

雑音環境下での音声認識に関する一考察

6N-8

山本浩通、磯部俊夫
航空宇宙技術研究所

1. まえがき

将来的な航空宇宙技術は、システムの巨大化・複合化が必然的となる。それに対処するため、搭乗員等人的要素を含むシステム全体の高度な信頼化が必須な要素技術的課題となる。ロケット/ジェットエンジン・空気力学的および機体構造的振動などによる環境雑音、通話・通信系統などにおけるシステム雑音等への対処は、今後の航空宇宙技術の研究開発における音声認識技術に課せられた不可避な基盤的技術課題である。

従来、雑音に埋もれた信号の抽出処理の研究は各々の応用分野においてそれぞれ行われて来ているが、雑音環境の下での音声信号の認識技術においては、幾多の課題が残されている。ここでは、雑音を含む信号データの修復等に対して研究開発が進められている弛緩(Relaxation)法を発声音の認識問題へ応用する事を試みる。

2. 音声認識における雑音除去

発声音の認識は音響学的にも、あるいは音声パターンの同定問題としても、内外における各種のプロジェクトをはじめ、広く研究開発が進められて来ている。しかし、従来の多くの音声認識法は、発声方法ある

いは話者個々人の差異に対する配慮が主体的に行われて来ており、雑音環境下、あるいは話者を特定しない場合等に対しては、多くの課題がある。

雑音環境下での発声音を、時系列的信号処理の一環として対処する方法として、適応フィルタリングによる音声信号の抽出、スペクトル空間での雑音成分の削除、等の種々の雑音除去処理方式が考えられるが、音声信号に雑音が相乗し、音声波形の歪、音声周波数の変動等が生じ、従来の音声認識方法では対処する事が容易でない事が洞察されて来ている。これに対して、音響センサー以外のマルチセンサーの併用による含雑音音声信号の認識方法¹⁾等の提案もなされている。

3. 確率的弛緩法の適用

発声音の特徴を表す各種のパラメータ、例えば、日本語母音においてその特徴が代表され得る幾つかの周波数(フォルマント周波数)、あるいは、周波数特性空間における各母音間の相対位置的な特性等、を利用した各種の認識アルゴリズムが報告されている。

しかし、このような特性パラメータは、発声音に必然的に存在する雑音成分により、

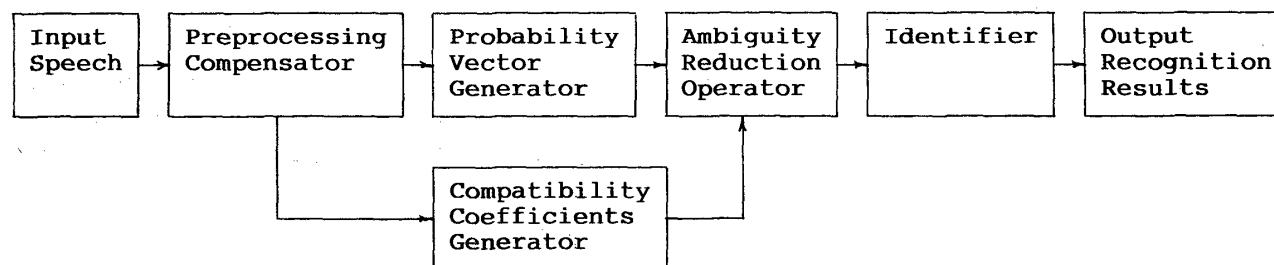


図1 弛緩法による音声認識システム

時には著しい悪影響を認識結果に与える。発声音の生成過程に含まれる雑音成分の特定が容易な場合には、前处理的な雑音除去過程を認識アルゴリズムに加える事で解決し得る可能性もあるが一般的ではない。したがって、雑音を含む信号においては、信号パラメータの同定に際しての”曖昧さ”を前提としたアルゴリズムが妥当となる。

すなわち、発声音は、各種の雑音要因から特徴パラメータの検出のみによる同定では、一意的に識別する事が不可能な場合が多く、第2次的に時空間的構造情報を用い、発声音の識別性を増強させる方式が有効となる。このための1つの手段が弛緩法²⁾の適用である。

ここで提案する弛緩法による音声認識法は、次の特徴を有する：

- 1) 音声特徴空間により、同定すべき発声音の可能性を確率的に算定し、認識処理過程の短絡化を可能とする。
- 2) 並列繰返し型弛緩演算処理を基調として、高雑音レベルの下での発声音の同定に対処する。

本方式における基本式は次の(1)式および(2),(3)式である。

$$P^{(0)}(\lambda) = p(x|\lambda)P(\lambda) / \sum_{\lambda} p(x|\lambda)P(\lambda) \quad (1)$$

$$P^{(k+1)}(\lambda) = \frac{P^{(k)}(\lambda)(1+q^{(k)})}{\sum_{\lambda} P^{(k)}(\lambda)(1+q^{(k)})} \quad (2)$$

$$q^{(k)} = r_{d\lambda} P^{(k)}(\lambda) \quad (3)$$

$$\lambda \in \{ \lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n \}$$

ここで、 $p(x|\lambda)$ は、発声音の特徴空間における認識音 λ に対する識別発声音 x の確率密度、 $P(\lambda)$ は λ の事前確率である。また、 $r_{d\lambda}$ は時空間的構造情報としての適合係数であり、2次的処理における中心的な役割を果たすもので、個々の認識音に対

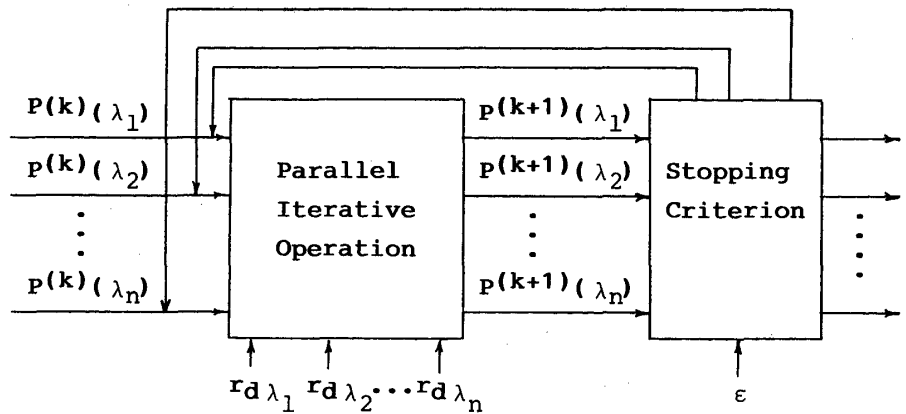


図2 処理基幹部

して算定される。(2),(3)式の並列型繰返し処理は、

$$| P^{(k+1)}(\lambda) - P^{(k)}(\lambda) | < \epsilon \quad (4)$$

なる収束判定条件が設定される。

4. 全体処理システム概要

弛緩法を利用する音声認識システムの全体処理システムのブロックを図1に示す。処理基幹部は、図2に示す如く、雑音成分を含む発声音個々に対応する並列型確率弛緩処理を基調とするもので、弛緩結果を次の如く定式化される識別部で発声音に対する認識音 λ の抽出を行う。

$$\lambda = \max_{\lambda_i} \{ P^{(k+1)}(\lambda_i) \mid \lambda_i \in \{ \lambda_1, \lambda_2, \dots, \lambda_n \} \} \quad (5)$$

5. あとがき

ここで提案した方式は、従来のな諸方式に比べ、広範囲な適応性を有するものであるが、システムの簡便性、連続発声音認識における処理時間的な対応性、等については今後の課題である。

参考文献

- 1) V. Viswanathan, C. Henry, R. Schwartz, and S. Roucos, ICASSP 86, Tokyo, pp.85-88, April 1986.
- 2) A. Rosenfeld, R. Hummel, and S. Zucker, IEEE Trans. Syst., Man, Cybern., vol.SMC-6, pp.420-433, 1976.