Stereo Echo Cancellation

Supervisor: Tsvi Dvorkind

Students: Yuval Dafna
Ido Loberman

2/8/2007
Stereo telecommunication systems are very common in daily life.

A major problem in stereo telecommunication is a returning echo from the far end room.

The project’s goal is to perform echo cancelling in a stereo channel.
The System

Fig. 1. Schematic diagram of stereophonic echo cancellation.

the returning echo (error) \[ e(n) = (h_1 - \hat{h}_1) * x_1(n) + (h_2 - \hat{h}_2) * x_2(n) \]
In Stereo case $e(n)=0$ implies:

$$ (h_1 - \hat{h}_1) * x_1(n) + (h_2 - \hat{h}_2) * x_2(n) = 0 $$

in Frequency domain:

$$ \left[ (H_1 - \hat{H}_1) G_1(f) + (H_2 - \hat{H}_2) G_2(f) \right] S(f) = 0 $$

Not necessarily $\hat{h}_1 \rightarrow h_1$, $\hat{h}_2 \rightarrow h_2$

We might track variations in the transmission room

**Solution:** Decorrelation reduces the cross-products and therefore minimizes MSE with respect to receiving channels.

$$ Ee^2 = E \left[ (h_1 - \hat{h}_1) x_1 \right]^2 + E \left[ (h_2 - \hat{h}_2) x_2 \right]^2 + 2E \left[ (h_1 - \hat{h}_1) x_1 (h_2 - \hat{h}_2) x_2 \right] $$

Fig. 1. Schematic diagram of stereophonic echo cancellation.
Transfer both of the channels through a nonlinear transformation in order to reduce correlation.

- Very simple, yet not highly effective technique

$$f(x[n]) = \begin{cases} x(n) & x(n) \geq 0 \\ 0 & x(n) < 0 \end{cases}$$

$$x'(n) = x(n) + \alpha f(x[n])$$

$$\alpha \in [0.3, 0.5]$$
Reduce correlation by cyclically inserting the original signal for a period $T/2$ and then inserting a delayed version of it.

$$c(k) = \begin{cases} 
1 & k \text{ mod } T < \frac{T}{2} \\
0 & \text{else} 
\end{cases}$$
New approach: reduce correlation by watermark embedding

- Speech signal divided into 512 samples sections, the DFT is analyzed and a masking threshold based on the Human hearing model is created.
- Here, the watermark is simply a generated white Gaussian noise, amplified to the threshold level.
- The purpose of this process is to decorrelate the two channels in order to minimize the required MSE.
- The method should not defect the audio performance.
The suggested Solution
Watermark Embedding + Sliding Input

- The combination of the two methods in order to decorrelate the two channels.
- Both channels go through a watermarking unit (WME) and one channel through the Sliding input unit (SI) afterwards.
- The new Technique gives much better results in the misalignment measure, while keeping a high audio quality.
The misalignment measures the convergence of the adaptive filter vector $\hat{h}(t)$ to the receiving room impulse response vector $h(t)$.

Mathematically, the misalignment is given by:

$$\text{misalignment}(t_0) = \frac{\|h(t_0) - \hat{h}(t_0)\|}{\|h(t_0)\|}$$
The effect of source location Change on the Misalignment Measure

- simulation for speaker’s location change from Point A to B in transmission room.

- The WME+SI is the Most robust technique for the location change.
**The Trade-Off**

**Misalignment Vs. Speech Quality**

- Perceptual Evaluation of Speech Quality (PESQ) is a measure for the speech quality ranked [0 5]. There is a tradeoff between the misalignment & the PESQ grade.

<table>
<thead>
<tr>
<th>PESQ</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Excellent</td>
</tr>
<tr>
<td>4</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
</tr>
<tr>
<td>2</td>
<td>Poor</td>
</tr>
<tr>
<td>1</td>
<td>Bad</td>
</tr>
</tbody>
</table>

![Graph showing misalignment (dB) vs. PESQ after N=60000](image)

- Watermark Embedding
- Sliding Input
- Watermarking + Sliding Input Q=1000
- Watermarking + Sliding Input Q=2000
Conclusions

- A new technique for stereo echo cancellation was developed, based on the sliding input technique and the watermark embedding approach.
- The new technique improves results in the misalignment measure, and therefore attains better results in the echo cancellation.
- The WME+SI technique is more robust to speaker location change in the transmission room.
- The new approach is capable of keeping a high-speech Quality (PESQ>3.6).
- For a given Misalignment constraint it has the lowest speech quality degradation.