**Fundamentals of Azure AI Speech**

**Speech Recognition och Speech Synthesis**

**Speech recognition** omvandlar tal till text genom att analysera ljudmönster och identifiera ord. Detta görs genom två huvudsakliga modeller:

* **Acoustic model** – Konverterar ljudsignalen till **phonemes** (ljudrepresentationer).
* **Language model** – Kartlägger phonemes till ord med en **statistisk algoritm**.

Taligenkänning används exempelvis för:

* **Closed captions** i videor
* **Transkription** av möten
* **Automatiserad diktering**

**Speech synthesis** gör det motsatta, dvs. omvandlar text till tal. Det kräver:

* **Text input**
* **Voice selection** (röst, tonläge, hastighet, volym)

**Speech synthesis** används för:

* **Röststyrda menyer** i telefonsystem
* **Uppläsning av e-post eller meddelanden**
* **Publika utrop** på stationer eller flygplatser

**Azure AI Speech**

**Azure AI Speech** erbjuder **speech to text** och **text to speech** via **API**:er och stödjer realtids- eller batchprocessering.

**Speech to text**

* **Real-time transcription** – Direkt omvandling av tal till text från en mikrofon eller ljudström.
* **Batch transcription** – Omvandling av lagrade ljudfiler via **SAS URI** i Azure.

Microsofts **Universal Language Model** används som standard, men du kan skapa **custom models** för **acoustics, language, pronunciation**.

**Text to speech**

* Konverterar text till tal och kan spelas upp direkt eller sparas som en ljudfil.
* **Neural voices** ger mer naturligt ljud genom avancerad intonering.
* Möjlighet att skapa **custom voices**.

**Azure-resurser för AI Speech**

För att använda **Azure AI Speech** i en applikation krävs en resurs:

* **Speech resource** – För enbart **Azure AI Speech**.
* **Azure AI Services** – För Speech i kombination med andra AI-tjänster.

Tjänsten kan användas via **Speech Studio, CLI, REST API** eller **SDK** i flera programmeringsspråk.