

MMDAvatar– An Online Voice Chat Robot with 3D Avatar and Artificial Intelligence

Jianming Liu and Steven L. Grant

Electrical and Computer Engineering Department of Missouri University of Science and Technology

Abstract

This paper proposes a web-based online voice chat robot, which supports speech recognition, speech synthesis, artificial intelligence and motion response with a 3D avatar. There are a lot of online chat robots which are based on Artificial Intelligence Markup Language (AIML), however, they usually only support text chat mode. Recently, a voice toolkit called MMDAgent is developed by Nagoya Institute of Technology, Japan, which supports speech recognition, synthesis and a 3D agent. However, the brain of MMDAgent is based on script instead of AIML and it could only support Japanese and run on local machine, which is not very convenient and intelligent. Therefore, we developed an online version voice chat robot called MMDAvatar, which also supports speech recognition, speech synthesis, and motion response with a 3D avatar. Furthermore, it is based on AIML and open to training online. Currently, for the speech recognition, we use the HTML5 speech recognition tag. The Festival speech synthesis engine is used to synthesize the speech. An AIML engine (program O) is used as the brain of the chat robot and the 3D avatar is drawn by WebGL. Compared to MMDAgent, with an AIML engine, our chat robot is more flexible. Meanwhile, since it is online and open to training, it will become more and more intelligent with the time.

Review of Chat Robot

There has been a long history of text-mode chat robot, especially the Artificial Intelligence Markup Language (AIML) based chat robot [1]. Siri of Apple has gained great success and has become more and more popular with the help of speech recognition/synthesis and artificial intelligence [2]. However, there is no avatar in Siri, and it can not support motion response. Recently, Nagoya Institute of Technology in Japan developed a voice toolkit called MMDAgent which supports speech recognition/synthesis and a 3D agent [3]. The disadvantage of MMDAgent is that, the brain of MMDAgent is based on scripts instead of artificial intelligence. Furthermore, all the speech engines (Julius for recognition and HTS for synthesis) are run on local machines, which consume a lot of computation efforts.

Structure of MMDAvatar

In order to improve the intelligence and reduce the computation load of MMDAgent, we developed an online version voice chat robot called MMDAvatar, which also supports speech recognition, speech synthesis, and motion response with a 3D avatar. Currently, for the speech recognition, we use the HTML5 speech recognition tag, which is supported by Google Chrome browser. The Festival speech synthesis engine is used to synthesize the speech on the server [4]. Similar to ALICE, we use an AIML engine named program O as the brain of the chat robot [5], and it is open to training online. For the 3D avatar, we use the WebGL based MikuMikuDance model engine [6] and extend the AIML engine to support motion response besides to speech. Compared to MMDAgent, with an AIML engine, our chat robot is more intelligent and flexible. Meanwhile, considering that both the speech recognition and synthesis are done by server, this will significantly simplify the requirements for the client. You could use your browser to access our chat robot online [7]. It should be noted that Google Chrome browser is recommended and you should enable its WebGL support.

Technical Accomplishments and Design Lessons Learned

In this design, we used the cutting-edge HTML5 speech recognition tag and WebGL3D technology. Meanwhile, we studied the speech synthesis solution and configured an online speech synthesis server based on Festival. Finally, we extend the traditional AIML engine to support motion response besides to speech. We also have the following design lessons learned, for example, the AIML engine needs a lot of efforts to train the robot, especially the motion response data set. Meanwhile, the HTML5 speech recognition tag and WebGL are still not well supported by many browsers.

Bibliography

1. ALICE A.I. Foundation, "Alicebot," (2012). Available WWW: <http://www.alicebot.org/>
2. Apple Inc., "Apple – iPhone 4S – Ask Siri to help you get things done," (2012). Available WWW: <http://www.apple.com/iphone/features/siri.html>
3. MMDAgent project team, "MMDAgent – Toolkit for building voice interaction systems," (2012). Available WWW: <http://www.mmdagent.jp/>
4. The University of Edinburgh, "The Festival Speech Synthesis System," (2012). Available WWW: <http://www.cstr.ed.ac.uk/projects/festival/>
5. Elizabeth Perreau, "Program O AIML Chatbot," (2012). Available WWW: <http://blog.program-o.com/>
6. MMD.js, "MMD on WebGL," (2012). Available WWW: <http://edv.sakura.ne.jp/mmd/>
7. Jianming Liu and Steven L. Grant, "MMDAvatar," (2012). Available WWW: <http://www.mmdavatar.com/>

Biographical Information

Jianming Liu is a PhD candidate of Missouri S&T, and he received his BS from Shandong University, China in 2005 and MS degree from Tsinghua University, China in 2009. He was an audio design engineer of Nokia in Beijing, China, and currently his research in Missouri S&T mainly focuses on speech signal processing, especially adaptive signal processing and doubletalk in echo cancellation.

Steven L. Grant is currently the Wilkens Missouri Telecommunications Professor at Missouri S&T, and he received his B.S.E.E from Missouri S&T in 1979, M.S.E.E from Caltech in 1981, and Ph.D. from Rutgers in 1994. In 1980, he was with Bell Laboratories. He was with International Telephone and Telegraph Corporation--Defense Communications Division from 1982 to 1984 and then returned to Bell Laboratories, where, in 2001, he became Manager of the Acoustics Research group. From 2002 to 2004, he was with the Massachusetts Institute of Technology Lincoln Laboratory, Lexington.

Abstract

❖ This poster proposes a web-based online voice chat robot, which supports speech recognition, speech synthesis, artificial intelligence and motion response with a 3D avatar. There are a lot of online chat robots which are based on Artificial Intelligence Markup Language (AIML), however, they usually only support text chat mode. Recently, there is a voice toolkit called MMDAgent developed by Nagoya Institute of Technology, Japan, which supports speech recognition, synthesis and a 3D agent. However, the brain of MMDAgent is based on script instead of AIML and it could only support Japanese and run on local machine, which is not very convenient and intelligent. Therefore, we developed an online version voice chat robot called MMDAvatar, which also supports speech recognition, speech synthesis, and motion response with a 3D avatar. Furthermore, it is based on AIML and open to training online. Currently, for the speech recognition, we use the HTML5 speech recognition tag. The Festival speech synthesis engine is used to synthesize the speech. An AIML engine (program O) is used as the brain of the chat robot and the 3D avatar is drawn by WebGL. Compared to MMDAgent, with an AIML engine, our chat robot is more flexible. Meanwhile, since it is online and open to training, it will become more and more intelligent with the time.

Review of Chat Robot

❖ There has been a long history of text-mode chat robot, especially the Artificial Intelligence Markup Language (AIML) based chat robot [1]. Recently, there is a voice toolkit called MMDAgent developed by Nagoya Institute of Technology, Japan, which supports speech recognition/synthesis and a 3D agent [2]. The disadvantage of MMDAgent is that, the brain of MMDAgent is based on scripts instead of artificial intelligence. Furthermore, all the speech engines are run on local machine, which consumes a lot of computation efforts.

Proposed MMDAvatar

- ❖ What's MMDAvatar:
 - ✓ MMDAvatar is an online version voice chat robot, which also supports speech recognition, speech synthesis, and motion response with a 3D avatar.
- ❖ Why MMDAvatar rather than MMDAgent?
 - ✓ In order to improve the intelligence.
 - ✓ Reduce the computation requirement of client.
- ❖ Target of MMDAvatar:
 - ✓ Thin client: web browser with WebGL and HTML5 support.
 - ✓ Intelligent and flexible: online and open to training.

Structure of MMDAvatar

- ❖ Main Components of MMDAvatar:
 - ✓ Speech recognition: HTML5 speech recognition tag, which is supported by Google Chrome browser.
 - ✓ Speech synthesis: Sever based Speech Synthesis with Festival [3].
 - ✓ Artificial Intelligence: AIML based Program O Engine [4].
 - ✓ 3D Avatar: WebGL based MikuMikuDance Model Engine [5].

Future work of MMDAvatar

- ❖ Possible future work of MMDAvatar:
 - ✓ Motion recognition input: Kinect based gesture recognition.
 - ✓ Siri-like personal assistant: SiriServerCore instead of simple AIML engine

Try MMDAvatar at <http://www.mmdavatar.com> Today!

Technical Accomplishments

- ❖ The main technical accomplishments are:
 - ✓ We used the cutting-edge HTML5 speech recognition tag and WebGL 3D technology.
 - ✓ We studied the speech synthesis solution and configured an online speech synthesis server based on Festival.
 - ✓ We extend the traditional AIML engine to support motion response besides to speech.

Design Lessons Learned

- ❖ There are also a lot of design lessons learned:
 - ✓ For the AIML engine, it needs a lot of efforts to train the robot, especially the motion response data set.
 - ✓ Meanwhile, the HTML5 speech recognition tag and WebGL are still not well supported by many browsers.

Bibliography

1. ALICE A.I. Foundation, "Alicebot," (2012). Available WWW: <http://www.alicebot.org/>.
2. MMDAgent project team, "MMDAgent – Toolkit for building voice interaction systems," (2012). Available WWW: <http://www.mmdagent.jp/>.
3. The University of Edinburgh, "The Festival Speech Synthesis System," (2012). Available WWW: <http://www.cstr.ed.ac.uk/projects/festival/>.
4. Elizabeth Perreau, "Program O AIML Chatbot," (2012). Available WWW: <http://blog.program-o.com/>.
5. MMD.js, "MMD on WebGL," (2012). Available WWW: <http://edv.sakura.ne.jp/mmd/>.