\documentclass{beamer}

\usepackage[utf8]{inputenc}

\usepackage{utopia} %font utopia imported

\usepackage{multimedia}

\usepackage{media9}

\usetheme{Madrid}

\usecolortheme{default}

\graphicspath{ {./images/} }

%------------------------------------------------------------

%This block of code defines the information to appear in the

%Title page

\title %optional

{Multimodal Speech Emotion Recognition Using Audio and Text}

% \subtitle{}

\author[Group J] % (optional)

{Group J: Jan Arvin Lapuz, Alyssa Lim, Le Van Nguyen, Ramil Zabala}

\date[Nov 2020] % (optional)

{COMP8240 Applications of Data Science, Nov 2020}

%End of title page configuration block

\begin{document}

\frame{\titlepage}

\begin{frame}

\tableofcontents

\end{frame}

\section{Paper Abstract}

\begin{frame}

\frametitle{Paper Abstract}

\textbf{Paper}:

\begin{itemize}

\item Multimodal Speech Emotion Recognition Using Audio and Text by Seunghyun Yoon, Seokhyun Byun, Kyomin Jung

\item Source code is available on author's Github repository but not pre-trained model

%\item Github link: github.com/david-yoon/multimodal-speech-emotion

\item 2018 IEEE Spoken Language Technology Workshop (SLT)

\item Rank 14 in Computational Linguistics Category on Google Scholar

\end{itemize}

\textbf{Four models}: Text only, Audio only, Multimodal, Multimodal-Attention

\textbf{Processed input data}:

\begin{itemize}

\item Speech Data (.npy file)

\begin{itemize}

\item Mel Frequency Crystal Coefficients (MFCC)

\item Prosodic Features

\end{itemize}

\item Text Data (.npy file): Word Tokens

\item Emotion Categories: Angry, Happy, Sad Neutral

\end{itemize}

\textbf{Set up for Google Colab}:

\begin{itemize}

\item tensorflow==1.4; python==2.7

\item scikit-learn==0.20.0; nltk==3.3

\end{itemize}

\end{frame}

% \section{Replication Results}

% \begin{frame}

% \frametitle{Replication Results}

% % \begin{itemize}

% % \item Results from their original results:

% % \item Results from our replication

% % \item Summary of the previous presentation

% % \end{itemize}

% \begin{center}

% \begin{tabular}{||c c{1.5cm} c{1.5cm} ||}

% \hline

% Model & Replication Accuracy & Research Accuracy \\ [0.5ex]

% % \hspace{5cm} YT \hspace{2cm} TESS \\ [0.5ex]

% \hline\hline

% Text Only & 0.00\% & 63.5\% & 62.8\%| \\

% \hline

% Audio Only & 51.1\% & 54.6\% & 62.8\% | \\

% \hline

% Multimodal & 0.00\% & 71.0\% & 71.8\% | \\

% Multimodal-Attention & 49.9\% & 69.0\% & 62.8\% | \\

% \hline

% \end{tabular}

% \end{center}

% \end{frame}

%---------------------------------------------------------

\begin{frame}

\frametitle{Replication Results on Original Dataset}

% \begin{itemize}

% \item Pre-trained model not available but codes and model parameters are available on GitHub: \href{https://github.com/david-yoon/multimodal-speech-emotion}{https://github.com/david-yoon/multimodal-speech-emotion}

% \item Replication Results: \\

% \begin{center}

\begin{tabular}{||c c{0.5cm} c{0.5cm} c{0.5cm}||}

\hline

Model & Repl Acc TESS & Repl Acc YT & Research Acc \\ [0.2ex]

\hline\hline

Text Only & 0\% & 62.8\% & 63.5\%\\

\hline

Audio Only & 0\% & 55.7\% & 54.6\%\\

\hline

Multimodal & 0\% & 71.0\% & 71.8\%\\

\hline

Multimodal-Attention & 0\% & 48.5\% & 69.0\%\\

\hline

\end{tabular}

% \end{center}

% \begin{center}

% \begin{tabular}{||c c{0.5cm} c{0.5cm} c{0.5cm}||}

% \hline

% Model & Replication Accuracy TESS & Replication Accuracy YT & Research Accuracy \\ [0.2ex]

% \hline\hline

% Text Only & 0\% & 62.8\% & 63.5\%\\

% \hline

% Audio Only & 0\% & 55.7\% & 54.6\%\\

% \hline

% Multimodal & 0\% & 71.0\% & 71.8\%\\

% \hline

% Multimodal-Attention & 0\% & 48.5\% & 69.0\%\\

% \hline

% \end{tabular}

% \end{center}

% \end{itemize}

\end{frame}

%---------------------------------------------------------

\section{New Data Capture}

\begin{frame}

\frametitle{New Data - Youtube Audio and Transcript}

% \textbf{Environment}: Google Colab

% \textbf{Requirements}:

Audio

\begin{itemize}

\item Youtube\\_dl Python package - extract the audio

\item Sox package - mp3 to wav conversion

\end{itemize}

Transcript

\begin{itemize}

\item savesubs.com to save the captions

\item Autogenerated captions

\end{itemize}

\end{frame}

% \section{New Data- Youtube Emotion}

\begin{frame}

\frametitle{New Data - Youtube Emotion}

Qualtrics Emotion Annotation Survey

\includegraphics[width=8cm]{images/Qualtrics\_Survey.png}

\end{frame}

\begin{frame}

\frametitle{New Data - Emotion Distribution}

Emotion Distribution

284 Transcripts Collected

\includegraphics[width=8cm]{images/Youtube Emotion Distribution.png}

\end{frame}

% \section{New Data - }

\begin{frame}

\frametitle{New Data - TESS data}

\begin{itemize}

\item Attach the structure of the test data – Tess data structure mismatch

\item Get description of the dataset

\item Google API to get the transcript

\item Tess data structure mismatch

\item https://tspace.library.utoronto.ca/handle/1807/24487

\end{itemize}

\end{frame}

\section{Movie snippet}

\begin{frame}

\frametitle{Movie snippet}

% \movie[options]{placeholder box}{images/DY008.wav}

% \movie {images/DY008.wav}

\\

% \includemedia[options]{alt content}{images/DY008.wav}

% \\

\includemedia[options]{alt content}{images/DY008.wav}

\\

\end{frame}

\section{New Data Pre-Processing}

\begin{frame}

\frametitle{New Data Preprocessing - MFCC }

\\

MFCC (Audio Features)

\\

\begin{itemize}

\item Extract features through Open Smile

\item Major MFCC Features

\begin{itemize}

\item Volume

\item Energy

\item Pitch

\item Zero Crossing Rate

\item Spectral Centroid

\end{itemize}

\end{itemize}

\\

\includegraphics[width=8cm]{images/MFCC File Layout.png}

\\

\end{frame}

\section{Model Evaluation}

\begin{frame}

\frametitle{New Data Preprocessing - Prosody}

\\

Prosody (Emotion Features)

\\

\begin{itemize}

\item Extract features through Open Smile

\item Major Prosody Features

\begin{itemize}

\item Intonation

\item Stress

\item Rhythm

\item Speech Rate

\item Pauses

\item Voice Quality

\end{itemize}

\end{itemize}

\\

\includegraphics[width=8cm]{images/MFCC File Layout.png}

\end{frame}

% \section{New Data Model Testing}

%---------------------------------------------------------

\begin{frame}

\frametitle{New Data Model Testing}

% This is a text in second frame. For the sake of showing an example.

\begin{itemize}

\item Absence of pre-trained model

\item Retraining on original IEMOCAP data - results close to the original

\item Evaluation and Training on New Data

\item Small modifications to the configuration

\end{itemize}

\end{frame}

\section{Summary and Reflections}

%---------------------------------------------------------

\begin{frame}[fragile]

\frametitle{Summary and Reflections}

% \bigskip

\begin{itemize}

\item Replication results

\item Effect of the Tess dataset variability

\item Weights Parameters in the Open Smile

% suited for IEMOCAP data

% \item The variability of the Tess dataset is not that good because they used the same transcripts for the emotions

\item Effect of music background

\item Controlled environment in the IEMOCAP data

% \item Small Sample size?

\end{itemize}

\end{frame}

\end{document}