

## 人間親和性を持つ音響反応装置の研究 II

### Development of Audio-response Machine having Human-friendliness II

松原 邦彦

Kunihiko Matsubara

#### 1. はじめに

本研究は人間が聴覚において持つ感応や感性と音響パラメータとの関係を調べ、人間親和性を持つ音響反応装置を実現しようとするものである。人間の発する警戒的な叫び声や、機械の故障時に発生する摩擦音、破壊音など非常事態を反映する音響情報は従来研究されてきた言語的な情報処理によっては認識できず、人間の聴覚的感応または感性に結びついた神経の機構を模擬することによって認識されよう。ここではこのような認識をすることができる音響反応装置の実現を目指している。人間親和性とは「より人間に近い反応」を意味しており、音響による判断において存在するあいまいな状態に対してもなんらかの反応を示し、判断が確定しなければそれなりに情報をあたえる装置を目標とする。本報告は前回の同名の報告 [1] に続くものである。

本研究では聴覚における「感応」及び「感性」という言葉を次のように使い分けている。

聴覚における感応：聴覚による刺激から誘発される生体的防衛的の反応で、高等動物に共通して見られる反応。(闘達、警戒、消沈、恐れ、落ち着き、いらだち)

聴覚における感性：聴覚による刺激から誘発される情緒的感情を意味し、記憶や理性と相まって発生する人間に固有の精神機能と考えられる反応。(喜び、悲しみ、怒り、いたわり、勇気、活

気)

上記のうち「感性」は一般に音響パラメータと一意的に結合するとは考えられず、個人の特殊経験や全人的性格に依存する部分を多く含むため現段階の技術レベルでは到達が困難なものが多い。これに対して「感応」は音響の持つ特性パラメータと精神的反応が直接的に結びついている。ここでは「感応」を主体に取り扱うこととなる。

本テーマの研究は次の4ステップを経て完成にいたる。

ステップ1：音響分析システムの開発整備

ステップ2：聴覚から受ける感応や感性と音響パラメータとの関係の研究

ステップ3：ファジイ理論を主体とした音響データの感応/感性の推論機構の研究

ステップ4：機械安全、特に作業ロボットにおける音響による危険状態察知への応用研究

ステップ1および2の研究はすでに報告 [1] で取り扱った。そこでは各種音響データの採取とスペクトル分析を行ない、警戒感を生起する音響信号を主体に、これら聴覚上の感応/感性を支配するパラメータを分析、抽出した。警戒感を生起する音響として次の3種を中心にデータ分析を行ってきた。

- (1) 破壊/打撃音
- (2) 自動車の警笛
- (3) 人の叫び声

これらの結果、最終的な応用の形態として、1) 警戒音として判別する、2) 人の叫び声を判

別する、の2点にしぼって if...then 形式の規則を見いだした。

今回の報告はステップ3に関するものであり、上記の if...then 形式の規則にもとづく推論の機構の開発を取り扱う。

ステップ4に関しては現在取り組んでいる課題であり、次回以降に報告の予定である。

## 2. 研究の内容

### 2.1 ランニングースペクトルの特徴抽出とパターン化

前回報告の警戒感を生起する音響についての分析結果をここに要約しておく。以後に述べるファジィ推論等の機構の研究はこれらをベースにして展開される。

ランニングースペクトルは図1に示すように、周波数軸と時間軸を持つ2次元空間上に表わされる。

ここにはゼロレベルから最大値までを8階調に分割した濃淡で示してある。1) 警戒音として判別する、2) 人の叫び声を判別する、の2観点から規則を見いだした。人の叫び声には言語的意味を持たないもの(アーッ、キャーッなど)と言語的意味を持つもの(ストップ!トマレ!アブナイ!など)とがあるが、ここでは言語的パラメータ抽出等の処理をすることなしに、発声による聴覚上の感応に関する性質を分析した。ファジィ推論を主体にするという前提に立って以下のように if...then 形式の規則でそれらを書き表しておく。

警戒音に共通のパターン上の特徴

#### (1) 周波数パターン上の規則

(1)ーa) 「連続的なスペクトル成分または非常に密に並んだ線スペクトル成分が広い範囲に渡って多いならば、その音響は警戒感が強い」

(1)ーb) 「線スペクトルの数が少ないならば、その音響は音曲的である」

(1)ーc) 「もし音響スペクトルピーク群が比較的少数の離散的傾向を示すならば、その音響は人の声である可能性が高い」

(1)ーd) 「大きなピークが100—1000Hz程度の周波数帯に離散的に現われるならば、その音響は人の声である可能性が高い」

なおスペクトルパターン上では人の声と楽器音とは非常に似ており、一般に区別するのが難しい。人の叫び声に特有な性質は主としてフレーム時間パターン上に現われる。

#### (2) フレーム時間軸パターン上の規則

(2)ーa) 「立ち上がりカーブが急激ならば、その音響は警戒感が強い」

(2)ーb) 「立ち上がり後の減衰カーブが急激ならば、警戒感が強い」

(2)ーc) 「音響の持続時間が約100msecより少ないならば、警戒感が弱い」

(2)ーd) 「終局時にバックグラウンドレベルに戻る時間が非常に短いならば、人の叫び声である可能性が高い」

警戒音に属する時間軸上のパターンとして図2に3種を掲げる。図3はその時間微分曲線で、パターンの判別の上で重要な役割をはたす。

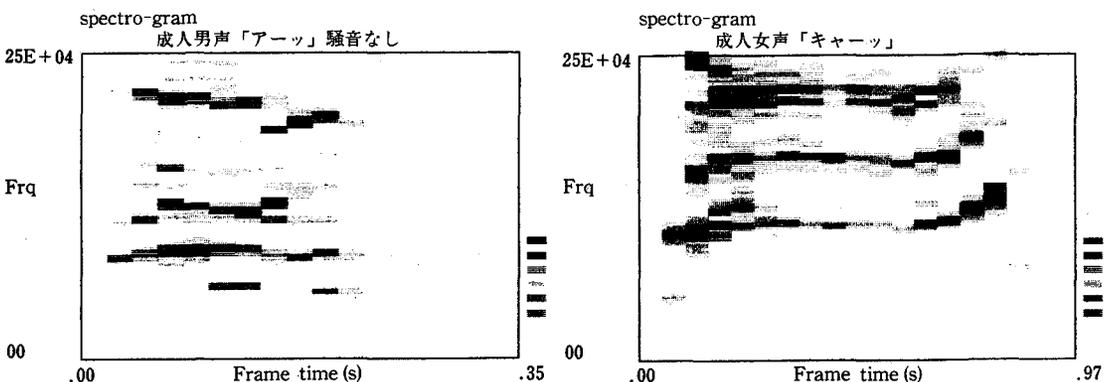


図1 ランニングースペクトルの例

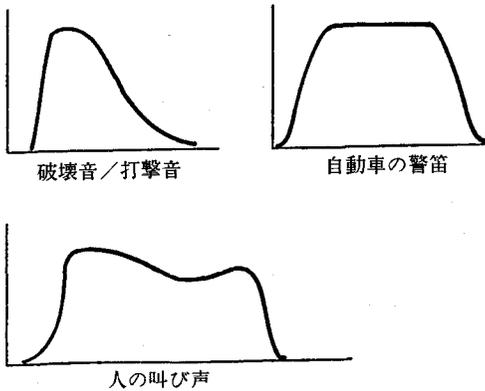


図2 警戒音のフレーム時間パターン

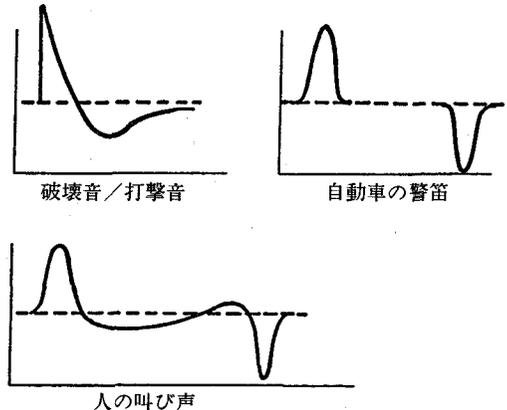


図3 警戒音のフレーム時間パターンの微分曲線

## 2.2 ファジィ推論によるパターンの判別

ファジィ集合とは1つの変量のとる値が特定グループに属するか、その帰属度を0から1の間の数値で示したものである。特定の部分集合に属するか属さないかあいまいさがある場合、その部分集合に属する度合を0から1の間の重みで表わす。メンバーシップ関数とは、確実に所属しないと切り切れる部分には0を、確実に所属すると言い切れる部分には1を当て、あいまいさを残す部分にはその程度に応じて0-1の中間の値を当てたものをいう。

ファジィ関係とは、2つ以上の多変量がつくる空間上に明確な境界をもたせないで部分集合を構成したものである。特定の部分集合に属するか属さないかあいまいさがある場合、その部分集合に属する度合を0から1の間の重みで表わす。

ファジィ推論規則はファジィ集合を用いて一般には次のような if...then 形式で与える。

前提1 If x is A then y is B (1)

前提2 x is A' (2)

推論 y is B' (3)

ここでxはこの推定系への入力となる信号で入力対象と呼ばれる。y同様に出力対象と呼ばれる。A、A'、B、B'はファジィ集合である。「'」のついたものは「'」のつかない関数の修飾系を示す。ここで修飾系とは「やや」、または「非常に」といった言語に対応したもので、一定の強調を意味する。

ファジィ集合は、単純な物理法則に乗らない複

雑なメカニズムによる事象を取扱うとき有効であり、技術的な多数の経験の結果を集約して効果的に曲線群に載せることができる。ファジィ集合の導入により推定を行なう方法には様々な手法が研究されており<sup>4)5)</sup>、これらを使って様々な応用形が考えられる。この研究では、大枠では既に知られた方法に基づくが、過去の筆者の経験<sup>6)</sup>にもとずき音響反応のために独自に工夫した推論法を用いる。

## 2.3 周波数軸上のパターンの判別

ここで行なう判別に使用するファジィ規則は次の2つに絞って適用する。

(1) a) 「連続的なスペクトル成分または非常に密に並んだ線スペクトル成分が広い範囲に渡って多いならば、その音響は警戒感が強い」

(1) c) 「もし音響スペクトルピーク群が比較的少数の離散的傾向を示すならば、その音響は人の声である可能性が高い」

ここで次の定義をしておく。

「警戒音のパターン」: 連続的なスペクトル成分または非常に密に並んだ線スペクトル成分を広い範囲に渡って多く持つパターンと規定する。

この定義を用いれば(1) a) は次の形式に書くことができる。

(1) a') 「もし音響スペクトルパターンが警戒

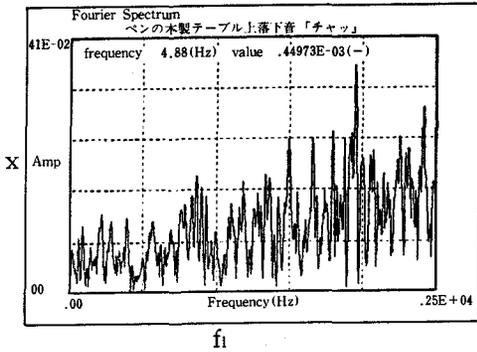


図4 周波数上のスペクトル入力

音のパターンに近いなら、警戒感を持つ」

これを式(1)、(2)、(3)の if...then 形式のファジイ推論に当てはめよう。

式(1)、(2)における入力対象  $x$  は音響スペクトルの特定周波数上の成分の大きさであり、周波数上の最大成分の大きさによって正規化しておく。その1例を図4に示す。式(1)における  $A$  は入力のスเปクトル上の成分  $x$  が上に定義した「警戒音のパターン」の範囲に入る度合(帰属度)を表わすメンバーシップ関数で、その1例を図5に示す。

式(1)、(3)における  $y$  は警戒感に寄与する度合とし、 $B$  はそれを決めるメンバーシップ関数とする。

音響のパターン認識に使うにはこのままではいくつかの不都合が生じる。第1にその入力値は一つの数値ではなく、離散的な多数の数値の集団、いわゆるパターンとして入力される。第2に上の形式ではメンバーシップ関数を4箇所 ( $A, A', B, B'$ ) も使用しているが、パターン上の一つの特徴を推論するのに同じレベルの曖昧さを2箇所以上に設定するのは複雑すぎる。

第1の問題については次の形式に拡張する事で解決する。周波数軸のメッシュ数だけ上記の if ... then 形式を与えてこれをベクトルの形式で記述する。

$$X = \{x_1, x_2, \dots, x_n\}, Y = \{y_1, y_2, \dots, y_n\} \quad (4), (5)$$

$$M = \{A_1, A_2, \dots, A_n\}, N = \{B_1, B_2, \dots, B_n\} \quad (6), (7)$$

第2の問題については次のように解決する。こ

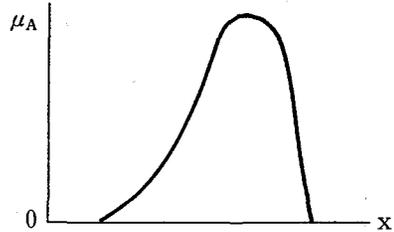


図5 パターン帰属度を表わすメンバーシップ関数

ここで修飾系を考えないことにし、さらに規定のパターンへの帰属度をメンバーシップ関数  $M$  に含めてしまい if ... then の後件部をファジイ集合にせず、通常関数関係  $y = wB(x)$  したがって  $Y = WN(X)$  で置き換えて簡略化する。また実際の計測値として得られる入力は  $X = X^\circ = \{x^\circ_1, x^\circ_2, \dots, x^\circ_n\}$  であり、入力に対する曖昧さを導入する必要が無い。従って次のように変形できる。

前提1 If  $X$  is  $M$  then  $Y = WN(X)$  (8)

前提2  $X = X^\circ$  (9)

推論  $Y = WN(X^\circ)$  (10)

この計算は次のように行なう。

- 1) 入力  $X = X^\circ$  のそれぞれのエレメントに対して if ... then の前件部の条件にどの程度適合するかを求める。ここで  $W = \{w_1, w_2, \dots, w_n\}$  として

$$W = M(X^\circ) \quad (11)$$

- 2) if ... then の後件部によりパターン帰属度を求める。

$$y_1 = w_1 B_1(x^\circ) \quad (12)$$

$y_1$  は各周波数メッシュ点上のパターン帰属度を表わす。ただしここでは  $B_1 = 1$  とする。

- 3) これらは  $n$  個の数値からなるので全体としての評価を  $z$  で表わして

$$z = \sum_{i=1}^n y_i / n \quad (13)$$

とする。以上の推論過程をブロック図で表わすと図6のようなになる。

実際に計測した多数のデータを集約して組み立

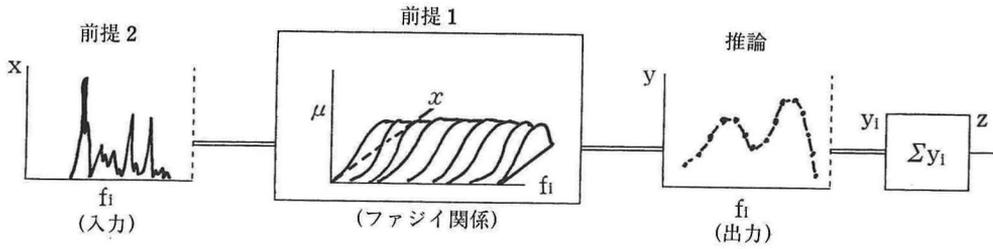


図6 ファジイ推論のアルゴリズムをブロック図で示す

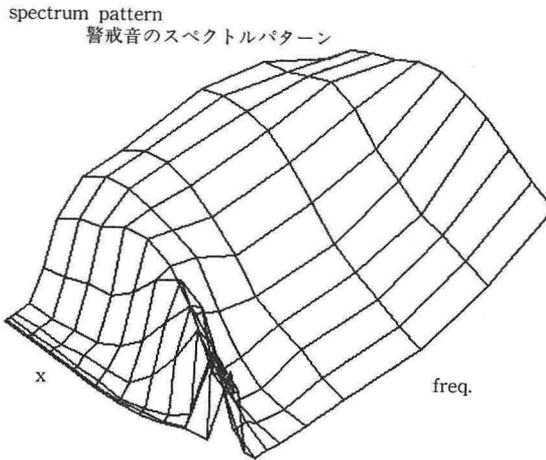


図7 パターン帰属ファジイ関係 (3次元鳥瞰図表示)

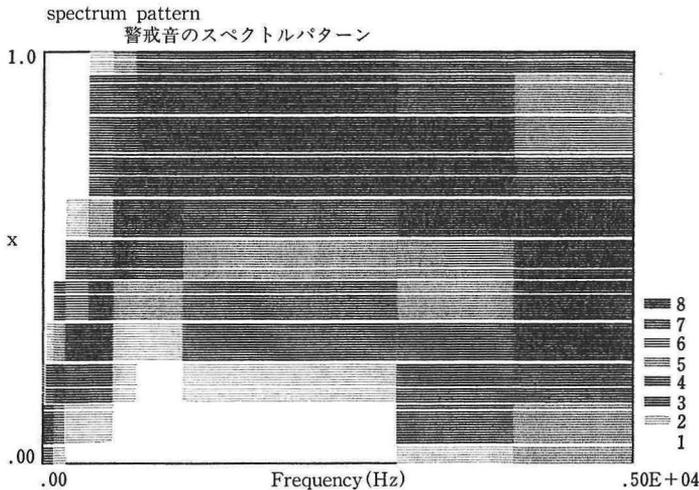


図8 パターン帰属ファジイ関係 (8階調表示)

てたファジイ関係を図7、図8に示す。図7には3次元鳥瞰図として示し、図8には8階調表示で示した。さらにこれらを組み立てたときのパターン帰属度をメンバーシップ関数として図9 a—9

cに示す。カバーする周波数帯を12のメッシュに分け、それぞれの周波数メッシュについて示してある。

これらは様々なデータの分析から得た結果を1

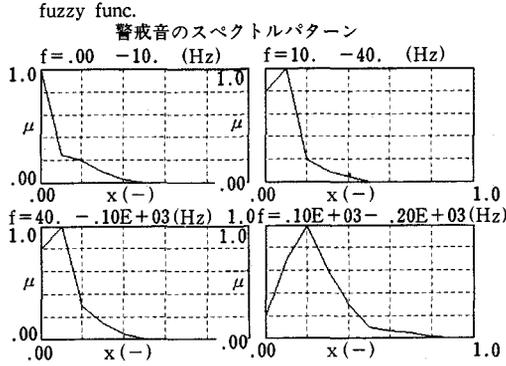


図9 a パターン帰属メンバーシップ関数 (周波数 0—200Hz)

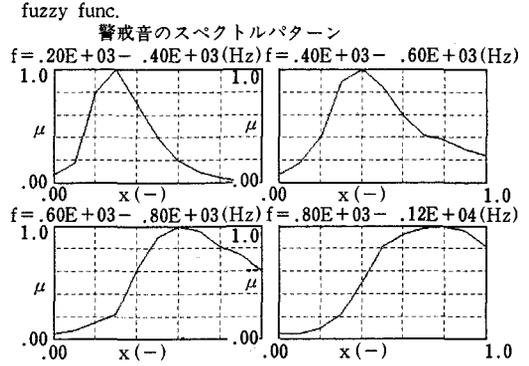


図9 b パターン帰属メンバーシップ関数 (周波数 200—1200Hz)

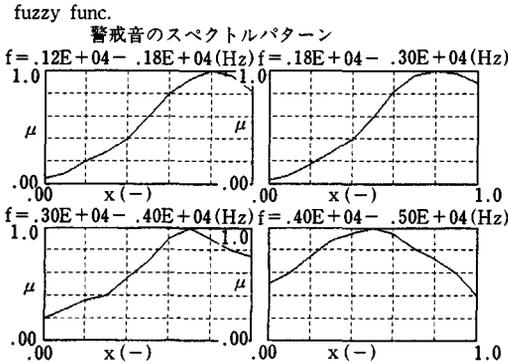


図9 c パターン帰属メンバーシップ関数 (周波数 1200—5000Hz)

つの経験則に仕立てたものであるが、ファジイ関係のグラフ図7のピーク値を辿ると、結果的に人間の最小可聴範囲を示す曲線<sup>2)</sup>にきわめて類似した曲線となる。

次に人の声に特徴的なパターンの判別には規則(1)一c)を使用する。ここでスペクトルが離散的か団塊のかを定量化しておく必要があり、次の値を定義しておく。

「ピーク間周波数差」：次の処理を経て得た数値をもって定義する。

- $y_1$  が有意のもの (パターンへの帰属度が0.20以上のもの) を捨てる。(12)式で得た結果を使い、これを  $y^*_1$  とする。
- 捨てられた相隣るピーク間の周波数メッシュ番号の差を調べる。上記で得た  $y^*_1$  の属する周波数メッシュ番号を  $i^*$  として相隣る番号との差  $di^*$  を求め、さらにそれからピーク間周波数差  $dF_1 = df \times di^*$  を求める。

次に例示をもって具体的に説明する。

例1) 団塊的な場合： $y_1$  が有意のものが次のように得られたとしよう。

$$i^* = 5, 6, 7, 8, 9, 13, 25, 26, 27, 28$$

$$di^* = 1, 1, 1, 1, 4, 12, 1, 1, 1$$

$$\text{周波数差は } dF_1 = df \times di^*, (i=1, \dots, 9)$$

例2) 離散的な場合：適当な間隔でバラついているとしよう。

$$i^* = 7, 12, 16, 17, 21, 26, 30, 35, 37, 38$$

$$di^* = 5, 4, 1, 4, 5, 4, 5, 2, 1$$

$$\text{周波数差は } dF_1 = df \times di^*, (i=1, \dots, 9)$$

$m$  個の  $y^*$  があるとすれば  $di^*$  したがって  $dF_1$  が  $m-1$  個得られる。

人間の声に特徴的な  $dF_1$  の帯域は経験的に男性成人では 100—250、女性、子供では 200—400 Hz であることが解析されている<sup>2), 3), 6), 7)</sup>。両者をカバーする範囲は 100—400 Hz となる。いくつかの叫び声の分析結果によれば上記の帯域より若干高い方へ移行する傾向がみられる。そこで叫びを主体に検出する目的に沿って  $dF_1$  の帯域を 250—500 Hz とする。また人の声に特徴的な周波数帯 100—1.8 kHz とすれば十分である<sup>2), 3), 6)</sup>。そこで人の叫び声に特徴的な離散スペクトルパターンを次のように規定しておく。

「人の叫び声の離散スペクトルパターン」：複数のピーク間周波数差  $dF_1$  が 250—500 Hz の帯域にあるスペクトルパターン

人の声のスペクトルはその基本周波数の倍数毎に離散的に大きな成分を示し、 $dF_1$  は実はこの基

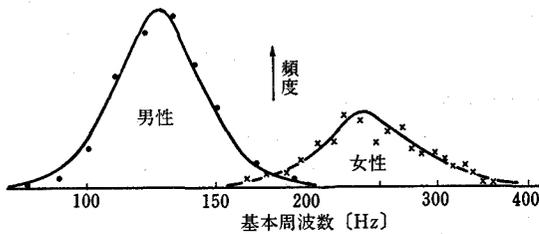


図10 発声音による基本周波数の分布

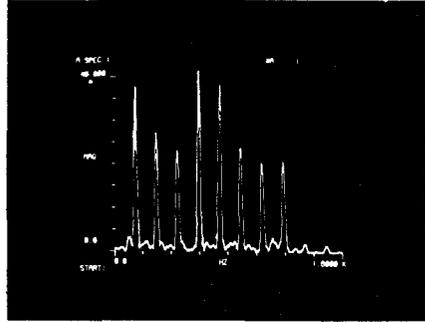


図11 人の声のスペクトルパターン  
最も離散的な例(男声 アー)

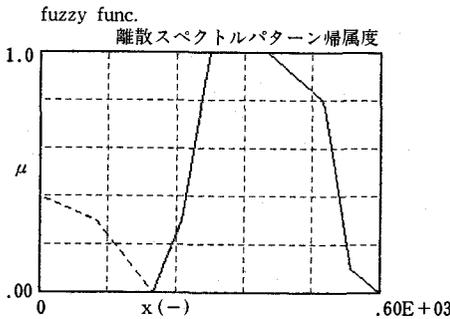


図12 人の叫び声の離散スペクトルパターン  
への帰属度

本周波数に等しい。人の声の基本周波数の分布はよく調べられており、図10に示す周波数分布をしている<sup>7)</sup>。図11には人の声の離散スペクトルの典型的な例を示す。ただしこれは通常会話音声中のものである。一般に叫び声は緊張した雰囲気で作られるのでこれら通常会話中より高い周波数へシフトする。女性・子供の場合に著しい。筆者の分析結果を加味して高音側へ偏らせたものが図12の実線である。

上記の様に定量化した上で if ... then ルールを書くと(1)-c)は

(1)-c) 「もし相隣るピーク間周波数差が人の叫び声の離散スペクトルパターンに属するならば、それは人の叫び声である」  
となる。これを式(1)、(2)、(3)の if ... then 形式のファジイ推論に当てはめよう。入力 x は上に定義したピーク間周波数差  $dF_1$  である。A は上に規定した「人の叫び声の離散スペクトルパターン」

の範囲に入る度合(帰属度)を表わすメンバーシップ関数とする。Y は人の叫び声に帰属する度合である。

以前行なったように修飾系を考えないことにし、さらに曖昧さを持つパターンへの帰属度をメンバーシップ関数 A に含めてしまい、if ... then の後件部にファジイ集合を使用せず、通常関数関係で置き換える。さらに入力に対する曖昧さを導入する必要が無いとして次のように書き下す。ここで同時に多数の  $dF_1$  が与えられ、その分布が上記のパターンに属するかを判定しなければならないためこれだけでは不足である。 $dF_1$  が均一に分布するときパターンに帰属しないが図11の実線のメンバーシップ関数だけではこれを除外できない。そこで上記のメンバーシップからはずれた値が存在するとき失格の度合を与え、帰属の度合から減ずることを考える。これをもう一つの if ... then ルール、前提 2 として以下のように与える。

$$\text{前提 1 } \text{If } dF_1 \text{ is } o \text{ then } e_1 = w_1 p(dF_1) \quad (14)$$

$$\text{前提 2 } \text{If } dF_1 \text{ is } q \text{ then } e_1' = u_1 r(dF_1)$$

$$\text{前提 3 } dF_1 = dF_1^\circ \quad (15)$$

$$\text{推論 } e_1 = w_1 p(dF_1^\circ) \ominus u_1 r(dF_1^\circ) \quad (16)$$

ここで o は各  $dF_1$  に対して離散とみなすのが適当かどうかを表わすメンバーシップ関数(図12の実線)、q はパターンからはずれたときに与える失格の度合を表わすメンバーシップ関数(図12の点線)、 $e_1$  は離散スペクトルのグループへの帰属度、 $w_1$ 、 $u_1$  はそれぞれ関数 o、q から読み取られた値である。P、r は一般には  $dF_1$  の関数であるがここではいずれも定数を用いる。ここで  $a \ominus$

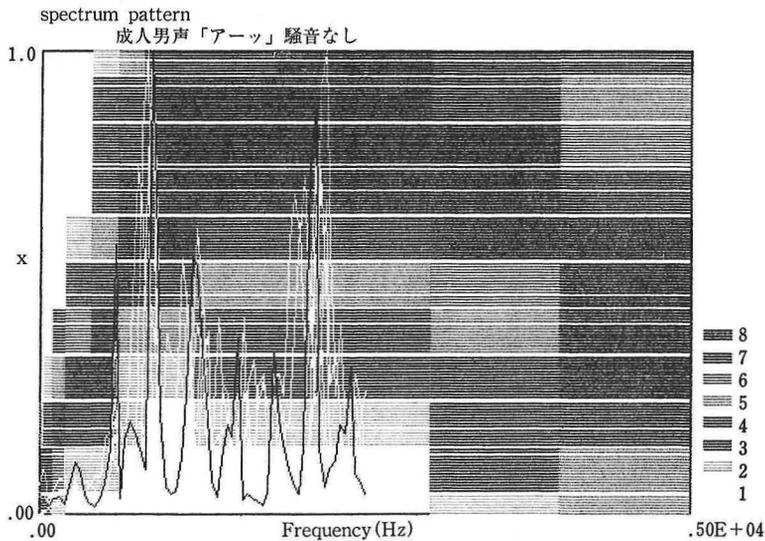


図13 警戒スペクトルパターン判定の例

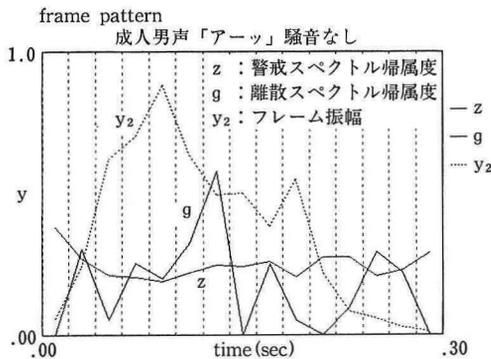


図14 z (13式)とg (21式)および振幅(y2)のフレーム時間上的変化

$b = (a-b) \vee 0$  で定義され、「限界差」と呼ばれる演算を意味する。計算は次のように行なう。

1) 入力  $dF_i = dF_i^\circ$ , ( $i=1, \dots, m-1$ ) に対し前件部の条件に適合する度合を求める。

$$w_i = o(dF_i^\circ) \tag{17}$$

2) if ... then の後件部によりパターン帰属度を求める。

$$e_i = w_i p_i(dF_i^\circ) \tag{18}$$

ただし  $p_i(dF_i^\circ) = 1$  とする。

3) 入力  $dF_i = dF_i^\circ$ , ( $i=1, \dots, m-1$ ) に対し前件部の条件に適合する度合を求める。

$$u_i = q(dF_i^\circ) \tag{19}$$

4) if ... then の後件部によりパターン帰属失格度を求める。

$$e_i' = u_i r_i(dF_i^\circ) \tag{20}$$

ただし  $r_i(dF_i^\circ) = 1$  とする。

5) 離散パターンへの帰属度を表わす  $e_i$  は  $m-1$  個の数値からなるので、全体としての評価を  $g$  として

$$g = \frac{\sum_{i=1}^{m-1} (w_i \ominus u_i)}{\sigma} \tag{21}$$

ここに  $\sigma$  は最も離散的とされるパターンにおける  $\sum e_i$  である。 $\sigma$  は次のようにして求める。まず図12の離散スペクトルパターンへの帰属度が最大値  $e_{max}$  (通常は1) となる周波数差  $dF_{imax}$  を求める。図12のように  $dF_{imax}$  がある範囲に渡っているときはその最大値または中間値をとる。この  $dF_{imax}$  で分析対象となっている周波数帯域を分割したときの個数を  $n_{max}$  として

$$\sigma = \sum_{i=1}^{n_{max}} e_{max} = e_{max} \cdot n_{max} \tag{22}$$

と求める。上記のアルゴリズムによる FORT-RAN プログラムを付録Aに掲げておく。

これらの手法により分析した例を図13に掲げる。スペクトルパターンに入力スペクトルを重ねていき、(11)、(12)、(13式)により各フレーム毎に  $z$  を計算する。図中白線は第1フレームから第4フレームまでの判定時のスペクトル、黒線は第5フレームの判定中のスペクトルである。図14は警戒スペクトルパターンへの帰属度  $z$  と人の声に特徴的

な離散スペクトルへの帰属度  $g$  の全フレームにわたる変化を示す。  $y_2$  はフレーム内のスペクトル振幅最大値で後述の時間フレーム上のパターン判定に使われる。

### 2.4 フレーム時間軸上のパターン判別

ここでは次の規則を使用して判定を行なう。

- (2) — a) 「立ち上がりカーブが急激ならば、その音響は警戒感が強い」
- (2) — b) 「立ち上がり後の減衰カーブが急激ならば、警戒感が強い」
- (2) — c) 「音響の持続時間が約 100msec より少ないならば、警戒感が弱い」
- (2) — d) 「終局時にバックグラウンドレベルに戻る時間が非常に短いならば、人の叫び声である可能性が高い」

フレーム時間軸上のパターンを表わすパラメータは振幅値である。振幅としてフレーム内の平均振幅を採るべきか最高振幅を採るべきかという問題があるが、生体的反応を考えると最高振幅を採るほうが自然である。

最高振幅の時間経過上のパターンは図2に示すように時間軸上に固定したのではなく、時間軸上で様々な伸縮がある。これらをフェジイ変数によりパターン化する方法は複雑化するのみである。そこで最近の神経回路網の研究の進展からヒントを得て、簡単なネットワークにより上記のパターンの特徴の検知モデルを考案した。ただしここではニューラルネットワーク理論を適用していない。その理由は上の検知モデルでは微分、遅延素子を持ち、アナログ回路のシミュレーションとして組立てたほうがはるかに簡単になるからである。

警戒感がフレーム時間軸上の変化の激しさに関係していることから人間の感覚受容器は音という波動の刺激をなんらかの形で図2に示した時間パターン曲線の微分値として検出していると考えられる。微分値は図3に示したように音響信号が入されるとまずプラス側にピークが現れ、次にマカイナス側にピークが現われる。ある場合には鋭く、ある場合にはなだらかに現われる。人間の脳では最初のピークが引金となり刺激が神経網に広がり、関連した様々な感覚器官を駆動する信号が生起されると想定できる。

聴ニューロンは音の変化分に対してインパルスが生じ、それに続いて抑制の効果が働き、結果として変化の無い音波に対してほとんど応じないという<sup>2)</sup>。脳生理学的な研究で明かにされているこのような大脳皮質聴ニューロンの働きにヒントを得て、時間微分値のピークを検出した後時間が長引けば次第に興奮がさめるように減衰してゆき、警戒感信号が抑制される過程を模擬する。これを（ピーク保持+減衰）の機能とよぼう。この分析に沿って規則(2) — a)、(2) — b) を次のように書き換える。

- (2) — a') 「フレーム毎の最高振幅の時間変化率の正のピーク値に時間的減衰を考慮した値の大きさに応じて警戒感が発生する」
- (2) — b') 「フレーム毎の最高振幅の時間変化率の負のピーク値に時間的減衰を考慮した値の大きさに応じて警戒感が発生する」

これらに対応した信号発生回路を図15に示すように組み立てる。

プラス、マイナスのピーク検出は微分値のプラス、マイナスのピーク値をそれぞれに調べている。ピークの保持と減衰には次式を用いる。

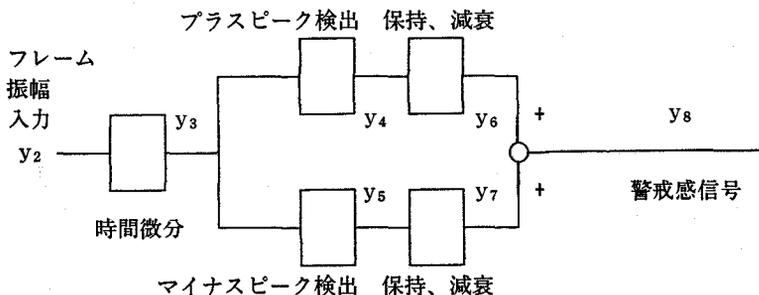


図15 フレーム時間軸上の警戒感信号発生回路

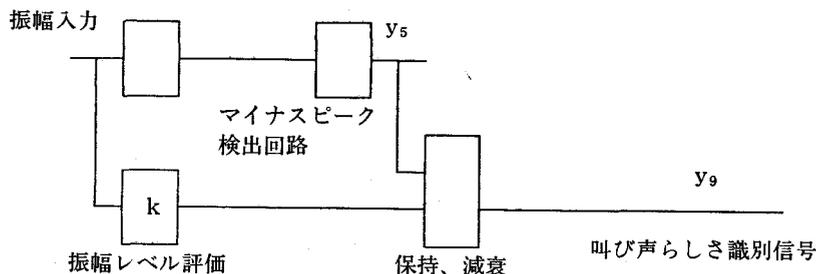


図16 フレーム時間軸上の人の呼び声らしき判別の回路

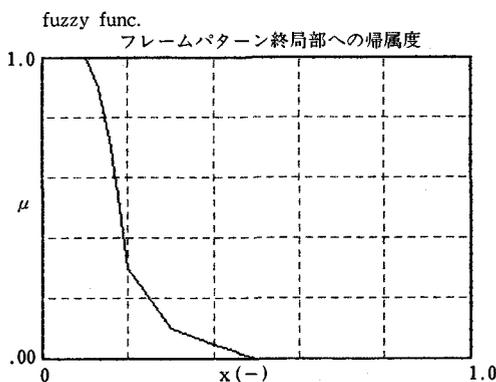


図17 フレーム終局部への帰属度

$$y_i = y_{i-1} + k \int_{t-\Delta t}^t u(t') dt' \exp(-\Delta t/\tau) \quad (23)$$

$$u(t') = x(t'), \quad (y_{i-1} \leq x_{i-1} \cdot \Delta t) \\ = 0 \quad (y_{i-1} > x_{i-1} \cdot \Delta t) \quad (24)$$

ここに  $\Delta t$  はフレーム時間メッシュ巾で、添え字  $i$  はメッシュ番号、 $\tau$  は減衰定数、 $x$  はピーク値の入力、 $k$  はゲインで通常は1である。(23)式は微分値のプラス（またはマイナス）の部分積分してこの回路の入力に戻す。ただしこのとき(24)式により1メッシュ前の微分値  $(x_{i-1}) \times \Delta t$  を下回る場合には積分に算入しない。こうする事によってピーク値に対応した数値が保持される。(23)式の exponential は信号のメッシュ巾内における減衰を計算する。

規則(2)一c) に関しては聴覚の神経経路の伝達特性と関係がある。耳で捉えられた聴覚信号は蝸牛神経、延髄のオリブ核、中脳の下丘、内側膝状体のシナプス伝達を経て大脳皮質側頭葉の聴覚野に至るが、これらの伝達遅れ時間の測定値として約 40msec という値が報告されている<sup>2)</sup>。特別の訓練でもない限り、大脳における判別速度などの個人差も併せて経験的に約 100msec 程度

以下の時間的特性の分解能は急激に落ちると考えられる。言い換えれば

- (2)一c) 「信号伝達特性および神経網の特性によって約 100msec より短い音響信号の時間分解能は低下する」

この規則にもとづき模擬回路を考えるならば、約 100msec の時定数をもつ1次遅れ伝達要素を時間微分の直後に挿入するのが妥当であろう。しかし1フレームの時間巾が 70—100msec にとられるため解析システムの時間分解能が同程度になってしまう。このような事情からフレーム巾が70 msec 程度より小さいときのみ考慮すればよい。

規則(2)一d) に関しては終局部において急激に切れる叫び声の特徴を判定しなければならない。急激に途切れる特徴は(2)一c) と同様の条件によって評価することができる。「声が途切れる直前にある」ことを判定するには負のピーク発生が終局部の低い振幅レベルにあることを把握すればよい。これらの分析に基づき、

- (2)一d) 「フレーム毎の最高振幅の時間変化率の負のピーク値に、その生じた時点の振幅レベルの低さを考慮した値に応じて人の叫び声らしさが発生する」

と書き換える。これに対応した信号発生回路を図16のように組立てる。

振幅レベル評価ではフレーム時間に沿った振幅値が0レベルに近い値であるかを調べ、もしそうならば(23)式の  $k$  を1に近くとり、そうでなければ  $k$  値を0に近くとる。このときの重みをファジイ変数（メンバーシップ関数）として設定する。図17にこのメンバーシップ関数を示す。

図15と図16を組合せてフレーム時間軸上のパターン判別回路とする。これを図18に示す。この回

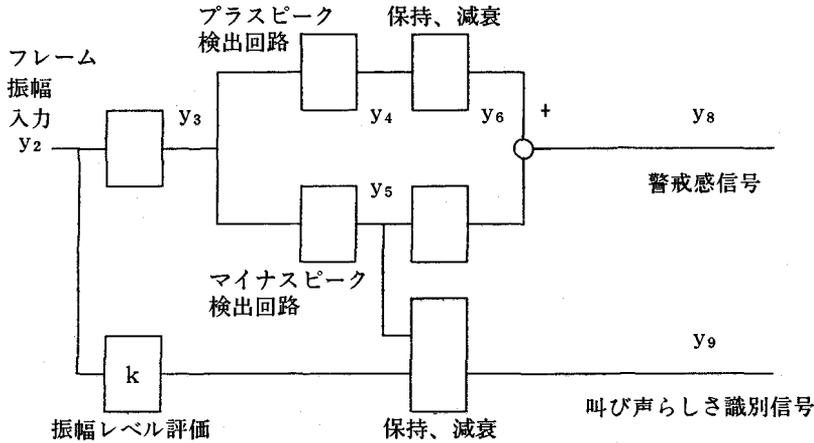


図18 フレーム時間軸上のパターン判別の回路

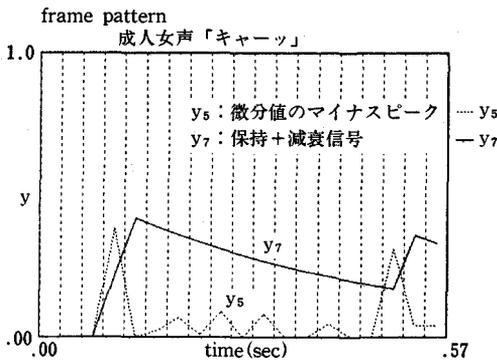


図19 ピークの保持、減衰の計算例

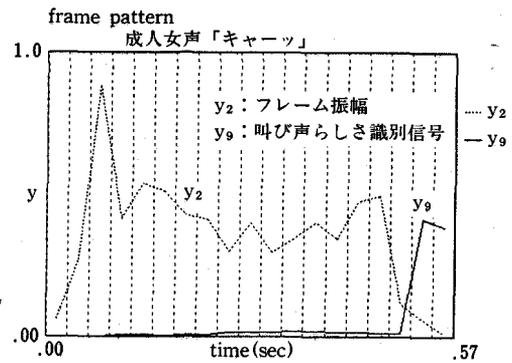


図20 叫び声らしき識別信号 (フレーム時間軸パターン)

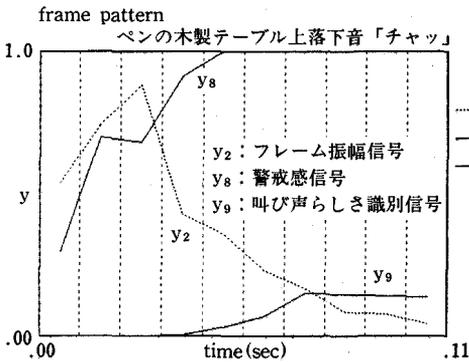


図21 フレーム時間軸上のパターン判別の回路の出力例

路に沿ってフレーム時間軸上のパターン判別を行なう FORTRAN プログラムを付録Bに掲げる。

図19に(2)式によりピーク保持、減衰の計算例を示す。係数は  $k=1$  に固定している。  $y_5$  は図15に示した変数で時間微分後のマイナスピークのみを

取り出したものである。

$y_7$  はその保持と減衰の信号である。

図20は終局部において急激に切れる叫び声の特徴を抽出した信号 (叫び声らしき識別信号) で図16の  $y_9$  の変化を示す。  $y_2$  はフレーム振幅信号で音声のフレーム毎の大きさを示している。

図21には別の音響例についてフレーム時間軸上のパターン判別回路 (図18) から得られる2つの信号、警戒感信号  $y_8$  と叫び声らしき識別信号  $y_9$  を示している。この場合は叫び声らしき信号は低く押さえられているのが見られる。

## 2.5 全パターンの判別

フレーム毎のスペクトルパターン判別とフレーム時間パターン判別を統合して音響パターン判別を行なう。その構成を図22に示す。まずフレーム毎のスペクトルパターンが調べられ、警戒音声

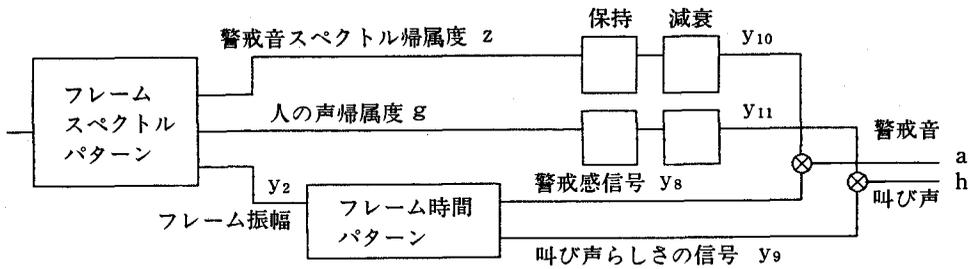


図22 全パターン判別回路

ペクトル帰属度  $z$  と人の声帰属度  $g$  およびフレーム振幅 (最高値)  $y_2$  が得られる。 $z$  および  $g$  はそれぞれフレーム時間毎の瞬時値であるから保持、減衰回路を経て  $y_{10}$ 、 $y_{11}$  として出力される。この回路はいわば神経系短期記憶保持に相当するもので、フレーム時間軸上のパターンと比較演算する必要性からつけ加えた。計算は(3)式である。

$y_2$  はフレームパターン判別回路に入り、警戒感信号  $y_8$  と人の叫び帰属度  $y_9$  を出力する。 $y_{10}$  と  $y_8$  のファジイ論理積 (どちらか小さい方をとる

$\min$  演算) をとって最終的な警戒感を表わす信号  $y_a$  とし、 $y_{11}$  と  $y_9$  とのファジイ論理積をとって人の叫び声らしさを表わす信号  $y_h$  とする。

上記の構成により各種のサンプリングデータについてテストを行い、評価を行なった。図23～図25には人の声と無関係な破壊音および警笛の例を示す。図23はガラス等の破壊音の例で警戒感信号が非常に高く判定される。一方、人の叫び声らしさは非常に低く判定される。図24、25は自動車の警笛の例で、図24は一回やや長く鳴らした場合、

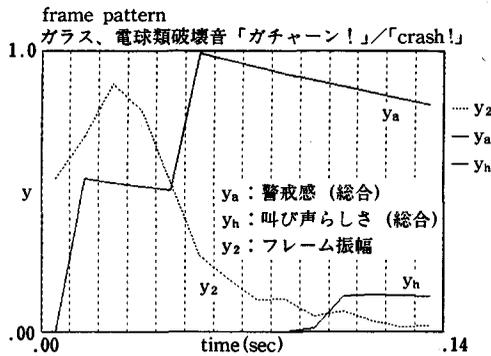


図23 破壊音における判定結果

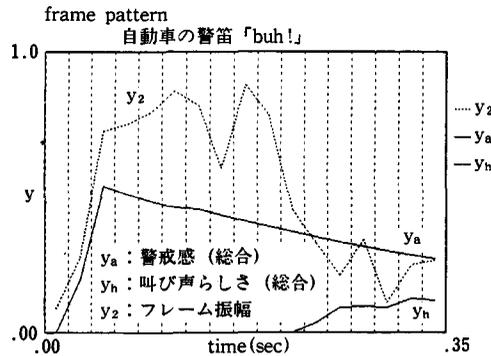


図24 警笛 (一回長く鳴らした場合)

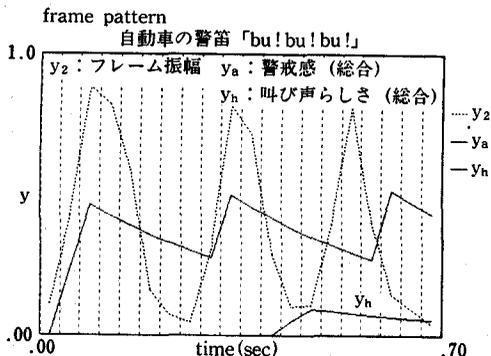


図25 警笛 (3回短く鳴らした場合)

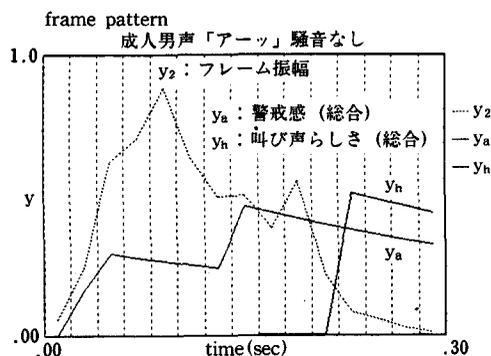


図26 男性の短い叫び声の判定結果

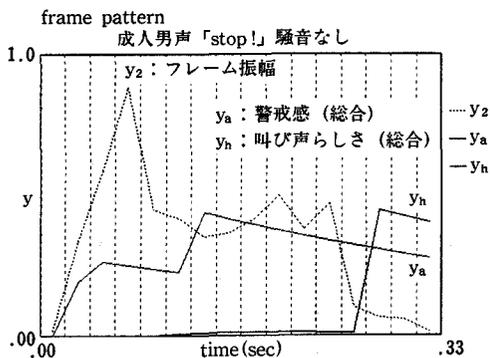


図27 男声「stop!」の判定結果

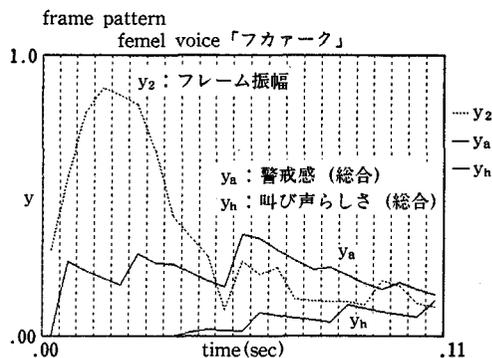


図28 女性ナレーション「ふか〜く」の場合

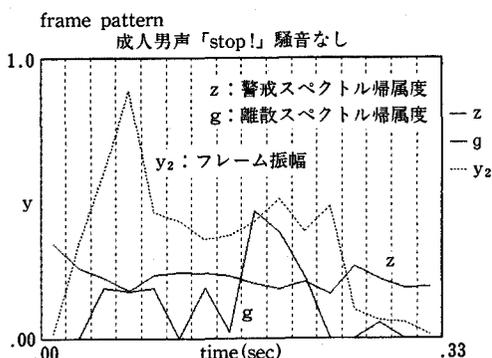


図29 男声「stop!」の判定の詳細 (スペクトル判定)

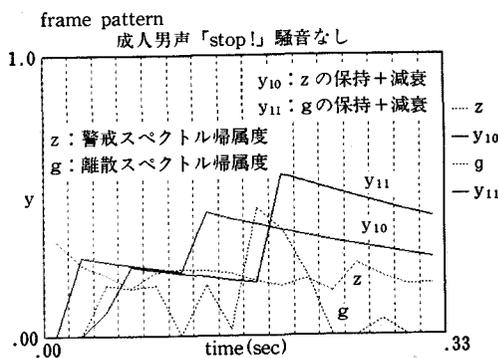


図30 男声「stop!」の判定の詳細 (z、gの記憶保持、減衰記号)

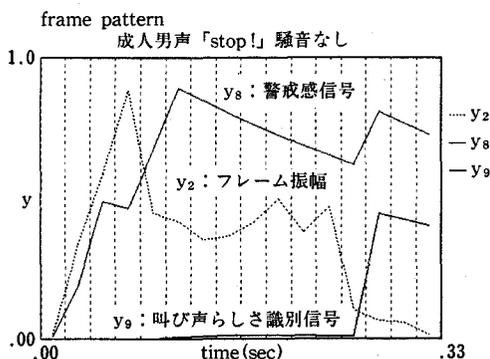


図31 男声「stop!」の判定の詳細 (フレーム時間パターン)

図25は短く3回鳴らした場合である。後者の場合は前者の場合より警戒感が強く判定される。これは音を激しく断続する事によりフレーム時間パターン判別での微分値のピークが効いて来ることによる。我々は断続的に警笛を鳴らすことにより緊急度の高い警告を行なうことがあり、この結果は我々の経験とよく一致する。

図26～図28は人の声に関連した音響の場合をで

ある。図26は男性の叫び声の例で、警戒感信号  $y_a$  は中位程度に、叫び声らしさを表わす信号  $y_h$  は高く検出される。図27は男声で「stop!」と叫んだ時の分析結果である。この場合も叫び声らしさは高く判定されるが警戒感はやや低い。図28には警戒感を持たない女声によるナレーションの1部分で、「ふか〜く」(深く)とゆっくり発声した部分の分析結果を示した。警戒感、叫び声らしさともに低く押さえられた。離散スペクトルへの帰属度は高く評価されるもののフレーム時間パターン上で急に途切れる叫び特有のパターン帰属度  $y_9$  が低いために「叫び」としては判定されない。

上記の分析例のうち図27に関して中途経過を詳細に示したものが図29～図31である。図29はスペクトルパターン上の判定で、警戒音パターン帰属度  $z$  と人の声に特徴的な離散スペクトル帰属度  $g$  の瞬時値変化を示す。図30はそれらの保持、減衰信号を、図31にはフレーム時間軸上のパターン判定を示す。フレーム時間軸上の警戒感信号  $y_8$  は非常に高いものの、スペクトルパターン上の警戒

信号が低い場合ファジイ論理積により、低い方が採用される。

## 2.6 まとめ

本研究のステップ3の目的は、音響データの感応/感性の推論機構の開発である。本研究では特に警戒音の判別、人の叫び声の判別に焦点をあてた。判別はスペクトル上のパターンとフレーム時間上のパターンで行ない、両者を組み合わせて行なった。

スペクトルパターンでは(1)警戒音パターン、(2)離散スペクトルパターンの2つを考え、それぞれ

のパターンへの帰属度をファジイ集合を使って計算した。フレーム時間パターンでは(1)振幅の変化率、(2)急激に切れる終局部のパターンを判定する方法を使用した。

破壊音、警笛、男性の叫び声、男声による叫びに近い短い言葉、女声のナレーションの一部などによりテストを行なった。結果は警戒音の感応及び人の叫び声の認識に関して概略我々の経験とよい一致を示した。しかし本推論機構の有効性について統計的に厳密な評価はまだこれからである。

演算時間は MPU80386+数値演算プロセッサ 80387のシステムで次のデータが得られた。

時系列データ長さ	$\Delta t$	フレーム分割数	FFT計算	判定演算
0.57sec	0.2msec	19	8sec	2.5sec
0.33sec	0.2msec	16	4sec	1.5sec
0.14sec	0.1msec	14	3sec	1.5sec

今後の課題として次の項目が挙げられる。

- (1) 人の叫び声あるいは危険を知らせる短い発声にはその時の緊迫感により発声パターンにかなりの差がある。このシステムではその程度に応じた応答を得ることが要求される。この要求を満たしているかどうかについて様々な状況設定下におけるテストと厳密な統計的評価が必要である。
- (2) 雑音レベルの高い環境におけるデータサンプリングとテストが必要である。また有効な前処理法を研究する必要がある。

これらはステップ4：機会安全、特に作業ロボットにおける危険状態察知の応用研究において行なわれる予定である。

(まつばら くにひこ 教授)

(1994. 8. 24 受理)

## 謝 辞

本研究は(財)長野県テクノハイランド開発機構

および(財)浅間テクノポリス開発機構から研究助成を受けて行なったものです。関係各位に深く感謝いたします。

## 参考文献

- 1) 松原邦彦、人間親和性を持つ音響反応装置の研究 I、長野大学紀要投稿中 (1994)
- 2) 電子通信学会編、聴視と音声、p.28、電子通信学会発行 (昭和45年)
- 3) 松原邦彦 (発明者)「音曲反応の方法」特許第1771641号 (分類G10L 5/00) 特許出願公告第050600号 (平成4年)
- 4) 寺野寿郎、浅居喜代治、菅野道夫共編「ファジイシステム入門」オーム社、1988) p.45
- 5) 水本雅晴「ファジイ理論とその応用」サイエンス社 (1989) p.307
- 6) 松原邦彦「音響スペクトルに関するデータ資料」(個人資料) (1974—1985)
- 7) 古井貞熙「デジタル音声処理」東海大学出版会 (1985) p.17

## 付録A スペクトルパターン判別推論のプログラム

スペクトル離散度を判定する2つのメンバーシップ関数は意味を持つ領域が完全に分かれているため、これを1つの関数(データ配列  $xmu, ymu$ ) に収め、パターン帰属失格の度合  $u_1$  をマイナスの値

として読み取り、加え合わせるという方法を採用、プログラムを簡単にしている。

```

c .....
      subroutine ptmach (nn, jstat, ndat, np, nfp, df, dx, Z, SUME)
c *****
c      frequency pattern matching by fuzzy functions
c      nn : total number of frequency data
c      np : mesh number of fuzzy function, jstat : data start no.
c      nfp : freq. mesh no., dx : mesh width, ndat : number of data per frame
c *****
      dimension xmu(10), amu(10)
      common fz(16, 32), h(32), data(8194), xh(8194), x(16)
      data wide, height/24.0, 17.0/
      data xmu/ 0., 100., 200., 250., 300., 350., 400., 500., 550., 600./
      data amu/ -0.4, -0.3, 0.0, 0.3, 1.0, 1.0, 1.0, 0.8, 0.1, 0.0/
      data nmu/10/
c scale factor of sca12 should be equal to height of plot graph
      sca12=data (nn+2)/ height
      sumw=0.0
      sume=0.0
      cc0=0.0
      istar0=0
      fmax=ndat*df
      if(fmax. gt. 1800.0) fmax=1800.0
      nmax=int (fmax/350.)
      jstop=int(1800./df)
      jend=jstat+ndat-1
      do 100 jd=jstat, jend
c serch for frequency band on fuzzy function
      do 40 j=1, nfp
          if (xh(jd). lt . h(j)) go to 50
40 continue
      j=j-1
50 js=j
c fuzzy evaluation of spectrum pattern for each frame
      cc=data(jd)*sca12
      do 60 i=1, np
          if(cc. lt. x(i)) go to 70
60 continue
      70 w=fz(i-1, js)+(fz(i, js)-fz(i-1, js))*(cc-x(i-1))/dx
          sumw=sumw+w
c count for frequency difference between adjacent peaks
      if(cc. lt. 0.30) go to 100
      cc2=data(jd+1)*sca12

```

```

if(jd. ge. jend) cc2=0.0
if(cc. lt. cc0. and. cc2. lt. 0.3) go to 100
if((jd-jstat). lt. jstop) go to 100
if(istar0. eq. 0) go to 90
    dfi=(jd-istar0)*df
c fuzzy evaluation of spectrum discreteness
do 80 m=2, nmu
    if(dfi. lt. xmu(m)) go to 85
80 continue
go to 100
85 e=amu(m-1)+(amu(m)-amu(m-1))*(dfi-xmu(m-1))/(xmu(m)-xmu(m-1))
sume=sume+e
90 istar0=jd
100 cc0=cc
    if(sume. lt. 0.0) sume=0.0
    z=sumw/float(ndat)
    g=sume/float(nmax)
return
end

```

#### 付録B フレーム時間軸パターン判別推論のプログラム

```

c .....
subroutine frmpat (nfrm, dt, y2p)
c *****
c pattern recognition of frame time history
c nfrm : number of frame, dt : frame time mesh width (sec)
c y2p : peak value for normalizeing frame amplitude
c *****
dimension xmu(10), amu(10)
common fz(16, 32), h(32), y(8194), xh(8194), x(16)
data xmu/0., 0.1, 0.13, 0.16, 0.2, 0.3, 0.4, 0.5, 0.6, 1.0/
data amu/1., 1.0, 0.9, 0.7, 0.3, 0.1, 0.05, 0.0, 0.0, 0.0/
data nmu/10/
c initial values
ak=0.5
c=exp(-2.5*dt)
j3=3*nfrm+1
y(j3)=y(2*nfrm+1)
caky=c*ak*y(j3)/y2p
y(4*nfrm+1)=y(j3)/y2p
y(5*nfrm+1)=0.0
y(6*nfrm+1)=caky
y(7*nfrm+1)=0.0

```

```

y(8*nfrm+1)=caky
y(9*nfrm+1)=0.0
do 100 i=2, nfrm
c differeciation
  j2=2*nfrm+j
  j3=j2+nfrm
  j4=j3+nfrm
  j5=j4+nfrm
  j6=j5+nfrm
  j7=j6+nfrm
  j8=j7+nfrm
  j9=j8+nfrm
  y(j3)=(y(j2)-y(j2-1))
c get plus and minus peak and normalization
  if(y(j3). ge. 0.0) then
    y(j4)=y(j3)/y2p
    y(j5)=0.0
  else
    y(j4)=0.0
    y(j5)=-y(j3)/y2p
  end if
c hold maximun of plus peak and attenation
  if(y(j4-1). ge. y(j6-1))then
    yint=ak*(y(j4-1)+y(j4))
  else
    yint=0.0
  end if
  y(j6)=c*(y(j6-1)+yint)
c hold maximun of minus peak absolute and attenation
  if(y(j5-1). ge. y(j7-1)) then
    yint=ak*(y(j5-1)+y(j5))
  else
    yint=0.0
  end if
  y(j7)=c*(y(j7-1)+yint)
c alarm sound recognition
  y(j8)=y(j6)+y(j7)
  if(y((j8). gt. 1.0) y(j8)=1.0)
c fuzzy evaluation for human voice
  do 70 m= 2, nmu
    yi=y(j2)/y2p
    if(yi. lt. xmu(m)) go to 75
70  continue

```

```
bk=0.0
go to 77
75 bk=amu(m-1)+(amu(m)-amu(m-1))*(yi-xmu(m-1))/(xmu(m)-xmu(m-1))
77 if(y(j5-1). ge. y(j9-1)) then
    yint=2.0*ak*bk*(y(j5-1)+y(j5))
    else
        yint=0.0
    end if
    y(j9)=c*(y(j9-1)+yint)
100 continue
return
end
```