ERASMUS+ ARTIE
ARTIFICIAL INTELLIGENCE IN EDUCATION

Co-funded by the
Erasmus+ Programme
of the European Union

ARTIE: Artificial Intelligence in Education - challenges and opportunities of the new era:
development of a new curriculum, guide for educators and online course for students
Project co-funded by European Union under Erasmus+ Programme, 2020-1-HR01-KA201-077800

TITLE: Speech recognition and generation for beginners in Scratch

LEARNING SCENARIO		
School:	*Duration (minutes):*	90
Teacher:	*Students' age:*	13-14

Essential Question:	What is Speech recognition and generation?

Topics:

- Speech recognition and generation for beginners in Scratch

Aims:

- To learn about Speech recognition and generation for beginners in Scratch

Outcomes:

- Understanding Speech recognition and generation with the help of a simple example in Scratch.

Work forms:
- *work in pairs, group work*

Methods:
- *presentation, talk, discussion, interactive exercise*

ARTICULATION
Course of action (duration in minutes)
INTRODUCTION

Ask students to try to define speech recognition and generation.
You can ask them to explain the difference between speech recognition and speech generation.
Ask students if they know any speech recognition systems. (Answer: Google NOW, SIRI, …)

Introduction to Speech Recognition

ARTIE: Artificial Intelligence in Education - challenges and opportunities of the new era:
development of a new curriculum, guide for educators and online course for students
Project co-funded by European Union under Erasmus+ Programme, 2020-1-HR01-KA201-077800

Speech Recognition is the ability to translate a dictation or spoken word to text. It is also known as Speech-to-Text (STT) and Voice Recognition.

It is achieved by following certain steps and the software responsible for it is known as a „Speech Recognition System". SR systems are usually implemented in the form of dictation software and intelligent assistants in personal computers, smartphones, web browsers and many other devices.

Introduction to Speech Generation

Speech generation or synthesis (also abbreviated as TTS, Text-to-Speech), unlike speech recognition, is not a technology that exploits the voice, it produces it. Synthetic voices are generally the final phase of the process and are becoming increasingly democratic because they are important in the overall experience of "voice".

Speech synthesis (TTS) is defined as the artificial production of human voices. The main use (and what induced its creation) is the ability to translate a text into spoken speech automatically.
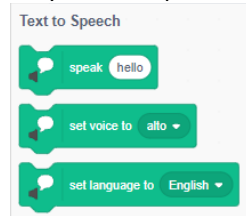
Defining the goal of the lesson:
Introduction to speech recognition and generation for beginners using an example of one simple program in Makeblock.
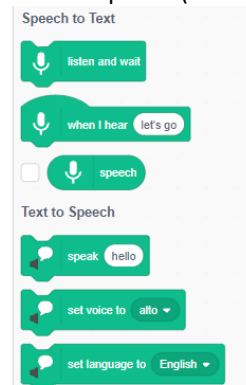
MAIN PART

Applications
Scratch (MIT) - https://scratch.mit.edu/projects/editor/
Only Text-to-Speech extension is available (3 blocks)

Scratch (ML4KIDS) - https://machinelearningforkids.co.uk/scratch3/
Text-to-Speech (3 blocks) and Speech-to-Text (3 blocks) extensions are available

Makeblock - https://ide.mblock.cc/
Text-to-Speech (3 blocks), Cognitive services (3 blocks related to Speec-to-Text)

2

ERASMUS+ ARTIE
ARTIFICIAL INTELLIGENCE IN EDUCATION

Co-funded by the
Erasmus+ Programme
of the European Union

ARTIE: Artificial Intelligence in Education - challenges and opportunities of the new era:
development of a new curriculum, guide for educators and online course for students
Project co-funded by European Union under Erasmus+ Programme, 2020-1-HR01-KA201-077800

speak (hello)

recognize speech in Mandarin(simplified) ▼ for 2 ▼ secs

set voice to alto ▼

set speech recognition result to hide ▼ punctuation marks

set language to Arabic ▼

text recognition result

* BONUS
Makeblock also includes blocks for language recognition and text sentiment analysis

analyze English ▼ text sentiment Help more people enjoy creating while having fun.

sentiment analysis result

recognize language: Cognitive Services

language recognition result

Sentiment analysis is the process of detecting positive or negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers. Sentiment analysis models focus on polarity (*positive, negative, neutral*) but also on feelings and emotions (*angry, happy, sad*, etc), urgency (*urgent, not urgent*) and even intentions (*interested v. not interested*). Learn more about Sentiment analysis: https://monkeylearn.com/sentiment-analysis/

To illustrate this, let's take an example with Makeblock.

Step 1: Open Makeblock page: https://ide.mblock.cc/
Step 2: Add extensions: Cognitive services and Text to Speech
Step 3: Check to display the following reporter type blocks:

✓ speech recognition result

✓ language recognition result

✓ sentiment analysis result

Step 4: Use speech recognition result in recognize language and analyse text sentiment blocks

Step 5: Use this sequence of blocks

3

ERASMUS+ ARTIE
ARTIFICIAL INTELLIGENCE IN EDUCATION

Co-funded by the
Erasmus+ Programme
of the European Union

ARTIE: Artificial Intelligence in Education - challenges and opportunities of the new era:
development of a new curriculum, guide for educators and online course for students
Project co-funded by European Union under Erasmus+ Programme, 2020-1-HR01-KA201-077800

Step 6: Turn on your microphone and speakers, start the code and say "I am happy"
(you will see the pop-up window recording your voice for 2 seconds - if you need more, increase the value in recognize speech in English for 3 secs (or more). You will hear "This is good" or "This is not good" depending on sentiment analysis results   .

Step 7: Start the code again, say something else and wait for sentiment analysis results   .

CONCLUSION

Basic Principles of Speech Recognition
The smallest unit of spoken language is known as a Phoneme. The English language contains approximately 44 phonemes representing all the vowels and consonants that we use for speech. We can take the example of a typical word such as moon which can be broken down into three phonemes: m, ue, n.

4

ARTIE: Artificial Intelligence in Education - challenges and opportunities of the new era:
development of a new curriculum, guide for educators and online course for students
Project co-funded by European Union under Erasmus+ Programme, 2020-1-HR01-KA201-077800

To interpret speech, we must have a way of identifying the components of spoken words and phonemes that act as identifying markers within a speech. An algorithm must be used to interpret the speech further. The Hidden Markov Model is a commonly used mathematical model used to do this. To create a speech recognition engine, a large database of models is created to match each phoneme.

Learn more: *https://www.ibm.com/cloud/learn/speech-recognition*

Basic Principles of Speech Generation

Unlike speech recognition systems that use phonemes (the smallest units of sound) in the first place to cut out sentences, TTS will be based on what are known as graphemes: the letters and groups of letters that transcribe a phoneme. This means that the basic resource is not the sound, but the text. This is usually done in two steps.

The first will cut the text into sentences and words (our famous graphemes) and assign phonetic transcriptions, the pronunciation, to all these groups. Once the different text/phonetic groups have been identified, the second step consists of converting these linguistic representations into sound. In other words, to read these indications to produce a voice that will read the information.

Try TTS online: *https://www.readspeaker.com/*

Speech recognition is the ability to translate a dictation or spoken word to text. It is also known as Speech-to-Text and Voice recognition. It is achieved by following certain steps and the software responsible for it is known as a "Speech Recognition System". Speech recognition systems are usually implemented in the form of dictation software and intelligent assistants in personal computers, smartphones, web browsers and many other devices.

Do the K.W.L. (**K**now, **W**ant, **L**earned) chart with your students .

What I **K**now	What I **W**ant to Know	What I **L**earned

Methods	Work forms
presentation	*work in pairs*
interactive exercise / simulation on the computer	*group work*

Material:

- https://scratch.mit.edu/projects/editor/
- https://machinelearningforkids.co.uk/scratch3/
- https://ide.mblock.cc/
- https://monkeylearn.com/sentiment-analysis/

5

ARTIE: Artificial Intelligence in Education - challenges and opportunities of the new era:
development of a new curriculum, guide for educators and online course for students
Project co-funded by European Union under Erasmus+ Programme, 2020-1-HR01-KA201-077800

- https://www.ibm.com/cloud/learn/speech-recognition
- https://www.readspeaker.com/

Literature

-

PERSONAL OBSERVATIONS, COMMENTS AND NOTES

6