*Future Technologies Conference (FTC) 2017*
*29-30 November 2017 | Vancouver, Canada*

# A Prototyping of BoBi Secretary Robot

Jiansheng Liu
Shanghai NewReal Auto-System Co., Ltd, NewReal
Shanghai, China
johnsonliu@newreal.net

Bilan Zhu
Department of Computer and Information Sciences
Tokyo University Agriculture and Technology, TUAT
Tokyo, Japan
zhubilan@cc.tuat.ac.jp

***Abstract*—We describe here a prototyping of intelligent personal robot named BoBi secretary. When it is closed, BoBi is a rectangular box with a smart phone size. Owner can call to BoBi to open to transform from the box to a movable robot, and then it will perform many functions like humans such as moving, talking, emoting, singing, dancing, conversing with people to make people happy, enhance people's lives, facilitate relationships, have fun with people, connect people with the outside world and assist and support people as an intelligent personal assistant. We consider BoBi is a treasure and so call the box moonlight box that is "月光宝盒" in Chinese. BoBi speaks with people, tells jokes, sings and dances for people, understands the owner and recognizes people's voices. It can do all works which a secretary is doing including scheduling of works, schedule reminders, sending emails, calling phones, booking, making reservations, searching information, etc. BoBi has three main functions: intelligent meeting recording, multilingual interpretation and reading papers. BoBi is a portable, transformable, movable and intelligent robot.**

***Keywords—Intelligent robot system; personal assistant robot; portable robot; transformable robot; movable robot***

## I. INTRODUCTION

Due to the development of artificial intelligence, machine learning, big data processing and pattern recognition technologies, intelligent robots are receiving more attention than ever before. Industrial robots have been applied to various industries, work instead of human, and have largely improved production efficiency [1]. Compared to industrial robots, intelligent service robots are still under development because they demand human-level intelligence resulting in great technique difficult.

Human-machine interaction is a key technology of service robots. In the service robot systems it is necessary to provide an intelligent voice interaction function that is able to assist and support people as an intelligent personal assistant. Siri by Apple [2]-[3] and Cortana by Microsoft [4]-[5] are such intelligent personal assistants. However, they are only smart phones and cannot move to make people feel touched like humans.

Toy robots Robi [6] and Plami [7] can walk and move like a person and they dance, sing and their quick, light and agile actions are gathering attention. However, they can only have a simple conversation and are far from intelligent robots. Plen robot [8] can do many agile actions but it is an unintelligent robot.

Humanoid robots Pepper developed by Aldebaran Robotics and SoftBank [9]-[10] and Nao developed by Aldebaran Robotics [11]-[13] have been designed to talk and make communication with people where their communication abilities are very simple compared to humans.

In this paper, we propose an intelligent assistant robot: BoBi secretary as shown in Fig. 1. When it is closed, BoBi is a rectangular box with a smart phone size. We consider it as a moonlight box. It automatically opens to transform from the moonlight box to a robot when its owner calls to it, "open," and then it performs functions like humans such as moving, talking, emoting, singing, dancing, and conversing with people to make people happy, enhance people's lives, facilitate relationships, have fun with people and connect people with the outside world. BoBi assists and supports people as an intelligent personal assistant. BoBi understands the owner and recognizes people's voices, and does all secretary works including scheduling of works, schedule reminders, sending emails, calling phones, booking, making reservations, searching information, etc. In the secretary works there are three main functions: intelligent meeting recording, multilingual interpretation and reading papers. BoBi is a portable, transformable, movable and intelligent robot. There are wheels under BoBi's feet so it can moves by them. We control BoBi mainly by voice interaction as well as by touch screen.

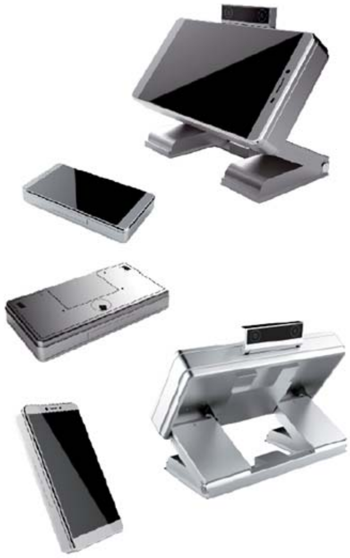
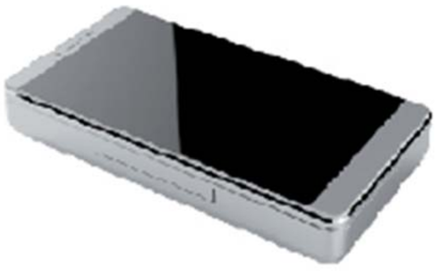
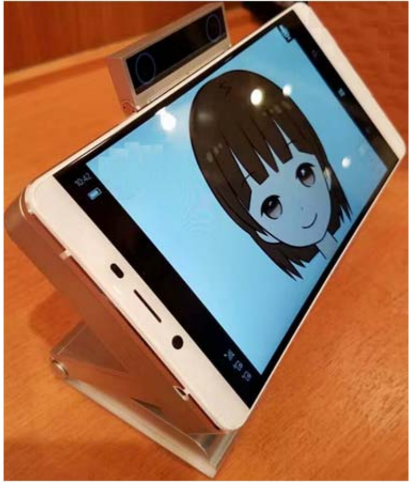


Fig. 1. Bobi RoBot.



(a) Moonlight box



(b) Opening

Fig. 2. Transformation of BoBi Robot.

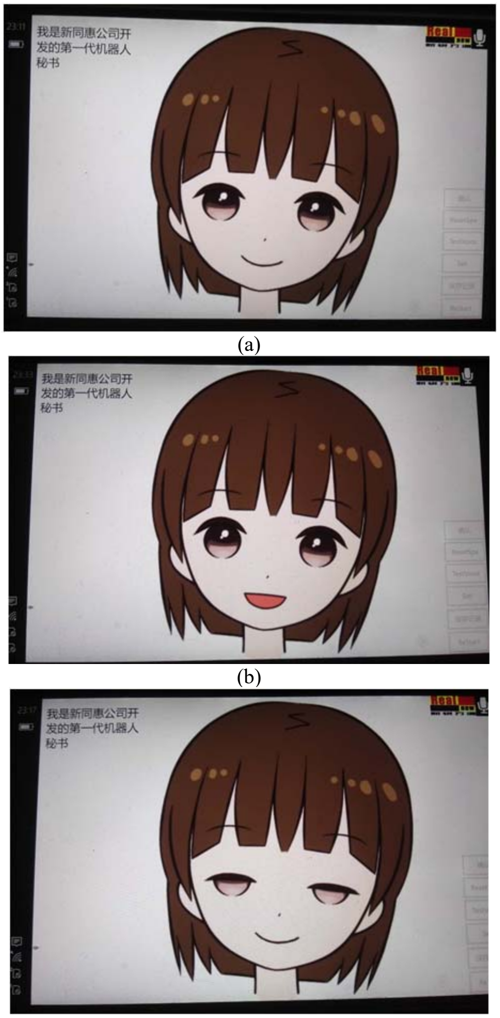
The rest of this paper is organized as follows: Section 2 begins with the design of the BoBi robot. Section 3 describes the basic functions of the BoBi robot, Section 4 presents the secretary works, Section 5 presents the results, and Section 6 makes our conclusions.

## II. DESIGN ON BOBI ROBOT

When it is closed, BoBi is a rectangular box with a smart phone size, and it automatically opens to transform from the box to a movable robot when its owner calls to it, “open” and then BoBi will perform many functions like humans such as moving, talking, emoting, singing, dancing, conversing with people to make people happy, enhance people's lives, facilitate relationships, have fun with people, connect people with the outside world and assist and support people as an intelligent personal assistant. We consider BoBi is a treasure and so call the box moonlight box that is “月光宝盒” in Chinese as shown in Fig. 2(a). There are wheels under its feet and it moves by them. We control BoBi mainly by voice interaction as well as by touch screen.

When it is opened BoBi shows a face on the screen and speaks with people. When speaking its eyes and mouth move like humans to emote and expresses feelings as shown in Fig. 2(b).

We make communication with BoBi by voice interaction. We give instructions to ask BoBi to sing, dance, move, tell jokes, make conversation and do secretary works including scheduling of works, schedule reminders, sending emails, calling phones, booking, making reservations, searching information, etc. It can remember and recognize people’s voices and understand its owner. We designed three main functions for the secretary works: intelligent meeting recording, multilingual interpretation and reading papers. Therefore, BoBi is portable, transformable, movable and intelligent.



(a)

(b)

(c)

Fig. 3. BoBi’s eyes and mouth moving like humans to emote and expresses feelings.

## III. BASIC FUNCTIONS BY VOICE INTERACTION

We make communication with BoBi and give instructions by voice interaction. BoBi works on a Windows Phone system. BoBi records people’s voice from Microphone when people speaks with BoBi, and then sends the voice data to a speech recognition server (Baidu speech recognition) via Web API and obtains a recognition result text. We apply a Voice Activity Detection (VAD) algorithm to detect the speech and the non-speech frames [14], and detect the stop of the people’s speech by the non-speech frames. When the stop time of the people’s speech is longer than a threshold, BoBi stops the voice recording and sends the voice data to the Baidu speech recognition server to obtain a result text. According to the result text, BoBi recognizes it as an instruction if it includes key words in the prepared key word list, otherwise, considers it as a chat text and sends the text to the Turing Robot web server by web API [15] to obtain an answer. After that BoBi replies the obtained answer by synthesizing the answer text into a human sounding speech using the Microsoft Windows Phone Speech Synthesizer and playing the synthesized speech. When it is an instruction, BoBi replies a prepared answer that is also synthesized into a human sounding speech by the Microsoft Windows Phone Speech Synthesizer so as to play the speech. Then BoBi acts according to the instruction such as singing, dancing and moving. When BoBi is speaking BoBi's eyes and mouth move like humans to emote and expresses feelings by displaying and changing continually the face pictures as shown in Fig. 3, where the face picture are changed from (a) to (b) and then to (c).

At the beginning, BoBi is a moonlight box with a smart phone size as shown in Fig. 2(a). When we say, "open", BoBi automatically opens soon and transforms from the box to a movable robot as shown in Fig. 2 (b). To do it, BoBi sends the people's voice data to the Baidu speech recognition server and obtains a result text. When the text is understood as open instruction, BoBi gives an instruction to the motor control board by a Bluetooth message that instructs motors to work to open.

The interface design is shown in Fig. 4, where we show BoBi's answers on the left field and show some function buttons on the bottom-right. We show some examples of Human-BoBi interaction in Chinese as follows:

Human: Says, "打开。(Open.)"

BoBi: Opens and replies, "你好， 我叫波比。(Hello, I am BoBi.)"

Human: Asks, "我能和你聊天吗！(Can I chat with you?)"

BoBi: Says, "当然可以！(Of course, you can!)"

Human: Asks, "你高兴吗？(Are you happy?)"

BoBi: Says, "很高兴见到你。(I am very happy to meet you.)"

Human: Says, "唱个歌吧！ (Sing please!)"

BoBi: Says, "好的！(Okay!)" and then sings a song.

Human: Says, "跳个舞吧！(Dance please!)"

BoBi: Says, "好的！(Okay!)" and then moves and rotates to show a dance while playing a music.

Human: Says, "红烧肉怎么做？(Please tell me how to cook the braised pork?)"

BoBi: Tells the cook method of the braised pork.

Human: Asks, "牛顿第一定律是什么？(What is Newton's first law?)"

BoBi: Describes Newton's first law.

Human: Says, "说个笑话。 (Tell joke please.)"

BoBi: Tells a joke.

BoBi can remember and recognize people's voices. We extract Mel-Frequency Cepstraum Coefficient (MFCC) features [16]-[19] that are widely used in the speech recognition field, and then apply Gaussian Mixture Models (GMMs) [20]-[22] to the MFCC features to recognize which speaker each voice belongs to. Some examples for speaker recognition are shown as follows:

Human: Says, "你知道我是谁吗？ (Do you know who I am?)"

BoBi: Replies, " 对不起， 我不认识你的声音。(I am sorry I do not know your voice.)".

Human: Says, "请帮我登记声音。 (Please register my voice for me.)"

BoBi: Replies, " 要登记声音吗？好的请说 2 秒以上的一段话。(Would you register your voice? Okay, please say for more than 2 seconds.)"

Human: Says something.

BoBi: Says, " 请问你叫什么名字？(Please tell me your name.)"

Human: Replies, " 刘建生 (Liu Jiansheng)"

BoBi: Says, " 刘建生对吗？如果不对请编辑更改。确认好以后， 请按确认键。(Is Liu Jiansheng right? If not, please edit it. Please press the 'confirmation' button after confirming it.)"

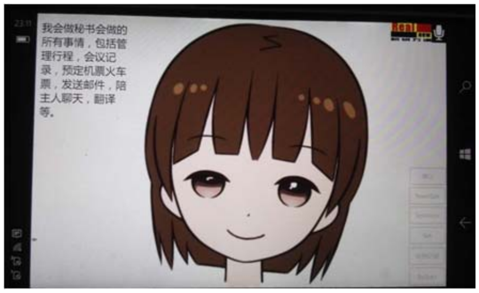
Human: Presses the 'confirmation' button.

BoBi: Says, " 您的声音正在登记， 请稍等。(Your voice is being registered, wait a moment please.)"

After a while BoBi says, " 您的声音已被登记， 谢谢。(Your voice has been registered, thank you.)"

Human: Asks, " 你知道我是谁吗？(Do you know who I am?)"

BoBi: Says, " 我认识你的声音， 你是刘建生。(I know your voice, you are Liu Jiansheng.)"



我会做秘书会做的所有事情，包括管理行程，会议记录，预定机票火车票，发送邮件，陪主人聊天，翻译等。

Fig. 4. Interface design.

## IV. SECRETARY WORKS

BoBi does secretary works including scheduling of works, schedule reminders, sending emails, calling phones, booking, making reservations, searching information, etc. We designed three main functions for the secretary works: intelligent meeting recording, multilingual interpretation and reading papers. In this section we present the main three functions.

### A. *Intelligent Meeting Recording*

We designed an intelligent meeting recording system for BoBi. Before starting the recording system, people who will attend the meeting needs to register his voice to BoBi. When BoBi is records meeting, a Microphone records voice that includes several persons' speeches while applying another process to recognize the voice at the same time to transform it into text, resulting in an online meeting recording. It is necessary to take the meeting voice from a Microphone and process the voice to transform it into text at the same time because it is inconvenient to wait a long time for the processing result if we process the voice after the meeting. Therefore, we apply a multithread processing to do both the two works (taking the voice from a Microphone and processing the voice to transform it into text) at the same time.

We apply the Voice Activity Detection (VAD) algorithm [14] to detect the speech and the non-speech frames to segment the recorded voice into some parts. Then we apply GMMs to the extracted MFCC features from each part to recognize which speaker it belongs to. After that we send each speaker's voice to the Baidu speech recognition server to obtain a result text.

We show some examples of the intelligent meeting record as follows:

Human: says, "请帮我们会议记。(Please help us to record a meeting.)"

BoBi: replies, "您要会议记录吗？好的，请开始说话。(Would you record a meeting? Okay, please start your meeting.)"

Then BoBi do the meeting record as shown in Fig. 5.

Human: says, "请结束会议记录。(Please close the meeting recording.)"

BoBi: Says, "您结束会议记录吗？好的。(Would you close the meeting recording? Okay.)"

### B. *Multilingual Interpretation*

BoBi can do translation among 25 kinds of languages as shown in Fig. 6. It sends people's voice to the Baidu speech recognition server to obtain a result text. Then the result text is sent to the Baidu translation server to get a translation result. Finally, BoBi says the translation result by synthesizing the translation result into a human sounding speech using the Microsoft Windows Phone Speech Synthesizer and playing the synthesized speech. We show some examples of the multilingual interpretation as follows:

Human: Says, "中文翻译成英语。(Please translate Chinese into English.)"

BoBi: Replies, "中文翻译成英语吗？好的，请说。(Would you like me to translate Chinese into English? Okay, please start.)"

Human: Says, "你叫什么名字？(What is your name?)"

BoBi: Says, "What is your name?"

Human: Says, "结束翻译。(Finish please.)"

BoBi: Says, "结束翻译吗？好的。(Would you finish the translation? Okay!)"

Human: Says, "英语翻译成中文。 (Please translate English into Chinese.)"

BoBi: Says, "英语翻译成中文吗？好的，请说。(Would you like me to translate English into Chinese? Okay, please start.)"

Human: Says, "Are you happy?"

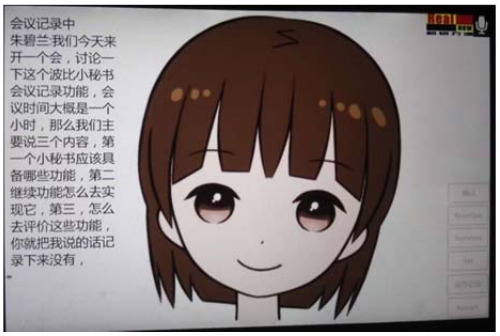
BoBi: Says, "你快乐吗？(Are you happy?)"

Human: Says, "Finish please."

BoBi: Says, "Okay!"

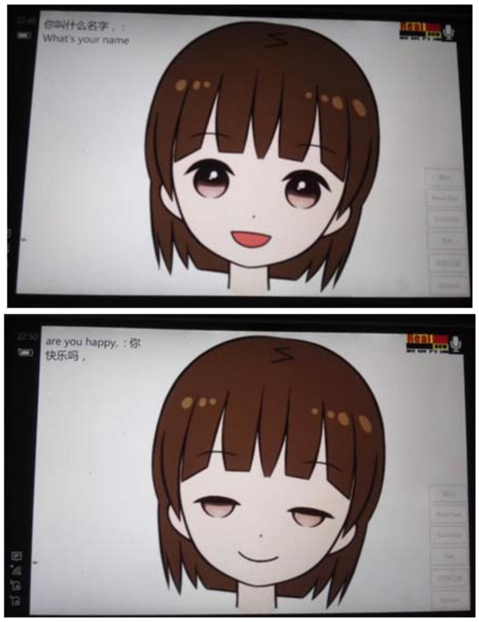
### C. *Reading Papers*

In our daily life, it is indispensable to read letters or papers. Because of it, we design BoBi to have a function to read papers. When we show BoBi a paper and say, "read please," BoBi will read the paper for us. To do it, BoBi uses its camera to capture a picture of the shown paper, and then applies an Optical Character Recognition (OCR) to recognize and transform the picture into text. Finally, BoBi reads the text to people by synthesizing the text into a human sounding speech using the Microsoft Windows Phone Speech Synthesizer and playing the synthesized speech.



会议记录中
朱碧兰:我们今天来开一个会，讨论一下这个波比小秘书会议记录功能，会议时间大概是一个小时，那么我们主要说三个内容，第一个小秘书应该具备哪些功能，第二继续功能怎么去实现它，第三，怎么去评价这些功能，你就把我说的话记录下来没有，

Fig. 5. Intelligent meeting recording.



你叫什么名字，:
What's your name

are you happy, : 你
快乐吗，

Fig. 6. Multilingual interpretation.

We show some examples of the function as follows:

Human: Shows a paper and says, “请帮我读一下。(Please help me to read the paper.)”

BoBi: Replies, “好的，请把纸张正对着我。(Okay! Please take the paper to me.)”

BoBi: Takes a picture, says, “我正在处理，请稍等。(I am processing this paper. Wait for a while please.)”

After a while BoBi reads the paper to people.

## V. RESULT

At voice interaction, BoBi can detect the people speech stops by the VAD method correctly and replies answers smoothly while expressing its feelings by moving its eyes and mouth and gives instructed the expected acts. It realized a natural human-BoBi interaction. The intelligent meeting record system can almost correctly segment each speaker voice and record meeting. The multilingual interpretation can translate among 25 kinds of languages and on the results the correct rate is about 90%.

## VI. CONCLUSION

We have presented an intelligent personal robot: BoBi secretary. BoBi opens and transforms from the moonlight box “月光宝盒” to a movable robot. It performs many functions like humans such as moving, talking, emoting, singing, dancing, and conversing with people to make people happy, and assist and support people as an intelligent personal assistant. BoBi does all works which a secretary is doing. BoBi is a portable, transformable, movable and intelligent robot. We realized the natural human-BoBi interaction. In the future, we will evaluate the usability of each function in detail.

## ACKNOWLEDGMENT

This work is partially supported by Japanese Grant-in-aid for Scientific Research C-15K00225.

## REFERENCES

[1] J. Wallén, “The history of the industrial robot,” in Linköping University Electronic Press, 2008.

[2] https://en.wikipedia.org/wiki/Siri

[3] M. Gurman, “iOS 9.2 to bring Arabic support to Siri following UAE store openings,” 9 to 5 Mac. November 6, 2015.

[4] https://en.wikipedia.org/wiki/Cortana_(software)

[5] Chris Lau, “Why Cortana Assistant Can Help Microsoft in the Smartphone Market,” The Street, March 18, 2014.

[6] https://deagostini.jp/rot/

[7] http://palmigarden.net/site/index.html

[8] https://plen.jp/

[9] https://en.wikipedia.org/wiki/Pepper_(robot)

[10] Sam Byford, “SoftBank announces emotional robots to staff its stores and watch your baby – Pepper will go on sale for under $2,000 in February,” theverge.com, Vox Media, June 5 2014.

[11] https://en.wikipedia.org/wiki/Nao_(robot)

[12] “Nao robot replaces AIBO in RoboCup Standard Platform League,” Engadget, August 16 2007.

[13] “Robot that walks, talks, emotes like humans ... ‘Nao’ ”, Times of India, February 4 2013.

[14] Javier Ramırez et al, “Effcient voice activity detection algorithms using long-term speech information,” In: Speech communication, 42.3, pp. 271–287, 2004.

[15] http://www.tuling123.com/help/h_cent_webapi.jhtml?nav=doc

[16] D. A. Reynolds, “Experimental evaluation of features for robust speaker identification,” IEEE Transactions on Speech and Audio Processing, Vol. 2, pp. 639-643, 1994.

[17] J.P. Openshaw, Z.P. Sun, J.S. Mason, “A comparison of composite features under degraded speech in speaker recognition,” Proc. 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 371-374, 1993.

[18] P. Aarabi et.al, “Phase-based speech processing,” Woned Scientific, 2006.

[19] G. Shi et.al, “On the importance of phase in human speech recognition,” IEEE Trans. Audio, Speech and Language Processing, Vol.14, No.5, pp1867-1874, 2006.

[20] D. A. Reynolds, “Speaker identification and verification using Gaussian mixture speaker models,” Speech Communication, Vol. 17, No. 1-2, pp. 91-108, 1995.

[21] S. Calinon and F. Guenter and A. Billard, “On Learning, Representing and Generalizing a Task in a Humanoid Robot,” IEEE Transactions on Systems, Man and Cybernetics, Part B. Special issue on robot learning by observation, demonstration and imitation, Vol. 37, No. 2, pp. 286-298, 2007

[22] Douglas A Reynolds and Richard C Rose, “Robust text-independent speaker identification using Gaussian mixture speaker models,” In: Speech and Audio Processing, IEEE Transactions on 3.1, pp. 72–83, 1995.