

高性能音響エコーキャンセラの開発

High-Performance Acoustic Echo Cancellation Development

小島 順治*
Junji KOJIMA

牧野 昭二*
Shoji MAKINO

羽田 陽一*
Yoichi HANEDA

島内 末廣*
Suehiro SHIMAUCHI

あらまし

マルチメディア時代の到来を迎え、遠隔地と通信しているにもかかわらず、あたかも同一の室内にいるように会話ができることや、マイクロホンやスピーカの位置を意識しないで会話ができるシームレスな音響空間の実現が望まれている。

NTTでは、このようなシームレスな音響空間の実現に必須の技術である音響エコーキャンセラ[†]の開発を進め、新しいES(Exponentially weighted Stepsize)射影アルゴリズムによる適応追従性や同時通話時の通話性能の向上、ならびに音声スイッチ[‡]併用方式を採用することにより、自動学習を実現した高性能音響エコーキャンセラを開発した。

Abstract

In the coming multimedia age, a seamless acoustic environment, in which two parties can talk to each other as if they were in the same room, regardless of the microphone positions, is desirable.

We have developed a new high-performance acoustic echo canceller that creates such a seamless acoustic environment. It achieves fast convergence using the new exponentially weighted stepsize projection algorithm, high speech quality during double talk, and training-free operation through the use of voice-switching technologies.

1 ま え が き

通信を利用して多人数が同時に会話をするテレビ会議の利用がビジネス分野で広がっている。また、マルチメディアの時代の到来を迎え、遠隔地と通信しているにもかかわらず、あたかも同一の室内にいるように会話ができることや、マイクロホンやスピーカの位置を意識しないで会話ができるシームレスな音響空間の実現が望まれている。

テレビ会議のような拡声通話では、スピーカから出た声が、相手側のマイクロホンで拾われて再び自分の側に戻るため、エコーが生じる。さらにひどい場合にはハウリングを起こす。音響エコーキャンセラは、このようなエコーやハウリングを取り除くための装置であり、スピーカからマイクロホンへ至るエコーを予測して差し引くことによりエコーを消去するものであるが、現状の装置では性能や使い勝手など種々の問題があった。

NTTでは、適応アルゴリズム[‡]などの音響信号処理の研究成

*NTT ヒューマンインタフェース研究所 NTT Human Interface Laboratories

[†]論文中のゴシック文字は、本論文に解説があります。

©日本電信電話株式会社 1995

果を生かし、適応追随性や同時通話時の性能に優れた高性能音響エコーキャンセラを開発したのでその概要を述べる。

2 開発の背景

エコーキャンセラはもともと衛星通信や長距離回線で生じるエコーを消去する装置として開発されていたものである。回線エコーキャンセラ⁽¹⁾がハイブリッドトランスなどによる電気的なエコー経路を対象にしているのに対し、音響エコーキャンセラ⁽²⁾はスピーカからマイクロホンへの音場を経由する音響的な経路を対象としている点に異なる。音響エコーキャンセラは、一般に、回線用よりもエコー消去時間³の長いものが要求されるため、回線用の装置に比べ規模が大きい点に特徴がある。

NTTでは、1987年にエコーキャンセラ方式の音声会議装置を開発⁽³⁾したが、装置も大きく高価であったため、同時に開発された安価な音声スイッチ方式の音声会議装置のほうがよく売れている。しかしながら、音声スイッチ方式では通話時に切断が起こるため、満足な通話品質は得られていない。

研究所では、その後、音響信号処理研究の一環として音響エコーキャンセラに適用可能な適応処理アルゴリズムの研究を進め、従来の方式に比べ、優れた適応追随性を有する新しい適応アルゴリズム(ES射影アルゴリズム⁽⁴⁾)を開発した。さらに、音声スイッチとの併用制御方式など、音響エコーキャンセラとして装置化をするうえで重要となるエコーキャンセラの制御法の研究を進め、装置試作による技術確認を行った結果、従来装置に比べ性能、経済性など種々の点で優れた音響エコーキャンセラを実現した。

3 エコー消去方式とその問題点

エコーを消去する方法については音声スイッチ、センタクリッピング、周波数偏移など種々の方法があるが、音声に歪みを与えない点でエコーキャンセラ方式が最も自然な会話を可能とするといえる。

エコーキャンセラの原理を図1に示す。エコーキャンセラは、スピーカからマイクロホンへ至るエコー経路に相当する伝達特性を推定し、その特性を適応フィルタにより実現して擬似エコー信号を生成することにより、エコー信号を消去するものである。

エコーキャンセラは、スピーカからマイクロホンへ至るエコー経路を常に推定するが、人の動きやマイクロホンの移動などの音響条件が急激に変化すると、適応するのに時間がかかる。このため、従来の装置は、音響条件が急激に変化すると通話中にハウリングを起こしたり、初期学習を仕直す必要があった。

また、ダブルトーク(同時通話)時には、スピーカからの音と

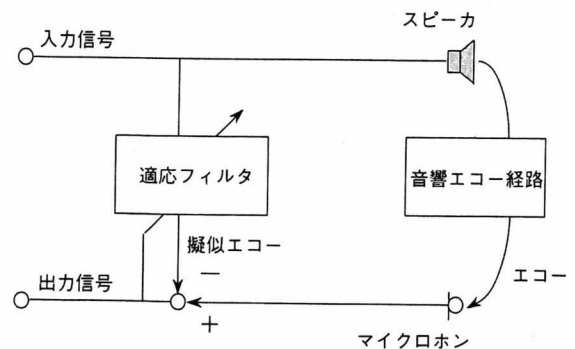


図1 音響エコーキャンセラの原理

ともに送話者の声も入力されるので、適応フィルタの収束が不安定になり、エコーが消去しにくくなることも問題であった。そのため、ダブルトーク時には適応動作を止めてしまうものが多かった。

さらに、エコーキャンセラは一般に、白色雑音や低域から高域までの周波数を含むスイープ信号等を利用した初期学習を必要とするものが多く、耳ざわりであるばかりか、電源を入れても学習が終了するまでは、使えないなど使い勝手が悪い難点があった。

一方、従来、エコーの抑圧には最も多く使われている技術である音声スイッチの原理を図2に示す。音声スイッチは送話と受話のレベルを比較し、送話のレベルが高い場合は送話状態と判断して受話路に損失を挿入し、反対に受話のレベルが高い場合には、受話状態であると判断して送話路に損失を挿入することにより、ハウリングやエコーを抑圧するものである。音声スイッチは、ハードウェアは簡単であるが、通話路に損失を挿入するため、同時に話す(ダブルトーク)と話頭や話尾が切断される点が問題である。

エコーキャンセラと音声スイッチ方式の比較を表1にまとめて示す。

音声スイッチ方式はエコーキャンセラに比べ切断がある分だけ通話品質が劣化するが、エコーキャンセラのように常に適応

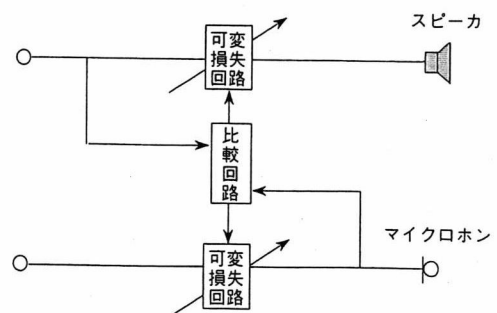


図2 音声スイッチの原理

表1 エコーキャンセラと音声スイッチ方式の比較

項 目	エコーキャンセラ	音声スイッチ
方式の概要	スピーカからマイクロホンへの伝達特性を擬似的に生成することにより、スピーカからマイクロホンへの音を消去する	送受話のレベルを常に比較して、レベルの低い方にその都度損失を挿入する
ダブルトーク時の動作	収束の安定性を欠くため、動作を止めるなどの処理が必要	話頭話尾切断、または相手の声が聞こえない場合がある
ハード規模	大	小

動作をしていないので動作は安定している。初期学習を行わないエコーキャンセラの開発には、音声で学習している間、音声スイッチの技術を利用して通話路に損失を挿入してハウリング・エコーを防止することが有効となる。

4 開発のねらい

音響エコーキャンセラの技術は、電話機からデスクトップ会議、テレビ会議装置まで、拡声通話系を有する通信機器にはすべて適用可能である。しかしながら、これらの機器は使用形態や設置（音響）環境がそれぞれ異なるため、装置化にあたっては対象を明確にする必要がある。

ハンドフリー電話機は個人が家庭、またはオフィスで使い、デスクトップ会議はオフィス、テレビ会議装置は特定の会議室で多人数で使う場合が多い。一方、設置環境はハンドフリー電話が送・受話距離50cm～1mくらいで普通の部屋で使うのに対し、デスクトップ会議では送・受話とも1～2m、テレビ会議では送話距離50cm～1mくらい、受話距離2～5mくらいでカーペット、天井の吸音ボードなどにより多少の吸音処理が成されている部屋で使う場合が多い。さらに、エコーの発生状況に密接に関係のあるスピーカ、マイクロホンの距離については、電話機で5～20cm、デスクトップ会議で50cm～1m、テレビ会議で2～5mと考えられる。これらの違いをエコーキャンセラの規模で考えてみると、おおむね図3のように書くことができる。図の横軸は、適応フィルタのタップ数でエコー消去時間長に相当



図3 エコーキャンセラの規模と適用域

する。縦軸は、必要なエコー消去量を示す。図の点線以下は音声スイッチの適用領域である。音声スイッチは原理的には消去すべき量が幾らでも動作するが、その挿入損失の量が切断感に影響を与え、通常10dB以下が望ましいとされている。

今回は、将来のマルチメディアのデスクトップ会議を想定したものが世の中にあまり出ていないことや、製造するうえでのハードウェア規模の妥当性から、デスクトップ会議用を開発し、技術確認を行うこととした。また、音響エコーキャンセラのみを装置化し、音声会議装置に具備されているマイクロホン切替え機能や、回線インタフェースは持たず、主としてISDN回線で使用する画像コーデックなどに接続して使用する形態とした。ISDNにおいてはG.722の音声符号化方式により高品質（7 kHz）の通話ができる。

学習機能は使用条件や対象にかかわらず、ないことが望ましい。そこで、今回の開発では音声による自動学習方式を対象とすることとした。

5 ハードウェアの構成

今回開発した音響エコーキャンセラ装置の構成を図4に示す。エコーキャンセラと音声スイッチ制御回路を組み合わせるとともに、エコーキャンセラ部分は適応フィルタから転送される係数を保持するエコー消去用フィルタと、常に適応処理を行う適応フィルタから構成されている。この構成はダブルトーク時の性能を考慮したものである。また、初期学習を行わず、会話音

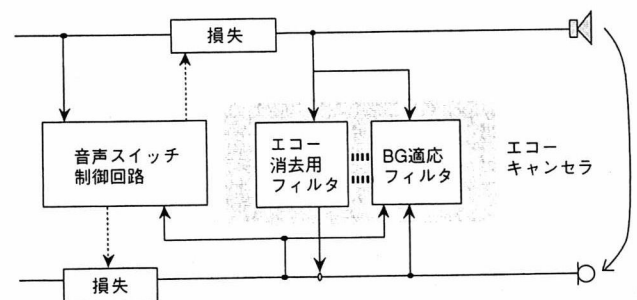


図4 音響エコーキャンセラ装置の構成

声で自動的に学習を可能とするため、音声スイッチとの併用制御を行っている。音声スイッチは電源立上げ時に働いて、ある程度適応動作が終了(収束)するまで損失を挿入している。

さらに、音場のモデル化の研究成果⁽⁵⁾を反映して、受話系に共通極・零モデルに基づくイコライゼーションフィルタを搭載し、受話特性の平坦化を図っている。この方法は従来に比べ、フィルタの規模を小さくできる利点がある。

開発した装置の機能仕様を表2に示す。帯域はISDNで7kHz音声を用いることを考慮して7.3kHzとした。エコー消去時間長は従来のテレビ会議や音声会議でも使用可能な150msとした。エコー消去量はS/N比等の条件が悪くなければ、通常35dB得られる。

ハードウェアは汎用の浮動小数点32ビットDSP(Digital Signal Processor)および固定小数点DSP複数個で構成され、デジタル部はほぼB5判1枚の大きさ(厚さは最大12mm)である。エコーキャンセラとともに音声スイッチもすべてデジタルで制御され、アナログ部はAD/DA、マイクロホンアンプ、スピーカアンプなどである。

エコーキャンセラも外部とのインタフェースはライン入出力で行うが、オーディオ入出力はマイクアンプおよびスピーカアンプを内蔵しておりスピーカ、マイクをそのまま接続すれば通話が可能となる。また、外付けでスピーカアンプ、マイクアンプを利用することも可能である。ライン入出力レベルはテレビ会議用の画像コーデックとの接続を考慮して決定した。

本装置の外観を図5に示す。大きさは270(W)×60(H)×242(D)mmであり、重さは約2.5kgである。表のパネルにはスピーカの音量調節のみとし、マイクロホンの入力レベルはあまり操作しないと考えられるため、後部パネルに配置した。

エコーキャンセラは使用される音響環境により異なった性能

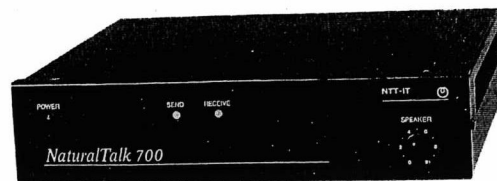


図5 音響エコーキャンセラ

を示す。本装置ではディップスイッチを装備し、部屋の大きさ、騒音の状況などの音響環境に合わせて適切なパラメータが設定できるようになっている。また、一方的な講義形式の会話、頻繁に同時通話が起る会話などの会話形態に合わせて音声スイッチの待時方式なども選択可能ようにした。

6 本装置の特徴および評価

本装置の特徴は大きく分けて、適応追従性、同時通話性能、自動学習の3つであり、表3にまとめて示す。以下にこれらの3項目について述べる。

6.1 新しい適応アルゴリズムの開発による適応追従性能の向上

テレビ会議において、エコーは会議室の音響条件に影響される。エコー経路はスピーカやマイクロホンの位置や向き、部屋の形や大きさ、参加人数や着席位置および姿勢などに影響され、変動し易く、この変動に適応してエコー経路を推定することが必要となる。いったん伝達系が変動するとエコー経路が推定され、エコーが消去されるまでハウリングやエコーを生じる状態が続くため、エコーキャンセラでは高速な適応追従性が要求される。

本装置では、NTTで開発した新しいES射影(Exponentially weighted Stepsize-Projection)アルゴリズムを採用することにより、従来ほとんどの製品に採用されているNLMS法(学習同定法)に比べ約4倍の適応追従性能を実現した。このアルゴリズムは室内音場における音声信号の減衰特性が指数関数的であることに着目して適応動作を行うESアルゴリズムと音声のスペクトルを白色化することにより高速性を実現する射影ア

表3 開発した音響エコーキャンセラの特徴

- ◇ES射影アルゴリズムによる適応追従性能の高速化
 - ◇適応フィルタとエコー消去用フィルタの構成によるダブルトーク性能の飛躍的向上
 - ◇初期学習不要かつ小型経済化を実現
- ES: exponentially weighted stepsize

表2 開発した音響エコーキャンセラの機能仕様

帯域	100Hz～7.3kHz
エコー消去時間	150ms
エコー消去量	35dB
ライン入力	端子形状 : RCAピンジャック レベル : -9dBm定格(最大+9dBm) インピーダンス: 11kΩ(不平衡)
ライン出力	端子形状 : RCAピンジャック レベル : -9dBm定格(最大+9dBm) インピーダンス: 600Ω(不平衡)
オーディオ入力	マイクロホンと外部ミキサの切替え可
オーディオ出力	スピーカと外部パワーアンプ切替え可
表示機能	送受話レベルの有無、電源ON/OFF
電源	AC100V 50/60Hz
消費電力	20W
寸法	270(W)×60(H)×242(D)mm
重量	約2.5kg

ルゴリズムを組み合わせたものである。このアルゴリズムの処理量は従来の学習同定法とほぼ同じであるため、ハード規模を増やさずに高速適応追従性が実現できる。

6.2 ダブルトーク時の通話性能改善

エコーキャンセラは、上記のように音響条件の変化を随時学習しながら適応動作をする。ところが、ダブルトーク時にはこの学習動作に悪影響を与え、エコーの除去が不完全になるという問題があった。

ダブルトーク時の適応フィルタを制御する方法については、種々考えられている。

例えばダブルトークを検出した後に適応フィルタの動作を停止する方法が考えられるが、この方法は検出遅れによるフィルタ係数の誤推定が生じる。

そこで、本エコーキャンセラでは、演算量は多少多いが誤推定を少なくすることを考慮し、適応フィルタとエコー消去用フィルタを持つ構成を基にダブルトーク制御法の検討を行った⁽⁶⁾。

この方式は、適応動作を行う適応処理部と、適応処理部の指示に基づいてエコーを消去するエコー消去用フィルタで構成される。適応処理部で逐次エコー経路を推定し、実際のエコー消去にはエコー消去用フィルタを用いる。適応フィルタ側では、ある程度エコーが消去されていると判断したときは常にフィルタ係数をエコー消去用フィルタへ転送する。ダブルトーク時は適応フィルタ側からエコー消去用フィルタ側に転送を行わず、ダブルトーク前の係数で消去を行う。これにより、ダブルトーク時に適応フィルタの動作が不安定となり通話性能が劣化することを防ぎ、良好な通話品質を得ることができた。

6.3 音声による自動学習

従来のエコーキャンセラは、電源投入時や、通信中に大幅に音響条件が変化した場合には初期学習をする必要があった。学習用の雑音(白色雑音など)は非常に耳ざわりなことや、学習が面倒であるため、あまり評判はよくなく、音声による自動学習の実現が望まれていた。本装置では音声スイッチとの併用制御方式を開発することにより、初期学習をしなくても会話音声による自動学習を可能とし、使い易さの向上を図った。

また、従来、電源投入時などの学習が必要な場合に不安定な動作をすることがあったが、この装置では、不安定時には音声スイッチが働いて損失を挿入(利得の減少)するとともに、適応フィルタの学習状況に合わせて損失を制御する新しい制御法を考案することにより、その問題を解決した。また、このような動作は適応追従性と密接な関係を持っており、開発した装置では音声を二言三言話す程度で収束し、挿入損失を抜くことができた。

さらに、この新しい制御法により、フィルタの規模に相当す

るエコー消去時間長が長くなくても、エコー消去が不十分のときは挿入損失でカバーすることができるので、同じ音響環境への適用域を保ちながら装置の小型化、経済化を実現することができた。

本エコーキャンセラの評価実験を行ったところ、優れた適応追従性を有するとともに、同時通話時にも歪みを与えずに通話ができることが確認できた。また、送話および受話の処理時間はどちらも3msであり、ほとんどリアルタイムで動作することが分かった。

さらに、本装置を用いて会議室、およびテレビ会議室等(残響時間200~700ms)で実(ISDN)回線を用いた会議を行った結果、部屋の隅から話しても十分よく聞こえ、相手の通話中に自然な割込みが可能となるなど、従来のテレビ会議装置よりも優れた通話性能が得られた。

本装置は、本来、マルチメディアのデスクトップ会議などを想定したものであったが、従来のテレビ会議などの替わりに用いても、十分使用可能であることが明らかになった。

7 あとがき

本エコーキャンセラを用いることにより十分な音量で相互に会話ができ、シームレスな音響空間を実現することが可能となった。また、エコーキャンセラ技術としては種々の適用形態が考えられるので、この技術を使ったハンドフリー電話機、テレビ会議装置、音声会議装置、およびパソコンベースのテレビ会議システムなど他システムへの展開を図る予定である。

一方、開発したエコーキャンセラは音響エコーキャンセラとしてシームレスな音響空間の実現を可能とするばかりでなく、回線用のエコーキャンセラへの適用も考えられる。

さらに、将来はATMやデジタル移動システムなどデジタルシステムが普及してくるので、それとともにエコー消去技術は、ますます重要となってくると思われる。今後はステレオエコーキャンセラなど高機能化、および雑音下でも消去できるアルゴリズムの開発などを行う予定である。

文 献

- (1) 例えばD.L.Duttweiler and Y.S.Chen: A Single Chip VLSI Echo Canceller, BSTJ, 59, No.2, pp.149-160, Feb., 1980.
- (2) 例えばY.Itoh, Y.Murayama, N.Furuya and T. Araseki: An Acoustic Echo Canceller for Teleconference, Proc. ICC85, pp.1498-1502, Jun., 1985.
- (3) H.Oikawa, S.Minami and T.Saeki: A New Echo Canceller Rearrized by High Performance Digital Signal Processor, Proc. ISCAS88, pp.1329-1332, Jun., 1988.
- (4) Makino and Kaneda: Exponentially Weighted Step-Size Pro-

jection Algorithm for Acoustic Echo Cancellers, IEICE TRANS. FUNDAMENTALS, E75-A, No. 11, pp.1500-1508, 1992.

- (5) Haneda, Makino and Kaneda: Common Acoustical Pole and Zero Modeling of Room Transfer Functions, IEEE Trans SP, 2, pp.320-328, 1994.
- (6) 中原・羽田他: 音響エコーキャンセラにおけるダブルトーク制御方式の検討, 日本音響学会講演論文集, 3-5-7, Mar., 1992.



小島 順治

ヒューマンインタフェース研究所主幹研究員
昭和48年入社。主に拡声電話機、音声入出力装置、およびISDN端末装置の研究実用化に従事。現在、オーディオ情報処理の研究実用化に従事。
昭和48年早稲田大学理工学部電子通信学科卒業。
電子情報通信学会・情報処理学会・日本音響学会会員。
昭和57年信学会学術奨励賞受賞。



牧野 昭二

ヒューマンインタフェース研究所主幹研究員
昭和56年入社。電気音響変換器、拡声電話機、および音響エコーキャンセラなどの音響信号処理の研究に従事。
昭和54年東北大学工学部機械工学第二学科卒業。56年同大学院機械工学修士課程修了。平成3年工学博士(東北大学)。
IEEE・電子情報通信学会・日本音響学会会員。



羽田 陽一

ヒューマンインタフェース研究所音声情報研究部研究主任
平成元年入社。主に音響エコーキャンセラの研究開発・実用化に従事。
昭和62年東北大学理学部物理学科卒業。平成元年同大学院物理学専攻修士課程修了。
電子情報通信学会・音響学会・IEEE会員。



島内 末廣

ヒューマンインタフェース研究所音声情報研究部社員
平成5年入社。音響エコーキャンセラの研究開発に従事。
平成3年東京工業大学工学部電子物理工学科卒業。5年同大学院総合理工学研究科物理情報工学専攻修士課程修了。
日本音響学会会員。

■用語解説■

音響エコーキャンセラ

エコーキャンセラはエコーの生じる伝搬経路の特性を推定し、擬似エコーを生成することにより、それを消去する装置である。伝送回線のエコー経路を対象とするものを回線用エコーキャンセラ、拡声通話系における音響経路のエコーを対象とするものを音響エコーキャンセラと呼んでいる。通常、音響エコーのほうが長く持続するため、回線用に比べ、ハードウェアの規模が大きくなることが多い。主としてテレビ会議装置に用いられる。

音声スイッチ

双方向通話において、送話と受話のレベルを比較して送話のレベルが高いときは送話状態と見なして受話側に損失を挿入し、反対に受話のレベルが高いときは受話状態と見なして送話側に損失を挿入することによりハウリングやエコーを抑圧するものである。同時通話時にはどちらかに必ず損失が挿入されるため、話頭切断や話尾切断を生じる。アナログの簡単な回路で構成できるため、ハンドフリー電話機でよく用いられている。

適応アルゴリズム

適応アルゴリズムとは、各時刻において観測される入力信号と、その入力信号に対する出力誤差に基づいて、適応フィルタの係数を逐次的に修正していく処理手順である。エコーキャンセラに使用される適応アルゴリズムとしては、学習同定法(NLMS(Normalized Least-Mean-Squares)アルゴリズム)が最も一般的であるが、音声に対する収束速度が遅いという欠点を持つ。近年検討されている高速RLS(Recursive Least-Squares)アルゴリズムは、数値的に不安定であり、まだ実使用に耐えない。現状では、射影アルゴリズムが最も有望である。

エコー消去時間長

エコーキャンセラの機能仕様を表す1つの尺度であり、どのくらいの長さのエコーを消去できるかを表している。標準化周波数とデジタルフィルタの規模(タップ数)で決まる数字で、例えば、標準化周波数が8 kHzでタップ数が1,000であれば125msのエコー消去時間長となる。一般に残響時間の長い(響きの多い)場所ではエコーが長くなるため、エコー消去時間長の長いものを使う必要がある。

タップ数

エコーキャンセラの規模を表す1つの尺度である。エコーキャンセラは、通常、FIR(Finite Impulse Response)型デジタルフィルタで構成されており、タップ数はフィルタ係数(乗算器)の数に相当し、タップ数の多いものは長いエコーに対しても消去可能となるが、ハードウェア規模も大きくなることになる。音響エコーキャンセラでは800から2,000、回線用では150から500程度のタップ数が普通である。