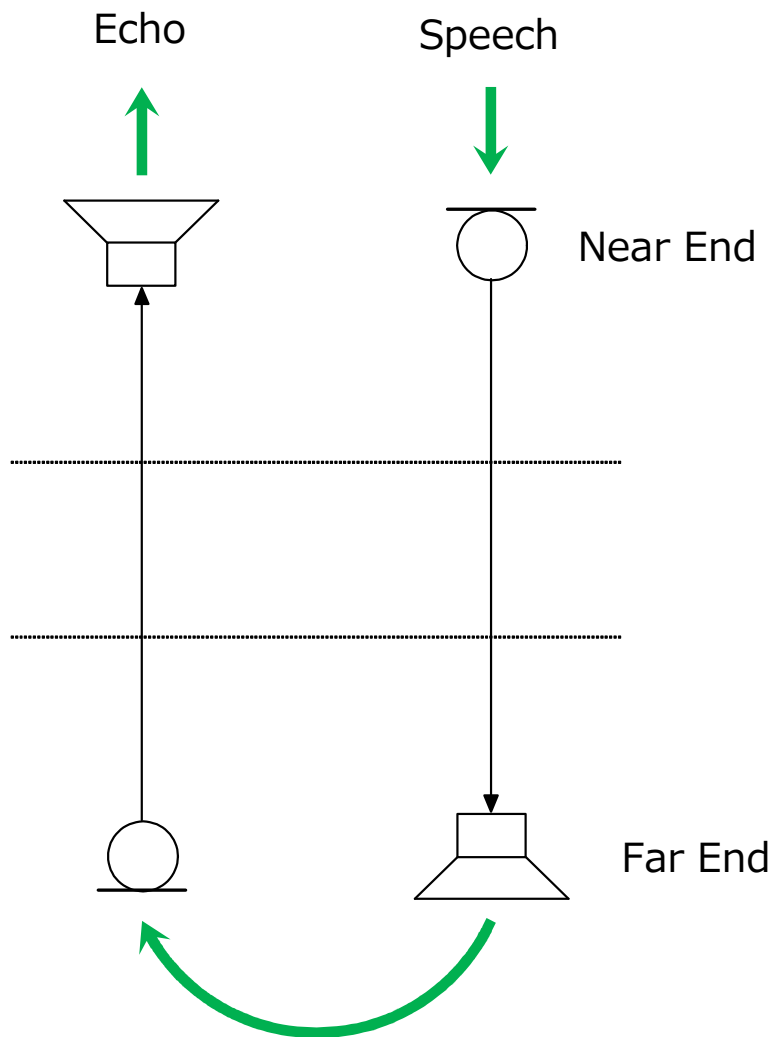
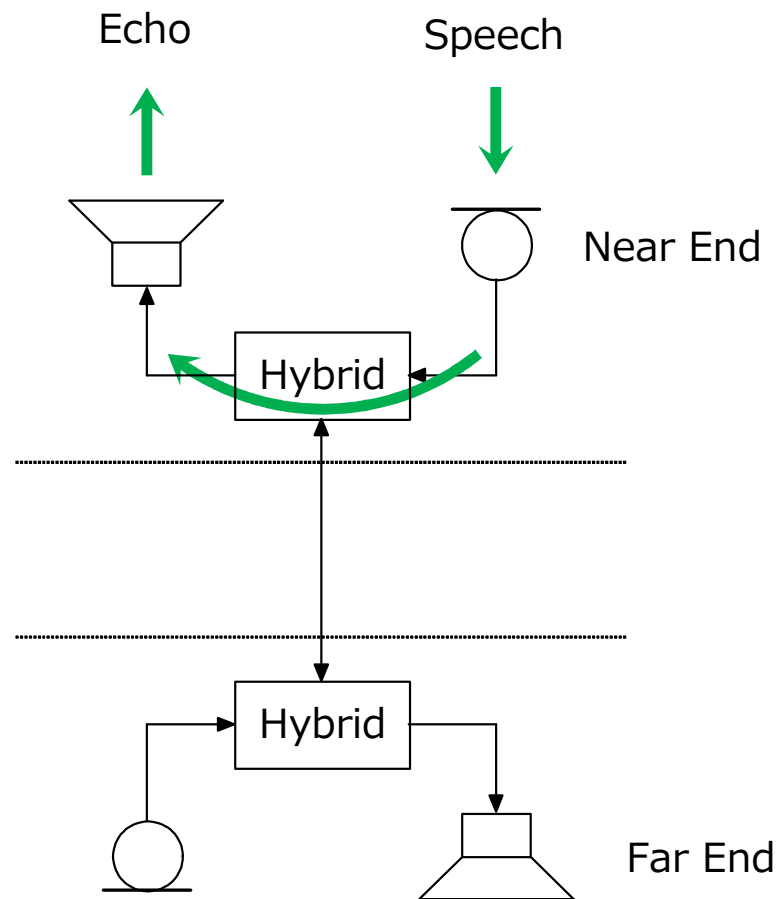

エコーキャンセラ向けの Normalized LMSアルゴリズムの 改良

■ エコー・キャンセラ



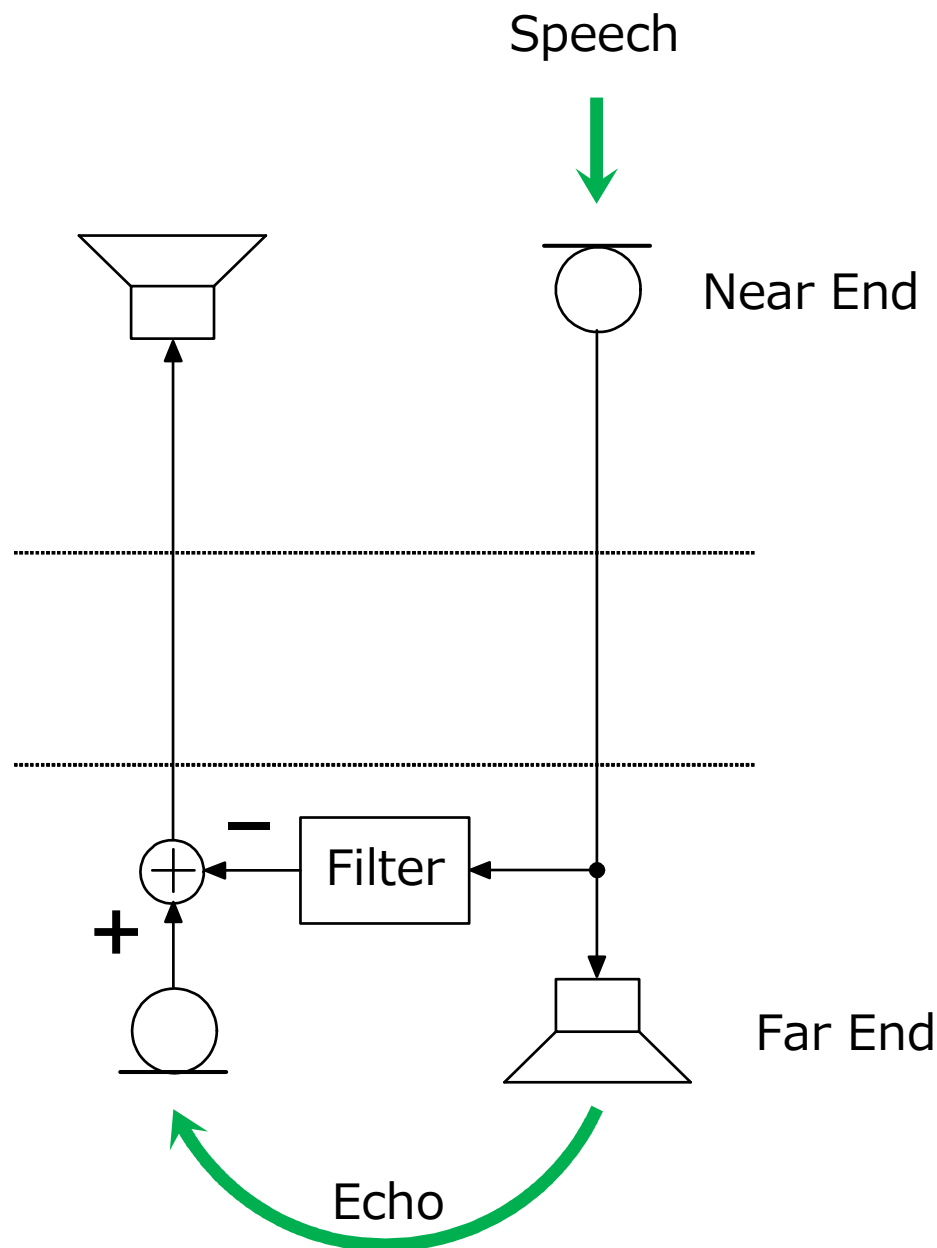
音響エコー・キャンセラ



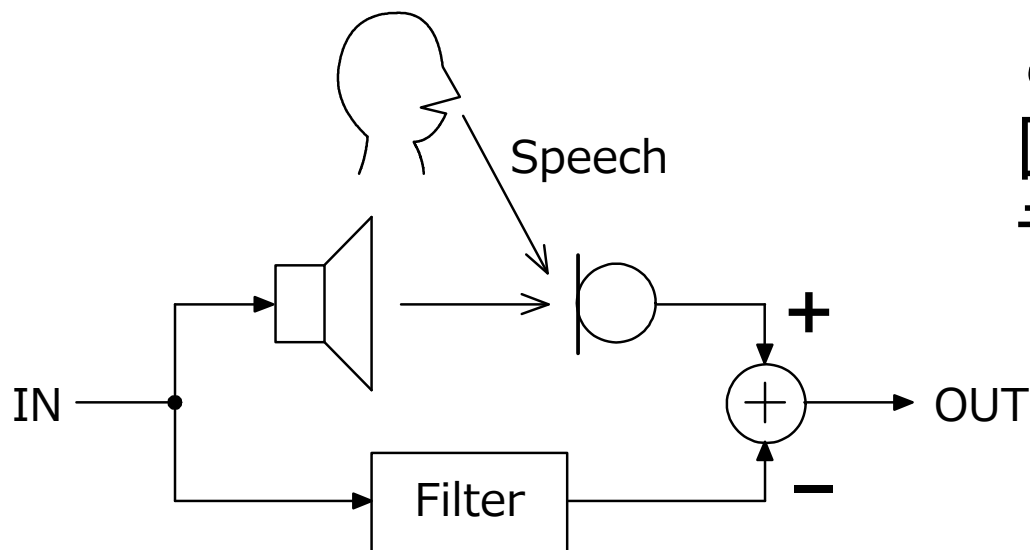
回線エコー・キャンセラ

■ 音響エコー・キャンセラ

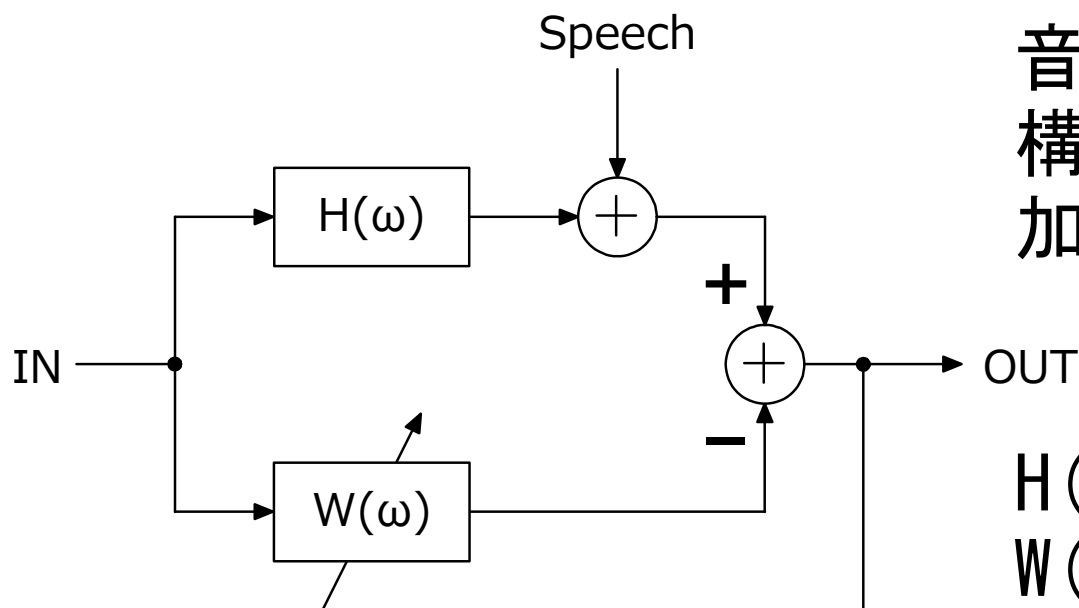
- ・ フィルタで疑似エコー信号を生成して引き算処理でエコーを打ち消す
- ・ フィルタはスピーカ、マイクを含む音響系の伝達特性を近似する
- ・ 音響系の伝達特性は常に揺らんでいる



■ 音響エコー・キャンセラのモデル



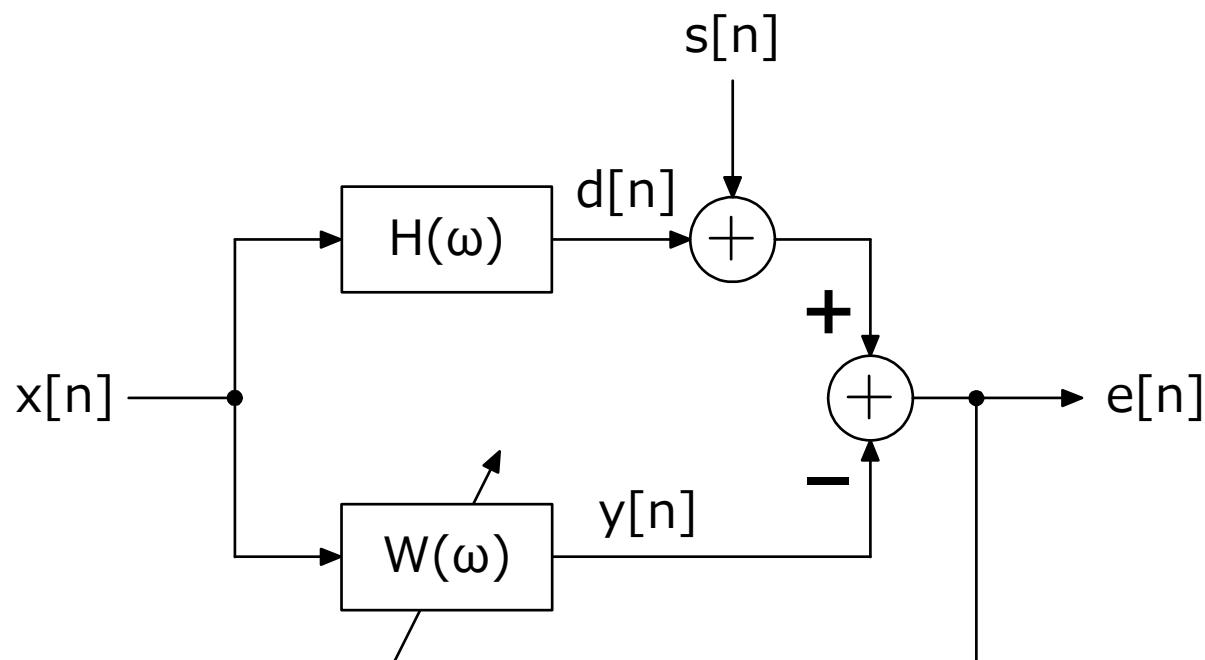
この図のspeechは前ページの図の通信相手 (Far End) の音声



音声はシステム同定の構成の適応システムへの加法性雑音となる

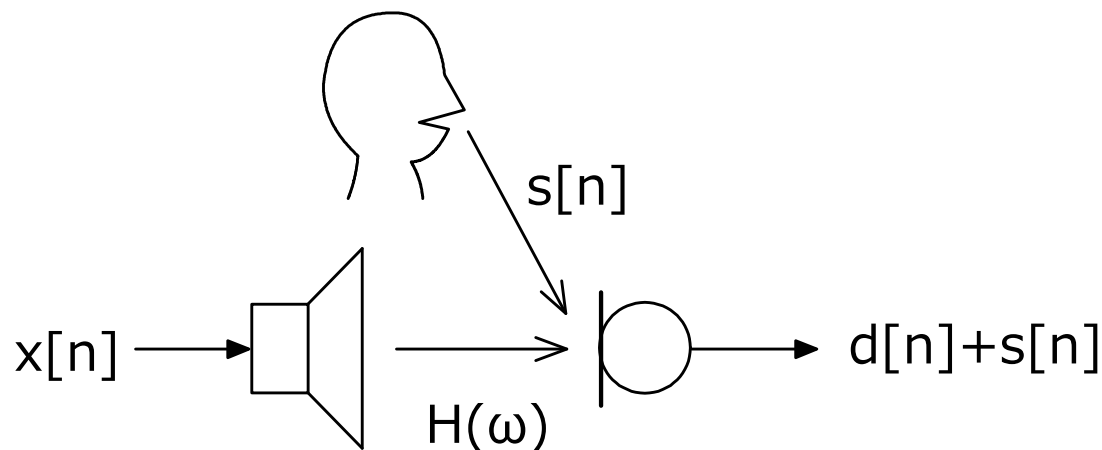
$H(\omega)$: 音響系の伝達特性
 $W(\omega)$: 適応フィルタ

■ エコー・キャンセラの技術的課題



- $s[n]$ によって収束が乱される (ダブルトーク)
- $x[n]=0$ であればエコー・キャンセラは不要 (?)
しかし, その間の $H(\omega)$ の変動に適応フィルタが追従出来ないので $x[n] \neq 0$ の状態に戻った時に困る
- ハウリング

■ ダブルトーク問題

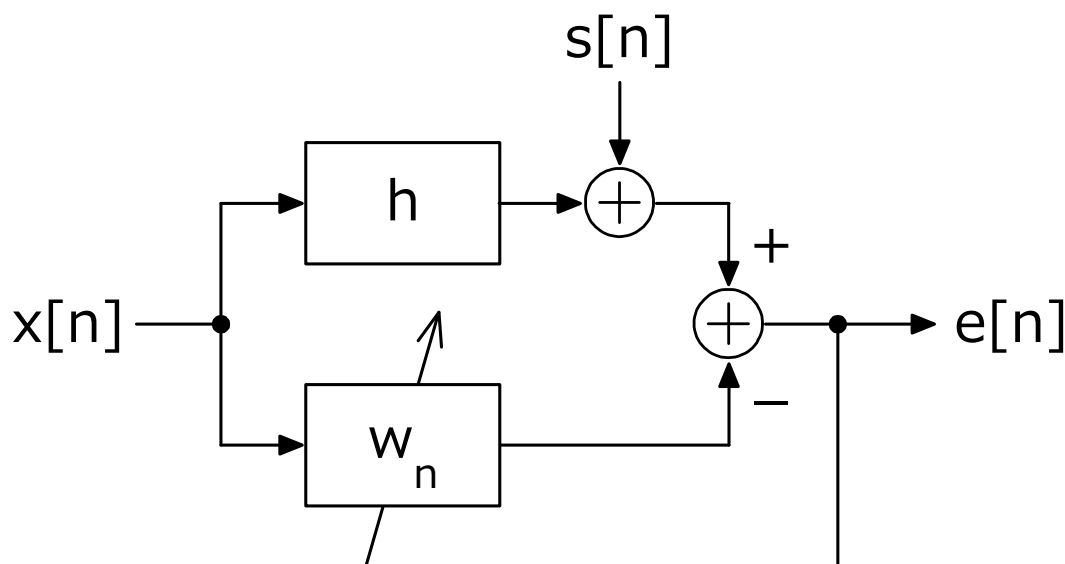


- 全二重通信システムでの自然な通話ではかなりの頻度でダブルトーク状態が生ずる ($s[n] \neq 0$ かつ $x[n] \neq 0$)
- $s[n]$ のみの観測は不可能 ($d[n] + s[n]$ は観測可能)
 - 加法性雑音が収束特性に与える悪影響の小さいアルゴリズムで対処？
 - $s[n]$ を推測して引き算で $d[n] + s[n]$ から除去？
 - $s[n]$ のパワーを推測して適応動作をオン/オフ？

■ エコー・キャンセラ製品の適応アルゴリズム

- ・ 基本はLMSアルゴリズム（時間領域での処理，または周波数領域でのブロック処理） + α

$$w_{n+1}[i] = w_n[i] + 2 \cdot \mu \cdot e[n] \cdot x[n-i]$$



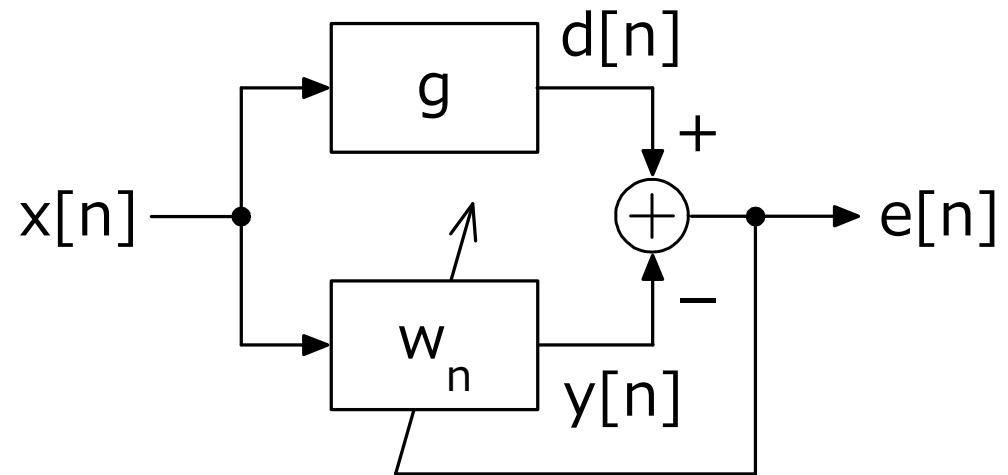
- ・ Normalized LMSはダブルトークに弱いのでN. G.
- ・ ローエンドの製品に高度なアルゴリズムは使われない

■ 1tap 適応フィルタ (LMS, NLMS)

$$d[n] = g \cdot x[n]$$

$$y[n] = w_n \cdot x[n]$$

$$e[n] = d[n] - y[n]$$
$$= (g - w_n) \cdot x[n]$$



$$[\text{LMS}] \quad w_{n+1} = w_n + 2 \cdot \mu \cdot e[n] \cdot x[n] \cong w_n + 2 \cdot \mu \cdot (g - w_n) \cdot x^2[n]$$

$$[\text{NLMS}] \quad w_{n+1} = w_n + 2 \cdot \mu \cdot \frac{e[n] \cdot x[n]}{x^2[n]} \cong w_n + 2 \cdot \mu \cdot (g - w_n)$$

- ・ 加法と乗法のみを用いて、漸近法で割り算 $d[n]/x[n]$ の結果を求めていることになる
- ・ LMS/NLMSはアバウトだがシンプルで実用的

■ エコー・キャンセラ研究の現状

- ・ 電話のエコー・キャンセラはとりあえずは現状の技術で間に合っている（抑圧しきれなかったエコーのために通話品質が常時著しく低下するほどではない）
 - 機器メーカーは対症療法的な処理手法改良に終始
- ・ 大学の研究者の多くは信号処理アルゴリズムの理論屋さん（聴覚・音声の専門家では無い）
 - 適応アルゴリズム研究のデモンストレーションの対象としてエコー・キャンセラを利用
 - 音声処理手法としての性能評価が非常に甘い
- ・ 「通信方式」ではないので規格化する必要が無い・規格化されない → 研究開発に競争原理が働かない

■ 音声圧縮伸張方式の研究開発

- ・ デジタル音声通信・音声伝送の基幹技術
 - 規格が定まらなければ製品開発，サービス開始が出来ない（自分の処理と相手の処理との整合性）
- ・ 公的な規格策定にあたって各社・各研究機関が開発した処理手法を持ち寄ってコンペを実施
 - 大規模な主観評価試験により白黒をつけた
- ・ 規格化作業を通じて研究成果が整理・定式化されて技術が急速に進歩した
 - 学術的研究と工学的応用・実用化が同時に進展

■ エコーキャンセラ向けに改良した NLMS アルゴリズム

NLMS
$$w_{n+1}[i] = w_n[i] + 2\mu \frac{e[i]x[n-i]}{x^2[n] + c}$$

改良 NLMS (1)
$$w_{n+1}[i] = w_n[i] + 2\mu \frac{e[i]x[n-i]}{\sqrt{x^2[n] e^2[n] + c}}$$

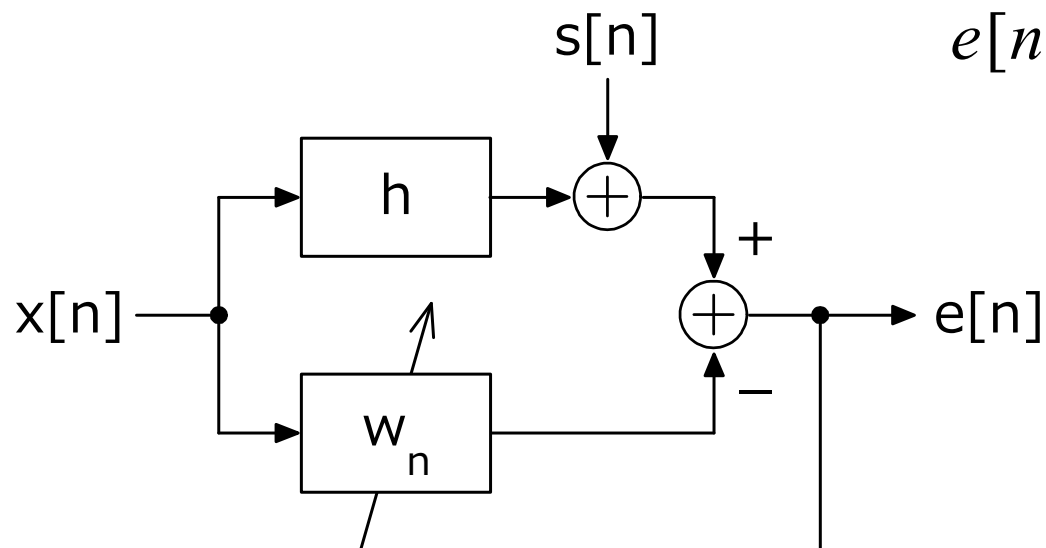
改良 NLMS (2)
$$w_{n+1}[i] = w_n[i] + 2\mu \frac{e[i]x[n-i]}{|x[n]e[n]| + c}$$

c は分母が0になって発散するのを防ぐ微少な値の定数

■ ダブルトーク問題への対処と改良 NLMS の発想

- ・ 理論的に外乱に強い優れたアルゴリズム（全最小二乗法, カルマンフィルタ etc）
 - ローエンドのアプリケーション向けでは無い（演算量, 必要とされる演算精度 etc）
- ・ エコー・キャンセラへのダブルトーク対応処理の追加
 - 現実の音声では成り立ちがたい仮定・前提条件に基づく複雑精緻な手法が多い（音声処理として??）
- ・ 「アバウト」な処理で良いから, ダブルトーク状態の時に自然と収束速度を低下させることは出来ないか？
 - 試行錯誤的にいろいろ試してみたらうまくいった

■ 改良 NLMS の特性（正規化処理の分母）



$$e[n] = x[n] * h[n] + s[n] - x[n] * w_n$$

$$= (h[n] - w_n) * x[n] + s[n]$$

ダブルトークなし
未収束 $h[n] \neq w_n$

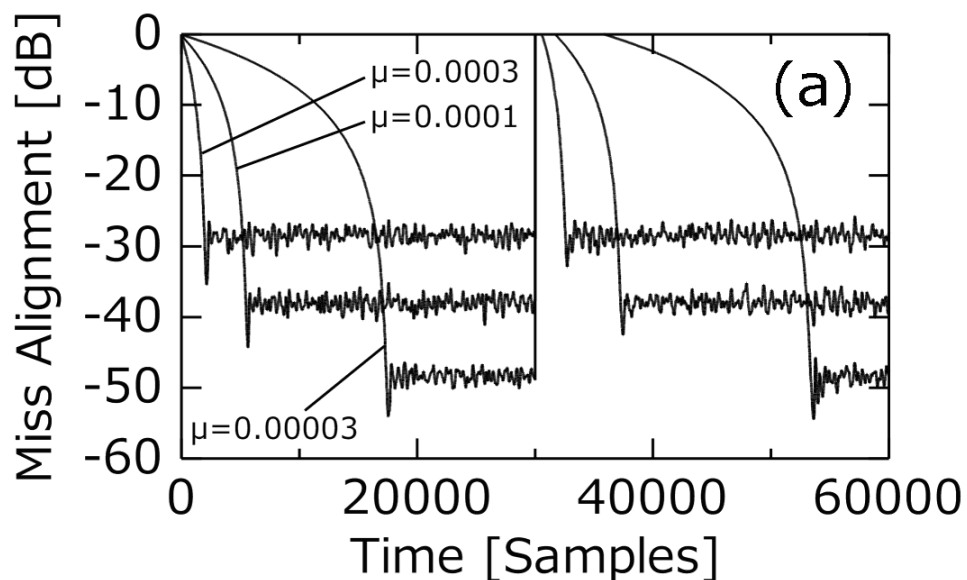
$$\sqrt{x^2[n] e^2[n]} = \sqrt{x^2[n] \{(h[n] - w_n) * x[n]\}^2}$$

$$\approx x^2[n] |h[n] - w_n|$$

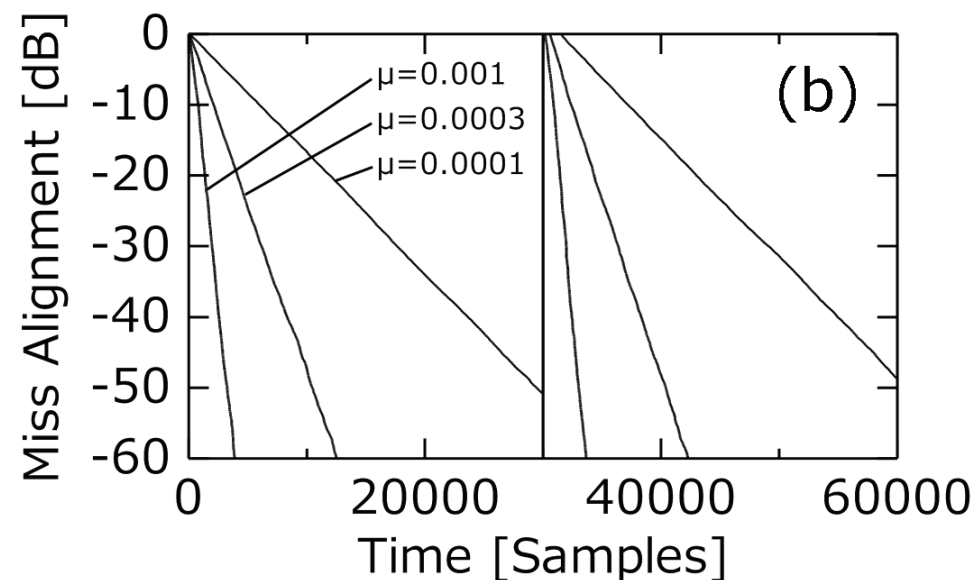
ダブルトーク状態
収束済 $h[n] = w_n$

$$\sqrt{x^2[n] e^2[n]} = \sqrt{x^2[n] s^2[n]}$$

■ シミュレーション結果（ダブルトークなし $s[n]=0$ ）



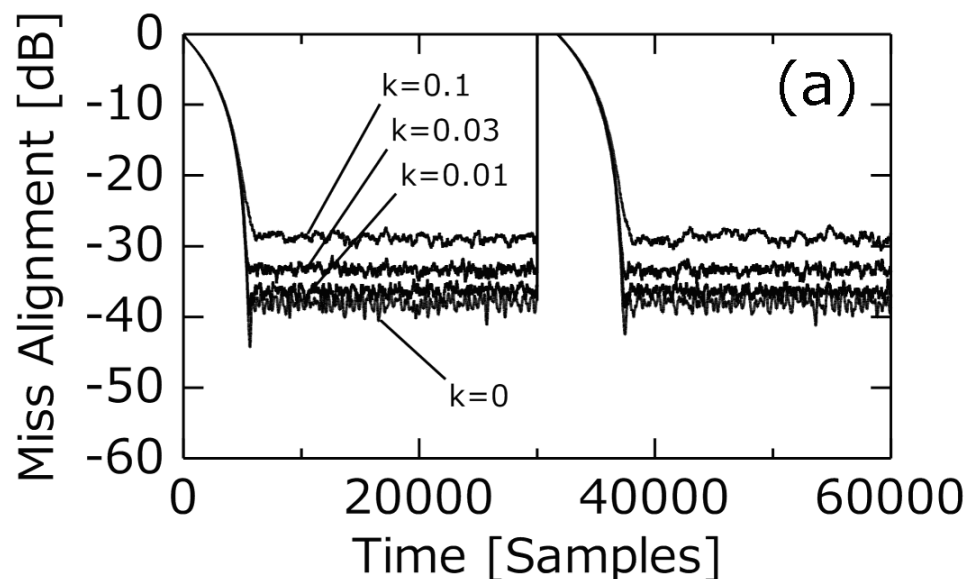
改良 NLMS



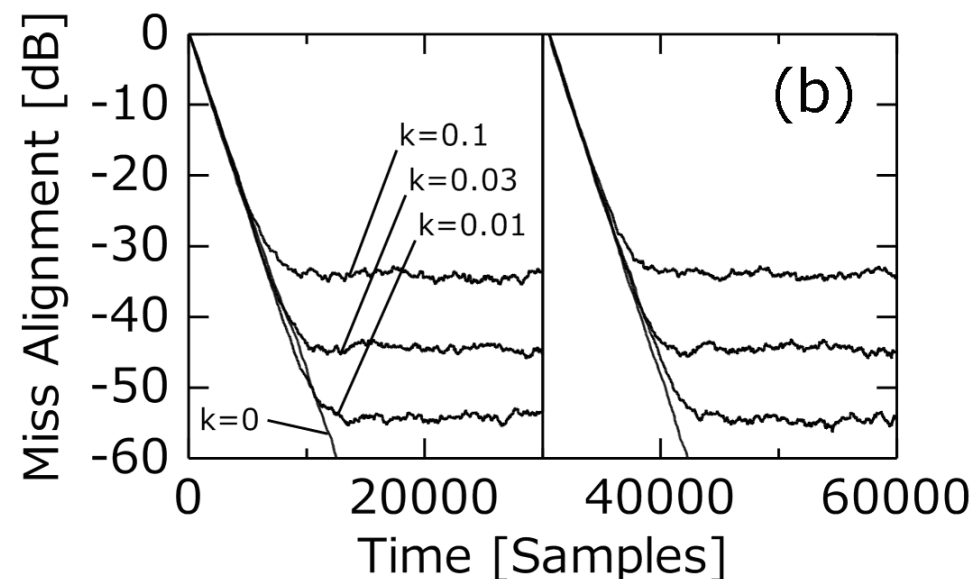
LMS

- ・ 改良 NLMS は過剰誤差が多い
- ・ しかしエコー減衰量にのみ着目するならば, LMS ほどの収束特性は必要無い~という見方も出来る
(電話等ならばエコー減衰量は -30dB 程度で十分)

■ シミュレーション結果（ダブルトーク状態 $s[n] \neq 0$ ）



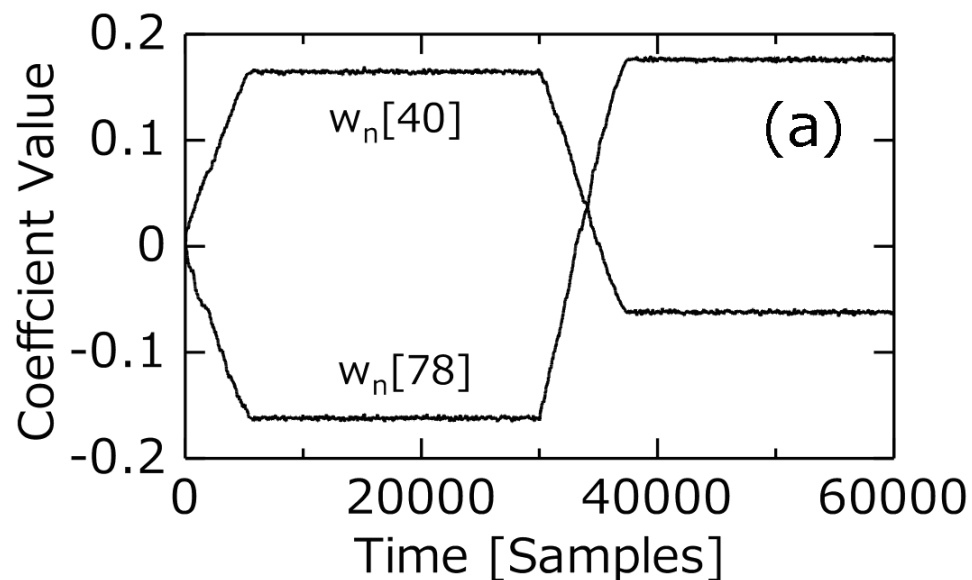
改良 NLMS ($\mu = 0.0001$)



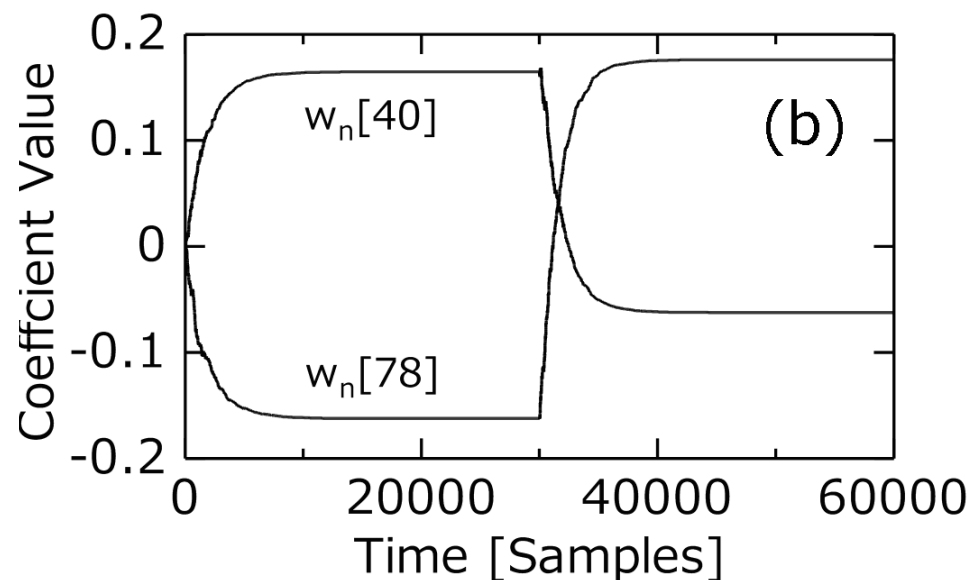
LMS ($\mu = 0.0003$)

- ・ k はダブルトークの程度（遠端信号 $x[n]$ と近端信号 $s[n]$ のパワー比）をあらわす。 $k=0$ の時 $s[n]=0$
- ・ 改良 NLMS はダブルトーク状態になっても、過剰誤差の増加がゆるやか

■ シミュレーション結果（適応フィルタ係数の変化）



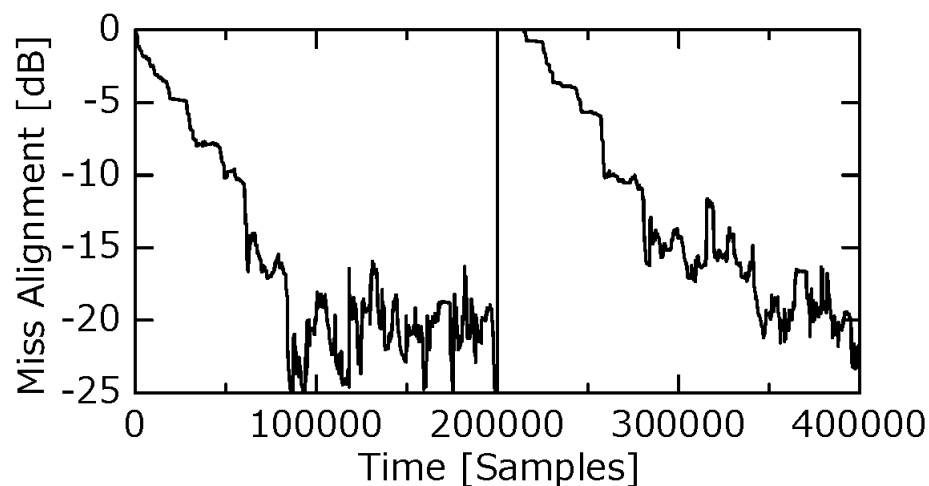
改良 NLMS



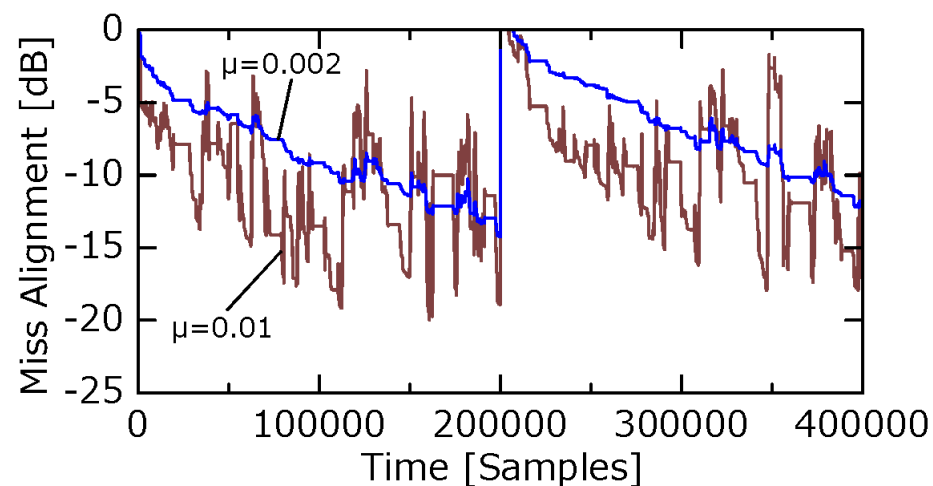
LMS

- ・ 改良 NLMS は適応フィルタ係数が直線的に変化する

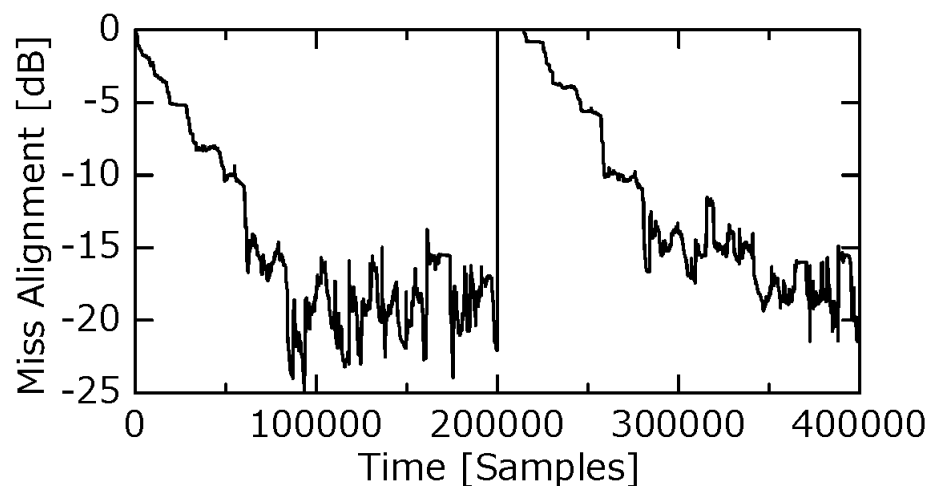
■ シミュレーション結果（音声入力，ダブルトーク）



改良 NLMS (1)



LMS

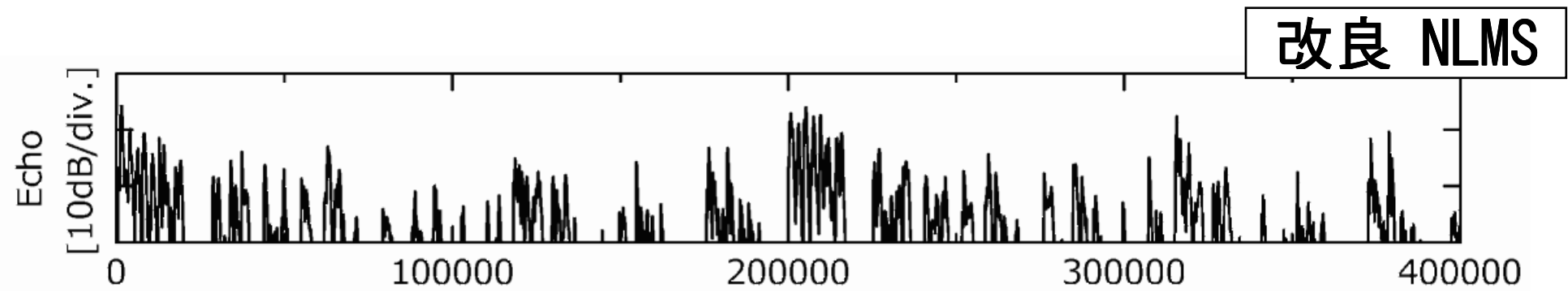
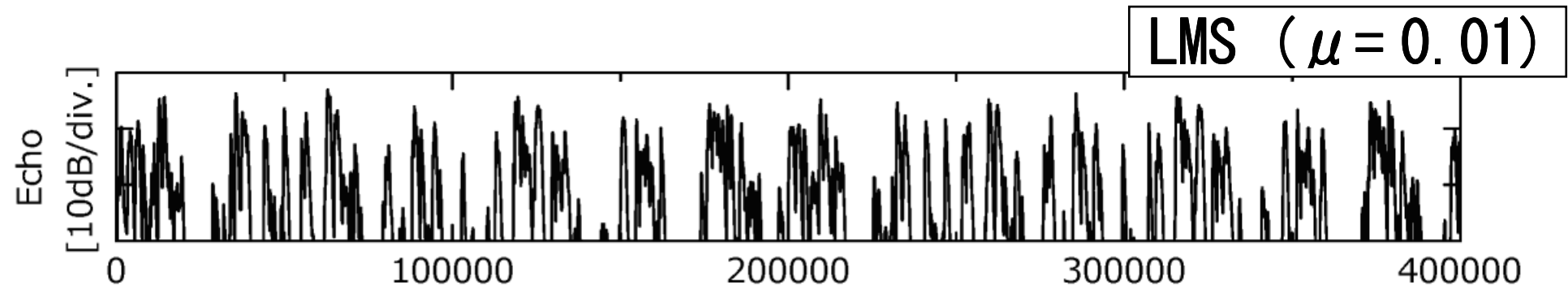
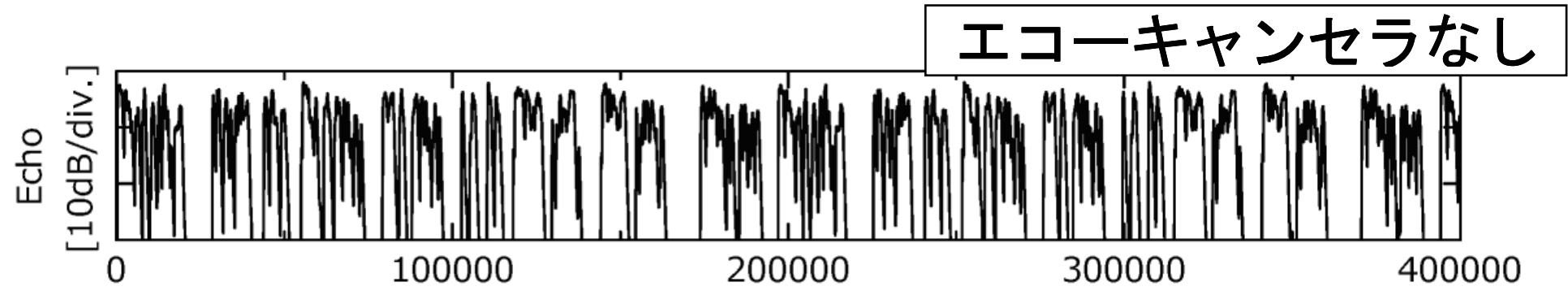


改良 NLMS (2)

改良NLMSは実際の音声
信号での特性が良好。
ただし初期の収束が
少し遅い。（ $f_s = 8\text{kHz}$ ）

■ シミュレーション結果 (エコー・レベルの変化)

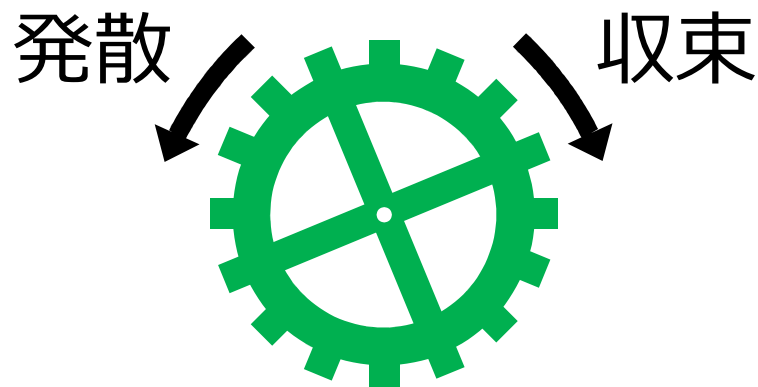
$f_s = 8\text{kHz}$



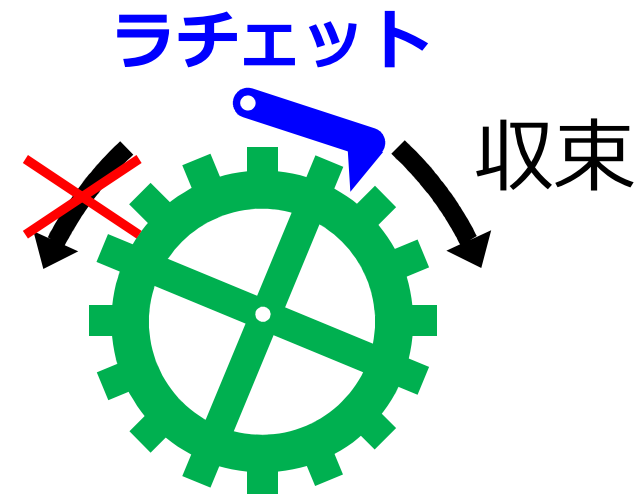
■ 模式的な説明

- ・ 誤差信号のパワー $e^2[n]$ の短時間平均による正規化処理は、歯車が一方向（収束）にしか回転しないようにするラチェットの働きをしている

従来の
適応アルゴリズム



改良 NLMS



- ・ 発散しそうになって、誤差信号のパワーが増えたとラチェットがかかる（適応動作の速度が低下する）

■ 改良 Normalized LMS のまとめ

- Normalized LMS アルゴリズムを小改良
 - 理論的にも素直なアルゴリズムの拡張
 - 演算量は微増
- 適応アルゴリズム自体を『ダブルトーク対応』に拡張
 - ダブルトーク検出処理は不要（になるはず）
- 実際の音声では成り立ちがたい仮定・前提に基づいた複雑精緻な処理ではない
 - 安定した動作が望める
- 今後さらに実使用条件での従来手法との詳細な比較評価が必要（通信事業者や端末機器メーカーでないとなかなか難しいが....）

■ 適応アルゴリズムの向き・不向き

- 理論的に「高度な」処理手法（LMS/NLMSより演算量が1ケタ増える代わりに、収束特性が1ケタ以上向上）
 - 「普通の電話」程度には過剰な性能（コストをかけて実装しても使用者は付加価値を認めてくれない、高度な機能を正當に評価する能力が無い）
 - 法人用TV電話会議システム等の高級な用途向け
- 改良 Normalized LMS
 - わずかな演算量増加でリーズナブルな性能改善（ダブルトーク検出処理が不要になれば効用大）
 - ハンズフリー・インターホン等の応用に有望？

■ 市販のエコーキャンセラ I P のカタログ表記

- ・ エーアイコーポレーション

「100msec以内瞬間収束の80dBシングルトークエコー削減」

「ユーザ調整可能なダブルトークエコー削減」

→ 手を抜いた説明文（嘘は言っていないが....）

- ・ ルネサステクノロジ

「周波数領域ブロック適応フィルタ」

「エコー消去時間：～60ms」

「エコー除去量：～30dB」

→ 正直なスペック表記（エコー消去時間は??）