

調 査

681.142 [534.78]

音声識別研究用計算機入力装置の検討

情報処理研究室

内 容

緒論

第1部 聴覚機能のアナログ・シミュレータの検討

1. 緒言
2. 聴覚機能とその機構の概要
3. 聴覚における周波数分析過程
4. 周波数分析過程のシミュレーション
5. 聴覚における時間的分析過程
6. アナログ・シミュレータの基本設計
7. 結言

第2部 音声スペクトル分析用 BPF 特性の検討

1. 緒言
2. 有声音特性
3. 無声音特性
4. 振幅過渡特性
5. 周波数過渡特性
6. 周波数配分と帯域分割
7. Pre-processing の特性
8. 結言

結論

緒 論

音声認識の研究および実験を計算機によって行なう場合、どのような形で音声波の情報を計算機に読み込ませるか、計算機内部で行なう認識のための情報処理の方法、必要な記憶容量、所要時間など、いいかえると最終的な音声の認識能力とそのための有効で能率的な計算機の使用という点に関連して非常に重要な問題である。

能率的な情報処理ということの問題にしなければ、原理的にいって音声の波形そのものを計算機に一度読み込み記憶させそのあとに必要な情報処理を行なえばよい。しかし実際的に考えて次のような理由から何らかの形で音声情報の“pre-processing”を行なうことが望ましくまた必要であろう。

まず、(1)音声波形そのものの持つ情報量が非常に多くしかも(2)音声の音韻としての識別には(人間の音声聞き取り結果から判断する限りでは)音声波形そのものは多くの冗長性をもっている。たとえば“vocoder”や“clipped speech”や“pitch synchronous speech”によっても音声の伝送が行なわれるということは、この冗長性の存在を立証している。

Pre-processing によってどのような音声情報を抽出するかはそのあとに計算機内部で行なう認識のための主情報処理過程とも関連するが、情報量が比較的少なくか

つ音韻識別に必要な情報はほとんど失われていないという観点からみて、現在考えられる可能性は短時間周波数スペクトルと零交叉波の利用である。しかしここでは次のような理由から周波数スペクトルの抽出についてのみ考察する。

(1) vocoder によってじゅうぶん明瞭な音声を送送しうること、また Sonagram のようなスペクトル分析による visible pattern によって視覚的にかかなりの程度までは音韻を読み取ることができることなどから、音声の短時間周波数スペクトル(およびその時間的な変化)が音声の音韻としての識別に必要な主要情報要素であると考えられ、しかもその情報量が比較的少ないこと。

(2) 音声発生理論との対応が明確であり、音響的あるいは生理的な段階で operational なスペクトル合成のモデル(数値的に計算機で計算可能な理論式)を作ることができること。

(3) 音韻識別のための主情報処理の過程において、音声についての従来の音声学、音韻論的な知識をじゅうぶん利用することができると考えられること。

(4) 聴覚における音声分析過程の一つとして周波数分析が行なわれていると考えられること。

(5) BPF 群によってアナログ的に比較的容易に時間おくれ少なくその情報を抽出することができること。

Pre-processing の一つとしてスペクトル分析を行なうとして、どのような方法でどのような特性のスペクトル分析を行なうかが次に問題になる。もし音声が単純に周期的な定常波形であるならばスペクトル分析は数学的な Fourier 分析の実現手段と考えられ、その優劣を比較することも簡単になるが、実際には比較的緩やかではあるがスペクトルが時間的に変化し、しかもその時間的な変化が重要な情報であり、さらに有音声においてはそのスペクトルが音源基本周波数いわゆるピッチ周波数の高調波成分からなる線スペクトルに近い構造をもっていることなどから問題は複雑になる。

周波数分析の仕方には原理的にいつて二つの異つた立場が考えられる。その一つは計算機を含めた全音声認識装置を人間の働きになぞらえ、入力装置に聴覚の働きを、計算機本体に大脳の働きを対応させたとき、その対応に単なるアナロジ以上の重要さをおく考え方である。少なくとも現在までのところ人間の音声識別能力の方が現存するいかなる機械的（計算機処理を含む）音声識別装置の能力よりもすぐれていることから、人間の機能に音声認識研究の目標とモデルを求めるのは一つの必

然的な考え方であろう。この立場からは入力装置の特性がいかに人間の聴覚機能のよりよい近似であるかという点からその特性が検討される。他の一つの考え方はできるだけ純物理現象的に考えて最適なスペクトル分析を行なおうとする立場である。物理現象的に考えるとはいつても、音声そのものが人間によつて発生され人間によつて聞き取られるということのくり返しによつて発展、形成されてきた人間相互間の情報伝達手段である以上、本質的に人間の機能に関係しないわけにはゆかないのだが、できる限り人間の機能との対応を抽象的概念的な範囲にとどめておこうとする考え方である。この立場からは入力装置の特性の一つ一つが適当に定められた物理的な判断基準（例えば最小電力誤差とか最大周波数分解能など）にしたがつて個別に検討されなければならない。

ここでは第1部に前者の立場から計算機の入力装置としての聴覚アナログ・シミュレータの基本特性とその基本設計を調査検討し、第2部に後者の立場から BPF 群による入力装置の諸特性を調査検討して、今後の研究の参考とした。

第1部 聴覚機能のアナログ・シミュレータの検討

中 田 和 男 中 野 康 明*

1. 緒 言

現在音声認識の研究は各国、各所で盛んに研究されつつあるが、今のところ人間に優る能力をもつた装置は実現されていない。その原因は人間の音声認識機構の特異性⁽¹⁾⁽²⁾にあると思われるので、人間の音声情報処理機能を無視して純粋に工学的に音声識別の研究を押し進めるというやり方は、必ずしも能率のよい見通しのはつきりした研究方法とはいえない。（一般的にいつて、生体の情報処理機能を解明することによつて、パターン認識の研究において新しい側面を開拓しうる可能性がある。）しかし大脳における一般パターン認識の機能は非常に複雑であり、しかも個々の脳細胞の機能が知られたとしても、このような徹底的な知識からだけでは脳全体としての総合的な機能を解明することはほとんど不可能に近い（人間の脳細胞の数は約150億個といわれている。）

そこでこの二つの立場、すべてを純粋に工学的に処理

しようとする立場と、すべてを生物（人間）の情報処理機能とのアナロジーで解決してゆこうとする立場の結合として、複雑で現在のところほとんど未知に近い大脳皮質の聴覚における音声認識機能は一応原理的にみて計算機で代行させるとして、その入力装置として感覚器である聴覚機能のアナログ・シミュレータを考え、それによつて能率的で有効な音声情報の抽出（pre-processing）を行なわせようとする立場をとつた。

しかし聴覚の機能にも不明な点は多く、ことに音声情報処理の面ではよくわかつていない。しかし単純な音響しげきに対する応答特性には知られているものも多く、また他の感覚から類推できるものもあるので、このような既知の応答様式を質的にも量的にも正しく実現できるように系を構成すれば、複雑な入力（音声）に対するその系の出力はじゆうぶん⁽³⁾に意味のある情報を与えてくれるものであると期待することができよう。以上のような観点から聴覚機能のアナログ・シミュレータについて基礎的な調査、検討を行なつたのでその結果をここに報告する。

*日立製作所中央研究所

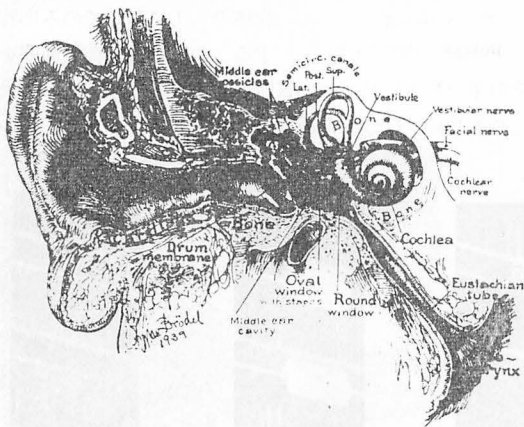
2. 聴覚機能とその機構の概要

2.1. 聴覚機能

人間の聴覚器官は入力音響しげきの情報処理にあつて、周波数的分析機能と時間的分析機能をもっている。音声の処理においては、主としてその短時間周波数スペクトル分析結果から音韻情報を、時間的分析結果から声性情報（たとえばピッチ高低の弁別など）を抽出しているものと考えられる²⁾。時間的分析と周波数的分析とは独立なものではなく相関連して音声認識に役立つものと思われるが、ここでは一応切り離して取扱い、総合的な処理は計算機で行なうものとした。

2.2. 聴覚機構⁽³⁾

人間の聴覚器官として、ここでは大脳皮質聴顔へ入るまでの外、中、内耳、聴覚神経系を考える。（第1図参照）



第1図 聴覚器官の説明図

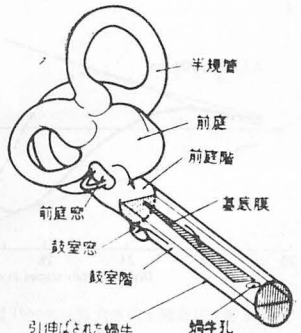
外耳は集音して音波を鼓膜に伝え、鼓膜は空気振動としての音波を機械的な変位に変換して中耳に伝える。中耳の入口は鼓膜であり奥は蝸牛につながる。この間に耳内三小骨の連鎖があり、これが外気と内耳（液体）の間の音響インピーダンスの整合、音圧増幅、ダイナミックレンジの圧縮（過大音に対する内耳の保護）などの役割を果たす。

内耳（蝸牛）はかたつむりに似た外形をもつ管で、内耳から伝えられた機械変位を神経パルスに変換する。

蝸牛内は基底膜で境され、上部が前庭階、下部が鼓室階であり、淋巴液で満されている。蝸牛両階の中耳端はそれぞれ前庭窓、鼓室窓であり、前庭窓は鐮骨足板で耳

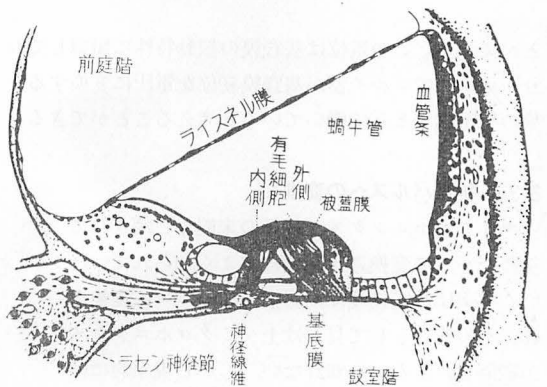
²⁾両耳による諸効果（たとえば方向感）はここでは考えない。

内三小骨につながり鼓室窓は薄膜である。また中耳端と反対側すなわち蝸牛頂には蝸牛孔があり、両階が通じている。第2図に蝸牛を引き伸ばした説明図を示す。



第2図 蝸牛を引伸ばした内耳

前庭階の一部はライスネル膜で斜に限られ基底膜、ライスネル膜 蝸牛外壁で囲まれた三角形断面の蝸牛管が形成され、その内部は淋巴液で満されている。基底膜上にコルチ器がついて基底膜振動を検出する。コルチ器の主要部分は有毛細胞であり、これは内側、外側の2群に分かれる。基底膜は薄いゼラチン状の膜で無入力の場合は張力はない。第3図に蝸牛内部（断面）の微細構造を示す。



第3図 蝸牛内部の微細構造（断面図）

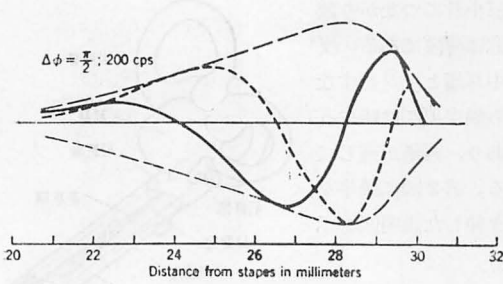
3. 聴覚における周波数分析過程

3.1. 基底膜振動

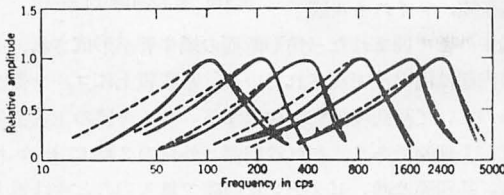
Békésy の観察によれば、正弦波振動を前庭窓に加えると基底膜上に進行波が生じ、蝸牛頂へ向つて進む。基底膜振幅の包絡線はある点で最大となり、その位置は周波数の関数となる。その様子を第4図、第5図に示す。この観測結果から逆に基底膜上の1点について各周波数に対する振幅を測り、周波数特性を画くことができる。これを第6図に示す。

3.2. マイクロホンツクスの発生

蝸牛内に微少電極を挿入し、外部から音しげきを加えると、マイクロホンツクスと呼ばれる電位を観測するこ



第4図 基底膜上の進向波パタンの1例(Békésy の観測による)



第6図 基底膜振動の周波数特性(実線は Békésy の実測値, 点線は Zwislocki の理論による計算値)

とができる。この電位は基底膜の振動特性に類似しており、簡単にはコルチ器が基底膜変位を電圧に変換する一種の圧電素子として働いていると考えることができる。

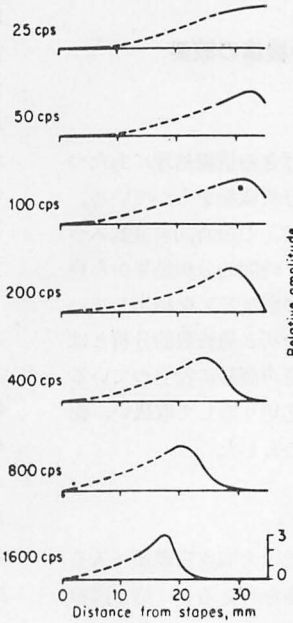
3.3. 神経パルスへの変換

マイクロホニックスが神経の末端をしげきして神経パルスに変換される過程はよく知られていない。事実は基底膜変位が直接神経パルスに変換され、その総合として見かけ上マイクロホニックスが観測されるのかも知れない。しかし現象的には振動がいつたんマイクロホニックスに変換されそれが神経パルスに変換されると見なすこともでき、またその方が電気的な等価回路表現を考えるうえでも考えやすい。

正弦波の音しげきが弱いときは、神経パルスとして自発性放電と呼ばれる不規則なスパイクが現われているが、しげきが強くなると入力周波数に同期してスパイクが現われるようになり自発性放電は消失する。周波数が低いときは同期したスパイクの数は1より大きい(2~3程度まで)が、周波数が増すにつれて同期的な対応は1対1になり、さらに周波数が高くなるとそのしげきの始めと終りのみにスパイクが出るような on-off response に近づく⁽⁵⁾(第7図参照)

3.4. 神経系での信号の伝達と処理

神経パルスはニューロンと呼ばれる神経線維の



第5図 基底膜上の進行波振動のエンベロープ

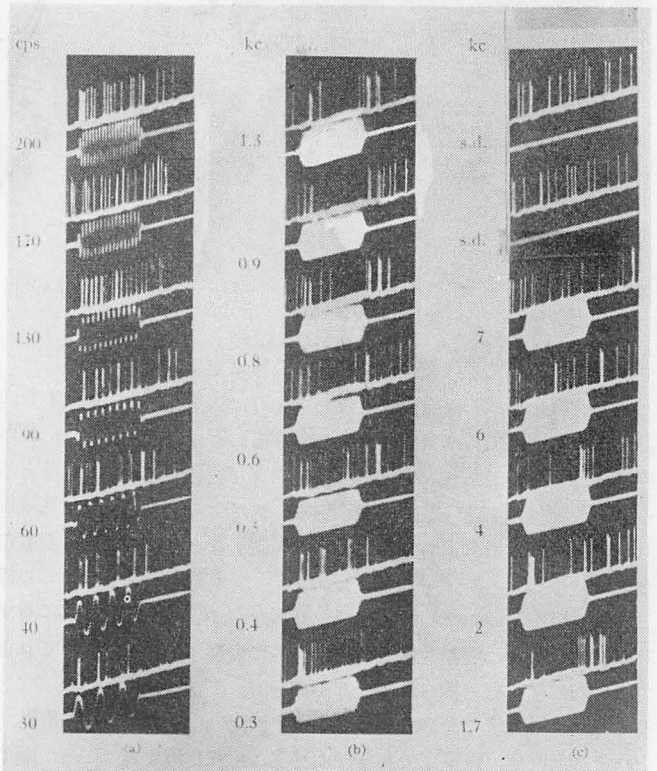
(Békésy の観測による)

その単一ニューロンの応答野という。

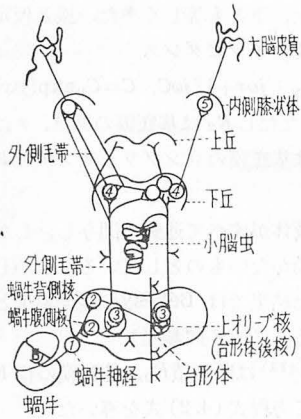
勝木⁽⁵⁾⁽⁶⁾は微小電極によりネコ、モルモット、サル

なかを伝わる。蝸牛を出る蝸牛神経線維は前庭神経、顔面神経と合して第八神経となる。聴覚神経の経路は第8図のようであり、聴覚神経の特徴としてニューロンの中継点が多いことで、大脳皮質聴領にいたるまでに5個の核を通る。この神経系での信号の伝達と処理の特徴の一つは、応答野の狭まり(周波数特性の sharpening)であり、他の一つは連続しげきに対する適応作用(adaptation)である。

(1) 応答野の狭まり(sharpening)：音響しげきに対する単一ニューロンの応答は周波数特性をもち、ある周波数範囲の入力に対してのみ応答する。この応答周波数範囲を



第7図 正弦波音しげきの周波数と神経パルスとの関係を示す実測データ(勝木による)

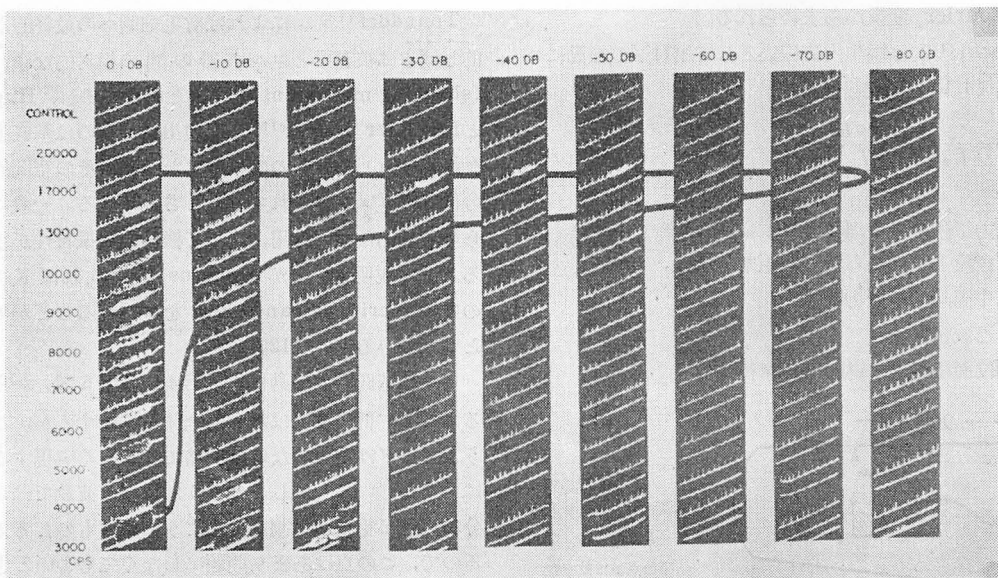
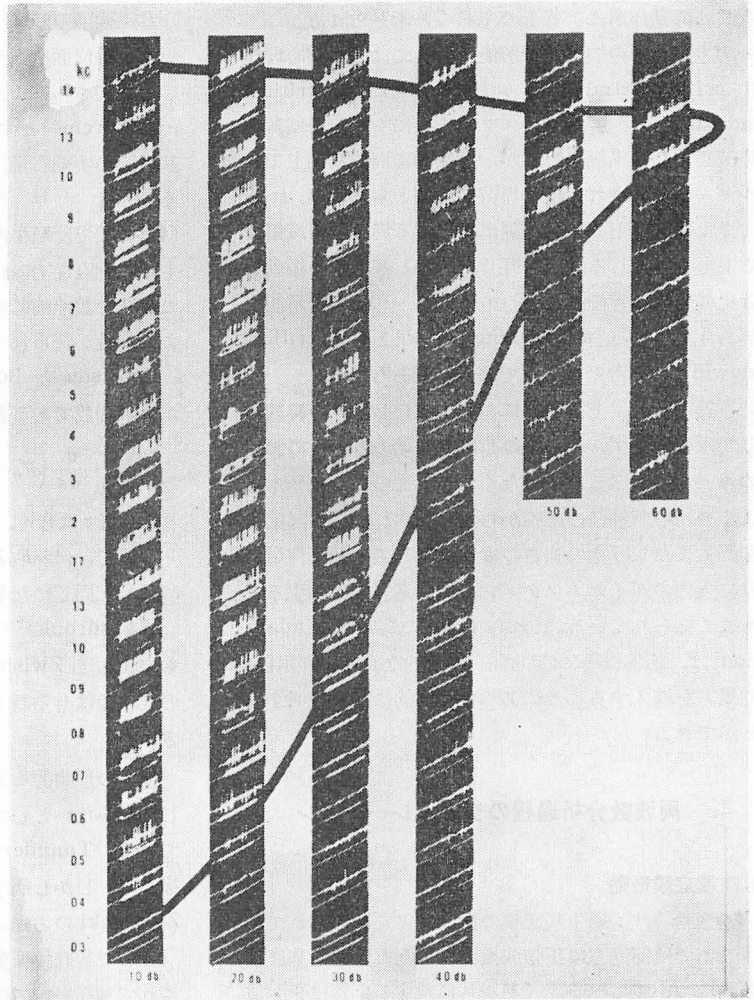


第8図 脳内聴覚神経路

第9図 猫末梢聴ニューロンの応答野

(勝木による)。縦軸は音の周波数 (kc), 横軸は音の強さ (dB), (太い実線が応答野を示す) →

などの聴覚神経活動を単一ニューロンのレベルで観測することに成功したが、その観測結果によれば、聴覚神経系を上にくにしたがいニューロンが代わるごとにその応答野が狭まり、間脳(内側膝状体)で最も鋭い周波数選択特性を有することがわかった。その1例を第9図と第10図に示す。このような応答周波数範囲の狭まり (sharpening) は、ニューロン



第10図 猫内側膝状体(間脳)における聴ニューロンの応答野 (勝木による)(縦軸は音の周波数(cps),横軸は音の強さ(dB)太い実線が応答野を示す)

相互間の抑制作用によるものであると考えられる。

一方人間の耳の周波数分解能については、聴覚心理学的に critical bandwidth が実測されている。critical bandwidth は、純音をマスクするのに有効な帯域雑音の帯域幅として実測されるが、聴覚的に周波数軸上でエネルギーが積分される積分周波数範囲とも考えられるし、また基底膜上の一定間隔に相当する周波数帯域幅であるとも考えられる。(7)(8)(9) しかし基底膜の振動特性と比較すると等価的な Q の値で6~10倍の分解能の差があり、むしろ sharpening の結果として critical bandwidth を考えるべきではないであろう。

(2) 適応現象：上部神経においては、しげきの持続に対しては応答せず、ただそのしげきの始めと終りにおいてのみパルスが出るにすぎなくなる。このことは勝木らによるサル聴神経の応答からも確認されている。(第7図参照) またしげきの強さの変化に対しても適応作用があり、入力ダイナミック・レンジを適度に圧縮することもよく知られている。これらすべての適応現象(adaptation)は、応答機構に時間的、強度的な負帰還作用(抑制効果)を導入することによって機能的によく説明することができる。

4. 周波数分析過程のシミュレーション

4.1. 基底膜振動

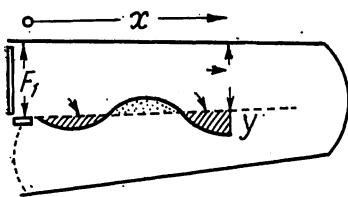
(1) 振動特性：蝸牛基底膜の振動はよく調べられている。ここでは前庭窓に正弦波振動を与えたときの基底膜の振動について述べる。(パルスに対する応答は正弦波応答の Fourier 変換から求められる。)

Zwislocki⁽¹⁰⁾ は蝸牛内方程式として第11図の振動に対応して(4.1)式を与えた。

$$\frac{\partial^2 p}{\partial x^2} = \frac{2}{F_0} \cdot \frac{j\omega\rho + R_1}{J} p$$

ここで

- ω : 入力正弦波の角周波数
- x : 基底膜上鐙骨から測つた距離
- p : 蝸牛両階間の圧力差
- ρ : 淋巴液密度
- R_1 : 単位面積当りの淋巴液の摩擦係数



第11図 基底膜振動(行進波)のおこり方 (F は前庭階, 左下の点線は鼓室窓の薄膜)

F_0 : 両階の断面積(上, 下とも等しくまた一様と仮定)

J : 単位面積当りの膜のインピダンス

ここでさらに、 $J = R_2 + j\omega J + 1/j\omega C$, $C = C_0 \exp(\beta x)$, $\beta = 1.5 \times \text{cm}^{-1}$ とする。ただし R_2 は基底膜の抵抗, σ は基底膜の実効質量, C は基底膜のコンプライアンスである。

(4.1)式では蝸牛内の液体がすべて運動に関与し、しかも膜に垂直な方向には動かないものとしている。しかしこの式を数値的に解いた結果では Békésy の観測結果とよい一致が得られている。(第6図参照)

Peterson と Bogert⁽¹¹⁾ は断面積 F_0 , 基底膜の幅 b が x の関数であるとして方程式(4.2)式を導いた。

$$\frac{1}{F_0} \frac{d}{dx} \left(F_0 \frac{dp}{dx} \right) + \frac{\omega^2}{c^2} \left(1 + \frac{2b\rho c^2}{j\omega F_0 Z} \right) p = 0 \quad (4.2)$$

ここで c は音速, Z は膜の単位長当りのインピダンス。上式で、 $c^2 \gg F_0 Z / 2b\rho$, F_0 および b が一定とすれば式(4.1)と同じになる。

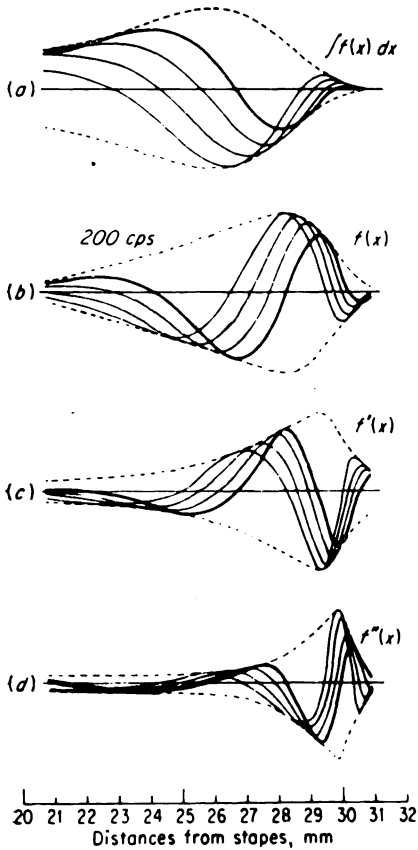
Wansdronk⁽¹²⁾ は基底膜のごく近くの液だけが動くものとして Zwislocki 型の方程式を立てているが、この理論では有効断面積が不明確なので数値的な検討はできない。

以上の理論的モデルでは流体は基底膜の長さの方向にしか動かないとしているが、実際には流体は渦運動をしている。Tonndorf⁽¹³⁾ は基底膜の運動を浅水波として扱った。しかし蝸牛内の非線型現象を考えない限りでは Zwislocki の方程式でじゅうぶんであろう。

ここで基底膜振動の結果によるいかなる物理量が検知されて神経パルスに変換されるのかを考えてみる。

Tonndorf⁽¹⁴⁾ によれば有毛細胞への最終的な機械的しげきは被蓋膜とコルチ器の間の相対的なずれ運動 (shearing movement) であるといわれるし、Hugging と Licklider の研究⁽¹⁵⁾ でも検出されるのは基底膜の振動振幅 $y(x)$ そのもののみではなくて、その空間的な微分値 dy^n/dx^n も含まれているとされている。一般にこれらの空間的な微分作用によって振動の共振特性が鋭くなることが説明されるが Huggins⁽¹⁶⁾ によれば dy^4/dx^4 の特性で critical bandwidth を説明することができる。とされている。(第12図参照)

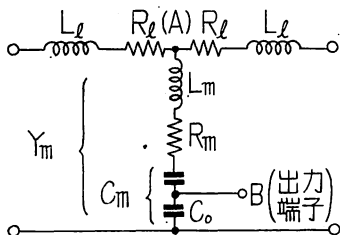
さらに基底膜の振動様式(第4図, 第5図)からわかるように、低周波では基底膜全体が振動するが、このようなモードは電気的な等価回路ではうまく実現することはむずかしい。しかしこのような低い周波数では時間的分析機能が重要な役割を果たすようになるものと考えられるので、この点はあまり問題にしないでよいと思われる。



第12図 空間微分による周波数分解能の向上
(a. 積分波形, b. 基底膜上の進行波形, c. d. 空間的な一次および二次微分波形)

(2) 基底膜のシミュレーション：基底膜の電氣的模擬回路の構法としては二通りの考え方がある。一つは蝸牛の力学的運動を直接電氣的にアナログ回路として表わすもので、通常集中定数による不均一伝送線路で実現される。他の一つは蝸牛を一つのブラックボックスとみなして、Békésy によつて観測された特性を示すようなフィルタを実現するものである。

まず前者の立場でアナログ回路を作ることを考える。蝸牛基底膜を微小部分に分割し、各微小部分のアナログ回路を作る考え方と、Zwislocki の方程式 (4.1) を Kirchhoff の方程式に対応させて回路を作る考え方とがあるが、いずれも同じ結果になる。第13図にその1区間



←第13図 蝸牛1区間のアナログ回路

の回路を第14図に全体の構成図を示す。図における諸定数と実際の諸量との対比は次のようである。

- $L_l, R_l \leftrightarrow$ 流体の質量, 抵抗
- $L_m, R_m, C_m \leftrightarrow$ 膜の質量, 抵抗, コンプライアンス
- A 点の電圧 \leftrightarrow 膜の両側の圧力差
- Y_m に流れる電流 \leftrightarrow 膜の振動速度

膜の変位は膜速度を積分すれば得られるので第14図でコンデンサ C_m の端子電圧をとりだせばよい。 C_m を二つのコンデンサの直列結合とし、一方を定容量としてその端子電圧 (第13図の B 点) を取りだせば場所によらず一定の比例係数で出力がえられる。なお空間微分は近似的に隣接区間の出力との差をとればよい。

Bauch⁽¹⁷⁾ は基底膜を分割し、Zwislocki の方程式で定数を、

$$\begin{aligned} \rho &= 1 \text{g/cm}^3, R_l \approx 100 \text{dyne} \cdot \text{sec/cm}^4, \\ R_m &= 5 \cdot 10^3 \text{dyne} \cdot \text{sec/cm}^4 \\ F_0 &= 8 \cdot 10^{-9} \text{cm}^2, C = C_0 e^{1.5x}, \\ C_0 &= 2 \cdot 10^{-10} \text{cm}^4/\text{dyne} \end{aligned}$$

と仮定したので第13図の各定数は、

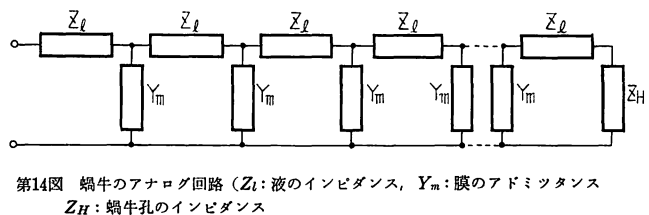
$$\begin{aligned} C_m &= 8.34 e^{1.5x} (\mu\text{F}), L_m = 0 \\ R_m &= 120\Omega \\ R_l &= 0.75\Omega \quad L_l = 7.5 \text{mH} \end{aligned}$$

となつた。その結果実験的によい近似特性を得ている。

Wansdronk は Fletcher の公式に従い、すべての素子の値が指数的に変化するものとして、100 区間からなるアナログ回路を作つた。しかし膜のごく近傍の流体だけが振動に関与するとしたので L_l および R_l の値がはつきりときまらない欠点がある。

Peterson と Bogert⁽¹¹⁾ は前庭階、鼓室階、基底膜をそれぞれあらわす三角型の立体的な回路を提案した。しかし実験としてはこれを梯子型回路に変換して実現しており⁽¹⁸⁾、実際には第14図とほぼ同様な回路となつている。したがつて、われわれは前にものべたように Zwislocki の方程式による第14図の回路構成でじゅうぶんであると考える。

第2の考え方にもとづく研究の一つとして、Flanagan は Békésy の観測結果をなるべく簡単な物理的に実現可

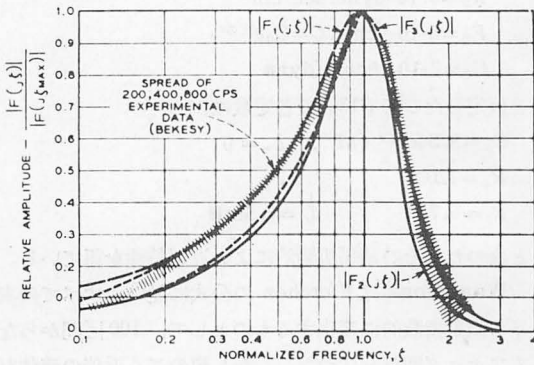


第14図 蝸牛のアナログ回路 (Z_l : 液のインピーダンス, Y_m : 膜のアドミッタンス Z_H : 蝸牛孔のインピーダンス)

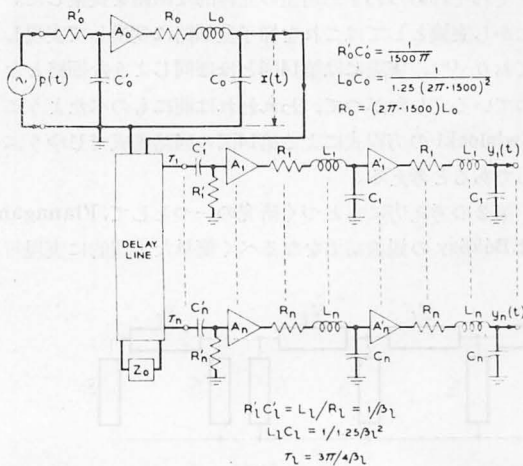
能な回路特性としてあらわすことを計算機によつて試みた。Flanagan⁽¹⁹⁾⁽²⁰⁾の求めた3種類の式のうちの一つを例示すれば、基底膜振動と同じ周波数特性を示すフィルターの伝達関数は β を位置のパラメータとして、

$$F_1(s, \beta) = C_1 \beta^4 \left(\frac{2000\pi}{\beta + 2000\pi} r \cdot \frac{(s + \varepsilon)}{s + \gamma} \right) \times \left[\frac{1}{(s + \alpha)^2 + \beta^2} \right]^2 e^{-\frac{3\pi}{4\beta} s} \quad (4.3)$$

ここで β は、基底膜上で尖頂からの距離 x (cm)とすれば、 $x = 0.75 \log_{10}(\beta/40\pi)$ となるような位置をあらわすパラメータであり、 $\varepsilon, \gamma, \alpha$ は位置のみに依存するパラメータで、 $\varepsilon = 0, \gamma = \beta, \alpha = \frac{1}{2}\beta$ ととられている。 $C_1, r(=0.83)$ は定数である。この特性をBékésyの実測結果と比較して示すと第15図のようである。



第15図 Flanagan の計算機シミュレーションによる基底膜振動特性と Békésy の実測値との比較

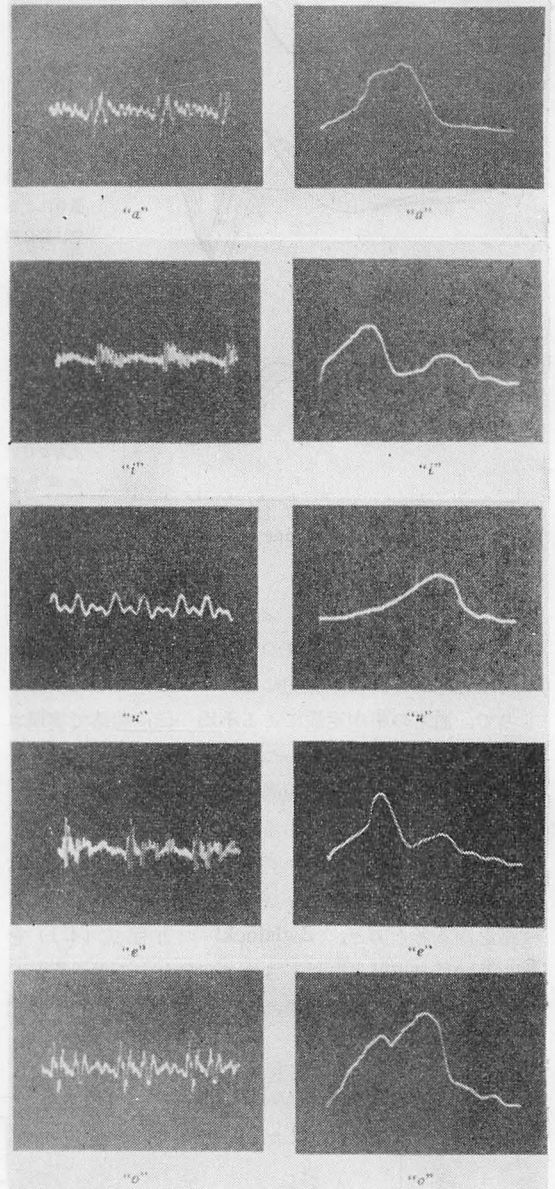


第16図 Flanagan の計算式をあらわす電気回路 $|G(s)F_1(s)|$
($G(s)$ は中耳の特性, $F_1(s)$ は基底膜の振動特性)

F_1 の回路的な実現は第16図のようになる。同図で $R_1' \cdot C_1' = L_1 / R_1 = 1/\beta_1, L_1 C_1 = \frac{1}{1.25} \beta_1^2, \tau_1 = 3\pi/4\beta_1$ である。

Flanagan はこの $F_1(s, \beta)$ の逆変換としてインパルス・レスポンス $f_1(t)$ を計算機によつて求めている。

Siebert⁽²¹⁾ は計算機シミュレーションによつて模擬する場合には物理的に実現可能なフィルター特性のみに限る必要はなく、インパルス・レスポンスの形で解を与



第17図 Wansdronek の基底膜振動特性のアナログ・シミュレータによる音声スペクトルの分解 (スペクトル・エンベロープ)
(オランダ語母音, 男声, ピッチ約100cps)
左: 音声波. 右: スペクトル

えればよいとして、かなり複雑な形のインパルス・レスポンス特性を与えた。

これらの第2の考え方による方法は、計算機による蝸牛機能のシミュレーションの場合に有利である。

いずれの型のシミュレーションでも、各区間の出力を整流してブラウン管面上に表示すれば、(水平軸には基底膜上の位置 x をとる) 基底膜振動による分析スペクトルが直視される。その2,3の例⁽¹²⁾を第17図に示す。音声波入力によつてどのようなスペクトル包絡線が得られるかについて Wansdronek, Caldwell⁽²²⁾らが研究しており、音韻識別に使用するつもりのものであるが、このままでは従来のソナグラムやフィルター群による分析にくらべて特にすぐれているといった特性はみとめられない。

4.2. 神経パルスへの変換

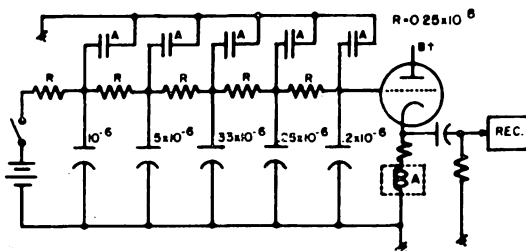
すでにのべたように、ここでは基底膜振動によつてマイクロホニックスが生じ、その電位が神経パルスに変換されるものとする。聴覚におけるマイクロホニックスの神経パルスへの変換過程はよくわかっていないが、他の感覚(受容)器や制御指令の神経パルスへの変換機能から類推することはできる。

Dallos と Jones⁽²³⁾らはカプトガニの光受容器やカエルの筋紡錘で、入力強度と神経パルスの周波数とが対数的な比例関係にあることを説明するために第18図のような回路を考えた。図でC-R回路は模擬伝送線路であり、その終端電圧が閾値 E_0 に達すると多極リレーが働いてすべてのコンデンサを短絡する。出力パルスの周波数は

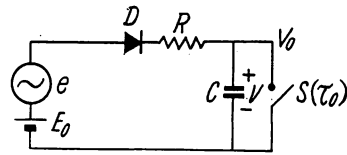
$$f_p = \begin{cases} k \ln(E/E_0) & E \geq E_0 \\ 0 & E < E_0 \end{cases}$$

で与えられる。

磯⁽²⁴⁾は Jones とは独立に聴覚に対して第19図の回路を考えスイッチに不応期を加えた。光の受容はエネルギー的であるから正の値だけを考えればよいが、基底膜の振動は正負の値をとるので正値のみを検出するように



第18図 模擬伝送線路による神経パルス変換回路(Aは多極リレー)



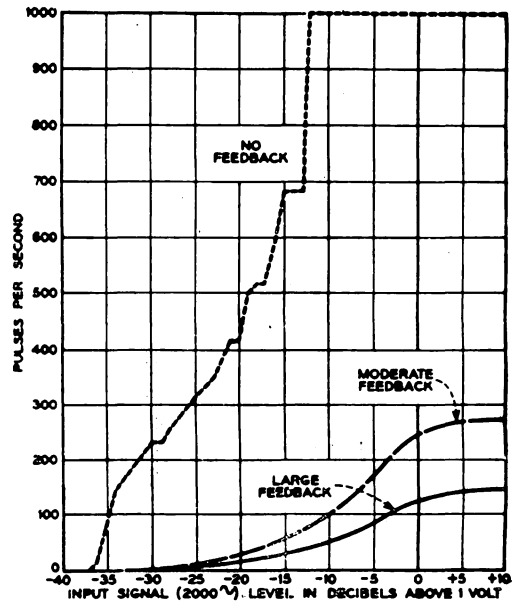
第19図 磯による弛張発振器型神経パルス変換回路 (D:diode, V_0 :閾値, τ_0 :不応期)

整流器を挿入した。Flanagan はコルチ器有毛細胞が基底膜振動の正方向の振幅のみを検知すると仮定して、クリック音に対する方向感の実験結果が説明できることを示しているし、勝木らの単一ニューロンの正弦波振動に対するレスポンスからもこのように考えることの妥当性がうかがわれる。磯の実験によればこの回路の正弦波入力に対する放電パターンは聴神経における神経パルス・パターンに類似し、とくに低周波における同期の様相がよく説明される。

Schott⁽²⁵⁾は相対不応期をもつ弛張振動回路により入力をパルス変換し、そのバイアス電圧に出力パルスの整流平滑値を負帰還させてパルス変換の閾値が変化するような回路をモデルとして実験した。

その結果入力振幅の値と出力パルス数との関係が第20図のようなS字特性を示すことが示された。

Weiss⁽²⁶⁾は閾値と不応期だけを考えた計算機モデルで、Kiangのクリック音に対する神経パルスの応答データを説明しようと試みており、特に信号がないとき、白色雑音を入力に加えることによつて自発性放電を説明しようとしている。しかしこの考察は、音響しげきとそれ



第20図 負帰還を加えたパルス変換回路の変換S字特性 (Schott による)

に対する神経パルスとの対応関係を巨視的に見たもので、基底膜、コルチ器、有毛細胞などの機能は総合して black box とみられている。

4.3. 抑制作用

すでにのべたように sharpening および adaptation の現象は、空間的（周波数的）、時間的なニューロン間の抑制作用によつて説明され、その抑制作用は神経パルスの再生中継点シナプスで行なわれる。

第21図に示すようにニューロンには細胞体と軸索および樹状突起があり、細胞体に発生した神経パルス（スパイク）は軸索をへて次のニューロンに伝達される。軸索は他のニューロンの細胞体および樹状突起と、シナプスと呼ばれる接合部を作る。すなわち樹状突起は入力端子、細胞体はスパイクの発生部、軸索は出力伝送路とみなせる。

軸索を伝わってきたパルスは末端で化学的物質を分泌させ、これがニューロンのシナプス下膜に作用する（シナプスは電気的には絶縁されているので伝導は化学的のみに行なわれる）。このためにニューロン内部には興奮性シナプス電位 EPSP を発生する。普通1個の神経パルスによる EPSP はニューロンを興奮させるために必要な電位の $1/10$, $1/20$ であり、興奮が次のニューロンに伝達

されるためには、入力パルスが空間的にまた時間的に加算される必要がある。またシナプス電位には興奮性のものほかに抑制性のものがあり、入力パルスが抑制性経路から入ると EPSP とは逆極性の電位 IPSP を発生する。

EPSP と IPSP を空間的、時間的に加算したものがシナプス電位であり、この値がある閾値を超えるとニューロンの放電が起り、新しい神経パルスが次へ伝達される。放電が起つたあとではシナプス電位は消失し、数10 m. sec の不応期がある。

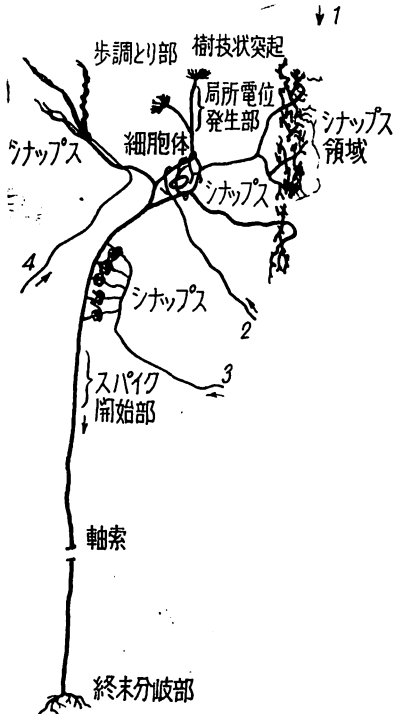
シナプスおよびニューロンのモデルとしては電氣的にまたは化学的に数多くのものが研究、実験されている。

空間的なニューロン相互間の抑制作用による周波数特性の sharpening については、鈴木、池田⁽²²⁾および Martin と Talavage⁽³⁰⁾らによつて実験されている。第22図に前者の実験モデルとその結果を、第23図に後者の実験モデルとその結果を示す。

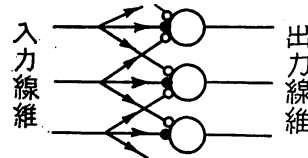
これらの結果は簡単なモデルについてではあるが、ニューロン間の空間的（周波数的）な相互抑制作用によつて周波数特性の sharpening が行なわれることを実験的に立証している。

このことは概念的には次のように考えることもできる。すなわち第21図のモデルでシナプス電位を与える鈴木らの式

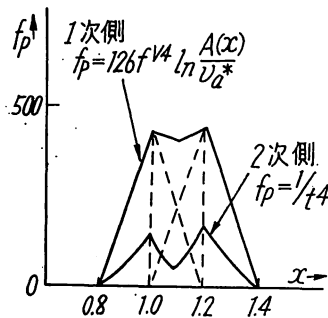
$$V_b(x_0, t) = V_s(x_0, t) - \gamma[V_s(x_0 + 4x, t) + V_s(x_0 - 4x, t)]$$



第21図 ニューロンの構造説明図

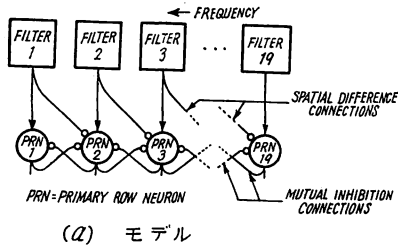


(a) 抑制効果のモデル
● 興奮性結合, ○ 抑制性結合



(b) 2信号の分離

第22図 抑制による信号の分離特性 (鈴木らによる)

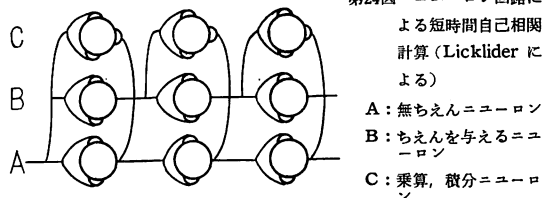


字特性を説明した。(第20図参照) また Harmon と Bergeijk も彼らの神経系モデル“neuromime”を用い、抑制機構が複雑になるほどダイナミックレンジが増大するというを示した⁽³¹⁾。またこの時間的な抑制効果によつてしげきの連続に対する応答神経パルスの間引き、いわゆる adaptation を説明することもできる。

5. 聴覚における時間的分析過程

聴覚のピッチ識別能力は、すでにのべたような周波数分析機能で考えられるものよりはるかに優れたものであり、その機能は波形の時間的な分析による周期性の検出過程にもとづいているものと考えられる。例えば高調波成分を多く含んだ波形からフィルターで基本波成分を除いても、聴覚によればその基本波の周波数のピッチが聞こえることはよく知られた事実である⁽³²⁾。この場合には基本周波数の存在という周波数分析的な過程ではなくて、周期性の検出という時間的な分析過程によるものと考えられる。

Licklider⁽³³⁾ はピッチ検知に周波数分析と短時間自己相関による周期性検出の二つの過程を考えた。短時間自己相関は第24図のような神経回路網により計算され、相関出力はそのピッチ周波数に相当するニューロンで相関最大となり、最大出力を与えるニューロンの位置として空間的に検出される。



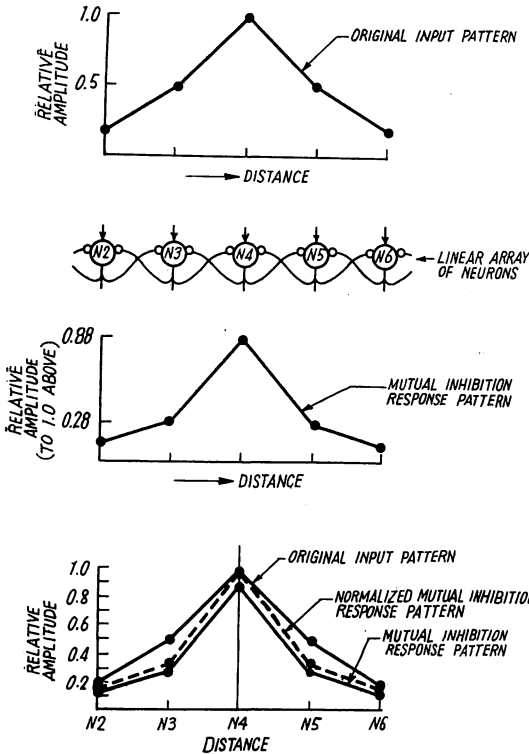
この神経回路網は通常の関相計の形をニューロンの形であらわしたものであるが、このような神経回路網の存在が神経生理学的にまだ裏づけられていない点問題である

しかし神経系のどこかで自己相関の測定または生体内の標準(周波数)との相互相関の測定が行なわれている可能性はじゆうぶんあるように思われる。

6. アナログ・シミュレータの基本設計

6.1. 基本的な構成

前節までの調査結果によつて考えられる聴覚のアナログ・シミュレータの基本的な構成を第25図にブロック図で示す。各部について簡単に説明すると、



(b) SYSTEM RESPONSE の SHARPENING

第23図 抑制効果による system response の sharpening (Martin と Talavage による)

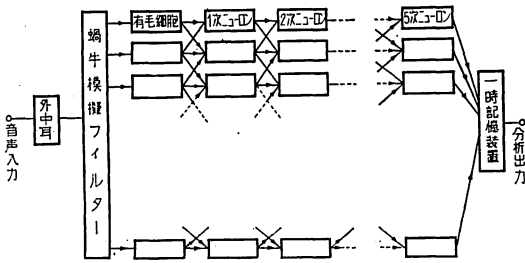
(V_s はシナプス電位であり、 x はニューロンの番号に相当する空間座標をあらわす) で抑制度 γ を λ とすれば、

$$V_s(x_0, t) = \frac{1}{2} [V_s(x_0, t) - V_s(x_0 - \Delta x, t) - V_s(x_0 + \Delta x, t) - V_s(x_0, t)]$$

$$\cong \frac{1}{2} \frac{\partial^2 V_s}{\partial x^2} \Delta x^2$$

となりシナプスで再生中継されるたびに空間的な二次微分がとられることとなり、そのため sharpening がおこる。

空間的な抑制だけでなく時間的な抑制(出力の積分値の負帰還)も考えられ、すでにのべたように Schott⁽²⁵⁾ はこの効果によつて放電パルスのしげき強度に対する S



第25図 聴覚アナログ・シミュレーターの基本構成

- (1) 外, 中耳フィルター: 簡単に低域通過フィルタと考えてよく, 遮断周波数 1.5kc, 減衰特性は 18db/oct となる。
- (2) 蝸牛模擬フィルター: Zwislocki 流の伝送線路型とするか, Flanagan 流のフィルタ・バンク型とするか簡単にきめることはむずかしいが, 特性を調整しやすいという点ではフィルタ・バンクの方が有利であろう。
- (3) 有毛細胞モデル: 磯氏のモデル(第19図)でよいと思われるが時定数周波数別の閾値などについては実験的に定めなければならない。
- (4) ニューロンモデル: McCulloch-Pitts のモデルとか Caiancillo のモデルとかがよく知られているが, ここで問題なのは徹視的な個々のニューロンの働きではなくて, 全体としての情報処理機能とくに sharpening と adaptation の機能を抑制効果によつて実現することであり, そのような観点から検討しなければならない。

6.2. 設計に必要な基礎実験

有毛細胞のモデルについては, 外, 中耳フィルター, 蝸牛模擬フィルタとの総合特性として, その放電の閾値を聴感度特性と合うよう調整しなければならないし, このような総合周波数特性としての聴感度特性をどこでどうとるかをきめなければならない。ニューロン・モデルとしては抑制効果によつていかに sharpening と adaptation の特性を実現するかが問題であり, ニューロン自体のモデルとしては比較的簡単なもので, むしろ抑制の範囲, 結合の強さの変わり方, 時定数などについて実験的に検討してきめることが必要であろう。そしてこのような簡単なニューロン・モデルとその結合系でどれだけの機能と特性を実現しうるかをまず検討してみなければならない。

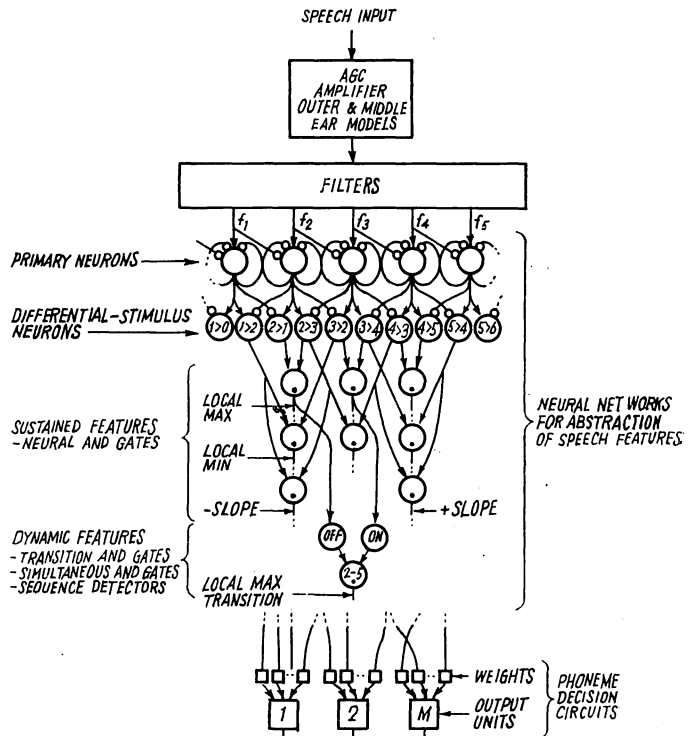
そのほか実際の設計上の問題としては蝸牛

模擬フィルタの区間数を何個にするかが問題であるが, 少なくとも 100 個程度は必要であろう。区間数が多いということになると伝送線路型の方が有利になつてくる。またニューロンの数も 5 層のすべてに 100 個ずつ必要ではなくて, おそらく層を増すごとにその数をへらしてゆくことも可能であろう。

最終的な出力としては第25図の基本ブロック図に示されているように第5次ニューロンの出力としてパルス周波数変調されたパルスの形で得られる。計算機によみこむ場合すでに単位パルスに変化されているのでA-D変換の必要はないが, 時間的には非同期的に出力が得られるので出力側に一時的な記憶装置が必要となる。

7. 結 言

ここでは脳における情報処理機能を電子計算機によつて代行させ, そのための最も能率的で整合のとれた入力装置の一つの可能性として聴覚機能のシミュレータを考えた。Martin と Talavage⁽³⁰⁾ は音韻まで識別しうるような神経回路をニューロン・モデルで構成しているが, (第26図参照)この考え方では脳における情報処理の機能までニューロン・モデルでシミュレートしようと



第26図 ニューロン・モデルによる音韻識別系 (RCAの研究)

しているもので、一貫したモデルとしては興味深い、音韻の識別能力という最終目標に対しては必ずしも有効な方法とはいえないと思われる。

われわれの立場としては単純なしげきに対して正しく応答するような系を作りあげ、複雑な入力である音声に対してこの受容器モデルがどのような応答を示すか、どのような出力が得られるかに興味がある。

工学的な方法だけからでは生体の情報処理機能におよばない現在、このような聴覚系の機能に学ぶところは非常に多いと思われ、このアナログ・シミュレータの研究、実験は有用であり、示唆に富む結果が得られるものと期待される。しかし実際の装置のハードウェアを作る前に、デジタル的にまたアナログ的にじゆうぶん計算機シミュレーションを行なつてその特性を検討することが必要である。

以上音声認識の点から聴覚のアナログ・シミュレータの重要性を論じたが、生体そのものではなくてその機能から学んだ点を工学上に応用することも“Bionics”の一つの方法であり⁽³⁴⁾、音声研究のみでなく他の応用面が開かれる可能性もないわけではない。

第1部 主要参考文献

- (1) D. B. Fry, et al: The identification and discrimination of synthetic vowel, *Language and Speech*, 5, part 4, 171~189(1962).
- (2) A. M. Liberman, et al: A motor of speech perception, *Speech Communication Seminar*, Stockholm, (1963).
- (3) G. von Békésy and W. A. Rosenblith: The mechanical property of the Ear, *Handbook of Experimental Psychology* (Ed. S. S. Stevens), 1075~1115, (1951).
- (4) G. von Békésy: *Experiments in hearing*. (Ed. E. G. Wever). McGraw-Hill, (1960).
- (5) Y. Katsuki, et al: Neural mechanism of the peripheral and central auditory system in monkeys, *JASA*, 34, 9-2, 1396~1410, (1963).
- (6) 勝木保次: 聴覚における音波の分析機構, *日本物理学会誌*, 17, 6, 401~410, (1962).
- (7) H. Fletcher: *Speech and hearing in communication*, Van Nostrand, (1953).
- (8) E. Zwicker: Subdivision of the audible frequency range into critical bandwidth, *JASA* 33, 2, 248, (1961).
- (9) D. D. Greenwood: Critical bandwidth and frequency coordination of the basilar membrane, *JASA*, 33, 10, 1344~1356, (1961).
- (10) J. Zwillocki: Theory of the acoustical action of the cochlea, *JASA*, 22, 6, 778~784, (1950).
- (11) L. C. Peterson and B. P. Bogert: A dynamical theory of the cochlea, *JASA*, 22, 3, 369~381, (1950).
- (12) C. Wansdronk: On the mechanism of hearing, *Philips Research Rep. Suppl.*, 1962. 1.
- (13) J. Tonndorf: Fluid motion in cochlea models, *JASA*, 29, 5, 558~568, (1957).
- (14) J. Tonndorf: Time/frequency analysis along the partition of cochlea models, *JASA*, 33, 9-2, 1337~1350, (1963).
- (15) W. H. Huggins and J. C. R. Licklider: Place mechanisms of auditory frequency analysis, *JASA*, 23, 3, 290~299, (1951).
- (16) W. H. Huggins: A theory of hearing, *Communication Theory*, (Ed. W. Jackson), 363~380.
- (17) H. Bauch: Die Schwingungsform der Basilar membran bei Erregung durch Impulse und Geräusche, Gemessen an Einem Elektrischen Modell des Innenohrs, *Frequenz*, 10, 2, 222~234 (1956).
- (18) B. P. Bogert: A network to represent the inner ear, *Bell Labs. Record*, 28, 11, 481~485, (1950).
- (19) J. L. Flanagan: Models for approximating basilar membrane displacement, *BSTJ.*, 39, 5, 1163~1191, (1960).
- (20) J. L. Flanagan: Models for approximating basilar membrane displacement part II, *BSTJ.*, 41, 3, 959~1009, (1962).
- (21) W. M. Siebert: Models for the dynamic behavior of the cochlea partition, *MIT. QRP.*, 64, 242~258, (1962).
- (22) W. F. Caldwell: Recognition of sounds by cochlear patterns, *IEEE. Trans.*, MIL-7, 2, 3, 179~185, (1963).
- (23) P. J. Dallos and R. W. Jones: A receptor analog having logarithmic response, *IEEE. Trans.*, BME-10, 1, 13~15, (1963).
- (24) 磯 泰行: 神経系の通信特性(I)パルス周波数変調の基本特性, *九工大研報*, 12, 13~23, (1962).

- (25) L. O. Schott : Signal translation in hearing
Bell Labs. Record, 30, 1, 2~8, (1962).
- (26) T. F. Weiss : A mode, firing pattern of
auditory nerve fibers, MIT. QPR., 69, 217~223,
(1963).
- (27) 若林 勲 : 細胞の電氣的活動, 科学, 29, 4, 176
~180(1959).
- (28) 伊藤正男 : 神経細胞をめぐる抑制と興奮, 科学,
33, 10, 512~516, (1963).
- (29) 鈴木良次, 池田泰子 : 聴覚神経のシミュレータ,
情報と制御の研究, 4, 37~45, (1963).
- (30) T. B. Martin and J. J. Talavage : Application
of neural logic to speech analysis and recogni-
tion, IEEE. Trans., MIL-7, 2, 3, 189~196, (1963)
- (31) L. D. Harmon et al : Analog models of
neural mechanism, IRE. Trans., IT-8, 2, 107
~112, (1962).
- (32) J. F. Schouten, et al : Pitch of the residue,
JASA., 39, 9-2, 1418~1424, (1962).
- (33) J. C. R. Licklider : Periodicity pitch and
related auditory process models, International
Audiology, 1, 1, 11~36. (1962).
- (34) L. M. Butsdh, Jr. and C. W. Gwinn : Bionics-
status and plans, IEEE. Trans., MIL-7, 2, 3,
261~266, (1963)

第2部 音声スペクトル分析用 BPF 特性の検討

中 田 和 男 松 野 辰 治*

1. 緒 言

音声波のスペクトル分析を行なう方法としては次のよ
うなことが考えられそれぞれすでに実用されている。

- (1) 音声波形を計算機内部で調和解析する。
- (2) 音声波形をヘテロダイン式に frequency sweep
する。
- (3) 音声波形を BPF 群によつて周波数分析する。
- (4) 特殊な分析法による。

(1)の方法の実例はベル電話研究所における“pitch
synchronous analysis”⁽¹⁾であるが、すでにのべたよ
うにこの方法は計算機の能率的な利用法とはいえない。
(2)の方法は通常のスペクトログラムに多く用いられてい
るが、“sonagram”にみられるように real time での
分析は困難である。(3)の方法はいわゆる vocoder 型の
スペクトル分析であり、分析のための時間おくれも少な
くほぼ real time での分析が可能であるのでここでは
この方法について検討する。(4)には最近いろいろのもの
が考えられているが、その中で注目すべきものは、Fe-
deral Scientific Co. の“SIMORAMIC” analyzer⁽²⁾
である。しかし現在のところ装置の複雑さに問題がある

のでここでは検討の対象としない。

音声スペクトルの BPF 群による分析において、検討
すべき諸特性は次のようなものと考えられる。

- (1) BPF 個々の特性
 - (a) 有声音特性 (periodic impulse response)
 - (b) 無声音特性 (random noise response)
 - (c) 振幅過渡特性 (transient response)
 - (d) 周波数過渡特性 (frequency transient
response)

(2) BPF 群としての特性

- (a) 周波数配分 (frequency scale の選択)
- (b) 帯域分割 (周波数分解能とチャネル数)

このほかスペクトル分析装置全体としては、各 BPF
出力の検波後の平滑特性、スペクトル分析する前の周波
数的な pre-emphasis の特性、振幅的な compressor
の特性などが問題になる。

以下これらについて検討する。

2. 有 声 音 特 性

有声音 (主として母音) に対する分析フィルターの特
性は、音源のくり返し周波数 (ピッチ周波数) の変化に

*東京電機大学

よる分析出力の変動と音声スペクトルの分析フィルターによる周波数分解能の点から検討される。

2.1. 分析出力のピッチ周波数の変化による変動⁽³⁾

分析 BPF $U\left(\frac{x}{d}\right)$ ($x=\omega-\Omega_0$, $F_0=\Omega_0/2\pi$ は通過帯域の中心周波数, d は等価帯域幅) の impulse response を $h(t)$ とすれば周期 t_0 の periodic impulse による BPF の定常出力 $H(t)$ は (2.1) 式によつて計算される。
($\Omega_0/d \gg 1$ と仮定する)

$$H(t) = \sum_{n=0}^{\infty} h(t+nt_0) \quad 0 \leq t \leq t_0, \quad (2.1)$$

$f_0 = 1/t_0$: ピッチ周波数

一般に $h(t) = R_e[h(t) \exp[j\Omega_0 t]]$ とかけられるから,

$$H(t) = \sum_{n=0}^{\infty} R_e[h(t+nt_0) \cdot \exp[j\Omega_0(t+nt_0)]] \quad (2.2)$$

(2.2) 式は振幅が $R_e[h(t+nt_0)]$ で位相が $j\Omega_0 nt_0$ であるような周波数 F_0 のベクトルの和と考えられる。そこで $R_e[h(t)]$ が負にならない場合には*, 分析出力は $\Omega_0 t_0 = 2\pi N$ (N : 整数) すなわち $F_0 = N \cdot \frac{1}{t_0} = N f_0$ (同調) のとき最大となり $\Omega_0 t_0 = 2\pi N \pm \frac{1}{2}$ すなわち $F_0 = \left(N \pm \frac{1}{2}\right) f_0$ (完全非同調) のとき最小となる。したがつてそのときの出力の変動比は (2.3) 式で計算される。

$$\left| \frac{[H(t)]_{max}}{[H(t)]_{min}} \right| = \frac{\sum_{n=0}^{\infty} R_e[h(t+nt_0)]}{\sum_{n=0}^{\infty} (-)^n R_e[h(t+nt_0)]} \quad (2.3)$$

実例として単一同調回路 1 段の BPF 特性について計算してみると次のようになる。

$$U\left(\frac{x}{d}\right) = \left[1 + (2Qx)^2\right]^{-\frac{1}{2}} = \left[1 + (x/d)^2\right]^{-\frac{1}{2}}$$

$d = 1/2Q$

$$\therefore h(t) = A \exp(-t/2Q) \exp(j\Omega_0 t) \quad t \geq 0$$

$$\begin{aligned} \therefore H(t) &= \sum_{n=0}^{\infty} A \exp[(-1/2Q + j\Omega_0)(t+nt_0)] \\ &= A \exp[(-1/2Q + j\Omega_0)t] \\ &\quad \times \frac{1}{1 - \exp[(-1/2Q + j\Omega_0)t_0]} \end{aligned}$$

したがつて分析出力のエンベロープ振幅は次のようになり, (2.3) 式に対応して (2.4) 式をうる。

$$|H_E| = \frac{A \exp[-t/2Q]}{\left[1 - 2\exp[-t_0/2Q] \cos \Omega_0 t_0 + \exp[-2t_0/2Q]\right]^{\frac{1}{2}}}$$

$$\left| \frac{[H_E(t)]_{max}}{[H_E(t)]_{min}} \right| = \frac{1 + \exp(-t_0/2Q)}{1 - \exp(-t_0/2Q)} \quad (2.4)$$

ここで $t_0/2Q$ は d/f_0 すなわち BPF の通過帯域中に含まれるピッチ高調波の数をあらわす。

結論的にいつて分析 BPF の中心周波数 F_0 とピッチ高調波との同調 ($F_0 = N f_0$), 非同調 ($F_0 = (N+k) f_0$, $\frac{1}{2} \geq k > 0$) の関係によつて出力は変動し, その変動の大きさは通過帯域内に含まれるピッチ高調波の数 (d/f_0) の減少関数である。

この有声音のピッチ周波数の変化による分析出力の変動をあらわす原理は, 定性的には他の多くの BPF 特性についても適用できる。ただしその変動の幅はその impulse response の real part すなわちエンベロープの減衰特性の関数であり, 個々の場合について計算しなければならぬ。

2.2. BPF による分析の周波数分解能⁽⁴⁾

声道の共振によるホルマントで特徴づけられる有声音波形は, 励振音源のピッチ周波数 $f_0 = 1/t_0$ cps を一定とすれば (2.5) 式で近似される。

$$\begin{aligned} f_V(t) &= \sum_{n=1}^{\infty} \left[\sum_{i=1}^r A_i \exp[p_i(t+nt_0)] \right] \\ &= \sum_{i=1}^r \frac{A_i}{1 - \exp(p_i t_0)} \exp(p_i t) \end{aligned} \quad (2.5)$$

$0 \leq t \leq t_0$

ここで A_i は i 番目のホルマントの振幅*, $p_i = -\delta_i + j\omega_i$ で, $f_i = \omega_i/2\pi$, $b_i = \delta_i/2\pi$ はそれぞれ i 番目のホルマントの周波数およびその帯域幅である。いま分析 BPF $U\left(\frac{x}{d}\right)$ の impulse response が $h_i(t) = R_e[h_i(t)] \exp(j\Omega_i t) = B_i \exp[-\delta_i(t)] \cdot \exp(j\Omega_i t)$ とあらわされるとすれば, (2.5) 式のような音声波入力に対する分析 BPF の出力は convolution integral から (2.6) 式のように計算される。

* 単一同調回路 ($R_e[h(t)] = A \cdot \exp(-\delta t)$), ガウス型 ($R_e[h(t)] = A \cdot \exp(-d^2 t^2)$) などがあるが, 矩形 BPF ($R_e[h(t)] = A \cdot \sin\left(\frac{bt}{t}\right)$) のようなものに対してはこのように簡単に論じられない。

* A_i は p_i ($i=1, \dots, r$) の複雑な関数である。

$$f_{01}(t) = \sum_{n=0}^{\infty} \left[\sum_{i=1}^r \frac{A_i}{1 - \exp(p_i t_0)} \int_{nt_0}^{nt_0+t} \exp[p_i(t + nt_0 - \tau)] B_i \cdot \exp[-\delta_i(\tau) + j\Omega_i \tau] d\tau \right] = \sum_{i=1}^r \frac{A_i \cdot B_i}{1 - \exp(p_i t_0)} \cdot \exp(p_i t) \sum_{n=0}^{\infty} \exp \int_{nt_0}^{nt_0+t} \exp[-\delta_i(\tau) + j\Delta\omega_i \tau] d\tau$$

$$\Delta\delta_i(\tau) = \delta_i(\tau) - \delta_i \tau$$

$$\Delta\omega_i = \Omega_i - \omega_i \tag{2.6}$$

実例として $U_i\left(\frac{x}{A}\right)$ の特性を単一同調回路とすれば、上式は次のように計算される。

$$f_{01}(t) = \sum_{i=1}^r \frac{A_i \cdot B_i}{1 - \exp(p_i t_0)} \exp(p_i t) \sum_{n=0}^{\infty} \exp(p_i n t_0) \int_{nt_0}^{nt_0+t} \exp[(-\Delta\delta + j\Delta\omega)\tau] d\tau$$

$$= \sum_{i=1}^r \frac{A_i}{1 - \exp(p_i t_0)} \cdot \frac{B_i}{1 - \exp(p_i t_0)} \cdot \frac{\exp[(-\Delta\delta + j\Delta\omega)t] - 1}{-\Delta\delta + j\Delta\omega} \cdot \exp(p_i t)$$

$$0 \leq t \leq t_0 \tag{2.7}$$

$$\Delta\delta = \delta_i - \delta_i$$

$$\Delta\omega = \Omega_i - \omega_i$$

ここで第1項は音声スペクトル自体のホルマント周波数とピッチ高調波との関係による変動を、第2項は分析BPFの中心周波数とピッチ高調波との関係による変動を、第3項は分析BPFによる周波数分解能を示す。

周波数分解能をあらわす第3項のエンベロープ部分を $\theta(t)$ としその $0 \sim t_0$ 秒間の平均値を計算すれば $(1/\delta_i \ll 1)$ とする、

$$\bar{\theta} \approx \frac{1}{\left(\frac{\delta_i}{\delta_i} - j\frac{\Delta\omega}{\delta_i}\right)} \left[\frac{e^{-\delta_i t_0}}{\delta_i} \cdot \frac{1 - e^{(-\Delta\delta + j\Delta\omega)t_0}}{t_0(\Delta\delta + j\Delta\omega)} \right]$$

$$= \frac{e^{-\delta_i t_0}}{\sqrt{\delta_i^2 + (\Delta\omega)^2}} \cdot \frac{(1 - e^{(-\Delta\delta + j\Delta\omega)t_0})}{\sqrt{(\Delta\delta)^2 + (\Delta\omega)^2}} e^{j\theta}$$

$$\theta = \varphi_1 + \varphi_2, \quad \tan \varphi_1 = -\frac{\Delta\omega}{\delta_i}, \quad \tan \varphi_2 = \frac{\Delta\omega}{\Delta\delta}$$

$$\therefore \bar{\theta}/\theta_0 = \frac{e^{-\delta_i t_0}}{\sqrt{\delta_i^2 + (\Delta\omega)^2}} \cdot \frac{(1 - 2e^{-\Delta\delta t_0} \cdot \cos \Delta\omega t_0 + e^{-2\Delta\delta t_0})^{\frac{1}{2}}}{\sqrt{\Delta\delta^2 + \Delta\omega^2}} \tag{2.8}$$

$\lim_{\Delta\omega \rightarrow 0} \bar{\theta}/\theta_0$ を $\bar{\theta}/\theta_0$ とすれば、

$$\bar{\theta}/\theta_0 \rightarrow \frac{e^{-\delta_i t_0} - e^{-\delta_i t_0}}{\delta_i \cdot \Delta\delta} = \frac{e^{-\delta_i t_0} (1 - e^{-\Delta\delta t_0})}{(\delta_i + \Delta\delta) \Delta\delta}$$

$\bar{\theta}/\theta_0$ で normalize した分解能を θ とすれば

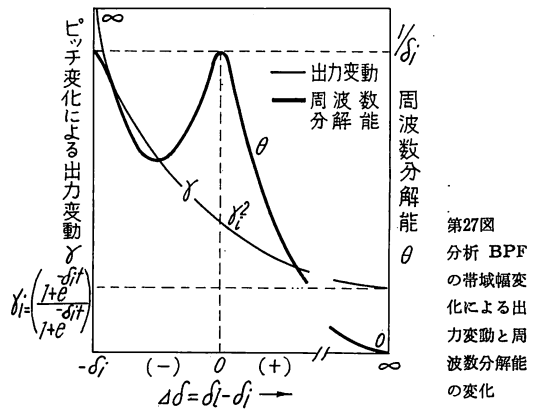
$$\theta = \frac{1}{\sqrt{1 + \left(\frac{\Delta\omega}{\delta_i}\right)^2}} \cdot \frac{1}{\sqrt{1 + \left(\frac{\Delta\omega}{\Delta\delta}\right)^2}} \times \frac{(1 - 2e^{-\Delta\delta t_0} \cdot \cos \Delta\omega t_0 + e^{-2\Delta\delta t_0})^{\frac{1}{2}}}{1 - e^{-\Delta\delta t_0}}$$

$$\approx \frac{1}{\sqrt{1 + \left[\frac{1}{\delta_i^2} + \frac{1}{\Delta\delta^2}\right] \Delta\omega^2}} \quad \Delta\omega \approx 0 \tag{2.9}$$

(2.9) 式が BPF による分析の周波数分解能をあらわすものと考えられる。

上述の考察の結果から、 $\delta_i \rightarrow 0$ すなわち完全 Fourier 分析のときと $\Delta\delta \rightarrow 0$ すなわち matched filter analysis のとき最大の周波数分解能が得られることがわかる。^{*}

(2.7) 式から分析BPFの帯域幅の変化による出力の変動と分解能の変化の様子を示すと第27図のようになる。



第27図 分析BPFの帯域幅変化による出力変動と周波数分解能の変化

この結果からみてわかるように、分析BPFの帯域幅は、ピッチ周波数の変化による出力変動と、周波数分解能とのからみ合いで $\delta_i = \delta_i \sim 2\delta_i$ の範囲が最適と考えられる。

以上の考察は主として単一同調回路による分析を例にとつて検討したが、BPFの特性があまりに sharp cut-off 特性でない限り、適当な等価帯域幅を考えることにより、上述の結果をほぼ適用することができるものと考えられる。

* 最大の周波数分解能はもとの (2.8) 式から $(\delta_i^2 + \Delta\omega^2)^{-\frac{1}{2}}$ に比例することもわかる。

3. 無声音特性⁽⁵⁾

無声音（主としてまざつ性の無声音）に対する分析フィルターの特性は、ある電力スペクトル密度をもつた random noise に対するスペクトル分析特性として検討することができる。

3.1. Random Noise のモデル

(1) 波形表示

$$N(t) = \sum_{n=1}^{\infty} \left(a_n \cos \frac{2\pi}{T}nt + b_n \sin \frac{2\pi}{T}nt \right)$$

(2) 振幅分布

$$p(a_n)da_n = (\pi E_n)^{-\frac{1}{2}} \exp[-a_n^2/E_n^2] da_n$$

E_n は mean square power

$$p(b_n)db_n = (\pi E_n)^{-\frac{1}{2}} \exp[-b_n^2/E_n^2] db_n$$

$P_n = a_n^2 + b_n^2$ の分布は

$$p(P_n)dP_n = E_n^{-1} \exp[-P_n/E_n] dP_n$$

(3) 電力スペクトル

電力スペクトル密度： $E(\omega)$ $E_n = \frac{2\pi}{T} E(\omega_n)$

$$\omega_n = n \cdot \frac{2\pi}{T}$$

3.2. スペクトル分解誤差の計算

分析BPFの特性を $U\left(\frac{x}{D}\right)$ とすれば、このBPFの分析出力の電力密度 $P(\omega_0)$ は、

$$P(\omega_0) = \frac{\sum_{n=1}^{\infty} P_n U^2\left(\frac{x_n}{D}\right)}{\int_{-\infty}^{\infty} U^2\left(\frac{x}{D}\right) dx}$$

ω_0 は分析BPFの中心周波数
 x_n は $\omega_n - \omega_0$

周波数 ω_0 における真の電力密度は $E(\omega_0)$ であるから分析誤差は次のようにならわされる。

$$P(\omega_0) - E(\omega_0) = \frac{\sum_{n=1}^{\infty} (P_n - E_n) U^2\left(\frac{x_n}{D}\right) - \psi}{\int_{-\infty}^{\infty} U^2\left(\frac{x}{D}\right) dx}$$

$$\psi = E(\omega_0) \int_{-\infty}^{\infty} U^2\left(\frac{x}{D}\right) dx - \sum_{n=1}^{\infty} E_n U^2\left(\frac{x_n}{D}\right)$$

上式から分析の mean square error を計算すれば、

$$\varepsilon^2 = \frac{\int_0^{\infty} \int_0^{\infty} \dots \int_0^{\infty} \left[\sum_{n=1}^{\infty} (P_n - E_n) U^2\left(\frac{x_n}{D}\right) - \psi \right]^2 \sum_{n=1}^{\infty} p(P_n) dP_n}{\left[\int_{-\infty}^{\infty} U^2\left(\frac{x}{D}\right) dx \right]^2}$$

$$= \frac{1}{I^2} \left\{ \sum_{n=1}^{\infty} E_n^2 U^4\left(\frac{x_n}{D}\right) + \psi^2 \right\} = \varepsilon_{\rho}^2 + \varepsilon_b^2 \quad (3.1)$$

ここで $I^2 = \left[\int_{-\infty}^{\infty} U^2\left(\frac{x}{D}\right) dx \right]^2$ であり、 $\varepsilon_{\rho}^2 = \frac{1}{I^2} \sum_{n=1}^{\infty} E_n^2 U^4\left(\frac{x_n}{D}\right)$ は statistical error であり、 $\varepsilon_b^2 = \psi^2/I^2$ は blurring error である。

Statistical error は \sum を f におきかえて次のように近似される。

$$\varepsilon_{\rho}^2 \approx \frac{\eta E^2(\omega_0)}{I^2} \int_{-\infty}^{\infty} U^4\left(\frac{x}{D}\right) dx$$

$$\eta = 2\pi/T \quad (3.2)$$

Blurring による error については $E(\omega)$ が有界で解析的であることから、平均値の定理を使い、 \sum を f で近似すると、

$$\psi \approx \left[\frac{dE}{d\omega} \right]_{\omega_0} \int_{-\infty}^{\infty} x \cdot U^2\left(\frac{x}{D}\right) dx + \frac{1}{2} \left[\frac{d^2E}{d\omega^2} \right]_{\omega_0} \times \int_{-\infty}^{\infty} x^2 U^2\left(\frac{x}{D}\right) dx$$

さらに $U^2\left(\frac{x}{D}\right)$ が ω_0 のまわりに対称であると仮定すると、

$$\psi \approx \frac{1}{2} \left[\frac{d^2E}{d\omega^2} \right]_{\omega_0} \int_{-\infty}^{\infty} x^2 U^2\left(\frac{x}{D}\right) dx$$

$$\text{したがって } \varepsilon_b^2 \approx \frac{2}{I^2} \left[\frac{d^2E}{d\omega^2} \right]_{\omega_0} \int_{-\infty}^{\infty} x^2 U^2\left(\frac{x}{D}\right) dx \quad (3.3)$$

3.3. Bandwidth の効果

上にのべた近似計算の結果、

$$\varepsilon^2 = \eta E^2(\omega_0) \frac{K}{I^2} + \frac{1}{2} E_{(2)}^2(\omega_0) \frac{J^2}{I^2}$$

$$= \eta E^2(\omega_0) \frac{K_1}{I_1^2} \frac{1}{D} + \frac{1}{2} E_{(2)}^2(\omega_0) \frac{J_1^2}{I_1^2} D^4 \quad (3.3)$$

ここで、 $E_{(2)}(\omega_0) = (d^2E/d\omega^2)_{\omega_0}$

$$I = \int_{-\infty}^{\infty} U^2\left(\frac{x}{D}\right) dx = \Delta I_1 \quad I_1 = \int_{-\infty}^{\infty} U^2(x) dx$$

$$J = \int_{-\infty}^{\infty} x^2 U^2\left(\frac{x}{D}\right) dx = D^3 J_1 \quad J_1 = \int_{-\infty}^{\infty} x^2 U^2(x) dx$$

$$K = \int_{-\infty}^{\infty} U^2\left(\frac{x}{A}\right) dx = \Delta K_1 \quad K_1 = \int_{-\infty}^{\infty} U^4(x) dx$$

この mean square error を最小にするような等価帯域幅 Δ を求めると,

$$\frac{d\varepsilon^2}{d\Delta} = 0 \quad \text{から} \quad \Delta^5 = \frac{1}{2} \cdot \frac{\eta E^2(\omega_0)}{E^{(2)}(\omega_1)} \cdot \frac{K_1}{J_1^2} \quad (3.5)$$

この最適等価帯域幅 Δ を (3.4) 式に代入すると,

$$\begin{aligned} \varepsilon^2 &= \frac{5}{2^{1.8}} \eta^{.8} E^{1.6}(\omega_0) E^{.4}(\omega_1) \frac{K_1^{.8} J_1^{.4}}{I_1^2} \\ &= K_\varepsilon E^{.8}(\omega_0) [\eta^2 E^{(2)}(\omega_0)]^{.2}; \quad K_\varepsilon = \frac{\sqrt{5}}{2^{.9}} \cdot \frac{K_1^{.4} J_1^{.2}}{I_1} \end{aligned} \quad (3.6)$$

この K_ε が各種 BPF の特性の無声音分析の merit を量的に与える measure となり, K_ε の小さいほどよい特性であるということが出来る。

3.4. 各種 BPF 特性の比較

S.S.L. Chang の計算⁽⁵⁾によると K_ε を最小にするような最適の BPF 特性 $U\left(\frac{x}{A}\right)$ の第 1 近似値として次のような特性が求められている。

$$\left[U^2\left(\frac{x}{A}\right) \right]_{opt} = \frac{A}{\left[1 + \left(\frac{x}{A}\right)^2 \right] \left[1 + \left(\frac{x}{A}\right)^4 \right]} \quad (3.7)$$

解析的に計算できるような BPF 特性について K_ε の値を求めそれらを比較すると第 1 表のようになる。

第 1 表 各種 BPF の無声音特性の比較

BPF の型	K_ε
同調回路 1 段	∞
同調回路 2 段 (結合なし)	.77
同調回路 3 段 (結合なし)	.71
矩形型フィルター	.68
最適フィルターの一近似	.68
ガウス型フィルター	.67
最適フィルター	.66

無声音 (random noise で近似) の分析においては, 単一同調回路 1 段では通過帯域のすその切れ方が緩やかすぎて J_1 の積分が無限大となり blurring による誤差が過大となる。

4. 振幅過渡特性

BPF の過渡特性についてはすでによく知られているように次のような検討が行なわれている。BPF の特性を $U(x/A)$, その等価低周波特性を $b(\omega)$, impulse

response を $h(t)$, その amplitude envelope を $f(t)$ とすると,

time response spread の mean square value : Δt^2

$$(\Delta t)^2 = \frac{\int_{-\infty}^{\infty} t^2 |f(t)|^2 dt}{\int_{-\infty}^{\infty} |f(t)|^2 dt} \quad (4.1)$$

amplitude response spread の mean square value : $\Delta \omega^2$

$$(\Delta \omega)^2 = \frac{\int_{-\infty}^{\infty} \omega^2 |b(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |b(\omega)|^2 d\omega} \quad (4.2)$$

過渡特性について最適なフィルターの特性は,

$(\Delta t)^2 + K(\Delta \omega)^2 = \min.$ (K : 常数), または $(\Delta t) \cdot (\Delta \omega) = \min.$ のいずれかの基準で選ばれる。そしてこの二つの判定基準のいずれによつても最適な BPF 特性は Gauss 型であることがすでに Gabor および Wiener によつて示されている。

解析的に計算できるような BPF 特性について Δt , $\Delta \omega$ の値を計算して比較すると第 2 表のようになる。

第 2 表 各種 BPF の過渡特性の比較

BPF の型	Δt	$\Delta \omega$	$\Delta t \cdot \Delta \omega$
同調回路 1 段	$\sqrt{2} Q$	∞	∞
同調回路 2 段 (結合なし)	$2\sqrt{3} Q$	$\frac{1}{2} \cdot \frac{1}{Q}$	$\sqrt{3}$
同調回路 3 段 (結合なし)	$\sqrt{30} Q$	$\frac{1}{2\sqrt{3}} \cdot \frac{1}{Q}$	$\sqrt{\frac{5}{2}}$
矩形型フィルター	∞	$\frac{1}{3} b^2$	∞
ガウス型フィルター	$\frac{1}{\sqrt{2}} \frac{1}{4\sqrt{\sigma}}$	$\frac{1}{\sqrt{3}} 4\sqrt{\sigma}$	$\frac{1}{2}$

第 2 表からわかるように単一同調回路 1 段では周波数分解能が不足であり, 矩形型フィルターでは時間分解能が不足である。また単一同調回路 2~3 段に対して Gauss 型フィルターでは約倍の周波数時間分解能がある。

5. 周波数過渡特性⁽⁵⁾

音声スペクトルの分析において最も重要な点の一つは, 分析対象のスペクトルが定常的で時間的に不変なものではなくて, 準定常的または過渡的で時間とともに変化するものでありしかもその変化の仕方そのものが重要な情報であるということである。さきにも述べた振幅過渡特性の検討もこのような観点からするものであつた。

スペクトル特性の時間的変化のなかで最も重要なものは, 音源の性質の変化すなわち有声音と無声音の区別と

有声音スペクトルのホルマント周波数の変化である。

前者の情報はスペクトル分析以外の方法たとえばピッチの検出などによつても抽出することができると考えられるので、ここではホルマント周波数の時間的な変化に対する分析出力すなわち周波数的な過渡特性について検討する。

5.1. 周波数過渡特性の計算

周波数が時間的に変化するような入力信号のモデルとして次のような波形（直線の変化）を考える。

$$f_i(t) = A \exp\left[j\left(\omega_1 t + \frac{1}{2} q t^2 + \phi_1\right)\right] = A \exp[j\psi] \quad (5.1)$$

ここで瞬時周波数 ω_i は $\frac{d\phi}{dt} = \omega_1 + q t$ であり、 ϕ_1 は初期位相、 q は周波数変化率 (radian/sec) である。

分析 BPF の特性は、 $\Delta/\Omega_0 \ll 1$ すなわち narrow band filter の場合、その impulse response で次のように近似される。

$$h(t) = \exp[-\Delta_0^2(t-\tau_0)^2 + j\Omega_0 t] \quad (5.2)$$

ここで τ_0 は impulse が加えられてから出力が最大となるまでの時間であり、 $1/\Delta_0$ はその envelope の等価継続時間である。

(5.2) 式のような impulse response をもつ分析 BPF の (5.1) 式のような入力に対する出力は次のように計算される。

$$\begin{aligned} f_o(t) &= \int_{-\infty}^t f_i(\tau) h(t-\tau) d\tau \\ &= \int_{-\infty}^t A \exp\left[j\left(\omega_1 \tau + \frac{1}{2} q \tau^2 + \phi_1\right)\right] \\ &\quad \cdot \exp[-\Delta_0^2(t-\tau_0-\tau)^2 + j\Omega_0(t-\tau)] d\tau \\ &\approx A \exp\left[j\left\{(\omega_1 - q\tau_0)t + \frac{1}{2} q t^2\right\}\right] \exp\left[j(\Delta\omega\tau_0\right. \\ &\quad \left.+ \frac{1}{2} q \tau_0^2 + \phi_1)\right] \int_{-\infty}^{\infty} \exp\left[-\left(\Delta_0^2 + j\frac{q}{2}\right)\theta^2\right] \\ &\quad \cdot \exp[j(\omega + q(t-\tau_0))\theta] d\theta \\ &= \frac{A}{\sqrt{2}} \exp\left[j(\omega_1 - q\tau_0)t + \frac{1}{2} q t^2\right] \\ &\quad \cdot \exp\left[j\left(\frac{1}{2} q \tau_0^2 + \Delta\omega\tau_0 + \phi_1'\right)\right] \frac{1}{\Delta_0 \sqrt{1 + \left(\frac{q}{2\Delta_0^2}\right)^2}} \\ &\quad \cdot \exp\left[-\frac{(\Delta\omega - q\tau_0)^2 + \{q t^2 + 2q(\Delta\omega - \tau_0)t\}}{4\Delta_0^2 \left(1 + \left(\frac{q}{2\Delta_0^2}\right)^2\right)}\right] \end{aligned} \quad (5.3)$$

上式で $q=0$ とおけば純正弦波に対する出力を与える。

$$\begin{aligned} \lim_{q \rightarrow 0} f_o(t) &= \frac{A}{\sqrt{2}} \exp[j\omega_1 t] \exp[j\Delta\omega\tau_0 + \phi_1] \\ &\quad \times \frac{1}{\Delta_0} \exp\left[-\frac{(\Delta\omega)^2}{4\Delta_0^2}\right] \end{aligned} \quad (5.4)$$

(5.3) 式と (5.4) 式をくらべてわかることは、

(1) 出力の envelope amplitude が $(1+q^2/4\Delta_0^4)^{\frac{1}{2}}$ 分の 1 となる。

(2) 等価帯域幅が $(1+q^2/4\Delta_0^4)^{\frac{1}{2}}$ 倍となる。

(3) 周波数が $q\tau_0 \approx q/\Delta_0$ radian per sec ずれる。ということである。

いずれにしても周波数の変化による出力の変化が少ないためにはこの BPF の等価帯域幅の 2 乗が q にくらべて大きいことが必要である。

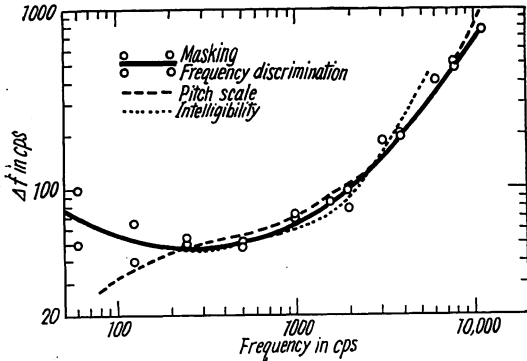
例えばホルマントの周波数が 10m. sec で最大 1kc 程度移行する（例えば \dot{y} 音部の第 2 ホルマント周波数）とすれば $q = 10^3 \times 2\pi / 0.1 = 2\pi \times 10$ radian/sec であり、したがって分析 BPF の等価帯域幅が 100cps 程度であれば出力の変化は 3db 以下と考えてよい。

また BPF 特性の差については Gauss 型が最もすぐれ、 $q \ll 14$ (-3db bandwidth)² といわれており、矩形型フィルターに対して 8 倍の分解能をもっているといわれている。

6. 周波数配分 (frequency scale) と帯域分割

第 2 部において音声スペクトル分析の立場から BPF 個々の特性について検討を加えてきたが、そのうえさらに BPF 群全体として周波数配分と帯域分割について検討しなければならぬ。このようなスペクトル分析装置全体としての特性は、個々の BPF の特性よりはるかに強くその使用目的によつて想定されるものであり、一般的に論ずることは困難になる。音声の音韻としての認識に対する周波数帯域の貢献度は、明瞭度指数のようなマクロスコピックな平均的な定量化のデータから帯域明瞭度指数として示されており、音声の伝送を目的とした vocoder などの周波数配分と帯域分割に対して一つの指標を与えている。しかし音声の識別を目的とするような立場ではもつとミクロスコピックな考慮が必要であろう。

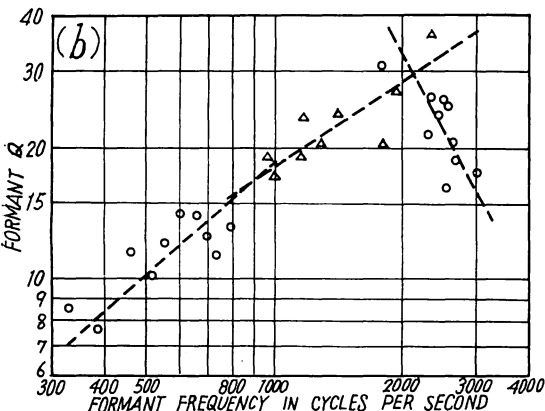
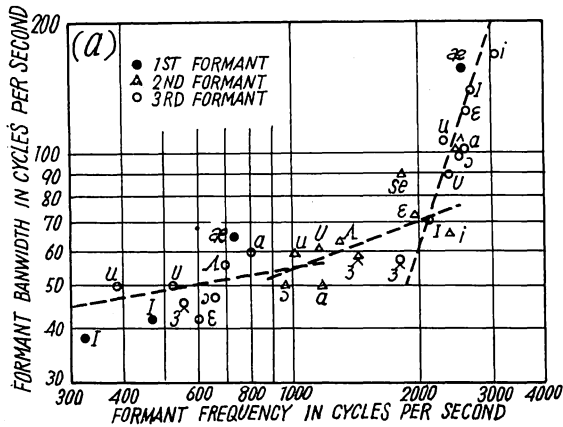
音声識別の場合には聴覚機能との比較なしにこの問題を考えることはできない。聴覚機能との比較でまず考えられる対応は critical bandwidth である。すでに第 1 部で論じたように critical bandwidth とは音響しげきの perception において energy 的な積分作用が行なわれると考えられる帯域幅であり、したがってスペクト



第28図 Critical band function (○—)

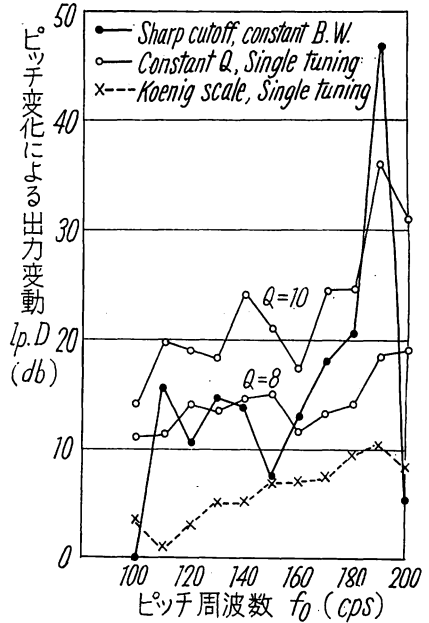
ル分析における帯域分割に重要な関係があると考えられる。その特性は第28図に示すようである。

一方入力としての音声自身の特性から考えると、母音、よう音、鼻音などは約3~4kc以下の帯域でじゆうぶんにその音韻情報の伝送が可能であるが、まさつ音、はさ



第29図 母音のホルマント帯域幅特性

- (a) ホルマント帯域幅特性
- (b) Qの値になおした特性



- Sharp cut-off, constant bandwidth : 23 channels
(F₁ 領域で 100cps, F₂ 領域で 150cps) (隣接チャネルの中心周波数で 20db 減衰)
- Constant Q, single tuning
Q=10 : 25 channels
Q= 8 : 20 channels
- Koenig Scale, Single tuning : 19 channels
1 kc 以下 constant bandwidth : 100cps
1 kc 以上 constant Q, Q= 10

第30図 ピッチ周波数の変化によるBPF群出力の変動 (200 cps ~ 3500 cps)

- l_p : local peaks の数
- D : 平均出力からの標準偏差

つ音、はれつ音のような子音の明瞭な伝送には3~7kcの高周波帯域が必要である。したがって低、中周波域においては母音、よう音、鼻音などのような明瞭なホルマント構造をもつた有声音の分析が主となり、その分析帯域幅のとり方に対して matched filter analysis の結果が一つの指標を与える。これに関連して母音のホルマント帯域幅特性⁽⁹⁾を示すと第29図のようである。

一方子音の分析を主目的とする高周波域では無声子音分析特性と過渡特性が問題になる。またまさつ音のような継続時間の比較的長い無声子音における子音ホルマントの帯域幅は分析、合成の実験結果⁽¹⁰⁾からQが5~10程度とされている。

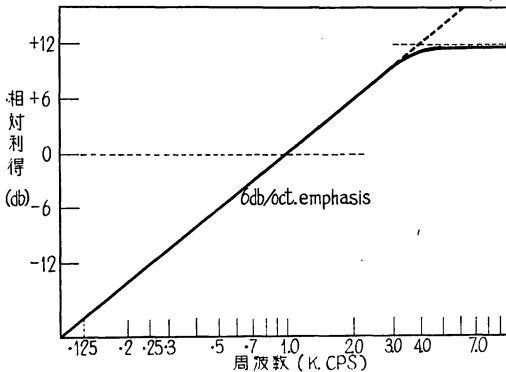
有声音スペクトルの分析においてはピッチ周波数の変化による出力の変動が問題になるが、分析BPF群としては個々のBPFの特性とその周波数配分に関し解析的に求めることは困難なので、2,3の実例について computer simulation によつて検討した。その結果の1例を第30図に示す。ここではBPF群としての出力変

動の measure として、各分析チャネルの出方の平均レベルからの標準偏差と出力の凹凸の個数をあらかず極大値 (local peaks) の数との積をとった。この結果から、単一同調回路のような cut-off 特性の緩いフィルターでその周波数配分が聴覚での周波数軸に近似の Koeing scale のものがよいということがわかった。

7. Pre-processing の特性

7.1. Frequency emphasis の特性

平均(標準)的にいつて有声音源のスペクトル・エンベロープは $-12\text{db/oct.}(1/f^2)$ の減衰特性をもつ。唇の放射特性は第1近似として $6\text{db/oct.}(f)$ の強勢特性をもつ。したがって音声スペクトルの一次的な情報源と考えられる声道のホルマント共振特性を至少なく抽出するためには平均的にいつて低、中周波域においては 6db/oct の frequency emphasis が必要である。しかし乱流によつて生ずる無声子音の音源スペクトル特性は $3\sim 7\text{kc}$ の範囲ではほぼ平坦と考えてよく、また分析出力の S/N 比を劣下させないためにも高周波域では emphasis は必要ないと考えられる。ただ弱い無声子音のスペクトルを検出するために全体として高利得であることは必要と考えられるので総合的な利得特性として第31図のようであればよい。



第31図 Pre-emphasis の特性 (総合特性)

7.2. Amplitude compress の特性

音声波形の dynamic range は非常に広く 60db (2進10ビット) 以上である。このような広い変化範囲をそのまま忠実に記録することは一般に困難でありまたある意味では不必要でもある。従来は分析 BPF に入る前にアナログ的に amplitude compressor が用いられてきた。しかし分析出力の処理を計算機でデジタル的に正確に行なうのに、その入力アナログ的な不確かさを

つているのは望ましいことではない。そこで PCM 伝送において開発されてきた non-linear coding の技術を A-D 変換にとり入れて、例えばデジタル的に対数変換して6ビットなり7ビットなりを有効に使うということが考えられる。

7.3. 検波後の LPF 特性

この点については特に検討を加えなかつたが、過渡特性について考えるとやはり Gauss 型に近いものが望ましい。またその cut-off 周波数については必要なまた利用しうる情報量との関係で考慮すべきであるが、vocoder による実験例、A-D 変換器の性能などの点からいつて $50\text{cps}\sim 100\text{cps}$ 程度が適当であろう。

8. 結 言

以上の検討結果を総合して次のような特性が一つのモデルとして考えられる。

- (1) 低、中周波域特性： $0.1\text{kc}\sim 3\text{kc}$ 程度、有声音の分析を主目的とする。
 - (1.a) 低周波域： $0.1\text{kc}\sim 1\text{kc}$ 程度 (主として第1ホルマントの周波数域)
 - BPF特性：単一同調回路1～2段程度
 - 帯域分割：等価帯域幅 $50\sim 100\text{cps}$ で直線の周波数配分 (10～20チャンネル)
 - (1.b) 中周波域： $1\text{kc}\sim 3\text{kc}$ 程度 (主として第2, 第3ホルマントの周波数域)
 - BPF特性：単一同調回路1～2段程度
 - 帯域分割：等価 Q の値で10～20で対数的周波数配分 (10～20チャンネル)
- (2) 高周波域特性： $3\text{kc}\sim 7\text{kc}$ 程度、無声子音の分析を主目的とする
 - BPF特性：近似的な Gauss 型とする
 - 帯域分割：等価 Q の値で10～6程度で対数的周波数配分 (5～10チャンネル)
- (3) その他の特性
 - (3.a) frequency emphasis 特性： $0.1\sim 3.0\text{kc}$ の間で 6db/oct. emphasis (総合特性)
 - (3.b) amplitude compressor 特性：non-linear A-D 変換による6～7ビットの対数変換 ($60\sim 100\text{db}$)
 - (3.c) 検波後の LPF 特性：近似的な Gauss 型とし、等価帯域幅 50cps とする。

第2部主要参考文献

- (1) M. V. Mathews, et al: Pitch synchronous analysis of voiced sounds, JASA, 33, 2, 179~186, (1961).
- (2) J. Capon: High-speed Fourier analysis with recirculating delay-line heterodyner feedback loops, IRE. Trans., 1-10, 32~37, (1961).
- (3) G. Fant and J. Liljencrants: How to define formant level. A study of the mathematical model of voiced sounds, STL-QPSR-2/1962, 1~9, (1962).
- (4) C. P. Smith: Speech data reduction, Appendix A some aspects of voice spectrum analyzer-synthesizer design, ASTIA Document No. AD 117290, (1957).
- (5) S. S. L. Chang: On the filter problem of the power-spectrum analyzer, IRE., 42, 8, 1278~1282, (1954).
- (6) J. Klapper and C. M. Harris: On the response and approximation of Gaussian filters IRE. Trans., AU-7, 80~87, (1959).
- (7) A. Gabor: Theory of communication, JIEE., Part III, (1946).
- (8) N. Wiener: Extrapolation, interpolation and smoothing of stationary time series, John Wiley & Sons Inc., (1949).
- (9) H. K. Dunn: Methods of measuring vowel formant bandwidth, JASA, 33, 12, 1737~1746, (1961).
- (10) J. L. Flanagan: Perceptual criteria in speech processing, Speech Communication Seminar, Stockholm, (1962).

結 論

以上音声識別研究用計算機入力装置の検討として二つの可能性、聴覚機能のアナログ・シミュレータとBPF群による分析について検討したが、いまのところいずれがより能率的でしかも有効な音声情報の抽出装置であるかを断定することはできない。前者は現在のところ未知の部分も多く、簡単に直に実験装置を作るということはできないが、生体の情報処理機能を工学的な目的に利用してゆくという点で“Bionics”の研究とも関連する問題であり、新しい将来の研究問題としてじゆうぶんに魅力的であり、また成果を期待しうる研究と考えられる。

最後にこの調査の機会と発表の便を与えていただいた河野次長、尾方室長に感謝するとともに、日立製作所中央研究所および東京電機大学の関係各位の御理解と御協力に深謝する次第です。

