Journal of The Korean Association of Information Education Vol. 20, No. 4, August 2016, pp. 323-332
http://dx.doi.org/10.14352/jkaie.2016.20.4.323

로봇보조언어교육을 통한 초등 영어 학습자의 운율 변화

인지영* · 한정혜**

충남대학교 언어학과* · 청주교육대학교 컴퓨터교육학과*

요 약

로봇의 발음인식과 진단 그리고 발음빠르기는 로봇보조언어교육의 가장 중요한 상호작용이다. 이 연구는 한국인 초등 영어 학습자를 위하여 음율적 오류를 수정함으로써 원어민과 같은 억양을 산출하기 위한 로봇음성합성기의 효과성을 측정하기 위한 것이다. 이를 위해 초등 4학년 영어학습자들의 F0 범위값과 발화 속도라는 음성음향적 변수를 측정하여 분석하였고, 그 결과를 정규 영어교육의 시작하지 않은 1학년 학습자와 비교하였다. 로봇음성합성기를 활용한 언어학습에서 두 집단은 F0값보다 발화속도 변인에 반응하였다.

키워드 : 로봇보조학습, 교육용 로봇, 음향음성학, 운율 변수, F0, 발화 속도

The Prosodic Changes of Korean English Learners in Robot Assisted Learning

Jiyoung In* · JeongHye Han**

Dept, of Linguistics, Chungnam National University* ·
Cheongju National University of Education**

ABSTRACT

A robot's recognition and diagnosis of pronunciation and its speech are the most important interactions in RALL(Robot Assisted Language Learning). This study is to verify the effectiveness of robot TTS(Text to Sound) technology in assisting Korean English language learners to acquire a native-like accent by correcting the prosodic errors they commonly make. The child English language learners' F0 range and speaking rate in the 4th grade, a prosodic variable, will be measured and analyzed for any changes in accent. We compare whether robot with the currently available TTS technology appeared to be effective for the 4th graders and 1st graders who were not under the formal English learning with native speaker from the acoustic phonetic viewpoint. Two groups by repeating TTS of RALL responded to the speaking rate rather than F0 range.

Keywords : RALL, TTS, Prosodic Parameter, F0 Range, Speaking Rate

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1A2039797;2015R1D1A1A09060450).
교신저자 : 한정혜(청주교육대학교 컴퓨터교육학과)
논문투고 : 2016-06-08
논문심사 : 2016-06-09
심사완료 : 2016-08-25

1. Introduction

The development of smartphone devices have helped foreign language education overcome spatial limit such as classroom or computer room and time limit. Such foreign language education which is based on mobile devices is called mobile-assisted language learning (MALL). Since mobile devices are easy to get in hands to anybody and they are portable, MALL is spotlighted for being a new way of learning foreign language[8]. For quite some time, robots have been used in various parts of our lives, and now there are efforts to use robots to communicate with humans in more effective ways. In particular, 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 foreign language(L2) learning other than native language(L1). This is called RALL, robot-assisted language learning [7]. As new technology such as robots or mobile devices got introduced, many studies have been progressed in order to verify whether such way of teaching foreign language can increase educational effects [4][7][8][11]. This study focuses on prosodic changes of Korean English learners in RALL.

Today, the focus of language learning is set on communication. In communication-oriented learning, the prosodic elements – the determinants of the overall speech's fluency – are viewed as the critical educational goal. Prosody includes accent, rhythm, intonation and others. It determines sentence types and takes role of delivering the attitude and emotion of the speaker. Therefore in communication-focused 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 [3][5][12].

Currently, there are studies on RALL, particularly for English learning, in Asian countries such as Korea and Japan. There are two types of RALL: autonomous RALL, which has its own artificial intelligence, and tele-presence RALL, which is equipped to provide the tele-presence of educational services through a remote control the instructor uses[6][7].

2. Related Works

2.1 RALL

Numerous studies in the field report that RALL is effective because of its capacity to play the role of a native speaker who communicates and interacts with the learner while making a connection at an individual level[7][18].

Lee et al.(2011) provided 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 [15]. Park et al.(2011) analyzed the lesson plan in after-school English learning utilizing ROBOSEM and reported the same result for effectiveness of learning [16].

The majority of existing studies on RALL tend to have a limited focus on academic achievement through the verbal activities offered by robot TTS at the expense of neglecting other components of communication, para-verbal and non-verbal properties. Given today's foreign language education trend, which emphasizes the acquisition of a proper accent, studies regarding the para-verbal aspect of robot TTS are more than necessary. However, such studies are non-existent so far.

2.2 Korean English language learners' prosody

What distinguishes a non-native English speaker from a native English speaker is not so much the phoneme segments but rather the prosodic elements such as accent and speaking rate[9][10]. The F0 range is a prosodic variable that shows the changes in stress within a sentence. The F0 range of the target language (L2) is known to be highly influenced by the speaker's native language (L1). English is a stress-timed language which has large degree of stress change within a sentence and produces accent through changing the stress.

On the contrary, Korean is a syllable-timed language which is not sensitive to stress and has smaller degree of stress change within a sentence compared to English. Therefore, as displayed 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 [17]. Non-native English speakers' narrow F0 range often leads to negative auditory perception by native English speakers.

<Table 1> Average F0 range differences[17]

	American (Hz)	Korean (Hz)	Achievement rate (%)
declarative	215	103	48%
interrogative	240	101	42%

In fact, native English speakers tend to perceive the pronunciation of Korean English language learners with a narrower F0 range to be poorer than their counterparts with a wider F0 range [9][17]. Speaking rate also influences the degree to which a non-native English speaker's accent approximates that of a native English speaker. In general, a foreign language learner with a lower L2 proficiency level has a slower speaking rate, which, in turn, picks up as he/she acquires a higher proficiency level[13].

As seen so far, research on Korean English language learners' accents indicate that, in order for Korean English language learners to produce a native-like English accent, language instruction and training need to be geared towards acquisition of the F0 range and speaking rate that approximate those of a native English speaker[14].

Based on the results of previous studies conducted on the accent of Korean English language learners, our study established the F0 range and speaking rate as variables with which to analyze the effectiveness of robot TTS technology in delivering English accent instruction. It was studied the changes of the F0 range and speaking rate repeating after the robot TTS that were targeting first grade of elementary school students (2 boys and 2 girls) [2]. They had changed the participants from the lower grades to higher grades, 4th grader (8 boys and 8 girls), and measured the F0 dynamic range of before and after repeating after the robot TTS as an intermediate result [1]. In this paper, we expanded the experiment of measuring speaking rate of 4th graders additionally.

3. Experiment Design

3.1 TTS for robots

This study used the educational service robot, ROBOSEM. While various types of TTSs, developed by Microsoft and Bell labs, may adopted to robots, we used the basic TTS engine installed on ROBOSEM for our experiment for the following reasons.

- Children who are subjects of our experiment have been exposed to the voice of ROBOSEM over along period of time for RALL.

- According to three Korean acoustics experts, the prosody of TTS installed on ROBOSEM was determined to be on par with other TTS engines.
- 5 Korean children do not recognize the difference among various TTS engines as a pilot experiment.
- ROBOSEM is installed with TTS engine which supports Korean language, the learners' native language[18].

The TTS engine installed on ROBOSEM supports three types of voice: male, female and children; it has range of 0.5~1.7 pitch and 0.6~2.0 for speed. <Table 2> below displays the average F0 range and speaking rate of adult TTS and child TTS installed in ROBOSEM.

<Table 2> Average F0 range and Speaking rate of Robot's TTS [2]

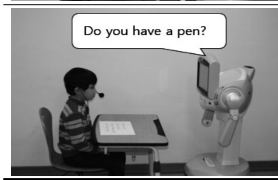
TTS	sen_type	F0 range	speaking rate
Adult	declarative	151.72	2.59
	interrogative	129.79	3.45
Child	declarative	182.75	2.57
	interrogative	149.76	3.41

3.2 Participants and test sentences

The study participants include Korean lower elementary students (4th graders) who speak standard Korean, with no history of residing in an English-speaking country. They had never interacted with an educational robot, although they had more exposure to English and native English speakers throughout the course of their school career. 16 students (8 boys, 8 girls) who expressed a desire to participate in the study were selected as the study participants. They were subsequently divided into two groups of adult TTS and child TTS.

For experiment sentences, considering the English capability of children learners in Korean elementary schools, a total of four most common types of sentences - declarative and interrogative sentences - were adjusted for our purpose in <Table 3>.

<Table 3> test sentences

Interrogative	declarative
	
	

This is because the F0 range is not identical in all the sentences produced by a native English speaker [17]. For this reason, different types of target sentences were required. A declarative sentence and an interrogative sentence were selected for the study, which the participants could learn with relative ease.

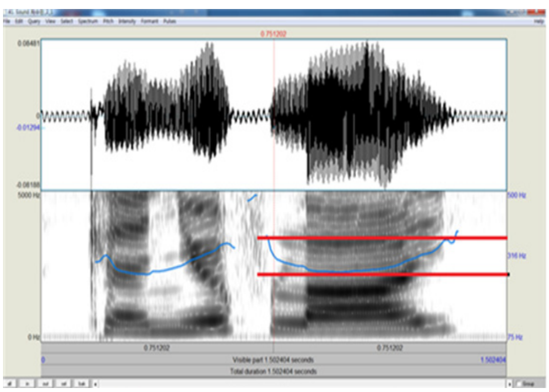
3.3 Experiment

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 children learners were asked to read the experiment sentence to be recorded. The second recording was conducted in order to analyze the average F0 range of children learners after the children repeated robots' TTS; the learners listened and repeated each sentence robot TTS spoke.

All recordings were conducted in a completely closed and quiet classroom and used Praat sound analysis program ver. 5.5.23 under the condition of standard rate of 44,050Hz, 16bit in two-ways. All target sentences were repeated a minimum of three times.

Prosodic variable value F0 range used in the analysis was checked with naked eyes for the maximum and minimum values to be measured. For speaking rate, the total speech time was divided by the number of syllables within the sentence. The total speech time was measured based on spectrogram and wave form.

Experimental voice data were analyzed using Praat voice analysis program ver.5.3.23. Praat voice analysis program is used by many researchers for prosody analysis because it is free for download and known to be useful.



(Fig. 1) Praat voice analysis screen shot: “Can you fly?” by the robot adult TTS

(Fig. 1) shows an example of analysis that actually measured F0 range with Praat. We analyzed F0 range by measuring the highest score and the lowest score of an accent curve in blue.

4. Result

4.1 F0 range

<Tables 4> and <Table 5> display the results of analysis regarding the 4th graders' average F0 range both before and after the robot TTS listen and repeat activity.

<Table 4> Average F0 range change for the robot adult TTS

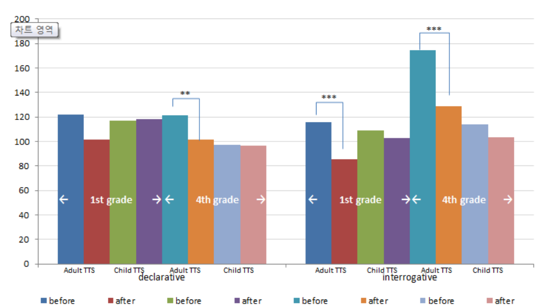
	F0 range(Hz)		T	p-value
	before	after		
dec	121.22	101.50	4.588	p<.01
int	174.26	128.99	2.580	p<.05

The average F0 range of the 4th graders who repeated after the adult TTS voice appeared to have become narrower than before the activity. This tendency was observed in both declarative and interrogative sentences. The results of the matching sample t-test indicated significant differences for both declarative and interrogative sentences; thus, the average F0 range was seen to be narrower after the TTS activity.

<Table 5> Average F0 range change for the robot child TTS

	F0 range(Hz)		T	p-value
	before	after		
dec	97.28	96.56	0.131	p<.05
int	114.22	103.19	1.656	p>.05

The average F0 range of the children who repeated after the child TTS voice showed different results. These children's average F0 range showed no significant change in either the declarative or interrogative sentences, and thus, their average F0 range was seen not to be affected by the robot TTS activity. We compared this result with [2] in (Fig. 2).



(Fig. 2) Comparison of the average F0 rage between [2] and this experiment

4.2 Speaking rate

<Tables 6> and <Table 7> display the analysis results regarding the changes in the average speaking rate of the 4th graders before and after the robot TTS activity.

<Table 6> Speaking rate change for the robot adult voice TTS

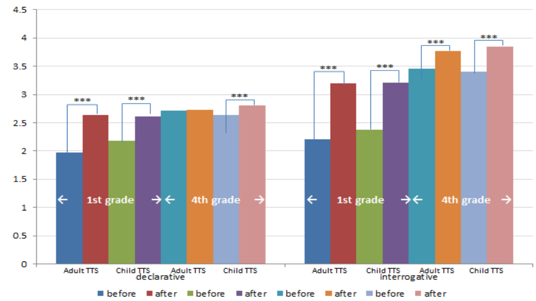
	speaking rate(syl/sec)		T	p-value
	before	after		
dec	2.72	2.73	-0.180	p>.05
int	3.46	3.77	-3.003	p<.01

As shown in <Table 6>, the average speaking rate of the 4th graders who repeated after the adult TTS voice increased. However, matching sample t-test results indicated that the changes were dependent on the sentence type. In the case of a declarative sentence, the change was insignificant, and thus, the average speaking rate is observed to be unaffected. On the other hand, the changes were significant in the case of the interrogative sentence, and thus, the average speaking rate was seen to have increased after the robot TTS speaking activity.

<Table 7> Speaking rate change for the robot child voice TTS

	speaking rate(syl/sec)		T	p-value
	before	after		
dec	2.63	2.80	-3.183	p<.01
int	3.40	3.85	-6.006	p<.01

As shown in <Table 7>, the average speaking rate of the children who repeated after the child TTS voice also increased. The match sample t-test results indicated significant changes in the average speaking rate for both the declarative and interrogative sentences. As such, the average speaking rate of these children was seen to have increased after the robot TTS speaking activity. We compared this result with [2] in (Fig. 3).



(Fig. 3) Comparison of the average speaking rate between [2] and this experiment

5. Conclusion

Since Korean English-learners' F0 range widens and speaking rate increases as their English level improves, F0 range and speaking rate can be the standard for assessing whether the pronunciation of learners' English accent is fluent[13][14][17]. This research analyzed whether there were changes in English accent for the children learners before and after repeating robot TTS through acoustic phonetic prosodic variables such as F0 range and speaking rate. The results indicated either no change or a decrease in the students' F0 range after the robot TTS speaking activity.

To be able to speak natural English accent, 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. 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 our results suggest, the average F0 range of the children learners decreased after repeating TTS. The unaffected F0 range, and in some cases, decreased F0 range, of the students may be attributable to the interference of their native language, which does not exhibit significant variety in intonation within a sentence. However, other possible explanations include that the average F0 range of the robot TTS failed to provide linguistic input sufficient enough to elicit changes in the learners' English accent, albeit temporarily.

However, there were changes after repeating TTS in speaking rate, the other prosodic variable. Compared to the average speaking rate of robot TTS, it was confirmed that the lower elementary students' speaking rate increased, close to that of robot TTS [2]. On the other hand, the average speaking rate of the 4th graders had already approximated that of the robot TTS. This may be explained by the fact that the 4th graders were generally more familiar with English. Some of them even exhibited a higher speaking rate than the robot TTS, but the increase was not as significant as that of the lower elementary students, who were generally less familiar with English. Changes in the average speaking rate within such a short period of time may be due to the fact that the robot TTS speaking rate registers more readily with the students (in their auditory perception) than the F0 range, a more subtle concept which concerns stress changes within a sentence. As such, robot TTS is a potentially effective educational tool with which to assist learners in acquiring a native-like speaking rate and, ultimately, a more natural English accent.

It is essential to understand and to train the stress changes within sentences in order for Korean English-learners to speak natural English accents. Particularly, even if listening and training the native English speakers' accent is sustained, it takes long learning time and persistent training to have the average F0 range close to that of the native English speakers. However, as shown <Table 1> and <Table 2>, the robot TTS showed maximum 85% and minimum 54% of the standard of achievement compared to the average F0 range of native English speakers. Thus, it may be difficult to achieve effective learning for English accent using robot TTS. In order to facilitate effective instruction on the prosodic features of English with the use of robot TTS, further research and technology development for language instruction robot TTS is quite necessary. Currently, the TTS engine is mostly optimized for use in broadcasting systems and news. To fully utilize the robot TTS for language instruction, however, its capacity to deliver sufficient linguistic input needs to receive as much attention as its capacity to produce a natural human voice and delivery. In other words, an effective robot TTS for foreign language instruction must be able to deliver the prosodic characteristics of the learner's target language while going beyond just producing natural human prosody, especially if it is to be used as an instructional tool for natural accent acquisition.

This study examined the potential of the robot TTS as an effective instructional tool in assisting child English language learners to produce a natural English accent. Subsequently, we proposed a minimum paralinguistic standard required of robot TTS for foreign language instruction based on our findings. The currently available TTS technology appears to be more effective for speaking rate improvement than accent improvement. As such, tele-presence RALL with a native English speaker on the other end seems to be more effective than an educational robot installed with TTS.

Our study has limitations in that only a particular type of TTS engine was concerned and the study participants were limited to Korean English language learners. Also, the number of samples were insufficient, thus it was difficult to generalize this result. Further research with an expanded scope to incorporate a variety of TTS engines and foreign languages will be beneficial in identifying the prosodic conditions required of effective robot TTS for foreign language instruction.

참고문헌

[1] Authors (2015). The Prosodic Conditions in Robot's TTS for Children as Beginners in English Learning, *Indian Journal of Science and Technology 8*, (S5), 48-51.

[2] Authors (2015). The Acoustic-Phonetics Change of English Learners in Robot Assisted Learning, ,HRI'15 ACM/IEEE International Conference on Human-Robot Interaction extended Abstracts, March 2-5, 39-40.

[3] B.F.Pennington (1991). Diagnosing Learning Disorders: A Neuro-psychological Framework. New York: Guilford Press.

[4] George M. Chinnery (2006). Emerging Technologies-Going to the MALL: Mobile Assisted Language Learning. *Language Learning & Technology, 10*(1), 9-16.

[5] H. S. Chung (2009). A Study on the Rhythm of Korean EFL Learner's English Pronunciation. *Phonetics and Speech Science 1*(2), 141-149.

[6] J. H. Han (2010). Robot-Aided Learning and r-Learning Services. Human-Robot Interaction, ISBN: 978-953-307-051-3.

[7] J. H. Han (2012). Robot Assisted Language Learning. *Language Learning & Technology, 16*(3), 1-9.

[8] J. S. Yang (2013). Mobile Assisted Language Learning: Review of the Recent Application of Emerging Mobile Technology, *English Language Teaching, 6*(7), 19-25.

[9] M. J. Munro, & T. M Derwing (2000). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners, *Language Learning, 45*(1), 73-97.

[10] M. J. Munro, & T. M. Derwing (2001). Modeling conceptions of the Accentedness and comprehensibility of L2 speech-The role of speaking rate. *Studies on Second Language Acquisition, 23*, 451-468.

[11] Robert Godwin-Jones (2011). Emerging Technologies-MOBILE APPS FOR LANGUAGE LEARNING. *Language Learning & Technology, 15*(2), 2-11.

[12] R. Wong (1985). Teaching Pronunciation: Focus on English Rhythm Intonation. New York: Prentice-Hall.

[13] S. G. Guion, J. E. Flege, S. H. Liu, and G. H Yeni-Komshian (2000). Age of learning effects on the duration of sentence produced in a second language. *Applied Psycholinguistics, 21*, 205-228.

[14] S. H. Kang, S. J. Rhee (2011). A study on the Suprasegmental Parameters Exerting an Effect on the Judgement of Goodness or Badness on Korean-spoken English. *Phonetics and Speech Science, 10*(4), 3-10.

[15] S. J. Lee, H. J. Noh, J. H. Lee, K. S. Lee, G. G. Lee, S. D. Sagong and M. S. Kim (2011). On the effectiveness of Robot-Assisted Language Learning. *ReCALL, 23*, 25-58.

[16] S. J. Park, J. H. Han, and B. H. Kang (2011) Teaching assistant robot, ROBOSEM, in English class and practical issues for its diffusion. Proceedings of the IEEE ARSO (Advanced Robotics and its SOcial impacts) conference, Oct. 2-4, Half-moon bay, California, USA.

[17] S. J. Rhee, C, H. Cho, and S. Y. Moon (2003), Korean & Native Speakers' High-low Range Differences in F0 and its Role in Pronunciation Assessment. *Phonetics and Speech Science, 10*(4), 93-103.

[18] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro (2007). A Two-month Field Trial in an Elementary School for Long-term HRI. IEEE Transactions on Robotics(Special Issue on Human-Robot Interaction), 23(5), 962-971.

저자소개

인 지 영

2004 충남대학교 언어학과(학사)

2007 충남대학교 언어학과(석사)

2014 충남대학교 언어학과(박사)

2014~현재 충남대학교 국제언어교육센터 한국어 강사

관심분야: 음성학, 운율, 언어교육

e-mail: ji-young29@hanmail.net

한 정 혜

1998 충북대학교 전자계산학과(박사)

1998~1999 연세대학교 산업 시스템공학과 포닥 연구원, 연세대학교 인지과학연구소 선임연구원

1999년~2001년 행정자치부 국가전문행정연수원 통계연수부 전산교육 전임교수

2001~현재 청주교육대학교 컴퓨터교육과 교수

2011.2~2012.2 미국 Stanford 대학교 인간과학기술진보연구소(H-STAR) 방문 교수

2015.7~2017.7 한국로봇산업진흥원 비상임이사

관심분야 로봇보조학습, 로봇활용 SW교육, 인간과 로봇 상호작용(HRI)

e-mail: hanjh@cje.ac.kr