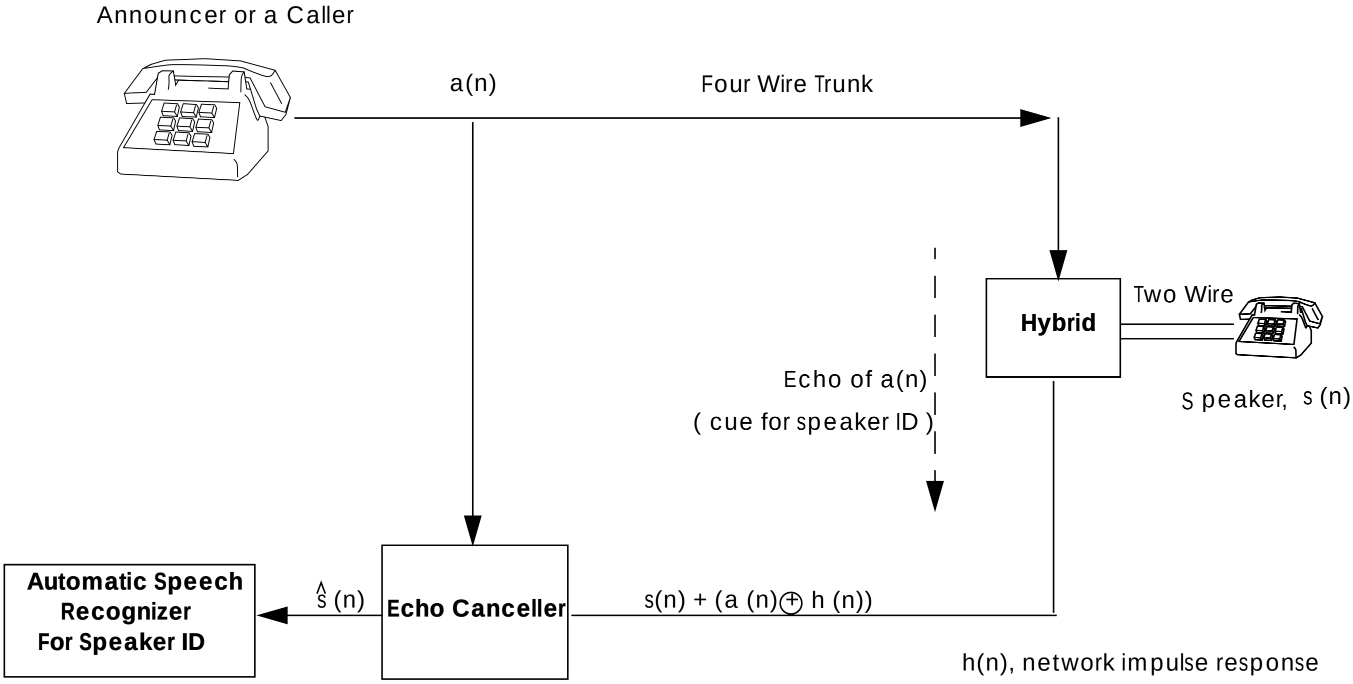
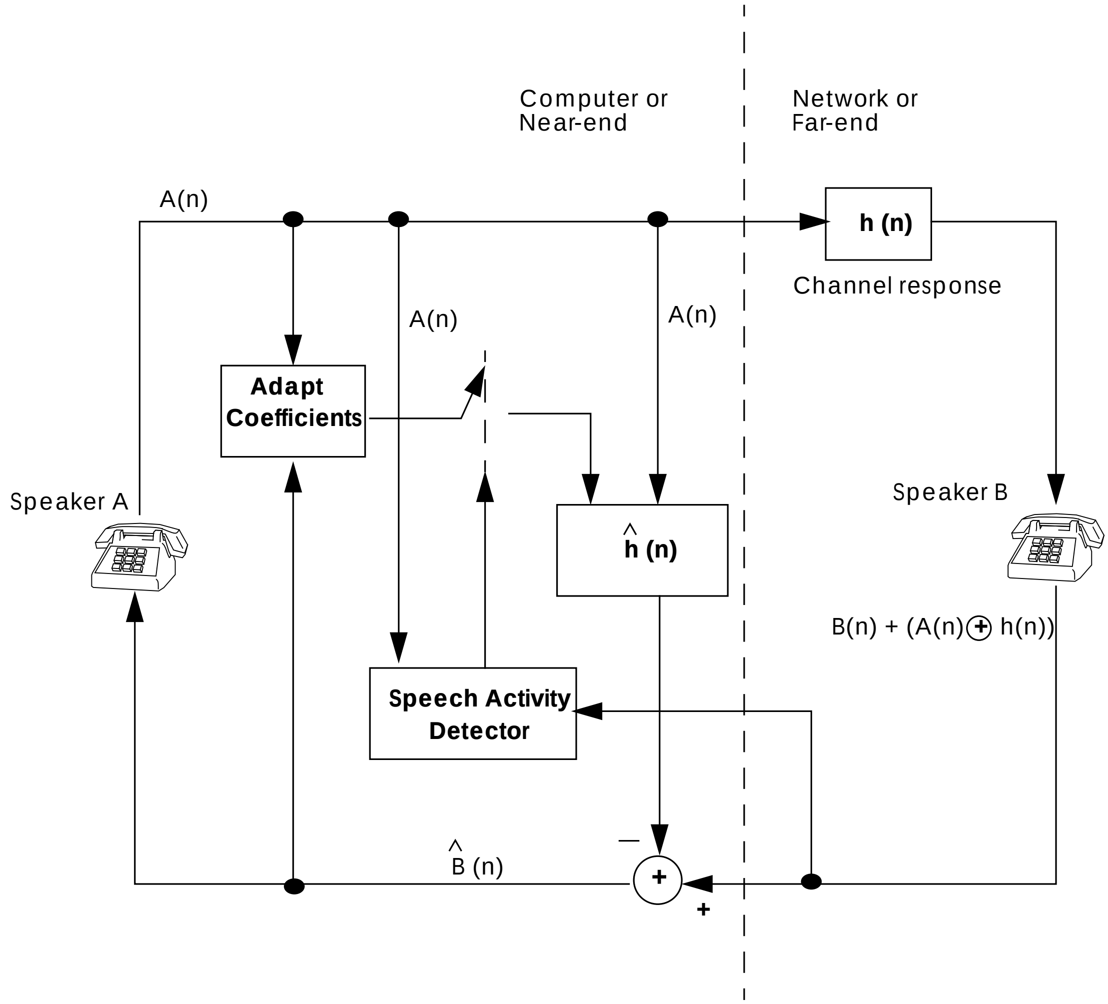
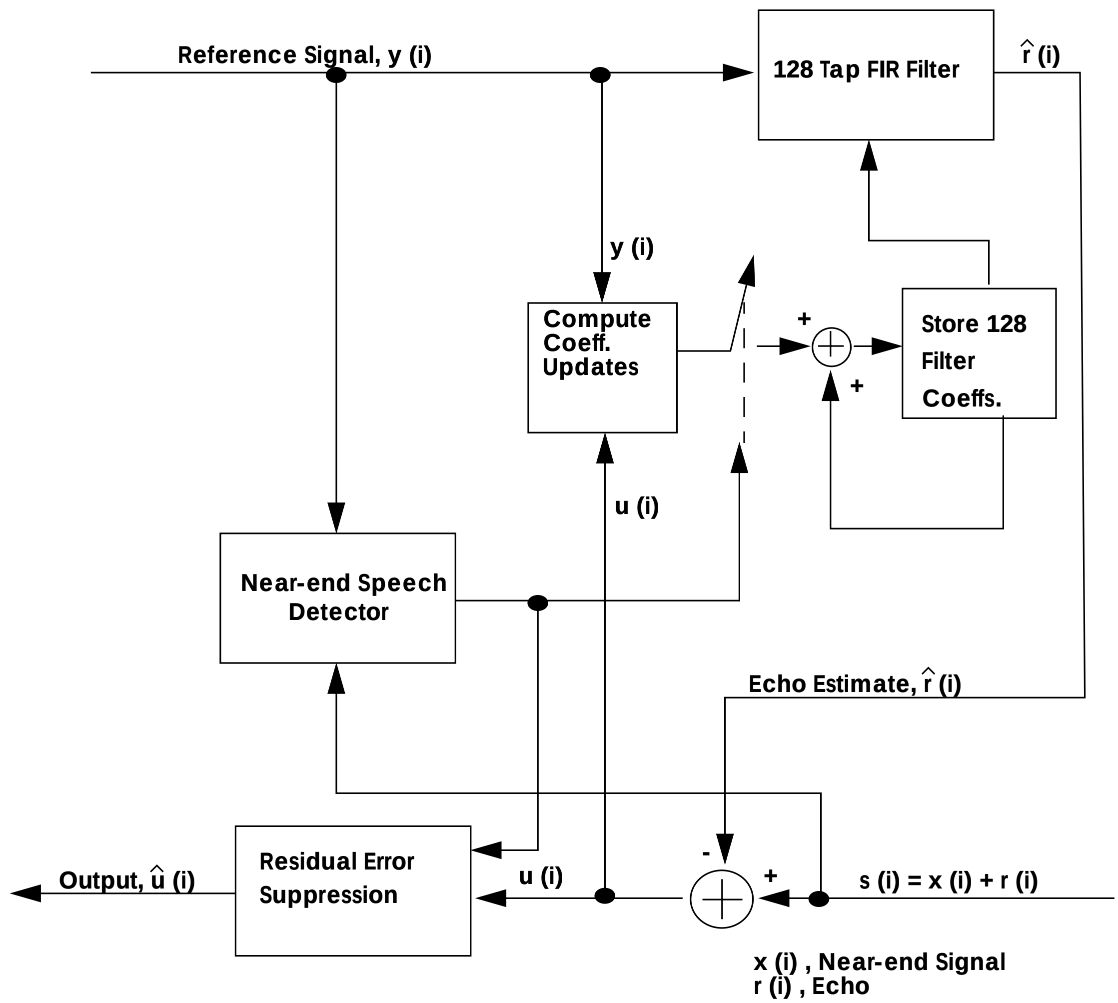



**Need For Echo Cancellation In SWITCHBOARD Data!**




**Echo Cancellation for Two-way Spontaneous Speech**




**Block Schematic of an FIR-Echo-Canceller**

This year at the ARPA Speech Recognition workshop, the Speaker-ID technology of the participating sites will be evaluated on the SWITCHBOARD data. SWITCHBOARD is a spontaneous two-way telephone speech data corpus. The data consists of one reference channel or the far-end channel and the other near-end channel. Due to the irregularities of the hybrids in the telephone networks, an echo of the far end speech gets added to the near-end speech.

This echo could be used by the speech recognizer to gather important cues regarding the ID of the speaker or the channel conditions, thereby making the job that much easier. To eliminate this problem, we need an efficient echo cancellation technology.

We have developed an FIR echo-canceller for this purpose. The block diagrams above give a good picture of the process of echo-cancellation. We have deviated from the standard implementation of an LMS echo-canceller at places to accommodate certain problems we face. Some of the main problems we encountered during the development of the system are:

- **Double talk**: This is a condition when both the speakers talk simultaneously. If we adapt the FIR filter coefficients during double talk, the filter will diverge, causing “blips” in the output. This can be avoided by having an efficient voice activity detector (VAD). When the VAD detects near-end speech the adaptation process is suspended. This avoids the divergence problem

- **Complex echo**: The echo-canceller performs poorly in some cases of double talk. It fails to cancel the far-end speech effectively. We attribute this to the possibility of the existence of complex echo patterns.

- **Residual Error Suppression**: We know that due to the non-linearities of the echo path of the telephone network the maximum suppression possible is limited to about 40dB. So, in cases when the return signal power falls below a threshold based on the reference signal power, it is suggested that we zero the output. This process however creates a choppyness in the background. To make the background more uniform, we decided to make the output equal to a scaled version of the reference signal when the near-end signal is not present.

- **Length of the filter**: Unfortunately the length of the FIR filter has to depend on the maximum delay in echo signal in the data set we are using. If we consider international telephone conversations, the round trip delay is typically an order of magnitude more than that for domestic calls. We would like our system to automatically choose the length of the filter depending on the maximum round-trip delay the user specifies. Also another unanswered question is the relationship between the adaptation rate constant and the filter length. From the experiments we performed, there seems to an inverse relationship between the two quantities. A solution to this problem is still being studied.