

短時間フーリエ変換に基づく非線形特性の近似と適応フィルタへの応用

Approximation of Nonlinear Characteristics Based on Short Time Fourier Transform and its Application to Adaptive Filter

小木戸 諒[†] 中山 謙二[†] 平野 晃宏[†]
† 金沢大学大学院 自然科学研究科 電子情報工学専攻

Ryo Kokido[†] Kenji Nakayama[†] Akihiro Hirano[†]
†Division of Electronics and Computer Science
Graduate School of Natural Science and Technology, Kanazawa Univ.
E-mail : kokido@leo.ec.t.kanazawa-u.ac.jp
nakayama@t.kanazawa-u.ac.jp

あらまし

非線形エコーキャンセラの方式について検討する。適応フィルタは時間領域と周波数領域で処理する方法がある。本稿では、時間領域と周波数領域における線形エコーキャンセラ（サブレッサ）による非線形エコーの抑圧特性について検討した。その結果、周波数領域における線形エコーサブレッサは非線形エコーをある程度抑圧できることを確認した。次に、周波数領域におけるエコーキャンセラの2つの方式を比較した。残留非線形エコーのスペクトルを線形擬似エコーのスペクトルの定数倍で近似する方法とフレーム間で平均化した擬似伝達特性を用いる方法である。線形エコーと非線形エコーの比率をいろいろ変えてシミュレーションを行った結果、ほぼ同じ特性が得られた。

ABSTRACT

This paper investigates performance of non-linear echo cancellers. Adaptive filters can be realized in the time domain and in the frequency domain. In the time domain, an FIR adaptive filter can be used. A spectral suppression (SS) method is one of the methods used in the frequency domain. In this paper, performance of suppressing non-linear echo components by using the FIR adaptive filter and the SS method is investigated. The SS method can suppress the non-linear echo components to some extent. Next, two kinds of echo cancellers are compared for suppressing combined linear and non-linear echo components. Their performance are almost the same.

1 まえがき

携帯電話の本体は小さく、軽量化のために振動しやすい部品を使わざるを得ず、振動によりエコーパスが非線形になり、スピーカ自体や、本体、機械的接点から非線形な歪んだエコーが発生する。これはハンズフリー通話時などの、本体に取り付けられた小型スピーカに大きいパワーの信号が送られた時に生じやすい。このエコーを抑圧するためにはエコーキャンセラが必要になる。

適応フィルタは時間領域と周波数領域で動作するものがある。線形適応フィルタとしては、前者では、FIR 形適応フィルタ等があり、後者では、スペクトルサブレッション法等がある。また、時間領域で非線形適応フィルタを構成する方法として、ボルテラフィルタを用いる方法がある。しかし、回路規模が大きくなり、収束も遅くなると言う問題がある。一方、周波数領域では、FIR 形線形エコーキャンセラと周波数領域におけるエコーサブレッションを組み合わせる方法が提案されている [2]。その方式では、残留非線形エコーのスペクトルが線形エコーキャンセラの擬似エコーのスペクトルの定数倍で近似できるという特徴を利用している。

本稿では、携帯電話等で必要とされる非線形エコーキャンセラの構成と非線形エコーの抑圧特性について解析を行う。まず、時間領域と周波数領域における線形適応フィルタで非線形エコーがどの程度抑圧できるかを解析する。前者は FIR 形適応フィルタを用い、後者はスペクトルサブレッション法を用いる。後者では、擬似伝達特性を用いてスペクトルゲインを求めている。また、擬似伝達特性を複数のフレームで平均化したときの有効性も解析する。

次に、FIR 形適応フィルタとスペクトルサブレッショ

ン法を組み合わせた NTT 方式 [1] と前述の NEC 方式 [2] による非線形エコーの抑圧特性をシミュレーションにより比較する。前者は、線形エコーキャンセラの方式として提案されたものであり、非線形エコーは対象としていない。エコーパスの伝達特性を推定してスペクトルゲインを求める方式であるが、非線形エコーキャンセラとして考えるときには伝達特性は擬似的なものとなる。

2 時間領域と周波数領域におけるエコーキャンセラの構成 (NTT 方式)

2.1 全体構成

従来の一体化方式を図 1 に示す [1]。時間領域でエコーを消去する線形エコーキャンセラ、周波数領域でノイズ、エコーを抑圧するノイズキャンセラ、エコーサプレッサーを組み合わせたものである。近端雑音の存在する環境でエコーを抑圧することを目的としている。

動作について簡単に説明する。まず、相手側 (遠端) からの遠端音声 $x(n)$ がエコーパス h を介して発生したエコー $y(n)$ 、自分側 (近端) の近端音声 $s(n)$ 、近端雑音 $n(n)$ がマイク入力信号 $z(n) = s(n) + y(n) + n(n)$ として入力されたとする。ここで、 n は離散時間を表す。

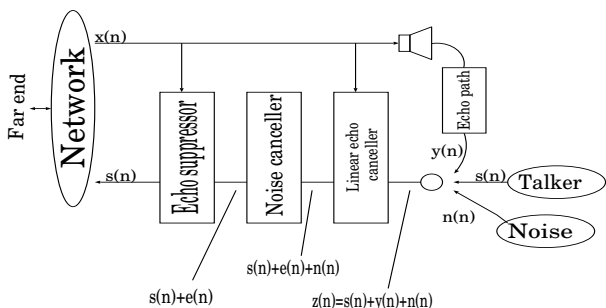


図 1: エコーキャンセラとノイズキャンセラの一体化方式 [1]

はじめに、マイク入力信号 $z(n)$ に対し、エコー $y(n)$ を線形エコーキャンセラを通すことにより消去している。この時、近端雑音の影響で消し残ったエコー $e(n)$ が近端雑音 $n(n)$ に埋もれて存在する。次に、線形エコーキャンセラの出力信号をノイズキャンセラに入力して近端雑音 $n(n)$ を消去し、さらにエコーサプレッサーにより残留エコー $e(n)$ を抑圧して、近端音声 $s(n)$ を出力する。

2.2 雑音抑圧部の構成

この方式ではノイズキャンセラとして、音声と雑音の統計的な性質を利用して雑音を抑圧するスペクトルサプレッション法を用いる。線形エコーキャンセラの出力を $za(n)$ とすると、入力信号は $za(n) = s(n) + n(n)$ となる。それぞれの周波数成分を $Za(l, k) = |Za(l, k)|e^{j\theta}$, $S(l, k) = |S(l, k)|e^{j\phi}$, $N(l, k) = |N(l, k)|e^{j\psi}$ とすると周波数領域で

の関係式は、

$$Za(l, k) = S(l, k) + N(l, k) \quad (1)$$

で与えられる。ここで l はフレーム番号、 k は周波数番号、 $|Za(l, k)|, |S(l, k)|, |N(l, k)|$ はそれぞれの振幅成分、 θ, ϕ, ψ はそれぞれの位相成分を表す。ノイズキャンセラの構成は図 2 のようになる。ここで、図の $E[S(l, k)], E[N(l, k)]$

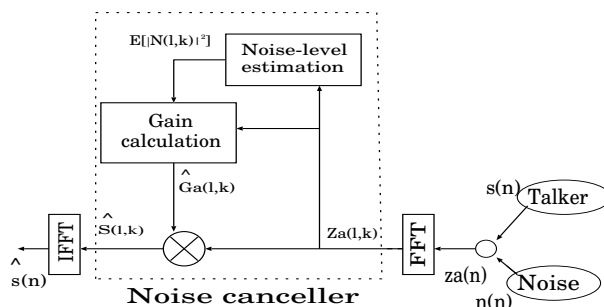


図 2: 図 5.1 の方式におけるノイズキャンセラの構成

は $S(l, k), N(l, k)$ の期待値 (時間平均) を表す。

処理の流れとしては、まず、入力信号を周波数領域に変換する。その後 $E[N(l, k)]$ を音声の存在しない区間 (無音区間) の入力信号の期待値から推定する。そして入力信号のスペクトル振幅成分に、ゲインを掛け合わせて、近端雑音を抑圧することができる。本研究では、音声がある場合と無い場合でのスペクトルエントロピーの差を利用して無音区間の検出を行う。先頭数フレームを無音区間と仮定して、先頭数フレームのスペクトルエントロピーの平均値に定数を掛けた値を閾値とする。以降のフレームで閾値よりスペクトルエントロピーが高い場合には無音区間と判断し、逆に低い場合には音声区間と判断する。ゲインはウィナーフィルタ法で計算される。

誤差の評価量として ε を $\varepsilon = E[S(l, k) - \hat{S}(l, k)]$ と定義してやると、 ε を最小にする条件を次のように求めることができる。

$$\varepsilon = E[S(l, k) - \hat{S}(l, k)] \quad (2)$$

$$\varepsilon = E[S(l, k) - Ga(l, k) \cdot Za(l, k)] \quad (3)$$

$$\frac{\partial \varepsilon}{\partial Ga(l, k)} = E[2Ga(l, k) \cdot |Za(l, k)|^2 - 2S(l, k) \cdot Za(l, k)] \quad (4)$$

ここで、 $\frac{\partial \varepsilon}{\partial Ga(l, k)} = 0$ より

$$Ga(l, k) = \frac{E[Za(l, k) \cdot S(l, k)]}{E[|Za(l, k)|^2]} \quad (5)$$

また、周波数領域上で $Za(l, k) = S(l, k) + N(l, k)$ の関係式が成立する。ここで、近端雑音と近端音声は無相関であることを考慮すると

$$Za(l, k) \cdot S(l, k) = Za(l, k) \cdot (Za(l, k) - N(l, k)) \quad (6)$$

$$= |Za(l, k)|^2 - |N(l, k)|^2 \quad (7)$$

よって、最終的にゲインの推定値は式 (8) で決定できる。

$$\hat{G}a(l, k) = \frac{|Za(l, k)|^2 - E[|N(l, k)|^2]}{|Za(l, k)|^2} \quad (8)$$

2.3 周波数領域におけるエコー抑圧部の構成

エコー抑圧処理は、雑音抑圧処理を拡張したような形で行う。ノイズキャンセラの出力が $zb(n)$ の時、 $zb(n)$ を周波数領域に変換し、ノイズキャンセラと同様にスペクトルサプレッション法に基づいてエコーを抑圧する。ゲインは式 (8) の近端雑音の期待値 $E[|N(l, k)|^2]$ を残留エコーの期待値 $E[|E(l, k)|^2]$ に置き換えて導出する。

$$\hat{G}a(l, k) = \frac{|Za(l, k)|^2 - E[|E(l, k)|^2]}{|Za(l, k)|^2} \quad (9)$$

エコーサプレッサーの構成は図 3 になる。

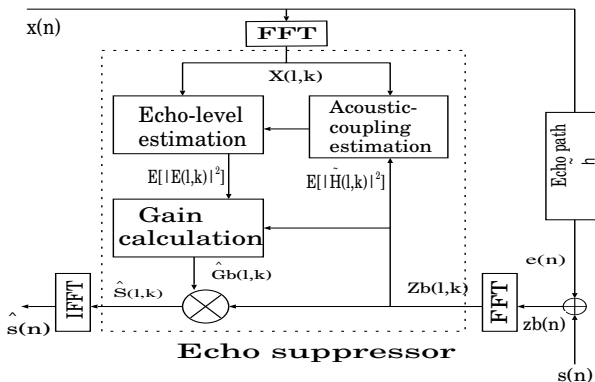


図 3: 図 5.1 の方式におけるエコーサプレッサーの構成

ここで \tilde{h} は遠端音声から残留エコーまでの伝達特性を表し、 $\tilde{H}(l, k)$ はその周波数特性を表す。残留エコーの期待値は、遠端音声のワースペクトル $|X(l, k)|^2$ にエコーパスの周波数特性 $E[|\tilde{H}(l, k)|^2]$ を乗じた値と、1 フレーム前に求めた残留エコーの期待値 $E[|E(l-1, k)|^2]$ を再帰的に加算して求める。

$$E[|E(l, k)|^2] = (1 - \beta) \cdot E[|\tilde{H}(l, k)|^2] \cdot |X(l, k)|^2 + \beta \cdot E[|E(l-1, k)|^2] \quad (10)$$

β は忘却係数で、残響時間に合わせた値に設定する。エコーパスの周波数特性には、遠端音声のワースペクト

ルと入力信号のワースペクトルの期待値 $E[|X(l, k)|^2]$, $E[|Zb(l, k)|^2]$ の比を用いる。

この時、残留エコーのワースペクトルが欲しいので $E[|Zb(l, k)|^2]$ としてはシングルトーク ($S(l, k) = 0$) 区間のワースペクトルを用いる。

従来の一体化構成において、以上の処理により近端雑音が存在する環境でもエコーの抑圧が可能であるといえる。しかし、この一体化方式では非線形なエコーの抑圧を意識していない。非線形なエコーの抑圧に対応できるエコーキャンセラ (NEC) について次節で説明する。

3 時間領域と周波数領域における非線形エコーの抑圧特性

エコーパスが線形特性のみからなる場合と、線形及び非線形特性からなる場合について、時間領域と周波数領域におけるエコー抑圧特性をシミュレーションにより調べる。入力信号として、遠端音声、近端音声共に 8kHz で標準化された女性の音声を用いた。近端雑音としては定常な白色雑音を付加した。

3.1 シミュレーション条件 1

時間領域で実現する適応フィルタでエコーの抑圧を考えた場合と、周波数領域で実現する適応フィルタでエコーを抑圧する場合での性能を比較する。そのために、時間領域の FIR 形エコーキャンセラのみを使用した場合と、NTT 方式における周波数領域におけるエコーサプレッサーのみを使用した場合について入力信号に近端音声 (ダブルトーク) が含まれない条件でエコーの低減量を計算した。線形エコーの 2 乗平均に対して、非線形エコーの 2 乗平均の割合が 0%~50% になるように加え、シミュレーションを行った。

また、近端音声があった場合に、エコーサプレッサーによる処理がエコー抑圧後の近端音声に与える影響について調べるために、エコーサプレッサーの入力信号に近端音声を含めて、ダブルトークが存在する区間での SNR_{seg} を計算した。エコーとしては非線形エコーのみを付加した。

3.2 評価方法

3.2.1 低減量

$$\text{低減量} = 10 \log_{10} \frac{B}{A} \quad (11)$$

ただし A, B は線形エコーキャンセラ、エコーサプレッサーのそれぞれにおいて A : 入力信号の 2 乗平均, B : 出力信号の 2 乗平均とする。さらに、十分に収束した部分で話を進めるために、低減量は最後の 1 周期分 (7 万サンプル) で計算する。低減量が低いほど良くエコーを抑圧できたことになる。

表 1: シミュレーション条件 1

線形エコーキャンセラ部分	
学習アルゴリズム	NLMS アルゴリズム
ステップサイズ, μ	0.1
FIR フィルタのタップ数	512
線形エコーパスのタップ数	512
非線形エコーパスのタップ数	60
近端音声 (ダブルトーク)	無し
遠端音声	サンプリング周波数 8kHz 21 万サンプル
近端雑音	無し
付加するエコー	線形エコー + 非線形エコー 0~50% 非線形エコーのみ
評価方法	
	低減量 (dB)
エコーサプレッサ部分	
分析フレーム長	Hanning 窓 $f=64,128,256,512,1024$ サンプル
オーバーラップ長	$f/2=32,64,128,256,512$ サンプル
近端音声 (ダブルトーク)	有り (サンプリング周波数 8kHz, 21 万サンプル) または無し
遠端音声	サンプリング周波数 8kHz 21 万サンプル
近端雑音	無し
付加するエコー	線形エコー + 非線形エコー 0~50% 非線形エコーのみ
忘却係数 β	0
評価方法	低減量 (dB), SNR_{seg}

3.2.2 音声推定誤差の評価

各区間の SNR の平均を求めるセグメンタル SNR で評価を行なう。細かい時間間隔で SNR を求め、その平均値である SNR_{seg} は次式で定義される。

$$SNR_{seg} = \frac{10}{L} \sum_{l=0}^{L-1} \log_{10} \frac{\sum_{n=N_l}^{N_l+N-1} s^2(n)}{\sum_{n=N_l}^{N_l+N-1} (\hat{s}(n) - s(n))^2} \quad (12)$$

ただし、 N は分析フレームの長さであり、近端音声の存在する区間で計算する。

3.3 シミュレーション結果 1

エコーパスが線形及び非線形特性からなる場合においてシミュレーション条件 (表 1) に従って、線形エコーキャンセラ (線形 EC) とエコーサプレッサ (ES) のエコーの低減量を比較する。結果は表 2 のようになる。エコーパスが線形特性のみの場合を除いて、エコーサプレッサの方が低減量が低くなっていることが分かる。

NTT 方式のエコーサプレッサでも非線形エコーを十分抑圧できることが分かった。実際にこれは NEC 方式のエコーサプレッサと同様な性能が得られていることが後に出るシミュレーション結果 2 より分かった (表 5~6)。これはフレームに亘る平均化により普遍的な関係 (非線形な畳み込み) を求めているからだと考えられる。これよりフレーム間の平均化が低減量にどのように影響するのかを調べる。シミュレーション条件 (表 1) において、近端音

声の無い条件でエコーパスの特性は非線形のみとし、忘却係数 β を変化させながらシミュレーションした結果を表 3 に示す。

また、近端音声が存在する場合にフレーム間の平均化が与える影響についてを調べる。今度は、シミュレーション条件 (表 1) において、近端音声のある条件でエコーパスの特性は非線形のみとし、エコーを抑圧したときのセグメンタル SNR を表 4 に示す。

表 2: 線形 EC, ES におけるエコー低減量 [dB] の比較

付加するエコー	線形 EC	ES				
		$f = 64$	128	256	512	1024
線形のみ	-13.09	-12.85	-12.56	-11.58	-10.26	-11.24
線形 + 非線形 10%	-8.75	-12.42	-12.19	-11.33	-9.99	-10.93
線形 + 非線形 20%	-6.86	-11.89	-11.69	-11.02	-9.37	-10.51
線形 + 非線形 30%	-5.43	-11.26	-11.10	-10.63	-9.41	-10.00
線形 + 非線形 40%	-4.94	-10.98	-10.83	-10.45	-9.26	-9.77
線形 + 非線形 50%	-4.36	-10.59	-10.46	-10.59	-9.04	-9.45
非線形のみ	0.072	-10.30	-9.65	-10.45	-10.25	-9.30

表 3: ES において忘却係数を変化させた時の低減量 [dB]

忘却係数 β	ES				
	$f = 64$	128	256	512	1024
0	-10.30	-9.65	-10.45	-10.25	-9.30
0.1	-10.06	-9.38	-10.28	-10.26	-9.31
0.2	-9.59	-8.90	-9.81	-10.05	-9.12
0.3	-8.24	-7.70	-8.35	-8.54	-7.81

表 4: ES において忘却係数を変化させた時の SNR_{seg} [dB]

忘却係数 β	ES				
	$f = 64$	128	256	512	1024
0	13.45	12.94	13.27	14.35	14.58
0.1	8.92	10.39	11.39	13.50	13.83
0.2	8.44	9.91	10.98	13.06	13.61
0.3	8.37	9.98	11.05	12.86	13.43

4 エコーキャンセラ (NEC 方式)

NEC 方式のエコーキャンセラは非線形エコーの抑圧を目的としている。図 4 にエコーキャンセラの構成を示す。マイク入力信号に、遠端音声 $x(n)$ がエコーパス (図中では省略) を介して発生したエコー (線形エコー+非線形エコー) $y(n)$ と近端音声 $s(n)$ を入力し、初段の線形エコーキャンセラでエコーの線形成分を除去する。線形エコーキャンセラで処理後、出力の残留エコー $e(n)$ としては非線形成分のみが残るとする。

そして、線形エコーキャンセラで計算した疑似エコー $\hat{y}(n)$ と線形エコーキャンセラの出力信号 $z(n) = s(n) + e(n)$ を周波数領域に変換し、疑似エコーと非線形残留エコーは周波数領域で相関があるということが報告されているので [2]、そのことを利用すると非線形エコーのスペクトルが得られ、非線形エコーを抑圧することができる。抑圧後に、時間領域に戻すことによってエコーを抑圧した近端音声 $\hat{s}(n)_{out}$ が得られる。疑似エコーと非線形残留エコーの周波数領域における関係式は式 (14) となり、比例定数 $a(l, k)$ は式 (15) で推定する。

$$|E(l, k)| = a(l, k) \cdot |\hat{Y}(l, k)| \quad (13)$$

$$a(l, k) = \frac{|E(l, k)|}{|\hat{Y}(l, k)|} \quad (14)$$

$$\hat{a}(l, k) = \frac{E[|Z(l, k)|_{single-talk}]}{E[|\hat{Y}(l, k)|]} \quad (15)$$

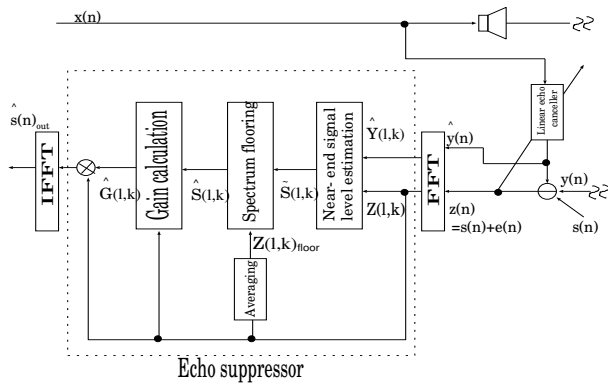


図 4: エコーキャンセラ (NEC) の構成 [2]

4.1 シミュレーション条件 2

NTT 方式 (図 1) においてノイズキャンセラを取り除き、線形エコーキャンセラとエコーサプレッサの 2 段構成にしたものと、NEC 方式 (図 4) のエコーキャンセラについて性能評価を行う。近端雑音が存在しない環境で、エコーパスが線形特性のみからなる場合と、エコーパスが線形及び非線形特性の場合について、エコーの低減量を調べる。シミュレーション条件は表 5 の通り。

4.2 シミュレーション結果 2

エコーパスが線形及び非線形特性からなる場合において、条件に従ってシミュレーションを行なった結果は表 6 のようになる。エコーパスが線形特性または非線形特性の両方の場合において、NTT 方式のエコーサプレッサの方が低減量が低くなり、エコーを良好に抑圧することができた。これより線形エコーキャンセラ用の NTT 方式をそのまま非線形エコーパスに用いた場合でも、NEC 方式と同様の性能が得られることが分かる。

表 5: 一体化構成の性能評価におけるシミュレーション条件

線形エコーキャンセラ部分	
学習アルゴリズム	NLMS アルゴリズム
ステップサイズ: μ	0.1
FIR フィルタのタップ数	512
線形エコーパスのタップ数	512
非線形エコーパスのタップ数	60
近端音声 (ダブルトーク)	無し
遠端音声	サンプリング周波数 8kHz 21 万サンプル (1 周期 7 万サンプルを 3 周期分繰返し)
エコーサプレッサ部分	
分析フレーム長	$f=512$ サンプルの Hanning 窓
オーバーラップ長	$f/2=256$ サンプル
近端雑音	無し
付加するエコーの量	条件 1. 線形エコーのみ 条件 2. 線形エコー + 非線形エコー 10% 条件 3. 線形エコー + 非線形エコー 20%
評価方法	低減量

表 6: 低減量 [dB]

付加量の条件	条件 1	条件 2	条件 3
EC 後	-13.09	-8.75	-6.86
ES 後 (NTT)	-8.83	-8.30	-8.47
ES 後 (NEC)	-7.95	-7.86	-8.04
合計 (NTT)	-21.92	-17.05	-15.33
合計 (NEC)	-20.94	-16.61	-14.90

5 まとめ

本稿では、線形処理で実現される時間領域の適応フィルタと周波数領域における適応フィルタによる非線形エコーの抑圧特性を解析した。時間領域では、FIR 形適応フィルタを用いたが、非線形エコーはほとんど抑圧できなかった。一方、周波数領域では、スペクトルサプレッション法を用いたが、非線形エコーをある程度抑圧することができた。次に、FIR 形適応フィルタと周波数領域におけるスペクトルサプレッション法を用いる 2 つのエコーキャンセラを比較した。一方は線形エコーキャンセラとして開発されたものであり、他方は携帯電話用の非線形エコーキャンセラとして開発されたものである。両者はスペクトルゲインを推定する方法が異なる。シミュレーション結果によれば、エコーキャンセラとしての特性はほぼ同じであった。

参考文献

- [1] 坂内 澄宇, 羽田 陽一, 田中 雅史, 佐々木 潤子, 片岡 章俊, “雑音抑圧及びエコー抑圧機能を備えた音響エコーキャンセラ”, 信学論 (A), Vol.J87-A No.4pp.448-457, April, 2004.
- [2] O.Hoshuyama, and A.Sugiyama, “An Acoustic Echo Suppressor Using a Frequency-Domain Model of Highly Nonlinear Residual Echo”, IEEE Proc. ICASSP2006, PP.V-269-272.
- [3] O.Hoshuyama, and A.Sugiyama, “Nonlinear Acoustic Echo Suppressor Based on Spectral Correlation between Residual Echo and Echo Replica”, IEICE Trans Fundamentals.2006; E89-A: 3254-3259.
- [4] 後藤 伸哉, “適応フィルタとスペクトルサブプレッジョンを併用するエコーキャンセラにおけるスペクトルゲイン制御法”, 金沢大学大学院, 修士論文, 2, 2007
- [5] 宝珠山 治, “非線形エコーキャンセラのための周波数領域モデルにおける回帰係数更新法”, 第 24 回信号処理シンポジウム (鹿児島), B2-3, 2009, 11
- [6] 小木戸 諒, “エコーキャンセラとノイズキャンセラの一体化構成”, 金沢大学, 卒業論文, 2, 2009