定常部抑圧処理のリアルタイム化へ向けて －ＤＳＰによる開発－

後藤 崇公、荒井 隆行、安 啓一
上智大学 大学院 理工学研究科 電気・電子工学専攻

Abstract

実際のコンサートホールや、多目的ホールにおいて音声明瞭度を減少させる一つの原因として、残響環境下では音声中の音素が時間的に後方に伸びる尾を伴い、その残響の尾が後の音素にかかってしまうために起こるoverlap-masking が報告されている[5]。

荒井らは音声の定常部を抑圧することでそのマスキング量を減らし、音声明瞭度の低下を防ぐ前処理技術を以前から提案している[1,2]。また、程島らは聴取実験において予めコンピュータ上で残響をたたみ込んだ音声に対しこの定常部抑圧処理がどれほどの効果をもたらすかを調べ、特定の残響時間においてこの処理が有効であることを報告している[3]。また、今までは実験室環境での聴取実験であったのに対し、後藤らは実際のホールでの実験を行い、荒井らの提唱している定常部抑圧処理はコンピュータ上で人工的にたたみ込まれた残響だけでなく、実際のホールの残響環境下においても音声明瞭度を保持するための前処理として有効であることを改めて確認した[4]。

我々は次にDSPにより定常部抑圧処理のリアルタイム化を進めている。リアルタイム化が実現することによって、多目的ホールの音響システムに組み込むことが可能となり、残響に強い音声を提示することができるようになる。そして、社会の高齢化とともに聞こえを改善する処理の需要は高くなってきている昨今の状況を踏まえ、「音のバリアフリー」を実現したいと考えている。

本論文では、DSPにより定常部抑圧処理のリアルタイム化を検討し、FFTに基づくフレーム処理によるアルゴリズムを完成させた。また それを用いて現在 TI DSP TMS320C6713DSKボードにて開発を試みた。

1. Introduction

1-1 定常部抑圧処理について

室内音響おいて残響の影響により音声の明瞭度が損なわれる原因として overlap-masking があげられる[4]。図 1 に音声が overlap-masking の影響を受けている様子を示した。こちらの音声は"October"と発話されている。左側が原音声で右側がそれに残響がかかった音声波形である(残響時間 1.1s)。右側の音素で、特にエネルギーの低い子音/k/, /t/, /b/ は前のセグメントにあるエネルギーが高い母音の残響によりマスクされているのがわかる。この overlap-masking 効果により残響環境下において音声の明瞭度が損なわれるのである。

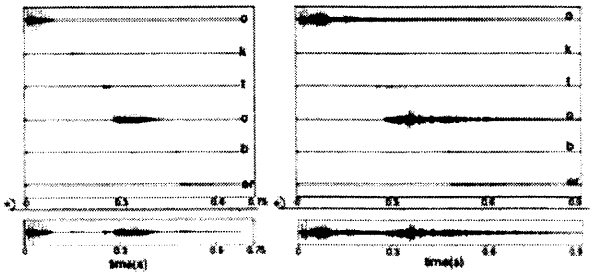

図 1 : "October" による overlap-masking 効果[4]

荒井らは overlap-masking による音声明瞭度を改善する前処理である「定常部抑圧処理」を提案している[1,2]。前処理とはホールなどにおいてスピーカから音声が放射される前に、ある信号処理を施すことによって残響の影響を受けづらい音声信号を作ることを意味する。特にこの処理では比較的エネルギーが高く、人間の音声情報処理に比較的重要でない母音の定常部のエネルギーを抑圧することにより、後続音素にかかるマスキング量を減らすことを目的としている。

1-2 従来の定常部抑圧処理

図 2 にあるように、従来の定常部抑圧処理は帯域分けによる処理を採用している。入力音声を 1/3 オクターブバンドにわけ、そこからヒルベルト変換により音声の包絡成分を取り出し、そこでダウンサンプルを施し 5 点間での回帰係数をもとめ、すべてのバンドに対してその係数の 2 乗和をとり、閾値 D を求める[6]。D 値がある閾値よりも低いか高いかでその音声部が定常部か否かを決め、定常部であるならばその振幅を例えば40%にまで抑圧する。なお、この抑圧率 40%は経験値である。

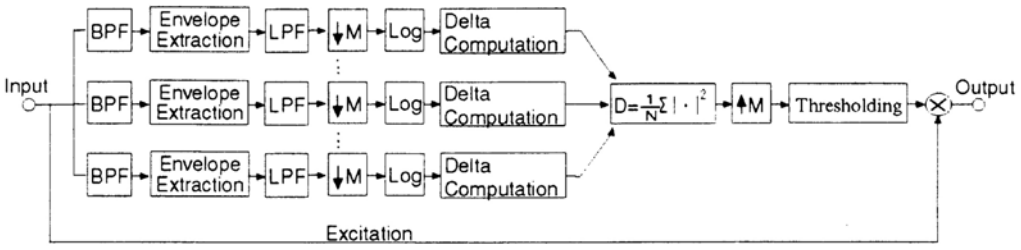

図 2: 「定常部抑圧処理」従来法ブロックダイアグラム[4]

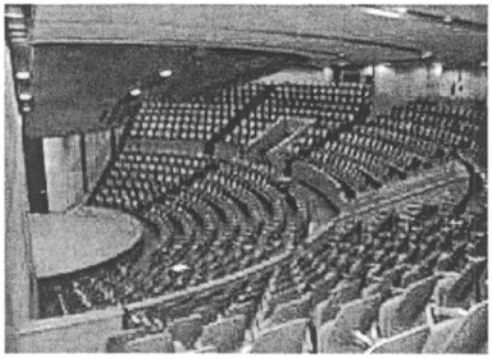

図 3: 上智大学 10 号館講堂[4]

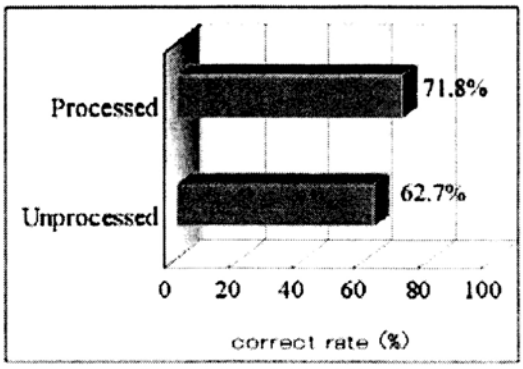

図 4: 聴取実験効果[4]

1-3 処理の効果

処理の効果を調べるために上智大学講堂(図３)にて聴取実験を行った。聴取実験は単音節明瞭度試験とし、単音節の母音は常に/a/とするのに対し、子音は/p/, /t/, /k/, /b/, /d/, /g/, /s/, /ʃ/, /h/, /dz/, / dʒ/, / tʃ /, /m/, /n/とした。計 14 単音節を処理あり・なしでランダムに 2 回繰り返し 56 刺激提示し、被験者(男性 12 名、女性 12 名)にどのように聞こえたかを解答してもらった。図 4 はそのときの正解率を処理ありと、処理なしを比較した結果である。処理ありの音声のほうが 9.1%聞こえが改善する結果になった($p<0.001$)。

このことを踏まえ、社会の高齢化とともに深刻化する残響環境下における音声の聞こえの問題に対し、我々の研究室ではこの定常部抑圧処理のリアルタイムシステムを提案できると考えた。以下にこの処理をリアルタイム化するための検討を行う。

2. Signal Processing

2-1 リアルタイム化の実現へむけて

従来のフィルタバンク法では音声ファイルを読み込み全体の音声信号に対して解析していたが、リアルタイム処理を考えるとこの手法は不利である。音声をフレームに分けて同じアルゴリズムと考えることができるが、そのままでは長い遅延時間が生じることから、FFT(高速フーリエ変換)に基づく周波数解析により母音の定常部を選出する同様のアルゴリズムを確立させた。

2-2 処理

図 5 に FFT に基づく定常部抑圧処理のブロックダイアグラムを示す。Ａ／Ｄ変換器により 16kHz 標本化でディジタル化された音声信号に対し、ハミング窓によってフレーム分けを施した。フレーム長は 20 ms で、フレームシフトは 10 ms とし、隣接するフレームは互いに 50%オーバラップするようにした。

次に、フレームごとに FFT を施し対数スペクトルを求めた。このとき、パワースペクトルを得るために、パワースペクトルをデシベルの尺度に変換して、出力の対数スペクトルとした。

次に、対数スペクトルに対して IFFT を施すことでケプストラム係数を生成する。生成されたケプストラム係数のうち、低い次元の係数が音声信号のスペクトル包絡を表す。そこで、ケプストラム係数に対しリフタリングを施すことにより、30 次までのケプストラム係数を取り出し次の解析に用いる。

後に各ケプストラム係数の時間軌跡に対し

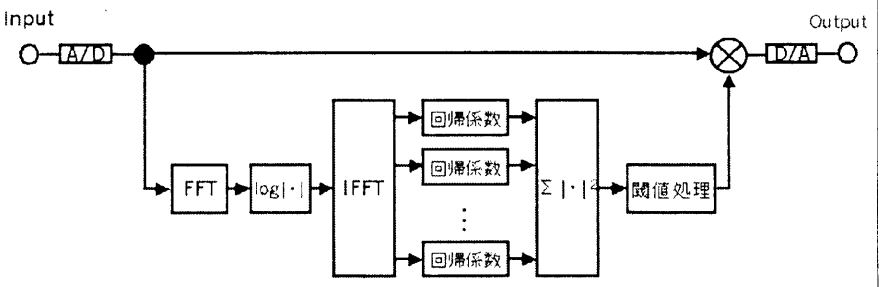

図 5：FFT ベースの定常部抑圧処理

て前後２点、計５点の回帰係数をフレーム毎に最小二乗法により計算する。ここでは連続する 5 つのフレームに対してそのケプストラム係数の時間軌跡を考える。

次に回帰係数の２乗平均 D を計算する。D 値は従来の処理同様、音声信号のスペクトル遷移を示すパラメータであり、フレーム毎に一つずつ求まる。D 値を閾値処理することで音声信号の定常部を求め、定常部と判定された区間に対して音声信号の振幅を 40%抑圧し、D／A 変換器を介して出力する。

従来法では音声信号をフィルタバンクによって帯域分割してから同様の処理を行っていたが、それに比べて FFT に基づく手法は処理遅延が非常に短く、実時間処理に適している。図 5 に示した音声処理装置は、A／D 変換器の出力から D／A 変換器までをＤＳＰ(Digital Signal Processor)を用いてリアルタイム処理することを最終目標とする。

3. DSP Implementation

3-1.開発環境

リアルタイム化へ向けて開発された FFT に基づく処理の検証は、MatLab でのシミュレーションによって行った。そこで DSP 実装に向けて、以下にどのようなプログラム設計で行うかを示す。

ハードウェアの仕様

- サンプリング周波数 A/D・D/A　8kHz
- McBSP からの割り込みにより EDMA を動かしデータの入出力を行う。またダブルバッファリングにより音声を入出力しているときと同時に音声処理をする設計(図 6 参照)

ソフトウェアの仕様

- 音声データ型 16bit short 型
- 処理系のバッファと元音声のバッファを定義し、抑圧は元音声のバッファに対して行う。
- 1 frame 音声データ 20ms のち 10ms オーバラップさせた音声データを処理系バッファに入力。オーバラップメソッド。
- a) 256 点 FFT( )　TI より提供されている cfftr2_dit( )を使用
- b) 20*Log10(|FFT|)を計算するメソッド
- c) 256 点 IFFT( )　TI より提供されている icfftr2_dif( )を使用
- 回帰係数を計算するメソッド
- D 値の閾値処理をするメソッド

以上 a), b), c)はフレームごとに処理をし、5 フレームの計算が終わった時点で回帰係数を計算し、D 値の閾値処理をする。その結果を踏まえ、元音声のバッファにある音声データ 480Samples を抑圧する・しないを決定する。(詳しい内部設計は図 7 参照)

図６より、EDMA とダブルバッファを採用していることで処理にかける時間を長く持つことが出来る。この仕様では McBSP に INT11 の割り込みが入ると EDMA が働き、受け側のバッファにデータを書き込んでいる最中に、CPU では定常部抑圧処理をするプログラムをまわすことが可能となる。データの書き出しも同じである。そこで、許される CPU 側の処理時間はサンプリング周波数が 8kHz なので 480samples を蓄える時間つまり 60ms 以内となる。60ms 以内で図 7 にある定常部抑圧処理の処理系ルーチンが処理できればよいことになる。そこで、それぞれのメソッドがどのくらいのサイクル数で回っているかを調査することにした。

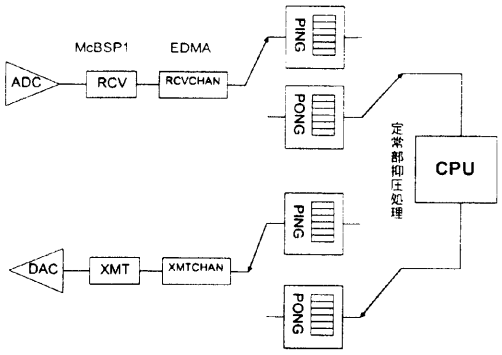

図６：ダブルバッファリング

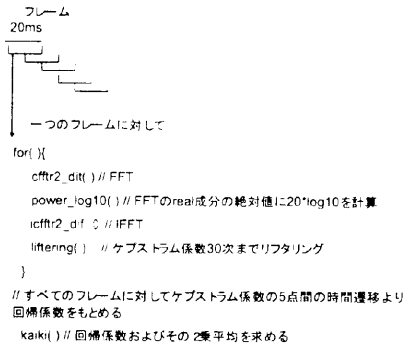

図７：定常部抑圧処理、内部設計

各メソッドのサイクル数を測る方法としてTimer を用いた手法にて行った。具体的にはDSP/BIOS の STS オブジェクトを用い CLK マネージャにより Timer の時間を調べる。今回は CLK マネージャの高解像度 Clock を測る関数 CLK_gethtime( )、STS_delta( )関数を用い各メソッドにかかるサイクル数を測定した。このとき C67 系のデバイスなので STS_delta( ) の値を 4 倍したものを実際のかかったサイクル数とした。

1. Overlap( ) = 4338 サイクル
2. Cfftr2_dit( ) = 4885 サイクル
3. Power_log10( ) = 391819 サイクル
4. Cifftr2_dif( ) = 4648 サイクル
5. liftering( ) = 507 サイクル
6. kaiki( ) = 1437 サイクル

2ー5 は for ループの中に入っているので実際にはこのサイクル数の 5 倍を要する。
その事を考慮し、全体のサイクル数は 2015070 で、連続する 5 フレームの音声が定常部かどうかという判断をする計算にかかるサイクル数が求まった。これから定常部と判断し、もしその場合は元音声のバッファにあるデータを 0.4 倍して出力側に渡す作業がのこっており、そこのサイクル数を同様に計算すると 3666 サイクルであった。

以上より全体のサイクル数は 2018736 サイクルとなり C6713 の動作周波数 225MHz より計算にかかる時間は約 9.0ms と計算できる。

よって、上限の 60ms を大きく下回っており、定常部抑圧処理の C6713 上での実装は安定して動いている事が立証された。

４．まとめ

我々が培ってきた定常部抑圧処理のリアルタイム化が DSP を用いて実現された。まだ少々デバッグが必要な箇所等あるが今後、実際に DSP を用いて今までと同様の聴取実験を行いこの処理のリアルタイム性が何処まで妥当かを調べたい。

５．謝辞

本研究の遂行にあたり技術的なサポートをしてくださいました日本テキサス・インスツルメンツ社ASP事業部DSP製品部アプリケーショングループの鴫谷さんに深く感謝するとともに、インターンシップの時に素晴らしい成長の機会を与えてくださったアプリケーショングループの方々、グループ長の河野さん、人事の竹村さんに深く感謝します。また、定常部抑圧処理に関して助言をくださったTOAの栗栖さんに感謝をいたします。

参考文献

[1] 荒井隆行・木下慶介・程島奈緒・楠本亜希子・喜田村朋子 “ 音声の定常部抑圧の残響に対する効果,” 日本音響学会講演論文集, pp. 449-450, 2001.10.

[2] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, “Effects on suppressing steady-state portions of speech on intelligibility in reverberant environments,” *Acoustical Science and Technology*, Vol. 23, No. 4, pp. 229-232, 2002.

[3] N. Hodoshima, T. Inoue, T. Arai and A. Kusumoto, “Suppressing steady-state portions of speech for improving intelligibility in various reverberant environments,” *Proc. of China-Japan Joint Conference on Acoustics* pp. 199-202, Nanjing, 2002.

[4] N. Hodoshima, T. Goto, N. Ohata, T. Inoue and T. Arai, “The effect of pre-processing for improving speech intelligibility in the Sophia University lecture hall,” *Proc. of the International Congress on Acoustics*, Vol. III, pp. 2389-2392, Kyoto, 2004.

[5] Bolt, R. H. and MacDonald, A. D., “Theory of speech masking by reverberation,” *J. Acoust. Soc. Am.*, 21: 577-580, 1949.

[6] Furui, S., “On the role of spectral transition for speech perception,” *J. Acoust. Soc. A*m., 80(4): 1016-1025, 1986.