

ステレオエコーキャンセラにおける 相互相関変動方式の検討

鈴木邦和* 杉山精* 阪内澄宇** 島内末廣** 牧野昭二**

*東京工芸大学大学院工学研究科

〒243-0297 神奈川県厚木市飯山1583

Tel:046-242-9520

E-mail:sugiyama@image.t-kougei.ac.jp

**NTTサイバースペース研究所

〒180-8585 東京都武蔵野市緑町3-9-11

Tel:0422-59-4207

E-mail:sakauchi.sumitaka@lab.ntt.co.jp

あらまし

ステレオ音声による拡声通話システムで必要となるステレオエコーキャンセラでは、ステレオの受話信号間の相互相関が高い場合が多く、適応フィルタが真のエコー経路に収束しない、真値への収束速度が遅い、といった問題がある。これらの問題を解決するために、人為的に相関を変動させる前処理方式が数多く提案されているが、これらの方式は音声に歪みを伴うという欠点がある。本報告では、実際の通話時における遠端の送話者の微小な移動に着目し、音像の定位を乱すことなく相互相関を変動させる方式を提案する。さらに聴覚特性を考慮した最適化により、収束性能の向上と音声品質保持の両立が可能であるという検討結果を示す。

キーワード

ステレオエコーキャンセラ, 相互相関, 音声品質, 聴覚特性

A study of decorrelation on a stereo echo canceller

Kuniyasu SUZUKI * Kiyoshi SUGIYAMA*

Sumitaka SAKAUCHI** Suehiro SHIMAUCHI** Shoji MAKINO**

*Tokyo Institute of Polytechnics

1583 Iiyama, Atsugi-shi, Kanagawa 243-0297, Japan

Tel:+81-46-242-9520

E-mail:sugiyama@image.t-kougei.ac.jp

**NTT Cyber Space Laboratories

3-9-11 Midori-cho, Musashino-shi, Tokyo 180-8585, Japan

Tel:+81-422-59-4207

E-mail:sakauchi.sumitaka@lab.ntt.co.jp

Abstract

A stereo echo canceller is required for a stereo teleconferencing system. The main problems are that the adaptive filters often misconverge or, if not, convergence speeds are very slow because of the cross-correlation between stereo signals. Several pre-processing methods which decorrelate stereo signals in order to overcome this problem have been proposed. But these methods introduce distortion resulting in low speech quality. In this paper, we focus on tiny movement of far-end talker and propose a new method of decorrelating with stereo signals without any confusion in sound image localization. We show that convergence can be further improved and speech quality maintained by optimizing using the characteristics of auditory perception.

key words

stereo echo canceller, cross-correlation, speech quality, auditory perception

1 はじめに

ハンドセットやヘッドセットの代わりに、スピーカとマイクロホンを使用する拡声通話系は、テレビ会議システムなどで、より自然な通話を実現している。この様な拡声通話には、エコーやハウリングを低減するためのエコーキャンセラが不可欠である。エコーキャンセラは、スピーカからマイクロホンまでのエコー経路のインパルス応答を適応フィルタにより推定して、疑似エコーを生成し、真のエコーから差し引くことにより、エコーを消去する。更に、ステレオ音声による拡声通話システムは、より臨場感の高い通話を可能とするが、同様にステレオエコーキャンセラ (SEC) が必要となる。しかし、SECでは、ステレオ受信信号間の相互相関が高い場合が多く、適応フィルタが真のエコー経路に収束しない、真値への収束速度が遅い、といった問題がある[1,2,3,4,5,6,7]。これら問題を解決するために、ステレオ信号間の相互相関を人為的に変動させる前処理方式が数多く提案されている[1,2,3,4]。文献[1]の方式は、左右の信号に互いに無相関な白色雑音を付加し、文献[2]の方式は、半波整流を付加している。文献[3]の方式では、音声の振幅に比例した大きさの白色雑音を付加している。また文献[4]の方式は、一方のチャンネルの信号を数千サンプル毎に、1サンプル遅延ありと遅延なしを切り替える。図1に相互相関変動前処理を含んだSECの構成を示す。既提案方式の多くが音声に雑音や歪みなどを付加し、スピーカから音を出すため、音声品質を劣化させる原因となっており、適応フィルタの収束特性と音声品質保持の両立が難しい。

そこで本報告では、遠端の送話者が微小に移動をしながら発話する際、受話者は音像の乱れを感じないが、マイクロホンに入力したステレオ信号間の相互相関が変動しているという現象に着目する。さらに、この送話者の移動を参考にして音声歪みを伴わない相互相関変動方式を提案する。

はじめに、遠端の送話者の微小な移動を模擬する信号を生成し、予備検討として主観評価実験による音声品質評価と、コンピュータシミュレーション

による収束特性評価を行い、送話者移動の有効性を明らかにする。次に、送話者移動を簡易に模擬する方法を提案し、更に聴覚特性に基づき、送話者位置別の最適化を検討する。

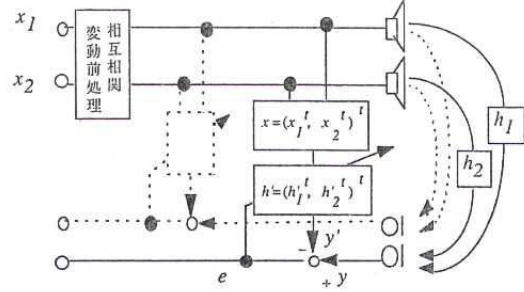


図1：ステレオエコーキャンセラの構成

2 ステレオエコーキャンセラの解不定の問題

SECには次のような問題点がある。図1に示すように、ステレオ拡声通話系では、左右信号 $x_1(k)=[x_1(k), x_1(k-1), \dots, x_1(k-L)]^T$ と $x_2(k)=[x_2(k), x_2(k-1), \dots, x_2(k-L)]^T$ (L は適応フィルタタップ数) とが、未知である真のエコー経路ベクトル $h_1(k)$, $h_2(k)$ を経ることによりエコー $y(k)$ は、

$$y(k) = h_1^T(k)x_1(k) + h_2^T(k)x_2(k) \quad (1)$$

として表せる。これに対してステレオエコーキャンセラでは、 $h_1(k)$, $h_2(k)$ を適応フィルタ $h_1'(k)$, $h_2'(k)$ により模擬し、疑似エコー $y'(k)$ を

$$y'(k) = h_1'^T(k)x_1(k) + h_2'^T(k)x_2(k) \quad (2)$$

として生成し、 $y(k)$ から $y'(k)$ を差し引くことにより、エコーを消去する。但し T はベクトルの転置を表す。しかし、実際のテレビ会議などにおいて、遠端の単一送話者の発話ではステレオ信号 $x_1(k)$, $x_2(k)$ の相互相関が高いため、次式

$$h_1'^T(k)x_1(k) + h_2'^T(k)x_2(k) = h_1^T(k)x_1(k) + h_2^T(k)x_2(k) \quad (3)$$

の一つの方程式では、 $h_1(k)$ と $h_2(k)$ の2つの未知系を求める事が難しい。この様に、ステレオ信号間の相互相関の影響により、エコー経路を模擬するための適応フィルタの解不定の問題があり、誤った特性に収束してしまう。解不定にならないためには、 $x_1(k)$ と $x_2(k)$ が無相関であるか、相互相関

の変動が必要である。

3 送話者の微小な移動とSECの収束特性

3.1 着眼点

実際のステレオ通話系において、送話者は無意識のうちに身体を微小に揺り動かしている。このような微小な移動で、送話者の音像の乱れは聴感上検知されにくい、ステレオ信号の相互相関は変動している。そこで、このような送話者の微小な移動に着目する。そして、聴感上問題とならない、送話者が移動できる最大幅（最大許容移動幅）を明らかにするため、予備実験（主観評価）を行う。さらにその結果を元に、適応フィルタの収束特性シミュレーションを行い、従来方式との収束性能比較を行う。

送話者の移動方向としては、左右のマイクロホンに正対した時、前後・左右・上下方向が考えられるが、ここでは、左右方向で基本原理を確認することとする(図2)。また、送話者位置によって音像の定位に影響しない移動幅が異なると予想されるので、図3の8カ所(A~H)の送話者位置について検討する。

予備実験及び収束特性シミュレーションに使用するための参照信号として、実際に移動している送話者の音声を用いると、移動幅などの正確な再現が困難である。そこで、Lcm幅の送話者の移動を以下の手法で模擬した。簡易無響室において各送話者位置にスピーカを設置し、そこからマイクロホンL及びRまでのインパルス応答(f_1, g_1)と、各送話者位置から左及び右にL/2cmの位置からマイクロホンL・Rまでのインパルス応答(f_2, f_3, g_2, g_3)を測定した(図2)。そして図4に示すように、3つの独立した時間窓(w_1, w_2, w_3)をかけた1チャンネル音声に、 f_1, f_2, f_3 及び g_1, g_2, g_3 を畳み込んだ後、足し合わせ、参照信号の左チャンネル(x'_1)、右チャンネル(x'_2)とした。なお、送話者の左右への移動周期は予備検討により聴感上と適応フィルタの収束特性の両面でバランスよく優れた性能を

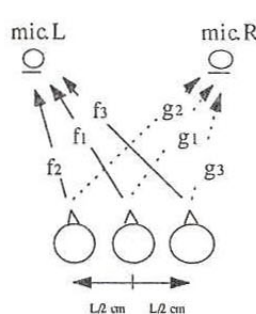


図2：送話者の移動

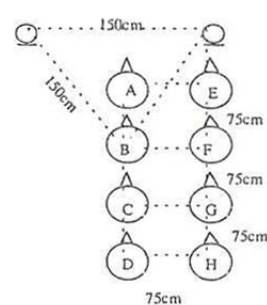


図3：送話者位置

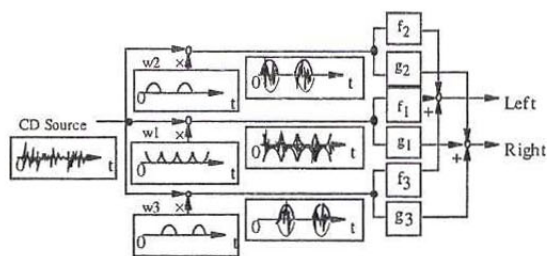


図4：送話者移動の模擬

持つ「2秒」とした。

3.2 最大許容移動幅の決定

各送話者位置における最大許容移動幅を明らかにするために予備実験を行った。ここで、所望の移動幅Lを得るために、インパルス応答 f_2, g_2, f_3, g_3 を、各送話者位置から左及び右にL/2cmの位置より測定した。簡易無響室において、図5に示すように、2つのスピーカを150cm離して設置し、評定者の頭部と2つのスピーカが正三角形を形成する位置で評価を行った。実験には7kHz帯域の音声を用いた。評定者は音声研究者男女10名である。評価は、音像の揺れと不自然さについて是非を問う2段階評価とした。ここでは、評定者の50%以上が容認した参照信号をもって、音像の定位に影響しない最大許容移動幅を決定した。

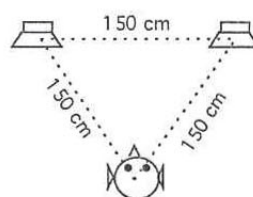


図5：実験システム

表1に、実験より得られた各送話者位置における最大許容移動幅を示す。この結果は、送話者位置によって音像の定位に影響しない許容移動幅が異なることを示している。

表1：送話者位置ごとの最大許容移動幅

送話者位置	移動幅(cm)	送話者位置	移動幅(cm)
A	5	E	60
B	10	F	40
C	10	G	40
D	20	H	60

続いて、収束性能の評価のために計算機による収束特性シミュレーションを行った。参照信号には、予備実験より明らかになった各送話者位置における最大許容移動幅での移動を模擬した信号を用いた。スピーカ・マイクロホン間のインパルス応答 h_1, h_2 を残響時間約200msの部屋において標準化周波数16kHzで測定し、これを模擬するフィルタ h'_1, h'_2 は各1000タップとした。式(4)に示す通り、評価は h'_1, h'_2 と h_1, h_2 との係数誤差 ϵ で行った。なお適応アルゴリズムとして2次のステレオ射影アルゴリズム[3]を使用した。

$$\epsilon = \frac{\|h_1 - h'_1\|^2 + \|h_2 - h'_2\|^2}{\|h_1\|^2 + \|h_2\|^2} \quad (4)$$

ここでは例として、送話者位置G(図3)における送話者移動を模擬した信号と、同じ位置Gで送話者が静止している場合に従来方式で処理した信号の収束特性を、図6に示す。従来方式としては、文献[2]の方式(MOS=2.8[5])、文献[4]の方式を用いた。図6の縦軸は係数誤差 ϵ [dB](収束性能)、横軸は所要時間である。図6から、送話者の移動は収束性能に対し有効であることが分かる。他の送話者位置においても、ほぼ同様の結果が得られた。

以上の結果から分かるように、音像の定位に影響しない送話者の移動は、ステレオ信号の相互相関を変動させ、真値への収束を向上させるために有効であることが確認できた。

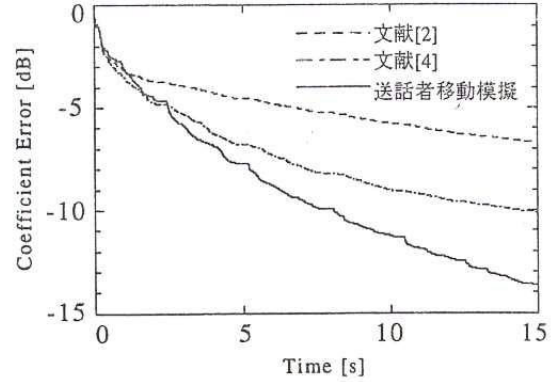


図6：送話者位置Gの収束特性

3.3 送話者移動模擬の簡易な実現

図4に示した方法は、モノラルで残響付加のないドライな音源に対する送話者移動模擬方法である。しかし、実際のSECで処理される受話信号は、ステレオ信号で発話位置からマイクロホンまでの空間的情報が含まれており、既に、送話者移動の中心である f_1 と g_1 が畳み込まれているものと同等である。図2の様に、送話者の左への移動の模擬を実現するには、厳密には $f_2f_1^{-1}, g_2g_1^{-1}$ というフィルタをかける必要がある。しかし、逆フィルタ f_1^{-1}, g_1^{-1} は不安定な場合もあり、算出方法も複雑である。そこで、 $f_2f_1^{-1}, g_2g_1^{-1}$ と同等の効果を簡易に実現できるフィルタ(F_{12}, G_{12})を畳み込む方式を検討する必要がある。ここでは、 F_{12}, G_{12} を、 f_1 と f_2 、 g_1 と g_2 の到達時間差と音圧レベル差によって近似し、簡易に送話者移動模擬を実現する方法を検討する。

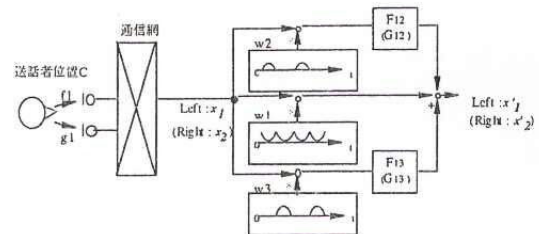


図7：送話者移動模擬の簡易な実現

送話者位置Cで発話したという空間的情報を含むステレオ信号 x_1, x_2 の各々に対し、図7に示すように、3つの独立した時間窓をかけ、到達時間差と音圧レベル差のフィルタをかけた後、足し合

わせて参照信号 x_1, x_2 を得る。この処理により得た信号に対する収束特性と、図4の処理により、送話者位置Cにおける移動を模擬した信号に対する収束特性を図8に示す。到達時間差と音圧レベル差で移動模擬を実現した場合でも、図4での送話者移動の模擬と、ほぼ同等な収束特性が得られた。なお、送話者位置ごとに最大許容移動幅が異なるため、図7でのF12~G13に、それらの位置に依存した異なるフィルタを用いることにより、より効果的な相互相関変動が可能である。

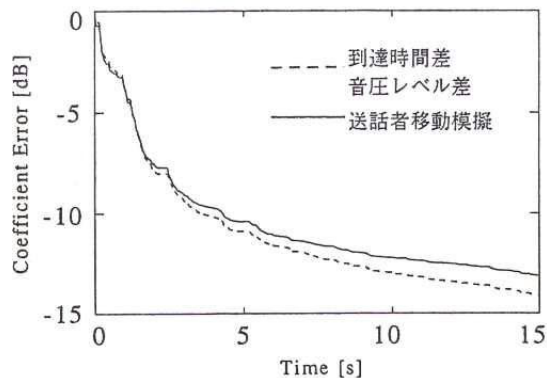


図8：到達時間差と音圧レベル差を用いた時の収束特性

4 聴覚特性を利用した提案方式

予備実験において、送話者移動模擬の簡易な実現により、送話者位置別に異なる処理ができ、音声品質を保ったまま優れた収束特性を得られることが分かった。しかし、

1. 図3の送話者位置A,B,C,Dの最大許容移動幅(表1)が与える到達時間差と音圧レベル差は全てほぼ同等である。
2. 図2に示すような左と右に同じ移動幅ではなく、左右に異なる移動幅を与える処理が有効な場合もある。
3. 到達時間差・音圧レベル差による音像の偏移方法に改善の余地がある。

という理由などから、次の様な方式を提案する。
遠端からの音声を、左右スピーカ上に定位する

音声と定位しない音声の2種類に分類する(4.1参照)。左右スピーカ上に定位する音声には、到達時間差を用いた処理を行い(4.3参照)、左右スピーカ上に定位しない音声には、到達時間差と音圧レベル差による音像偏移を互いに相補する(4.2参照)という方式を提案する。

4.1 左右スピーカ上に

定位する / 定位しない音声

ある音声は左右スピーカ上に定位するか否かは、受聴位置に依存する。最も一般的な2chステレオ通信システムにおけるマイクロホン・スピーカ・送話者及び受話者の各位置を図9に示す。

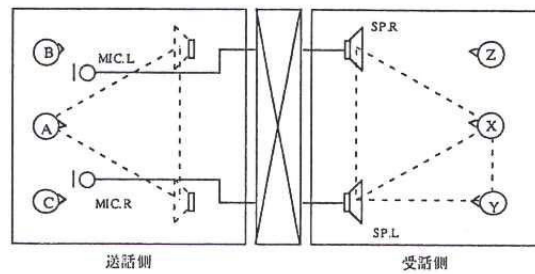


図9：2chステレオ通話システム

ここで、「左右スピーカ上に定位する音声」と「左右スピーカ上に定位しない音声」の定義をする。図9の送話位置A-B間のある位置からの音声に対し、受話位置X,Y,Zにおいて最も音像を内側に感じるのは受話位置Yである。従って、受話位置Yで、音声は右スピーカ上に定位していると感じる音声は、全ての受話位置で右スピーカ上に定位していると感じる。よって「左右スピーカ上に定位する音声」とは、音像を最も内側に感じる受話位置Y又はZにおいて右又は左スピーカ上に音像を感じる音声と定義する。それ以外の音声を「左右スピーカ上に定位しない音声」とする。

なお、図9の送話側Aの位置は、左右スピーカと正三角形をなす位置で、B,CはAの真横でスピーカの正面である。通常、参加者はA,B,Cのラインより後方に位置することが想定される。B,Cは最も前方隅の位置に相当する。受話側のX,Y,Zの位置についても同様である。

4.2 左右スピーカ上に定位しない

音声への処理

図10に示すように、実際のステレオ通話系において、左及び右マイクロホンに正対した送話者が左方向へ移動すると、移動前に比べ、左マイクロホンへ入力される音声は到達時間が早くなり、音圧レベルは大きくなる。ここでは、到達時間差と音圧レベル差が音像を同方向に偏移させるように変化している。これに対し例えば、音圧レベル差を到達時間差に対して左右逆に偏移させるとする。到達時間差と音圧レベル差による相関の変動は上記の場合と同等なので、収束性能は同じである。しかし、音像の偏移は互いに補償され通話品質の向上が期待できる。そこで、到達時間差と音圧レベル差各々の影響を互いに補償すること（相補）に着目する。相補は、等しい音像の偏移幅を与える到達時間差及び音圧レベル差を得て、互いに音像を逆方向に偏移させれば可能となる。

到達時間差及び音圧レベル差による音像偏移方法を図11に示す。マイクロホンで收音されたステレオ信号 x_1 , x_2 の各々に対し3つの独立した窓関数(w_1, w_2, w_3)及び、到達時間差又は音圧レベル差のフィルタ(T又はL)をかけた後、足し合わせて参照信号の左チャンネル(x'_1), 右チャンネル(x'_2)を得る。なお、音像偏移方向は、処理前の音像に対し左右両側とし、左右への偏移周期は2秒とした。

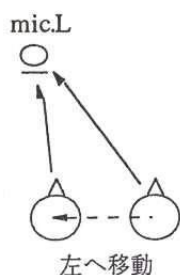


図10：送話者の移動

4.3 左右スピーカ上に定位する音声への処理

通常の2chステレオシステムでは、音像は左ス

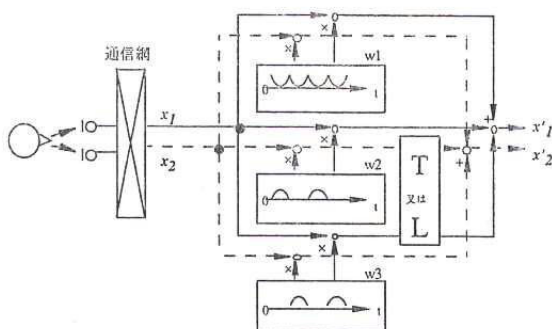


図11：音像偏移方法

ピーカと右スピーカ間の外側へは定位し難いことに着目する。遠端からの音声が入る左又は右スピーカ上に定位する場合、その音像がスピーカ間の外側へ移動するように左右信号の到達時間差・音圧レベル差を変化させる。実際には音像が左右スピーカ間の外側へは移動しないので、定位の乱れは起きにくく到達時間差・音圧レベル差をより大きく変化可能で、相互相関を大きく変動させられると予想される。

「左右スピーカ上に定位する音声」の中にも、図9の送話側Bの位置から、A-B間のある位置（受話位置Qとする）まで一定の範囲があると考えられる。ここではA-B間のある位置で、定位する/しない、の境界にある送話者位置Qの音声に対する、聴感上問題とならない到達時間差の最大量（最大許容変化量）を明らかにし、「左右スピーカ上に定位する音声」への処理に適用する。予備検討をして、次の事柄を確認・決定した。

1. 受話者位置Qの決定。
2. 到達時間差による遅延周期は聴感上2000サンプル程度が良い。
3. 1, 2に加え、音圧レベル変動を用いると大幅に音声品質が劣化する。

以上を参考にして決定した処理方法を図12に示す。

図12中の時間差のフィルタTは、音像が定位しているスピーカとは逆のスピーカに入力する信号に対して掛けられる。予備検討の確認事項によ

り、変動周期は2000サンプルとし、音圧レベル差のフィルタは用いない事とした。

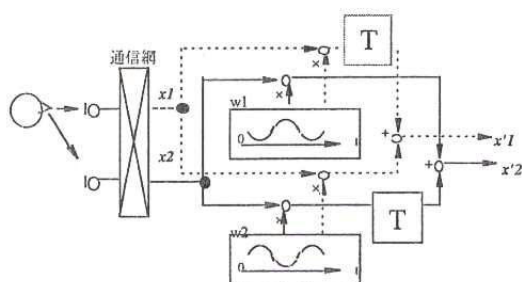


図12：時間差変動方法

4.5 主観評価実験

4.5.1 主観評価実験Ⅰ

まず、「左右のスピーカ上に定位しない音声」について、図11により得た参照信号を用いて、到達時間差及び音圧レベル差の最大許容変化量を明らかにするために主観評価実験を行った。簡易無響室において、2つのスピーカを150cm離して設置し、評定者の頭部と2つのスピーカが正三角形を形成する位置で評価を行った（図5）。評価に用いた音声は、図9のAの送話者位置の音声である。実験条件は表2に示す通りである。ここでは、評定者の50%以上が容認した参照信号をもって、音像の定位に影響しない最大許容変化量を決定した。これより得られた到達時間差及び音圧レベル差による音像偏移幅は、ほぼ等しいと仮定する。表3に実験より得られた結果を示す。また簡単な受聴テストにより、到達時間差と音圧レベル差で相補した音声は、送話者移動模擬よりも音声品質が向上している事を確認した。

表2：主観評価実験条件

音声帯域	100Hz - 7kHz
評定者数	音声研究者男女10名
評価方法	音像の揺れ、不自然さについては是非を問う2段階評価

表3：到達時間差及び音圧レベル差の最大許容変化量

到達時間差の変化量	3 サンプル (標準化周波数16kHz)
音圧レベル差の変化量	1.6倍

4.5.2 主観評価実験Ⅱ

続いて、「左右のスピーカ上に定位する音声」について、到達時間差の最大許容変化量を明らかにするために、同様に主観評価実験を行った。但し、評価に用いた音声は、図9の右スピーカ上に定位する音声（受話位置Q）で、評価受話位置はYである。表4に実験より得られた結果を示す。

表4：到達時間差の最大許容変化量

到達時間差の変化量	100サンプル (標準化周波数16kHz)
-----------	--------------------------

5 収束特性シミュレーション

5.1 実験条件

主観評価実験Ⅰ・Ⅱの結果を元に信号を生成し、ステレオエコーキャンセラの収束特性シミュレーションを行い、これらの信号と従来方式との収束性能との比較を行なった。参照信号には、「左右のスピーカ上に定位しない音声」には、主観評価実験Ⅰよりそれぞれ独立して決定した到達時間差及び音圧レベル差の最大許容変化量を使って相補した信号を用いた。また、「左右のスピーカ上に定位する音声」には、主観評価実験Ⅱより決定した到達時間差の最大許容変化量を使って生成した信号を用いた。ここでは実際の通話を想定し、図9の送話者位置A,B,Cそれぞれの送話者がA,B,Cの順番で5秒間ずつ発話するとした。スピーカ・マイクロホン間のインパルス応答 h_1, h_2 を残響時間200msの部屋において標準化周波数16kHzで測定し、これを模擬するフィルタ h'_1, h'_2 は各1000タップとした。式(4)に示す h'_1, h'_2 と h_1, h_2 との係数誤差 ϵ により評価を行う。なお、

適応アルゴリズムとして2次のステレオ射影アルゴリズム[3]を使用した。なおここでは、送話者位置判別は、左右ステレオ信号の音圧レベル比を検出して実現した。

5.2 実験結果

図9のA,B,Cの送話者が順番に発話した時の、提案方式・従来方式で処理した信号の収束特性を図13に示す。縦軸は係数誤差 ϵ [dB](収束性能)、横軸は所要時間である。これより、従来方式に比べ、提案方式の収束性能が優れている事が分かる。

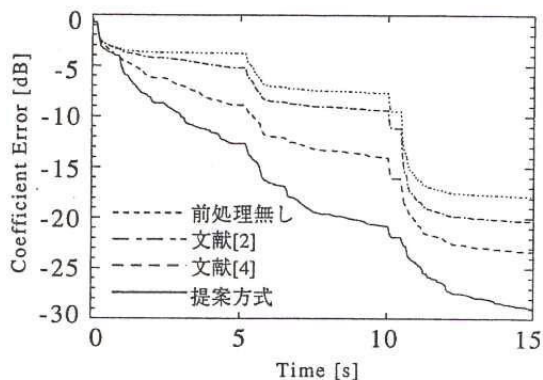


図13：提案方式と従来方式の収束特性

6 まとめ

通話品質を保ったまま、ステレオエコーキャンセラの収束性能の優れた相互相関変動方式を確立するために、実際の通話時における送話者の微小な移動に着目した。まず、送話者の微小な移動が相互相関を変動させ、真値推定に有効であることを確認した。次に、送話者の移動の模擬を、音の到達時間差と音圧レベル差で簡易に実現できることを示し、さらに聴覚特性を考慮することで、送話者別の最適な処理が可能な方式を提案した。そして、主観評価実験により、その提案方式の到達時間差及び音圧レベル差の最大許容変化量を決定し、最後にコンピュータシミュレーションにより、収束性能が従来方式よりも優れていることを明らかにした。

「謝辞」

日頃御指導頂くNTT研究所山森プロジェクトマネージャー、金田グループリーダーを始めとする研究グループの皆様に感謝致します。

「参考文献」

- [1] M. M. Sondhi, D. R. Morgan and J. Hall, "Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem", *IEEE SP letters*, vol. 2, no.8, pp. 148-151(Aug.1995).
- [2] J. Benesty, D. R. Morgan and M. M. Sondhi, "A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation", *Proc. ICASSP97*, pp. 303-306 (1997).
- [3] S. Shimauchi and S. Makino, "Stereo Prejection Echo Canceller with True Echo Path Estimation," *Proc. ICASSP95*, vol. 5, pp. 3059-3062 (1995).
- [4] Y. Joncour and A. Sugiyama, "A New Stereo Echo Canceller with Pre-processing of One Input Signal: I. Strict Identification of the Echo Paths", 信学ソ大, A-4-10 (1997).
- [5] 鈴木邦和, 阪内澄宇, 島内末廣, 羽田陽一, "ステレオエコーキャンセラにおける収束改善のための前処理方式の検討", 音講論集, 3-5-10 (Mar.1998).
- [6] 鈴木邦和, 阪内澄宇, 島内末廣, 牧野昭二, "音響系の変動に着目したステレオ信号の相関低減方法", 音講論集, 1-6-12 (Mar.1999).
- [7] 鈴木邦和, 阪内澄宇, 島内末廣, 牧野昭二, "音響系の変動に着目したステレオ信号の相関低減方法 - (第2報) 聴覚特性を考慮した最適化", 音講論集, 1-6-12 (Sep-Oct.1999).