

6D-6

音声認識における背景雑音の影響

糟谷敏宏 村上憲也
NTTデータ通信(株)

1. はじめに

現在、音声認識装置における耐雑音性能向上を目的とした研究が各機関により盛んに行われている。音声認識の背景雑音として挙げられるのは、計算機室内ノイズ、駅のコンコースでの騒音、自動車内騒音等であるが、実際の研究の多くは計算機シミュレーションにより得られる白色雑音を用いた例が多く、実雑音が用いられるようになったのはごく最近の事である。白色雑音をもとに開発された雑音除去法では充分にその機能を達成していないのが現状である。しかも実雑音が音声認識に与える影響を深く解析した例はまだ見られない。そこで本稿では、実雑音が音声認識に与える影響を解析し、テンプレート変形による耐雑音性能向上方式の検討を行った。

2. 背景雑音の分析

白色雑音の特徴を利用した雑音除去手法が、過去に数多く提案されている。計算機により生成される白色雑音は、スペクトルが平坦であり定常的であるという特徴を持っているため、白色雑音に対する除去法が実際の背景雑音に対してどの程度有効であるかの検討が必要であるが、一般には実雑音の特徴が明白ではないために難しい。

本稿では、音声認識時に特徴量として用いられるパワー、スペクトル、LPCケプストラムを種々の雑音から抽出し、それらのパラメタの定常性、分散について分析を行った。

分析の対象としたのは、駅コンコース、駅広場、駅前電話ボックス内、及び駅周辺の電話ボックス内で収録された雑音である。収録したサンプルの中からそれぞれ2分間のデータを切り出して使用した。

図1は、雑音のパワー変化を示している。白色雑音のパワーやスペクトルの変化は理論的には零であるが、分析窓の影響のため多少変動している。実雑音は白色雑音に比してパワー変動が大きく

(10dB程度)認識に悪影響を及ぼすと考えられる。

背景雑音のスペクトルは様々で、図2に示されるようにコンコース、駅前等では比較的平坦な特性を示しているが、電話ボックス内では共鳴によると考えられる1000Hz付近の成分が強く現れ、全体的にも低域成分が強い結果となった。従って電話音声の認識を考えたとき、電話ボックス内雑音特有のスペクトルを考慮する事は有効であると考えられる。

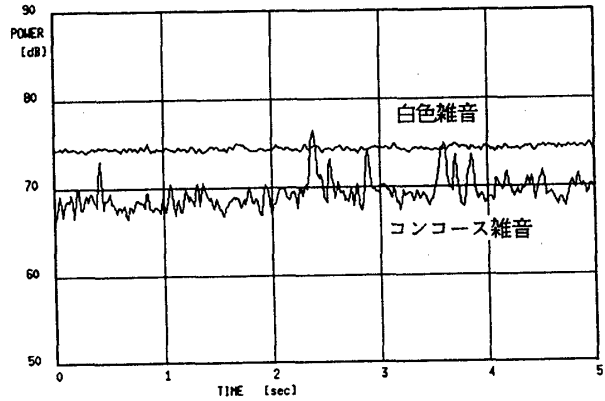


図1. 5秒間のパワー変化

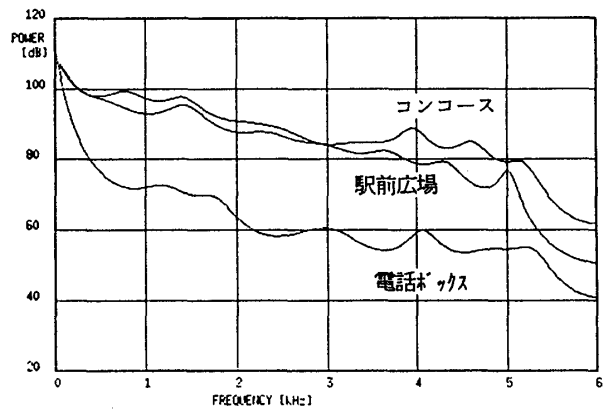


図2. 背景雑音のスペクトル

図3は、雑音の重畳されていない母音100音節から抽出された特徴ベクトル(LPCケプストラム)に対して主成分分析を施しその最大分散軸と第2分散軸を算出し、雑音を含まない音声、音声にSNR 20 dB, 10 dB, 0 dBで雑音を重畳させた音声、雑音のみのサンプルについてそれぞれの軸方向成分をプロットした結果である。雑音の重畳によりクラスターが図の上では右下に移動した形となって現れている。この図からわかるように、雑音により音声のスペクトル変形を受け特徴ベクトルが大きく変化し、SNRの劣化に伴い雑音のみのサンプルから得られる特徴ベクトルの近傍に分布が近づいてゆき、カテゴリ間の距離が小さくなるため認識を困難にすると考えられる。また、カテゴリ内の特徴ベクトルの分散が雑音の分散値に近づいてゆく傾向を示している。

これは、比較的不安定な音声スペクトルの谷部

がノイズにより底上げされることによりスペクトルが変形を受けた為であると考えられる。

次に、母音に様々な背景雑音を重畳させた時のパラメタ分散値を示す。図4は、母音 /a/ に実雑音を加えた時の分散値を示している。このように実雑音においても白色雑音程顕著ではないが、雑音の重畳によりパラメタ分散値が低下していることが理解できる。

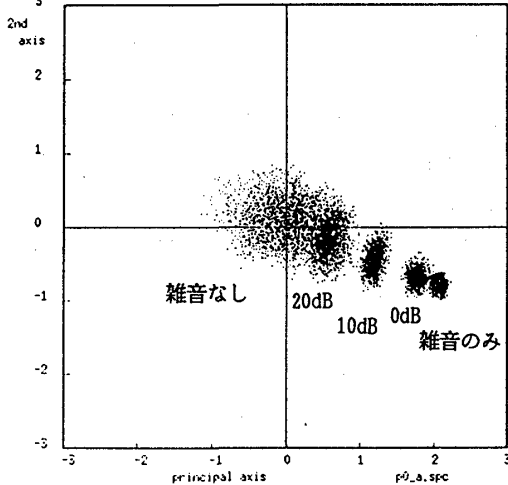


図3. 雑音重畳時のパラメタ変化

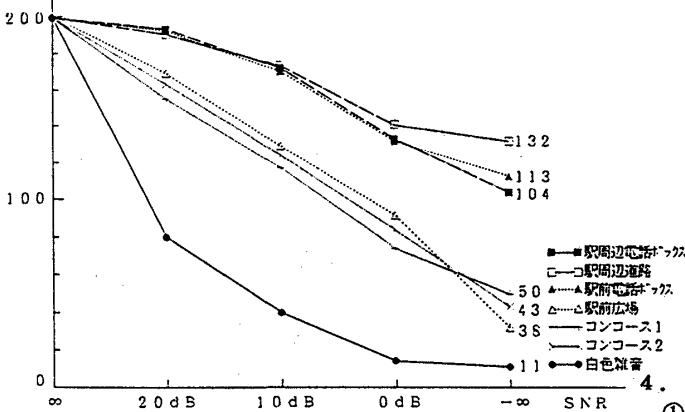


図4. 重畳音声のパラメタ分散の変化

3. 耐雑音性能向上方式の検討

以上の実雑音の特徴を考慮し、認識に用いるテンプレートに変形を加える事により雑音パワー変動の吸収が可能ではないかという想定のもとに、簡単な母音認識実験を行った。

マルチテンプレート法は、様々なSN比で雑音の付加した音声から得られる複数の標準パターンを用いて、雑音によるスペクトル変形を救う方法である。しかし、テンプレート数が、増すことにより認識時に時間を要する点や、テンプレートに無いSNRの音声はそれに近いいずれかのテンプレートを用いて認識されるという問題点を抱えている。そこで、雑音のパワー変化方向へパターンを変形することにより、認識性能の向上を考えた。

[方法]

パターンの変形は、実際には変形処理を施さない場合の距離 (d) に雑音のパワー変化を考慮した距離 (dn) の和を新たな距離とする事によって行っている。雑音変化を考慮した距離 (dn) は、標準ベクトルに対し雑音パワーが微小変化したと

きの特徴ベクトルの変化方向を算出し、それと垂直な方向へサンプルが離れたときに大きな値を示すような方向重み付け尺度である。

距離尺度

$$距離 = d + \alpha \times dn$$

d : 通常の変化を考慮した距離
dn : 雑音係数

雑音パワー変化方向の計算

(雑音の無いパターンの場合)

$$c^{s+n} = c^s + \Delta c$$

$$\Delta c = F^{-1} [F \frac{c^n}{F_s} - c^n]$$

($F_n \ll F_s$)

c : ケブストラム
s : 音声信号
n : 付加雑音
F : フーリエ変換

[認識結果]

認識率
標準パターン
距離尺度

母音認識実験結果
標準パターンにないSNRの音声の認識率の向上のみならず、標準パターンとして挙げたSNRの音声に関しても認識率は向上した。これは、本手法を用いない場合に別のSNRの異カテゴリのテンプレートにマッチングするサンプルが多く見られたが、それを救済することができたためと考えられる。

表1. 母音認識実験結果

認識率 (%)	処理前		処理後	
	◇	◆	◇	◆
白色雑音	91.5	94.7	94.2	93.9
駅コンコース	94.1	95.5	93.4	96.5
駅前広場	94.3	96.6	94.7	97.1
駅前電話ボックス	82.4	96.3	82.1	96.4

4. 検討

- ① 白色雑音、実雑音いずれの場合にも認識率は処理前に比べ向上し、本手法の有効性が示された。
- ② 標準パターンにないSNRの音声の認識率の向上のみならず、標準パターンとして挙げたSNRの音声に関しても認識率は向上した。これは、本手法を用いない場合に別のSNRの異カテゴリのテンプレートにマッチングするサンプルが多く見られたが、それを救済することができたためと考えられる。
- ③ 標準パターンとして挙げたSNRの音声の認識率が、標準パターンにないSNRの音声の認識率より悪いのは、0dBの認識率が他に比して大幅に低かったためと考えられる。
- ④ SNRの異なるテンプレート間の補完が可能となるため、テンプレート数の削減が計れると考えられる。

参考文献

[1] 水谷、北村：“雑音下の数字音声認識における参照パターンと距離尺度の検討”，信学論SP88-121