

◎淵脇陽介, 荒井隆行, △穴見繁幸, △中島健晴, 村原雄二(上智大・理工)

1. はじめに

音声強調はさまざまな音声信号処理を実用化の際の前処理として重要なものであり、Bollによって提唱されたスペクトルサブトラクション法(SS法)[1]はその代表的なものである。人間の聴覚は優れた雑音抑圧機構を備えているため、これを音声強調に組み込むことでその効果が高まることが期待される。一方、Hermanskyは臨界帯域、等ラウドネス曲線、べき法則の3つの聴覚特性を用いた線形予測法PLP (perceptual linear prediction) [2]を提案した。

本報告ではこのPLPで用いられている聴覚特性を用いてSS法を拡張した音声強調アルゴリズムを提案する。さらに、10種類の音声にcar noiseを加えた音声を手手法を用いて処理を行い、そのSNR<sub>seg</sub>を測定する実験を行うことによりこのアルゴリズムの有効性を検討した。

2. アルゴリズム

SS法に基づいて拡張したアルゴリズムのブロックダイアグラムをFig.1に示す。

2.1 臨界帯域に基づく帯域通過フィルタ

雑音を付加した音声信号は帯域通過フィルタバンクにより18のチャンネルに帯域分割する。この帯域通過フィルタバンクの特性は臨界帯域に基づいて設計し、各帯域の帯域幅が中心周波数における臨界帯域幅に一致するように定めた。

2.2 等ラウドネス曲線による変換

人間が感じる音の大きさには周波数依存性があり、同じ強さの音でも周波数が異なると感じる音の大きさも異なる。等ラウド

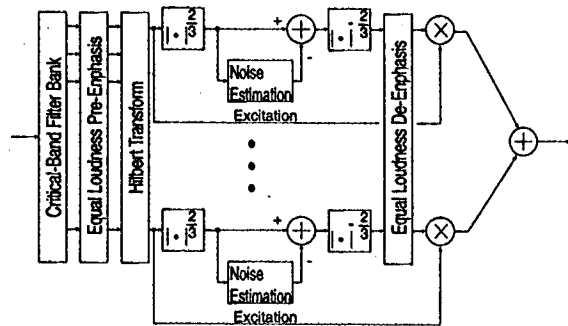


Fig.1 Block diagram of PLP-based spectral subtraction

ネス曲線は実験的に求められた音の大きさの等感曲線である。本実験では聴覚の周波数依存性を模倣するため、各チャンネルの振幅を中心周波数に応じて等ラウドネス曲線に対応する尺度に変換した。

2.3 包絡情報の抽出

包絡情報の抽出はヒルベルト変換を用いて行なった。ヒルベルト変換によって信号は包絡情報と位相情報に分離される。分離された信号のうち包絡情報に対してのみ処理を行い、位相情報は処理せず、信号を再合成する際の位相情報としている。

2.4 べき法則

Stevensによれば人間の聴覚が受ける感覚量は刺激量の2/3乗に比例する[3]。そこで得られた包絡情報を2/3乗することで処理にべき法則を組み込んだ。

2.5 雑音の予測と減算

雑音の予測は非音声区間における振幅の平均を用いた。各チャンネルごとに雑音の予測を行い、それぞれのチャンネルにおいて包絡情報から予測した雑音を減算することで音声信号を強調する。

\* Speech Enhancement Based on the Human Auditory System Used in PLP  
by Y. Fuchiwaki, T. Arai, S. Anami, T. Nakajima, Y. Murahara (Sophia University)

減算後の包絡情報が負である場合は雑音が過大に推定されているため、負の包絡情報については0にした。

## 2.6 音声波形の再合成

包絡情報は等ラウドネス曲線、べき法則を組み込むために振幅尺度の変換がされているため、それぞれの逆変換をおこなうことで元の尺度に戻した。その後、各帯域で包絡情報に位相情報をかけ合わせる。ここで位相情報は、元の信号の位相情報を用いる。そして、各帯域で得られた信号を加算し、強調後の音声信号を合成する。

## 3. 実験結果及び検討

実験は10種類の音声に $SNR_{seg}$ が-6, -4, -2, 0, 2, 4, 6 dBとなるようによくcar noiseを加えたものを用いた。これらの音声それぞれを今回提案する手法とSS法の2通りで処理し、処理後の信号の $SNR_{seg}$ を測定した。入力信号の $SNR_{seg}$ ごとに処理後の信号の $SNR_{seg}$ を平均した結果をFig.2に示す。

この結果を見ると今回用いた手法は $SNR_{seg}$ については従来のSS法に比べ、改善がみられなかった。この理由としてはべき法則を組み込むために2/3乗してから雑音の減算を行っていることがあげられる。予

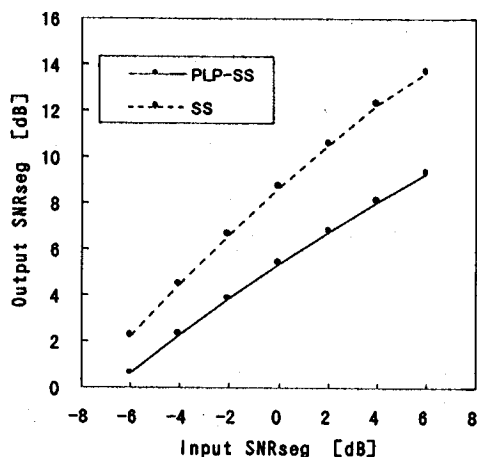


Fig.2 PLP-based spectral subtraction  
 $SNR_{seg}$  comparison

測した雑音減算した時点では取り出したい信号との差がもっとも小さかったとしても、べき法則の逆変換を行う際に3/2乗するため、誤差が広がってしまうことが考えられる。

しかし、Fig.2を見ると入力の $SNR_{seg}$ が低いほど従来法と結果が近くなる傾向があるため、より低い $SNR_{seg}$ 環境では結果が逆転する可能性がある。そのため、より劣悪な雑音環境についても実験を行う必要があるといえる。

## 4. まとめ

今回はSS法を聴覚特性を用いて拡張した手法を提案し、その有効性を $SNR_{seg}$ を用いて検証した。その結果、 $SNR_{seg}$ には改善が見られなかった。しかし本手法は聴覚の特性を組み込んでいるため、聴取実験により聞こえに対しての有効性についても検討する必要がある。

また、car noise以外の雑音環境についても実験を行い、その結果にどのような違いが現れるかについて調べることも今後の課題である。

## 参考文献

- [1] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. ASSP-27, No.2, pp.113-120, 1979.
- [2] H. Hermansky, "Perceptual Linear Predictive (PLP) analysis of speech," J. Acoust. Soc. Am., 87 (4), pp. 1738-1752, April 1990.
- [3] B. C. J. Moore, "An Introduction to the Psychology of Hearing (3rd edition)," Academic Press, 1989.
- [4] P. M. Clarkson, and S. F. Bahgat, "Envelope Expansion Methods for Speech Enhancement," J. Acoust. Soc. Am., 89, pp. 1378-1382, 1991.