スペクトル軸変形を用いた雑音環境音声認識*

波多野　志郎　　秋田　昌憲　　緑川　洋一
大分大学工学部

１．まえがき

現在、音声認識を利用、応用したシステムが多いが、完全な音声認識というのは未だ実現されていない。実際に音声を認識するためにはクリアしなければならない様々な問題がある。

そのひとつに、雑音がある。音声を発する人の周囲が無音であれば問題ないのだが、そういうわけにはいかず、他人の声だとか、その他様々な雑音が混同してしまう。今回の実験ではそういった雑音を数字音声に擬似的に付加した音声データを使い、それと雑音を付加する前の数字音声との認識率を見ることによってより認識率の高い方法を知る事が大きな目的である。

そこで、有声部が無声部より雑音の影響を受けにくいことに着目し、有声・無声判断[1]を行い、さらに低域周波数領域を強調することによって認識率の向上を試みた。その方法として帯域制限が挙げられる。これは、高域周波数領域をカットし認識率の向上を図るものであり、それに加え周波数軸変換を用いる。これは周波数軸変換係数αを用いて低域周波数の部分を強調するものである。

さらに、これらの変形を行ったデータにスペクトル強度軸変形法を適用し、その認識率の変化を観察した。

２．周波数軸変換法によるスペクトル包絡の低周波数領域の強調

雑音環境下では図１のような変化が起こる。この場合、低レベル部は埋まってしまうが、低周波数部分は影響が少ない。このことから、周波数軸変換法と帯域制限法によって低域周波数領域の特徴を強調する。周波数軸変換は、Oppenheim の再帰式[2]を用いる。この再帰式は、通常の最小位相のケプストラムを$c(i)$、周波数変換された後のケプストラムを$\tilde{c}(i)$とすると次式のようになる。

$$w_0^{(m)} = c(m) + \alpha w_0^{(m+1)}$$
$$w_1^{(m)} = (1-\alpha^2) w_0^{(m+1)} + \alpha w_1^{(m+1)}$$
$$w_j^{(m)} = w_{j-1}^{(m+1)} + \alpha\left[w_j^{(m+1)} - w_{j-1}^{(m)} \right]$$
$$\tilde{c}(i) = w_i^{(0)}$$

ここで、周波数軸変換係数をαとする。

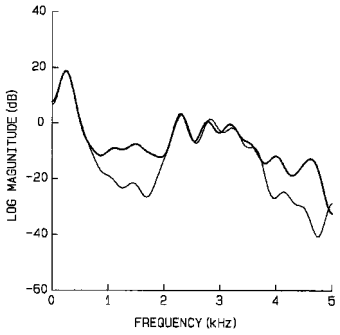

図１　男性話者 /i/のスペクトル包絡の雑音による影響

３．帯域制限法によるスペクトル包絡の特徴抽出

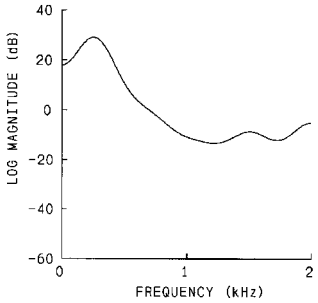

図２　男性話者/i/（無雑音）の帯域制限スペクトル（2kHz）

＊ Noisy Digits Recognition Using Spectral Envelopes Evaluated on Warped Magnitude Axis.
Shiro HATANO , Masanori AKITA , and Yoichi MIDORIKAWA (Oita University)

帯域制限スペクトル[3]は、スペクトル包絡の高周波数域を定められた周波数以後をカットし、これを再サンプリングすることで求められている。この包絡の例を図２に示す。

３．スペクトル包絡強度軸の非線形変換法

低域強調に加えて、雑音による包絡全体の平滑化を補正する意味で、ここではスペクトルのピーク域の強度レベルはそのままにしながら、低レベル強度軸を圧縮する（１）の非線形変換とスペクトルのピーク域の強度レベルを強調する（２）の非線形変換を行う。

β、γをそれぞれの変換係数とし、元のスペクトルを $S(k)$、変換後のスペクトルを $S'(k)$とする。

$$S'(k)=\begin{cases} S(k) & (S(k)\geq TH1) \\ TH1-\beta(TH1-S(k)) & (S(k)<TH1) \end{cases} \quad (1)$$

$$S'(k)=\begin{cases} S(k) & (S(k)\leq TH1) \\ TH1+\gamma(S(k)-TH1) & (S(k)>TH1) \end{cases} \quad (2)$$

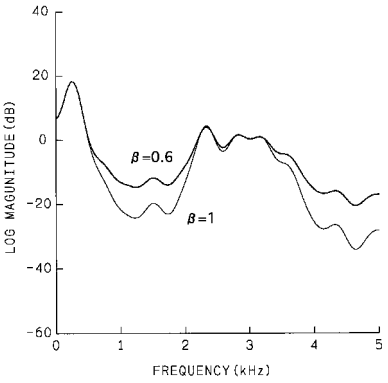

図３-(a)　スペクトル強度軸変化例（閾値=0）

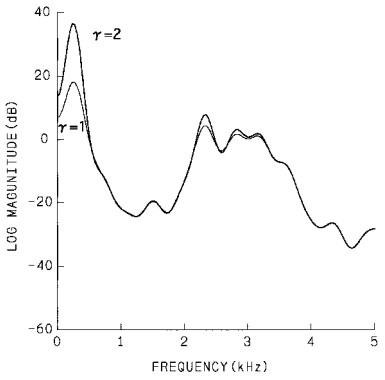

図３-(b)　スペクトル強度軸変化例（閾値=0）

図３にスペクトル強度軸の非線形変化例を示す。(a)が、スペクトル包絡の低レベル強度軸を圧縮した例で、(b)がスペクトル包絡のピーク域の強度レベルを強調した例である。

４．認識実験

本実験で使用する音声データは８人の男性がそれぞれ１０数字３回発声し、それに擬似的に自動車騒音を再現した自動車ノイズとピンクノイズを付加する。標準パターンには無雑音のものを用い、入力パターンには雑音を付加したものを用いた。

スペクトル変形はまず帯域制限と周波数変換を併用する。実験結果として、以下に全区間におけるスペクトル変形の認識率の変化例を示す。

図４は帯域制限スペクトルを使用せず、全区間において周波数軸変換を用いており、図５、図６はそれぞれ４kHz、３kHz 帯域制限スペクトルを周波数軸変換している。周波数軸変換係数はα=0～0.7 の間で変化させ、認識実験を行っている。

それぞれ pink はピンクノイズで、auto は自動車ノイズとする。

このように、帯域制限スペクトルの使用は自動車ノイズにはあまり効果は見られないが、ピンクノイズについては多少の認識率の向上が見られることが解る。このことは自動車ノイズが高周波域において標準の音声データにあまり影響を与えていないことに起因するものと思われる。

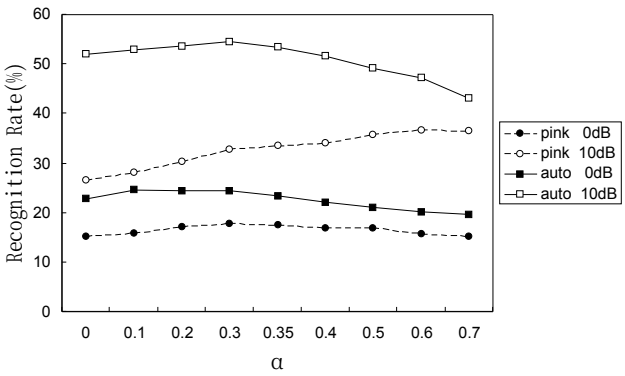

図４　周波数変換スペクトル
（帯域制限なし）の認識実験
αは周波数変換係数

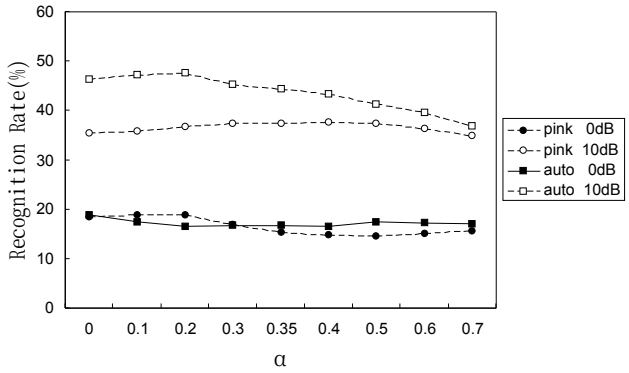

図５　周波数変換スペクトル
４kHz 帯域制限の認識実験

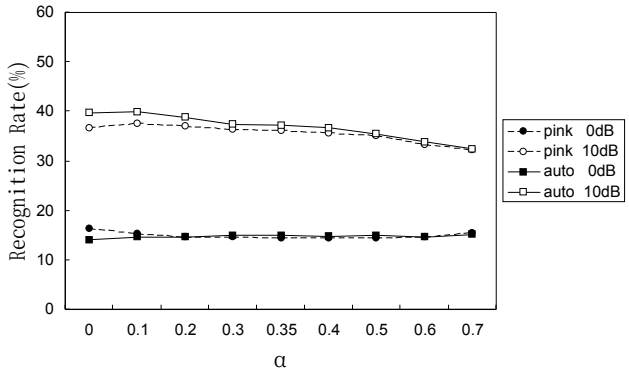

図６　周波数変換スペクトル
３kHz 帯域制限の認識実験

次に、子音部の場合、母音分と比べて高域周波数にその特徴がある可能性が大きいのでここで、子音部には低域周波数を強調する帯域制限をかけないで認識実験を行う。

つまり、音声の有声・無声判断をした後、有声部にのみ帯域制限をかけ、さらに周波数変換係数αを用いて変換する。そうすることによって認識率の向上を図る。ここで無声部のαは常にα=0 を保たせる。

また、ここで言う有声・無声判断は、無雑音のデータの最小位相ケプストラムを$c(i)$とするとき

$$UV = \sum_{i=0}^{3} c(i) \qquad (3)$$

とし、(3) 式で UV が-1.0 以上の場合を有声部、-1.0 より小さければ無声部と判断するものとする。

図７に帯域制限スペクトルを使用せず、有声部にのみ周波数軸変換を使用した認識率を示す。図８，図９、図１０はそれぞれ４kHz、３kHz、２kHz の帯域制限スペクトルを使用し、周波数軸変換を行っている。

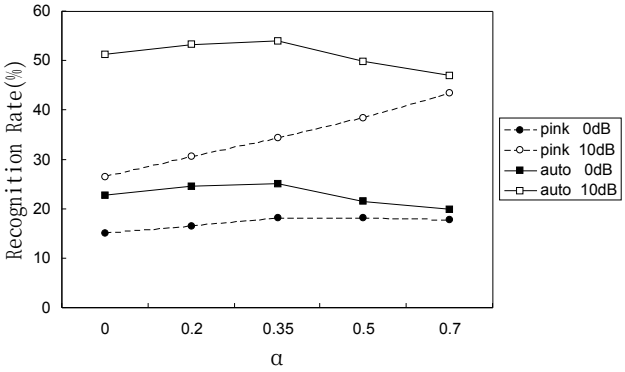

図７　有声部のみ周波数軸変換を
用いた認識結果（帯域制限なし）

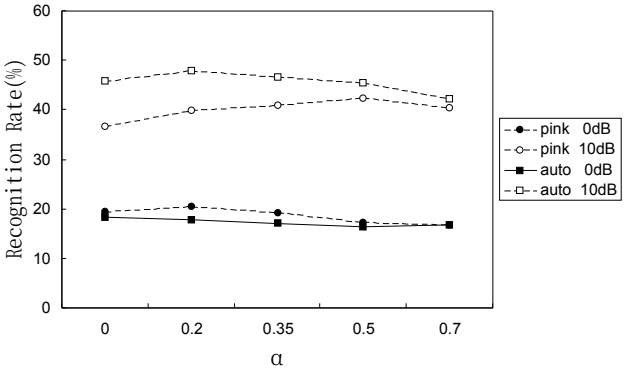

図８　４kHz 帯域制限スペクトルを用い、有声部
のみ周波数軸変換を行った認識結果

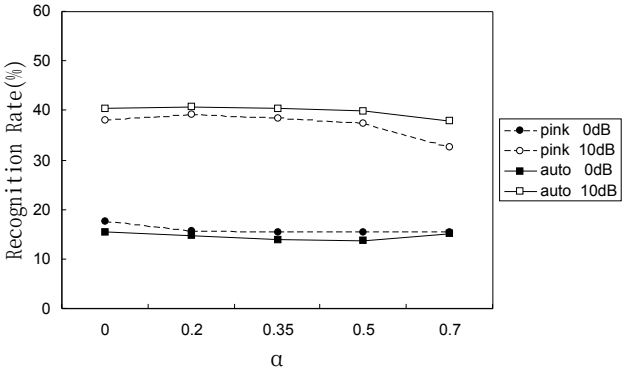

図９　３kHz 帯域制限スペクトルを用い、有声部
のみ周波数軸変換を行った認識結果

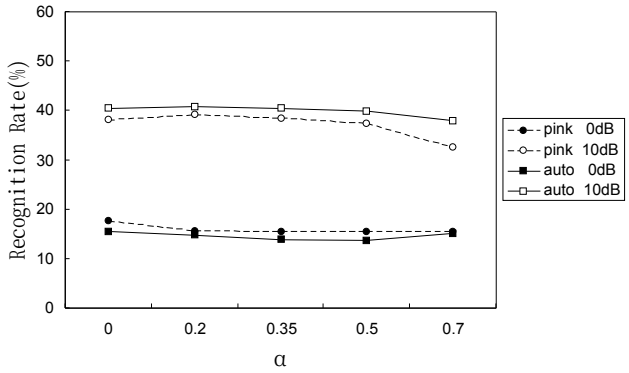

図１０　２kHz 帯域制限スペクトルを用い、有声部のみ周波数軸変換を行った認識結果

先の有声・無声判断を行わない実験より認識率の向上が見られる。しかしながらその効果は、僅かなもので、大きな効果は見込めない。

そこで図１０、図１１では、スペクトル包絡強度の非線形変換法を用いた認識結果を示す。

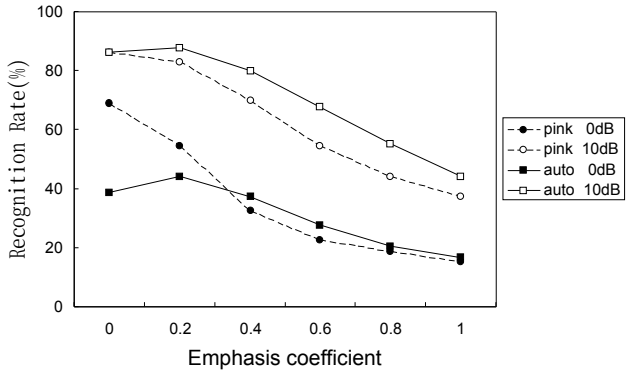

図１１　低レベル強度軸を圧縮した場合の認識結果（閾値=0dB）

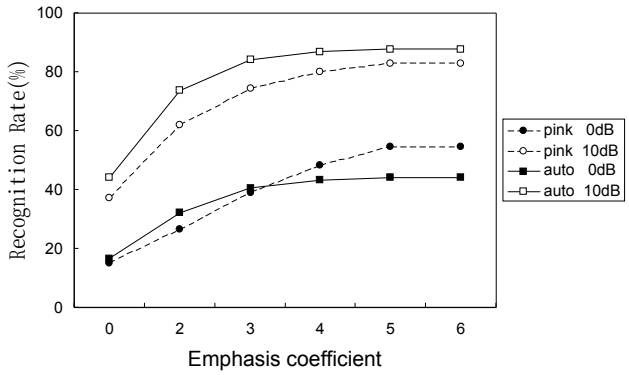

図１２　ピーク域の強度レベルを強調した場合の認識結果（閾値=0dB）

５．むすび

本報告では、雑音環境下の音声認識における低域強調である時変周波数変換と、帯域制限ケプストラムを用いた認識実験の併用とスペクトル強度軸の非線形変換法を用いた認識結果について比較を行った。

その結果、低域強調については、ピンクノイズについては有効であるが、自動車ノイズについてはあまり効果がないことが解った。また、スペクトル強度軸の非線形変換と低域強調の併用についてはその効果は明らかで、今後、低レベル強度軸を圧縮とピーク域の強度レベルの強調の併用などの実験も進行中である。

６．参考文献

[１] 秋田、長尾：“時変低域強調を用いた雑音環境音声認識”秋音講論 2-Q-15,pp,159-160(1996)

[２] A.V.Oppenheim and D.H.Johnson：“DiscreteRepresenration of Signals“,Proc.IEEE60 pp.681-691(1989)

[３] 秋田、玉井：“時変部分周波数特徴を用いた雑音環境音声認識”信学技報 EA99-79(1999)