Stereophonic acoustic echo cancellation
- an overview and recent solutions -

Shoji Makino and Suehiro Shimauchi

NTT Lifestyle and Environmental Technology Laboratories

3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa, 243-0198 Japan
Problem:

- No noise, strong and fixed crosscorrelation
  ⇒ non-uniqueness problem

Solutions:

(1) Inaudibly-processed independent noise
    - variations in crosscorrelation

(2) Adaptive algorithms
  ⇒ emphasize the independent noise
    - non-fixed crosscorrelation

(3) Subband structure
1. INTRODUCTION

- Stereo system: realistic presence

- Spatial information: distinguish the talker

- Stereo echo cancellers: full-duplex communication
Why Multi-Channel?

Single-channel cannot distinguish the talker!
Multi-channel can distinguish the talker!
2. FUNDAMENTAL PROBLEM OF STEREO ECHO CANCELLATION

2.1 Stereo echo cancellation

\[ x_1(k) \text{ and } x_2(k): \text{strong and fixed crosscorrelation.} \]
2.2 Non-uniqueness problem

- Misconvergence, slow convergence
  [Sondhi, SP Letters95]

- Suffers from
  - near-end echo path
  - far-end transmission path

- Goal: identify the true echo path
• Minimization of the weighted least squares criterion,

\[ J(k) = \sum_{l=1}^{k} \lambda^{k-l}e^2(l), \quad (1) \]

leads to the normal equation

\[ \mathbf{R}(k)\hat{\mathbf{h}}(k) = \mathbf{r}(k), \quad (2) \]

where

\[ \mathbf{R}(k) = \sum_{l=1}^{k} \lambda^{k-l}\mathbf{x}(l)\mathbf{x}^T(l) = \begin{bmatrix} \mathbf{R}_{11}(k) & \mathbf{R}_{12}(k) \\ \mathbf{R}_{12}(k) & \mathbf{R}_{22}(k) \end{bmatrix}, \quad (3) \]

\[ \mathbf{r}(k) = \sum_{l=1}^{k} \lambda^{k-l}\mathbf{y}(l)\mathbf{x}(l). \quad (4) \]

• If input signals are denoted as

\[ x_1(k) = g_1(k) \ast s(k) \quad \text{and} \quad x_2(k) = g_2(k) \ast s(k), \quad (5) \]

• If there is no noise and \( g_1(k) \) and \( g_2(k) \) are time-invariant
  – (3) is not full rank
  – (2) has an infinite number of solutions

• \( \hat{\mathbf{h}}(k) \) is "uniquely" determined in the sense of a minimum-norm solution

• \( \hat{\mathbf{h}}(k) \neq \mathbf{h}(k) \).
- Simplest example,

\[ x_1(k) = a_1 s(k) \quad \text{and} \quad x_2(k) = a_2 s(k), \quad (6) \]

- If the initial value \( \hat{h}(0) \) is set to zero vector,

\[ \hat{h}_{1a}(k) = \frac{a_1^2}{a_1^2 + a_2^2}[h_1(k) + \frac{a_2}{a_1} h_2(k)] \neq h_1(k) \quad (7) \]

\[ \hat{h}_{2a}(k) = \frac{a_2^2}{a_1^2 + a_2^2}[\frac{a_1}{a_2} h_1(k) + h_2(k)] \neq h_2(k). \quad (8) \]
2.3 Performance of conventional stereo echo canceller

(a) ERLE

(b) Coefficient error
3. A CLUE FOR TRUE ECHO PATH ESTIMATION

(a) ERLE

(b) Coefficient error
Three clues:

(1) Independent noise

(2) Impulse response tail [Benesty, ICASSP95]

(3) Variations in the crosscorrelation
3.3 Variations in the crosscorrelation

- Crosscorrelation varies slightly

- "new" convergence starts from the "old" misconverged solution.
- Coefficient error becomes smaller
- Converge to the "true" solution [Shimauchi, ICASSP95]
4. RECENT SOLUTIONS

4.1 Addition of independent noise and variations in crosscorre

- Function block [Shimauchi, ICASSP95]
4.2 Nonlinear processing

- Nonlinear processing [Benesty, ICASSP97]

\[ x'(k) = x(k) + \alpha f[x(k)]. \tag{9} \]
- Exclusive adaptive filters [Shimauchi, ICASSP98]
4.3 Noise shaping

- Addition of spectrally shaped random noise
  [Gilloire, ICASSP98]

Figure 2. Principle of the proposed method
• Utilization of the MPEG audio coder

[Gansler, ICASSP98]

Figure 1: Audio coder and Stereophonic AEC. Only one return part is shown.
4.4 Decorrelation filters

- Decorrelation filters,

\[ x'_1(k) = x_1(k) - f_2(k)x_2(k) \quad (10) \]
\[ x'_2(k) = x_2(k) - f_1(k)x_1(k) . \quad (11) \]

- \( x'_1(k) \) and \( x'_2(k) \):
  filtered versions of the same signal \( s(k) \)

- No noise and fixed-crosscorrelation:
  \( x'_1(k) = 0, \; x'_2(k) = 0 \)
• Noise and variations:

Adjustment vectors $x'_1(k) \perp x_2(k)$, $x'_2(k) \perp x_1(k)$

$$x'_1(k) = x_1(k) - \frac{x_2^T(k)x_1(k)}{x_2^T(k)x_2(k)}x_2(k) \quad (12)$$

$$x'_2(k) = x_2(k) - \frac{x_1^T(k)x_2(k)}{x_1^T(k)x_1(k)}x_1(k) \quad (13)$$

• $x'_1(k) \not\perp x'_2(k)$

• Removes the correlated speech

• Emphasize the noise and variations
4.5 Time-varying filtering

• Time-varying all-pass filtering [Ali, ICASSP98]

Figure 2: Configuration of the modified stereophonic echo cancellation systems
• Two-tap time-varying filter [Joncour, ICASSP98]

Fig. 1: The proposed stereo echo canceler for teleconferencing.
4.7 Comparisons

- Additive noise: degradation of speech

- Additive variations: degradation of stereo perception
4.8 Subband processing

- Convergence improved [Makino, ICASSP97]
5. ADAPTIVE ALGORITHMS

5.1 Emphasis of independent noise and variation in crosscorre

- Noise and the variations:
  adapt the filters toward convergence

- Speech: disrupts the adaptation

- Emphasize the noise and variations
  accelerate the convergence
• We can treat $x(k)$ as,

$$x(k) = v(k) + \tilde{v}(k) \quad [v(k) \in S, \, \tilde{v}(k) \perp S], \quad (14)$$

• $\hat{h}(k)$: adjusted in the direction of $\tilde{v}(k)$
Stereo algorithms

(1) RLS algorithm

(2) Stereo projection algorithm
   [Shimauchi, ICASSP95], [Benesty, SP Letters96]

(3) Extended algorithms
5.4 Extended algorithms

- Extended LMS algorithm [Benesty, ICASSP95]

- Extended projection algorithm
  [Benesty, SP Letters96]

- Use the input-output relationships for reversed stereo signals [Shimauchi, ICASSP96]
6.1 Performance of the stereo projection algorithm

(a) Residual echo

(b) Coefficient error
6.2 Effect of subband processing

- Convergence improved considerably [Makino, ICASSP97]
6.3 Subband stereo projection algorithm [Makino, ICASSP97]

(a) Residual echo

(b) Coefficient error
- frequency range: 100 Hz to 20 kHz on DSPs
- number of taps: 1200 (0.1 - 4 kHz), 800 (4 - 8 kHz)
- adaptive algorithm: stereo projection algorithm
- 8 to 20 kHz: stereo voice switch
- Duo-filter control system [Haneda, EUSIPCO96]
7.1 Convergence at change of far-end talker position

(a) right channel $S_{out}$

(b) left channel $S_{out}$
High-Presence Communication
8. CONCLUSIONS

- Fundamental problem: non-uniqueness

- Three clues:
  - independent noise
  - impulse response tail
  - variations in crosscorrelation

- Recent solutions
  - function block
  - adaptive algorithms
  - subband structure

- Real world:
  - non-uniqueness ⇒ slow convergence

- Next step: sophisticated control