Part A

- **Find where the confidence score is located.**

  When the dialogue system activates ASR and the user speaks, `speechstate` sends a `RECOGNISED` event to the dialogue manager. This event contains a field called `value`, which is an array of recognition results.

  The confidence score is stored inside:

  `event.value[0]?.confidence`

  It is not stored directly in the DM context unless we explicitly assign it.

  ```
  event.value = [
    {
      utterance: "what you said",
      confidence: 0.63,
      ... other fields
    }
  ]
  ```

- **How is its value stored?**

  It is stored as a floating-point number between 0 and 1.

- **How can you use/call that value via TypeScript in your projects?**

  Inside the listening state of the DM:

  ```
  RECOGNISED: {
    actions: assign(({ event }) => {
      console.log("Utterance:", event.value[0].utterance);
      console.log("Confidence:", event.value[0].confidence);
      return { lastResult: event.value };
    }),
  },
  ```

- **ASR Performance Before Custom Speech Training**

  1. **Fictional places, people, or objects**

     Table 1 shows the utterances produced by the recognizer and their corresponding confidence scores.

Spoken Word	Recognized As	Confidence
Rivendell	Rivendell	0.39
Targaryen	Targaryen	0.85
Hogwarts	Hogwarts	0.83
Dumbledore	Dumbledore	0.80
Wakanda	Wakanda	0.18
Asgard	Asgard	0.80
Moria	Moria	0.19
Isildur	Is soldier	0.11
Eärendil / Isildur variant	Easelder	0.07
Númenor	Numenor	0.17
Silmarillion	Simarillion	0.06
Braavos	Bravos	0.06
Khal Drogo	Call Drogo	0.48
Grindelwald	Grindelwald	0.62
McGonagall	Mcgonigle	0.43

Observations

– Fictional names from popular media (e.g., *Hogwarts*, *Dumbledore*, *Targaryen*) were recognized surprisingly well, likely due to their frequent appearance in online text.

– Fantasy names from Tolkien (e.g., *Rivendell*, *Númenor*, *Silmarillion*) performed poorly, with confidence scores between 0.06 and 0.39.

– Scientific and Latin names such as *Magnolia lilliflora* were recognized with moderate accuracy but still low confidence.

– Foreign place names (e.g., *Reykjavik*) showed inconsistent performance, reflecting difficulty with non-English phonology.

– Accent influence was noticeable: several words were transcribed into phonetically similar English phrases (e.g., "Isildur" → "Is soldier").

2. **Scientific names or difficult real names**

Spoken Word	Recognized As	Confidence
Rosa canina	Rosa Canina	0.50
Archaeopteryx lithographica	Archaeopteryx Lithographica	0.76
Magnolia lilliflora	Magnolia Lillifora	0.46
Xiangling ("Shawn Ling")	Shawn Ling	0.16
Yeganeh (variant)	Yagana	0.07
Wojciech	Wojciech	0.07
Erdoğan	Erdogan	0.07
Xi Jinping	Xi Jinping	0.66
Mahler	Mahler	0.06
Shostakovich	Shostakovich	0.79
Rimsky-Korsakov	Rimsky, Korsakov	0.39
Rachmaninoff	Rachmaninoff	0.06

Observations

These results further highlight the limitations of the baseline ASR model when handling rare, foreign, or phonetically complex names. Scientific Latin terms (e.g., *Archaeopteryx lithographica*) were recognized with relatively high confidence, likely because their pronunciation is more regular and appears in specialized corpora. In contrast, personal names with Slavic, Persian, or Turkish

phonology (e.g., *Wojciech*, *Yeganeh*, *Erdoğan*) received extremely low confidence scores, reflecting the model's difficulty with non-English phonotactics.

Composer names showed large variation: *Shostakovich* was recognized well, while *Mahler* and *Rachmaninoff* were assigned very low confidence. This inconsistency suggests that the model's familiarity with proper nouns is strongly tied to their frequency in the training data. The misrecognition of "Xiangling" as "Shawn Ling" also illustrates how the ASR attempts to map unfamiliar sounds onto the closest English-like pattern.

3. **Do you think any specific accent you are using makes words difficult to process?**

Some words were recognized correctly but with low confidence, which may be influenced by my accent. Certain phonemes (like "r", "ll", or non-English vowels) may be pronounced differently, causing the ASR system to guess less confidently.

- **Think about how this problem (transcription of something we did not intend to say) could be solved. Why do you think recognition falters in the examples that you tried?**

Recognition likely falters because the baseline ASR model is trained on large general-purpose datasets. Rare fantasy names, scientific terms, and non-English words appear less frequently in training corpora. Accent variation further increases phonetic ambiguity.

This problem could be improved by:

- Using Custom Speech training with domain-specific vocabulary.
- Adding pronunciation dictionaries.
- Including user-specific adaptation data.
- Using confirmation strategies in dialogue systems.

Part VG A

- **ASR Performance After Custom Speech Training**

Endpoint ID:

9d99579c-a534-43ac-96d1-0d179cfb6a61

1. **Fictional places, people, or objects**

Spoken Word	Recognized As	Confidence
McGonagall	Mcgonagall	0.94
Grindelwald	Grindelwald	0.95
Khal Drogo	Khal Drogo	0.61
Braavos	Bravos	0.38
Silmarillion	Silmarillion	0.97
Númenor	Numenor	0.89
Eärendil	Earendel	0.17
Isildur	Is solder	0.50
Moria	Moria	0.96
Asgard	Asgard	0.96
Wakanda	Wakanda	0.89
Dumbledore	Dumbledore	0.95
Hogwarts	Hogwarts	0.95
Targaryen	Targaryen	0.83
Rivendell	Rivendell	0.52

Observations:

The Custom Speech model shows a clear improvement across most of the tested words. Many previously low-confidence items now achieve confidence scores above 0.90, including *McGonagall*, *Grindelwald*, *Silmarillion*, *Moria*, *Asgard*, *Dumbledore*, and *Hogwarts*. These gains indicate that the language model successfully incorporated the domain-specific vocabulary provided during training.

Some names that were previously misrecognized (e.g., "Simarillion", "Mcgonigle") are now transcribed correctly with high confidence. Even words that remain imperfect, such as *Rivendell* or *Khal Drogo*, show substantial improvement compared to the baseline model.

However, a few items still perform poorly. Names like *Eärendil* and *Isildur* continue to be misrecognized ("Earendel", "Is solder"), suggesting that the model still struggles with rare phonological patterns or insufficient training examples. These cases highlight that Custom Speech training improves recognition most effectively when the target words appear clearly and consistently in the training data.

2. **Scientific names or difficult real names**

Spoken Word	Recognized As	Confidence
Rachmaninoff	Rachmaninoff	0.95
Rimsky-Korsakov	Rimsky, Korsakov	0.96
Shostakovich	Shostakovich	0.94
Mahler	Mahler	0.84
Xi Jinping	She Jinping	0.83
Erdoğan	Erdogan	0.29
Wojciech	Wycheck	0.27
Yeganeh	Yeganeh	0.40
Xiangling ("Sean Ling")	Sean Ling	0.55
Magnolia lilliflora	Magnolia lilliflora	0.97
Archaeopteryx lithographica	Archaeopteryx Lithographica	0.96
Rosa canina	Rosa Canina	0.73

Observations

The Custom Speech model shows substantial improvement for most scientific and real-world names. Classical composer names such as *Rachmaninoff*, *Rimsky-Korsakov*, and *Shostakovich* now achieve confidence scores above 0.94, compared to baseline values around 0.06–0.41. Scientific Latin terms (*Magnolia lilliflora*, *Archaeopteryx lithographica*, and *Rosa canina*) also show strong gains, with confidence scores between 0.73 and 0.97.

However, some names remain challenging. Non-English personal names such as *Erdoğan*, *Wojciech*, and *Yeganeh* still receive relatively low confidence scores, despite being included in the training data. This suggests that the model struggles with phonological patterns that are rare in English or that differ significantly from the orthographic form.