

論文 / 著書情報
Article / Book Information

論題(和文)	音声現象におけるゆらぎ
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	第2回コーワシンポジウム'85「ゆらぎと科学技術 - その現在と未来 - 」, , , pp. 35-47
Citation(English)	, , , pp. 35-47
発行日 / Pub. date	1985,

音声現象におけるやらぎ

古井 貞熙

1. はじめに-音声の生成過程

人間が音声言語を生成するプロセスの第1段階は、相手に伝えたい内容を決め、それを表現する語句を選択し、文法の規則に合うように並べることにより、言語的形式に変えることである。これができると、脳から発声器官へ運動神経指令が発せられ、発声器官の種々の筋肉が動いて、空気振動としての音波が作り出される。このプロセスは、神経や筋肉の活動する生理学的段階と、音波の発声と伝搬が行われる物理的段階に分けることができる。情報伝達のための符号体系としての言語は離散的な単位をもって構成された形式を有しているが、物理現象としての音声の性質は連続的である。

人間の発声器官の構造は図1に示すように、肺、気管、喉頭、咽頭、口腔、鼻腔などからなっており、これらは全体として1つの連続した管をなしている。喉頭より上の部分は声道とよばれ、顎、舌、口唇等を動かすことにより、種々の形に変化する。通常の呼吸のときは、声門すなわち左右の声帯の間隙は大きく開いているが、声を出そうとすると声帯が接近し、この間を肺からの空気が通り抜けようとするため、空気流と声帯の相互作用により、声帯が周期的に開いたり閉じたりして、ほぼ規則的な空気の断続が生ずる。これが音声の音源となる。声帯の緊張が大きく、かつ肺からの空気圧が高いと、声帯の開閉周期（基本周期）が短くなって、音源の音の高さすなわちピッチが高くなり、逆のときは低くなる。基本周期の逆数を基本周波数と呼ぶ。基本波と高調波成分からなる音源波は、声道を通るときにその形に対応した音色を付与され、/a/や/o/の音声となる。このようにして母音の音声生成される。これに、基本周期と振幅の時間的变化により、アクセントやイントネーションが付与される。

摩擦音や破裂音のような子音は、舌などに

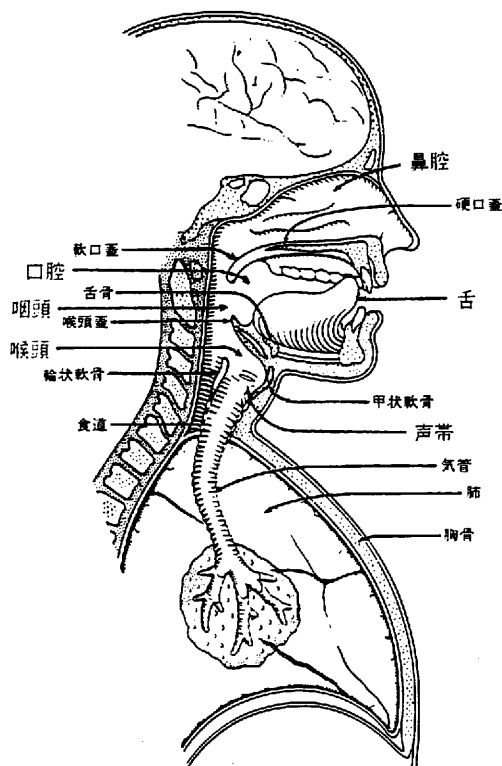


図1. 人間の発声器官

よって声道のある部分に狭い場所を作り、そこを空気流が通り抜けるときに乱流を生じたり、声道の一部を遮断することによって空気流を一時的に止め、圧力が十分高まったところで、これを急に開放してインパルス的な音を生ずることによって生成する。子音には、この他に破擦音、半母音、鼻音などがあるが、いずれも過渡的な性質を有する特徴がある。過渡部におけるスペクトルの遷移速度は発声器官の運動速度に制約されている。子音の生成は声帯振動の有無とは独立して行われ、声帯振動を伴う子音を有聲子音、伴わない子音を無声子音とよぶ。母音と子音をあわせて音素とよぶ[1][2]。

2. 音声波形とスペクトル

図2に、男性が「朝鮮南部に」と発声したと

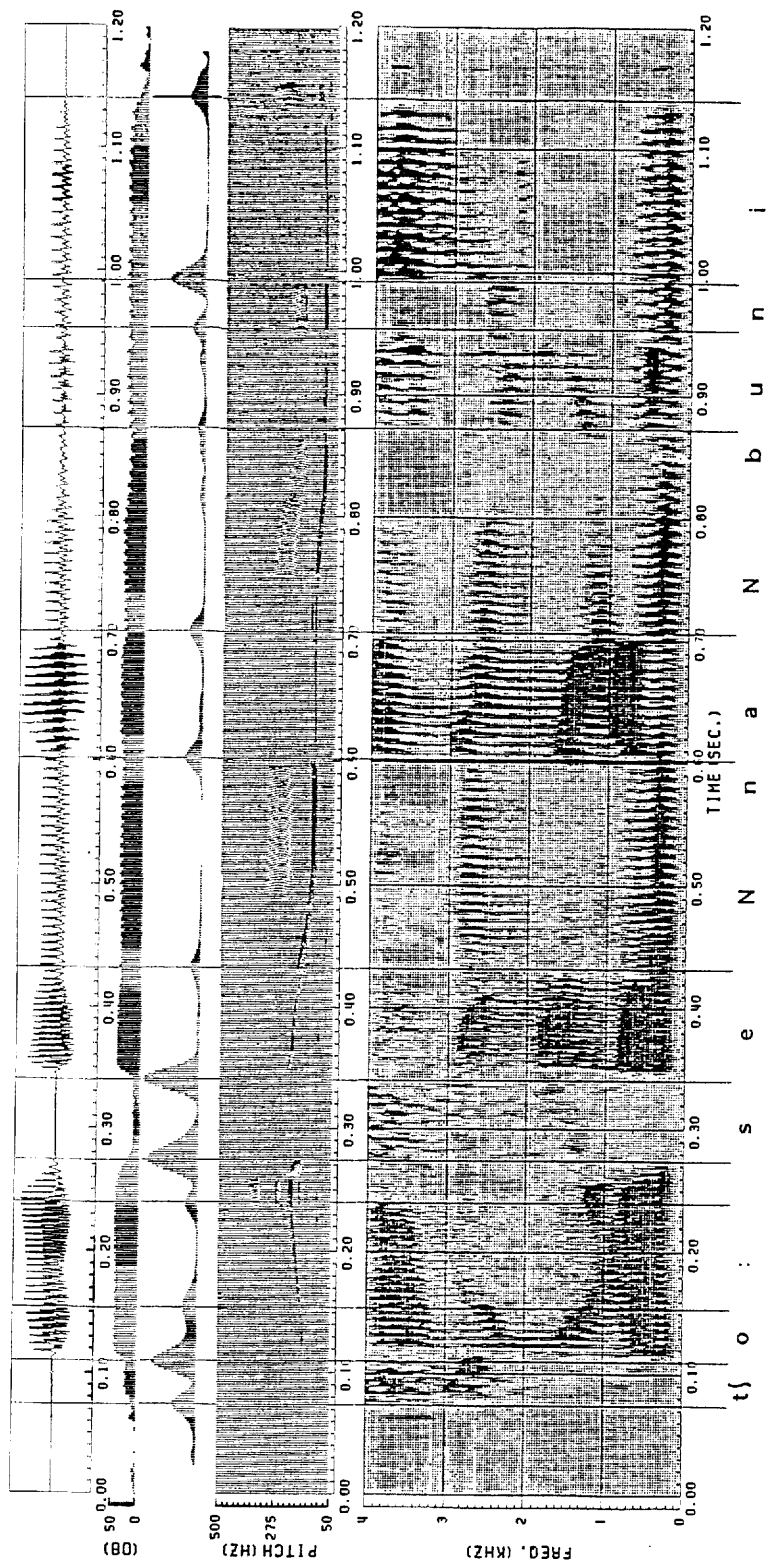


図2.「朝鮮南部に」と発声したときの波形、短時間当りのスペクトル変化量、基本周波数、およびスペクトルの時間的变化（上から）

きの、波形、短時間パワー、単位時間当りのスペクトル変化量、基本周波数（ピッチ）、およびスペクトルの時間的変化の様子を示す。この図からわかるように、スペクトルの変化量は各音素の境界付近で局所的に最大となるが、スペクトルは母音の区間でも常に変化しており、非定常プロセスと見做される。ただし、20～30ms 程度の適当な区間に分ければ、それぞれの区間内ではほぼ一定の性質をもつと見做すこともできる。無声子音 /p/ /t/ /k/ /s/ などに対応する区間では、音声波形は小振幅で不規則な波形を示し、スペクトルも不規則な模様を示す。一方 /i/ /e/ /a/ /o/ /u/ などの母音区間や /n/ /N/ などの有声子音区間では、音声波形は大振幅でほぼ周期的な波形を示し、スペクトルも比較的大きな濃淡模様の繰り返しを示す。図2のスペクトルの縦縞は基本周期に対応する。

波形の振幅のダイナミックレンジはかなり大きく、50dB 程度以上ある。音声振幅レベルの密度分布は図3に示すとおりで、母音群に対応する大振幅部分はほぼ指数分布で、子音群に対応する小振幅部分はほぼ正規分布で近似することができる。

音声の周波数スペクトルの長時間平均は図4に示す通りで、基本周波数が関連する低い周波数域を除いて、男女の差や国語による差はほとんどない。音声スペクトルの平均的な形を表現するときには、この結果をもとに、800Hz まではほぼ平坦で、800Hz 以上はオクターブ当たり -10dB の傾斜を持つスペクトルで近似することが多い。ただし、このような周波数間隔の粗い global な長時間平均スペクトルはほとんど個人差がないが、さらに細かい周波数分解能で求めると、個人差が抽出されることが確認されている。(6. 参照)

音声を知覚する上で重要な周波数成分は 4～5 kHz から下の帯域にあり、音声をデジタル化する場合、少なくとも 8～10kHz の標本化周波数 (100～125 μs の間隔) で標本化する必要がある。

3. 最尤スペクトル推定法

音声波形に関して、次の仮定をおくことがで

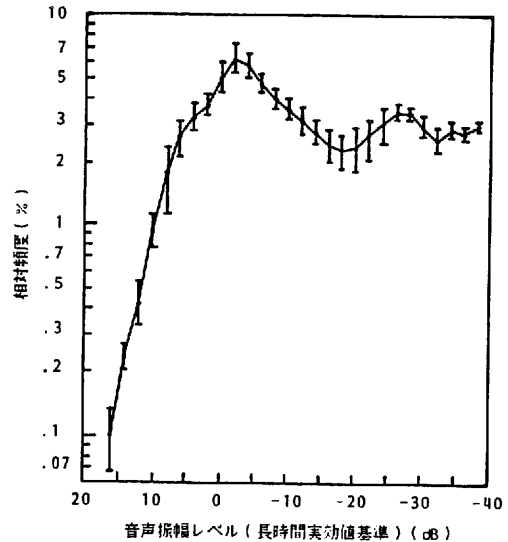


図3. 会話 (文章) 音声の振幅レベルの密度分布 (女性5名の平均と標準偏差)

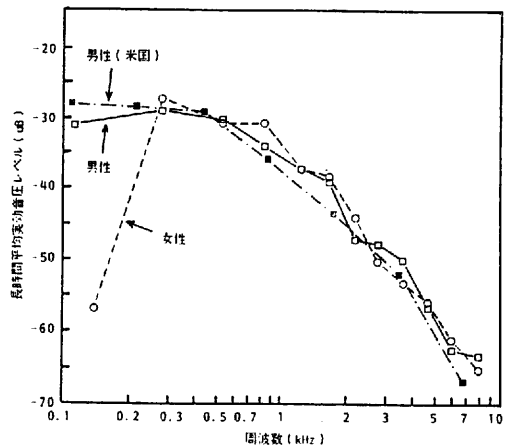


図4. 音声の長時間平均スペクトル

きる[3][4]。

- (1) 標本値 x_i は、あるスペクトル密度 $f(\lambda)$ をもつ定常ガウス過程からの標本とみなせる。(λ : 規準化角周波数, $-\pi \leq \lambda \leq \pi$)
- (2) スペクトル密度 $f(\lambda)$ は零点をもたない全極形の有理スペクトル密度で表わされ、次の関数形をもつ。

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{\left| \prod_{i=1}^p \left(1 - \frac{z_i}{z_i^*} \right) \right|^2}$$

$$= \frac{\sigma^2}{2\pi} \frac{1}{\left| 1 + \sum_{i=1}^p \alpha_i z^{-i} \right|^2}$$

$$= \frac{\sigma^2}{2\pi} \frac{1}{A_0 + 2 \sum_{i=1}^p A_i \cos i\lambda} \quad (1)$$

ただし, z_i は $1 + \sum_{i=1}^p \alpha_i z^{-i} = 0$ の根であり,

$$A_i = \sum_{j=0}^{p-|i|} \alpha_j \alpha_{j+|i|}, \quad \alpha_0 = 1 \quad i=0, \pm 1, \dots, \pm p$$

(2)

である。 σ^2 はスペクトル密度の強さのスケールファクタ, p は実際のスペクトル密度を近似するために必要な極の数 (一対の共役極は 2 個と数える) である。 α_i を線形予測係数と呼ぶ。

(1) の仮定は, 無声子音については抵抗なく受け入れられるが, ピッチ構造を有する母音や有聲子音では必ずしも納得できない。しかし実際の音声では, 声帯音源はゆっくりした変化と細かいゆらぎを持っており, 高調波はそれに応じて広がっているので, スペクトル全体としては許容できる仮定である。スペクトル密度に零点を入れないのは, 人間の聴覚が極に敏感で, 零点でしか表現できないようなスペクトルの急峻な谷には鈍感であることと, この方が数理的な取り扱いが容易になることによる。

このとき N 個の標本 $x = (x_0, x_1, \dots, x_{N-1})$ から計算したはじめの p 単位時間の遅延における短時間自己相関関数を用いた連立一次方程式により, σ^2 と $\{\alpha_i\}_{i=1}^p$ の各パラメータを最尤推定することができる。このようにして求まるパラメータを用いて表現されるスペクトル $f(\lambda)$ は, 元のスペクトル $f(\lambda)$ の山形部分を重視して, 適合誤差が最小になるように推定したスペクトル包絡になっている。図 5 に, 元の音声の短時間スペクトルと最尤法による推定スペクトル包絡の例を示す。音声の周波数帯域が 0 ~ 4 kHz 程度であれば, p を 10 程度にすれば短時間スペクトル $f(\lambda)$ の顕著な山形部分は $f(\lambda)$ にほぼ完全に反映される [5]。

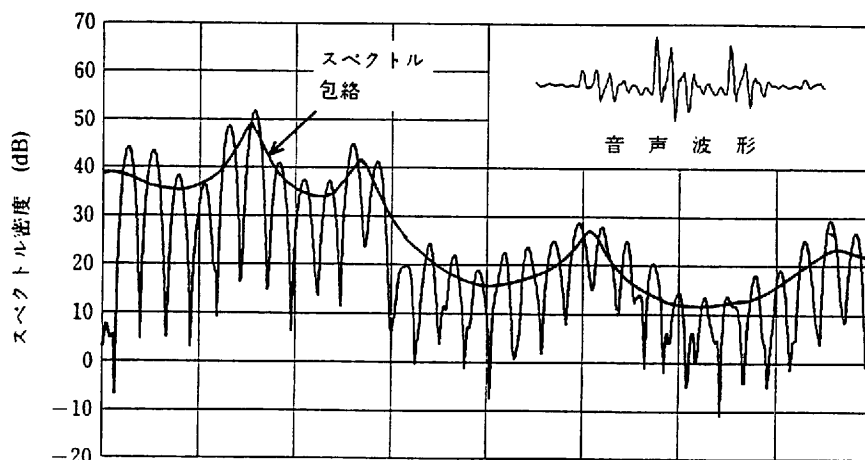
入力信号の短時間スペクトル $\hat{f}(\lambda)$ をスペクトル包絡 $f(\lambda)$ で正規化することによって得られるスペクトルの微細構造は, 大域的には周波数軸に対してほぼ平坦で, 入力信号に周期性があれば, 強さがほぼ均一の高調波構造を有する。従って $\hat{f}(\lambda)/f(\lambda)$ をフーリエ展開してできる相関関数 (変形相関関数) は, 有声音の場合は基本周期の整数倍の遅延で大きい相関を示し, 無声音の場合は顕著な相関を示さない。変形相関関数のこの性質を利用すれば, 入力信号のスペクトル包絡の形に無関係に, 音源の周期性の有無と基本周期を抽出することとができる。

こうして抽出した音源情報および σ^2 でパルスと白雑音の発生器を制御し, この音源をスペクトル包絡情報すなわち線形予測係数 $\{\alpha_i\}_{i=1}^p$ で制御した音声合成回路に通すことにより, 元の音声を再生することができる。一般的な音声分析合成系の構成を図 6 に示す。これらの各制御パラメータは, 通常 5 ~ 20ms 程度の時間間隔で更新すればよい。このようにして再生される音声波は, 元の音声のスペクトルの性質は保存しているが, 位相情報は取り除かれている。これが再生された音声の品質に劣化をもたらさないのは, 人間の聴覚が位相情報に対して鈍感であるためである。

4. 音声知覚におけるスペクトル変化情報の役割

合成音を用いた子音のスペクトルに関する実験によれば, 子音部から母音部への遷移におけるホルマント (エネルギーの強い周波数成分) の動きによって, 各子音の特徴が表現できるとされている。ただし, このホルマント遷移は, 着目している子音の前後の母音などが変わると, 調音結合 (音素から音素へ物理特徴が相互に影響を及ぼしあいながら, なめらかに変化する現象) のために大きく変化する。母音に関しても, 連続音声中では定常部分をほとんど持たずに, 隣接の音素へ向かってスペクトルが変化するため, その母音を単独に発声したときの値には到達しない。この際に, スペクトルの動特性からその目標値としての母音が知覚される一種の予測あるいは補償作用に関して, いくつかの研究が行われている [6]。

(a) スペクトル



(b) スペクトル
微細構造

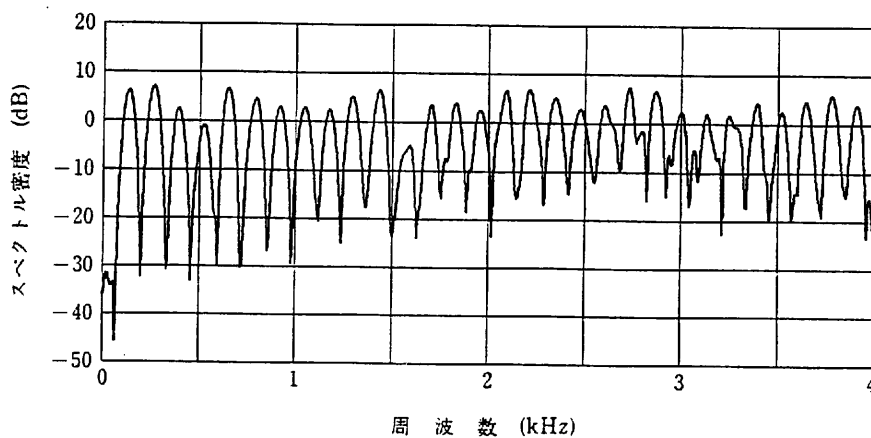


図5. 男性母音/a/のスペクトル, スペクトル包絡, およびスペクトル微細構造

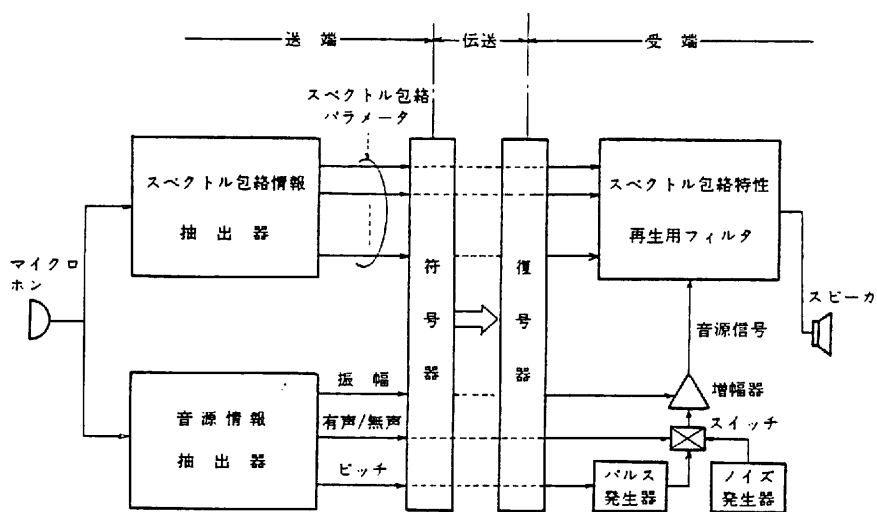


図6. 音声分析合成系の構成

生理学的には、聴覚神経系において、下位の神経系では持続形ニューロンがほとんどであるのに対し、上位の神経系では刺激の過渡的性質に対して応答するニューロンが多くなり、大脳皮質や内側膝状体では持続形ニューロンはほとんど見られなくなる。持続形ニューロンでも、刺激の冒頭ではパルス周波数が高いが時間とともにパルス数が減少する順応形ともいえるものが多い。これらはニューロン単体の性質ではなく、ニューロン間の相互接合や抑制作用による。猫の神経系においては、周波数変化の方向性（周波数上昇、下降）に応じるニューロンも見出されている[7]。FM音を用いた人間の聴覚のマスティング心理実験によれば、周波数が変化する時点でマスティングパターンが大きなオーバーシュートを示し、FM音に対する聴感度が高いことが確認されている。

図7に示すように、日本語単音節の話頭（先頭部）または話尾（終端部）を種々の時点まで切断除去した音を作成し、この明瞭度試験を行って、この結果を音声スペクトルの各時点での動的特徴と対照することによって、スペクトルの動的特徴の役割を調べる実験が行われている。その結果、図8に／njo／の場合について

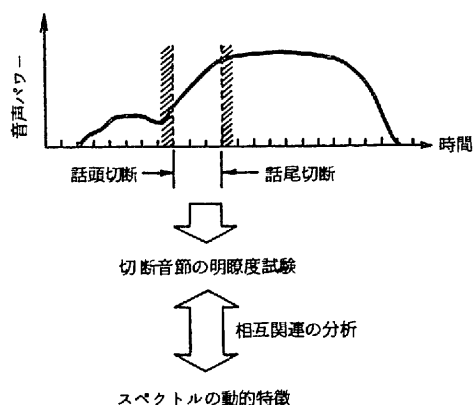


図7. 話頭／話尾切断した日本語単音節の明瞭度試験によるスペクトルの動的特徴の聴覚的役割の分析

示すように、話頭／話尾切断音節の明瞭度は、切断位置がスペクトル変化極大時点の前か後かで急激に変化し、この時点を含む10ms程度の区間が、最も重要な音韻性情報を担っていることが明らかになった。また図9に示すように、多次元空間内でのスペクトルの動特性を調べた結果によれば、後続音素ターゲットへの運動方向が定まり、その方向への変化速度が最大になった時点で、子音と母音からなる音節を基本単位として、子音が知覚されていると推定される。（限界点：音節明瞭度が80～90%となる時点）[8]

実際に音節を知覚するためには、これらの区間を含む30～40msの音声区間長が必要である。この長さは人間の情報処理の最小単位時間（知覚瞬間）に相当している。なお、子音の知覚のためには、上述の動特性のみでは必ずしも十分ではなく、破裂時点の2～数ms区間での波形の急激な変化や雑音性スペクトルの概形、破裂時点と有音声の開始時点との時間間隔（voice onset time, VOT）なども、子音を弁別する手掛かりとして用いられている。聴覚の感覚記憶の保持時間は200ms程度であり、これが言語的情報の知覚単位となっている。

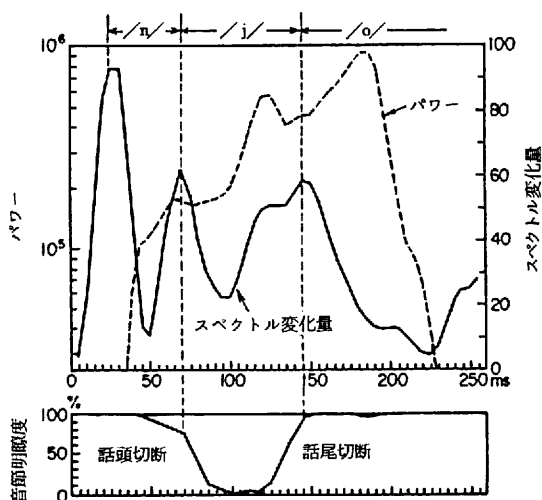


図8. スペクトルの動的特徴と切断音節の明瞭度の関係

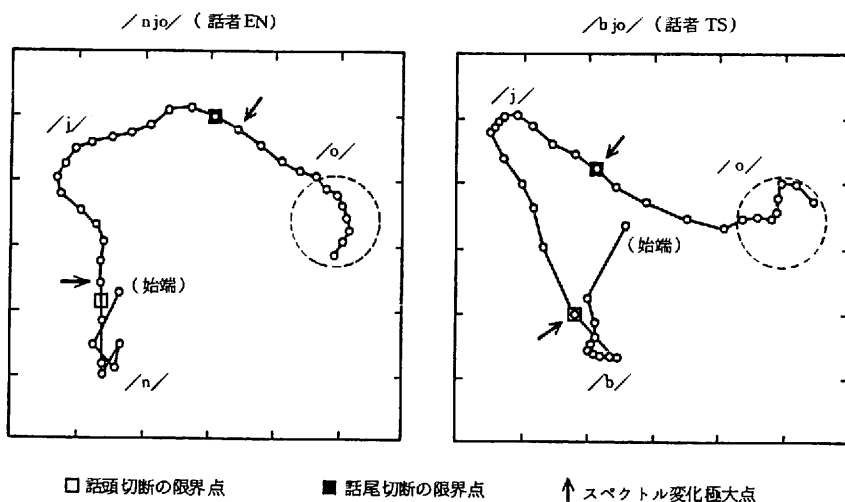


図9. 多次元尺度構成法による2次元平面に射影したスペクトルの動特性

5. 韻律性情報の動特性

アクセント、イントネーションなどの韻律性情報は、基本周波数や振幅の相対的变化(動特性)によって表現される。これらの物理特性の変動速度は比較的ゆっくりしており、その時定数は数100ms～数sである。基本周波数に関して、無声子音や無音の区間もなめらかに結んだ変動パターンについて周波数分析を行うと、変動周波数は10Hz以下になる。

基本周波数の時間的推移分布は、上昇が約18%、下降が約50%で、下降傾向が著しく、基本周波数の時間パターンは、ほぼ「へ」の字で近似することができる。このうち、声の出始めでは高く次第に時間とともに低下する特性を基本イントネーション成分または話調(声立て)成分、この特性の上に乗る単語、文節固有のアクセントを与える特性をアクセント成分とよぶ。図10にこの様子を例で示す[9]。

会話音声の中の基本周波数の時間的变化について、発声者ごとに統計的性質を調べた結果によれば、図11に示すように、女声は男声に比べて平均値、標準偏差ともに2倍程度であることがわかっている。なお、声帯の開閉周期を詳細に調べると、韻律性情報を表現する比較的ゆっくりした変化に重畳して細かい変動を有しており、これが音声に自然性を与える役割を果たしている。

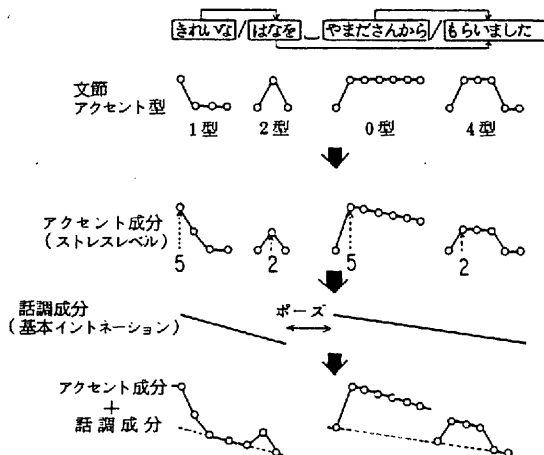


図10. 文章音声における基本周波数の時間パターンの生成

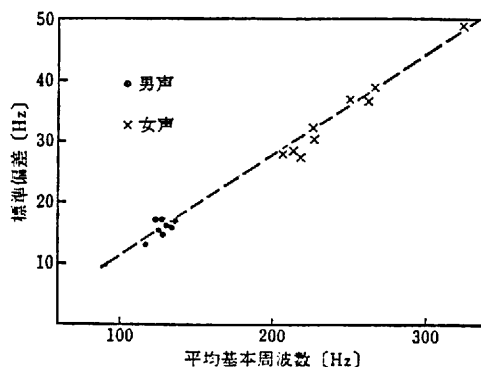


図11. 基本周波数の平均値と標準偏差

6. 話者認識におけるパラメータの長期間変動の影響

音声波に含まれる第一義的な情報は言葉の意味内容に関連した音韻性情報であるが、第二義的情報として重要なものに、誰が話しているかという個人性情報がある。音声波から個人性情報を抽出して誰の声を判定することを話者認識という。話者認識系の基本的構成は図12に示す通りで、音声波から個人の特徴を抽出して、あらかじめ各話者について標準パターンを登録しておく。未知入力に対しては、音声波から特徴抽出をしたのち、各登録話者の標準パターンとの類似度あるいは距離を調べ、その度合いにより判定を行う[10]。

話者認識は話者識別と話者照合にわけることができる。図13に示すように、話者識別とは、入力音声と全登録話者の標準パターンとの類似度を調べて、最も類似度の高い標準パターンの話者が発声したものと判定するものである。一方話者照合とは、入力音声とその発声者が名乗った本人の標準パターンとの類似度を調べて、ある一定の値以上であればその本人の音声であるとして受理し、そうでなければ他人の音声であるとして棄却するものである。

個人性を表現する特徴としては、言葉を限定したときのスペクトル包絡、基本周波数、パワーなどの時間パターンを用いるものと、その時間パターンから抽出した統計的特徴を用いるものとがある。後者の場合は発声内容を限定せず、任意の言葉から特徴を取り出すことが可能となるが、発声内容を限定しない場合は、まだ高い認識性能が得られていない。

音声波から抽出する種々の物理的特徴は、同じ話者が同じ言葉を発声してもそのつと変化する。音声スペクトルの時系列を濃淡図形に表わしたサウンドスペクトログラム(図2)は声紋とよばれるが、これと指紋との相違点は、前者が変化しやすい点にある。同じ時に繰り返し発声された音声に含まれる物理特徴の変化は少ないが、1日以上、あるいはさらに長い期間をおくと、発声の仕方や発声器官の細かい部分が変化して、音声の物理的特徴が変化し、これが話者認識の性能に影響を及ぼす。すなわち、標準

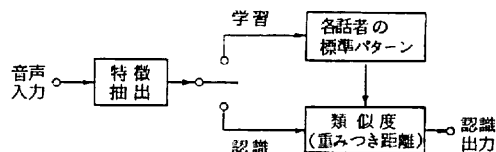


図12. 話者認識系の基本的構成

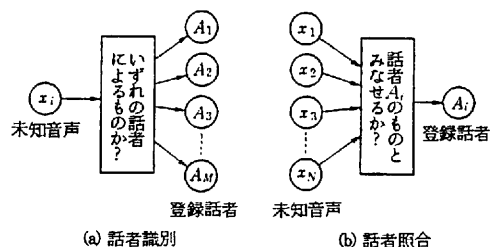


図13. 話者識別と話者照合

パターンを作成するのに用いる各話者の音声は、たとえば1日～10日間の短期間に収録したものであると、それから3週間程度までの間の音声に関しては高い話者認識率が得られるが、それ以上の時期がたつと、スペクトルの時期的変動のために急速に認識率が低下する。

発声内容に依存しない音声の普遍的特徴として長時間平均スペクトルを取り上げ、そのケプストラム(対数スペクトルの逆フーリエ変換、スペクトルパターンの大局的特徴と局所の特徴を分離して扱うことができる特徴がある)の時期的変動の様子を調べた結果について以下に述べる[11]。

まず、図14に示すような、短期間(2～3日から3か月までの種々の間隔を置いて約5か月にわたって収集したもの)と長期間(3か月おきに約2年間にわたって収集したもの)の2種類の男性9名の音声サンプルセットを用いて、各話者ごとに0～12次のケプストラムの時期的変動に関する主成分分析を行った。第1-第2主成分からなる平面上に各サンプル点を射影した結果の一部を図15に示す。2名の話者について、短期間の結果を(a)に、長期間の結果を(b)に示した。他の7名の話者も類似した結果を示す。各図とも分散を正規化して示しているため、短期間と長期間の変動の大きさは比較できないが、変動の特徴がはっきり現われている。すな

- わち、短期間の結果を見ると、10日間以内に収録した b, c, d, e の点は他の点に比べて互いに関わり合いがかなり接近しており、短い期間ではスペクトルが安定しているが、長い期間たつと大きく変わってくる性質を示している。一方、長期間の場合の各サンプル点は、3 か月の間隔をおいてかなり不規則に変動している。短期間の場合でも長期間の場合でも、その変動にははっきりした一方向性のようなものは見出されない。

次に、長時間平均スペクトルのケプストラムベクトルに関して、音声サンプルの収集時期の時間間隔と変化量の関係を調べた。ここで変化量は、次の式に示すように、多次元空間内のユークリッド距離の二乗によって計算した。 $(X_t$: 時刻 t におけるケプストラムベクトル)

$$d(\tau) = E \{ \|X_t - X_{t+\tau}\|^2 \} \quad (3)$$

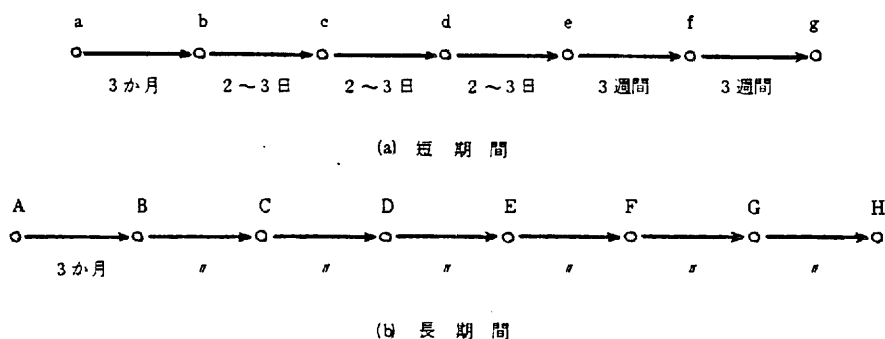


図14. 分析に用いた音声サンプルの時的関係

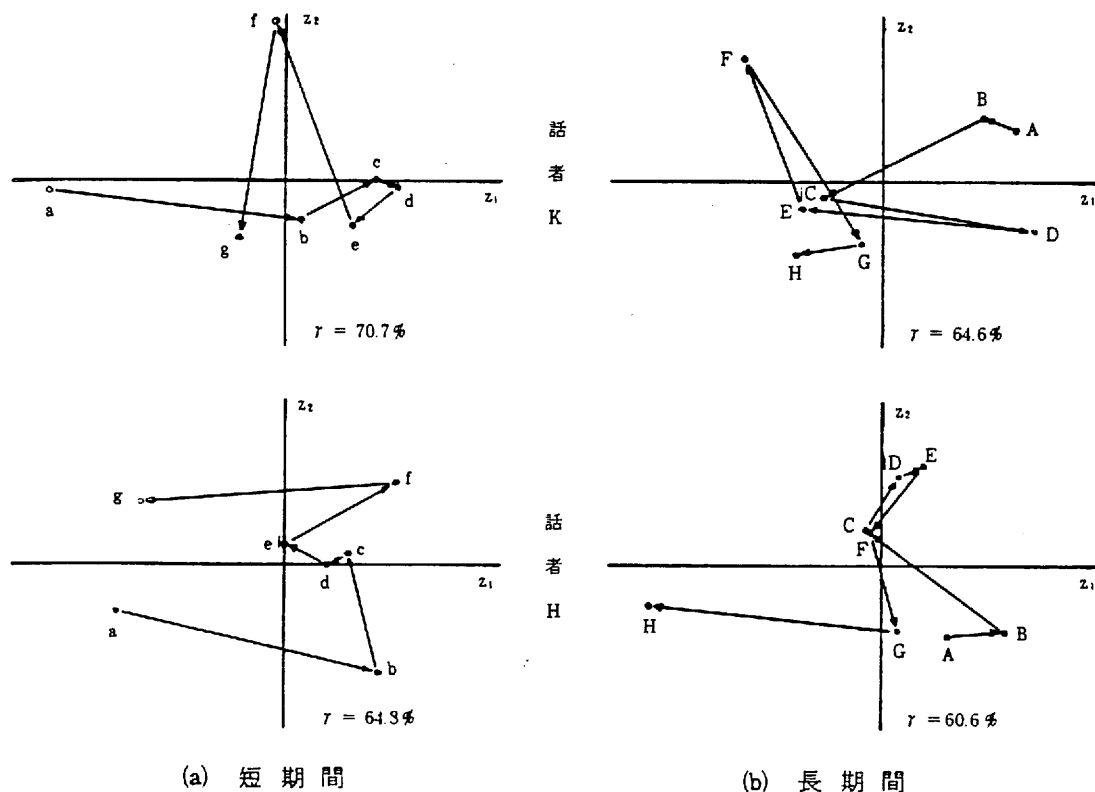


図15. 第1-第2主成分からなる平面に投影した長時間平均スペクトルのケプストラムの時的変動

図16は、9名の話者に関して平均した結果を示したものである。この図をみると、時間間隔が3か月程度になるまでは、時間間隔が長くなるに従って単調に変化量が大きくなる傾向があり、それ以上になると、時間間隔の長さに無関係にはほぼ一定になっている。この結果は、長時間平均スペクトルが、3か月程度までの間では時間の経過とともに元のパターンから平均的に遠ざかる方向にゆっくり変化する多次元ランダムウォーク的傾向を有する変化をするのに対し、3か月以上の長い期間においてサンプリングしてみると、ある平均パターンを中心に不規則な変動をする性質を有することを示している。

ユークリッド距離の二乗の期待値 $d(\tau)$ は、仮に1次元で考えれば、パラメータの自己相関関数 $\phi(\tau)$ と次のように対応づけることができる。
(τ : 時間遅れ)

$$d(\tau) = 2 \{ \phi(0) - \phi(\tau) \} \quad (4)$$

そこで、図16の関係を指数曲線と見做してあてはめを行ったところ、同図の破線の結果が得られた。これは、不規則信号をほぼ1.5～2か月の時定数を有するRCフィルタに通した成分と、通さない小振幅の成分の和からなる信号の自己相関関数から変換して求まる距離変化曲線に相当する。前者の成分は、スペクトルの時期的変動（時間間隔に依存した変動）に対応し、後者の成分は、時間間隔の長さに無関係な変動に対応すると考えられる。

各話者ごとに、ケプストラムの各成分の長期間における変動の標準偏差を求め、その値で正規化したのちフーリエ変換を施せば、長期間にわたって比較的安定した長時間平均スペクトルパターンを得ることができる。図17は、2名の男性話者が8か月にわたる4時期に発声した音声の長時間平均スペクトル(a)と、正規化したケプストラムから求めたパターン(b)を示したものである。(b)のパターンでは、スペクトルの全体的な傾斜のような概形の変動と細かい構造の変動が除かれて、よいまとまりを示しており、こうして残ったパターンに安定した個人性情報が

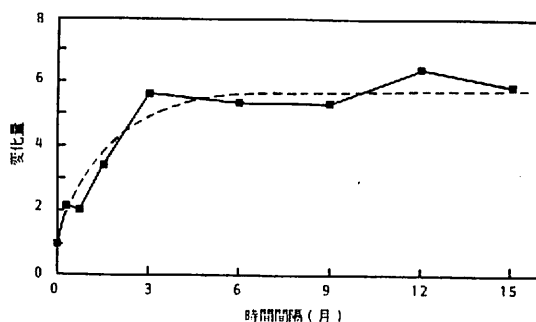
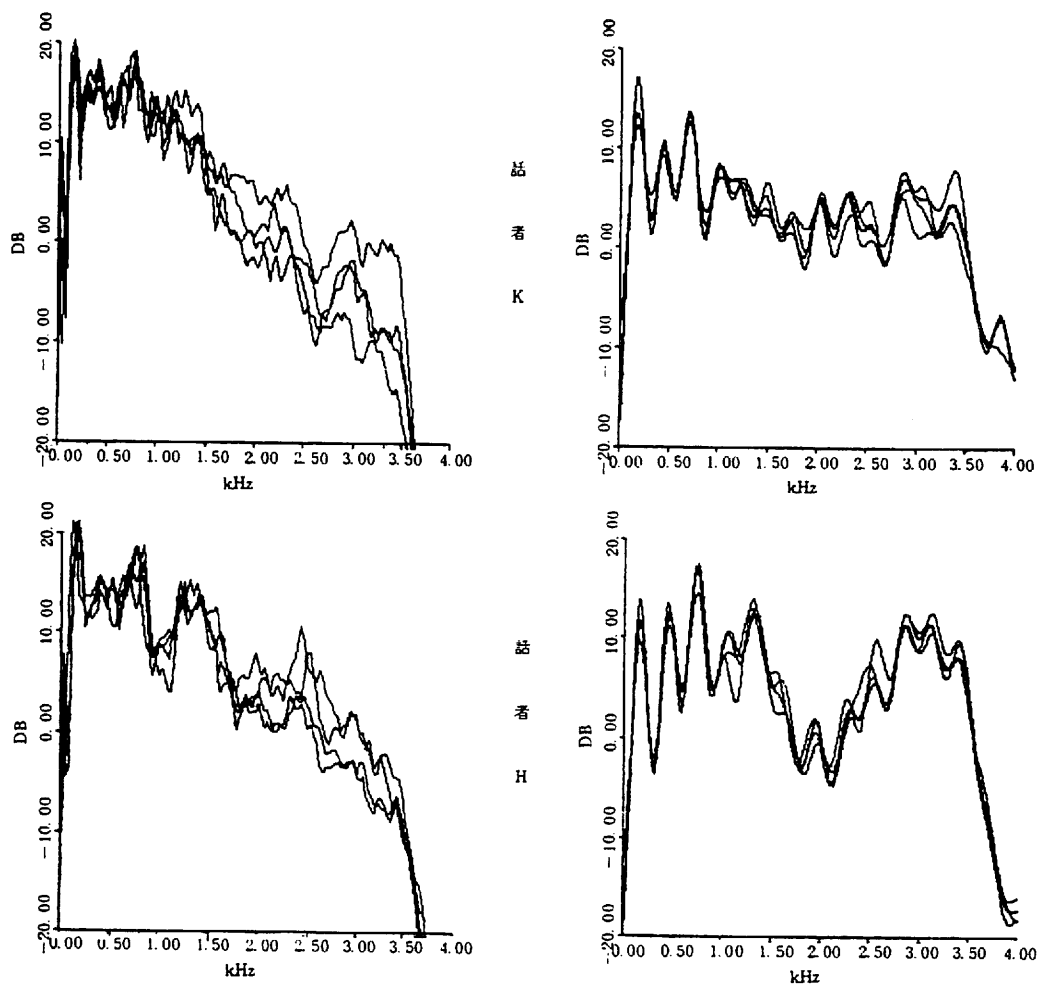


図16. 長時間平均スペクトルのケプストラムの変化量と時間間隔の関係

存在していることを示している。

スペクトルの全体的な傾斜は、声帯波スペクトル特性を表現していると考えられる。すなわち、音声スペクトルに含まれる成分のうち、声帯波スペクトル特性は時期による変動が大きく、これを除去して声道の特性のみにすることにより、比較的長期間にわたって安定した話者の特徴を得ることができる。具体的な処理としては、単語または短文章の全音声区間の平均スペクトルを1次系あるいは2次臨界制動系で近似し、音声をこの逆フィルタに通すことにより、話者認識において、各話者の音声スペクトルの時期的変動の影響を軽減し、長期間たった後の音声に対して高い認識率を維持することができる。

図18は、単語音声スペクトルの統計的特徴(対数断面積比と基本周波数の各単語区間内での平均値、標準偏差、および相互相関係数)を用いた話者認識系において、スペクトル等化処理の有効性を実験的に調べた結果を示したものである。各話者の標準パターンを作成する学習サンプルとして、10日間程度の短期間の間の4日間に発声したサンプルを用い、未知入力音声までの時期差を2～3か月から10か月まで変えて、話者照合の誤り率を調べた。この結果から、逆フィルタを適用すれば、6週間～10か月後の音声に対する誤り率を1/3程度に低下させることができることがわかる[12]。



(a) 長時間平均スペクトルパターン

(b) クラスタ化されたパターン

図17. ケプストラムを用いた長時間平均スペクトルの正規化の効果

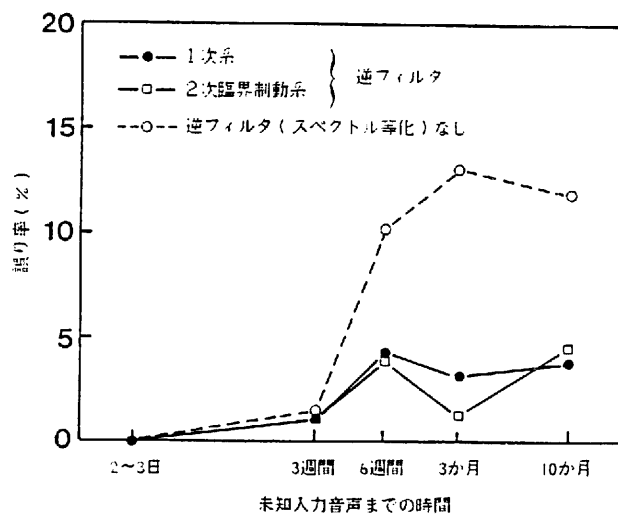


図18. 単語音声スペクトルの統計的特徴を用いた話者認識系におけるスペクトル等化処理の有効性 (短期間の学習サンプルによる話者照合の誤り率)

7. まとめ-音声現象に含まれる種々のゆらぎ

これまでに、「ゆらぎ」を広く動的性質と解釈して、音声の示す種々の現象を説明してきた。音声現象は基本的、本質的に動的現象であり、 μs オーダーから数年のオーダーにおよぶ種々の時定数をもった変動(ゆらぎ)を有している。以下にこれまでに述べたことを、時定数のオーダーの順にまとめて示す。

(1) μs オーダーのゆらぎ：音声は声帯によって断続された肺からの空気流を音源とし、声道での共振によって変調を受け、音韻固有の周波数特性を与えられた空気の疎密波である。音韻性を正しく伝えるためには、少なくとも 4 kHz 程度の周波数帯域が必要であり、従って $100\mu\text{s}$ 程度以下の時間間隔で標準化する必要がある。位相は伝送系などによって変動しやすいが、音声波形の位相の知覚的寄与は小さく、周波数スペクトルの寄与が圧倒的に大きい。

(2) ms オーダーのゆらぎ：普通に発声した音声には定常部はほとんどなく、すべての時点で非定常であり、この変化(動特性)そのものに音韻性情報がある。人間の聴覚は定常的信号よりも非定常的信号に対して高い感度を有し、動特性からその目標値を推測する能力を有する。 $2\sim\text{数 ms}$ の区間での波形の急激な変化と、スペクトル変化が極大となる時点における 10ms 程度の区間での動特性に重要な音韻性情報がある。音節の知覚には、これらの情報を含む 30ms 程度以上の音声区間が必要であり、この長さは人間の知覚瞬間に対応している。聴覚の感覚記憶の保持時間は 200ms 程度であり、これが言語的情報の知覚の単位となっている。音声による言語的情報の伝達速度は、発声器官(声道)と聴覚の両方の動的処理能力に制約されている。このため音声の分析合成系では、 $20\sim 30\text{ms}$ の区間では定常確率過程と見做して処理を行っても、聴覚的には原音に近い音声を得ることができる。

(3) s オーダーのゆらぎ：音韻性情報のほかにそれと密接に関連した音声の重要な情報として、アクセント、イントネーションなどの韻律性情報がある。これは基本周波数や振幅の数 $100\text{ms}\sim\text{数 s}$ の時定数の変化によって表現され

る。この特徴もある絶対的な値に意味があるのではなく、変化(動特性)そのものに情報がある。基本周波数は、韻律性に対応したゆっくりした変化に重畳した細かいゆらぎを有しており、これが音声に自然性を与えている。

(4) 数時間～数年オーダーのゆらぎ：同じ人が同じ言葉を発声しても、その波形とスペクトルはそのつど変化する。同じ時に繰り返し発声したときの変化は小さいが、1日以上時間をおくと変化量がそれに伴って増大する。ただし限りなく増大することではなく、 $1\sim 2$ 年の間の変化の様子は、不規則信号をほぼ2か月の時定数をもつ RC フィルタに通した成分と通さない成分の和で近似することができる。個人を識別する手掛かりとして音声を用いる話者認識の際には、この変化の正規化が重要となる。音声生成のメカニズムにもとづいて分析すると、音声波に含まれる情報のうち、声道に関連する情報の方が声帯波に関連する情報よりも長期的には安定しており、逆フィルタによって声帯波情報を近似的に除去することが、正規化に有効であることが確認されている。数年以上にわたる声質の変化は、人間の成長や老化に関連している。

これらの種々の時間変動(ゆらぎ)特性について、筆者自身 $1/f$ 雑音として解析を試みたことはまだないが、音声パワーの変動などが $1/f$ 雑音として取り扱うことができることはすでに報告されており[13]、ここで述べた現象の多くも $1/f$ 雑音として解析することが可能であると考えられる。今後試みてみたいと考えている。

・ [参考文献]

- [1] 斎藤, 中田: “音声情報処理の基礎”, オーム社 (昭56)
- [2] Flanagan, J. L.: “Speech analysis, synthesis, and perception”, 2nd Edition, Springer-Verlag (1972)
- [3] 板倉, 斎藤: “統計的手法による音声スペクトル密度とホルマント周波数の推定”, 信学論, 53-A, 1, pp. 35-42 (昭45)
- [4] 板倉文忠: “スペクトル符号化に基づく音声分析合成”, 音学誌, 37, 5, pp. 197-203 (昭56)
- [5] 東倉洋一: “偏自己相関係数を用いた音声分析合成系における音声品質向上の研究”, 東京大学学位論文 (昭55)
- [6] Lindblom, B. E. F.: “On the role of formant transitions in vowel recognition”, J. Acoust. Soc. Amer., 42, 4, pp. 830-843 (1967)
- [7] 境, 中山: “聴覚と音響心理”, 音響工学講座6, 日本音響学会編, コロナ社 (昭53)
- [8] 古井貞熙: “音声知覚におけるスペクトル変化情報の役割”, 音学聴覚研資, H-85-6 (昭60)
- [9] 佐藤大和: “音声合成”, 数理科学, 252 (6月号), pp. 42-48 (昭59)
- [10] 古井貞熙: “話者認識の現状と展望”, 信学誌, 67, 5, pp. 101-107 (昭59)
- [11] 古井貞熙: “音声波に含まれる個人性情報の研究”, 東京大学学位論文 (昭53)
- [12] 古井貞熙: “音声の個人性のパラメータの時期的変動と話者認識”, 信学論, 57-A, 12, pp. 880-887 (昭49)
- [13] Voss, R. F. and Clarke, J.: “ $1/f$ noise in music and speech”, Nature, 258, pp. 317-318 (1975)