

1. まえがき

音声認識率の低下の原因は音声に雑音が混入することである。雑音環境において、音声データをあらわすスペクトル包絡は、谷の部分がその雑音で埋もれレベル上昇を起こす。

改善法として今回はケプストラムの次数が大きくなった場合、その値が小さくなることに着目した。ケフレンシ方向に重みを付け強調する実験を行った。

そして、低周波数帯域を利用する周波数軸変換法⁽¹⁾、雑音によって消えてしまった谷を回復するスペクトル包絡変形法、⁽²⁾スペクトルの形を滑らかにするスペクトル平滑化法⁽³⁾などを利用した。ケフレンシ方向強調と3つの変形法の組み合わせで実験を行った。さらに、ケフレンシ方向強調した重み付けケプストラムとケプストラムの併用実験での認識率の検討も行った。

2. ケフレンシ方向強調法

ケプストラムへの重み付けは、入力・標準の音声データのケプストラムを $c(m)$ とすると、重み付けケプストラム $cw(m)$ は次式になる。

$$cw(m) = \left(1 + \frac{m-1}{u}\right) c(m) \quad (m=1,2,\dots,N) \quad (1)$$

ここで m はケプストラム次数で u は強調の度合いを決める強調係数である。ケプストラムでスペクトル包絡を表すと(2)式になる。

$$S(k) = \sum_{m=0}^{N-1} c(m) \cos\left(\frac{2\pi k m}{N}\right) \quad (2)$$

となり、これを周波数成分 k で微分すると

$$\frac{d}{dk} S(k) = -\frac{2\pi}{N} \sum_{m=0}^{N-1} mc(m) \sin\left(\frac{2\pi k m}{N}\right) \quad (3)$$

となるので、これを群遅延スペクトルという。これから式(1)で $u=1$ とした場合は群遅延スペクトルのマッチングを行っていることになっている。図1にケフレンシ方向強調のスペクトル包絡変化例を示す。

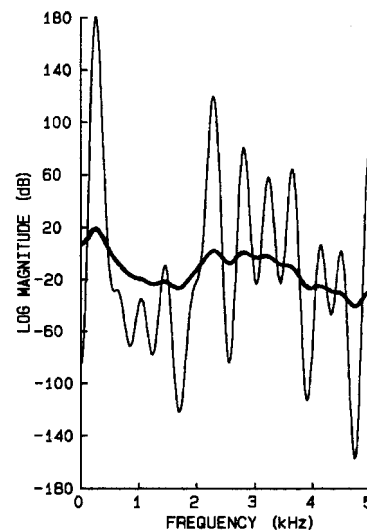


図1 ケフレンシ方向強調のスペクトル包絡変化例

(太線: 変形前 細線: 群遅延スペクトル)

3. ケフレンシ方向強調法の組み合わせと併用

スペクトル包絡変形法、スペクトル平滑化法、周波数軸変換法の3つをケフレンシ方向強調法と組み合わせで実験を行った。この3つのスペクトル変形法について説明する。

図2-1のように雑音環境下ではスペクトル包絡に変化が起きている

・周波数軸変換法

ノイズによる影響は高周波数領域より低周波数領域の方がノイズの影響が少ない。周波数軸変換法とは図 2-2 のように、低周波数領域の割合を大きくするスペクトル変形法である。

・スペクトル包絡変形法

雑音環境下でスペクトル包絡の谷がノイズで埋もれてしまい認識率の低下の原因である。これを改善するために、スペクトル包絡変形法により図 2-3 のように欠落した谷の規則変形を行い、谷を付加する作業を行う方法である。

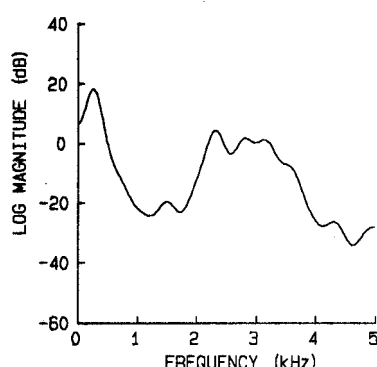


図 2-1 変形する前のスペクトル

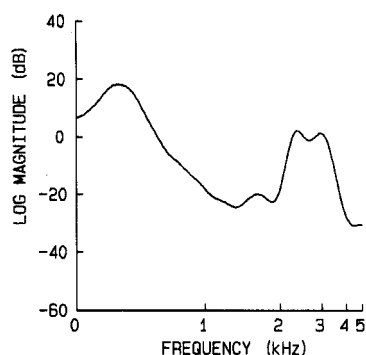


図 2-2 周波数軸変換したスペクトル

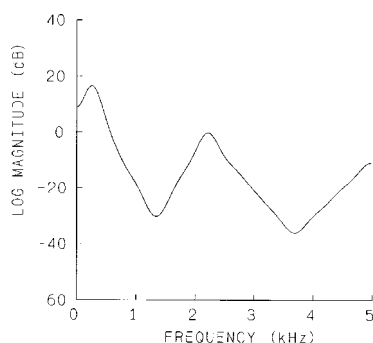


図 2-3 スペクトル包絡変形法により変形したスペクトル

・スペクトル平滑化法

スペクトル包絡変形法によって谷付けした後で得られるスペクトル包絡は図 3-(a) のように乱れている。改善法として、スペクトル平滑化法を用いることで図 3-(b) のようにスペクトルの平滑化を行う。実験ではスムージングフレーム 7 フレームを用いて実験を行った。

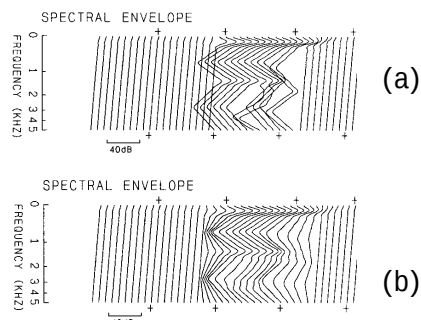


図 3-(a) 谷付けしたスペクトル包絡
(b) 平滑化したスペクトル包絡

これら 3 つのスペクトル変形法とケフレンシ方向強調法を組み合わせることで強調係数 u を変化させて実験を行った。

次の実験として、通常のケプストラムとケフレンシ方向強調により重みを付けたケプストラムを併用させて認識実験を行った。入力音声データのケプストラム $c(i)$, $cw(i)$ とし、標準となる音声データのケプストラムを $c'(i)$, $cw'(i)$ とする。また、 $d(i)$ を標準データと入力データの距離とすると次式になる。

$$d^2(i) = \sum_{i=1}^8 \{ [c(i) - c'(i)]^2 + (k [cw(i) - cw'(i)])^2 \} (4)$$

ここで k は併用係数である。この式を用いて、強調係数 u と併用係数 k を変化させて実験を行った。

3. 認識実験

本実験では、男性話者 8 名の各 3 回発声した 10 数字音声を 2 種類の有色雑音ピンクノイズと自動車ノイズを S/N 比 0dB, 10dB で重畳したものを用いた。

ケフレンシ方向強調を使用して、強調係数の

値を 0.1～ 5 まで変化させた時の認識率の変化を図 4 に示す。マッチング次数 1 次から 1 2 次でマッチングした。

結果はすべての雑音音声においても認識率の向上が見られた。しかし、逆に強調しすぎて認識率は低下してしまう。具体的に見ると自動車ノイズは $u=0.5$ 付近で認識率の向上が見られ、ピンクノイズ 10dB は $u=1$ 、ピンクノイズ 0dB は $u=0.2$ で認識率が良くなっている。このことから認識率の向上する u の値は各ノイズ違うが $u = 1$ 近辺で認識率がケフレンシ方向強調法の有効性が分かる。

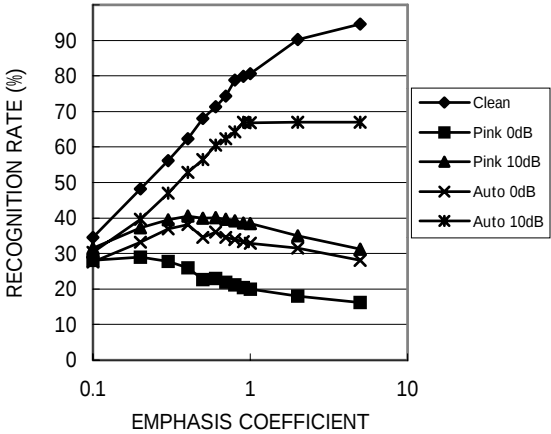


図 4 ケフレンシ方向強調を用いた認識率の変化

次の実験ではスペクトル包絡変形法、スペクトル平滑化法、周波数軸変換法の 3 つ変形法をケフレンシ方向強調法による重み付けと組み合わせた。表 1 の A～E の 5 パターンで実験を行った。

A. 重み付け
B. 周波数軸変換 + 重み付け
C. 包絡変形法 + 重み付け
D. 包絡変形法 + 周波数軸変換 + 重み付け
E. 平滑化 + 重み付け

表 1 ケフレンシ方向強調法の組み合わせた場合のパターン

マッチング次数は 1 次から 1 2 次でマッチングさせた。実験結果を図 5 - 1.2.3.4 にノイズの種類ごとに示す。

ズの種類ごとに示す。

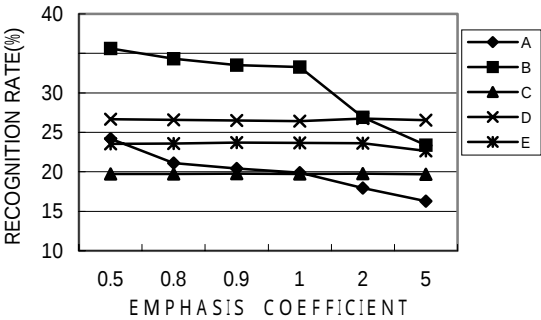


図 5-1 ケフレンシ強調法を組み合わせた場合の認識率(ピンクノイズ 0 dB)

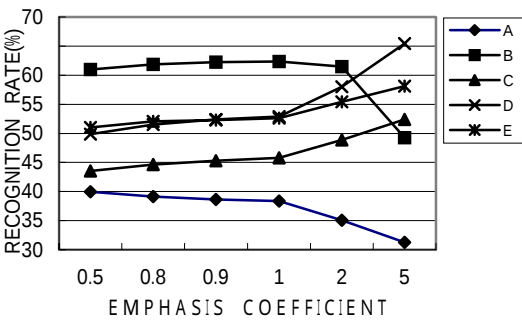


図 5-2 ケフレンシ強調法を組み合わせた場合の認識率(ピンクノイズ 1 0 dB)

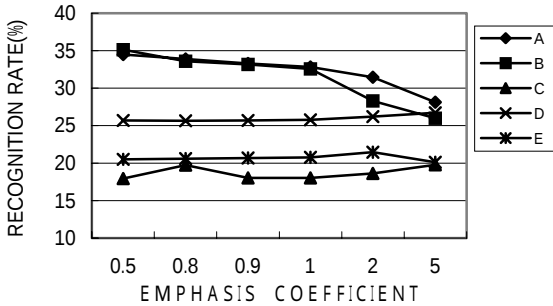


図 5-3 ケフレンシ強調法を組み合わせた場合の認識率(自動車ノイズ 0 dB)

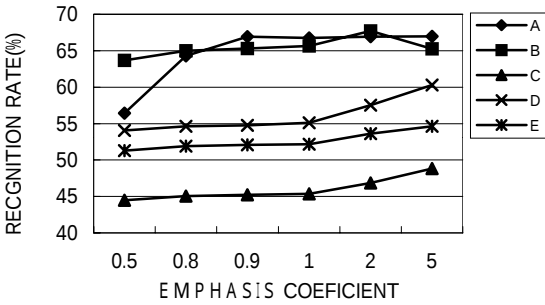


図 5-4 ケフレンシ強調法を組み合わせた場合の認識率(自動車ノイズ 1 0 dB)

周波数軸変換法とスペクトル包絡変形法とスペクトル平滑法との3つの方法をそれぞれケフレンシ強調法と組み合わせて5パターンの認識率の実験を行った。

ピンクノイズと自動車ノイズ結果を比較すると、ピンクノイズの場合はケフレンシ方向強調だけ使うよりも他の方法を組み合わせた実験結果のほうが全体的に認識率が向上するのがわかる。しかし、自動車ノイズの場合はケフレンシ方向強調だけと周波数軸変換法を組み合わせた場合が良い結果が出ている。

全体的に比較してみると周波数軸変換法とケフレンシ方向強調を組み合わせた場合が最も良い結果がでている。認識率低下の原因の1つには音声の特徴のスペクトルは低周波数領域で大きく、高周波数領域で小さくなっているため、雑音の付加した音声はスペクトル包絡の谷が欠落するために認識率が落ちてしまう。そこで、周波数軸変換法を使うことで高周波数領域では元の音声特徴があまり残らないので、音声特徴がある低周波数領域の強調の度合いを高める。だから、周波数軸変換法によって強調したケプストラムをケフレンシ方向強調によって高次数のスペクトルを特徴づけたことによって認識率の向上につながっていると考えられる。

最後の実験は、ケプストラムとケフレンシ方向強調により重みを付けたケプストラムを併用させて認識実験を行った。マッチング次数は1次から16次でマッチングを行う。実験結果は図6に示す。

ケフレンシ方向強調を使うと標準データと入力データの距離が大きくなるので併用係数 k を掛けることによってその距離を小さくすることで、ケプストラムと併用したときに認識率を向上するのを図ったが、結果として併用係数 k を変化させたが認識率の向上は1%ぐらいしか見られなかった。

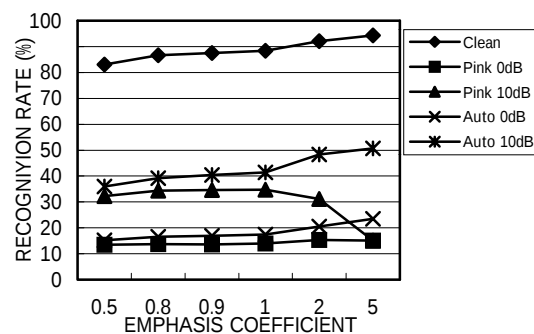


図6 ケフレンシ方向強調を併用した場合の認識率(併用係数 $k=0.7$)

5. むすび

実験より周波数軸変換とケフレンシ方向強調の組み合わせると認識率に良い結果が得られた。これは、周波数軸変換によって低周波数領域を強調し、ケフレンシ方向強調法でスペクトルの特徴を効率よく強調できたからだと考えられる。今後は、周波数軸変換法とケフレンシ方向強調を組み合わせることによってその特徴を他の改善法に生かすこと等が挙げられる。また、ケフレンシ方向強調法を他の方法の組み合わせや併用なども考えられる。

参考文献

- (1)秋田、長尾、柴山：信学技法 EA96-61(1996)
- (2)秋田、大倉：信学技法 EA95-57(1995)
- (3)秋田：電気情報通信学会秋季大会 D-382
(1994)

