Results

Lip-Reading Model Training
- Peak validation accuracy occurred at 21st epoch at 86.2%
- Trained on 12 classes with 1000 training and 50 validation samples per class; 15 samples per batch
- Used Stochastic Gradient Descent as an optimizer with learning rate 0.0003
- Can efficiently and accurately identify English words from video alone

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train Accuracy</td>
<td>99.7%</td>
</tr>
<tr>
<td>Val Accuracy</td>
<td>86.2%</td>
</tr>
<tr>
<td>Val Loss</td>
<td>0.618</td>
</tr>
<tr>
<td>Val Precision</td>
<td>91.1%</td>
</tr>
<tr>
<td>Val Recall</td>
<td>91.3%</td>
</tr>
<tr>
<td>Val F1 Score</td>
<td>91.1%</td>
</tr>
</tbody>
</table>

Table 5.2: Training Results: Epoch 23

Compression Training
- Compression used to send video to server
- For video compression, we chose to use the results found on the 100th epoch
- The accuracy slowly increased to 90 percent and plateaued

Research Challenges
- Head movement and unstable video - Solution: Use facial detection and isolate mouth region in each frame
- Stagnant classifier performance; accuracy in range of 30-50% - Solution: Replace Adam optimizer with SGD; improved to 80+% (using Keras)
- Host a classification server, and provide results for the user - Solution: Use node.js to create a lightweight REST API server.

Acknowledgement
We would like to thank Pingchuan Ma from Imperial College London for assistance with the model architecture, and Rob Cooper from BBC for providing access to news clips for training.

References