

## The learning data management of the utterance learning system using lip movements recognition

- Learning data management using SQLite and feedback to a learner -

Yuko Hoshino<sup>\*</sup>, Tomoki Yamamura<sup>\*</sup>,  
Mitsuho Yamada<sup>\*</sup>

### **Abstract**

We previously proposed an utterance recognition system that uses the tracking and analysis of speakers' lip movements as words and sentences are uttered. Also, we compared the lip movements in the SQLite database with those of utterances produced by students learning the method. We accumulated lips movement history of the utterance learner in the conventional utterance learning system and added a function to manage the change of the lips movement of the learner. In this report, we describe development and the result of the system.

*Keywords:* utterance training method, lip movements, database, SQLite, faceAPI

### **1 Introduction**

A device and a technique using much utterance information are developed now. The sound recognition is used for input interfaces such as a search site or the email software a lot, because speech input system came to be usable easily by the spread of smartphone and tablets etc. The system of such a sound recognition continues developing, but use under the noisy environment has a problem. And it means that the biggest problem of this system is users must pronounce loudly and clearly. This may be the function that it is easy to use for the people who can pronounce like that, but cannot use it for a person that is a deaf-mute and the foreign language learner who cannot pronounce it. In our research, we hypothesized that the use of image reorganization may be used to help voice recognition systems; that is, if a system can distinguish what the user is saying by recognizing the movements of their lips. The recognition system came to be used without clear voice sound under the noisy situation. Such a recognition system may also be beneficial for utterance training for the deaf-mute and new foreign language learners. The system recognizes the movement of lips and can use this information for learners when they learn how to move lips when they pronounce words and sentences.

This technique is already used for the Japanese students who learn English [1]. A report which described voice recognition technology using multiple modes including sound and images by Hayamizu et al. in 1998 [2]. But only one image was mentioned in the report. It was investigated by Nagata et al. that the use of two microphones with directivity in an attempt to reduce the effects of ambient noise [3]. However, these two systems are not easily adopted.

---

<sup>\*</sup> Tokai University, Tokyo, Japan

An attractive and more easily used method was proposed by Yanagi and Yamada in 2008 [4]; it involves multimodal recognition based on the recognition of lip movements. At first, The lips are detected from the image. And then, the motion of the lips is analyzed. And finally the sound or word that has been spoken is recognized. The method uses five points placed on the mouth and chin of the speaker. When the speaker utters vowel sounds, words, phrases and sentences, the movements of these five markers are tracked in the time order, creating a movement history of the coordinates. In addition, the power spectrum of each point of the lip movement is calculated using Fourier transform. And then, the correlations between the power spectra and the spectra of an existing dictionary are computed. It is enabling the recognition of words and sentences based on the correlations.

This method has been applied for lip-reading education for deaf children [5], but it is not easy to imitate the movement of the lip and mouth, especially for beginners. We has already improved the system so that it can compare the learner's data on a display with the correct utterance movements [1]. In this particular application's system, the dictionary data is a compilation of CSV (comma-separated values) files of the correct movement data, and the user's data file can be recorded and compared. Therefore we have developed the application system to replace the CSV files, such as dictionary data and user utterance data, with SQLite database [6]. It reduced taking effort for the program development and easily updated dictionary data to embed database. The application system can accumulate data to the database, but the data is not enough useful for users in this system. Therefore, in this study, we have developed an application system to show the history information that based on the data that accumulated to the database to a learner.

## 2 Utterance Recognition Using The Lips Movement

The utterance recognition by the lips movement is a method to recognize utterance contents like lip-reading such as acquiring the movement of lips from a movie. Because this method is a recognition method without voice sounds, to resist a noise is a characteristic of this method. The utterance recognition by the lips movement repeats those are to extract lips part from a picture, to digitize the position of the form and characteristic point of the mouth and to recognize them. To extract of the lips part includes some methods such as using skin and lip color ingredient, using characteristic points of a face shadow (including zygomatic form and the form of the nose), and etc. The method using the characteristics point such as the shadows of the face are used to extract in this study.

### 2.1 Recognition of The Face

It is common for a method of the facial recognition to detect a skin color in a color image, and then to analyze the form and color [7][8][9][10]. The gray scale is also used for this. Processing speed of using the image of the gray scale is faster than that using the image of RGB and YUV, because the gray scale image has less information than the color image. However, this method may have a low probability to recognize, because it has small data. Viola et al. suggested technique to have with high probability and high-speed facial recognition using a gray scale image [11]. Important points of this technique are to perform high-speed facial recognition by the identification of the Cascade structure using high-speed operation of Haar-Like characteristic quantity by Integral-Image, and to learn using Ada-Boost. Their facial recognition method has become the base of much facial recognition algorithm. We have used a software program faceAPI (which is sold by Seeing Machines Corp., Canberra, Australia) which are developed using Viola's algorithm.

## 2.2 The Acquisition of The Movement of Lips

The study on lips movement is conducted for a long time. A method by Saito and others is using pattern matching of the lips outline and the quantity of trajectory movement [12]. Ishii et al. reported the method that subspace method using parameters such as the length of the lips outline, height, width and etc., and the other method that vectors which include some elements (the distance of the top and bottom of lips, the width distance of right and left of lips, and etc.) are arranged in chronological order to recognize an individual [13]. The method which is vector analysis of movements of lips from movie using optical flow is reported by Ootsuki et al. [14]. An interface apparatus which named the lips characteristic point extraction headset, to study the lips movement has been developed [15].

We have recognized lips movements using 5points which include the most moving points which are top, bottom, left, right of lips and the point of chin. The procedure of the utterance recognition by the lips movement is consists of three parts mainly. Firstly, the characteristic points of each lip are arranged in chronological order to make a movement history. Then, All movement history of each characteristic points are applied Fourier transform as expression (1) to demand power spectrums. Finally, the correlation of the power spectrum by expression (1) and the dictionary power spectrum (consisting of a word and the sentences) that have been uttered earlier and acquired is demanded like (2). If correlation is high, it is recognized the word that is uttered.

$$X_k = \sum_{n=0}^{N-1} X_n W^{kn} \quad (1)$$

X : power spectrum

W : movement history of characteristic points

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

r : correlation

$x_i$  : power spectrums x coordinate

$y_i$  : power spectrums y coordinate

$\bar{x}$  : average of power spectrums x coordinate

$\bar{y}$  : average of power spectrums y coordinate

## 3 The Way to Create The Dictionary Database

The dictionary data for the utterance training method such as described above was created using utterances in both native Japanese-speaking television announcers of Japanese and native English-speaking teachers of English. Each lip movement dictionary was made from movie image data recorded by a digital video recording system with a camera built-in PC, at 640\*480 (VGA) resolution. The software program faceAPI is embedded in the recording application software that we had developed for the recognition of the position of the speaker's face, eyes and mouth as the speaker utters words and sentences. This recording application has already been included in our utterance training method system. A sample data acquisition screen is shown as Figure 1. A speaker sits in front of the camera built-in PC, utters a word toward the camera. The height of the camera is the same or slightly lower degree of face roughly (height to put the PC on the desk).

We used this application software to record data for a dictionary of Japanese words, phrases and sentences. The speakers were two television announcers (one male, one female) who are native



Figure 1: The display menu at the time of data acquisition.

Japanese speakers, from the NHK (Japan Broadcasting Corp.) Communication Training Institute. The application software was used to record sentences from two Japanese textbooks: "[Easy training for a good voice]" [16] and "Tanpopo [dandelion]" [17], a primary school textbook. And we also added many Japanese words that are frequently used in daily life, such as the names of railway stations.

The application software was also used to record English-language dictionary data. The speakers were one male and one female English teachers who were native speakers of English. We recorded 12 sentences

consisting of 65 phrases which were uttered by the speakers. These sentences are included in an NHK language module named "Three-Month Topic English Conversation" for learners of English [18]. Our application software, which we developed previously, maintains all of these dictionary data and the learner data in CSV text files. We selected CSV file, because it is not difficult that the CSV text files are managed when the volume of words and sentences is not high. However, the number of the files increases in proportion to the increase of the kind of the dictionary. In addition, it is overwritten that the CSV file recorded the utterance data of the users (learners).

## 4 Utterance Recognition with Our Application System Using CSV

We compared the lip movement history of a learner and the ability of the CSV dictionary data files to recognize an uttered word, using the application system described above. For the first experiment, some station names of the Odakyu Railway in Japan as the dictionary data were used. In the experiment, the learner uses the recording application, which faceAPI and the dictionary data have been embed in. And they utter a word (in this case, a station name) toward the web camera with the built-in PC. The movement of the five points of the learner's lips and mouth were outputted to a file by the application. And then, sample data file of the left point were obtained and outputted from data file above to calculate the power spectrum. This application obtains the power spectrums from the sample data by Fourier transform and outputs the power spectrums to other CSV file. The same process was done for the other points of lips (i.e., left, right, top, bottom), all of which went to a CSV file. Next, the learner's power spectrum was compared with the dictionary's spectrum. Figure 2 shows the structure of the dictionary which named "Odakyu" for the word some station names.

The power spectrum files of all lips characteristics points must be put in the subfolders of the word name to add a dictionary. Furthermore, these subfolders must be put under the folder of the dictionary name. When learner users operate the application, the procedures of the recognition are as follows. To record the lips movement history is started when "Start recording" button is pushed, and it is stopped when "END" button is pushed. In this time, the history data of lips movement is written in a CSV file. The left-mouth spectrum data is read by the system when the learner pushes the 'check' button to call up the utterance recognition function. Then, the application reads the left-mouth power spectrum data of the first word in the dictionary. And then, the

correlation between the power spectrum of the learner and that of the dictionary is calculated and outputted. Other the learner's right, top and bottom point markers are done same.

The application compares the learner's power spectrum and the dictionary's power spectrum, and writes in correlation at a different CSV file. The processing is repeated for all of the words in the dictionary.

When the value of the correlation is big enough, it is admitted that the utterance data resemble dictionary data. When the value of the correlation is big enough, the utterance data is accepted as similar dictionary data. All the marker points' correlation values of user data and dictionary data are obtained, and they are added up. The word in the dictionary which has the biggest total value is judged with an uttered word. Based on this result and similar outcomes, it is found that the reorganization of utterances from lip movements is possible at the level of the single words dictionary.

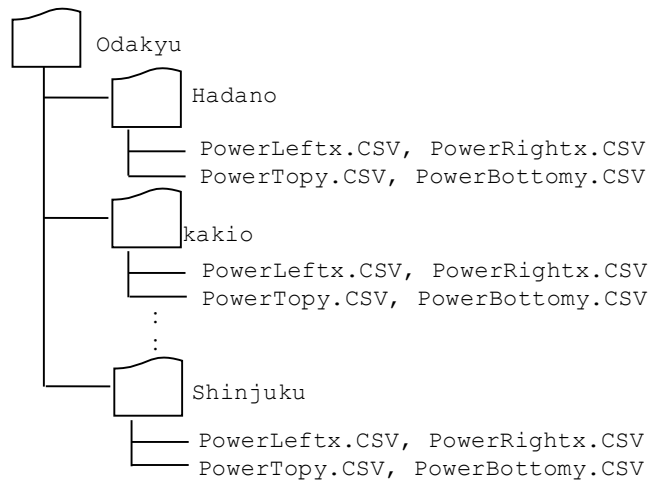


Figure2: Structure of the dictionary "Odakyu" version

## 5 Programming with CSV Files

### 5.1 Former Conditions Using CSV Files

We created a folder for every dictionary and several subfolders in each dictionary folder for the words enrolled. The group of CSV files which had been made from data uttered by announcers and teachers were put into a subfolder of each word in the dictionary folder. A folder had been made for every learner too, and his or her utterance data are saved in the subfolder as a CSV file group.

The kinds of dictionary data are increased by adding CSV file groups and folders. When the number of dictionary and words included in there increased especially, the problem occurs easily. If they overwrite any files with the wrong word's data or put the file to wrong folder, users may not notice. When a word is not recognized, another problem may occur. A cause of this problem is the dictionary data includes just numerical value of the power spectrum. Therefore the dictionary data is used to calculate the correlation even if one of the power spectrum files of a word is replaced to one of other words. This may be solved by adding ID numbers to each dictionary and word when they are saved. However, CSV files are easy rewritten without special software. And as the number of files increases, the number of times that files are input or output also increases, which could cause delay in processing.

### 5.2 To Accumulate and Analyze Learner's Data

Our utterance training application which has been developed already, stores the latest lip movement history, and the application can reproduce the lip movement for the pronunciation of the vowel sound by line drawing on the system's screen display. The system has been able to provide advice to the learner such as "put the lower lip higher" by comparing the lip movement of the dictionary with the learner's one. However, this advice picture is provided only for the latest utterance.

If the learner repeats an utterance continuously, the system should provide confirmation of the learner's progress by using the learner's prior stored utterance data. The training data is not only used for this purpose; the data should also be analyzed and used to help provide specific training feedback, such as "weak in the pronunciation of 'A'," "weak in a certain field," and "tendency to lower the lower lip." All prior utterance data can be stored in CSV files, but it is difficult that all data in CSV files are managed by the application. The data for a specific purpose (e.g. vowel sound utterance is feed backed) must be selected from among the stored data. When using CSV files, the data processing is complicated. We have adopted Database Management System (DBMS) to solve this problem.

### **5.3 Environment of Using The Application**

Our system currently works on Windows as the platform, and the system is considered with using with smartphones and tablets (e.g., for Android and iOS). Most of smartphones and tablets have both their main memory and external memory (such as an SD card) for media storage. Permission may be necessary to use of the external memory depending on a smartphone; when the developer puts the some data in external memory or an internal storage device using a program, to put that may thus not be possible. In addition, as noted above, when a new word is registered, the number of files to be written increases each time when a dictionary is updated.

## **6 To Embed The Database Management System in The Application**

In light of the issues described above, we embed a database management system to the application.

### **6.1 Choice of The Database Management System**

The former dictionary is aggregate of some words, and the word is a group of plural CSV files which consist of the numerical value of the utterance power spectrum. Each dictionary is given a dictionary name, such as 'Odakyu', the name of a vowel sound or a sentence group such as book title. The Relational Database Management System (RDBMS) is desirable, because these data require accuracy and consistency. A very much large DBMS has not been able to be used with a smartphone or tablet. When a database exists on a server computer, quantity of traffic between a smartphone and the server is increased. In addition, when communication situation of the smartphone is bad, they cannot connect to the server. It would be better if the recognition method could be used without network communication. Therefore, a DBMS that can be operated in the smartphone without network communication is as small as possible to embed. Also, when the dictionary is updated, a few numbers of files are better. If database has been used, the number of file can be only one. Based on these reasons, we have chosen the public-domain database SQLite.

The SQLite is a type of RDBMS which uses one database as one file, similar to DBMS which is Microsoft Access. When developers develop an application with SQLite, it is used as a function call of the API library. Moreover, the stored files are available in the differing application software programs among the different operating systems and architectures, because the SQLite database uses storage that does not depend on the byte order. Varieties of applications use and operate SQLite in the native code, and its transaction speed is comparable to those of the most commonly used databases.

## 6.2 Embed SQLite Database to The Application

We have incorporated SQLite in an utterance training application for Windows as follows. At first, the dynamic link library of SQLite and header files of the computer programming language C had been downloaded. And then, it had been added to call DLL and the operation schema for the database as a substitute for the input and output CSV files into the program. The Meta data of database file is made using the general SQL sentence (i.e., 'Create Database,' 'Create Table') by a program or a command-line tool. The processing times of the database have been compared with those of the CSV files. The processing times have not shown major differences very much, when the quantity of data is not too large. The quantity of coding has been also compared. It has been necessary to use every data file to create a folder path and to process open files when CSV files are used. An example program for a CSV file is shown as Figure 3. With the use of the SQLite database, the programming code to connect the dictionary data is needed only one time. Moreover, the dictionary has been able to gather into one database file. An example program to be connected to SQLite database is shown as figure 4.

It has been noted that the problems which accompany with the maintenance of the dictionary and learner data in the application software, and the SQLite database has been incorporated as a means to address these problems.

```
((CEdit*)GetDlgItem(IDC_EDIT1)->GetWindowText(FolderPath);
paths = (const char*)(FolderPath + "\\full2.CSV");
fp = _tfopen(paths, _T("r"));
```

Figure 3: Sample program for one CSV file.

```
sqlite3_open("dic.db", &dicDB);
```

Figure 4: Sample program for all dictionary data in the SQLite.

## 7 Learning Data Management to Feedback

The application mentioned above was released from agony of the file management by using a database. This application is used repeatedly to practice on utterance, because a user of this is an utterance learner. The conventional application overwrote the users' data, namely history data of the learner, whenever it is used. Thus, it could feed back that only information about the latest utterance.

It is important to practice repeatedly for utterance learning, and it is useful that the information learned in the past is read whenever a learner practices. If a learner can confirm that he or she can improve the utterance or cannot do that, it leads to maintenance of the motivation. For such a reason, we have decided to add the functions which are accumulation of the history of user utterance data and its feedback. In this report, the function is developed about words such as railway stations name.

### 7.1 A Review of The Table Columns of Dictionary Database

In an existing system, the data of users' lips characteristic points are saved in chronological order. This system pulls out the necessary number of the sample for Fourier transform from these sequential histories, then makes power spectrum by Fourier transform, and then stores power spectrum data to a table. It is necessary to record "which word of which dictionary was uttered? When it was uttered?" to follow a history.

Firstly, we thought that we registered all dictionaries, which included plural words and coped

only with a word, with one database. In this case, dictionary master table included dictionary ID and a dictionary name, word master included word ID registered with each dictionary and a word name, and a table for the power spectrum of each word (dictionary ID and the word ID were included in this, too) were necessary. However, the number of the records becomes enormous. When all words and power spectrum were stored on one table in this way, malfunction, such as the insufficiency of CPU power to process the data with a smartphone and the tablet, happens. In addition, a big database entirely must be transmitted and received when a dictionary was added and updated it. Therefore, we have decided to make the database for the dictionary master which registered some dictionary names. In addition, we thought that the choice of the dictionary to use could be facilitated by dictionary master database including dictionary ID, a dictionary name and the kind (e.g., vowel sound, word, the sentence) of the dictionary. Each dictionary (e.g., "Odakyu") is assumed different database file one by one, and made database file, and then registered by ID and name with a dictionary master. The application reads dictionary data based on this dictionary name, and then compare user's data with dictionary's data. The number of the database files increase by distributing every dictionary, but increasing the number is less than the way which plural CSV files have been needed for every word. The number of database files is around a quarter of the number of CSV files. Users cannot overwrite and change unjustly that database. The folder constitution of dictionary database and the example of the table are shown in Table 1.

Table 1: the example of the tables for the dictionary

Table 1-1: Dictionary Master

DicID	DicName
001	Odakyu
002	Tampopo
:	:

Table 1-2: Word Master for the dictionary "Odakyu"

WordID	Word
001	Hadano
002	Kakio
:	:

Table 1-3: Power Spectrum table of the word "Kakio" in dictionary "Odakyu"

left	right	top	bottom
4.1948	0.0036	10.9897	436.3297
11.2565	0.0057	4.7877	2.9588
6.2338	0.0004	4.6497	55.9463
0.4302	0.0021	0.4758	1.1813
0.1987	0.0031	0.4919	0.2668
0.0722	0.0006	0.7744	1.2403
0.0375	0.001	0.2132	0.0455
0.0054	0.0001	0.0105	0.2448
:	:	:	:

## 7.2 A Review of The Table Columns of User Database

User's database file which embeds in the existing application includes three tables. The purposes of each table are to store the chronological order data of the lips characteristic point of the utterance, the sample data which are necessary for Fourier transform and power spectrum for the comparison with the dictionary. The folder constitution of users' database, an example of the existing table constitution include in user database is shown in Table 2.

Only power spectrum should be saved for the feedback to a user because only power spectrum is used to recognize the utterance word and sentence in comparison with the dictionary. Therefore, the information about date and utterance; when a word is uttered, which word of which dictionary is uttered; and the power spectrums are necessary as data for the learning history preservation. In addition, it has been decided that user's database is divided into each user's database files like a dictionary. The number of the user database files also increases, but increasing the number is less than the way which plural CSV files have been needed for every users. The sum of each user's database files are around a quarter of the number of CSV files. Users cannot



overwrite and change unjustly those databases, too. In this way, DB (it stores a user ID and a DB name for the user) is prepared to record who is registered as a user, and then the database of each user's utterance practice history is made, like dictionary database. The utterance exercise history database for each user has two kinds of tables. One of them is the table which consists of exercise date-time, a list of exercised dictionary ID and word ID. Others are also table which named practice date-time include the power spectrum of each lips characteristic point. The table of the power spectrum is generated one table per every exercise date-time.

Table 2: The example of the tables for a user

Table 2-1: User Master

UserID	UserName
001	Hoshino
002	Yamada
:	:

Table 2-2: Utterance Practice History for the user ID "001"(db file name is "001.db")

PracticeDateTime	DictionaryID	WordID
20150323135354	001	002
20150402142432	001	002
20150403102321	001	001
:	:	:

Table 2-3: One of Power Spectrum table (named " 20150323135354") of the word "Kakio"

DicID	WordID	UtterOrder	left	right	top	bottom
1	2	0	26.4751	18.6256	132.0735	528.5968
1	2	1	2.2964	0.5963	5.4305	1.8951
1	2	2	1.1791	1.1241	39.2681	4.9472
1	2	3	0.8152	1.0721	12.0876	13.9748
:	:	:	:	:	:	:

When a learner wants to read the practice record that has been uttered, the learner chooses one's name among user master to get user's database ID and name. And then the application connect user's database to get some power spectrum of utterance history data in the past. When the learner choose a word among a list of exercised word table, some data which is date-time of the practiced word are got. Then, from the table of the name same as the date-time included in the provided practice date-time group, the power spectrum of each lips characteristic point are read to process. And then, these power spectrums are processed like the utterance recognition. By displaying chronological order of this processing result, a learner can show the exercise result history and know the growth of utterance in a few minutes. The calculation time is slightly changed by the length of words and sentences.

## 8 Conclusion

We have developed utterance recognition application system using the lips movement. And we have embedded SQLite database to decrease the coding quantity in the application. It has improved the application not to change and overwrite unjustly of data. Furthermore, we have changed the constitution of the database and added the function to confirm the practice history for an utterance learner. The existing system has a function to provide lips movement of the utterance by line drawing, but this function does not support stored exercise history information. Therefore, we are going to develop this function. In addition, we would like to develop this application for other mobile OS (e.g. android, iOS) , because this system supports only Windows PC and Windows tablet. The faceAPI is with a library for Windows basically, but the Android version after 4.0 has a library including the FaceDetector class to search a face. If a system rec-

ognize a face using this android library and combine it with an extraction technology of lips characteristic point by a color and the form, we think that the utterance recognition may be done without faceAPI. We investigate this method and want to inspect it in a future.

## Acknowledgement

We thank Mr. Atsuyuki Takahashi, the NHK announcers and the English teachers who cooperated in the creation of the utterance database. This work was supported by JSPS KAKENHI Grant No. 25330418.

## References

- [1] Wakamatsu E, Hoshino Y, Yamada M. Proposal for an utterance training method based on lip movements. IMQA2014 (The Seventh International Workshop on Image Media Quality and its Applications), September 2014, Chiba, Japan. pp. 44-47.
- [2] Hayamizu S, Takezawa T. Trends in research on multimodal information integration system. Trans Jpn Soc Artif Intell. 2006;13(2):206-211, Mar. 1998. (in Japanese)
- [3] Nagata Y, Fujioka T, Abe M. Target signal detection system using two directional microphones. IEICE Trans A. 2000;J83-A(12):1445-1454. (in Japanese)
- [4] Yanagi T, Yamada M. Proposal and verification of lip movement model about utterance recognition interface without voice. HIS2008 (8th International Conference on Hybrid Intelligent Systems), Sept. 2008, p. 2422. (in Japanese)
- [5] Oda M, Ichinose S, Oda S. Development of a pronunciation practice CAI system based on lip reading techniques for deaf children. Technical Report of IEICE. 2007;107(179):53-58. (in Japanese)
- [6] Y.Hoshino, E.Wakamatsu, M.Yamada, "The Utility of the SQLite Database for An Utterance Training Method Based On Lip Movements", Information Engineering Express International Institute of Applied InformatiCSVol. 1, No. 1, pp.1-9, 2015
- [7] K.Sung, T.Poggio, "Example - based learning for view - based face detection", IEEE Patt. Anal. Mach. Intell., volume 20, pp.39-51, 1998.
- [8] H.Rowley, S.Baluja, T.kanade, "Neural network - based face detection", IEEE Patt. Anal. Mach. Intell., volume 20, pp.22-38, 1998.
- [9] H.Schneiderman, T.kanade, "A statistical method for 3D object detection applied to face and cars", International Conference on Computer Vision, 2000.
- [10] D.Roth, M.Yang, N.Ahuja, "A snowbased face detector", Neural Information Processing12, 2000.
- [11] Paul Viola, Mihael Jones, "Robust Real-Time Object Detection", Second International Workshop on Statistical and Computational Theories of Vision - Modeling, Learning, Computing and Sampling, pp.1-25, Vancouver, 2001 .

- [12] S.Saito, R.Konishi, "Lip Reading based on Trajectory Feature of Lip and Mouth Cavity Regions", the Institute of Electronics, Information and Communication Engineers Information and Systems Society, Human Communication, J90-D, 4, pp.1105-1114, 2007 (in Japanese)
- [13] M.Ishii, K.Sato, M.Nishida, Y.Kageyama, "Feature Extraction and Analysis for Lip Reading in Continuous Image", The transactions of the Institute of Electrical Engineers of Japan. D, A publication of Industry Applications Society 119(4), pp.465-472, 1999. (in Japanese)
- [14] T.Otsuki, T.Ohtomo, "Automatic Lipreading of Station Names Using Optical Flow and HMM", Institute of Electronics, Information, and Communication Engineers, Technical Report of IEICE PRMU, Vol.102, No.471, pp.25-30, 2002 (in Japanese)
- [15] T.Yoshida, K Fukushi, S.Hangai, "Development of lip movement sensor and its application to spoken word recognition", Institute of Electronics, Information, and Communication Engineers, FIT2006, pp.125-126, 2006. (in Japanese) .
- [16] Fukushima M. Easy training for a good voice. Seibido Publishing, Tokyo. 2006. in Japanese
- [17] Hirayama K. Dandelion. First Volume of Revised New National Language Two, Tokyo Shoseki, Jan. 1985. (in Japanese)
- [18] Three-month topic English-conversation. NHK publication, 2009. (in Japanese)