Google Cloud Speech-to-Text Machine Learning Program

Group 4 Project Proposal
Matt Davis
Drew Murphy
Soi Kang

1. What is the topic and why have you selected this topic?

Google's Cloud Speech-to-Text Machine Learning Program enables developers to convert speech to text. It utilizes powerful neural network models and an API that recognizes over 120 languages and variants. It is able to compute in real time or prerecorded audio. It can be used in call centers, for live subtitling in TV, automatic transcription, voice biometrics, and many others. Because of the growing list of applications of speech to text, we felt knowledge of its uses, strengths, and shortcomings to be useful in our careers in IT.

2. Outline objective and project plan (what questions to cover in your research or what scope of development work)?
   1. What is Google Cloud Speech-to-Text Machine Learning Program
      a. Supported Languages and variants
      b. Language autodetection
      c. Speaker Diarization/Multichannel Recognition
      d. Pricing
   2. How does it work?
      a. Neural Network Models / API
   3. What are its strengths?
      a. Configure your own recognition requests
   4. What shortcomings does it have?
      a. The Client Libraries are released as Alpha and will likely be changed in backward-incompatible ways.
   5. Current applications it is best suited for.
      a. Sending audio from Android devices to provide streaming speech recognition
   6. Future applications

What is GC Speech-to-Text Machine Learning--15pts
- Supported Languages and variants (3pts)
- Language autodetection (3pts)
- Speaker Diarization/Multichannel Recognition (3pts)
- Pricing (3pts)
- Pre-built models (3pts)

How does it work --5pts
- Neural Network Models (2pts)
- API (3pts)

What are its strengths---6pts
- Configure your own recognition requests (1pt)
- Noise robustness (1pt)
- Global vocabulary (1pt)
- Real-time streaming (1pt)
- Inappropriate content filtering (1pt)
- Phrase hints (1pt)

What are its shortcomings---4pts
- The Client Libraries are released as Alpha and will likely be changed in backward-incompatible ways (1pt)
- Auto-detect language (Beta) (1pt)
- Speaker Diarization (Beta) (1pt)
- Automatic punctuation (Beta) (1pt)

Current applications--10pts
- Streaming speech recognition (2pt)
- Live TV captioning (2pt)
- Call Centers (2pt)
- Google Assistant (2pt)
- Transcription (2pt)

Future Application --5pts
- Voice Authentication (2.5pts)
- Thought-to-Text (2.5pts)