

両チャンネルに前処理ユニットを有するステレオエコーキャンセラ

Stereo Echo Canceller with Pre-Processing Units in Both Channels

水野裕介 カズダフリ・ロラン[†] 杉山昭彦[‡]
平野晃宏 中山謙二

金沢大学大学院 自然科学研究科 電子情報工学専攻

[†] Eurecom Institute

[‡] NEC 共通基盤ソフトウェア研究所

Yuusuke MIZUNO Laurent Kazdaghli[†] Akihiko SUGIYAMA[‡]
Akihiro HIRANO Kenji NAKAYAMA

Division of Electronics and Computer Science

Graduate School of Natural Science and Technology, Kanazawa Univ.

[†] Eurecom Institute

[‡] Common Platform Software Res. Labs., NEC Corp.

E-mail: y_mizuno@leo.ec.t.kanazawa-u.ac.jp

Abstract

This paper proposes a new stereo echo canceller with pre-processing units in both channels. Periodic delays with different phases by the two pre-processing units provides an additional condition for echo cancellation, leading to faster convergence. Signal sliding in both channels guarantees symmetrical shift of a sound image around the original position for better subjective quality of the pre-processed signal. Simulation results show that the convergence speed of the proposed echo canceller is 20% faster than that of the conventional echo canceller. Subjective test results show that the proposed echo canceller achieved a 0.37 point higher grading difference in the ITU-R five-grade impairment scale than the conventional echo canceller with a statistically significant difference.

概要

前処理ユニットを両チャンネルに有する構成のステレオエコーキャンセラを提案する。2つの前処理ユニットは、チャンネル毎の遅延挿入を異なる位相で行うことで新たなエコー消去の条件をつくり、収束を高速化する。また、両チャンネルの信号を周期的に遅延させることで、音像が前処理を行う前の位置を中心に移動するため、前処理した信号の音質が優れている。シミュレーションによって、提案するエコーキャンセラは、従来のエコーキャンセラよりも、20%高速に収束することを示す。また、主観評価結果

を用いて、提案するエコーキャンセラにおけるITU-R 5段階劣化率の評価差分値が、統計的に有意な差で、従来のエコーキャンセラより0.37ポイント高くなることを示す。

1 まえがき

エコーキャンセラは、通信システムにおいてエコーを消去するため、広く使われている。エコーキャンセラが使用されるシステムの一例として、遠隔会議システムがある。遠隔会議システムを利用すれば、離れた場所においても同じ部屋にいるかのように会議を行うことができる。近年、遠隔会議システムでは、臨場感溢れる会議を行いたいという要求が高まっている。遠隔会議を臨場感溢れるものにするためには、しっかりした音像定位をもつように、システムを多チャンネル化、少なくともステレオ化が必要になる。遠隔会議システムをステレオ化すると、音響エコーを消去するためのエコーキャンセラもステレオ化に対応する必要がある。

ステレオエコーキャンセラの基本となる構成は、2つのスピーカーと2つのマイクの間にある4つのエコー経路を、4つの適応フィルタで推定する線形結合型である[1]。しかし、線形結合型のステレオエコーキャンセラでは、チャンネル間相互相関が強い入力信号に対して、正しいエコー経路を推定できないという問題がある[2]。そのため、話者の交替、移動などにより、遠端側の音響特性が変化したとき、ERLE(Echo Return Loss Enhancement)が低下する[4]。

この問題に対して、片方のチャンネルの信号を周期的に

遅延させる前処理ユニットを配置することにより、正しいエコー経路を推定する方法が提案されている [3][4]。しかしながら、強いチャンネル間相関がある信号に対するフィルタ係数の収束速度は十分ではない。また、片方のチャンネルのみに前処理が入るので、正面にある音像が片方向のみに移動して、音像の移動が知覚されやすい。

本稿では、前処理ユニットを両チャンネルに有する構成のステレオエコーキャンセラを提案する。従来法の前処理について 2 節で説明する。3 節で提案法の前処理を説明する。4 節で従来法と提案法の係数収束速度を比較したシミュレーション結果を示す。5 節で音質の主観評価結果を示す。

2 従来法の前処理

図 1 に、従来法の構成図を示す。4 つのエコー経路に対応した 4 つの適応フィルタを並べた線形結合型のエコーキャンセラ部に加えて、片チャンネルの信号を周期的に遅延させる前処理ユニットを有する。前処理ユニットは、周期 Q の関数にしたがって係数が変化し、2 タップの FIR フィルタである。図 2 に前処理フィルタの周期関数 $c_L(n)$ を示す。

$c_L(n) = 1$ のとき、図 1 の前処理ユニットは遅延なしの状態となり、入力信号がそのまま出力される。 $c_L(n) = 0$ のとき、入力信号は 1 サンプル遅れて出力される。 $c_L(n)$ を 0 から 1、または 1 から 0 へ滑らかに変化させることで、音質の劣化を避ける。この前処理により、遅延がある状態とない状態とで、エコーが消去されるための 2 つの条件を得ることができる。その 2 つの条件に共通の解が、真のエコー経路のインパルス応答となる。

従来法における係数収束の様子を図 3 に示す。2 本の直線は、それぞれ遅延がある状態とない状態の不定解集合を表しており、その交点が理想解である。ある初期値から NLMS アルゴリズムで適応フィルタを学習させ始めると、係数は不定解集合である直線に垂直に向かっていく [5]。遅延がない状態では (1) の直線に向かって学習を行い、遅延がある状態では (2) の直線に向かって学習を行う。これを繰り返しながら、適応フィルタの係数は、徐々に理想解に近付いていく。

従来法には、線形結合型のステレオエコーキャンセラにはなかった、音像の移動という問題がある。線形結合型のステレオエコーキャンセラでは、音像が正面から移動しなかった。しかし、従来法では片チャンネルの信号を周期的に遅延させる前処理ユニットを配置するため、片チャンネルの音が遅れ、音像が移動する。図 4 に、従来法における音像の移動を示す。図 4 の場合、遅延ユニットが左チャンネルに配置されているため、遅延があるときは右チャンネルの音が早く耳に届き、音像は向かって左の、あ

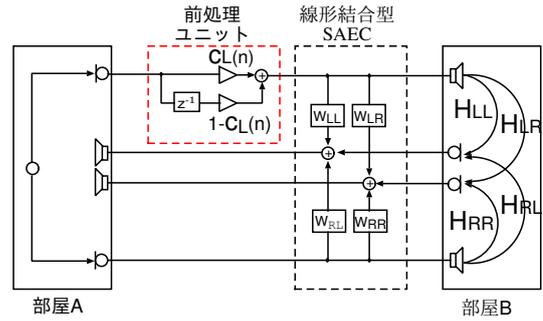


図 1: 従来法の構成図

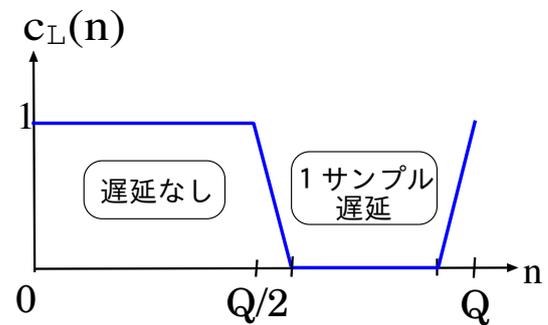


図 2: 周期関数 $c_L(n)$

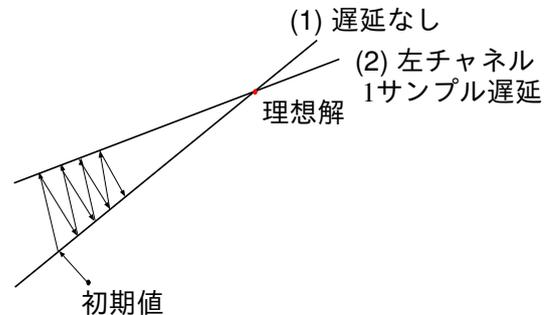


図 3: 従来法における係数収束の様子

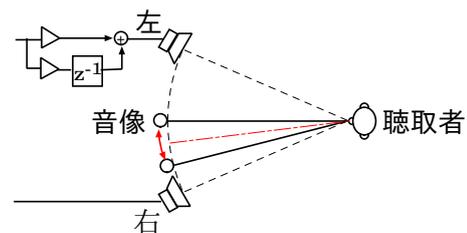


図 4: 音像の移動 (従来法)

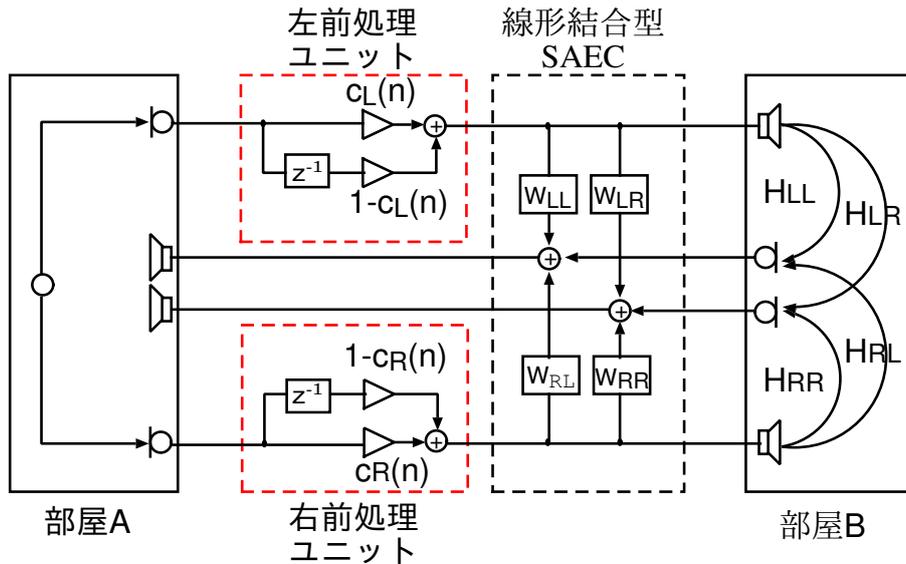


図 5: 提案法の構成図

る点に移動する。周期的に遅延がある状態とない状態を繰り返すため、音像は、正面と左に移動した点の間を往復する。このとき、図 4 の一点鎖線で示す通り、正面と左に移動した点の中間を軸に、音像が往復するため、知覚されやすい。

3 提案法の前処理

提案法は図 5 に示す通り、前処理ユニットを左右両方のチャンネルに有する。各チャンネルの前処理ユニットは、従来法と同様の周期関数にしたがって係数が変化する、2 タップの FIR フィルタである。図 6 に周期関数を示す。左チャンネルの周期関数 $c_L(n)$ は従来法と同じであるが、右チャンネルの周期関数 $c_R(n)$ は $c_L(n)$ の位相を $Q/4$ だけ進めている。

図 6 から、両チャンネル遅延なし、左チャンネル 1 サンプル遅延、両チャンネル 1 サンプル遅延、右チャンネル 1 サンプル遅延の 4 つの状態があることがわかる。両チャンネル遅延なしの状態と、両チャンネル 1 サンプル遅延の状態を、相対的に遅延がないとみなすと、状態は 3 つとなる。それに対して、従来法では遅延の有無という 2 状態しかないため、エコーが消去されるための条件数は 2 である。提案法では、前処理の両チャンネル化によって、エコーが消去されるための条件数が 3 が増えるので、収束の高速化が可能となる。

提案法における係数収束の様子を図 7 に示す。前処理ユニットが 1 つ増えたことで、新たな不定解集合である直線ができる。そして、(1) 遅延なし、(2) 左チャンネル 1 サンプル遅延、(3) 右チャンネル 1 サンプル遅延に対応する 3 つの直線の間を行き来しながら、理想解に近付いていく。

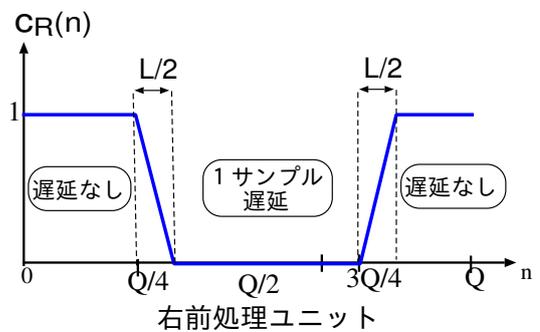
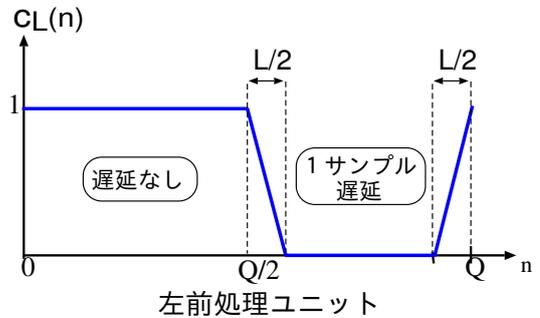


図 6: 周期関数 $c_L(n), c_R(n)$

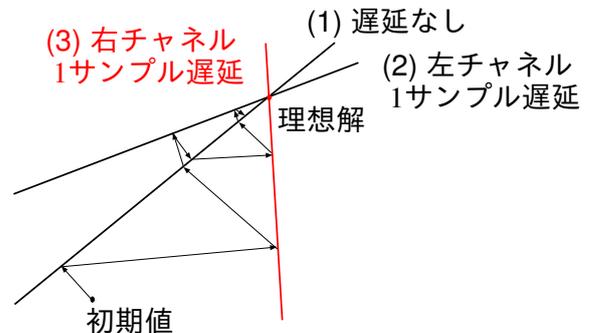


図 7: 提案法における係数収束の様子

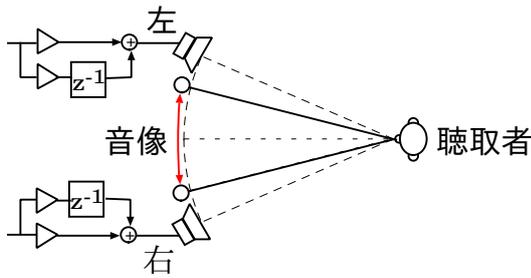
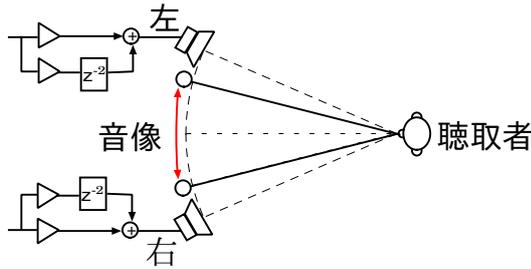
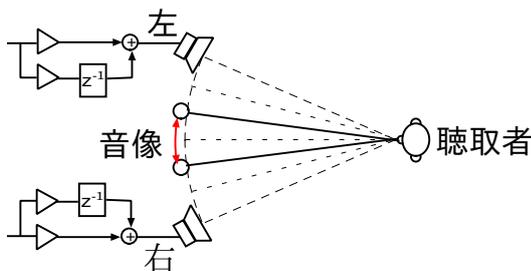


図 8: 音像の移動 (提案法)



(a) 2サンプル遅延の場合



(b) 1サンプル遅延の場合

図 9: 遅延サンプル数が音像の移動幅に与える影響

図 7 と図 3 と比べると、提案法の方がより速く理想解に収束する。

図 8 に提案法における音像の移動を示す。提案法では前処理ユニットが両チャンネルにあるので、音像は左右両方向に移動する。前処理を行う前の音像が正面にある場合は、移動の中心が正面に来るので、従来法に比べて音質が良くなる。ただし、両チャンネル化により音像の移動幅が従来の 2 倍に広がり、音像の移動が知覚されやすくなる。提案法では遅延量を半分にして、音像の移動幅を従来法と同じにすることで、音像の移動を知覚されにくくする。

図 9 に遅延サンプル数を変化させたときの音像の移動幅を示す。(a) が 2 サンプル遅延の場合、(b) が 1 サンプル遅延の場合であり、(b) では音像の移動幅が (a) の半分になる。そのため、従来法の 2 サンプル遅延と、提案法の 1 サンプル遅延では音像の移動幅が同じになり、移動の中心のみが変化する。

表 1: シミュレーション条件

方式	LC	LC-SPP1	LC-SPP2	LC-DPP
タップ数	64 または 1000			
ステップサイズ	0.5			
Q	なし	4000		
L	なし	400		
遅延サンプル	なし	1	2	1
背景雑音レベル	-40dB			

4 収束速度の客観評価

線形結合型 (Linear Combination structure: LC), 従来法 (LC structure with Single Pre-Processing unit: LC-SPP), 提案法 (LC structure with Double Pre-Processing units: LC-DPP) について、シミュレーションによって係数収束特性と ERLE を比較した。シミュレーションは、表 1 の条件で行った。入力信号には、サンプリング周波数 16kHz の音声 10 種類 (男声 5 種類, 女声 5 種類) を用いた。適応フィルタの係数更新は、NLMS アルゴリズムで行った。パラメータ Q , L は、遅延を入れる周期を決めるパラメータで、大きな値に設定することで音質の劣化を防ぐ。背景雑音として、エコーに対して -40dB の白色雑音を加えた。LC-SPP は、遅延サンプル数の異なる 2 例について、シミュレーションを行った。LC-SPP1 では 1 サンプル遅延、LC-SPP2 では 2 サンプル遅延を与えた。2 サンプル遅延の場合についてシミュレーションを行ったのは、音像の移動幅が提案法と等しい場合における従来法の性能を確認するためである。LC-DPP は 1 サンプル遅延としており、これは LC-SPP2 の音像の移動幅と等しくなる。

各方式において、係数が不定解に収束する問題を回避して、正しいエコー経路に収束しているか否かの評価指標として、係数誤差ベクトルのノルム (Norm of the Coefficient-Error Vector: NCEV) を用いた。NCEV は、エコー経路のインパルス応答とフィルタ係数ベクトルの誤差ノルムとして定義される。エコー経路を \mathbf{h} , 時刻 n におけるフィルタ係数を $\mathbf{w}(n)$ とすると、次の式で表される。

$$NCEV(n) = 10 \log_{10} \frac{\|\mathbf{h} - \mathbf{w}(n)\|^2}{\|\mathbf{h}\|^2} \quad (1)$$

エコー経路に 64 タップのインパルス応答を用いた場合の NCEV を図 10 に示す。LC では NCEV が、-4.0dB までしか低下しておらず、フィルタ係数が正しいエコー経路に収束していない。そのため、このままでは話者の交替な

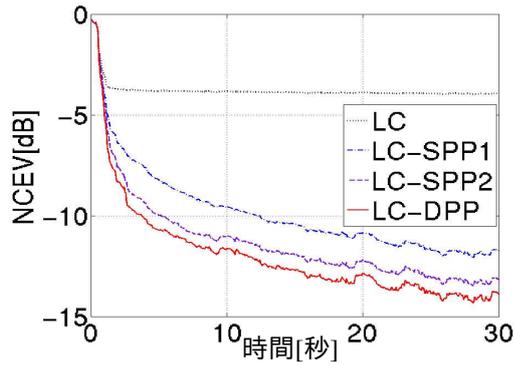


図 10: NCEV(64 タップ)

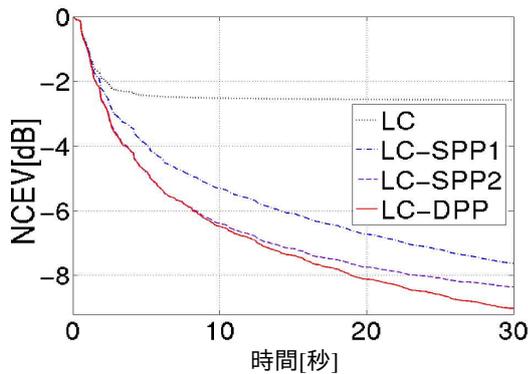


図 11: NCEV(1000 タップ)

どがあったとき、ERLE が低下する。LC-SPP1 は 30 秒後の NCEV が、 -12.1dB まで低下しており、正しいエコー経路に向かって収束している。遅延サンプル数を 2 サンプルとした LC-SPP2 は 30 秒後の NCEV が、 -13.5dB まで低下しており、遅延サンプル数が大きいほど収束が速くなるのが分かる。これは遅延サンプル数が 2 倍に増えたことで、1 サンプル遅延のときより左右のチャンネル間相関が低くなったことが原因である。ただし、LC-SPP2 は画像の移動幅が LC-SPP1 の 2 倍になっている。LC-DPP は画像の移動幅が LC-SPP2 と等しいが、30 秒後の NCEV が、 -14.3dB となり、さらに収束が高速化している。

エコー経路に 1000 タップのインパルス応答を用いた場合の NCEV を図 11 に示す。64 タップの場合と同様に、LC では NCEV が、 -2.5dB までしか低下しておらず、フィルタ係数が正しいエコー経路に収束していない。LC-SPP2 では、30 秒後の NCEV が、 -8.4dB まで低下した。それに対して、LC-DPP は 30 秒後の NCEV が、 -9.0dB であり、高速化の効果がみられる。また、NCEV が、 -8.0dB に低下するまでに必要な時間を比べると、LC-SPP2 では 24 秒であるのに対して、LC-DPP では 19 秒であり、収束速度が 20% 高速化している。

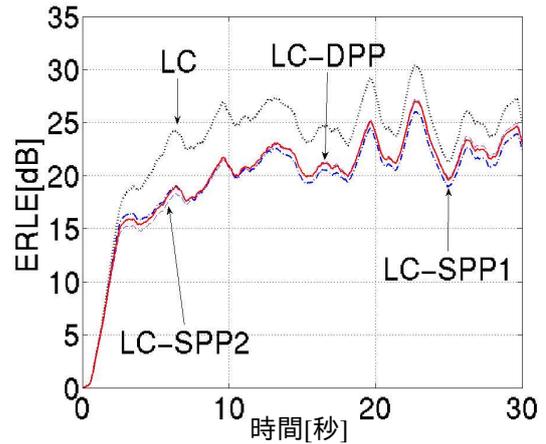


図 12: ERLE(1000 タップ)

表 2: 評価方法

得点	内容
5	原音と違いが分からない
4	違いが分かるが不快ではない
3	少し不快
2	不快
1	かなり不快

束速度が 20% 高速化している。

ERLE を図 12 に示す。時間が約 10 秒経過して以降、LC-SPP1 の ERLE は、LC-SPP2 と LC-DPP の ERLE より、約 1dB 低下している。これは、LC-SPP1 の遅延サンプル数が他の 2 方式に比べて少ないため、適応フィルタの係数が真の解に十分近付いていないことが原因である。

5 音質の主観評価

前処理を両チャンネルに入れたことによる音質の変化を確認するため、主観評価を行った。主観評価の手法として、隠れ基準付三刺激二重盲検法 (triple stimulus/hidden reference/double blind approach) を用いた [6]。被験者は参照音 (原音) を聞いた後、隠れ参照音と前処理を行った処理音を聞き、原音との違いに応じて 2 つの音を評価する。評価方法は、表 2 の ITU-R 5 段階劣化率に基づいている [7]。隠れ参照音と処理音の提示順は、ランダムである。

評価は図 13 に示す構成の部屋で行った。被験者は 8 人である。音声は、シミュレーションで用いたものを約 10 秒、切り出して使った。また、LC-SPP は、遅延サンプル数を 2 サンプルとした。その他のパラメータは、前節

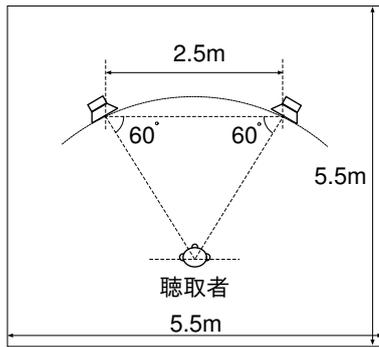


図 13: 部屋の構成

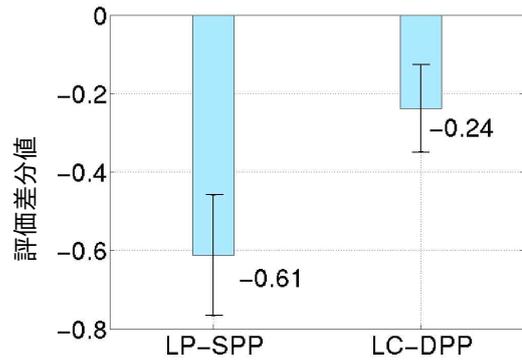


図 15: 評価差分値

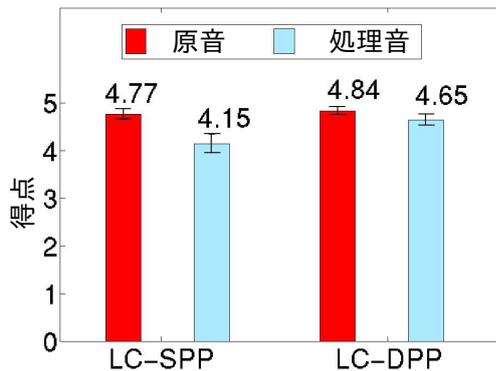


図 14: 主観評価結果

と同様である。

評価結果を図 14 に示す。各方式について、すべての被験者と音声に対して平均した得点を棒グラフで示している。棒グラフの上端にある縦線の部分は、正規分布による 95%信頼性区間である。LC-SPP, LC-DPP ともに、処理音に対する得点の平均値は参照音から低下しており、前処理によって音質が劣化している。しかし、LC-SPP に対する得点の平均値が 4.15 であるのに対して、LC-DPP に対する得点の平均値は 4.65 であり、参照音の得点に近い。

図 15 に、処理音に与えられた得点から参照音に与えられた得点を差し引いた値である、評価差分値 (Grading Difference) を示す。この値は、負の方向に絶対値が大きくなるほど、音質が劣化していることを表す。LC-SPP の評価差分値は、 -0.61 であるのに対し、LC-DPP の評価差分値は、 -0.24 であり、提案法は従来法よりも 0.37 ポイント高い。提案法と従来法の評価差分値は統計的な有意差をもっており、提案法は従来法よりも音質の劣化が少ない。これは、提案法では両チャンネルに前処理を行っており、音像が前処理を行う前の位置を中心に移動するため、従来法よりも音像の移動を知覚しにくいことが原因である。

6 まとめ

前処理ユニットを両チャンネルに有する構成のステレオエコーキャンセラを提案した。2つの前処理ユニットは、チャンネル毎の遅延挿入を異なる位相で行うことでエコー消去のための新たな条件式をつくり、収束を高速化する。提案法は、両チャンネルの信号を周期的に遅延させており、音像が前処理を行う前の位置を中心に移動するため、従来法よりも音質が優れている。シミュレーションによって、提案したエコーキャンセラは、従来のエコーキャンセラより 20%高速に収束することを確認した。また、主観評価によって、提案法における ITU-R 5 段階劣化率の評価差分値が、統計的に有意な差で、従来法より 0.37 ポイント高くなることを確認した。

参考文献

- [1] 藤井哲朗, 島田正治, “多チャンネルエコーキャンセラの一構成法” 電子通信学会技術報告, 26.Mar.1985.
- [2] A. Hirano and A. Sugiyama, “Convergence Characteristics of a Multichannel Echo Canceller with Strongly Cross-Correlated Input Signals - Analytical Results -,” Proc. DSP Symposium of IEICEJ, pp. 144-149, Nov. 1991.
- [3] Y. Joncour and A. Sugiyama, “A Stereo Echo Canceller with Correct Echo-Path Identification,” Proc. of ICASSP’98, pp3677-3680, May 1998.
- [4] A. Sugiyama, Y. Joncour, and A. Hirano, “A Stereo Echo Canceller with Correct Echo-Path identification Based on Input Sliding Technique,” IEEE Trans. SP, pp.2577-2587, Nov. 2001.
- [5] S. Haykin, “Adaptive Filter Theory, Third Edition,” Prentice-Hall, 1996
- [6] S. Bergman, C. Grewin and T. Rydén, “The SR Report on The MPEG/Audio Subjective Listening Test Stockholm April/May 1991,” ISO/IEC JTC1/SC2/WG11 MPEG 91/010, May 1991.
- [7] CCIR Recommendation 562.