波形で構成されるので、数個あるいは数十個の16次元ベクトルの時系列としてまとめられる。つまり、音声は

$$(\text{音声}) = \{\vec{a}_1, \vec{a}_2, \vec{a}_3, \dots, \vec{a}_N\}$$

となる。これをその音声の特徴パターンとし、コンピュータに事前に登録する。

3 音声識別

指示命令として新しく入力されてきた音声についても、上と同様に音声分析し、特徴パターンの抽出する。その音声は、既に登録されている音声のどれであるかを識別するには、互いの音声の特徴パターンを照合する必要がある。このとき、各音声の時系列の長さが極端に違う場合は除外するが、それ以外のすべての音声の特徴パターンと照合しなければならない。なぜなら、同一音声であっても、個人差や発声状況の差などにより、音声の時系列の長さに差異が必然的に生じるからである。そこで長さの異なるものを照合する方法が必要になる。その方法として、ここでは動的計画法⁽²⁾を採

用する。

特徴パターンがそれぞれ N 、 M 個のベクトルで表される次のような入力音声と登録音声があるとする。

$$(\text{入力音声}) = \{\vec{a}_1, \vec{a}_2, \vec{a}_3, \dots, \vec{a}_N\}$$

$$(\text{登録音声}) = \{\vec{b}_1, \vec{b}_2, \vec{b}_3, \dots, \vec{b}_M\}$$

これらを照合するとき、単に順番に照合するのではなく、重複をも考慮しながら、次式の値

$$\sum_{m,n=1}^{M,N} |a_n - b_m| w_{n,m} / \sum_{m,n=1}^{M,N} W_{n,m} \quad (w: \text{重み})$$

が最小になるように照合する。そのときの値を、入力音声と登録音声との類似性を表す距離とみなす。その様子を図10に示す。

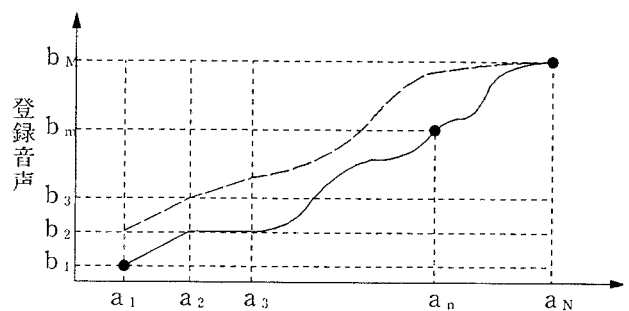


図10 特徴パターン照合

他のすべての登録音声とも照合し、その中で距離が最も短くなる登録音声を入力音声であると判定する。

V 統合結合

RS232Cは音声認識部のコンピュータとロボットアーム制御部のコンピュータを図11のように接続する。この接続で目的の音声制御ロボットシステムが構築できる。

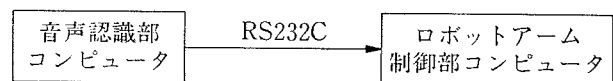


図11 統合結合

音声認識部のコンピュータが入力音声を認識し、その結果をアスキーコードに変換し、RS232Cに送り出す。

一方、ロボットアーム制御部のコンピュータはRS232Cに信号があるかどうかを調べ、信号があればそれを即時に受信する。受信したアスキーコードの種類によって、ロボットアームの制御方法を変える。

RS232Cの伝送条件を表1に示す。

表1 RS232Cの伝送条件

通信速度	9600 bps	ストップビット	1 bit
ビット長	8 bits	Xパラメータ	なし
パリティ	なし		

VI 作動実験結果

音声制御ロボットシステムを制御するために登録した音声は次の15単語である。

(登録音声単語) = { はじめ、終了、止まれ、
右、左、上、下、右上、
右下、左上、左下、
前、後ろ、
離せ、つかめ }

それぞれの登録単語の制御意味は

はじめ：初期設定位置からスタート位置に移す
 終了：任意位置から初期設定位置に戻す
 止まれ：作動中に任意の位置で止める
 右、左、上、下、右上、右下、左上、左下、前、後ろ
 ：現在位置からその向きに移す
 離せ、つかめ：図6の点Pのハンド爪を動かす
 である。

ロボットアーム作動の実験結果を表2に示し、以下にまとめる。

1) 音声の登録者が指示命令したとき、音声認識の結果が正しく認識されたかどうか(ロボットアームの作動結果で判定する)の正解認識率の平均は96.7%である。

2) 指示命令を発声し終えてから、ロボットアームが追従し始めるまでの演算処理に要した時間は最大で1秒以内である。

3) また、ロボットアームが移動する最小単位の距離は0.1 mmである。

表2 システムの作動実験結果

認識率	平均96.7%
演算処理時間	最大1秒以内
移動距離	最小単位0.1 mm

VII おわりに

音声制御ロボットシステムには、音声認識率の向上などまだ改良しなければならない課題が多々ある。しかし、指示命令に対して最大で1秒以内の応答であり、リアルタイムに近い応答を示している。以下に今回の研究開発を通しての関連成果を掲げる。

1) 音声入力で作動するロボットシステムが構築されたことにより、今後の研究開発になお一層の発展性を求めることができる。

2) 高度な機能をもつシステムとしてシステム教育訓練に役立てることができる。

3) 多数のモータを迅速に確実にコンピュータで制御していることから、モータのデジタル制御学習に還元できる。

4) 音声認識では音声波形データを様々な加工処理しているため、デジタル信号処理学習の一つに有効活用できる。

参考文献

- (1) Rabiner, L. R. and Schafer, R. W.: "Digital processing of speech signals", Prentice-Hall, 1978, p 132-141
- (2) 古井貞熙: "デジタル音声処理", デジタルテクノロジー6、東海大学出版会、1992、p 162-165