_____

TEXT-TO-SPEECH IN HUMAN-ROBOT COMMUNICATION

Elena Mogoş

Department of Computer and Information Technology, "Dunarea de Jos" University of Galati, Romania
(email: Elena.Mogos@ugal.ro)

Abstract: Human-robot communication is an important field of robotics. In order to facilitate the communication and interactions between humans and robots, in this work we propose a Text-to-Speech system, which allows robots to speak naturally and as close as possible to the human voice. The robot's decision system must eliminate the ambiguities and archive the data obtained in the human-robot communication. The proposed solution involves a voice synthesizer, which generates voice messages starting from text messages, adding environmental sounds for a more natural impression.

Keywords: Text-to-Speech, voice synthesizer, Matlab

1. INTRODUCTION

Text-to-Speech, known as TTS is a form of speech synthesis that converts text into spoken voice output. The first Text-to-Speech systems were designed to help the visually impaired people communicate with a computer. Nowadays TTS systems (software) are different from the voice response systems, which use a database with prerecorded words and a language's graphemes and phonemes to synthesize the human speech in the communication between a robot and a human (https://www.webopedia.com).

In this paper, we present some of the existing Text-to-Speech systems and, starting from these solutions, we created our own TTS system. This involves a Voice Synthesizer which generates voice output, and combines the sounds from the environment of the robots in order to produce a natural impression for the listeners.

2. TEXT-TO-SPEECH SYSTEMS IN THE COMMUNICATION BETWEEN A HUMAN AND A ROBOT

A decade ago, Text-to-Speech software (Fig.1) was rather expensive, but these days there are excellent Text-to-Speech tools available free of charge.
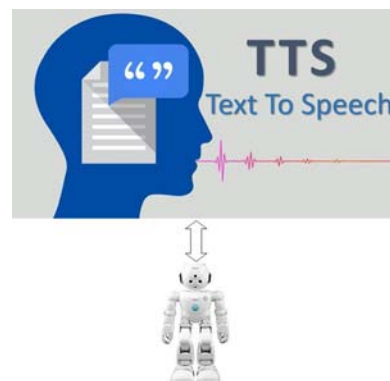

Fig.1. TTS software (https://blog.usejournal.com)

_____

2.1 Google Text-to-Speech software

Google Text-to-Speech is a software tool developed by Google for its Android Operating System. This Text-to-Speech software applies groundbreaking research in speech synthesis and uses a powerful neural network to delivery high fidelity audio (Fig.2).

The new version of Google TTS is Google Cloud Text-to-Speech. The Google Cloud Text-to-Speech converts any text in more than 100 voices across 20+ languages and variants. The speech synthesis used by Google Cloud TTS is based on WaveNet. Wavenet is a deep neural network for generating raw audio created by DeepMind company, which introduces mimics in robot speech and reduces the gap between this tool and human performance with 50%.

Google Cloud TTS easily integrates all applications and devices from phones and tablets to humanoid robots and it is an alternative way for a robot naturally speech (https://cloud.google.com).
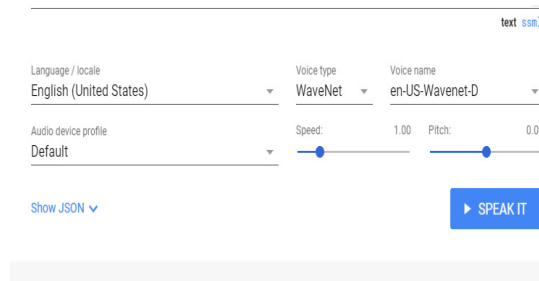

Fig.2. Google Cloud Text-to-Speech software (https://cloud.google.com)

2.2. IBM Watson Text-to-Speech software

IBM Watson Text-to-Speech uses IBM speech synthesis to convert written text into natural sounding audio in different languages and voices (https://cloud.ibm.com).

For each language IBM Watson TTS supports the voice of female or male and both. This TTS software offers a minimal delay in the communication with the humans, a WebSocket interface and a customization interface. A WebSocket interface can make possible the communication between a human and a robot in both ways without polling the robot's server. The customization interface allows defining human's own sounds and human's own sounds translation in the standard International Phonetic Alphabet representation (IPA) or in the proprietary IBM Symbolic Phonetic representation (SPR) (Fig.3).

IBM Watson Text-to-Speech software can:

-make all the devices to "speak";

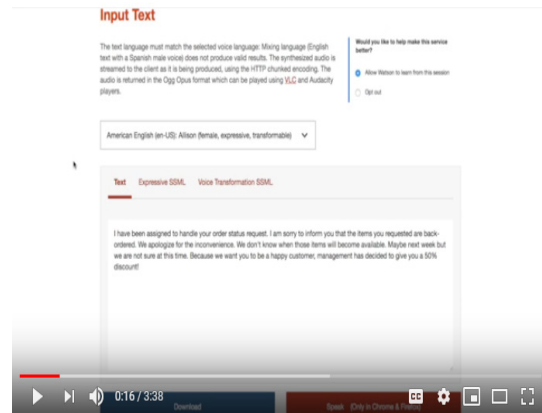-customize the pronunciation in robot's speech;

-go across languages and voices;

Fig.3. IBM Watson TTS on Youtube (https://www.youtube.com)

2.3 Microsoft Azure Text-to-Speech Software

Microsoft Azure Text-to-Speech software allows robots to convert text to speech in natural responsiveness as humans and allows humans to choose from 75+ standard voices in 45 languages and 5 neural voices in 5 languages for the voice of the robots or to personalize by themselves the voice of the robots.

The standard voices used by Microsoft Azure Text-to-Speech software are created with different techniques to sound naturally and to be understood by humans. The neural voices used by Microsoft Azure Text-to-Speech software are based on deep neural networks which introduce the patterns of stress and intonation in the voice of the robots. The custom voices used by Microsoft Azure TTS are adjusted voices by humans in the communication between the humans and robots. The human can match the voices of the robots with SSML (Speech Synthesis Markup Language), an XML-based Markup language (https://docs.microsoft.com).
Microsoft Azure TTS can help humans with impaired hearing and to proofread texts on the go (https://stackify.com).

2.4 Other Speech-to-Text software

Amazon Polly
Amazon Polly turns the text into lifelike speech and offers an intelligent text to Polly, a humanoid robot.

The benefits of Amazon Polly (https://aws.amazon.com) are:

-natural sounding voices;

-store and redistribute speech;

-real time streaming;

-customize and control speech output;

-low costs;

Voice Reader Home 15

Voice Reader home 15 is another commercial TTS software for private users. It converts any text into audio files and sends them directly to a PC or to a smart phone (https://www.linguatec.de).

Voice Reader Home 15 is:

- easy;
- quick;
- convenient;

Capti Voice

Capti Voice is a TTS software for the education sector. It is very efficient for students who can read more effectively their subjects and can be used by students after their graduation and at their job (https://www.captivoice.com).

Natural Reader

Natural Reader is a TTS software based on cloud computing. The cloud computing is the delivery of different services through the Internet. Natural Reader offers high quality to human-robot communication, reads from scanned photos or documents and converts past uploads in mp3 files (https://www.naturalreaders.com).

Voice Dream Reader

Voice Dream Reader is an accessible TTS software for the mobile devices and used by anyone who wants any text to be read out loud (https://www.voicedream.com).

3. TEXT-TO-SPEECH IN THE HUMAN-ROBOT COMMUNICATION

The TTS software from our work uses an assembly language and the HSV to convert the human's written text into speech for our humanoid robot. The speech synthesizer of our humanoid robot is natural and intelligible for the humans. For these we create a program in Matlab code which is presented below:

Matlab code

```
prompt="???";
```

```
name='Text to Speech';
numlines=1;
defaultanswer={'20','hsv'};
answer=inputdlg(prompt,name,numli
nes,(defaultanswer));
if isempty(answer)
  return;
end
answer = char(answer);
NET.addAssembly('System.Speech');
obj=
System.Speech.Synthesis.SpeechSyn
thesizer;
obj.Volume =100;
Speak(obj,answer);
```

The above Matlab codepresents a window, whereinthe user can enter any text. Our humanoid robot can tell us the written text loud and clear in our native language, Romanian. In despite of this success our humanoid robot cannot play the audio file with written numbers in Romanian, only in English (Fig.4).

Beside of our TTS software we use in our work two another software and two hardware:

1. Windows 10, Matlab R2016a and I3 Intel Core Processor;

2. Windows 7, Matlab R2018b and AMD E-350 Processor 1.6 Hz.

In both cases we obtain the same results. Our humanoid robot can speak loud and clear in Romanian any written text.
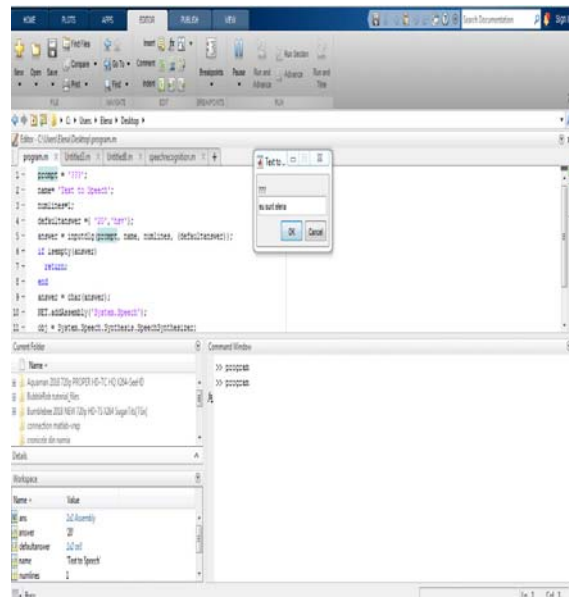
Fig.4. TTS Matlab code

4. CONCLUSIONS

The TTS software are assistive technologies which can be used in many ways:

1.E-learning and training

TTS software provides an easy way and less costly way to roll out across multiple languages.

2. Customer Service

TTS improve the quality of human-robot communication.

3. Media and Entertainment

TTS software increases the efficiency of human-robot communication.

5. REFERENCES

https://www.webopedia.com;
https://blog.usejournal.com;
https://cloud.google.com;
https://cloud.ibm.com;
https://dzone.come;
https://www.youtube.com;
https://stackify.com;
https://docs.microsoft.com;
https://aws.amazon.com;
https://www.captivoice.com;
https://www.naturalreaders.com;