Full Combination Multistream System\cite{1-2}:

\begin{itemize}
  \item \(N\) streams result in \(2^N-1\) neural networks
  \item \(N\) streams \rightarrow \(2^N-1\) networks
  \item \(5\) streams \rightarrow \(31\) networks
  \item \(7\) streams \rightarrow \(127\) networks
\end{itemize}

Not suitable for bottleneck feature (BNF) extraction

Different weights results in different transforms

BNFs are crucial for many speech applications, e.g. Speaker ID/Language ID

Proposed architecture:

Replace multiple DNNs in Full combination architecture with single DNN

Training stage:

\begin{itemize}
  \item \(Z_i\) is a binary switch
  \item Block Dropout features of a stream
  \item DNN randomly sees one of the following input patterns:
      \begin{align*}
      &\text{[stream1; 0]} \\
      &\text{[0; stream2]} \\
      &\text{[stream1; stream2]}
      \end{align*}
\end{itemize}

Testing stage:

\begin{itemize}
  \item Use performance monitor\cite{3-4} to select best streams
\end{itemize}

Results in Synthetic band-limited noises:

\begin{itemize}
  \item \(20\) dB: DNN, Mstrm DNN, Mstrm DNN + Perf. Monitor
  \item \(10\) dB: 42.65, 37.49, 35.39
  \item \(0\) dB: 55.45, 46.26, 38.57
\end{itemize}

Results in Additive noises:

\begin{itemize}
  \item \(\text{subway}\): DNN, Mstrm DNN, Mstrm DNN + Perf. Monitor
  \item 71.3, 68.8, 64.1
  \item \(\text{volvo}\): 79.6, 77.8, 73.9
  \item \(\text{factory}\): 78.0, 75.8, 74.8
  \item \(\text{babble}\): 78.2, 76.6, 75.6
\end{itemize}

Results:

\begin{itemize}
  \item IARPA BABEL Year4 Languages:
    \begin{itemize}
      \item Baseline: 60.5, 58.0, 46.7, 43.6, 52.2
      \item Proposed: 59.7, 57.3, 46.1, 43.5, 51.5
    \end{itemize}
  \item Aurora4:
    \begin{itemize}
      \item Baseline: 3.16, 5.09, 5.70, 16.32, 9.83
      \item Proposed: 2.43, 4.54, 3.72, 14.03, 8.42
    \end{itemize}
\end{itemize}

\noindent\textbf{References}

\begin{itemize}
\end{itemize}