

## 目次

1. 緒言
  2. 研究の背景と目的
  3. 研究の手順
    3. 1 方式検討
      3. 1. 1 母音について
      3. 1. 2 自己相関法
    3. 2 ハードウェア構成
      3. 2. 1 ブロック図
      3. 2. 2 CPU 周辺のハードウェア構成
      3. 2. 3 リセットスイッチ
      3. 2. 4 入力スイッチ
      3. 2. 5 LED 表示器
      3. 2. 6 マイク入力とプリアンプ
      3. 2. 7 スイッチト・キャパシタ・フィルタ
      3. 2. 8 基板外観
      3. 2. 9 回路図
      3. 2. 10 部品表
    3. 3 ソフトウェア
  4. 研究の結果
    4. 1 動作確認手順
    4. 2 動作確認の結果
  5. 今後の展望
- 謝辞
- 参考文献

## 1. 緒言

本研究においては、安価なマイクロコントローラを実装した特定話者による母音認識装置のプロトタイプを作成して動作を確認し、その有効性を確認、考察した。

## 2. 研究の背景と目的

現在、個人を特定するための本人認証方式は、これまで主流であったパスワード方式のみの運用から、パスワード以外の方法の使用、あるいはパスワードとそれ以外の方法を組み合わせる方式が増えてきている。

パスワード以外の方法としては主に指紋や静脈などの生体認証方式が主流となっている。身近なところでは銀行のキャッシュカードに指紋認証を付加したものが導入されている。

音声認識の本人認証への応用は、指紋認証に比して被認証者の心理的抵抗が低い事から注目されてきた。

最近では乗算器あるいは積和乗算器を同一ダイ上に実装したマイクロプロセッサあるいはマイクロコントローラが比較的安く入手できるようになった。そのため単一基板上でスタンドアロンである程度の処理を実現するシステムを安価に実現可能になってきた。

本研究の目的は、個人認証や家電用リモコンのコマンド変換に利用可能な不特定話者音声認識装置のプロトタイプを作成する事にある。

本研究においては、安価なマイクロコントローラを実装した特定話者による母音認識装置のプロトタイプを作成して動作を確認し、その有効性を確認、考察した。

### 3. 研究の手順

#### 3. 1 方式検討

##### 3. 1. 1 母音について

安価なマイクロコントローラを使用した音声認識装置を作成するにあたって母音に特化した音声認識装置を検討する。

文の構成単位は単語であり、単語は音節(syllable)からなる。日本語の音節は通常、子音と母音からなる。日本語では、外来語の表記も含めて約110個の音節が考えられる。

[1]

母音は子音に比して、通常長い継続時間長を持ち、スペクトルも比較的明確である。よって、母音は通常容易にかつ確実に認識されることができるので、人間による音声認識でも機械による音声認識でも、重要な役割を果たしている。

##### 3. 1. 2 自己相関法

音声信号を解析するにはFFTなどで周波数成分を解析する方法がある。その場合得られる関数の横軸は周波数になる。自己相関関数は解析すべき音声信号自身をシフトしながら内積を得てその信号の周期性を調べる関数である。基本的な定義は次式で表される。

$$r(m) = \sum_{n=0}^{N-m-1} x(n)x(n+m) \quad (0 \leq m \leq N-1)$$

また、最大値が1になるように正規化した自己相関関数は次式で表される。その場合、シフト量  $m=0$  では必ず1となる。

$$r(m) = \frac{\sum_{n=0}^{N-m-1} x(n)x(n+m)}{\sum_{n=0}^{N-1} x(n)x(n)} \quad (0 \leq m \leq N-1)$$

本基板の録音時においてはサンプリング周波数  $F_s=5.9\text{kHz}$  で音声データを配列  $x[300]$  に格納し、シフト量を  $m=8$  として自己相関値を  $aa[8], ii[8], uu[8], ee[8], oo[8]$  の配列のうち該当するところに収納してテンプレートデータとする。

認識時においては入力中の音声データの自己相関値  $cor[]$  と各テンプレートデータとの2乗誤差を求め、もっとも小さい2乗誤差を持つテンプレートデータに相当する母音を認識結果として該当するLEDを点灯させている。

### 3. 2 ハードウェア構成

#### 3. 2. 1 ブロック図

ハードウェアのブロック図を Fig1 に示す。

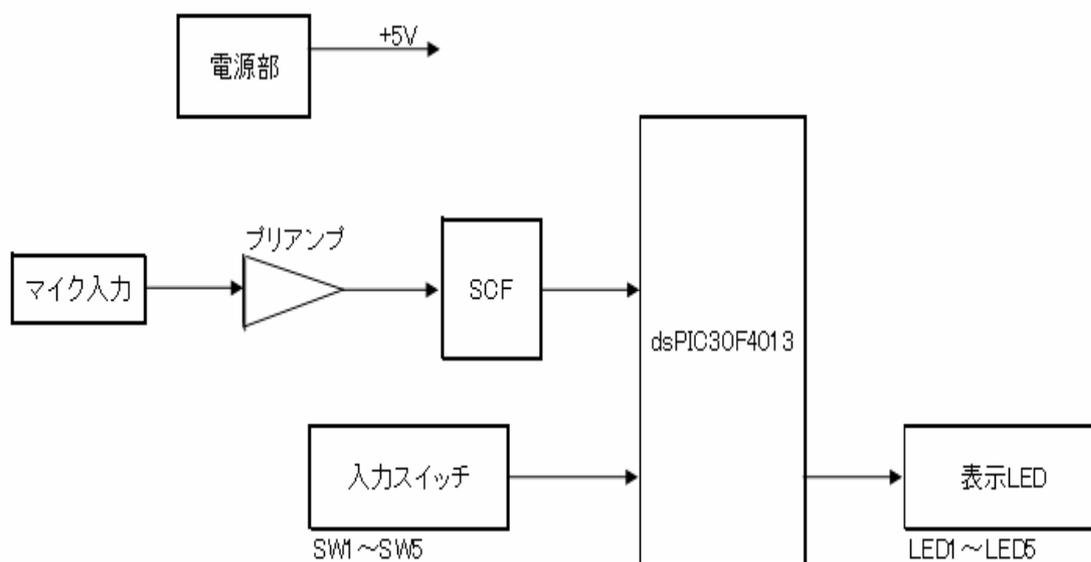


Fig1

### 3. 2. 2 CPU 周辺のハードウェア構成

CPU : dsPIC30F4013

クロック周波数 : 20MHz (セラミック振動子を使用)

プログラムメモリ : 48k byte(CPU に内臓)

SRAM : 2 k byte(CPU に内臓)

### 3. 2. 3 リセットスイッチ

SW6 を CPU のリセットスイッチとして使用  
押下で!CLR アサート

### 3. 2. 4 入力スイッチ

入力スイッチはモーメンタリーのタクトスイッチを使用。

シンボル	ポート	プログラム上の扱い
SW1	RB0	押下で"1"
SW2	RB1	押下で"1"
SW3	RB2	押下で"1"
SW4	RB3	押下で"1"
SW5	RB4	押下で"1"

### 3. 2. 5 LED 表示器

シンボル	ポート	プログラム上の扱い
D1	RF0	"1"で点灯
D2	RF1	"1"で点灯
D3	RF2	"1"で点灯
D4	RF3	"1"で点灯
D5	RF4	"1"で点灯

### 3. 2. 6 マイク入力とプリアンプ

マイク入力回路はカットオフ周波数 0.72Hz ゲイン 26.59dB(21.36 倍)とし、+5V 単一電源での使用のため、+2.5V をオフセット電圧としている。

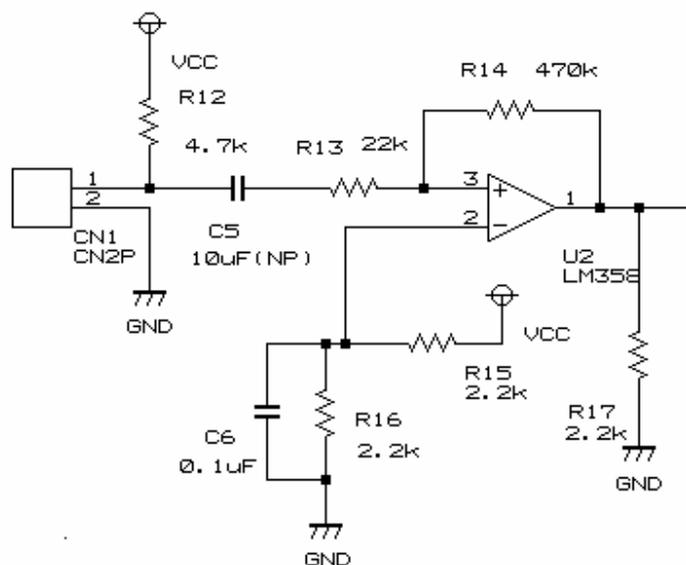


Fig 2

### 3. 2. 7 スイッチト・キャパシタ・フィルタ

A/D 変換器のエリアシングノイズを低減するため、サンプリング周波数の 1/2 以下のコーナー周波数のフィルタを挿入した。

本装置においては母音の認識に機能を限定しているので電話器の帯域を参考にしておおよそ 3kHz を目標とし、2.83kHz を最高周波数とした。サンプリング周波数はその 2 倍以上必要であるので 5.9kHz とした。

スイッチト・キャパシタ・フィルタは専用 IC の MAX7403 を使用した。コーナー周波数は 8 番ピンに接続するコンデンサの容量により下式で決定される。

$$f_c = (K \cdot 10^3) / (C_{osc} \cdot 100)$$

$f_c$  : コーナー周波数(kHz)

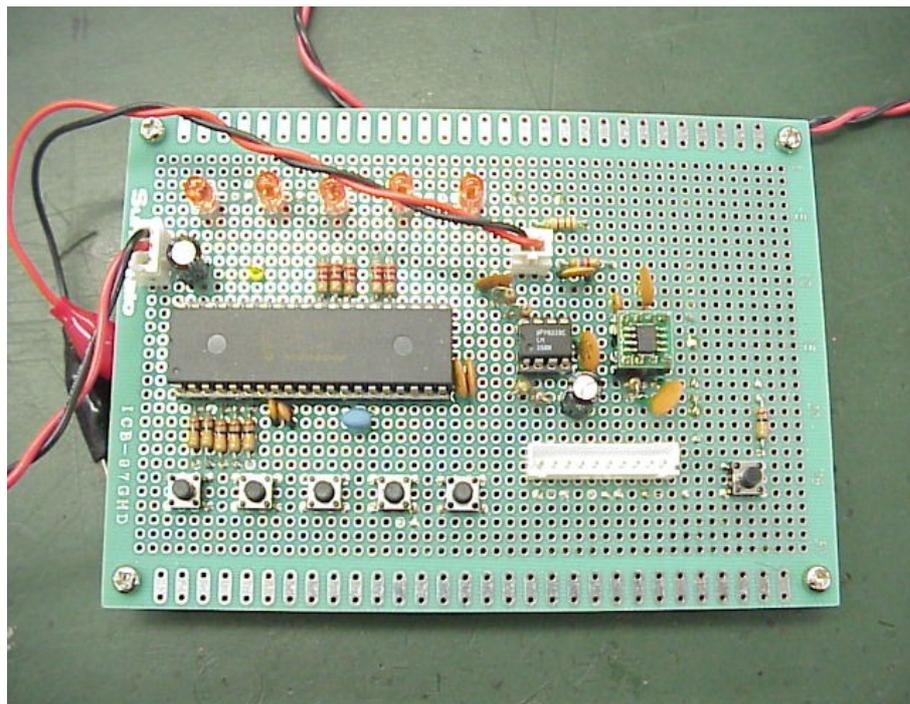
$K$  : MAX7403 の定数

$C_{osc}$  : 8 番ピンに接続するコンデンサの容量(pF)

本基板においては  $C_{osc} = 120$  (pF) としている。

### 3. 2. 8 基板外観

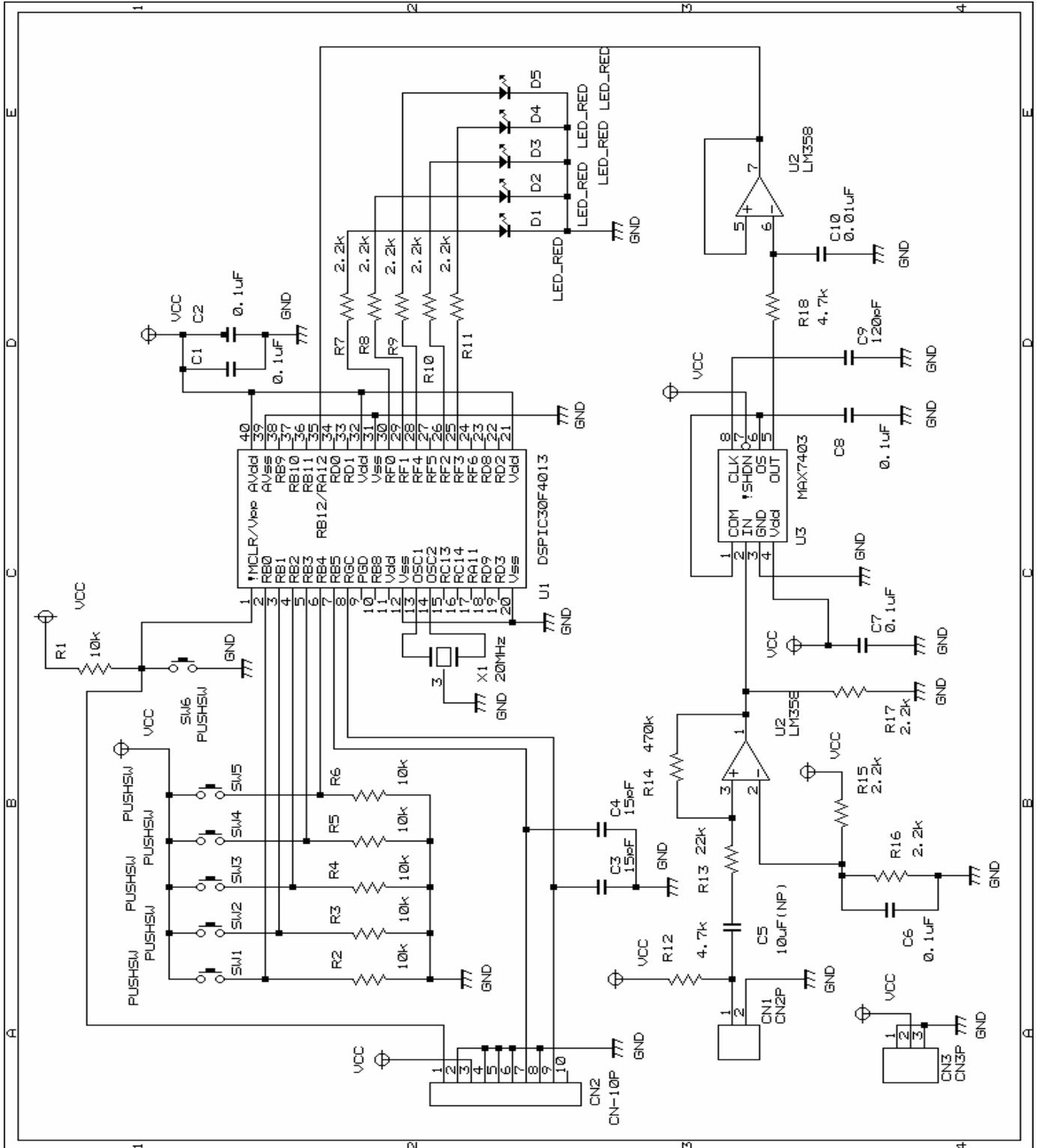
基板外観をF i g 3に示す。



### 3. 2. 9 回路図

本基板の回路図をF i g 4に示す。

F i g 4



### 3. 2. 10 部品表

本基板の部品表をF i g 5に示す。

品名	型式	ロケーション	数量	単位
CPU	dsPIC30F4013	U1	1	個
スイッチトキャパシタフィルタIC	MAX7403	U3	1	個
オペアンプ	LM358	U2	1	個
抵抗器 2.2k		R7,R8,R9,R10,R11,R15,R16,R17	8	本
抵抗器 4.7k		R12,R18	2	本
抵抗器 10k		R1,R2,R3,R4,R5,R6	6	本
抵抗器 22k		R13	1	本
抵抗器 470k		R14	1	本
セラミックコンデンサ 15pF		C3,C4	2	個
セラミックコンデンサ 120pF		C9	1	個
セラミックコンデンサ 0.01uF		C10	1	個
セラミックコンデンサ 0.1uF		C1,C2,C6,C7,C8	5	個
無極性アルミ電解コンデンサ 10uF		C5	1	個
セラミック振動子 20MHz		X1	1	個
LED (赤)		D1,D2,D3,D4,D5	5	個
タクトスイッチ		SW1,SW2,SW3,SW4,SW5,SW6	6	個
基板用コネクタ 2P		CN1	1	個
基板用コネクタ 10P		CN2	1	個
基板用コネクタ 3P		CN3	1	個
SOP→DIP 変換基板		(U3)	1	枚
ユニバーサルプリント配線板		(PCB)	1	枚
その他配線材料等				

F i g 5

### 3. 3. ソフトウェア

本基板のソフトウェアはCQ出版社のトランジスタ技術のサポートページより2006年10月号連載「クローズアップ・ワンチップ・マイコン〈第6回〉関連データなど」のページからダウンロードしたものを使用した。

## 4. 研究の結果

### 4. 1 動作確認手順

動作確認は下記の手順で行った。

- 1) マイクロフォンで「あ」を長めに発音しながらSW1を押下して録音する。
- 2) 同様に以下「い」「う」「え」「お」を其々別のスイッチを押下しながら録音する。

「い」 => SW2

「う」 => SW3

「え」 => SW4

「お」 => SW5

- 3) 2)が全部終わった状態で「あ」「い」「う」「え」「お」のいずれかを任意に発音する。このとき、LED D1~D5のうち下記に示すものが点灯することを確認する。

「あ」 => D1

「い」 => D2

「う」 => D3

「え」 => D4

「お」 => D5

### 4. 2 動作確認の結果

一定の条件を満たせば概ね認識できる事を確認した。一定条件とは同一話者が録音時と同じ声の高さで発音する事を意識するという事である。

また、その条件を満たしても「あ」と「え」をそれぞれ誤認識しやすい傾向がある。これは、発音するときの口の形、舌の位置に大きな差異が無いからと考えられる。

## 5. 今後の展望

以上の研究において、CPU内蔵メモリのみを使用したシステムでも自己相関法によって母音のみの特定話者音声認識装置を実現できる可能性を確認できた。

本研究の目的は特定話者音声認識の本人認証への応用であるが、誤認識を低減するためには複数の認証方法を併用する事を検討している。

また、本機をネットワークへ接続することによりより高度な認識、応用が可能となる。

例と挙げればビルの出入り口すべてにおいて認証を行う場合などでは、全ての出入り口の認証装置で被認証者のテンプレートデータを蓄積しておくことは現実的ではない。また、昨今防犯カメラ等のセキュリティー機器の多くが LAN に接続されて監視と動画の蓄積まで行われている現状を考えれば音声認識装置にも LAN のインターフェイスを接続する事になると考えられる。その場合、ノードとしての音声認識装置の役割は音響データの抽出に留まり、認証はホスト側で行う形になっていくと考えられる。

また、今回は自己相関法による簡便な方法のみであり、単語の認識、文章の認識を実現するには統計的な音声認識手法を使用する必要がある、その場合もホスト側で処理するシステムを提案することが現実的であると考えられる。

## 謝辞

本研究を行うにあたりご指導を頂きました信州大学工学部情報工学科カワモト ポーリン ナオミ准教授に謹んで感謝申し上げます。また、大学院での就学の機会を与えていただいた信州大学の関係各位に心より感謝申し上げます。

## 参考文献

- [1] 鹿野 清宏、伊勢 史郎、中村 哲 “音声・音情報のデジタル信号処理”  
昭晃堂 (1997/11)
- [2] 山口晶大 “汎用マイコンのように使える DSP 「dsP I C」誕生”  
トランジスタ技術 2006 年 10 月号
- [3] 青木直史 “デジタル・サウンド処理入門” CQ 出版社 (2006/4)