The echo does not really affect the intelligibility of the conversation.

3.1 Echo Detector

from the near-end $x$ and one coming from the far-end $y$, supposedly belonging to the same speaker. The VAD flag informs us where the speech is present. 

First estimate of network delay $\hat{\tau}$:

$$\arg \max_{\tau} \sum_{i}(1)E[x(n + \tau) - \mu_x]y(n) - \mu_y)]^2$$

calculated for each possible delay $\tau = 0, \ldots, 50$ on each $i$th feature and averaged.

Iterative updating of the network delay estimate $\hat{\tau}$

$$cc_L(n, \tau) = \frac{1}{1 + \tau} \sum_{i=0}^{\tau} \frac{1}{\sqrt{E[x(n)g(y(n + \tau))]}E[y(n)]}$$

with $\tau = 0, \ldots, 20$. Done if at the $n$th subframe, $VAD_L(n) = 1$ and $VAD_L(n + \tau_0) = 1$. The two important values obtained are:

$$cc(n) = \text{max[cc_L(n, \tau)]}$$
$$\delta_\tau = \arg \max_{\tau} \text{cc}_L(n, \tau)$$

$\delta_\tau$ is used to update the delay, and $cc(n)$ is considered as the echo-likelihood parameter. The update of the delay by $\delta_\tau$ will be only done if $cc(n) > 0.85$.

3.2 Double Talk Detection

Two Gaussian pdfs for $cc(n)$ in the presence and absence of double talk are defined and then weighted by $P(DTD) = 0.05$ and $P(DTDD) = 0.95$. An optimal fixed threshold is then found $cc_{DTDD} = 0.42$.

Cancellation algorithms will only work for $cc(n) > cc_{DTD}$.

4 AEC Algorithms

The conditions for the AEC algorithm to be operative are that the voice activity detectors on the aligned temporal axis are both high $VAD_L(n + \tau_0) = 1$ and $VAD_L(n) = 1$ and only the echo is present $cc(n) > cc_{DTD}$.

4.1 $g_p$ and $g_n$ modifications

Use of Normalized Least Mean Square algorithm with step-size $1.5 \cdot cc(n)$ (for convergence)

General assumption: $g_p(n) \approx g(n) + g_0(n) + g_{0n}(n)$

Considering

$$g_p(n) = \frac{L-1}{L} g(n) - \frac{1}{L} h(l) = h^T g_i(n)$$

$h$ is being adapted at time $n+1$ with the following NLMS procedure:

$$h(n + 1) = h(n) + 1.5 \cdot cc(n) \frac{g_p(n) - \hat{g}(n)}{\sum_{i} g_i(n) g_i(n)}$$

Thus, the signal $g_p$ coming out of the canceller will be:

$$g_p(n) = g(n) - g_0(n) = g(n) - h(n) g_i(n)$$

4.2 $T_p$ and $\{L_i\}$ modifications

Eliminate the long-term information by randomizing the value of $T_p$:

$$T_p = \tau$$
$$\Omega_T = \{17, 18, \ldots, 142, 143\}$$

Whitening of the signal by morphing the LPC spectrum (done in the LF domain):

$$L_s(n) = cc(n) \frac{1}{\sum_{i} g_i(n) g_i(n)} (1 - cc(n))L(n)$$

5 Results

The main problems of the AEC algorithm implemented happen as the $ERL$ becomes too high or the $SNR$ becomes too low, however in these cases the echo does not really affect the intelligibility of the conversation.

6 Conclusion

It is possible to transpose AEC operations from time domain to parameters domain

Suitable for implementation in speech enhancement equipments in voice networks and using AMR coded speech

A good alternative to the existing AEC procedures

References
