ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 8(S5), 48–51, March 2015* DOI : 10.17485/ijst/2015/v8iS5/61476

The Prosodic Conditions in Robot's TTS for Children as Beginners in English Learning

Jiyoung In[1] and Jeonghye Han[2*]

[1]Department of Linguistics, Chungnam National University, Daejeon, South Korea; injy0729@cnu.ac.kr
[2]Department of Computer Education, Cheongju National University of Education, Cheongju, South Korea; hanjh@cje.ac.kr

Abstract

A robot's recognition and diagnosis of pronunciation and its speech are the most important interactions in RALL. This paper analyzes para verbal characteristics, F0 range and speaking rate, in English TTS engine speech which falls in the category of robot speech for Korean learners as beginners. In the actual result of this research, beginners did not change before and after repeating robot TTS, however it was confirmed that the speaking rate as the other prosodic variable increased close to the robot TTS's speaking speed. Therefore robot TTS can be a suitable educational means to correct errors and train for speaking rate of English learners. However, it is still inadequate for the average F0 range of native speakers.

1. Introduction

Communication between robots and humans is the most important component in utilizing the robots for educational purposes. Educational service robots refer to intelligent robots used in the teaching and learning environment. When educational service robots are used for learning, they are primarily used language learning other than native language. This is called robot-assisted language learning, RALL[1]. Currently, there are studies on RALL, particularly in English learning, in Korea and Japan. There studies report on not only triggering of motivation but also on the increased effectiveness of learning[1,6–8].

Today, the focus of language learning is set on communication. In communication-oriented learning, the importance of prosody learning is emphasized in order to precisely deliver the intention of the speaker and to effectively communicate with the native speakers in the actual speaking situation[2–4]. In keeping with the foreign language learning trend which emphasizes prosody learning to enhance the fluency of learners and help them communicate effectively, prosody should be given important weight also in the English education through robots' TTS. However, existing works are focused on improving the educational effectiveness of educational service robots in the autonomous type while there is no experimental study available on prosodic aspects.

Therefore, the main objective of our study to analyze if English accent education through robots' TTS is capable of correcting and training prosodic errors made by Korean English learners and if learners can pick up the most natural English accent, in a country where RALL is most widely used. Through this analysis, we intend to see whether English accent education through robots' TTS can be effective means for Korean English learners in acoustic phonetic perspective. Based on our analysis of the results, we provide suggestions on what should be included in prosodic conditions for robot TTS in RALL in order to provide effective foreign language accent learning for the language learners.

2. Related Work

2.1 RALL

The robots take the role of native speaker communicating with the learners in most RALL. According to Kanda et

**Author for correspondence*

al., RALL is effective because the robots not only conduct language learning but also establish relationships[7]. Lee et al. provides results from its experiment on English language learning of elementary school students using two types of robots. They reported a major improvement in the participants' speaking skills after activities with robot MERO which is capable of automatic scoring of pronunciation quality for students' speech and with robot Engkey taking the role of store clerk and students taking the role of customers in a role playing act[6]. Park et al. provide analysis on lesson plan in after-school English learning utilizing ROBOSEM and report the same result for effectiveness of learning[8].

Yet, previous RALL studies on the three message types in communication, namely verbal, paraverbal and nonverbal, have only provided academic achievement level which is limited in the verbal activities of robot TTS but have not analyzed acoustic phonetic differences by TTS paraverbal.

2.2 Average F0 Range and Speaking Rate among Korean English-Learners

What distinguishes the accent of native English speakers from non-native speakers is not the accuracy of segmental sound but prosodic elements such as accent and speaking rate[9,10]. F0 range is prosodic variable that shows the degree of stress change within a sentence and because F0 of objective language is known to be affected a lot by the native language of foreign language learners[11–13]. As shown in Table 1, Korean English-learners have narrower F0 range compared to English native speakers due to the fact that they are affected by their native language which does not have a lot of stress changes within a sentence[14]. The narrow F0 value gives auditory consciously negative affects to the English native listeners[9].

In addition, speaking rate is a standard which allows making a judgment on how the foreign language learner is producing the natural accent of the language intended to learn; foreign language learners tend to have slower speaking rate when they are not used to the target language; however they develop to speak faster as they get used to the language[15].

Table 1. Average F0 range differences between the Korean English-learners and English natives14

	Native (Hz)	Learner (Hz)	Standard of achievement(%)
Declarative	215	103	48
Interrogative	240	101	42

Therefore, based on the previous studies on prosody where subjects were Korean English-learners, we set F0 range and speaking rate as our analysis variables in order to analyze whether English accent learning is effective through robots' TTS.

3. Method and Participants

3.1 Characteristics of Robot TTS

This study used the ROBOSEM. Each experimental sentence was designed with a value based on the basic level default of robots' TTS. Table 2 presents average F0 range and average speaking rate of Robot's TTS.

3.2 Experimental Design

We selected eight students – both boys and girls (first and second graders) among students in the after-school English program who have never lived abroad in English speaking countries and who can read the short sentences in English provided by our researchers. They have already been taking English lessons with RALL on a regular basis over the past eight months in their after-school programs. Thus, we can assume that there is no novelty effect in robot supplementary learning in English learning for our subjects. We divided the selected eight children experiment group into TTS type (adult, child) and assigned four (2 boys and 2 girls) in each group.

All recordings were used Praat sound analysis program. We conducted two rounds of recording processes. The first recording was conducted in order to analyze the average F0 range of children learners before they repeat after the robots' TTS. The second recording was conducted in order to analyze the average F0 range of children learners after the children repeated robots' TTS.

For experiment sentences, 2 declarative sentences (Can you fly?/Do you have a pen?) and 2 interrogative sentences (Yes, I can fly./ Yes, I have a pen.) were used. All

Table 2. Average F0 range and speaking rate of Robot's TTS

TTS	Sentence type	Range(Hz)	Speaking rate
Adult	Declarative	151.72	2.59
	Interrogative	129.79	3.45
Child	Declarative	182.75	2.57
	Interrogative	149.76	3.41

experiment sentences were repeated at least three times and total sentences that were used for the analysis were 192 sentences.

4. Results

4.1 Average F0 range by TTS types

First, changes in the average F0 range of children learners before and after repeating after the robot TTS were observed as shown in Table 3.

We discovered that the average F0 range of children learners who repeated after the adult voice TTS actually decreased both in declarative and interrogative sentences. As a result of Wilcoxon signed rank test, there were no significant changes in the average F0 range in declarative sentences [W(23)=300, z=-1.657, p=0.097] However, there were significant changes in the average F0 range in interrogative sentences. [W(23)=300, Z=-2.714, p<0.01]. For the children learners who repeated after the child voice TTS, there were different outcomes for the average F0 range by sentence type. For declarative sentences, the average F0 range was found to be wider after repeating, however it was found to be narrower for the interrogative sentences. Nonetheless as a result of Wilcoxon signed rank test, neither declarative nor interrogative sentence showed any significant change in the average F0 range [decW(23)=300, z=-0.343, p=0.732; intW(23)=300, z=-0.714, p=0.475].

4.2 Speaking Rate by TTS

Table 4 shows the result of our analysis on the changes in the average speaking rate of children learners after repeating after the robot TTS. After repeating after the adult voice TTS, the average speaking rate of children learners increased for declarative sentences and interrogative sentences.

Based on Wilcoxon signed rank test, both declarative and interrogative sentences showed significant changes in the average speaking rate; it was confirmed that the average speaking rate increased after repeating adult voice TTS for the children learners [dec W(23)=300, z=-4.286, p<0.001; intW(23)=300, z=-4.286, p<0.001]. In case of child voice TTS, the average speaking rate of children learners also increased. In Wilcoxon signed rank test, both declarative and interrogative sentences showed significant changes in the average speaking rate, and we found that the average speaking rate increased after repeating TTS [decW(23)=300, Z=-3.714, p<0.001; intW(23)=300, Z=-3.771, p<0.001].

Table 3. The changes in the average F0 range

TTS	Sentence type	F0 range(Hz)	
		before	after
Adult	Declarative	122.05	101.82
	Interrogative	115.62	85.22
Child	Declarative	116.72	118.40
	Interrogative	108.93	102.77

Table 4. The changes in the average speaking rate

TTS	Sentence type	Speaking rate(syl/sec)	
		before	after
Adult	Declarative	1.97	2.63
	Interrogative	2.21	3.19
Child	Declarative	2.18	2.61
	Interrogative	2.37	3.21

5. Conclusion

It is essential for Korean English-learners to widen the average F0 range to the level of native speakers by clearly expressing the stress changes within sentences and to learn and train in order to obtain fast speaking rate[13–15]. However, as shown in Table 2, based on the average F0 range of robot TTS which was narrower than the average F0 range of native English speakers, it is difficult to expect that there will be effective results from English accent learning through robot TTS. Actually, as this our results suggest, the average F0 range of the children learners either did not change or even decreased after repeating TTS. Even if the fact that there was no change in the average F0 range for children learners was temporary, the average F0 range of robot TTS could not change the English accent of the learners.

However, speaking rate is increased after repeating TTS in speaking rate. The fact that the children learners showed changes in a short period of time for average speaking rate was probably because the speed of speech of robot TTS was more easily recognized than the stress changes within sentences in auditory conscious sense. Nevertheless, it is essential to understand and to train the stress changes within sentences in order for Korean English-learners to speak natural English accents.

However, the robot TTS have narrower F0 range compared to English native speaker.

This conclusion is drawn from the fact that the most TTS engines are implemented based on broadcastings and news. Thus, there has to be technological development in TTS engines for language learning in order to achieve the stress changes within the average F0 range of the native speakers. In other words, the prosody of robot TTS for foreign language learning must be able to implement the prosodic characteristics of target language for the foreign language learners on top of the capability to produce natural human prosody, in order to be fully functional as a learning tool for foreign language accent learning.

6. Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (NRF-2012R1A1A2039797). Special thanks to teachers and students from CharmSaem elementary school and Yujin Robotics, Inc. for their support.

7. References

1. Han JH. Robot assisted language learning. Lang Learn Tech. 2012 Feb; 16(3):1–9.
2. Pennington BF. Diagnosing learning disorders: A neuro-psychological framework. J Autism Dev Disord. 40(9):1165–6.
3. Wong R. Teaching pronunciation: focus on english rhythm intonation. New York: Prentice-Hall; 1985.
4. Chung HS. Study on the rhythm of Korean EFL learner's english pronunciation. Phonetics and Speech Science. 2009; 1(2):141–9.
5. Han JH. Robot-aided learning and r-learning services. In: Chugo D, editor. Human-Robot Interaction. 2010. ISBN: 978-953-307-051-3.
6. Lee SJ, Noh HJ, Lee JH, Lee KS, Lee GG, Sagong SD, et al. On the effectiveness of robot-assisted language learning. ReCALL. 2011; 23:25–58.
7. Kanda T, Sato R, Saiwaki N, Ishiguro H. A two-month field trial in an elementary school for long-term hri. Robot IEEE Trans. (Special Issue on Human-Robot Interaction). 2007; 23(5):962–71.
8. Park SJ, Han JH, Kang BH. Teaching assistant robot, ROBOSEM, in English class and practical issues for its diffusion. Proceedings of the IEEE ARSO (A workshop on advanced Robotics and its Social impacts) conference; 2011 Oct 2-4; Half-moon bay, California, USA.
9. Munro MJ, Derwing TM. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. Lang Learn. 2000; 45(1):73–97.
10. Munro MJ, Derwing TM. Modeling conceptions of the accentedness and comprehensibility of L2 speech-The role of speaking rate. Stud Sec Lang Acquis. 2001; 23:451–68.
11. van Benjooizen R. Sociocultural aspects of pitch differences between Japanese and Dutch women. Lang Speech. 1995; 38(3):253–65.
12. Scherer K. A cross-cultural investigation of emotional inferences from voice and speech: implications for speech technology. INTERSPEECH; 2000. p. 379–82.
13. Kang SH, Rhee SJ. A study on the suprasegmental parameters exerting an effect on the judgement of goodness or badness on Korean-spoken English. Phonetics and Speech Science. 2011; 10(4):3–10.
14. Rhee SJ, Cho CH, Moon SY. Korean & native speakers' high-low range differences in f0 and its role in pronunciation assessment. Phonetics and Speech Science. 2003; 10(4):93–103.
15. Guion SG, Flege JE, Liu SH, Yeni-Komshian GH. Age of learning effects on the duration of sentence produced in a second language. Applied Psycholinguistics. 2000; 21:205–28.