I.J. Modern Education and Computer Science, **2015, 7, 1-7**
Published Online July 2015 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2015.07.01

English Pronunciation Practice Method with CG Animations Representing Mouth and Tongue Movements

Kohei Arai[1] and Mariko Oda[2]
[1]Graduate School of Science and Engineering, Saga University – Japan
[2]Contemporary Society Faculty, Hagoromo University – Japan
Email: arai@is.saga.ac.jp, moda@hagoromo.ac.jp

Abstract—Method for English pronunciation practice utilizing Computer Graphics: CG animation representing tongue movements together with mouse movements is proposed. Pronunciation practice system based on personalized 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. In addition to the mouth movement, tongue movement is represented by CG animation. Experimental results show 20 to 40 % improvement is confirmed by adding tongue movements for pronunciations of "s" and "th".

Index Terms—English pronunciations, CG animation, mouth movement, tongue movement.

I. Introduction

English pronunciation practice system with voice and CG animation is proposed by Yamada [1] together with spectral representation (formant) of voices [2]. Meanwhile, it is well known that mouth movement helps voice recognitions [3], [4]. In particular, strong pronunciation accent is well recognized by seeing mouth movements [5]. Lip reading is also well known as helping voice recognitions by looking at mouth movements [6].

Computer Aided Instruction: CAI based English pronunciation is getting more popular [7]. CAI based English pronunciation practice method with mouth movement representation with CG animation is proposed already [8].

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 [9]. 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 [10]. 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 [11]. Some of lip-reading methods and systems are proposed [3]-[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

[1] http://www.advancedmedia.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/

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.

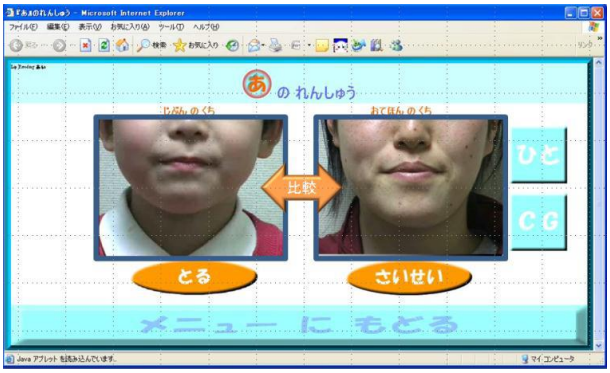


Fig.1. Example of display image of the previously proposed Lip Reading "AI".

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 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.

In order to create resemble mouth moving picture to the user in concern, correct mouth movements are extracted from moving picture by using Dipp-MotionPro2D3. 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.

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

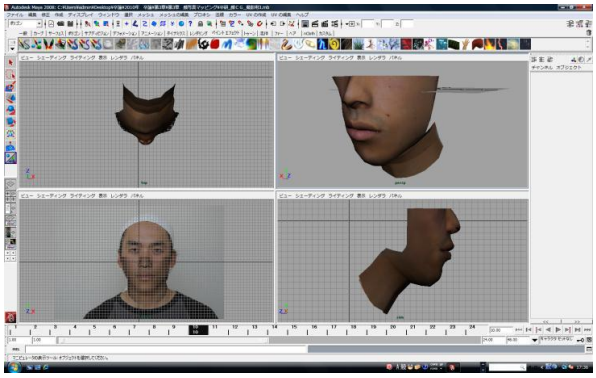


Fig.3. Example of Maya of CG animation software generated reference mouth and lip moving picture

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

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

(a)Close the mouth

(b)Open the mouth

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

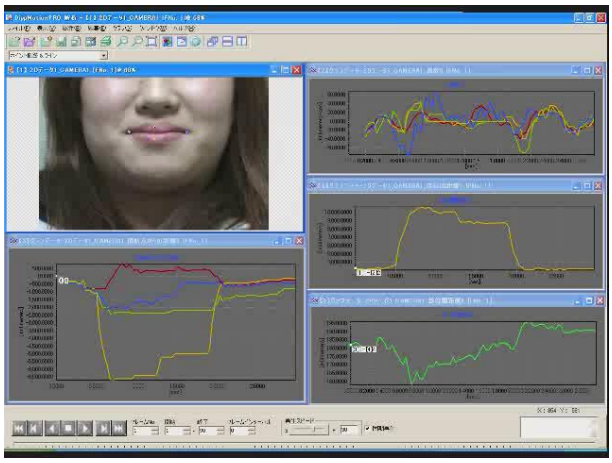


Fig.6. Example of motion analysis of four feature points when the lecture pronounces "a"

As the results of the experiments, it is found that the pronunciation practice system by means of mouth movement model based personalized CG animation is proposed.

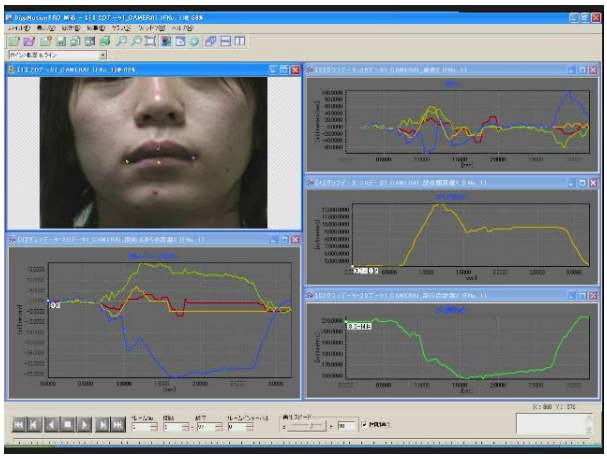


Fig.7. Example of motion analysis of four feature points when the lecture pronounces "u"

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.

C. CG Animation for inside mouth

The proposed English pronunciation practice system allows users to look not only outlook of mouth but also inside mouth movement for instruction. Inside mouth movements implies relations among teeth, tongue, and shape of inside mouth as shown in Fig.8.

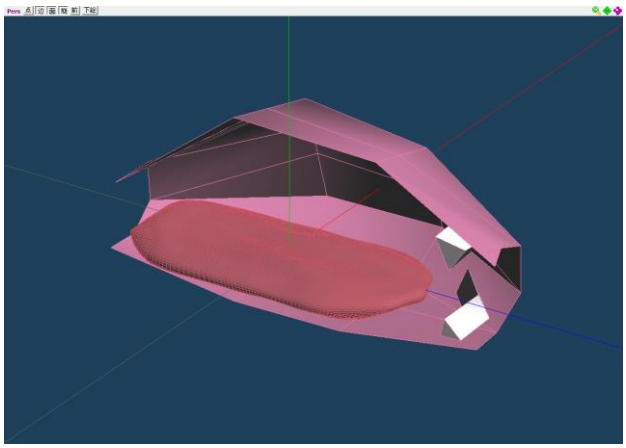


Fig.8. An example of the CG for inside mouth movements.

Also, aspect of the CG for mouth inside movements can be modified as shown in Fig.9. In the case of Fig.9, side view of the CG for inside mouth movement is shown as an example. Users can change the viewing direction of the CG image as user likes.

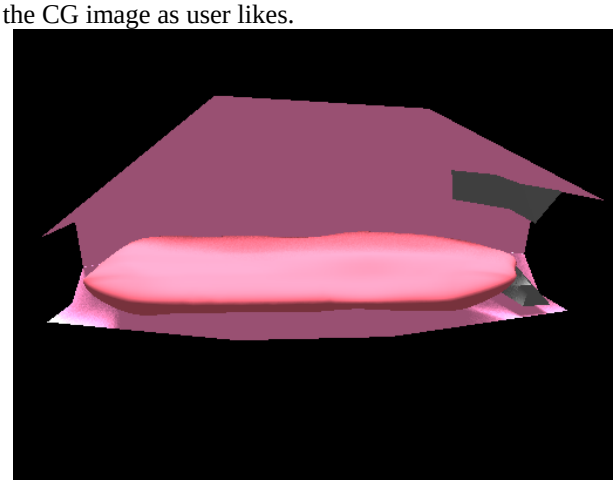


Fig.9. Example of side view of the CG image of inside mouth movements.

Thus users have instruction of inside mouth movements other than just outlook. Users can imagine their inside mouth movement. Then their pronunciation is improved.

D. Pronunciation Instructions with the CG animation for side view of inside of the mouth

Top menu is sh9wn in Fig. 10. “s”, “ʃ”, and “θ” can be practiced in this case.

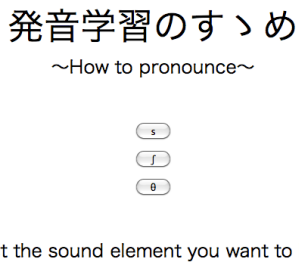


Fig.10. Top menu of the proposed English pronunciation practice system

If students selected one of these candidates of pronunciation practice, then instruction of the selected pronunciation is appeared. Not only the outlook of mouth movement shown in Fig. 11, but also CG animation of inside mouth movement (Fig. 12) is available to see for the selected pronunciation. Meanwhile, students' mouth movements can be also seen on the computer screen. By comparing between two mouth movements moving pictures, one is the template and the other one is students' mouth movement of moving pictures, pronunciation practices can be done efficiently and effectively.

On the other hand, students' also check their inside mouth movement through looking at the CG animation of the selected pronunciations.

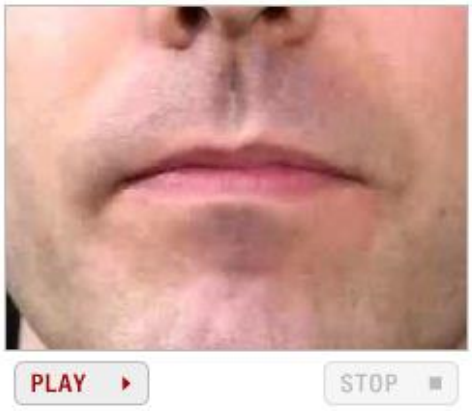


Fig.11. Outlook of mouth movement for pronunciation of "s"

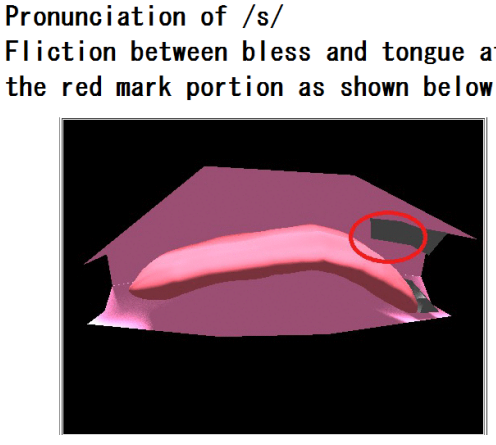


Fig.12. CG animation of inside mouth movement is available to see for the selected pronunciation

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.13. 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.

Fig. 12 Three types of pronunciation evaluations

In addition to the mouth movement animations, tongue movement animation is used for improvement of their pronunciations. 10 of university students participate the experiments.

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.

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 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%

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.

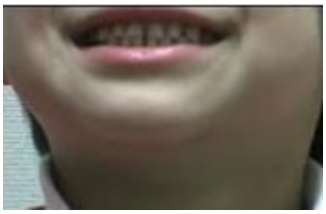


Fig.13. Moving picture for mouth and lip of L8 learner before and after the pronunciation practice of /sa/.

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. 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.

Fig.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.

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%

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

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

V. Conclusion

Method for English pronunciation practice utilizing CG animation representing tongue movements together with mouse movements is proposed. 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. In addition to the mouth movement, tongue movement is represented by CG animation. Expremental results show 20 to 40 % improvement is confirmed by adding tongue movements for pronunciations of “s” and “th”.

Acknowledgement

Authors would like to thank Mr. Sakai of Saga University for his efforts on the experiments conducted.

References

[1] R.Yamada, et al., Evaluation of the pronunciation of /r/, /l/ by Japanese: An effect of listening training, Proc. Of the Fall Conference of the Acoustic Society of Japan, 3, 3, 9, 397-398, 1995.

[2] T.Yamada, et al., How to improve English speaking scientifically, ATR Lab., BlueBacks, B-1263, 1999.

[3] W.Sumby and I.Pollali, Visual contribution to speech intelligibility in noise, Journal of Acoustic Society of America, 26, 2, 212-215, 1954.

[4] Q.Summerfield, Lip-reading and audio-visual speech recognition, Phil. Trans. Royal Society of London, 71-78, 1992.

[5] D. Reiseberg, et al., Easy to hear but hard to understand: A lip-reading advantage with impact auditory stimulus in hearing by eye, The Psychology of Lip Reading, Lowrence Erlbawin, 746-748, 1987.

[6] P.L. Silsbee, et al., Computer lip reading for improving accuracy in automatic speech recognition, IEEE Trans. Speech and Audio Processing, 4, 5, 337-350, 1996.

[7] M.Oda, S.Oda, K.Arai et al., Web based CAI system for /l/ and /r/ English pronunciation with voices and mouth movement pictures, Journal of Education System Society of Japan, 7, 6, 162-173, 2000.

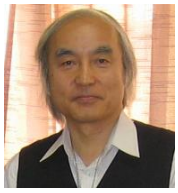
[8] K.Arai, S.Matsuda, M.Oda, English Pronunciation System Using Voice and Video Recognition Based on Optical Flow, Proc. Of the 2nd International Conference on Information Technology Based Higher Education and Training, 1-6, 2001.

[9] 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).

[10] 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 ー 75(2002).

[11] McGurk H, Hearing lips and seeing Voices, Nature, 264, 746-748, 1976.

Authors' Profiles



Kohei Arai 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 April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 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 moved to Saga University as a Professor in Department of Information Science in April 1990. He was a counselor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008. He wrote 33 books and published 500 journal papers.

Mariko Oda received BS and MS degrees of Computer Science from Saga University in 1992, 1994, respectively. She received PhD degree from Saga University in 2013. She was with Kurume Institute Technology from 1994 to 2014. She moved to Hagoromo International University in April 2014. Her major concern is education systems in particular for disable persons.

