Effects of Pronunciation Practice System Based on Personalized CG Animations of Mouth Movement Model

Kohei Arai [1]
Graduate School of Science and Engineering
Saga University
Saga City, Japan

Mariko Oda [2]
Faculty of Engineering,
Kurume Institute of Technology University,
Kurume City, Japan

***Abstract*— Pronunciation practice system based on personalized Computer Graphics: CG animation of mouth movement model is proposed. The system enables a learner to practice pronunciation by looking at personalized CG animations of mouth movement model , and allows him/her to compare them with his/her own mouth movements. In order to evaluate the effectiveness of the system by using personalized CG animation of mouth movement model, Japanese vowel and consonant sounds were read by 8 infants before and after practicing with the proposed system, and their pronunciations were examined. Remarkable improvement on their pronunciations is confirmed through a comparison to their pronunciation without the proposed system based on identification test by subjective basis.**

Keywords- Pronunciation practice; mouth movement model; CG animation.

I. INTRODUCTION

There are many pronunciation practice systems [1] which allow monitoring voice waveform and frequency components as well as ideal mouth, tongue, and lip shapes simultaneously. Such those systems also allow evaluations of pronunciation quality through identification of voice sound. We developed pronunciation practice system (it is called "Lip Reading AI") for deaf children in particular [1]. The proposed system allows users to look at their mouth movement and also to compare their movement to a good example of mouth and lip moving picture. Thus users' pronunciation is improved through adjustment between users' mouth movement and a good example of movement derived from mouth movement model.

Essentially, pronunciation practice requires appropriate timing for controlling mouth, tongue, and lip shapes. Therefore, it would better to show moving pictures of mouth, tongue, and lip shapes for improvement of pronunciations [2]. Although it is not easy to show tongue movement because tongue is occluded by mouth, mouth moving picture is still useful for improvement of pronunciations. McGurk noticed that voice can be seen [3]. Some of lipreading methods and systems are proposed [4]-[6].

One of the key issues for improvement of efficiency of the pronunciation practice is personalization. Through experiments for the proposed "Lip Reading AI" with a number of examiners, it is found that pronunciation difficulties are different by examiner. Therefore, efficient practice needs a personalization. The proposed pronunciation practice system in the paper utilizes not only mouth movement of moving pictures but also personalization of moving picture by user.

The following section describes the proposed system followed by some experimental results with 8 examiners. Then effectiveness of the proposed system is discussed followed by conclusion.

II. PROPOSED PRONUNCIATION PRACTICE SYSTEM

A. *Lip Reading "AI"*

Previously proposed Lip Reading "AI" allows comparison between learner's mouth movement and reference movement with moving picture in real time basis. An example of display image is shown in Fig.1.

Figure 1 Example of display image of the previously proposed Lip Reading "AI".

B. *CG Animation of Reference Moving Picture*

In the system, real mouth images are used as reference movement of moving picture. Not only real mouth moving pictures, but also CG animation of mouth images can be used

[1] http://www.advanced-media.co.jp/products/amivoicecallpronunciation.html
http://www.prontest.co.jp/soft/
http://www.english-net.co.jp/~pros/1/ppower/progfeat.htm
http://shop.alc.co.jp/course/hc/
http://www.smocca.co.jp/SMOCCA/English/HatsuonRyoku/index.html
http://sgpro.jp/demo/

as shown in Fig.2. Using CG animation of mouth moving picture, much ideal reference could be generated. "Maya"[2] of CG animation software (Fig.3) is used to create reference mouth moving picture. Then it can be personalized. Namely, resemble CG animation to the user in concern can be created as shown in Fig.4.

Figure 2 CG animation based reference moving picture for monitoring mouth movements.

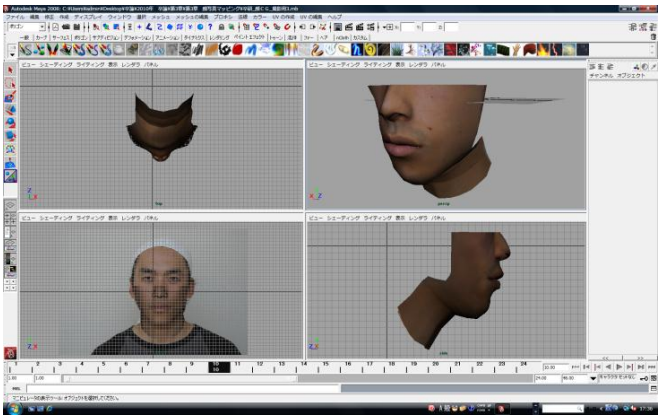


Figure 3 Example of Maya of CG animation software generated reference mouth and lip moving picture

In order to create resemble mouth moving picture to the user in concern, correct mouth movements are extracted from moving picture by using Dipp-MotionPro2D[3]. 12 of lecturers' mouth moving pictures are acquired with video camera and then are analyzed. Every lecturer pronounced "a", "i", "u", "e", and "o". Four feature points, two ends of mouth and middle centers of top and bottom lips are detected from the moving pictures. Four feature points when lecture close and open the mouth are shown in Fig.5 (a) and (b), respectively.

An example of motion analysis of four feature points when the lecture pronounces "a" is shown in Fig.6.

[2] http://ja.wikipedia.org/wiki/Maya

[3] http://secure.shanon.co.jp/ipjbio2008/exhidir/BIO/ja/company/exhibitor_1032.html

Figure 4 Resemble mouth and lip moving picture to the user in concern could be created with the CG animation software derived reference moving picture.

(a)Close the mouth

(b)Open the mouth

Figure 5 Four feature points when lecture close and open the mouth.

In the figure, red, yellow, light green, and blue lines show the top lip, the bottom lip, right end of mouth and left end of mouth, respectively. When the lecture pronounces "a", the bottom lip moves to downward direction remarkably while other three feature points (top lip and two ends of mouth) do not move so much. On the other hand, when the lecture pronounces "u", two ends of mouth moves so much in comparison to the other two (top and bottom lips) as shown in Fig.7.

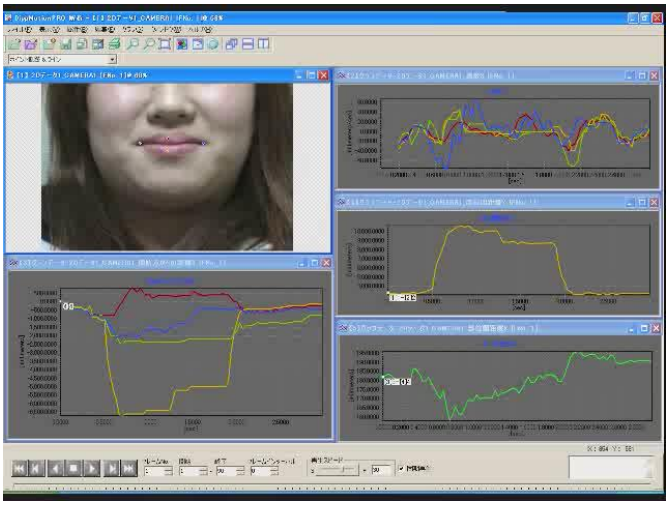


Figure 6 Example of motion analysis of four feature points when the lecture pronounces "a"

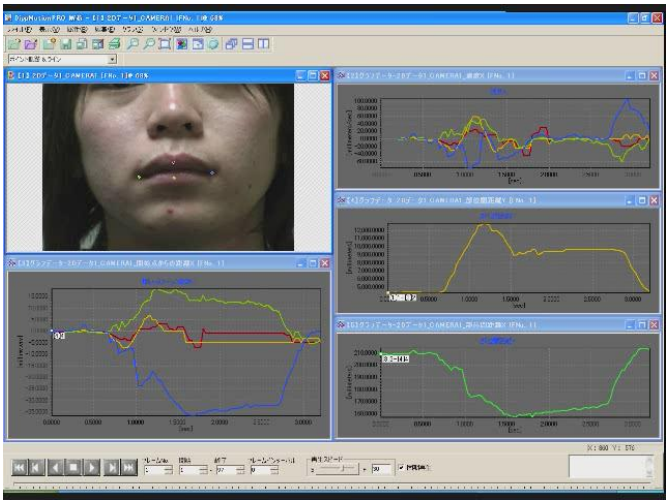


Figure 7 Example of motion analysis of four feature points when the lecture pronounces "u"

C. CG Animation Method

3D model with wireframe has to be created first as shown in Fig.8. Then rendering is followed by. Once basic 3D shape model is created, and then the target 3D models are created as shown in Fig.9.

After that, intermediate pictures are created with basic and the target 3D models as key frame method as shown in Fig.10.

During the process for creation of intermediate frames, chin movement has to be realistic. In this process, jaw bone is assumed as shown in Fig.11. Fig.11 (a) and (b) shows model derived lectures mouth and lip images when lecture closes and opens the mouth, respectively. On the other hands, Fig.11 (c) and (d) shows the assumed jaw bones when lecture close and open the mouth, respectively. The mouth can be opened when the jaw bone angle is increased.

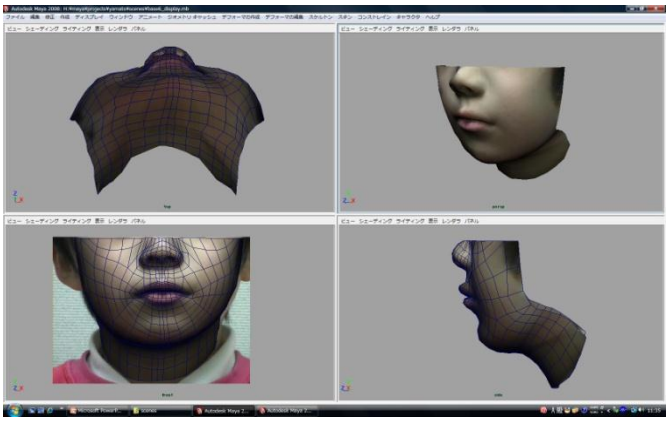


Figure 8 3D model with wireframe of mouth and lip

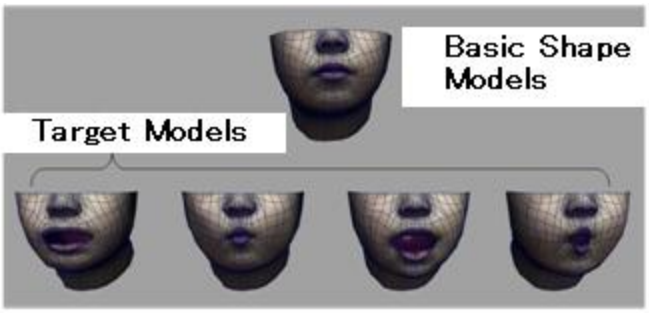


Figure 9 Target models of target 3D models are created from the basic 3D shape model.

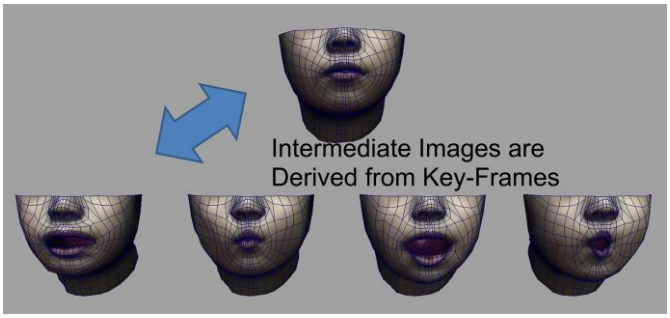
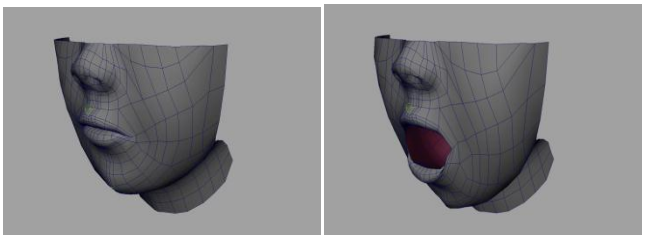
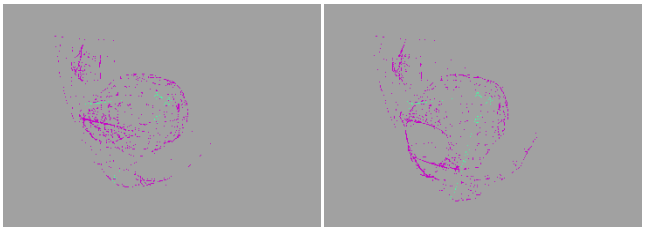


Figure 10 Intermediate pictures which are created with basic and the target 3D models as key frame method

Thus CG animated moving picture of mouth movements can be created as shown in Fig.4. All the feature points and picture for rendering are derived from the learners and lecturers so that CG animated moving picture is resemble to them.



(a)Close the mouth (b) Open the mouth



(c) Close the mouth (d) Open the mouth

Figure 11 Model derived lectures mouth and lip images together with assumed jaw bone when lecture closes and opens the mouth.

III. EXPERIMENTS

A. *Experiment Procedure*

8 kindergarten boys and girls (L1 to L8) whose age ranges from five to six are participated to the experiment. Pronunciation practice is mainly focused on improvement of pronunciation of vowels and consonants, /s/,/m/,/w/. Before and after pronunciation practice with the proposed system, 12 examiners identify their pronunciations with their voice only and with their moving picture only as well as with their voice and moving picture as shown in Fig.12. Thus identification ratio is evaluated together with mouth and lip shapes difference between before and after pronunciation practices. Pronunciation practice for vowels and consonants are conducted. Those are E1: vowels, E2: /a/-/sa/-/ma/-/ta/-/wa/, E3: /i/-/mi/, E4: /u/-/mu/, E5: /e/-/se/-/me/, E6: /o/-/so/-/mo/, respectively. All these pronunciations are Japanese.

Figure 12 Three types of pronunciation evaluations

B. *Experiment Results*

Table 1 shows experimental results of identification ratio for before and after the pronunciation practice.

TABLE 1 IDENTIFICATION RATIO FOR BEFORE AND AFTER THE PRONUNCIATION PRACTICES

	Moving Picture Only		Voice Only		Moving Picture and Voice	
Exercise	Before	After	Before	After	Before	After
E1	81%	87%	90%	94%	95%	98%
E2	70%	80%	87%	94%	92%	98%
E3	69%	73%	95%	95%	98%	99%
E4	70%	79%	93%	93%	98%	98%
E5	71%	78%	94%	99%	90%	99%
E6	64%	76%	92%	94%	91%	95%
Average	71%	80%	92%	95%	94%	98%

It is noticed that identification ratio for after pronunciation practice is improved by 3-9% from before pronunciation practice for all three cases of evaluation methods. Evaluation results with voice only show 3% improvement. This implies that their pronunciation is certainly improved.

Much specifically, pronunciation of /sa/ for L3 learner before pronunciation practice is 100% perfect while L8 learner has difficulty on pronunciation of /sa/ before pronunciation practice as shown in Table 2.

TABLE 2 IDENTIFICATION RATIO BEFORE PRONUNCIATION PRACTICE.

Learner	a	sa	ta	ma	wa	Unclear
L1	0.0%	87.5%	12.5%	0.0%	0.0%	0.0%
L2	12.5%	87.5%	0.0%	0.0%	0.0%	0.0%
L3	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
L4	0.0%	87.5%	12.5%	0.0%	0.0%	0.0%
L5	25.0%	62.5%	12.5%	0.0%	0.0%	0.0%
L6	0.0%	87.5%	0.0%	0.0%	0.0%	12.5%
L7	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
L8	12.5%	25.0%	37.5%	0.0%	0.0%	12.5%

In particular, pronunciation of /sa/ for L8 learner is used to confuse with /a/ (12.5%), /ta/ (37.5%), and unclear (12.5%) before pronunciation practice as shown in Table 3 (a). This situation is remarkably improved as shown in Table 3 (b). Identification ratio of pronunciation of /sa/ for L8 learner is changed from 25% to 100% perfect after the pronunciation practice.

TABLE 3 IDENTIFICATION RATIOS FOR BEFORE AND AFTER PRONUNCIATION PRACTICE OF PRONUNCIATION OF /SA/ FOR L8 LEARNER

(a) Before pronunciation practice

Sound (Sa)	a	sa	ta	ma	wa	Unclear
L8	12.5%	25.0%	37.5%	0.0%	0.0%	12.5%

(b) After pronunciation practice

Sound (Sa)	a	sa	ta	ma	wa	Unclear
L8	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%

Fig.13 shows one shot frame image of moving picture for mouth and lip when L8 learner pronounces /sa/ before and after the pronunciation practice. Although he could not open his mouth when he pronounces /sa/ before the pronunciation practice, he made almost perfect mouth shape for /sa/ pronunciation after the practice.

Another example for identification ratios for before and after pronunciation practice for L6 learner is shown in Table 4. L6 learner has difficulty on pronunciation of /a/ due to his mouth shape. Although identification ratio of /a/ is 100% before the practice when it is evaluated with voice only, it is 62.5% before the practice when it is evaluated with both voice and moving picture.

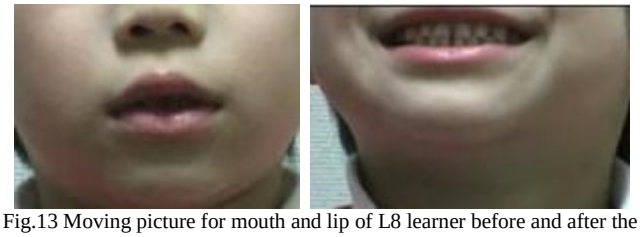


Fig.13 Moving picture for mouth and lip of L8 learner before and after the pronunciation practice of /sa/.

This implies that his mouth shape is resembled to that of /sa/, /wa/, and unclear even though his voice sound can be heard as /a/. It is improved to 100% perfect after the practice. Therefore, it may say that the practice is effective to improve not only voice but also mouth and lip shapes.

TABLE 4 IDENTIFICATION RATIOS FOR BEFORE AND AFTER PRONUNCIATION PRACTICE OF PRONUNCIATION /A/ FOR L6 LEARNER

(a) Voice only before practice

Sound (a)	a	sa	ta	ma	wa	Unclear
L6	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%

(b) Voice and moving picture before practice

Sound (a)	a	sa	ta	ma	wa	Unclear
L6	62.5%	12.5%	0.0%	0.0%	12.5%	12.5%

(c) Voice and moving picture after practice

Sound (a)	a	sa	ta	ma	wa	Unclear
L6	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%

His mouth and lip shapes before and after the pronunciation practice of /a/ are shown in Fig.14.

Figure 14 Mouth and lip shapes of L6 learner before and after the pronunciation practice of /a/

Another example is shown in Table 5. L3 learner has difficulty on pronunciation of /e/ due to his mouth and lip shapes. Although identification ratio of /e/ is 75% before the practice when it is evaluated with voice only, it is 62.5% before the practice when it is evaluated with both voice and moving picture. This implies that his mouth and lip shapes are resembled to that of /i/. Although it is improved to 87.5% after the practice when it is evaluated with voice only, it is improved to 100% perfect after the practice when it is evaluated with both of voice and moving picture. Therefore, it may say that the practice is effective to improve not only voice but also mouth and lip shapes.

His mouth and lip shapes before and after the pronunciation practice of /e/ are shown in Fig.15.

TABLE 5 IDENTIFICATION RATIOS FOR BEFORE AND AFTER PRONUNCIATION PRACTICE OF PRONUNCIATION /A/ FOR L6 LEARNER

(a) Voice only before practice

Sound (e)	a	i	u	e	O	Unclear
L3	0.0%	25.0%	0.0%	75.0%	0.0%	0.0%

(b) Voice and moving picture before practice

Sound (e)	a	i	u	e	o	Unclear
L3	0.0%	37.5%	0.0%	62.5%	0.0%	0.0%

(c) Voice only after practice

Sound (e)	a	i	u	e	o	Unclear
L3	0.0%	12.5%	0.0%	87.5%	0.0%	0.0%

(d) Voice and moving picture after practice

Sound (e)	a	i	u	e	o	Unclear
L3	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%

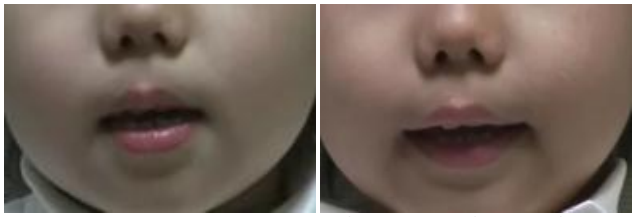


Figure 15 Mouth and lip shapes of L3 learner before and after the pronunciation practice of /e/

IV. CONCLUSION

Pronunciation practice system by means of mouth movement model based personalized CG animation is proposed. The system allows users to look at both users' mouth movement and model based CG animation of moving picture. Therefore, users practice pronunciation by looking at both moving pictures effectively. 8 infants examined pronunciation practices of vowels and consonants, in particular for /s/,/m/,/w/ by using the proposed system. Remarkable improvement (3-9%) on their pronunciations is confirmed.

ACKNOWLEDGMENT

Authors thank to Mr. Isao Narita and Mr. Hiroshi Kono of Faculty of Engineering, Kurume Institute of Technology for their great effort to experiments. Also authors thank to Professor Dr. Seio Oda of Fukuoka Institute of Technology, Junior College for his great supports and comments and suggestions to our research works.

REFERENCES

[1] Mariko Oda, Shun Ichinose, Seio Oda, Development of a Pronunciation Practice CAI System Based on Lip Reading Techniques for Deaf Children, Technical Report of the Institute of Electronics, Information, and Communication Engineers of Japan, vol.107, No.179 WIT2007-25(2007).

[2] Mariko Oda, Seio, Oda, and Kohei Arai, Effectiveness of an English /l/ -/r/ Pronunciation Practice CAI System Based on Lip Reading Techniques, Journal of Japan Society of Educational Technology, 26(2), pp.65－75(2002)

[3] McGurk H, Hearing lips and seeing Voices, Nature, 264, 746-748, 1976.
[4] Silsbee P., Computer lipreading for improved accuracy in automatic speech recognition, IEEE Trans. Speech & Audio Processing 4(5), 337-350, 1996
[5] Reiseberg D., Easy to hear, but hard to understand: A lipreading advantage with intact auditory stimuli, Hearing by Eye, 1987.
[6] Sumby W., Visual contributions to speech intelligibility in noise, J. Acoust. Soc. Amer. 26(2), 212-215, 1954

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science, and Technology of the University of Tokyo from 1974 to 1978 also was with National Space Development Agency of Japan (current JAXA) from 1979 to 1990.

During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He was appointed professor at Department of Information Science, Saga University in 1990. He was appointed councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was also appointed councilor of Saga University from 2002 and 2003 followed by an executive councilor of the Remote Sensing Society of Japan for 2003 to 2005. He is an adjunct professor of University of Arizona, USA since 1998. He also was appointed vice chairman of the Commission "A" of ICSU/COSPAR in 2008. He wrote 30 books and published 332 journal papers.

Mariko Oda, She received BS degree in 1992 and MS degree in 1994 from Saga University. She was subsequently appointed assistant professor at Department of Information and Network Technology, Kurume Institute of Technology.